VRIJE
UNIVERSITEIT
BRUSSEL

Master thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Applied Sciences and Engineering: Computer Science

EFFICIENTLY LEARNING OPTIMAL VACCINE ALLOCATION STRATEGIES FOR THE MITIGATION OF DENGUE EPIDEMICS

A Multi-objective Multi-armed Bandit based Approach

Lennert Saerens

2023-2024

Promotor: Prof. Dr. Pieter Libin

Advisor: Bram Silue

Science and Bio-Engineering Sciences

VRIJE
UNIVERSITEIT
BRUSSEL

Proefschrift ingediend met het oog op het behalen van de graad van Master of Science in de ingenieurswetenschappen: computerwetenschappen

EFFICIËNT LEREN VAN OPTIMALE VACCINATIESTRATEGIEËN VOOR HET ONDERDRUKKEN VAN DENGUE-EPIDEMIEËN

Een aanpak gebaseerd op Multi-objective Multi-armed Bandits

Lennert Saerens

2023-2024

Promotor: Prof. Dr. Pieter Libin

Advisor: Bram Silue

Wetenschappen en Bio-ingenieurswetenschappen

Abstract

Dengue, a mosquito-borne viral disease, poses a significant global health threat, particularly in tropical and subtropical regions. With an increasing incidence and geographic spread, effective vaccination strategies are crucial for mitigating its impact. This dissertation explores the use of multi-objective multi-armed bandit (MOMAB) algorithms to identify optimal vaccine allocation strategies for the mitigation of dengue epidemics, balancing medical efficacy and monetary costs.

The research investigates whether MOMAB algorithms can efficiently pinpoint a subset of optimal vaccination strategies based on stochastic simulations. By extending the 2009 Recker et al. dengue model to incorporate vaccination strategies and age heterogeneity, we simulated the effects of 53 different strategies. The simulations included Gaussian noise to reflect real-world unpredictability, aligning with the stochastic reward functions used by MOMAB algorithms.

We adapted several MOMAB algorithms for Pareto front identification (PFI), and also propose a completely novel Top-two Pareto Front Thompson Sampling (TTPFTS) algorithm for the PFI setting. To evaluate the quality of the recommendations made by the considered algorithms, we developed three metrics: the Bernoulli metric, the Jaccard similarity metric, and the Hypervolume metric. Testing across 100 experimental repetitions with a limited budget of 30,000 arm pulls revealed that the Pareto UCB1, TTPFTS, and Pareto Thompson Sampling algorithms consistently performed excellently in terms of efficiency and stability, drastically outperforming the currently used Uniform Sampling method.

The findings demonstrate that PFI MOMAB algorithms are effective in identifying optimal vaccination strategies for dengue mitigation in a sample-efficient manner. This research contributes to the optimization of vaccination programs, providing a robust decision-making framework for public health officials facing the challenge of dengue epidemics. The study underscores the potential of MOMAB algorithms to enhance strategic deployment of vaccines, ultimately improving disease management and control.

Contents

Part I

Introduction

Chapter 1

Dengue

1.1 Current global situation

Dengue is a major disease transmitted by mosquitoes, impacting tropical and subtropical regions worldwide. It poses a serious public health threat, with 3.6 billion people at risk and over half a million individuals requiring hospitalization annually [24].

The incidence of dengue has increased 30-fold over the last 50 years, with the virus and its vectors expanding geographically [24]. Specifically in the last year, a notable surge was observed in the Americas. By April 2024, this region had already surpassed seven million cases, exceeding the previous annual peak of 4.6 million cases in 2023 [192]. Current estimates indicate that up to one third of the worlds population could be at risk of acquiring the disease [32, 191].

As of April 30, 2024, the World Health Organization (WHO) has reported over 7.6 million cases of dengue globally this year, with 3.4 million confirmed cases, more than 16,000 severe cases, and over 3,000 deaths [192]. Currently, active dengue transmission has been documented in 90 countries for 2024, although not all cases are formally reported. Many endemic countries lack robust detection and reporting systems, resulting in likely underestimations of the global burden of dengue [192]. To enhance global surveillance and monitor disease trends, the WHO has launched a comprehensive dengue surveillance system with monthly reporting across all regions, accessible through a new live dashboard[1].

1.2 Transmission and medical implications

The transmission of dengue is determined by a series of complex interactions between the virus, the mosquito vector, the human host, and environmental factors. The dengue virus (DENV) that causes the disease is an arbovirus for which there exist four related but distinct genetic variants called serotypes (DENV-1 to DENV-4) [5, 177, 36, 181]. The virus is a mosquito-borne, positive single-stranded RNA virus of the *Flaviviridae* family (genus *Flavivirus*). The virus is mainly spread by the *Aedes* mosquito, a mosquito that is active during the day and reproduces in stagnant water, such as in water containers. These mosquitoes have adapted

[1] `https://worldhealthorg.shinyapps.io/dengue_global/`

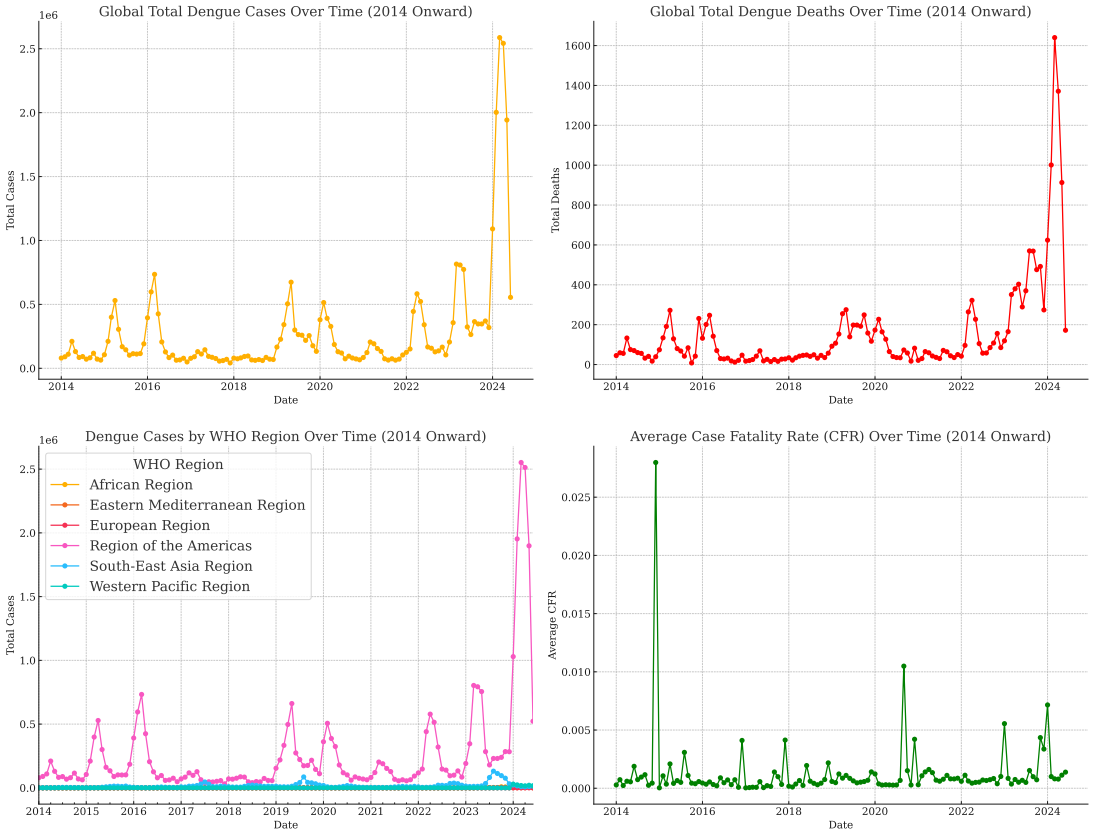


Figure 1.1: Number of dengue cases (by region), deaths as a result of dengue infection, and average case fatality rate per year between 2014 and 2024. Data retrieved from the WHO at `https://worldhealthorg.shinyapps.io/dengue_global/`.

to urban environments and are highly efficient in transmitting dengue virus [17]. Urbanization, globalization, and lack of effective mosquito control have led to increased dengue epidemics [120].

Socio-economic variables, including education level, housing conditions, and urban infrastructure, are associated with increased dengue risk [108]. Environmental factors such as temperature, precipitation, and mosquito breeding sites also play a crucial role in dengue transmission dynamics [108].

Dengue infection presents itself in various clinical forms, ranging from asymptomatic to mild disease [74, 180]. It is typically characterized by an acute feverish viral illness, often accompanied by symptoms such as headaches, bone or joint pain, and muscle discomfort [191, 74, 180]. In more severe cases, it can lead to a potentially life-threatening state [191, 74, 180], or dengue shock syndrome, marked by intense internal bleeding, and a significant decrease in platelet count [180]. Typically, primary dengue infection is less likely to cause severe symptoms. However, a subsequent infection with a different serotype significantly increases the risk of progressing to a more severe form of the disease [165, 161, 156, 77]. People who have been infected with one serotype of the virus acquire a lifelong immunity against that specific homologous serotype.

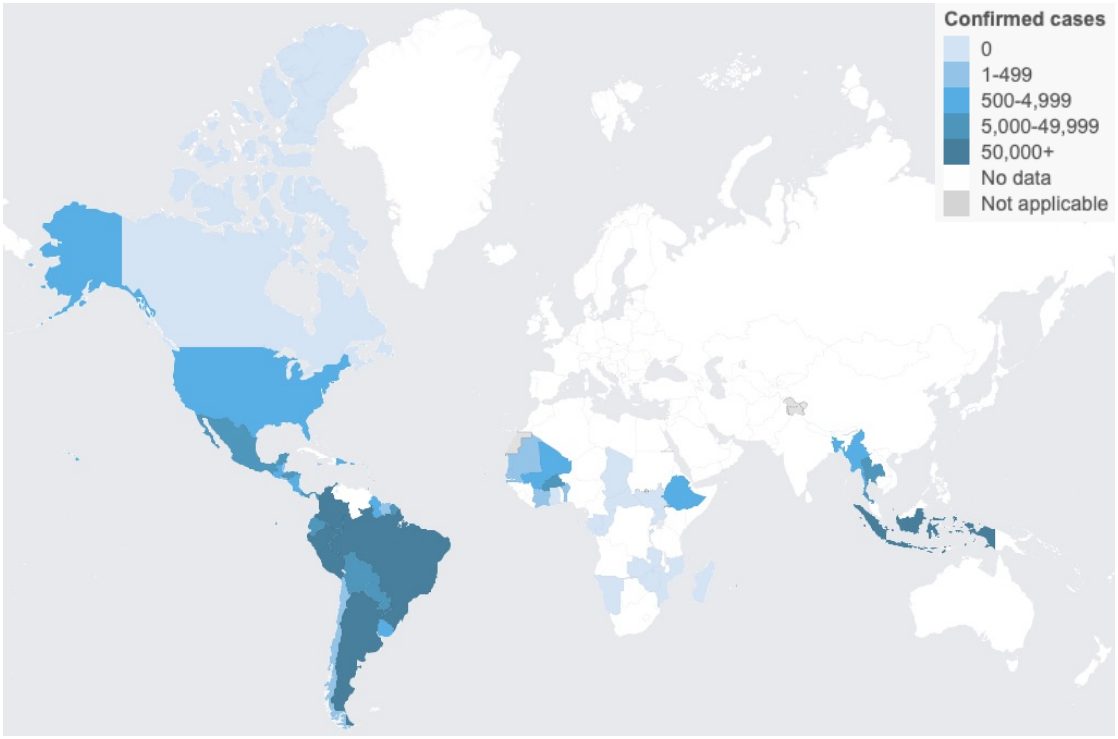


Figure 1.2: Global dengue situation: Confirmed cases between January and July of 2024 per geographical region.

However, the immunity they develop against other heterologous serotypes is only temporary [180]. As this temporary cross-protection diminishes, individuals encountering a secondary infection with a different dengue virus serotype face an increased risk of developing severe illness because of the antibody-dependent enhancement (ADE) phenomenon [165, 159, 150, 44, 149, 80, 75, 48]. The ADE theory suggests that when a secondary dengue infection occurs with a different serotype, it is identified by the antibodies created during the initial infection. However, instead of neutralizing this new strain, the existing antibodies inadvertently assist in exacerbating the infection. This phenomenon, which leads to an increased severity of the disease, is used to explain the cause of more severe dengue cases [165, 31, 76]. Numerous studies have shown that this effect induces higher viral loads in patients, increasing infectiousness [174, 99, 73, 178].

1.3 Treatment and vaccination

Dengue infection lacks a targeted treatment [192]. Supportive care suffices for uncomplicated cases of dengue, but severe cases necessitate hospital admission. Dengue diagnosis and management are challenged by overlapping clinical features with other diseases, such as COVID-19 [172] and other influenza-like illnesses.

Given the unique challenges posed by dengue, the development of vaccines is geared towards creating a tetravalent vaccine. This vaccine aims to offer lasting protection against all four dengue virus serotypes. The development of a dengue vaccine that is safe, efficacious, and

affordable, and that protects against all four strains would be a major breakthrough in disease control. Such a vaccine could play a crucial role in decreasing the transmission of the disease and reducing mortality rates [5]. Numerous potential tetravalent vaccines are currently in different phases of development [49, 184]. Two tetravalent dengue vaccines have successfully finished phase 3 clinical trials: Dengvaxia [39, 176, 78], created by Sanofi Pasteur, which is now authorized in over 20 countries, and DENVax [28, 142], developed by Takeda Pharmaceutical Company.

In April 2016, the World Health Organization's Strategic Advisory Group of Experts (SAGE) on Immunization advised the administration of the Dengvaxia vaccine for individuals aged between 9 and 45 years in areas with high disease prevalence. This recommendation was based on a mathematical modeling study assessing the vaccine's impact [5, 188, 189]. Mass vaccination campaigns were launched in the Philippines and Brazil, where around 1 million children and adolescents received the Dengvaxia vaccine without testing for prior infection. It was observed that Dengvaxia led to an increased incidence of severe dengue cases requiring hospitalization in children who had not previously been exposed to the virus (seronegative) [78], compared to similar children who did not receive the vaccine. The risks associated with administering Dengvaxia have been widely debated and analyzed [82, 9, 81, 10, 13]. After evaluating long-term safety data [164], the WHO issued a revised recommendation [190]. This new guideline suggests conducting a screening test prior to vaccination to ensure that only individuals who have previously been exposed to the virus (seropositive) receive the vaccine [184]. The DENVax vaccine demonstrated a more balanced efficacy in preventing dengue disease and hospitalizations between those who had never been exposed to the virus and those who had [27, 29]. However, similar to the observations with Dengvaxia, the serostatus of an individual before vaccination remains a key factor in determining the vaccine's effectiveness [143, 5]. Recent findings indicate a gradual decline in the protection offered by the vaccine over time [143, 11]. Consequently, it is essential to maintain long-term surveillance, involving thorough monitoring of individuals who received the DENVax vaccine during phase 3 trials.

Chapter 2

Epidemiological modelling

Epidemiological models, including compartment models and individual-based models, play a crucial role in examining the effects of preventive measures in silico [65, 25, 111]. Although individual-based models typically exhibit greater complexity and higher computational demands compared to compartment models, they offer a more precise assessment of preventive strategies [55]. To fully leverage these advantages and ensure the practical application of individual-based models, it is vital to optimize the use of available computational resources [111].

In the literature, preventive strategies are typically assessed by simulating each strategy an equal number of times [64, 59, 42]. However, this method is inefficient for identifying the optimal preventive strategy, as it consumes significant computational resources exploring suboptimal options [111]. Moreover, there is no consensus on the required number of model evaluations per strategy [185], and research indicates that this number varies with the complexity of the evaluation problem [111]. It is also important to recognize the necessity of planning epidemiological modeling experiments and specifying a computational budget in advance.

Given that running an individual-based model is computationally intensive, ranging from minutes to hours depending on the model's complexity [111, 110], minimizing the number of required model evaluations significantly reduces the total time needed to assess a set of preventive strategies. This makes the use of individual-based models feasible in studies where it might otherwise be computationally impractical [111, 110]. Additionally, reducing the number of model evaluations frees up computational resources in studies that already utilize individual-based models, allowing researchers to explore a broader set of model scenarios [111, 110]. This is crucial, as considering a wider range of scenarios enhances the confidence in the overall utility of preventive strategies [193].

This need for sample efficiency in epidemiological modeling underscores the importance of optimizing the allocation of computational resources. To address this challenge, techniques from the field of reinforcement learning offer promising solutions.

Chapter 3

Objectives and organization of this dissertation

3.1 Objectives

The objective of this dissertation is to contribute to the decision making process of selecting optimal vaccination strategies for the mitigation of dengue epidemics. To this end, a reinforcement learning approach is applied to address the need for sample-efficiency when investigating mitigation policies in epidemiological models.

Techniques from the field of reinforcement learning have already been applied to improve sample efficiency in optimization problems in various fields when the evaluation or simulation of strategies is computationally expensive. Examples include the optimization of power delivery and turbine life span in wind farms [96], and optimal planning of route choice with respect to individual travel time and overall system efficiency [47]. Other applications of reinforcement learning for finding optimal strategies that have been developed over the last years include traffic signalling [91], bidding and pricing [106], and intelligent manufacturing [109].

In recent years, reinforcement learning techniques have also already been applied to the identification of optimal mitigation strategies for various epidemics. Notable examples within this epidemiological setting include the use of single-objective multi-armed bandits (MABs) to efficiently evaluate influenza mitigation strategies [112, 111, 110], the use of Proximal Policy Optimization and Deep Q-Networks to evaluate school closure policies for the mitigation of influenza epidemics [110, 113], and the application of deep multi-objective reinforcement learning to learn a set of optimal deconfinement strategies for the mitigation of COVID-19 epidemics [141].

While dengue epidemics provide a particularly challenging and relevant setting, reinforcement learning techniques have not yet been employed to identify the optimal vaccination strategies to mitigate them. In this dissertation, a novel technique is contributed to evaluate vaccination strategies for the mitigation of dengue epidemics as a fixed budget multi-objective multi-armed bandit (MOMAB) Pareto-front identification (PFI) problem. As would be the case for uniform evaluation, the choice of budget is left to the decision maker, as we study a framework that works independently of the selected budget. The novel setting proposed in this dissertation is inspired by, and builds upon the use of single-objective MABs for the mitigation of influenza epidemics

from [112, 111, 110], extending it to the realm of multi-objective reinforcement learning inspired by [141], and applying it to dengue epidemics.

To this end, the first main research question that is examined in this dissertation is whether MOMAB algorithms for the PFI setting can be used to identify the subset of optimal vaccination strategies for the mitigation of dengue epidemics, and the trade-offs between them, in a sample efficient manner within the allocated budget, based on the output of computationally expensive stochastic simulations. The identified subset of optimal vaccination strategies can then be presented to the decision maker. Knowledge of the complete set of optimal preventive strategies, and the trade-offs between them, helps them significantly in making an informed decision about which strategy to implement. Note that the objective is not to learn the preferences of the decision maker or to make any assumptions about their preferred strategies. Selecting the most suitable vaccine allocation strategy from the subset of optimal vaccine allocation strategy is left to the discretion of the decision maker, where they can use the output of the bandit algorithm to support their decision making.

Both the extension to multi-objective reinforcement learning and the new application to dengue epidemics present intriguing areas for study. In real-life scenarios, decision-makers typically need to consider several potentially conflicting objectives. The example that will be studied throughout this dissertation is the trade-off between the effectiveness of a vaccination strategy and the monetary cost associated with its implementation. Although this dissertation focuses on two objectives, the proposed framework operates independently of the number of objectives, allowing for evaluation with respect to other objectives as necessary. Furthermore, the adaptability of the (multi-objective) multi-armed bandit framework for identifying optimal prevention strategies for both influenza and dengue highlights one of its key strengths: general applicability. When modeling infectious diseases, and when creating models in general, many setting-specific assumptions about the underlying real-world processes are made. While these assumptions are necessary, as models cannot be developed without them, they also serve the important function of making our assumptions explicit, thereby encouraging rigorous reasoning. However, the bandit framework works independently of the assumptions underlying the models, relying solely on the stochastic outputs to learn the optimal strategies (arms). This characteristic enables the (multi-objective) multi-armed bandit framework to identify optimal policies across a wide range of models, making it a highly valuable area for study and the generation of new insights. Due to the general applicability of the (multi-objective) multi-armed bandit framework, another primary objective of this dissertation is to gain new insights into this framework and contribute to its development.

To summarize, the two main goals of this dissertation are:

1. To contribute to the decision making process of selecting optimal vaccination strategies for the mitigation of dengue epidemics. To achieve this goal, a DENV MOMAB setting will be proposed in Chapter 11 within which the sample-efficiency and viability of various PFI MOMAB algorithms will be evaluated.

2. To study and gain new insights into the MOMAB framework, specifically within the context of Pareto-front identification (PFI), and to contribute to its development by proposing a completely novel PFI MOMAB algorithm in Section 10.3.

3.2 Organization

Following this first introductory part that delineates this dissertation's objectives and structure, Part II provides background information on the most important concepts that are used throughout the rest of the study. Chapter 4 aims to provide the reader with an introduction to the modelling of infectious diseases, gradually building up the complexity by starting from the most basic compartmental models and moving to stochastic and individual based models, as well as meta-population models. In Chapter 5, the multi-armed bandit (MAB) framework is introduced, giving the reader insight into both the regret minimization and the best-arm identification setting as the distinction between these settings is of particular importance for Chapter 10. This is achieved through the discussion of three MAB algorithms for both distinct settings. In Chapter 6, the final chapter of Part II, the multi-objective extension of the MAB framework is introduced.

After thorough background has been provided in Part II, the study then moves on to the literature review in Part III. The first part of the literature review in Chapter 7 is dedicated to the previous 13 years in developments in the modelling of dengue epidemics. In the second part of the literature review, Chapter 8, the literature on MOMAB algorithms is studied. There, two main categories of MOMAB algorithms are defined based on the current literature. The first category of algorithms incorporates the user's preferences, either by making assumptions about them or by interacting with the user at runtime. The second category comprises algorithms that strive to identify the entire set of Pareto optimal arms without considering the utility function or preferences of the decision maker. In this study these are referred to as the preference-unaware MOMAB algorithms. The literature on each of these classes of algorithms is subsequently discussed in Section 8.1 and Section 8.2. Section 8.2 also presents the reproduction and experimental verification of four distinct preference-unaware MOMAB algorithms from the literature.

After the relevant literature has been reviewed, this dissertation's methodological contributions and the associated experiments and results are discussed in Part IV. Part IV is split into three main chapters:

1. In Chapter 9, all contributions this dissertation makes within the field of dengue epidemiological modelling are analyzed. The methodological contributions made within this domain are threefold: in Section 9.1 the partial reproduction of the 2016 Ferguson et al. model is discussed, in Section 9.2 we show the process of reproducing the 2009 Recker et al. model, and in Section 9.3 the extension of the 2009 Recker et al. model with support for vaccination and age-heterogeneity is explained. Throughout these sections, various experiments and their corresponding model outputs are visualized.

2. Chapter 10 deals with all contributions made to MOMAB framework with the aim of completing objective (2) of this dissertation. In Section 10.1, three performance metrics for the MOMAB PFI setting are proposed: the Bernoulli metric, the Jaccard similarity metric, and the hypervolume metric. These performance metrics are tailor-made to quantify the quality of recommendations made by MOMAB PFI algorithms. Moving on from the performance metrics, Section 10.2 discusses the reproductions of 9 variants of MOMAB algorithms for the regret minimization setting previously presented in Section 8.2. Furthermore, the adaptation of four of these algorithms to the PFI setting is explained. We also benchmark the performance of these four algorithms adapted to the PFI setting with respect to the previously proposed performance metrics. Finally, Section 10.2 also discusses a number of insights into the relations between the different (MO)MAB settings and the algorithms designed to solve them. Last but not least, in Section 10.3, a completely novel preference-unaware PFI MOMAB algorithm is proposed: Top-Two Pareto Fronts

Thompson Sampling (TTPFTS). Apart from this methodological contribution, we also experimentally verify the performance of the novel TTPFTS algorithm compared to the four other PFI MOMAB algorithms.

3. In Chapter 11, the novel DENV MOMAB setting proposed by this dissertation is presented. This setting was specifically developed to complete goal (1) of this dissertation. In Section 11.1 the composition of various previously discussed elements of this study into the experimental DENV MOMAB setting is discussed. This section represents a major methodological contribution in which all previously made insights and results, into the modelling of dengue epidemics and the MOMAB framework applied to the PFI setting, are combined into a single experimental setup. Section 11.2 details the experiments conducted using the described setup, and in Section 11.3 the obtained results are visualized and analyzed, resulting in a number of interesting and valuable insights.

Following the extensive part dedicated to the numerous contributions, Part V is used to debate a number of interesting observations made throughout this study, ruminate about possible future work, and to reflect critically on the conducted research. Finally, Part VI concludes this dissertation.

Part II

Background

Chapter 4

Epidemiological modelling

Epidemiology, deriving its name from the Greek words *epi* ("upon"), *demos* ("people"), and *logos* ("study"), is dedicated to understanding the distribution and determining factors of health-related conditions within populations [119]. This discipline, often associated with Hippocrates, the "father of epidemiology", for his pioneering recognition of the relationship between diseases and environmental influences [87], has evolved significantly since its inception. The formal acknowledgment of epidemiology, particularly in the context of epidemics, was first documented by the Spanish physician de Villalba in *Epidemiologia Espanola* from 1802 [119]. Despite advancements, infectious diseases like lower respiratory infections, HIV, and dengue continue to pose major global health challenges.

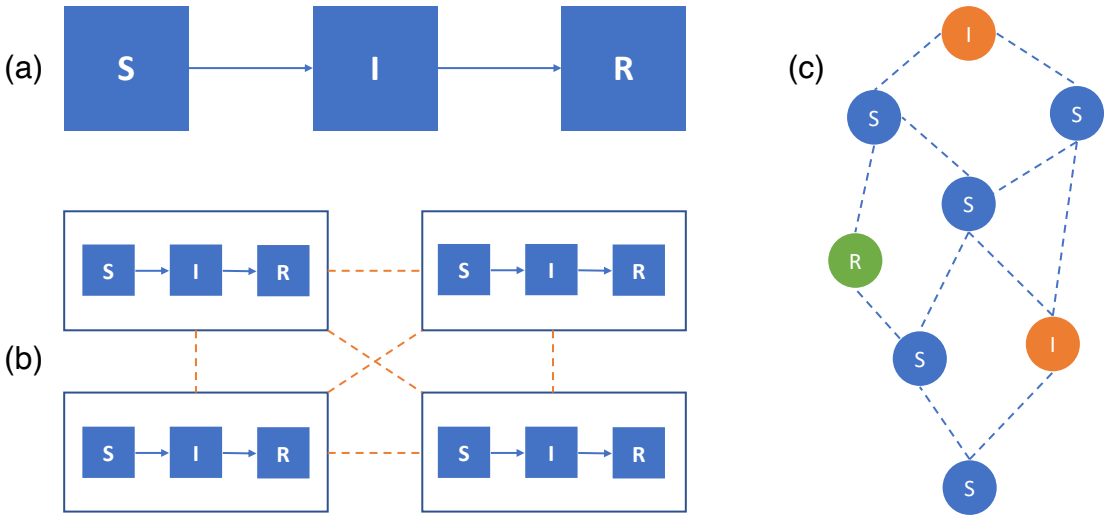


Figure 4.1: Overview of different epidemiological models: (a) A compartmental SIR model, (b) a slightly more complex meta-population model, and (c) an even more complex and fine-grained individual-based model. Reproduced from [110].

The mathematical modeling of infectious diseases, the focal point of this chapter, necessitates the identification of a suitable epidemiological model structure. This choice is dependent upon

various factors, including the nature of the pathogen, the pattern of social interactions, vector ecology, and the specific mitigation policies under consideration. A critical aspect of model selection is its level of detail or granularity. For instance, compartment models simplify the population into discrete, homogeneous groups, facilitating the analysis of transitions between these states [50]. Alternatively, individual-based models offer a detailed representation by simulating each person and their interactions, thereby tracing the pathogen's spread through the population [186]. Bridging these approaches, meta-population models encompass a spectrum of structures that cater to diverse public health questions, as illustrated in Figure 4.1. This chapter[1] will provide the reader with an introduction and insight into three primary modeling frameworks: compartment models, individual-based models, and meta-population models.

4.1 Compartment models

4.1.1 SIR model

Compartmental models categorize a population into a limited number of states, or compartments, allowing for interaction among them. Specifically, in the context of a pathogen that confers immunity post-infection (e.g., pandemic influenza), the population can be divided into three categories: susceptible individuals, those currently infected, and those who have recovered and thus gained immunity. This framework, known as the SIR model, short for Susceptible-Infected-Recovered, was first proposed by Kermack and McKendrick in 1927 [97]. Transition between the Susceptible and Infected compartments is facilitated by the rate of infection, denoted as $\beta\chi$, where β represents the likelihood of infection and χ denotes the rate of contact. The rate of recovery, symbolized by γ modulates the flow from the Infected compartment to the Recovered compartment. This can be seen visualized in Figure 4.2

Figure 4.2: A SIR model consisting of three compartments for susceptible (S), infected (I), and recovered (R) individuals. Flow between compartments is modulated by the transmission rate $\beta\chi$ and the recovery rate γ. Reproduced from [110]

.

The SIR model is particularly relevant for analyzing epidemics that emerge abruptly and conclude within a short span, negating the need to account for demographic changes due to births and deaths. Epidemics fitting this model include seasonal influenza, and the Ebola virus [110].

The SIR model can be mathematically represented as a system of ordinary differential equations like the one in Equation (4.1).

$$\begin{aligned} \dot{S}(t) &= -\beta\chi S(t)\frac{I(t)}{N(t)} \\ \dot{I}(t) &= \beta\chi S(t)\frac{I(t)}{N(t)} - \gamma I(t) \\ \dot{R}(t) &= \gamma I(t) \end{aligned} \tag{4.1}$$

[1] For this chapter I drew significant inspiration from chapter 3 of Prof. Dr. Libin's PhD thesis [110] as their analysis of epidemiological modeling provided an excellent foundational framework.

Here, β represents the probability of infection when a contact takes place between an infected and a susceptible individual, χ is the rate at which contacts occur, and γ is the recovery rate, with the following initial conditions,

$$S(0) > 0, I(0) > 0, R(0) = 0, \tag{4.2}$$

and the total population N is defined as the sum of the number of people in each compartment:

$$N(t) = S(t) + I(t) + R(t). \tag{4.3}$$

The definition of the SIR model in Equation (4.1) shows that each individual susceptible to infection encounters χ persons each day, representing the contact rate. A portion $\frac{I(t)}{N(t)}$ of these contacts is infectious. The probability of transmission per contact, β, varies with the pathogen and the transmission pathway. Therefore, the expression $\beta\chi S(t)\frac{I(t)}{N(t)}$ quantifies the rate at which susceptible individuals transition to the Infected compartment per unit of time. Subsequently, those infected recover at a rate γ, signifying that $\gamma I(t)$ represents the flow of individuals moving from the infected to the recovered compartment per unit of time [110]. A visualization of the dynamics of a SIR model with transmission rate $\beta\chi = 0.2$ and recovery rate $\gamma = 0.1$ can be seen in Figure 4.3.

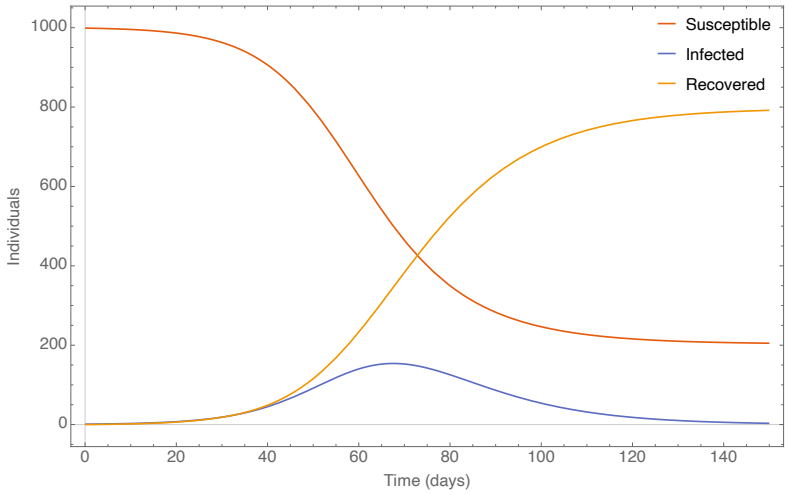


Figure 4.3: The number of susceptible (S), infected (I), and recovered (R) individuals as a function of time when considering a SIR model with $\beta\chi = 0.2$ and $\gamma = 0.1$. The size of the population $N = 1000$ and $I(0) = 1$. Reproduced from [110].

Moreover, Equations (4.1) to (4.3) highlight three foundational assumptions of this model. First, it considers the population to be closed, disregarding any births, deaths, or migrations. Second, it presupposes uniform mixing within the population, implying no significant variations due to spatial or age-related factors. Third, it posits that at the beginning of an epidemic, the number of infected individuals will rise exponentially, reflecting the initial rapid spread of the infection under these assumptions [110].

The basic reproductive number, denoted as R_0 in epidemiological models, is a crucial parameter that quantifies the average number of secondary infections produced by a single infected individual in a wholly susceptible population [110]. This metric is instrumental in understanding

the potential spread and control of infectious diseases, as it signifies the initial rate of spread of the epidemic [95]. To intuitively derive R_0 from the SIR model, consider the dynamics of transmission and recovery. The rate at which an infected individual contacts susceptibles and potentially transmits the disease is represented by $\beta\chi$ where β is the probability of transmission per contact and χ is the contact rate. The average duration an individual remains infectious is the reciprocal of the recovery rate, $\frac{1}{\gamma}$. Therefore, R_0 can be conceptualized as the product of the infection rate ($\beta\chi$) and the infectious period ($\frac{1}{\gamma}$). Mathematically, this is expressed as:

$$R_0 = \frac{\beta\chi}{\gamma} \tag{4.4}$$

This derivation from the SIR model underscores R_0's significance in gauging the initial spread of an epidemic, where a value of $R_0 > 1$ implies that the infection will likely propagate through the population, while $R_0 \leq 1$ suggests that the outbreak will eventually subside [110].

The SIR framework can be adapted to address more sophisticated epidemiological questions through two principal methods. Firstly (i), the introduction of additional compartments can enhance the model's complexity by incorporating new dimensions or characteristics pertinent to the disease's transmission dynamics. Secondly (ii), duplicating the basic model can depict population heterogeneity, reflecting variations in susceptibility, behavior, or other relevant factors.

The SIR model is formalized through a set of ordinary differential equations, as outlined in Equation (4.1), which suggests a deterministic approach to system analysis. Nonetheless, for forecasting purposes, stochastic models are often favored because they can incorporate random variations and facilitate the assessment of uncertainty [98]. Additionally, acknowledging the stochastic nature of epidemic spread is crucial for the assessment of intervention strategies [65]. Consequently, Section 4.1.4 will explore the methods by which the SIR model, along with other compartmental frameworks, can be interpreted through a stochastic lens.

4.1.2 SEIR model

Numerous pathogens are characterized by a latency period during which individuals, though infected, are not yet infectious. The SEIR model accommodates this aspect by incorporating an Exposed (E) compartment into the SIR framework, introducing an additional transition, denoted by ζ, the rate of latency, facilitating the progression from exposed to infectious states [110]. Figure 4.4 shows a representation of a SEIR model with the additional Exposed compartment when compared to the SIR model.

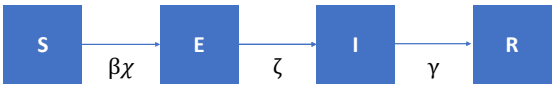


Figure 4.4: A SEIR model consisting of four compartments for susceptible (S), exposed (E), infected (I), and recovered (R) individuals. Flow between compartments is modulated by the transmission rate $\beta\chi$, the latency rate ζ, and the recovery rate γ.

This expanded model, the SEIR model, is also articulated through a series of ordinary differential

equations, detailed in Equation (4.5).

$$\begin{aligned}\dot{S}(t) &= -\beta\chi S(t)\frac{I(t)}{N(t)}\\ \dot{E}(t) &= \beta\chi S(t)\frac{I(t)}{N(t)} - \zeta E(t)\\ \dot{I}(t) &= \zeta E(t) - \gamma I(t)\\ \dot{R}(t) &= \gamma I(t)\end{aligned} \tag{4.5}$$

Here, β represents the probability of infection when a contact takes place between an infected and a susceptible individual, χ is the rate at which contacts occur, γ is the recovery rate, and ζ is the latency rate, with the following initial conditions,

$$S(0) > 0, E(0) > 0, I(0) \geq 0, R(0) = 0, \tag{4.6}$$

and the total population N is defined as the sum of the number of people in each compartment:

$$N(t) = S(t) + E(t) + I(t) + R(t). \tag{4.7}$$

A visualization of the dynamics of a SEIR model with transmission rate $\beta\chi = 0.2$, recovery rate $\gamma = 0.1$, and latency rate $\zeta = 1$ can be seen in Figure 4.5.

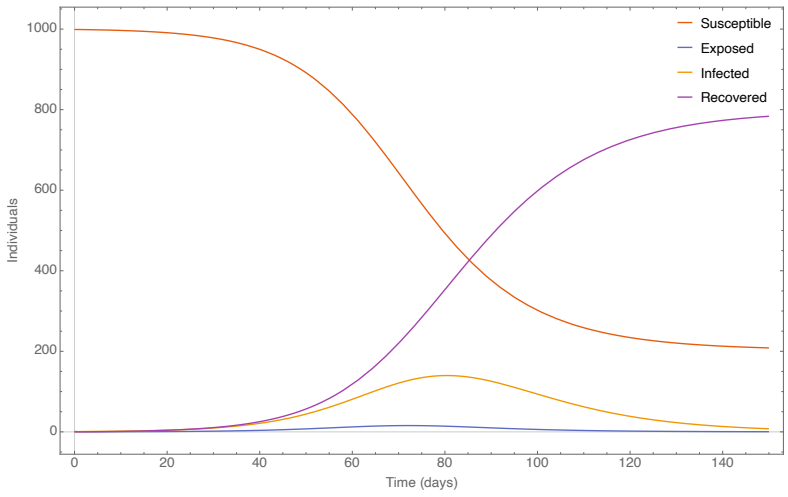


Figure 4.5: The number of susceptible (S), exposed (E), infected (I), and recovered (R) individuals as a function of time when considering a SEIR model with $\beta\chi = 0.2$, $\zeta = 1$, and $\gamma = 0.1$. The size of the population $N = 1000$ and $E(0) = 1$. Reproduced from [110].

4.1.3 Age-heterogeneous SIR model

The SIR model presupposes homogeneous mixing among all individuals within the population, an assumption that becomes implausible when evaluating policies such as school closures or vaccine distribution, which necessitate consideration of age-specific interactions [110]. To integrate age-dependent mixing into the SIR framework, one might divide the population into n distinct age

categories, establishing a separate SIR model for each group. These age-specific SIR models are subsequently interconnected to simulate the age-dependent mixing among the various age cohorts [110]. Such a model can be formalised as a system of ordinary differential equations like the one presented in Equation (4.8).

$$\begin{aligned} \dot{S}_i(t) &= -\beta S_i(t) \sum_{j=0}^{n} M_{ij} \frac{I_j(t)}{N_j(t)} \\ \dot{I}_i(t) &= \beta S_i(t) \sum_{j=0}^{n} M_{ij} \frac{I_j(t)}{N_j(t)} - \gamma I_i(t) \\ \dot{R}_i(t) &= \gamma I_i(t) \end{aligned} \tag{4.8}$$

Here, β gives the probability that an infection will take place, γ is the recovery rate, and M_{ij} is the average frequency of contacts between individuals in age groups i and j [110], with the initial conditions

$$S_i(0) > 0, I_i(0) > 0, R_i(0) = 0, \tag{4.9}$$

and the total size of a certain age group N_i within the population is given by the sum of the number of individuals in each compartment for that age group:

$$N_i(t) = S_i(t) + I_i(t) + R_i(t) \tag{4.10}$$

Compared to the formalisation of the SIR model in Equation (4.1), it can be seen that the age-heterogeneous SIR model presented here has a separate model for each of the age groups. For each of these models, the χ term, representing the rate of contact, has been incorporated in a weighted sum weighed by the average mixing frequency between age group i and j, M_{ij} [110]. Information about the average mixing frequency between the model's age groups can be gathered through surveys [127].

The age-heterogeneous SIR model will be demonstrated through a reproduction of one of Sherry Tower's lectures [171], the same way as in [110]. In this lecture, the population is divided into two age classes: children and adults. 25% of the population are children and 75% are adults. The proposed model can be seen visualized in Figure 4.6. For this model, the exact equations described in Equation (4.8) are used.

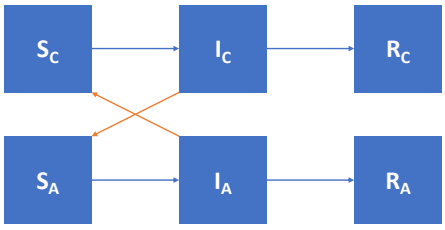


Figure 4.6: Age-heterogeneous SIR model with two age classes: children (C) and adults (A). Mixing between age groups is indicated by the orange arrows. Reproduced from [110].

The following initial conditions are used for the system:

$$\begin{aligned} S_C(0) &= N_C(0) - 1 \\ S_A(0) &= N_A(0) - 1 \\ I_C(0) &= I_A(0) = 1, \end{aligned} \tag{4.11}$$

and the contact matrix that descibes the average mixing between the two age classes M is given by

$$M = \begin{array}{c} \\ C \\ A \end{array} \begin{array}{c} \begin{array}{cc} C & A \end{array} \\ \begin{pmatrix} 18 & 9 \\ 3 & 12 \end{pmatrix} \end{array}. \tag{4.12}$$

Solving the system of ordinary differential equations and using the resulting functions to create a plot of the prevalence of infection as a function of time yields the graph shown in Figure 4.7.

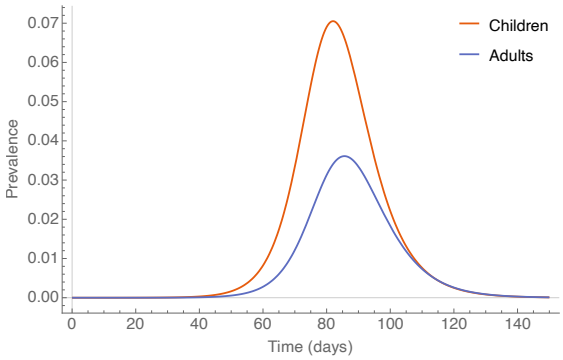


Figure 4.7: Prevalence of infection in both children and adults as a function of time. The recovery rate $\gamma = 1/3$ and the probability of infection β is calculated based on the R_0 value, γ, and the largest eigenvalue of the contact matrix M. Reproduced from [171].

4.1.4 Stochastic SIR model

Various methodologies have been developed for sampling trajectories within compartment models, among which the Gillespie algorithm stands as a notable technique [69]. This method conceptualizes the epidemiological system as akin to a chemical mechanism, treating individuals as reactants categorized by their compartmental affiliations, with transitions such as infection or recovery modeled as chemical reactions. The Gillespie algorithm facilitates the generation of precise stochastic trajectories for these reactions, employing Monte Carlo methods to ascertain the subsequent reaction and its timing. The selection of a reaction is governed by the probability proportional to the reactant count (i.e., the number of individuals available), and the time intervals are determined by an exponential distribution defined by the aggregate reaction duration. Although Gillespie's initial proposition [69] was predicated on unchanging rates, subsequent enhancements have accommodated time-variant reaction propensities and delays [37, 19]. Despite the algorithm's capacity for exact trajectory sampling, its computational demands have spurred the development of various adaptations, both for precise [19, 37, 68, 162] and approximate trajectory generation [38, 18].

An alternative principal approach involves stochastic differential equations (SDEs). To derive stochastic trajectories from a compartment model, the system of ordinary differential equations (ODEs) is converted into a system of SDEs. This conversion, as elucidated by Allen et al. [16], integrates noise terms for each transition within the original ODE framework, thereby incorporating stochastic elements into the model. For each transition $(X \to Y)$ from compartment X to compartment Y, with rate $\xi x(t)$, there is a noise term:

$$\sqrt{\xi x(t) X(t)} \dot{\mathcal{W}}_{(X \to Y)}(t), \tag{4.13}$$

where $\dot{\mathcal{W}}_{(X\to Y)}(t)$ is a Wiener process. This term is subtracted from the outgoing compartment and added to the incoming compartment [110]. For demonstration, this process is applied to the SIR model defined in Equation (4.1), resulting in the system of stochastic differential equations in Equation (4.14).

$$\begin{aligned}
\dot{S} &= -\beta\chi S(t)\frac{I(t)}{N(t)} - \sqrt{\beta\chi S(t)\frac{I(t)}{N(t)}}\cdot\dot{\mathcal{W}}_{(S\to I)} \\
\dot{I} &= \beta\chi S(t)\frac{I(t)}{N(t)} + \sqrt{\beta\chi S(t)\frac{I(t)}{N(t)}}\cdot\dot{\mathcal{W}}_{(S\to I)} - \gamma I(t) - \sqrt{\gamma I(t)}\cdot\dot{\mathcal{W}}_{(I\to R)} \\
\dot{R} &= \gamma I(t) + \sqrt{\gamma I(t)}\cdot\dot{\mathcal{W}}_{(I\to R)}
\end{aligned} \tag{4.14}$$

To simulate the system of stochastic differential equations (SDEs) and generate stochastic trajectories, one can employ the Euler-Maruyama approximation method, as delineated in the works of Allen et al. [16] and Rasmussen et al. [136]. The Euler-Maruyama method entails a process where, for each compartment within the model, deterministic components are multiplied by a small time increment Δt, and stochastic components are scaled by the square root of this time increment $\sqrt{\Delta t}$. Applying this method to the SDE system presented in Equation (4.14) results in a series of simulation equations that facilitate the numerical approximation of the system's dynamics:

$$\begin{aligned}
\Delta S &= -\Delta t\cdot\beta\chi\frac{SI}{N} - \sqrt{\Delta t}\sqrt{\beta\chi\frac{SI}{N}}\cdot\mathcal{N}(0,1) \\
\Delta I &= \Delta t\cdot\beta\chi\frac{SI}{N} + \sqrt{\Delta t}\sqrt{\beta\chi\frac{SI}{N}}\cdot\mathcal{N}(0,1) - \Delta t\cdot\gamma I - \sqrt{\Delta t}\sqrt{\gamma I}\cdot\mathcal{N}(0,1) \\
\Delta R &= \Delta t\cdot\gamma I + \sqrt{\Delta t}\sqrt{\gamma I}\cdot\mathcal{N}(0,1)
\end{aligned} \tag{4.15}$$

4.2 Individual-based models

In compartment models, individuals are aggregated based on shared characteristics, such as infection status or age group. Conversely, individual-based models distinctly delineate each person and their specific attributes [110]. Within these models, individuals are interconnected through networks that can be either static or dynamic, facilitating the simulation of epidemic spread across these connections [110].

At its core, an individual-based model would explicitly track each person, cataloging their infection status as susceptible (S), infected (I), or recovered (R). A basic illustration of this concept might involve arranging these individuals within a static network, as visually represented in Figure 4.8.

While this rudimentary model focuses on a singular aspect of an individual, it is possible to extend the model to encapsulate numerous attributes of both the individuals and their surroundings. These attributes can range from those directly influencing the course of an infection, such as infectiousness levels in influenza cases [41] or viral load in HIV infections [86], to those affecting preventive measures, like vaccination status [42] or the usage of condoms [94]. Furthermore, factors influencing network dynamics, such as workplace locations [41] or an individual's propensity for concurrent sexual partnerships [157], can also be incorporated to enhance the model's complexity and realism.

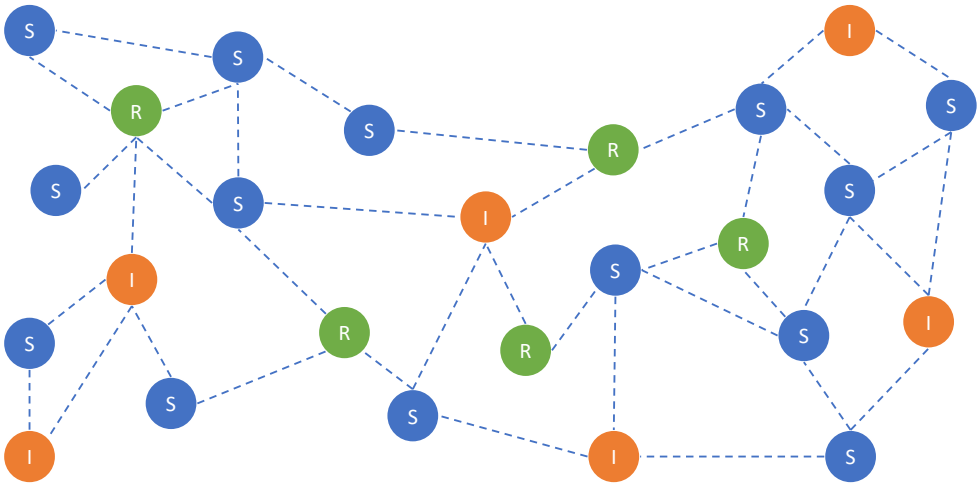


Figure 4.8: Visualization of a static network of individuals with their infection status: susceptible (S), infected (I), or recovered (R). An epidemic can be simulated over this network by studying the interactions between the nodes of the individual-based model.

Additionally, the network framework linking individuals can manifest as either static or dynamic. In static networks, the selection of a particular network configuration is influenced by the transmission pathway, for instance, sexual contact networks are often found to be scale-free [114]. Furthermore, these networks may exhibit overlapping characteristics; for example, during daytime hours, adults might connect with colleagues while children attend school, whereas evenings are typically reserved for familial interactions [70]. In contrast, dynamic networks require a mechanism for the dissolution and establishment of connections among individuals, as exemplified by the approach developed by Schmid and Kretzschmar [157], which dynamically constructs sexual networks by modulating ties in accordance with an individual's propensity for multiple concurrent sexual partnerships.

This exposition underscores the capability of individual-based models to dissect epidemic dynamics with remarkable granularity. Nonetheless, three significant considerations emerge. First, there exists a direct correlation between the level of detail integrated into the model and its computational demands [41]. Second, the development of such models necessitates a comprehensive understanding of the statistical distributions of various attributes. While certain data, such as population distribution, are readily accessible via geographic information systems (GIS), procuring data for other aspects, like zoonotic disease modeling, proves more challenging [110]. Third, the voluminous data produced by these models may complicate the extraction of insights regarding the principal factors influencing model outcomes [23].

4.3 Meta-population models

Compartment models are known for their computational efficiency, necessitating the segmentation of populations into broad categories. Conversely, individual-based models offer a more detailed perspective but at the expense of increased computational requirements. To navigate these trade-offs, meta-population models are often employed, particularly for simulating epidemic processes within a spatial framework [110]. Originally conceptualized within the realm

of ecology by Hanski et al. [83], these models are designed to depict sub-populations that are geographically distinct. An illustration of a basic meta-population model, featuring a separate patch for four geographical locations in Belgium, is presented in Figure 4.9.

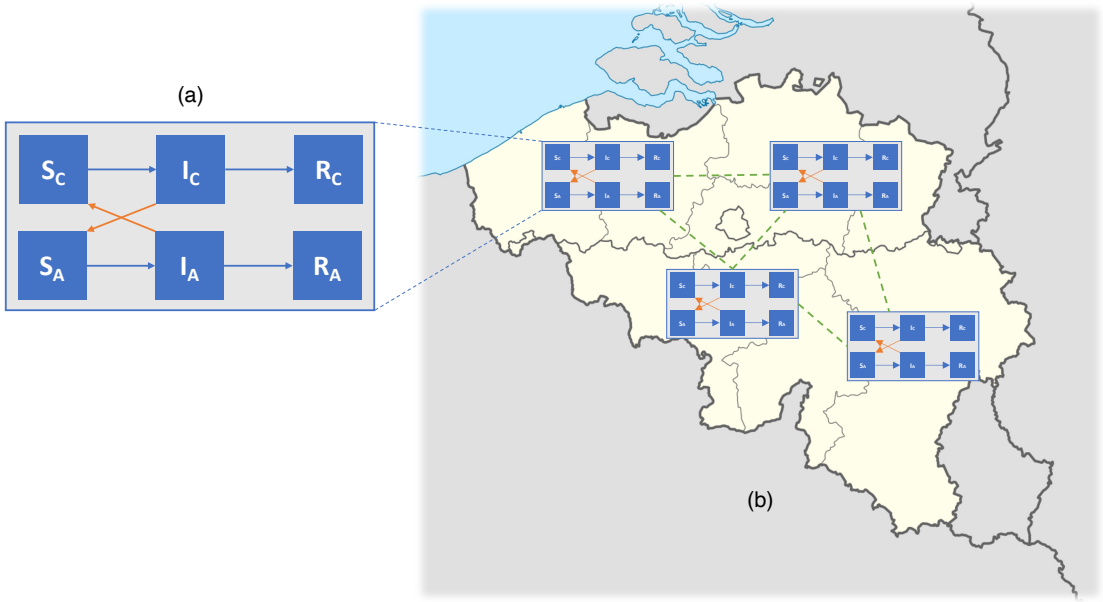


Figure 4.9: Expansion of the age-heterogeneous SIR model with two age classes (children (C) and adults (A)) proposed in Section 4.1.3 (a) to a meta-population model incorporating four age-heterogeneous SIR models for four geographical locations in Belgium (b).

Chapter 5

Multi-armed bandits

In this chapter[1], the concept of the *multi-armed bandit* (MAB) will be introduced. Intuitively MABs can be compared with the slot machines, also known as one-armed bandits, that can be found in casinos. In this analogy, the MAB is equipped with K arms, each denoted as a_k, where the actuation of a lever a_k yields a stochastic reward r_k [163, 175, 110]. This reward r_k is conceived as a realization from the reward distribution associated with arm a_k [175, 110]. The expectation of this reward, an unknown parameter, is represented as $\mu_k = \mathbb{E}[r_k]$ and denotes the expected reward of lever a_k. Multi-armed bandits can be adapted to encapsulate various problem domains, with the optimization of cumulative regret (regret minimization) and the identification of the optimal arm (best arm identification) being the most prominent applications [110].

The best arm a_* of a MAB can intuitively be thought of as the arm with the highest expected reward:

$$\mu_* = \max_k \mu_k. \tag{5.1}$$

Cumulative regret $R_C^{(T)}$ on the other hand can be thought of as the overall difference between the obtained rewards and the largest possible rewards that could have been obtained at each timestep:

$$R_C^{(T)} = \sum_{t=1}^{T} \mu_* - \mu_{a^{(t)}}. \tag{5.2}$$

In this equation defining the cumulative regret, $\mu_* - \mu_{a^{(t)}}$ is know as the instantaneous regret of pulling arm a_k at time t [175, 110, 163].

To effectively minimize cumulative regret within the multi-armed bandit framework, the intuitive strategy would be to consistently select the optimal arm. However, the inherent challenge lies in the fact that the identity of this optimal arm is not initially known, necessitating a phase of exploration to ascertain the most rewarding arm. Yet, excessive exploration can paradoxically augment cumulative regret [22], by diverting opportunities away from exploiting known profitable arms. Therefore, the quintessential strategy for minimizing cumulative regret involves a judicious balance between the exploration of untested arms and the exploitation of arms that are currently understood to yield favorable rewards [163, 175]. This dynamic interplay between exploration

[1]For this chapter I drew significant inspiration from sections 2.1 and 2.2 of Prof. Dr. Libin's PhD thesis [110] as their analysis of MABs provided an excellent foundational framework.

and exploitation is pivotal in devising effective strategies within the multi-armed bandit paradigm [110].

In contrast to the challenge of minimizing cumulative regret, the field also explores the best-arm identification problem, which focuses solely on pinpointing the most rewarding arm. This problem falls under the broader category of pure-exploration problems, as described by Bubeck et al. [33]. Approaches to best-arm identification vary, including strategies that operate within a predetermined budget [21], those that make a decision upon reaching a specified confidence level [56], and methods that offer a recommendation for the best arm at each time step [90]. The primary aim here is to minimize the so called simple regret $R_S^{(T)}$ [34] which can be defined as follows:

$$R_S^{(T)} = \mu_* - \mu_{J^{(T)}} \tag{5.3}$$

where μ_* is the average reward of the best arm and $\mu_{J^{(T)}}$ is the average reward of the recommended arm $J^{(T)}$ at time T [110].

5.1 Regret minimization algorithms

In this section, three popular MAB algorithms that are often applied within the context of regret minimization will be discussed: ϵ-greedy, Upper confidence bound (UCB), and Thompson sampling. Each of the algorithms is explained in detail and in order of increasing complexity, aiming to provide the reader with intuition on the variety of ways bandit algorithms can handle the exploitation-exploration trade-off.

Each of the algorithms is evaluated with respect to the cumulative regret using a running example. In this example, a bandit with $K = 31$ arms is used. Each of the arms represents what happens in a big population of potentially-sick individuals when vaccination strategies are implemented. Pulling one arm maps, in the real world, to the infection of a couple of people, and the assignment of vaccines to specific age groups. The reward given by the arm is computed from the total amount of people who became sick in 2 or 3 years. For the purpose of this example, thousands of outcomes for every arm have been generated and every time an arm was pulled, a random sample was taken from the samples corresponding to the selected arm. To obtain results that allow for a comparison of the different algorithms, each bandit was run for 10,000 steps, or arm pulls. The obtained cumulative regret was then averaged over 100 runs of the experiment.

5.1.1 Epsilon-greedy

An effective strategy for addressing the challenge posed by the cumulative regret in multi-armed bandit problems involves a predetermined allocation of the arm selection opportunities towards exploration and exploitation [110]. This methodology is precisely encapsulated by the ϵ-greedy algorithm, as delineated in Algorithm 1. Given that ϵ is generally small, it follows that the algorithm predominantly (with a probability of 1 - ϵ) opts for the arm perceived to be superior based on current estimations, thereby exploiting the available information [110, 107]. Conversely, a selection is made randomly with a probability of ϵ, facilitating the exploration of other options. It is important to note that the ranking of the arms is based on their empirical means, denoted as $\hat{\mu}_k^{(t)}$. Although this technique may not represent the pinnacle of efficiency, its intuitive nature and simplicity render it a popular choice in practical applications.

The algorithm necessitates the selection of a critical hyper-parameter, ϵ, which delineates the amount of exploration used by the algorithm. The optimal setting of this parameter is not

Algorithm 1: ϵ-greedy

Input: ϵ and a MAB with K arms

for $t \leftarrow 1$ **to** $+\infty$ **do**

$a_{\max} = \arg\max_k \hat{\mu}_k^{(t)}$

$$a^{(t)} = \begin{cases} a_{\max}, & \text{with probability } 1-\epsilon \\ \text{random element of } \{1,\dots,K\}, & \text{with probability } \epsilon \end{cases}$$

Play $a^{(t)}$ and observe its reward $r^{(t)}$

Update $\hat{\mu}_{a^{(t)}}^{(t)}$ with $r^{(t)}$

straightforward, as it is dependent upon the problem's complexity. In Figure 5.1, we illustrate the algorithm's functionality for ϵ values of 0.01 and 0.1. It is observed that a minimal exploration setting ($\epsilon = 0.01$) may lead to prolonged commitment to a suboptimal arm, thereby accelerating the accumulation of regret. Conversely, a higher exploration rate ($\epsilon = 0.1$) initially mitigates regret accumulation; however, it perpetuates exploration even post the identification of the optimal arm. This phenomenon is depicted in Figure 5.1, where the ϵ-greedy algorithm's performance at $\epsilon = 0.01$, in terms of cumulative regret, is surpassed by its counterpart set at $\epsilon = 0.1$.

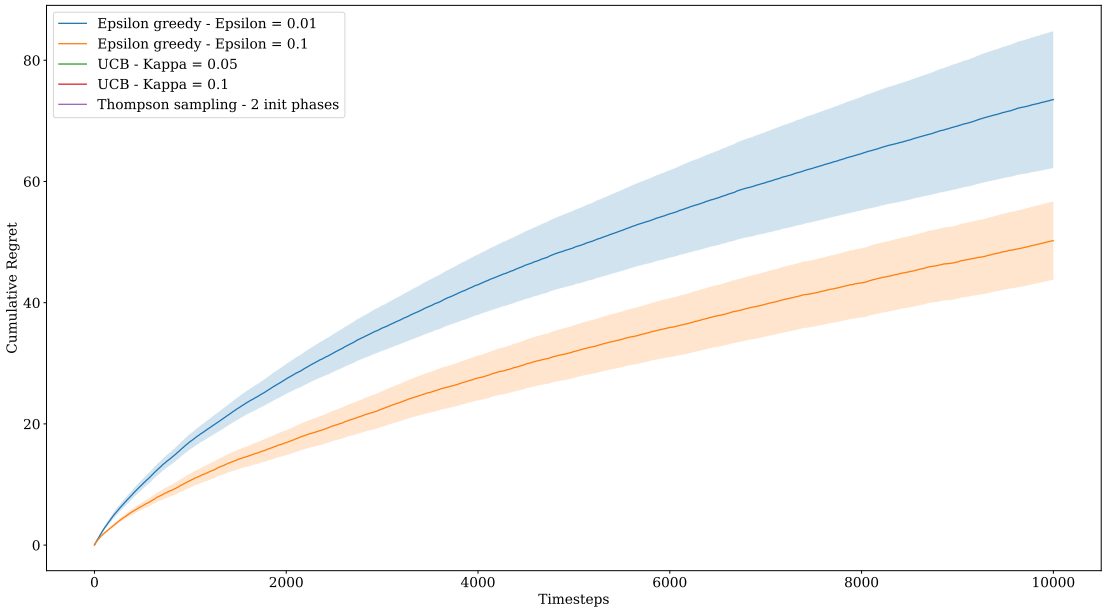


Figure 5.1: The accumulation of cumulative regret for the ϵ-greedy bandit over 10,000 arm pulls averaged over 100 runs, together with the 95% confidence interval around the mean. In this specific setting, the bandit with a higher exploration rate of 0.1 (orange line) outperforms the bandit with a lower exploration rate of 0.01 (blue line).

Such observations underscore two fundamental constraints associated with the conventional ϵ-greedy approach. Firstly, it maintains a constant exploration rate, irrespective of the progression of the algorithm [110, 175]. Secondly, the approach's exploration mechanism is relatively rudi-

mentary, as it resorts to uniform random exploration, disregarding the potential insights offered by the empirical mean estimates which could highlight significant disparities among the arms [110, 175]. These insights suggest a need for a dynamic exploration strategy, where the exploration rate diminishes over time, and exploration decisions are informed by the prevailing uncertainties surrounding the mean estimates. In pursuit of an advanced solution, we proceed to look at the Upper Confidence Bound (UCB) algorithm, as proposed by Auer et al. in 2002 [22], which addresses these limitations by incorporating a more nuanced exploration strategy.

5.1.2 Upper confidence bound

To balance between exploration and exploitation, the Upper Confidence Bound (UCB) algorithm quantifies the uncertainty about the reward distribution of each arm [110, 147]. It does this by maintaining a confidence bound of the following form for each arm:

$$B_k^{(t)} = \hat{\mu}_k^{(t)} + \kappa \cdot \sqrt{\frac{ln(t)}{n_k^{(t)}}} \tag{5.4}$$

Here, $\hat{\mu}_k^{(t)}$ is the empirical mean of arm k, t represents the total number of times an arm has been pulled, and $n_k^{(t)}$ is the number of times arm k was pulled. The κ parameter can be used to configure the amount of exploration. At each decision point, the arm selected for pulling is the one that maximizes the value of $B_k^{(t)}$, as delineated in Algorithm 2. This principle underlying the UCB approach inherently suggests a preference for arms with promising yet less frequent plays, thereby achieving a balance between exploration and exploitation efforts [110, 147].

Algorithm 2: Upper confidence bound

Input: κ and a MAB with K arms
for $t \leftarrow 1$ **to** $+\infty$ **do**
  $a^{(t)} = \arg\max_k \left(\hat{\mu}_k^{(t)} + \kappa \sqrt{\frac{\ln(t)}{n_k^{(t)}}} \right)$
  Play $a^{(t)}$ and observe its reward $r^{(t)}$
  Update $n_{a^{(t)}} : n_{a^{(t)}} = n_{a^{(t)}} + 1$
  Update $\hat{\mu}_{a^{(t)}}^{(t)}$ with $r^{(t)}$

A pivotal aspect of implementing UCB is the determination of the hyper-parameter κ, which governs the exploration intensity [110]. The efficacy of UCB is showcased through its application to the previously introduced 31-armed vaccine allocation bandit scenario, utilizing κ values of 0.05 and 0.1, as depicted in Figure 5.2. With a good choice of hyperparamter κ, the performance of UCB is notably superior to that of the ϵ-greedy algorithm with a lower exploration parameter ($\epsilon = 0.01$) and surpasses the ϵ-greedy variant with a higher exploration parameter ($\epsilon = 0.1$) swiftly. With both κ settings, UCB's cumulative regret trajectory stabilizes, reflecting growing confidence in the optimal arm selection - a desirable outcome. Nevertheless, the optimal selection of κ, akin to ϵ in the ϵ-greedy algorithm, depends on the complexity of the problem, making it difficult to determine in advance.

Furthermore, UCB's reliance solely on the empirical mean, $\hat{\mu}_k^{(t)}$, and the count of arm pulls, $n_k^{(t)}$, omits consideration of other reward distribution characteristics, which might be insightful, especially when prior knowledge exists. Such knowledge, even if rudimentary or intuitive, can significantly inform decision-making [147].

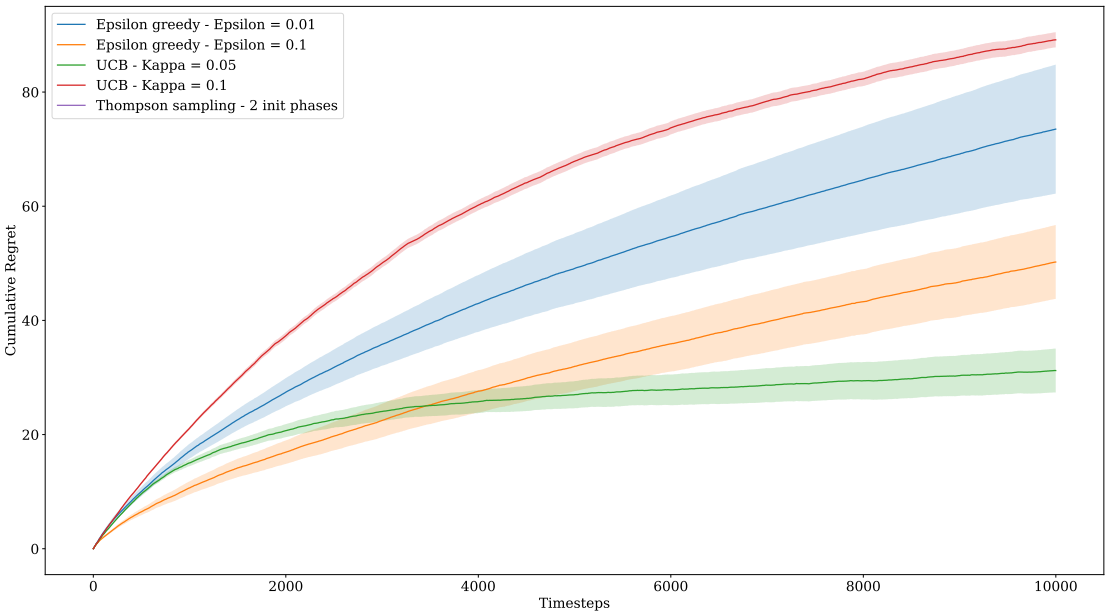


Figure 5.2: The accumulation of cumulative regret for the UCB based bandit over 10,000 arm pulls averaged over 100 runs, together with the 95% confidence interval around the mean. In this specific example, the UCB based bandit with the lower exploration setting outperforms the other bandits.

The Bayesian statistical paradigm offers a coherent framework for integrating this prior knowledge. Within the multi-armed bandit context, the Thompson sampling algorithm emerges as a Bayesian solution to the exploration-exploitation dilemma. Noted for its commendable performance [158, 43] and broad applicability [4, 132, 152], Thompson sampling operates by drawing from the Bayesian posterior distributions of the arm means [110]. This discussion sets the stage for delving into the Bayesian underpinnings essential for understanding Thompson sampling's mechanics.

5.1.3 Thompson sampling

To make the concept of integrating prior knowledge more intuitive, the example of flipping a coin to assess its fairness is considered. In the context of Bayesian inference, our objective is to determine the posterior distribution of a given hypothesis, articulated as a parameter vector $\boldsymbol{\theta}$. This process involves integrating our prior knowledge about the hypothesis with empirical data D, acquired through experimentation. In this scenario, the hypothesis pertains to the coin's bias, represented by the parameter θ, which signifies the probability of flipping a head. The empirical data D is gathered by flipping the coin and recording the outcomes, namely heads or tails.

In this framework, the posterior distribution, which reflects the probability of a hypothesis post-data observation, is articulated.

$$P(\theta|D) \tag{5.5}$$

To calculate the posterior, we introduce two foundational concepts: the prior belief concerning the hypothesis,

$$P(\theta), \tag{5.6}$$

and the likelihood of witnessing the observed data given a hypothesis θ,

$$P(D|\theta) \tag{5.7}$$

According to Bayes' theorem, the posterior distribution can be computed as the product of the prior and the likelihood, normalized by the marginal likelihood:

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(D|\theta')P(\theta')d\theta'}. \tag{5.8}$$

When the form of the posterior distribution aligns with that of the prior distribution, within the same family of probability distributions, such prior and posterior are termed conjugate distributions, a concept established by Schlaifer and Raiffa in 1961 [134]. The running example of coin flips presented here is an example of Bernoulli trials. In this case the beta distribution is commonly adopted as the conjugate prior, as highlighted by Robert in 2007 [144].

$$P(\theta) = \mathcal{B}eta(\theta|\alpha_0, \beta_0) \tag{5.9}$$

With:

$$\mathcal{B}eta(\theta|\alpha_0, \beta_0) = \frac{\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}}{B(\alpha_0, \beta_0)} \tag{5.10}$$

The beta function, denoted as $B(.,.)$, alongside hyperparameters α_0 and β_0, facilitates the incorporation of prior knowledge. Initially, we adopt a uniform prior, equivalent to a beta distribution where $\alpha_0 = 1$ and $\beta_0 = 1$.

The utility of conjugate priors lies in their provision for a closed-form solution for posterior updates upon data acquisition [144]. This obviates the necessity for numerical methods such as Markov Chain Monte Carlo for posterior approximation, a technique introduced by Hastings in 1970 [84]. In the context of our Bernoulli trials, the posterior is derived via Bayes' theorem, resulting in a Beta posterior that accounts for the occurrences of heads and tails.

To empirically validate this posterior distribution, two experimental setups are considered: one with a fair coin ($\theta_f = 0.5$) and another with a biased coin ($\theta_b = 0.7$). The evolution of the posterior is depicted in Figure 5.3 for the fair coin and in Figure 5.4 for the biased coin, starting from a uniform Beta prior in both instances. The initial few coin toss outcomes reveal the coin's propensity towards fairness or bias in Figures 5.3 and 5.4, respectively, albeit accompanied by considerable uncertainty. As the number of observations increases, this uncertainty diminishes, and post 500 coin tosses, the posterior distributions narrow significantly, offering robust evidence for the coin's fairness or bias.

Thompson sampling [170] is characterized by the establishment and updating of a Bayesian belief regarding the mean rewards of the bandit's arms. Initially, a prior probability distribution is imposed to represent the belief about the expected rewards, which is subsequently revised to a posterior probability distribution upon the acquisition of reward observations. At every decision point within the sampling process, a sample is drawn from the posterior distribution associated with each arm of the bandit. These samples are then ordered, and the arm corresponding to the highest-ranking posterior sample is selected for play. The reward observed from this action serves to refine the posterior distribution of the bandit.

Before discussing the Thompson Sampling algorithm in detail, first the posterior distribution over the bandit arms' means will be examined. Suppose a stochastic MAB, for which there

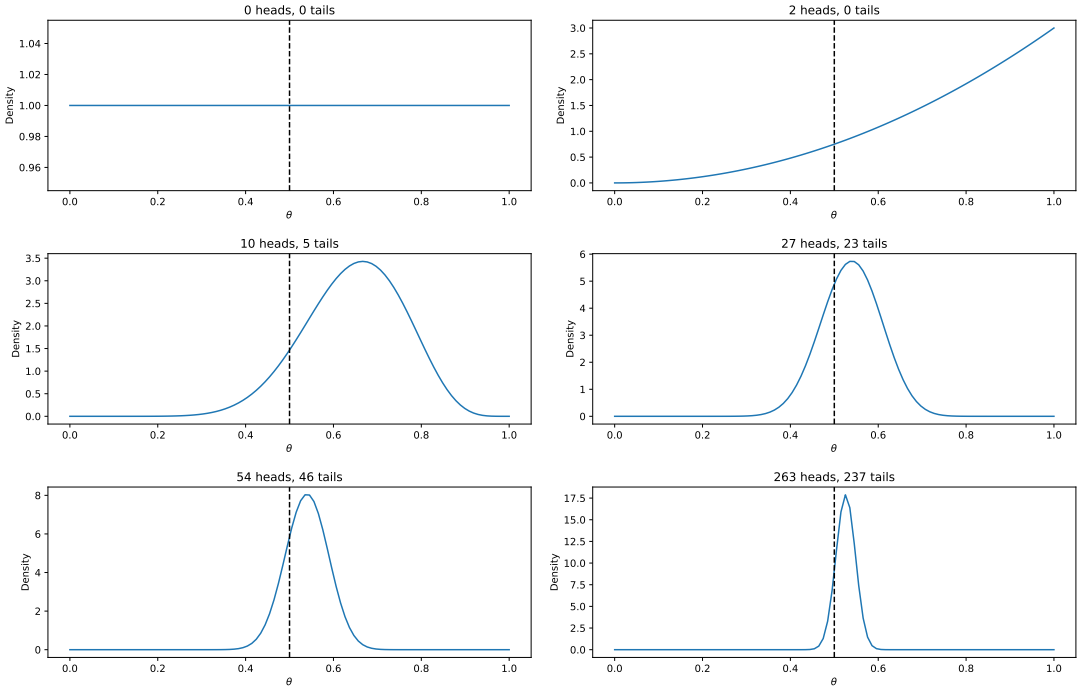


Figure 5.3: Evolution of the posterior distribution for the experiment with the fair coin. The x-axis represents θ. The vertical dotted line where $\theta = 0.5$ represents a fair coin. Reproduced from [110].

exists a distribution $\pi(.)$ that gives the prior belief over the means of the rewards associated with each arm. Given the obtained rewards for arm pulls up to time $t-1$,

$$\mathcal{H}^{(t-1)} = \left\{ a^{(i)}, r^{(i)} \right\}_{i=1}^{t-1}, \tag{5.11}$$

the posterior over the means of the bandit can be defined as

$$\pi(\cdot|\mathcal{H}^{(t-1)}). \tag{5.12}$$

For each of the means $\mu_{1..K}$, an estimate $\tilde{\boldsymbol{\mu}}^{(t)}$ is sampled at time step t from the posterior $\pi(\cdot|\mathcal{H}^{(t-1)})$. In order to determine the best obtained sample, a ranking operator is used. This way, the arm associated with the highest ranking sample can be written as:

$$\Psi_1(\tilde{\boldsymbol{\mu}}^{(t)}). \tag{5.13}$$

The exploration-exploitation dilemma is navigated by Thompson sampling through its consideration of the posterior's uncertainty. Initially, the scarcity of reward observations makes it so that there exists significant uncertainty, leading to an exploration strategy that closely approximates uniformity. As the more arms are pulled, enhancing the certainty regarding the superior arms, a gradual shift towards their preferential selection is observed, culminating in an almost exclusive focus on these arms. It should be noted, however, that as long as the posterior distribution employed possesses infinite support, a non-zero probability of selecting what are deemed sub-optimal arms persists, ensuring that exploration is never completely abandoned.

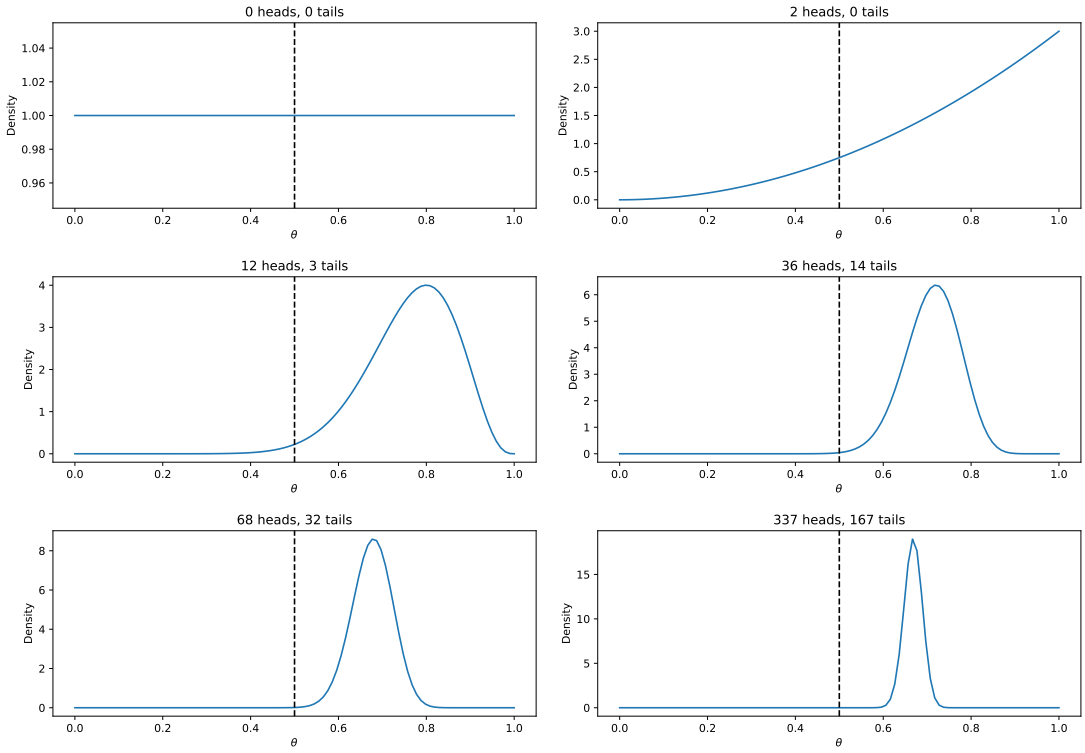


Figure 5.4: Evolution of the posterior distribution for the experiment with the biased coin. The x-axis represents θ. The vertical dotted line where $\theta = 0.5$ represents a fair coin. Reproduced from [110].

Algorithm 3: Thompson sampling

Input: A MAB with K arms, a prior $\pi(\cdot)$ and history $\mathcal{H}^{(0)} = \emptyset$
for $t \leftarrow 1$ **to** $+\infty$ **do**
  $\tilde{\boldsymbol{\mu}}^{(t)} \sim \pi(\cdot|\mathcal{H}^{(t-1)})$
  $a^{(t)} = \Psi_1(\tilde{\boldsymbol{\mu}}^{(t)})$
  $r^{(t)} \leftarrow$ Pull arm $a^{(t)}$
  $\mathcal{H}^{(t)} \leftarrow \mathcal{H}^{(t-1)} \cup \{a^{(t)}, r^{(t)}\}$

Thompson sampling is amenable to incorporating any form of prior knowledge that may be available to quantify the uncertainty associated with the means of the arms, ranging from a non-informative prior that merely delineates the reward distribution's family, to more informative priors that specify the distribution's family and variance, and even to priors that articulate dependencies between arms, as proposed by Gopalan et al. in 2014 [72]. The incorporation of such prior knowledge can significantly improve the learning process.

From a Bayesian standpoint, Thompson sampling is recognized for its intuitive approach to balancing the trade-off between exploitation and exploration by drawing samples from its posterior belief about the bandit. This belief, formalized as a posterior distribution, enables the articu-

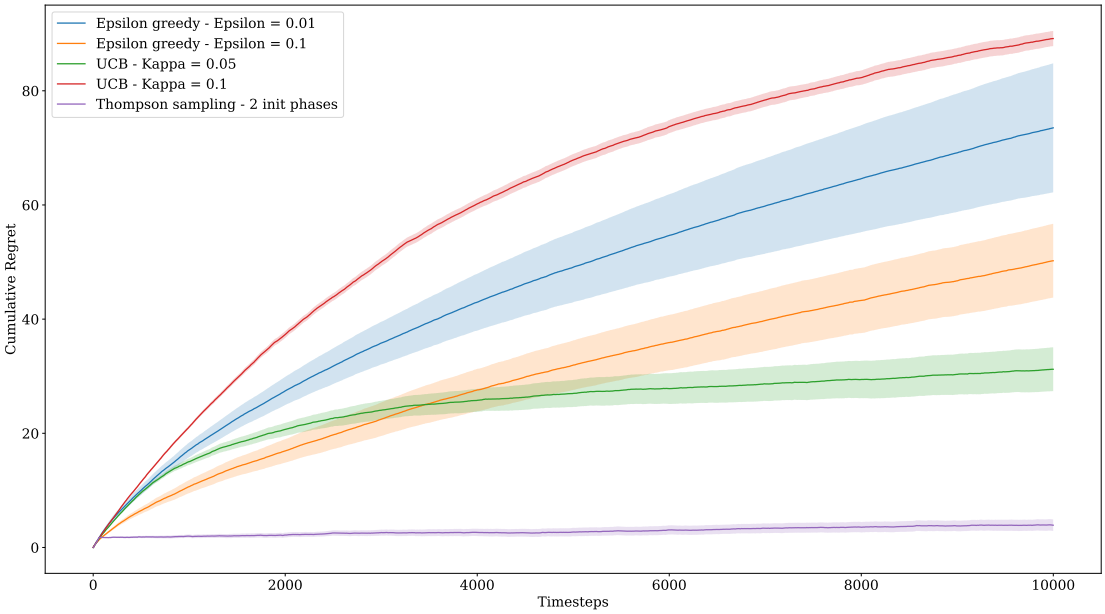


Figure 5.5: The accumulation of cumulative regret for the Thompson sampling bandit over 10,000 arm pulls averaged over 100 runs, together with the 95% confidence interval around the mean. In this specific setting, the bandit using Thompson sampling clearly outperforms all previously presented bandits.

lation of the uncertainty inherent in the decision-making process. To exemplify this approach, Thompson sampling will be applied to the Bernoulli bandit scenario introduced at the beginning of this section.

The initiation of Thompson sampling necessitates the selection of a suitable prior. In the ensuing example, the Jeffreys prior, a non-informative prior conjugate to the Bernoulli likelihood, will be employed. This prior is represented by a Beta distribution with parameters $\alpha_0 = \beta_0 = 0.5$, as suggested by Lunn et al. in 2012 [1].

As evidenced in Figure 5.5, a comparative analysis reveals that Thompson sampling markedly surpasses the performance of both the ϵ-greedy and UCB algorithms. Figure 5.6 shows the evolution of the rewards obtained by each of the different bandits. The Thompson sampling bandit consistently obtains the largest rewards, resulting in a lower cumulative regret compared to the other bandit algorithms that were discussed.

5.2 Best-arm identification algorithms

In this section, a number of bandit algorithms that are often applied within the best-arm identification setting will be discussed. Within this setting the goal is to find the optimal policy within a fixed budget of policy evaluations or arm pulls [21, 140].

The ability to recommend the correct arm to the user at each time step t of each of the algorithms was also examined using a Bernoulli bandit with 21 arms. The arms' success rates varied from a minimum of 0,3 to a maximum of 0,72 in increments of 0,02. The results of this experiment

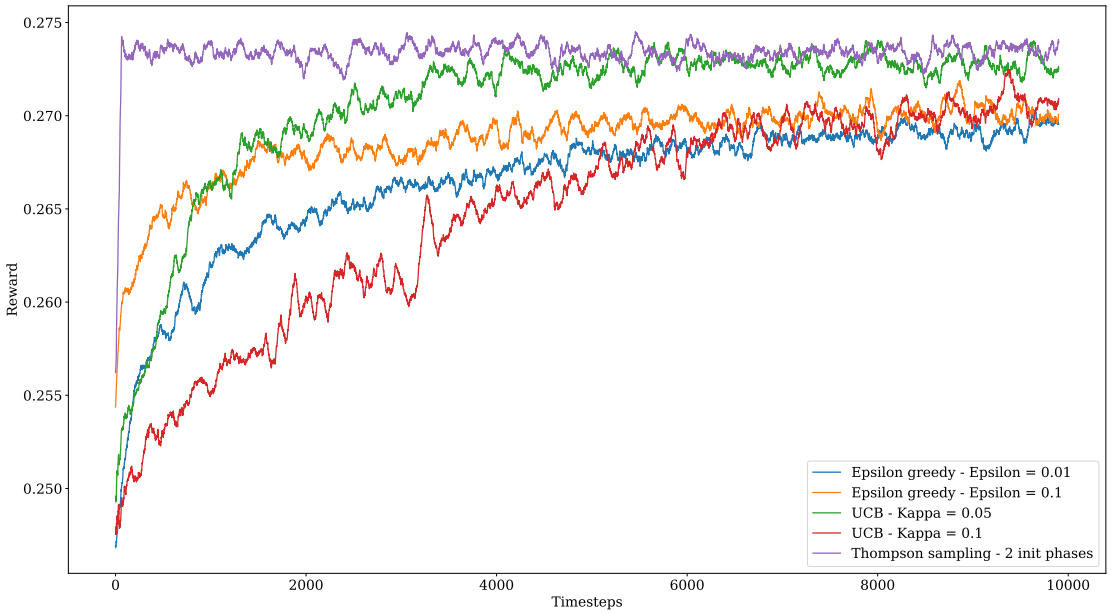


Figure 5.6: Evolution of the rewards obtained by each of the different bandits over 10,000 arm pulls, averaged over 100 runs of the running vaccine allocation experiment.

can be seen in Figure 5.7.

5.2.1 Uniform sampling

A simplistic method involves allocating an equal number of pulls to each arm a until the budget is depleted, after which the arm with the highest estimated mean μ_* is selected. This technique is known as the Uniform sampling strategy, also sometimes referred to as Round-robin [140]. The primary disadvantage of this method is that it wastes valuable resources on arms that are increasingly likely to be suboptimal, as indicated by their lower estimated means compared to others. These resources could be more effectively utilized to enhance our confidence in the top-performing arms, i.e., those with similar high estimated means [140]. Since this strategy does not leverage any existing knowledge of the estimated means, Uniform sampling is classified as a pure exploration strategy [140]. The detailed algorithm is presented in Algorithm 4.

Algorithm 4: Uniform sampling

Input: A MAB with K arms, arm pull budget T
for $t \leftarrow 1$ **to** T **do**
  $a^{(t)} \leftarrow t \mod K$ `// select next arm`
  $r^{(t)} \leftarrow$ Pull arm $a^{(t)}$
  $n_{a^{(t)}} \leftarrow n_{a^{(t)}} + 1$ `// increment pull count`
  Update $\hat{\mu}^{(t)}_{a^{(t)}}$ with $r^{(t)}$
return $\arg\max_{a \in A} \hat{\mu}^{(T)}_{a^{(T)}}$

5.2.2 Upper confidence bound

The UCB algorithm that was previously introduced within the regret minimization setting in Section 5.1.2 can also be adapted for use within the best-arm identification setting [21, 140]. To achieve this, the amount of exploration conducted by the bandit needs to be increased.

Instead of relying solely on pure exploration, like in the Uniform sampling algorithm, it is advantageous to leverage the knowledge accumulated from previous arm pulls. One of the most renowned algorithms for addressing MABs is the UCB algorithm, which adeptly balances exploration and exploitation using a frequentist approach [22, 140]. UCB operates on the principle that actions frequently selected are unlikely to exhibit significant changes in their estimated means. Conversely, it prioritizes actions with high potential rewards: those selected infrequently enough that, based on their current estimated means, they still have the potential to be optimal. This is all identical to the previously discussed implementation of the UCB algorithm for regret minimization for which pseudo code can be found in Algorithm 2. At each time t, an arm $a^{(t)}$ is selected as follows:

$$a^{(t)} = \arg\max_k \left(\hat{\mu}_k^{(t)} + \kappa \sqrt{\frac{\ln(t)}{n_k^{(t)}}} \right)$$

where k is the number of arms, $\hat{\mu}_k^{(t)}$ is the current estimated mean of arm a_k, and $n_k^{(t)}$ is the number of times arm a_k has been pulled [110, 140]. The hyperparameter κ can be used to tune the exploration rate of the parameter. Higher values for κ being more suitable for the best-arm identification setting [22]. The two main drawbacks of the UCB algorithm are the need of hyperparameter tuning for κ and the inability to incorporate prior knowledge that might be available on the arms' reward distributions [140].

5.2.3 Top-two Thompson Sampling

In contrast to the frequentist approach employed by UCB, a Bayesian approach can be utilized, which inherently incorporates prior knowledge [170, 140]. A prominent Bayesian algorithm for identifying the best arm is known as Top-two Thompson Sampling (TTTS) [151, 140]. The primary objective of TTTS is to differentiate the best arm from the second-best arm. The greater this distinction, the higher the confidence that the estimated best arm is indeed the optimal one [151, 140]. Since the other arms are inferior to the second-best arm, their ranking is not relevant within this setting, and thus, allocating arm pulls from the budget to them should be avoided.

Initially, analogously to the previously discussed Thompson Sampling algorithm, a prior probability distribution is imposed to represent the belief about the expected rewards of each arm, which is subsequently revised to a posterior probability distribution upon the acquisition of reward observations. At every timestep t within the sampling process, a sample is drawn from the posterior distribution associated with each arm of the bandit. It is at this point that the approach taken by the TTTS algorithm starts to deviate from the previously discussed Thompson Sampling algorithm.

Algorithm 5: Top-two Thompson Sampling

Input: A MAB with K arms, a prior $\pi(\cdot)$ and history $\mathcal{H}^{(0)} = \emptyset$, arm pull budget T
for $t \leftarrow 1$ **to** T **do**
  $\tilde{\boldsymbol{\mu}}^{(t)} \sim \pi(\cdot|\mathcal{H}^{(t-1)})$
  $b \sim \mathcal{B}(0.5)$
  if $b = 1$ **then**
    $a^{(t)} = \Psi_1(\tilde{\boldsymbol{\mu}}^{(t)})$
  else
    while $\Psi_1(\tilde{\boldsymbol{\mu}}^{(t)}_{alt}) = \Psi_1(\tilde{\boldsymbol{\mu}}^{(t)})$ **do**
      $\tilde{\boldsymbol{\mu}}^{(t)}_{alt} \sim \pi(\cdot|\mathcal{H}^{(t-1)})$ `// resample until a different best arm is found`
    $a^{(t)} = \Psi_1(\tilde{\boldsymbol{\mu}}^{(t)}_{alt})$
  $r^{(t)} \leftarrow$ Pull arm $a^{(t)}$
  $\mathcal{H}^{(t)} \leftarrow \mathcal{H}^{(t-1)} \cup \{a^{(t)}, r^{(t)}\}$
return $\arg\max_{a \in A} \hat{\mu}^{(T)}_{a^{(T)}}$

To decide if it should pull the best arm, TTTS samples from a Bernoulli distribution with the probability of success equal to 0.5 [140, 151]. If this sample is a success, the best arm is pulled. Otherwise, if the sample is not a success, TTTS needs to resample until it finds a best arm different from the original best arm. Just like with the normal Thompson sampling algorithm, the relative order between arms is computed using the ranking operator Ψ.

Initially, and assuming that our beliefs contain no information about the reward distributions, each arm is equally likely to be selected as the best. However, as the arms are pulled and the algorithm progresses, the belief distributions become increasingly informative, and the likelihood that the highest-ranked arm corresponds to the optimal arm also increases [140]. This process is repeated as long as there are arm pulls remaining in the budget T. Finally, the arm associated with the belief distribution with the highest mean is identified as the best arm and returned to the user. Pseudocode for TTTS can be found in Algorithm 5.

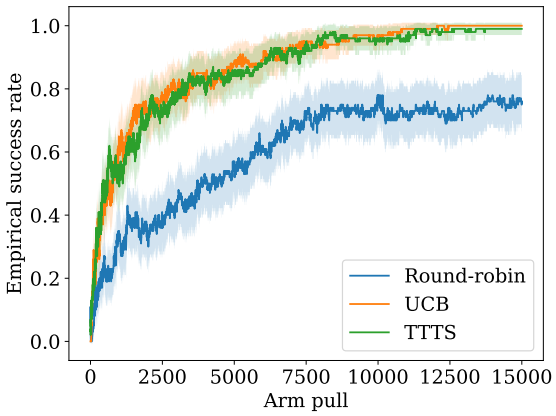


Figure 5.7: Empirical success rate of the recommendations made by the best-arm identification algorithms after each arm pull. Average over 100 runs of 15,000 arm pulls, together with the 95% confidence interval around the mean.

Chapter 6

Multi-objective multi-armed bandits

In this chapter, the *multi-objective multi-armed bandits* (MOMABs) framework proposed in [51] is introduced. This is the multi-objective extension of the single objective multi-armed bandits (MABs) that were discussed in the previous section. The need for multi-objective multi-armed bandits arises from the fact that inherently, the real world presents problems with several distinct and possibly conflicting objectives.

One of the principal distinctions between the single-objective MABs previously discussed and the MOMABs explored in this chapter lies in the stochastic reward's nature observed by the agent upon playing an arm $a \in \mathcal{A}$. In the context of single-objective MABs, at each time step t, the observed reward was a scalar $r^{(t)}$. Conversely, in MOMABs, the observed reward manifests as a vector $\mathbf{r}^{(t)} \in \mathbb{R}^D$, where D denotes the number of objectives [51, 140]. This vector is commonly referred to as the reward vector. Analogous to the scalar reward in the single-objective scenario, which was derived from a stationary reward distribution, the reward vector for MOMABs is generated by sampling from a multi-dimensional stationary reward distribution encompassing D dimensions, each corresponding to an objective [51, 140, 199]. This process yields a reward vector comprising a stochastic reward for each objective. To build intuition, an example difference in respective reward distributions is visualized in Figure 6.1. As the MOMAB framework was introduced by Drugan and Nowe in [51], the rest of this section is based on [51], using slightly modified notation to be consistent with the rest of this dissertation, but otherwise respecting the original structure and proposed terminology.

6.1 Ordering relations for reward vectors

In the analysis of scalar rewards, the identification of the optimum can be straightforwardly facilitated through the application of the existing partial order relation among real numbers [51, 199]. However, the identification of the optimal arm becomes more complex when addressing reward vectors. The absence of a partial order precludes a definitive best among multiple arms. To navigate this complexity, two predominant order relations are typically employed: (i) scalarization functions [54], including linear and Chebyshev functions, and (ii) the Pareto partial order [206], which facilitates the direct maximization of reward vectors within the multi-objective reward

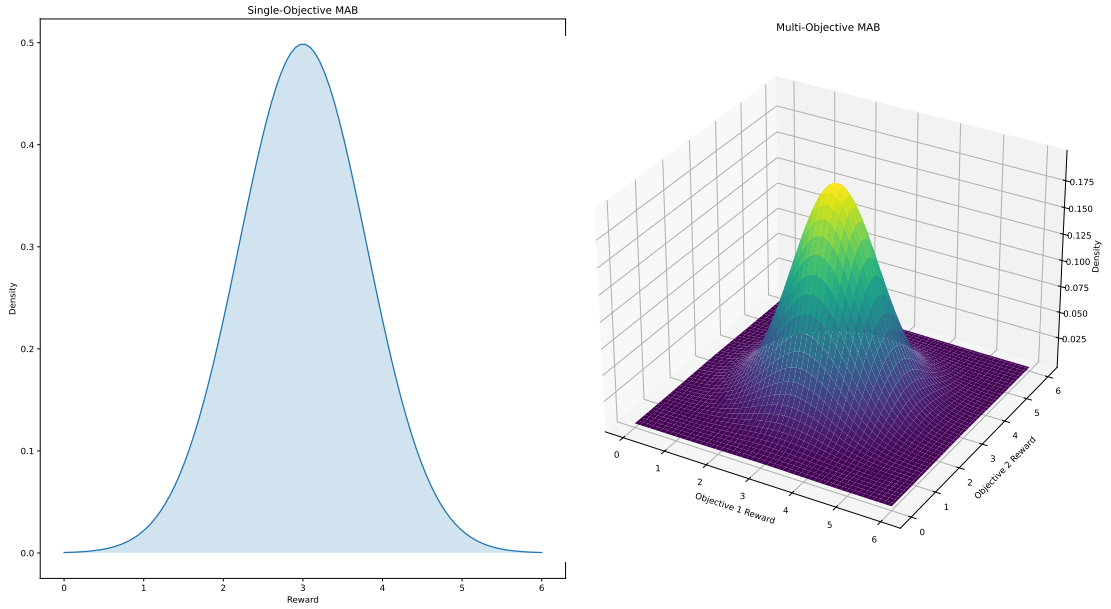


Figure 6.1: Graphical representation of possible reward distributions for a single-objective MAB (left panel) and a multi-objective MAB (right panel). The presented possible reward distribution for the single-objective case is a Gaussian distribution with $\mu = 3$ and $\sigma^2 = 0.8$. The multi-objective case is identical but uses a 2-dimensional multivariate Gaussian distribution with the same mean and variance for both dimensions.

space [51].

For conceptualizing these order relations in the context of reward vectors, consider a K-armed bandit scenario. Within a multi-objective framework, the expected reward for each arm a, denoted as $\boldsymbol{\mu}_a = (\mu_a^1, \cdots, \mu_a^D)$, is inherently multi-dimensional, where D represents a predetermined number of objectives. This scenario posits a general case where a reward vector may surpass another in one objective yet fall short in another, indicating the potential conflicting objectives [51, 199].

6.1.1 Pareto partial order

In the context of the Pareto partial order [206], "a reward vector $\mathbf{r}_1$ is considered better than, or dominating, another reward vector $\mathbf{r}_2$, $\mathbf{r}_2 \prec \mathbf{r}_1$, if and only if there exists at least one dimension j for which $\mathbf{r}_2^j < \mathbf{r}_1^j$, and for all other dimensions o we have $\mathbf{r}_2^o \leq \mathbf{r}_1^o$" [51]. $\mathbf{r}_1$ weakly dominates $\mathbf{r}_2$, $\mathbf{r}_2 \preceq \mathbf{r}_1$, "if and only if for all dimensions j, we have $\mathbf{r}_2^j \leq \mathbf{r}_1^j$" [51]. A reward vector $\mathbf{r}_1$ is incomparable with another reward vector $\mathbf{r}_2$, $\mathbf{r}_2 \parallel \mathbf{r}_1$, "if and only if there exists at least one dimension j for which $\mathbf{r}_2^j < \mathbf{r}_1^j$, and there exists another dimension o, for which $\mathbf{r}_2^o > \mathbf{r}_1^o$" [51]. Finally, $\mathbf{r}_1$ is non-dominated by $\mathbf{r}_2$, $\mathbf{r}_2 \not\succ \mathbf{r}_1$, "if and only if there exists at least one dimension j for which $\mathbf{r}_2^j < \mathbf{r}_1^j$" [51]. These Pareto relationships are summarized in Table 6.1.

Consider the *Pareto optimal reward set* $\mathcal{O}^*$ as the collection of reward vectors that are not dominated by any other reward vectors [51]. Let the *Pareto optimal set of actions* $\mathcal{A}^*$ denote

Table 6.1: Relations between reward vectors. Reproduced from [51].

Relationship	Notation	Formalised Relationships
μ dominates ν	$\nu \prec \mu$	$\exists j, \nu^j < \mu^j$ and $\forall o, j \neq o, \nu^o \leq \mu^o$
μ weakly dominates ν	$\nu \preceq \mu$	$\forall j, \nu^j \leq \mu^j$
μ is incomparable with ν	$\nu \| \mu$	$\nu \not\succ \mu$ and $\mu \not\succ \nu$
μ is non-dominated by ν	$\nu \not\succ \mu$	$\nu \prec \mu$ or $\nu \| \mu$

the set of actions corresponding to reward vectors in $\mathcal{O}^*$ [51]. In this context:

$$\forall \mathbf{r}_\ell \in \mathcal{O}^*, \quad \text{and} \quad \forall \mathbf{r}_o \notin \mathcal{O}^*, \quad \text{we have} \quad \mathbf{r}_\ell \neq \mathbf{r}_o$$

The entire set of Pareto optimal rewards are incomparable:

$$\forall \mathbf{r}_\ell, \mathbf{r}_o \in \mathcal{O}^*, \quad \text{it holds that} \quad \mathbf{r}_\ell \parallel \mathbf{r}_o$$

We further posit that it is infeasible to ascertain a priori the superiority of any specific arm in $\mathcal{A}^*$ over others. Consequently, it is posited that all reward vectors within the Pareto optimal reward set $\mathcal{O}^*$ are equally optimal [51, 199].

6.1.2 Scalarization functions

The transformation of a multi-objective environment into a single-objective framework can be achieved through the use of scalarization functions [140, 51, 199, 125]. As single-objective environments typically yield a single optimum, multiple scalarization functions are needed to create a diverse array of elements comprising the Pareto optimal set [51]. Owing to its simplicity, the linear scalarization function has attained popularity. It is characterized by the allocation of weights to each component of the reward vector, with the sum of these weighted components constituting the result:

$$f(\mathbf{r}_i) = \omega^1 \cdot r_i^1 + \cdots + \omega^D \cdot r_i^D, \forall i.$$

This weighted sum can also be defined as the dot product $\mathbf{r}_i \cdot \boldsymbol{\omega}$ between the reward vector and a predefined weight vector. The set of predefined weights used here can be written as $(\omega^1, \cdots, \omega^D)$, where $\sum_{j=1}^{D} \omega^j = 1$ [51]. A known disadvantage of using linear scalarization to transform the reward vector into a single scalar reward is its potential inability to find all arms belonging to a non-convex Pareto set [51, 199, 125]. In certain conditions however, Chebyshev scalarization can find all arms belonging to a non-convex Pareto set [122, 125]. Figure 6.2 visualizes such a case. This kind of scalarization was originally meant for minimization problems, but was adapted by [51] for use with MOMABs. Mathematically, the Chebyshev scalarization can be defined as:

$$f(\mathbf{r}_i) = \min_{1 \leq j \leq D} \omega^j \cdot (r_i^j - z^j), \forall i$$

In this formula, $z \in (z^1, \cdots, z^D)$ is a point that is dominated by all optimal reward vectors in $\mathcal{O}^*$ and is used as a reference point [51, 125]. This point z^j is the minimum of the current best rewards with a small positive value $\epsilon^j > 0$ subtracted [51]:

$$z^j = \min_{1 \leq j \leq D} r_i^j - \epsilon^j, \forall j$$

We identify the Pareto optimal set of actions ascertainable through linear scalarization as $\mathcal{A}_L^*$, and through Chebyshev scalarization as $\mathcal{A}_C^*$ [51]. The associated set of Pareto optimal reward

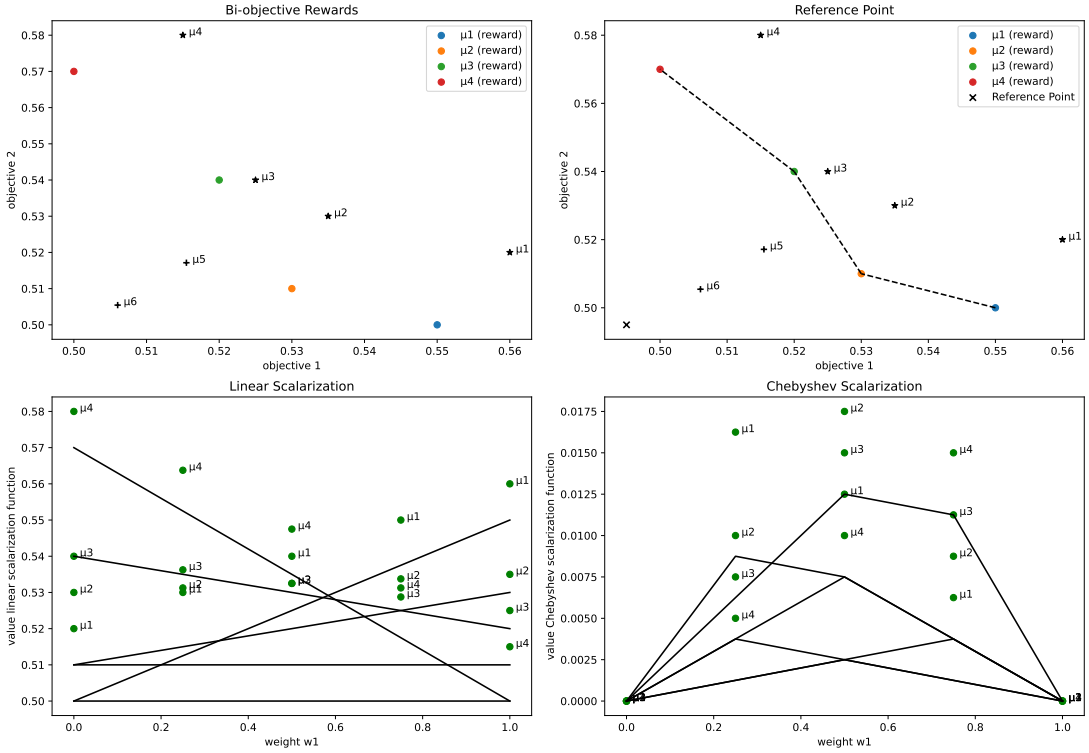


Figure 6.2: Example of both linear and Chebyshev scalarization where linear scalarization fails to identify all rewards in a non-convex set of optimal rewards $\mathcal{O}^*$. The top-left panel shows the true values of the (sub)optimal arms, as well as the observed reward vectors. The top right panel shows the non-convex nature of $\mathcal{O}^*$, as well as the reference point z used for Chebyshev scalarization. The bottom panels show the values of linear scalarization and Chebyshev scalarization. The weights used for each scalarization are defined as $(\omega_1, 1-\omega_1)$. From the scalarized points above the scalarization functions, it can be seen that $\mathcal{O}_L^* = \{\mu_1, \mu_4\}$ and $\mathcal{O}_C^* = \{\mu_1, \mu_2, \mu_3, \mu_4\} = \mathcal{O}^*$. Reproduced from [51].

vectors is denoted $\mathcal{O}_L^*$ for linear scalarization and $\mathcal{O}_C^*$ for Chebyshev scalarization [51]. Note that the set of optimal actions may vary depending on the scalarization technique employed.

Drugan and Nowe [51] argue that in typical scenarios where the shape of the Pareto front remains undetermined, it is useful to experiment with various weight combinations within a scalarized multi-objective MAB framework [51].

When employing a linear scalarization function, it becomes apparent that not every reward vector within any Pareto optimal reward set is accessible through this scalarization [51, 125]. As a result, there is a persistent positive regret when contrasting $\mathcal{O}^*$ and $\mathcal{O}_L^*$. This algorithm's inherent unfairness with respect to selecting each optimal arm equally often escalates with more frequent arm selections, as an arm deemed optimal in $\mathcal{A}_L^*$ is chosen more often, while other arms, equally optimal in $\mathcal{A}^*$, are less frequently recognized and engaged [51].

When considering the Chebyshev scalarization function, Drugan and Nowe [51] acknowledge the potential to pinpoint all solutions in $\mathcal{A}^*$ by adjusting the reference points [51, 125]. However, the

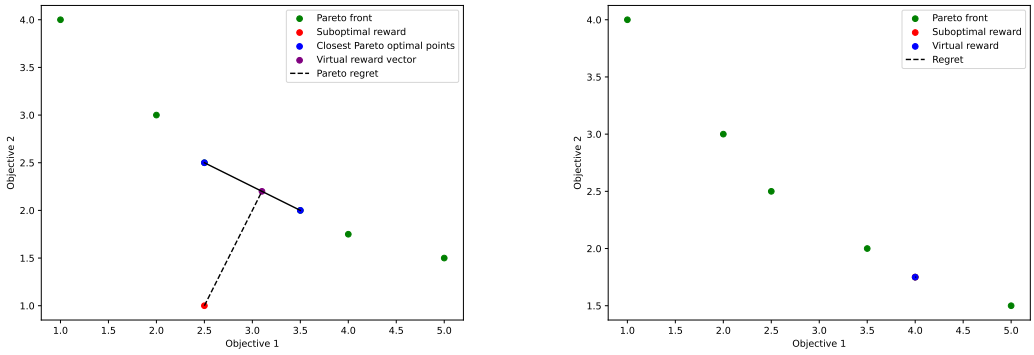
strategy for discovering these reference point sets remains unspecified, and their identification becomes a new challenge, especially when aiming to diminish the unfairness regret (see Section 6.2.3). If the Chebyshev multi-objective MAB reveals more Pareto optimal arms than the linear approach, then the former boasts a reduced unfairness compared to its linear counterpart [51].

6.2 Performance metrics for MOMAB regret minimization

The goal of multi-objective MABs is to minimize the regret in all objectives by selecting all arms from the Pareto optimal set equally frequently using a policy π [51]. This section suggests three metrics to assess multi-objective MABs effectively. The *Pareto regret metric* quantifies the gap between a reward vector and the Pareto optimal set [51]. In contrast, the *scalarized regret metric* gauges the disparity between the peak value of a scalarized function and the scalarized outcome of an arm [51]. Lastly, the *unfairness metric* associates with the variability in selecting all the optimal arms, since all arms $a \in \mathcal{A}^*$ should be played with the same frequency [51].

6.2.1 Pareto regret

Drugan and Nowe [51] argue that, from an intuitive point of view, "a regret metric measures how far a suboptimal reward vector $\mathbf{r}_i$ is from being an optimal arm itself".



(a) Illustration of the Pareto regret with a suboptimal reward vector, showcasing the computed virtual reward vector and the regret distance ε_i^*.

(b) Demonstration where the observed reward is optimal, resulting in zero Pareto regret.

Figure 6.3: Visualization of Pareto regret in a multi-objective multi-armed bandit scenario. The left subfigure (6.3a) depicts the scenario with a non-optimal arm and its associated regret, while the right subfigure (6.3b) presents a case with no regret when the reward is Pareto optimal.

Taking inspiration from the ε-dominance concept, which gauges the discrepancy between $\mathbf{r}_i$ and the Pareto optimal reward set $\mathcal{O}^*$, Drugan and Nowe propose a novel regret metric [51]. This metric assesses the regret associated with Pareto optimal rewards. To estimate the nearest distance between $\mathcal{O}^*$ and $\mathbf{r}_i$, they introduce the concept of a virtual reward vector $\boldsymbol{\nu}_{i,\varepsilon}$, which cannot be directly compared with any vector in $\mathcal{O}^*$ [51]. They define $\boldsymbol{\nu}_{i,\varepsilon}$ by enhancing $\mathbf{r}_i$ with a positive ε in every objective, yielding the so-called virtual reward vector $\boldsymbol{\nu}_{i,\varepsilon}$ [51]. Thus:

$$\boldsymbol{\nu}_{i,\varepsilon} = \mathbf{r}_i + \varepsilon \quad \text{where} \quad \forall j,\ \nu_{i,\varepsilon}^j = r_i^j + \varepsilon, \quad \varepsilon > 0$$

The arm i's optimal virtual reward, $\boldsymbol{\nu}_i^*$, is defined as having the smallest ε such that $\boldsymbol{\nu}_{i,\varepsilon}$ is incomparable to all rewards in $\mathcal{O}^*$ [51]. Formally:

$$\boldsymbol{\nu}_i^* \leftarrow \min_{\varepsilon \to \infty} \boldsymbol{\nu}_{i,\varepsilon}, \quad \text{for which} \quad \forall \mathbf{r}_\ell^* \in O^*,\ \boldsymbol{\nu}_i^* \nparallel \mathbf{r}_\ell^*$$

The particular ε that satisfies $\boldsymbol{\nu}_i^*$ is denoted as ε_i^* [51].

The regret associated with observing $\mathbf{r}_i$ is thus the distance between the arm's virtual optimal reward vector, $\boldsymbol{\nu}_i^*$, and its actual reward vector, $\mathbf{r}_i$. Therefore:

$$\Delta_i = \boldsymbol{\nu}_i^* - \mathbf{r}_i = \varepsilon_i^*$$

From this formalisation, it can be noted that the regret of the Pareto optimal arms will be equal to zero as the optimal reward vector itself will be the same as the virtual reward vector [51].

6.2.2 Scalarized regret

Apart from the previously discussed Pareto regret, Drugan and Nowe [51] also present a regret measure that may be utilized alongside scalarization functions. The *scalarized regret* for a chosen scalarized function f^j and arm i is denoted as:

$$\Delta_i^j \stackrel{\text{def}}{=} \max_{k \in \mathcal{A}} f^j(\mathbf{r}_k) - f^j(\mathbf{r}_i)$$

This scalarized regret represents the gap between the highest scalarization function value over the set of arms $\mathcal{A}$ and the scalarized value for a specific arm i [51].

For an arm i, the *linear scalarized regret* at time t is defined as:

$$\Delta_i^j = \sum_{t=1}^{D} \omega_j^t \cdot (r^{*t} - r_i^t), \quad \text{where} \quad f^j(\mathbf{r}^{*t}) = \max_{k \in \mathcal{A}} f^j(\mathbf{r}_k^t)$$

and f^j is a linear scalarization function characterized by a set of weights $\boldsymbol{\omega}_j = \{\omega_j^1, \ldots, \omega_j^D\}$ [51]. The *Chebyshev scalarized regret* for an arm i using the Chebyshev scalarization function f^j is given by:

$$\Delta_i^j = \max_{1 \le t \le D} \omega_j^t \cdot (r^{*t} - r_i^t), \quad \text{where} \quad f^j(\mathbf{r}^{*t}) = \max_{k \in \mathcal{A}} f^j(\mathbf{r}_k^t)$$

Demonstrating that the maximum value for any weight combination in both linear and Chebyshev functions aligns with one of the Pareto optimal arms is relatively straightforward [51]. Hence,

$$\forall j \in S, \exists! i \in \mathcal{A}^* \text{ such that } f^j(\mathbf{r}_i^*) = \max_{k \in \mathcal{A}} f^j(\mathbf{r}_k)$$

The set S represents the weights used. Although the concept of scalarized regret appears straightforward, it is not ideally suited for the aims of MOMABs, as it accumulates independent regrets rather than minimizing the regret of a comprehensive multi-objective approach across all objectives [51]. Scalarization functions do not consistently recognize all the optimal arms in $\mathcal{A}^*$. Consequently, a metric that encourages the exploration of every optimal arm is needed.

6.2.3 Unfairness regret

As defined in [51], consider $N_{i^*}(t)$ as the count of how often an optimal arm i^* is activated up to time t, and $\mathbb{E}[N_{i^*}(t)]$ the expected frequency of selecting an optimal arm [51]. The *unfairness*

[51] of a multi-objective multi-armed bandit algorithm is then defined as the *variance* in the choice distribution over the arms in $\mathcal{A}^*$, given by

$$\phi = \frac{1}{|\mathcal{A}^*|} \cdot \sum_{i^* \in \mathcal{A}^*} \left(N_{i^*}(t) - \mathbb{E}[N_{i^*}(t)]\right)^2.$$

If the algorithm fairly allocates opportunities to optimal arms, ϕ approaches 0. A multi-objective strategy that favors a subset of $\mathcal{A}^*$ will manifest a larger ϕ. This variance is reminiscent of risk metrics employed in financial analyses, offering a gauge for equitable arm selection by a multi-objective MAB algorithm, ensuring balanced exploration among the best-performing arms [51].

Another, alternative, definition for the unfairness regret that is slightly different from the one proposed in [51], but identical in spirit, is explained by Yahyaa and Manderick [199]. In their paper, the unfairness regret metric utilizes the *Shannon entropy* R_{SE}, which quantifies the randomness in the selection frequencies of the optimal arms from the Pareto front $\mathcal{A}^*$ [199]. A higher entropy value signifies greater disorder. The Shannon entropy at a particular time step t, $R_{SE}(t)$, is defined as:

$$R_{SE}(t) = -\frac{1}{N_{\mathcal{A}^*}(t)} \sum_{i^* \in \mathcal{A}^*} p_{i^*}(t) \ln(p_{i^*}(t)),$$

where $p_{i^*}(t) = N_{i^*}(t)/N(t)$ represents the relative selection frequency of the optimal arm i^* at time step t, with $N_{i^*}(t)$ indicating the number of times arm i^* has been activated, $N(t)$ the total activations for all arms, and $N_{\mathcal{A}^*}(t)$ the sum of activations for all optimal arms at time step t [199].

Part III

Literature review

Chapter 7

Dengue epidemic modelling

In this chapter of the literature review, we delve into the extensive body of literature on models for dengue fever epidemiology. Over the past two decades, we have witnessed a significant surge in interest and research on dengue fever, leading to the development of numerous epidemiological models. Consequently, the vast volume of proposed models, papers, studies, and books on this topic presents a considerable challenge in identifying the most pertinent models while maintaining the focus and scope of this dissertation. Indeed, a comprehensive review tracing the evolution of dengue epidemic models since their inception in the 1970s could itself constitute an entire study.

A key resource that underpins the literature review presented here is the paper titled "Mathematical models for dengue fever epidemiology: A 10-year systematic review", which surveys dengue epidemic models from 2010 to 2021 [5]. This paper provides an extensive overview of the existing body of work, offering a valuable foundation for our exploration. However, it is important to note that the authors of this review exhibit a noticeable bias towards their own models and publications over the decade in question. To ensure a more balanced perspective, this dissertation also incorporates additional literature from the same period that is not covered in the 10-year review.

Moreover, since the systematic review concludes in 2021, it does not encompass the most recent developments in dengue epidemic modeling. Given that this dissertation was written in 2023-2024, there was a clear need to include state-of-the-art papers and models published after 2021. To address this, a comprehensive database of publications was created from two major scientific libraries: ScienceDirect and PubMed. ScienceDirect was queried using the term "dengue model", while PubMed searches included "agent-based dengue model" and "compartment dengue model". All relevant papers published post-2021 were added to this database.

This rigorous approach resulted in a collection of 64 papers on dengue epidemic models, which were reviewed as thoroughly as possible within limited time and extremely extensive scope of this dissertation. The aim was to incorporate as much relevant literature as possible, while ensuring the review remained concise and within the scope of this dissertation.

7.1 Mathematical models

Epidemiological studies have long utilized mathematical models as essential tools. These models serve as structured methods to articulate hypotheses regarding the interactions between hosts,

vectors, and pathogens in epidemics. They play a crucial role in understanding and forecasting the trends of infectious diseases under varied scenarios, and are pivotal in assessing the effectiveness of public health strategies aimed at disease management [5].

The mathematical modeling of dengue fever's epidemiological patterns traces back to the 1970s [63]. Dengue fever, a globally significant public health issue, is a viral disease transmitted by mosquitoes. Most cases are either asymptomatic or mild. The disease can lead to severe conditions due to the *Antibody-Dependent Enhancement* (ADE) phenomenon. This phenomenon exacerbates the disease, as pre-existing antibodies, instead of neutralizing, intensify the subsequent infection [5].

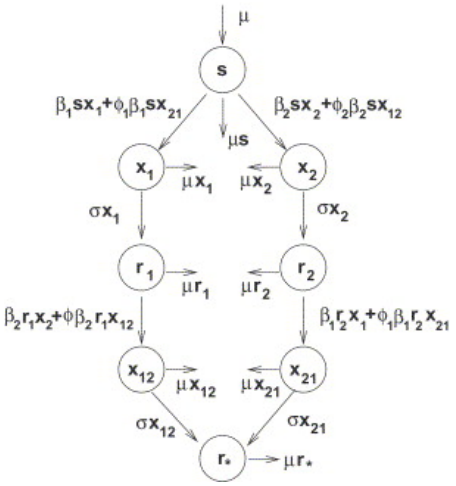
Mathematical models for dengue fever strive to include various elements of the disease and its vectors. This inclusion often results in models exhibiting complex behaviors, even at their most fundamental levels. These models have been designed to investigate various factors, such as the simultaneous presence of multiple viral strains, the immunological pathways influencing disease severity, and the effects of vaccination programs [5].

7.2 Host-to-host Transmission Models

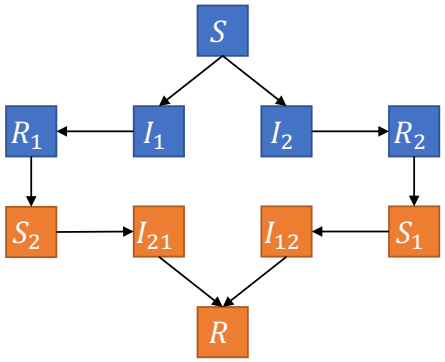
Models depicting multi-serotype host-to-host dynamics of dengue fever have been developed, extending the traditional susceptible-infected-recovered (SIR) frameworks discussed in Section 4.1.1 [5]. This methodology emphasizes the multi-serotype characteristic of the disease and its impact on host populations. It incorporates the effects of vector dynamics indirectly in the SIR-type model, such as seasonal variations in infection rates, while not explicitly representing mosquito dynamics [5]. From a mathematical perspective, this approach is valid, considering that in vector-borne illnesses, the epidemiological dynamics of human hosts typically operate on a larger timescale compared to that of the disease-transmitting mosquitoes, a discrepancy attributed to the significant difference in their respective lifespans [5].

Rocha et al. [145] conducted a study to explore the extent to which the rapid lifecycle of mosquitoes and their capability to transmit viruses are influenced by the slower dynamics of human infection and immune response. They examined a compartmental model comprising susceptible (S), infected (I), and recovered (R) humans, as well as susceptible (U) and infected (V) mosquitoes [145]. The findings revealed that the human timescale is the predominant factor in understanding long-term incidence data, with the mosquito dynamics exerting only minor perturbations [145]. This analysis of the SIRUV model aligns qualitatively with findings from a previously studied SISUV model. The main difference between the two is that the SIRUV model includes a recovered (R) class modeling temporary cross-immunity, whereas in the SISUV model individuals become susceptible (S) again immediately after their infection (I) has ended. This suggests a common characteristic of vector-borne diseases in general [146, 62]. These biological insights are valuable given that existing vector control strategies for dengue are only minimally effective, with broad-scale implementation proving challenging to attain and even more difficult to maintain [126]. Consequently, Aguiar et al. [5] argue that employing simplified mathematical models to plan the implementation of interventions against a complex, multi-strain pathogen with intricate transmission dynamics, such as dengue, could significantly enhance the practical predictability of the dynamical system [2]. We note that these arguments apply to the prediction of long-term disease case data [145]. We also note that in some cases it might be desirable to simulate the disease dynamics with a higher degree of accuracy by incorporating vector dynamics. Furthermore, for the simulation of vector control measures and effects like seasonality, vector dynamics need to be modeled explicitly.

In an effort to explain and understand the erratic patterns of dengue epidemics, mathematical models have been devised to represent the transmission dynamics of dengue viruses, focusing on elements such as multi-strain interactions, antibody-dependent enhancement (ADE), and temporary cross-immunity (TCI) [5]. A notable multi-strain mathematical model by Ferguson et al. [58], designed to examine the impact of ADE on the dynamics of dengue transmission, revealed deterministically chaotic behavior when a high level of infectivity for secondary dengue infections was used. Notably, this model did not account for the period of cross-immunity, allowing for the possibility of co-infection. Consequently, individuals could be concurrently classified in various compartments of infection, each corresponding to different dengue strains. Billings et al. [26] described chaotic desynchronization in a multi-serotype dengue model incorporating ADE [5]. Similar to the model by Ferguson et al., this model also did not incorporate the cross-immunity period. However, it differed in that cross-infection was not feasible as long as an individual was infected. Consequently, this led to a scenario where all compartments within the model were distinct and separate shown in Figure 7.1a.



(a) Visualization of the compartmental model proposed by Billings et al. in [26] to study the chaotic nature of dengue epidemics. Figure from [26].



(b) Visualization of the compartmental model proposed by Aguiar et al. in [121, 15] as an extension of the one proposed by Billings et al. in [26].

Figure 7.1: Comparison of compartmental models by Billings et al. and Aguiar et al.

Aguiar et al. [121, 15] conducted research on a minimalist two-infection dengue model shown in Figure 7.1b, which was an extension of the models initially proposed and analyzed in studies by Ferguson et al. [58] and Billings et al. [26]. In this model, it was assumed that a secondary dengue infection could only be caused by a different serotype from the one responsible for the primary infection. To account for this, TCI was integrated into the model through the introduction of additional compartments. These compartments were designated for individuals recovering from a primary infection who would become susceptible again after a brief period of cross-immunity. Remarkably, the inclusion of a TCI period in these models revealed a new chaotic window in a broader and unexpected parameter range. This range even accounted for reduced infectivity in secondary infections [5]. Previously, it was thought that significantly higher transmission rates for secondary infections were required to produce complex dynamics similar to oscillations observed in empirical data [5]. However, this new finding suggested that such assumptions could

be significantly relaxed. This discovery highlighted the greater significance of deterministic chaos in multi-strain models than previously recognized. It opened new avenues for analyzing existing data sets, suggesting that the dynamics of dengue transmission might be more complex and nuanced than previously understood [5].

In 2011, the minimalist two-infection dengue model initially introduced by Aguiar et al. [15] was further developed to include seasonality factors, thereby simulating the impact of vector dynamics [7]. This seasonal adaptation of the model exhibited complex dynamics and demonstrated notable alignment with empirical data. This congruence was particularly evident when a brief period of cross-immunity was integrated alongside the effects of ADE [7]. This approach was reinforced by epidemiological observations that indicated an increased risk of severe disease in secondary dengue infections. The model's assumptions included the premise that individuals with a primary dengue infection would likely exhibit asymptomatic or mild symptoms, maintaining mobility and the potential to transmit the disease [5, 7]. Conversely, those experiencing a secondary infection with a different dengue serotype were presumed more likely to develop severe symptoms, necessitating hospitalization. This hospitalization, in turn, was hypothesized to reduce their likelihood of transmitting the disease compared to individuals with a primary infection [5, 7].

The inclusion of stochastic elements became necessary to account for the variations observed in certain dengue data sets, unveiling a situation where noise and a complex deterministic framework interact intensively [14]. Stollenwerk et al. [166] revisited the framework for parameter estimation in dynamical systems similar to biological populations and applied this methodology to calibrate the dengue model proposed in [14]. The application of this model resulted in broad likelihood profiles for some parameters, indicating that the maximum likelihood iterated filtering technique presents an encouraging approach for inferring parameter values from dengue case notifications [5]. This technique enhances the understanding of the disease dynamics by incorporating the variability and unpredictability inherent in real-world data.

Utilizing bifurcation theory, Kooi et al. [104] conducted an analysis and comparison of three multi-strain dengue models previously proposed in studies [58, 26, 7], providing insights into the origins of their long-term dynamic behaviors. In their approach, they maintained a consistent parameter set across all models, while varying the duration of the cross-immunity period and the ADE factor as bifurcation parameters [5]. Their analysis identified not only endemic equilibria and periodic solutions but also chaotic behavior emerging through various routes. To quantify this complex behavior in the models, the calculation of Lyapunov exponents was employed [5]. A particularly notable dynamic aspect discovered by Kooi et al. [104] was the occurrence of a torus bifurcation as a pathway to chaotic dynamics in the model presented in [7]. This type of dynamical behavior had not been previously described in the field of epidemiology, marking a significant advancement in the understanding of epidemiological model dynamics [5]. The identification of a torus bifurcation in this context demonstrates the intricate and often unpredictable nature of disease transmission dynamics, especially in the case of complex infections like dengue.

The integration of the cross-protection assumption with the ADE effect has been modeled in various ways. Despite the inclusion of TCI in relatively complex models, the ADE effect consistently increased transmissibility or susceptibility in secondary infections [92, 195, 169, 138, 182]. Woodall and Adams [187] assumed partial cross-protection following a primary dengue infection, whereas Reich and colleagues [138] proposed a model where the enhancement factor influenced individual susceptibility, suggesting that individuals immune to one serotype would be more susceptible to a second infection [5, 138]. Upon evaluating their model against dengue data from a hospital in Thailand, they confirmed that models incorporating short-term cross-protection more accurately fit their data compared to models without it, estimating the optimal duration of

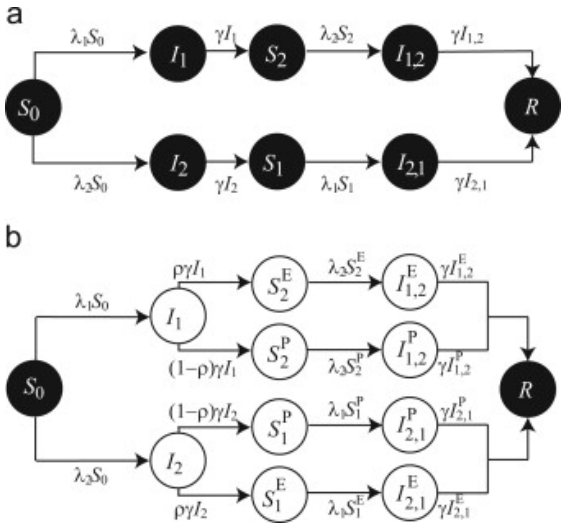


Figure 7.2: Figures from [187] showing (a) the structure for two-serotype SIR models with cross-protection or cross-enhancement and (b) the modified structure proposed in [187] incorporating partial cross-enhancement. Figure from [187].

the cross-protection period to be two years [5]. However, adding a serotype-specific transmission rate to their model reduced its fit to the data [5].

Aguiar et al. in studies such as [7, 8, 15], combined a brief period of temporary cross-immunity between primary and secondary infections. In their model, a secondary infection contributed less to the force of infection than a primary infection. They argue this was justified by the observation that disease severity and hospitalization reduced human interaction and, consequently, disease transmission [5]. However, it is important to note that, while human interactions play a role in the transmission of dengue epidemics, dengue is a vector-borne disease. A comparison of the basic two-strain dengue model, which differentiates between primary and secondary infections including TCI, with the four-strain model introducing the concept of multiple strain competition in dengue epidemics, revealed that the combination of TCI and ADE effects is a crucial driver of complex dynamics in the system, surpassing the impact of the specific number of dengue serotypes included in the model [8]. Until then, the TCI factor had not been actively explored, but it is now recognized as a vital component for developing a realistic dengue model [5]. Aguiar et al. [6] also detailed the impact of the number of subsequent infections versus the detailed number of dengue serotypes in the model framework, along with human immunological responses related to disease severity. In this survey, they compared extensions of the two-infection multi-strain model proposed in [8], clarifying the implications of additional compartments on model dynamics [5].

Chaotic dynamics were identified within the same parameter region of interest, which corresponds to the fluctuations observed in empirical data, for both two-strain and four-strain dengue models. Subsequently, the minimalistic two-strain model underwent further development to incorporate vaccination. This extension was aimed at assessing the impact of the sole licensed, albeit imperfect, dengue vaccine available at the time [12]. This inclusion of vaccination in the model reflects an essential step in understanding and predicting the dynamics of dengue transmission and control, particularly in the context of real-world vaccine efficacy and coverage [5].

Despite the relatively low occurrence of tertiary and quaternary dengue infections, models that

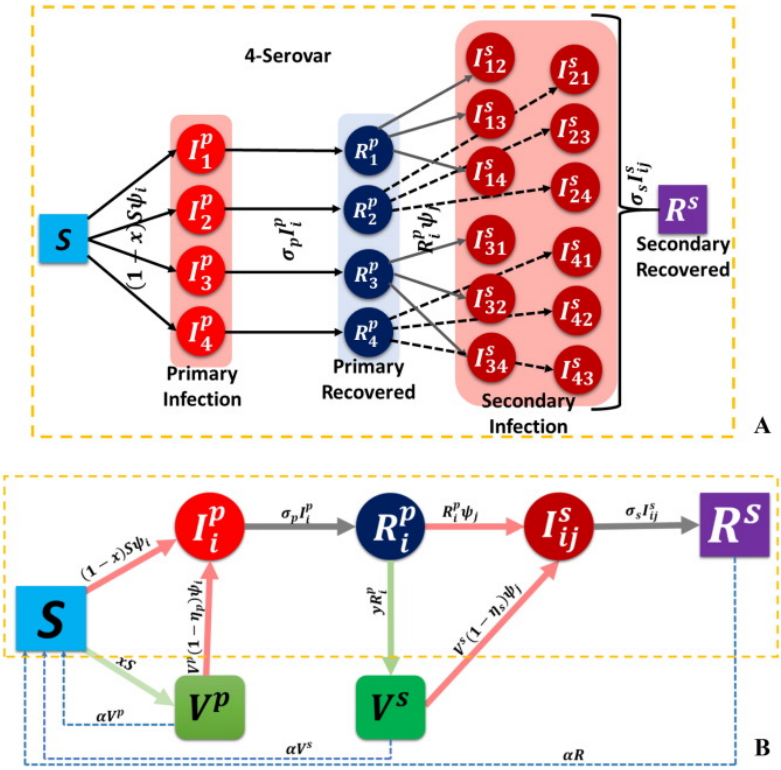


Figure 7.3: Compartmental model with support for vaccination proposed by Kabir et al. in 2020 in [92]. Their model allows for the investigation of the interactions of vaccination and ADE. Figure from [92].

account for third and fourth heterologous infections have frequently been developed. Wikramaratna et al. [183] devised a framework specifically to investigate the impact of third and subsequent infections on the epidemiology of dengue. Their findings suggested that the qualitative nature of the dynamical behavior in models, whether they include or exclude third and fourth infections, is largely similar. This observation implies that the fundamental dynamics of dengue transmission and its epidemiological patterns remain consistent even when the possibility of multiple infections beyond the primary and secondary stages is considered.

Many models examining multiple serotypes of dengue have traditionally presumed uniform infection rates across these serotypes. However, variation in transmission rates among different serotypes is addressed in literature, notably in [195]. This study explores varying transmission rates between serotypes, revealing that disparities among serotypes enhance the endurance of all strains. It is further noted that a specific value of the ADE factor optimally increases the probability of persistent serotype existence. Conversely, Kooi et al. [105] analyze the epidemiological differences among strains by varying the rates of infection force, while maintaining consistency in other epidemiological parameters. The resilience of a symmetric two-strain dengue fever model against such asymmetries is examined through bifurcation analysis [5]. This investigation indicates that, unlike symmetric models where a two-strain system perpetually prevails, asymmetric models can result in the endemic presence of a single strain. Furthermore, this strain asymmetry contributes to the stabilization of long-term dynamics, with chaotic patterns, necessary to

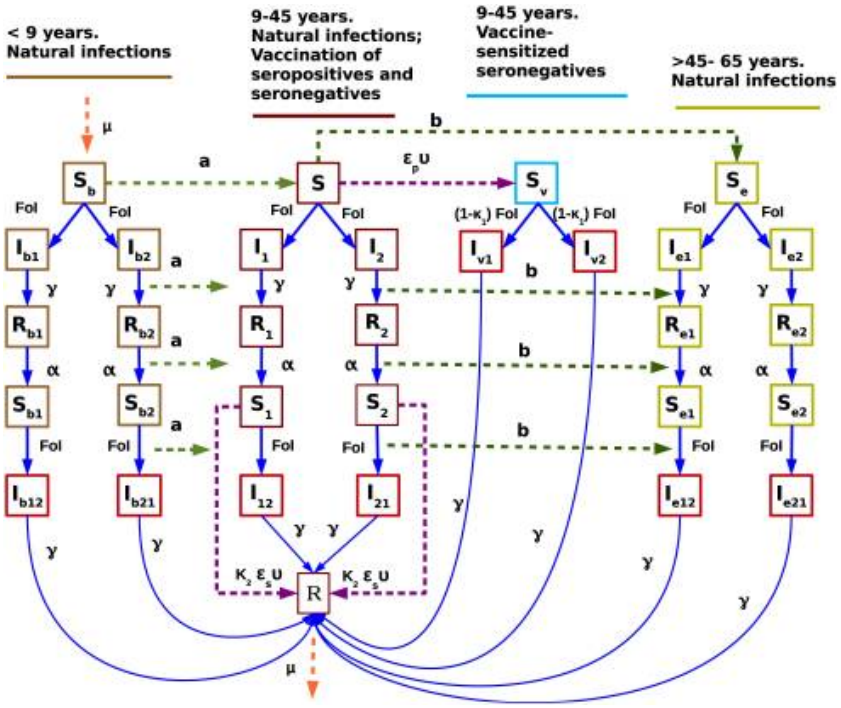


Figure 7.4: Minimalist two-strain model from [7] extended with support for different age groups and support for vaccination proposed in [12]. Figure from [12].

replicate the fluctuations seen in empirical observations, emerging only within limited parameter scopes [5].

Woodall and Adams [187] have developed an innovative method for dengue modeling that incorporates partial cross-enhancement in secondary infections. This approach involves the modification of three established models to include partial cross-enhancement: a foundational model without temporary cross-immunity (TCI) [58], a model integrating stochastic seasonality and co-infection [3], and a vector-host model featuring cross-temporary class with deterministic seasonality [7]. The authors posit that the multi-annual oscillations observed in dengue dynamics are not solely influenced by enhancement effects. Instead, these oscillations are a result of the interaction between enhancement and other contributing factors [5].

In parallel, Bosch et al. [169] have applied a pattern-oriented modeling (POM) strategy to evaluate various dengue models. These models are characterized by a blend of temporary cross-protective immunity, cross-enhancement, and seasonal driving forces [5]. The objective was to determine the models' effectiveness in reflecting the primary aspects of dengue dynamics. Their findings suggest that to accurately mimic observed patterns in environments with low seasonality, an extended period of cross-protection is necessary [169]. Conversely, in scenarios with pronounced seasonal influences, the incorporation of an Antibody-Dependent Enhancement (ADE) parameter becomes essential [169].

In their analysis of the impact of control strategies, Pandey and Medlock [133] utilized SIR-based deterministic models specific to dengue to examine the immediate and enduring outcomes following the introduction of vaccines [5]. The model's straightforward approach revealed potential

large temporary surges in the number of disease cases post-vaccination introduction, anticipated to manifest within a 15-year timeframe, assuming moderate levels of vaccine efficacy and coverage [5]. Despite vaccinations' potential to reduce the overall number of infections over an extended period, the authors highlighted the critical need to account for the likelihood of significant case spikes in the formulation of health policies. Notably, these findings were disseminated prior to the availability of any vaccine trial data [5]. Further, Aguiar et al. [12] devised an age-specific model, which, upon validation and calibration using existing vaccine trial data [78], indicated that a substantial decrease in hospital admissions would be feasible only if the vaccine were administered to individuals previously infected by dengue. Conversely, the model predicted a marked rise in hospital admissions if the vaccine were administered indiscriminately, including to those without prior exposure to the virus [5]. In their more recent investigation, Kabir et al. [92] introduced a model to assess the implications of Antibody-Dependent Enhancement (ADE) in the context of two distinct vaccination scenarios: one for seronegative individuals (primary vaccination) and another for seropositive individuals (secondary vaccination). Utilizing a vaccination game framework, the study explored the interplay between ADE and the cost-effectiveness of voluntary vaccination strategies [5]. The findings of the model suggested that vaccination of seropositives may be unnecessary if the vaccines provide effective protection to seronegative individuals. However, it is important to note that, as of the current state of affairs, the only authorized dengue vaccine is not recommended for seronegative individuals, rendering the study's conclusions currently inapplicable to actual practice [5].

7.3 Vector-host Transmission Models

Addressing the intricate interactions between vectors, hosts, and pathogens, along with devising effective disease management strategies such as vector control and immunization, necessitates the incorporation of additional variables and assumptions into the modeling framework, grounded in empirical data. Hu et al. [88] conducted a comparative analysis of three models, each varying in their level of complexity: a host-to-host model, which omits vector dynamics; a vector-host model that includes a latency class solely for vectors, introducing a separate compartment for mosquitoes that are infected but not yet infectious, thereby unable to transmit the disease during their latency period; and a more comprehensive vector-host model that incorporates latency compartments for both vectors and hosts. The findings demonstrated that the explicit inclusion of vector dynamics tends to stabilize the complex interactions observed within the host-to-host model [5]. The study concluded that the mere assumption of cross-protection was insufficient to account for the complexities observed in the data. Instead, it was necessary to incorporate the effect of ADE, which intensifies the transmission of secondary infections, to accurately depict the epidemiological data, challenging the conclusions drawn in previous studies [105, 8, 7].

Incorporating vector dynamics into host-to-host models can be efficiently achieved using a basic Susceptible-Infected (SI) type model, as proposed by Rashkov et al. [135]. This approach is exemplified in their work with a minimalistic multi-strain vector-host transmission model that considers two dengue serotypes and accounts for only two possible infections. Rashkov et al. [135] provide a comprehensive description of the model, detailing its structure and the underlying assumptions that facilitate the integration of vector dynamics into traditional host-centric frameworks [5].

In their study on a host-vector model for dengue encompassing two strains, TCI for hosts, and the potential for secondary infections, Rashkov et al. [135] delve into the analysis of endemic equilibria within both single-strain and dual-strain frameworks. They conducted a bifurcation

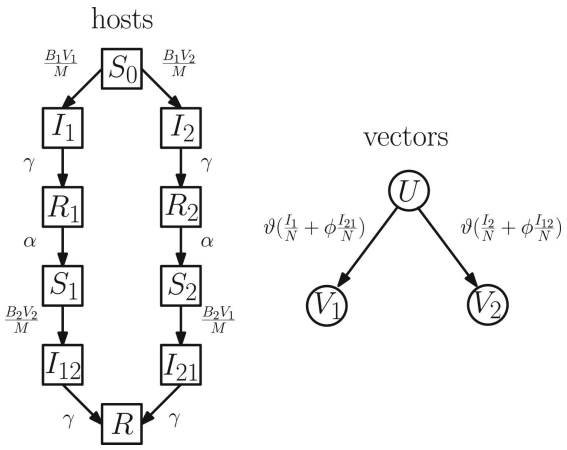


Figure 7.5: Minimalistic multi-strain vector-host transmission model that considers two dengue serotypes and accounts for two possible infections proposed in [135] by Rashkov et al. Figure from [135].

analysis and drew comparisons with the dual-strain host-to-host model outlined in prior research. Echoing findings similar to those reported by Hu et al. [88], Rashkov et al. [135] discovered that incorporating vector dynamics directly into host-centric models markedly simplifies the system's complexity [5]. Nonetheless, the introduction of annual seasonal variations in mosquito biting rates, achieved through sinusoidal adjustments to mosquito population numbers, reintroduces complex dynamics into the model [5]. This complexity arises under the same parameter conditions that lead to complex dynamics in the study referenced earlier, illustrating the nuanced impact of vector behavior on disease transmission models [5].

Murillo et al. [128] introduced a model incorporating two dengue virus genotypes within the vector-host framework. A genotype refers to the genetic makeup of the virus, indicating the specific genetic variations within a serotype. These variations are found within the RNA sequences of the virus and can influence its behavior and characteristics. The research presented by Murillo et al. [128] highlights how competitive interactions between dengue virus genotypes can significantly influence the long-term dynamics of dengue fever [5]. Their findings indicate that even a minimal likelihood of vertical transmission, transmission from an infected female mosquito to her eggs, can critically affect the outcomes of dengue fever outbreaks, influencing whether an outbreak fails, becomes invasive, or leads to endemicity within a population [5]. Subsequently, Anggriani et al. [20] developed a multi-serotype dengue model that considers the possibility of reinfection with the same dengue serotype. Through sensitivity analysis, they demonstrated the significant impact of the reinfection parameter on the dynamics of both primary and secondary infections, although the biological premise of increased infectivity during secondary homologous serotype infections remains subject to debate [5].

Lourenço et al. [116, 115] explored the dynamics of dengue through a multi-strain vector-host model that incorporates an explicit mosquito vector component, TCI following primary infection, and seasonal variations in mosquito biting rates [5]. In one of their studies [115], stochastic simulations were utilized to examine the invasion dynamics of a new dengue genotype within a population already hosting four dengue serotypes [5]. This work assessed the determinants of successful genotype fixation and its epidemiological impacts, underscoring the importance of the epidemiological context over viral fitness for the emergence of new serotypes [115, 5]. In another

study [116], they investigated both spatial and non-spatial aspects of multi-strain dengue models, revealing that spatially explicit models can exhibit a wide range of epidemiological dynamics even in the absence of immunological interactions among different strains [5].

Coudeville and Garnett [45] advanced a sophisticated age-structured model that incorporates both vector-host interactions and serotype-specific compartments, with the inclusion of seasonal variations [5]. Their analysis highlighted the significance of short-term cross-protection, lasting between 6 to 17 months, as essential for accurately mirroring empirical observations [5]. The model's ability to accommodate sequential infections by all four dengue serotypes, as opposed to merely two, was particularly effective in fitting the data. The findings from their study suggest that vaccination could reduce both the frequency and severity of dengue outbreaks and modify the age distribution of disease incidences [5, 45]. Furthermore, by integrating vaccine dynamics into their model, the researchers demonstrated that the net effect of vaccination is contingent upon both its efficacy and the duration of protection it offers [45, 5].

Building upon the framework similar to that of Coudeville and Garnett [45], Knerer et al. [100] formulated a model that also incorporates seasonality, age structure, and the possibility of successive infections by all four dengue serotypes. This model successfully aligned with Thailand's national data by factoring in seasonal patterns and TCI among the serotypes [5]. The study assessed the influence of integrated vector control and vaccination strategies on the transmission of dengue. The findings indicated that while the combination of vaccination and other control strategies yields greater efficacy, even a vaccine with limited effectiveness could beneficially impact dengue transmission dynamics. Utilizing this model, the authors further explored the cost-effectiveness of various well-established dengue prevention strategies and their combinations [101]. An exploratory analysis was conducted to assess the impact and cost-efficiency of employing the Wolbachia bacterium as a biological vector control strategy. This offered insights into the potential benefits of widespread adoption of such measures [5]. Ndii et al. [130, 131] conducted an analysis on the impact of Wolbachia on the dynamics of dengue transmission through a mathematical model that integrates vector-host interactions and accounts for two dengue serotypes [5]. Their research indicates that Wolbachia's introduction markedly diminishes dengue transmission [130] and presents advantages when considering the coexistence of multiple dengue serotypes [131].

Knipl and Moghadas [102] introduced a vector-host model for dengue that encompasses two serotypes, taking into account factors such as Antibody-Dependent Enhancement (ADE), cross-protection, and seasonal variations. Their analysis, informed by vaccine efficacy data from studies [39, 176], concluded that the eradication of the disease is unattainable [102]. Although increased vaccination coverage could reduce infection rates, the model predicts a rise in severe cases, attributed to ADE or the decline in immunity over time [102, 5].

Maier et al. [117] constructed a model that incorporates ADE to determine the most effective vaccination age against dengue in Brazil. By analyzing Brazilian epidemiological data, they calculated the basic reproduction number and identified the optimal vaccination ages for the four dengue serotypes, revealing that these ages vary according to the prevalent serotypes [117, 5]. González Morale et al. [71] proposed a mathematical model featuring two dengue virus strains, a vector mosquito population, and temporary cross-immunity (TCI). This model was used to assess the impact of different vaccination strategies, considering variables such as vaccine efficacy, transmission rates, and the duration of cross-immunity [71]. The findings underscored the significant role of cross-immunity duration in reducing disease incidence and influencing the overall dynamics of disease transmission [5].

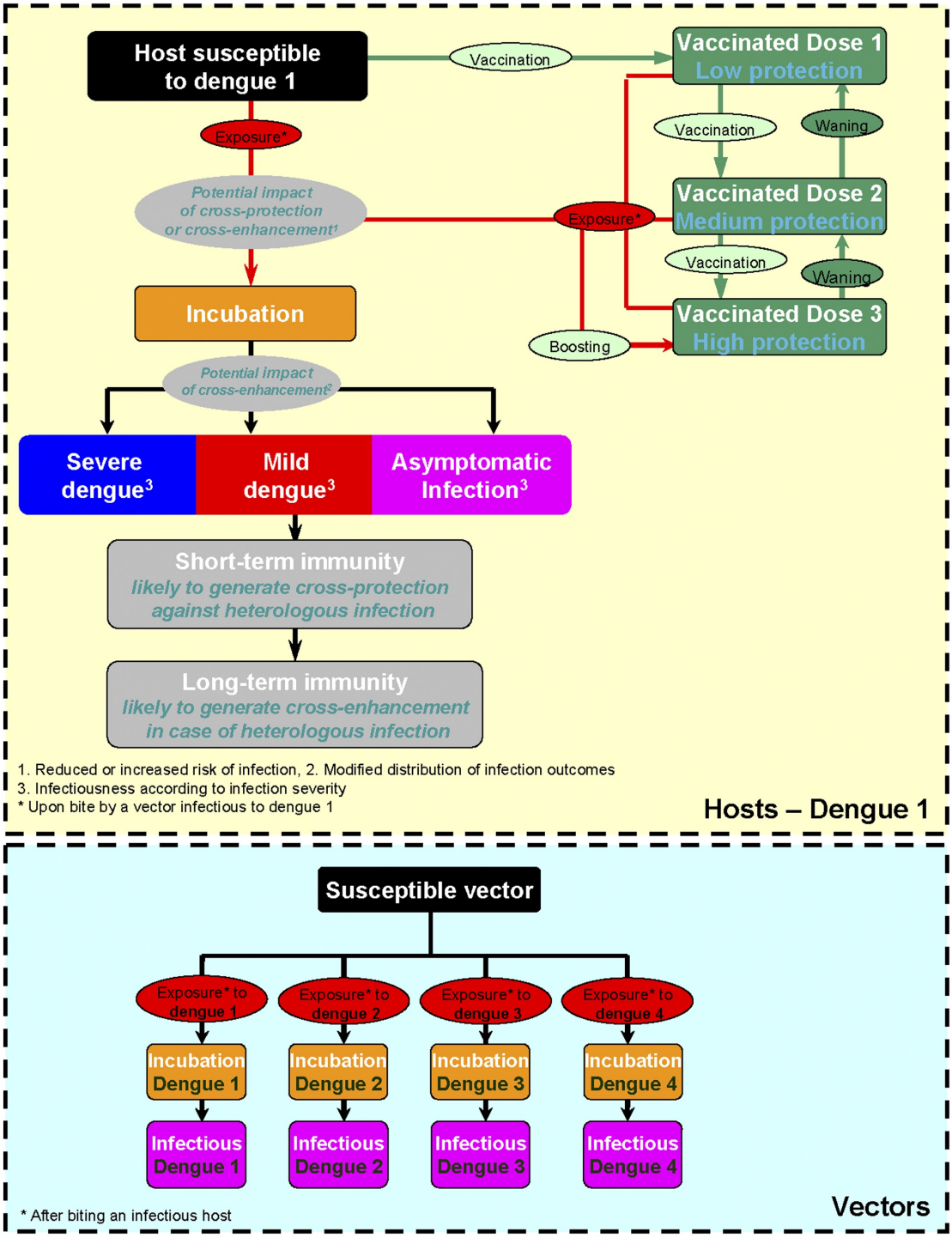


Figure 7.6: Vector-host compartmental model proposed by Coudeville and Garnett in [45]. As per the description from [45], rounded rectangles correspond to compartments and ellipses to factors influencing the transition from one compartment to another. Figure from [45].

Hendron and Bonsall [85] developed an epidemiological model to study the transmission of the dengue virus between vectors and hosts [5]. Their research focused on the synergistic effects of vector control, specifically through genetically modified sterile insect techniques (SIT), combined with vaccination within small networks [5]. They found that while strategies combining multiple interventions are generally more effective, the deployment of imperfect control measures could lead to adverse outcomes. The model also highlighted how host movement within these limited networks could increase the intensity of outbreaks [5, 85].

Falcón-Lezama [57] proposed a model that elucidates the critical roles played by demographic and spatial structures in the initiation, expansion, and control of epidemics. By segmenting the human population into distinct areas based on mobility patterns and incorporating separate submodels for both host and vector populations, the study highlights the influence of highly mobile groups [5, 57]. It identifies local dilution effects and spatial connectivity, represented by the vector-host ratio and the range of regular movement patterns, respectively, as primary drivers of disease spread [5, 57].

Mishra and Gakkhar [123, 124] explored the vector-host dynamics of dengue using two distinct models. In [123], Mishra and Gakkhar developed a model that considers two geographical regions each harboring different dengue serotypes, aiming to assess the impact of human migration on dengue prevalence [5]. The findings suggest that emigration reduces the basic reproduction number in the originating region, whereas immigration exerts an inverse effect. In another study [124], a model incorporating two dengue serotypes and the concept of Antibody-Dependent Enhancement (ADE) was scrutinized. This analysis indicated that an elevated ADE factor could enhance the persistence of the second serotype [5, 124].

Champagne and Cazelles [40] undertook a comparative analysis of deterministic and stochastic models in the context of dengue transmission [5]. Their study examined five compartmental models of increasing complexity, which included features such as vector-borne transmission, explicit representation of asymptomatic infections, and interactions between different virus serotypes [40]. These models were evaluated for their ability to replicate dengue data from rural Cambodia. The results demonstrated that while deterministic models offer a close approximation of average trends at a lower computational cost, stochastic models provide a more nuanced reflection of the uncertainty inherent in parameters and simulations [5, 40].

Bock and Jayathunga [30] formulated a multi-patch vector-host dengue model that integrates the role of mosquitoes carrying the Wolbachia bacterium as a biological control strategy, aimed at either diminishing viral levels within mosquitoes or curtailing their lifespan [5]. The numerical analyses from this model indicate that Wolbachia-infected vectors contribute to a reduction in the mosquito population, thereby curtailing the spread of dengue [30]. In a separate study, Shim [160] devised a mathematical model to explore dengue transmission and vaccination dynamics, accounting for the effects of ADE and varying immunological responses in humans [5]. The model sets forth an optimal control problem to minimize the costs associated with both dengue infections and vaccinations over a specified timeframe [5, 160].

Ghosh et al. [66] presented a qualitative assessment and optimal control strategy for a dengue model encompassing multiple strains and the possibility of co-infections. The model proposes three control measures aimed at diminishing infection rates in both humans and mosquitoes: public awareness campaigns to prevent mosquito bites, medical treatment for infected individuals, and efforts to reduce mosquito populations [5, 66]. The effectiveness of these strategies was evaluated under constant and time-varying conditions using Pontryagin's Maximum Principle, revealing that a combination of human awareness and treatment is more efficacious than pairing

mosquito eradication efforts with treatment [66]. Xue and colleagues [194] developed a comprehensive two-infection multi-strain vector-host dengue model that includes latent classes for both hosts and vectors. Through sensitivity analysis, the study pinpointed key factors influencing dengue's transmission dynamics [5]. The model, which accounts for temporary cross-immunity (TCI), did not delve into the specifics of this parameter. It further assessed two control measures—enhancing public awareness and improving mosquito control strategies—under the lens of optimal control, applying Pontryagin's Maximum Principle and considering the associated economic implications [194, 5].

Chapter 8

Multi-objective multi-armed bandits

In the current literature on multi-objective multi-armed bandit (MOMAB) algorithms, two primary categories exist. The first category comprises algorithms that strive to identify the entire set of Pareto optimal arms without considering the utility function or preferences of the decision maker. The second category includes those that integrate the utility function or preferences. In this study, these categories are referred to as *preference-unaware MOMAB algorithms* and *preference-incorporating MOMAB algorithms*, respectively. The preference-incorporating MOMAB algorithms either require some prior definition of the preferences of the user to establish an order between the arms within the Pareto front, or attempt to learn the preferences of the decision maker through query based interactions with the user at runtime.

An interesting observation that stems from this literature review, and is confirmed by [140], is the lack of literature on pure MOMAB algorithms for Pareto front identification (PFI) where the preference of the user is not taken into account. It is, however, important to note that within the space of multi-objective optimization problems, Bayesian Optimization algorithms exist that incorporate MABs for their acquisition function [140]. These Bayesian Optimization algorithms can be used to solve MOMAB settings like the one we propose but are beyond the scope of this dissertation. This PFI setting is essentially the multi-objective extension of the best-arm identification setting for single-objective MABs discussed in Section 5.2. This setting is of particular interest since it corresponds exactly to the objective of this dissertation: to provide the decision maker with the complete set of possible trade-offs between different optimal vaccination strategies, without making any assumptions about their preferences. The contributions in this area that were made within this study can be found in Chapter 10.

8.1 Preference-incorporating MOMAB algorithms

8.1.1 Constrained lower confidence bound

In their work, Kagrecha et al. [93] introduce a novel algorithm termed *Constrained Lower Confidence Bound* (Con-LCB) designed to address the complexities inherent in multi-objective multi-armed bandit (MOMAB) problems, particularly those that entail constraints. Central

to their contribution is the provision of a logarithmic regret guarantee, signifying that the algorithm's frequency of selecting non-optimal arms remains logarithmically proportional to the decision-making horizon [93]. This characteristic ensures an efficient convergence towards optimal decisions over time. Furthermore, Con-LCB is distinguished by its capability to ascertain, with a high degree of certainty, the feasibility of a given problem instance in relation to predefined constraints, thereby enhancing decision-making accuracy.

The algorithm's optimality is asserted to be within a universal constant, suggesting that advancements in algorithmic complexity are unlikely to yield significantly superior results [93]. This assertion underlines the robustness of Con-LCB in navigating the decision space of MOMAB problems. The authors advocate for a constrained optimization approach to multi-criterion decision-making, where one attribute is optimized while adhering to constraints on others, a perspective that is relatively underexplored in the domain of MOMABs.

Kagrecha et al. elaborate on the operational framework of Con-LCB, where the attributes of objectives and constraints are encapsulated by functions g_0 and g_1, respectively. These functions map a set of choices C to real values R with the constraint threshold denoted by $\tau \in \mathbb{R}$. The algorithm classifies arms into feasible and infeasible categories based on their compliance with the constraint $g_1(v(\cdot)) \leq \tau$, thereby guiding the selection process towards viable solutions.

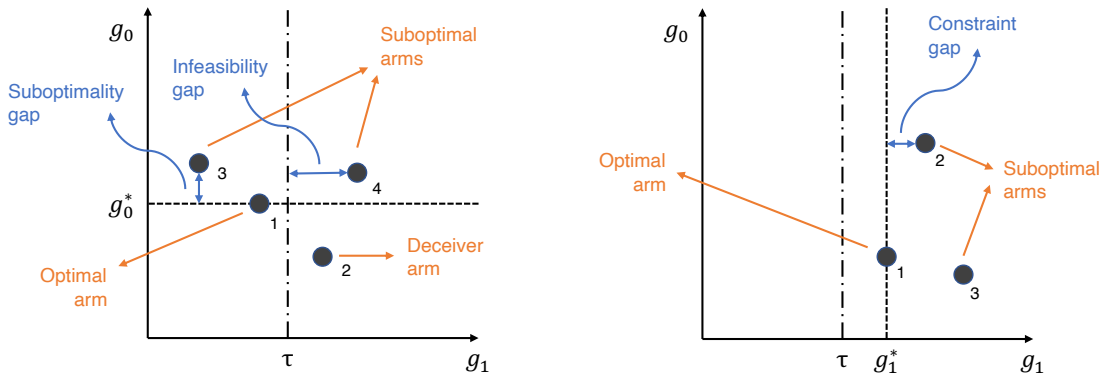


Figure 8.1: Representation of both the feasible and infeasible instance for the Con-LCB algorithm. The left panel shows four distinct arms where the optimal arm respects the constraint $g_1(v(\cdot)) \leq \tau$ and thus is considered feasible. Arm 2 has a lower value for g_0, so could be a better arm than arm 1. However, it does not respect the constraint set by the user with respect to τ. The right panel shows a case where none of the arms respect the constraint, so the one with the smallest constraint gap is considered the optimal arm.

In the context of a feasible instance, an optimal arm is characterized by its minimization of $g_0(v(\cdot))$ subject to the constraint $g_1(v(\cdot)) \leq \tau$. Conversely, for infeasible instances, the optimal arm is defined by the minimal value of $g_1(v(\cdot)$ that necessitates a relaxation of the constraint until at least one arm meets the criteria. This nuanced approach to defining optimality underscores the algorithm's adaptability to varying problem instances.

The Con-LCB algorithm operates on the principle of optimism under uncertainty, employing lower confidence bounds (LCBs) to maintain a set of potentially feasible arms and to facilitate the arm selection process [93]. This methodology underscores the algorithm's reliance on probabilistic bounds to navigate the uncertainty inherent in MOMAB problems. The feasibility flag, set at

the conclusion of the decision-making horizon, serves as a binary indicator of the presence of viable solutions, further exemplifying the algorithm's utility in practical scenarios.

Algorithm 6: Constrained Lower Confidence Bound, from [93]

Input: A MOMAB with K arms, a constraint τ, arm pull budget T
Play each arm once
for $t = K + 1$ **to** T **do**
  Set $\hat{\mathcal{K}}_t = \left\{ k : \hat{g}_{1,N_k(t-1)}(k) - \sqrt{\frac{\log(2T^2)}{a_1 N_k(t-1)}} \leq \tau \right\}$
  if $\hat{\mathcal{K}}_t \neq \emptyset$ **then**
    Set $L_0(k) = \hat{g}_{0,N_k(t-1)}(k) - \sqrt{\frac{\log(2T^2)}{a_0 N_k(t-1)}}$
    Play arm $k_t^\dagger \in \arg\min_{k \in \hat{\mathcal{K}}_t} L_0(k)$
  else
    Set $L_1(k) = \hat{g}_{1,N_k(t-1)}(k) - \sqrt{\frac{\log(2T^2)}{a_1 N_k(t-1)}}$
    Play arm $k_t^\dagger \in \arg\min_{k \in [K]} L_1(k)$
if $\hat{\mathcal{K}}_T \neq \emptyset$ **then**
  Set `feasibility_flag = true`
else
  Set `feasibility_flag = false`

In summary, Kagrecha et al.'s 2023 work [93] on the Con-LCB algorithm marks a significant advancement in the field of constrained multi-objective multi-armed bandits. By introducing a robust, optimally bounded solution that adeptly navigates the complexities of constrained decision environments, this research contributes valuable insights and methodologies to the domain, promising enhanced decision-making efficiency in multi-criterion contexts.

8.1.2 uMAP-UCB and interactive Thompson sampling

In the exploration of multi-objective multi-armed bandit (MOMAB) algorithms, the 2017 study by Roijers et al. [147] provides an insightful examination of the interactive Thompson Sampling (ITS) and its comparison with the uMAP-UCB algorithm. The principal aim of these algorithms is to optimize user interaction within an online interactive multi-objective reinforcement learning (MORL) setting, thereby minimizing user regret. This involves a dual interaction with both the environment and the user, necessitating a learning process about the reward at time t, denoted as $\mathbf{r}^{(t)}$, as well as the user's preferences and weightings, u and ω, to ultimately optimize $u(\mathbb{E}[\mathbf{r}^{(t)}], \omega)$.

Drawing upon the work of Zoghi et al. [207], who explored relative bandits in a model where reward vectors are not directly observable, the approach adopted by Roijers et al. [147] involves interacting with users through pairwise comparison queries either before or after an arm pull. This interaction diverges from Zoghi et al. [207] by presenting users with estimations of expected reward vectors for comparison, rather than outcomes from single arm pulls. Users are then asked to express a preference between two vectors, $\mathbf{r}_1$ and $\mathbf{r}_2$, with the preference denoted as $\mathbf{r}_1 \succ \mathbf{r}_2$. At any given timestep t, the algorithm has access to a dataset C of j such preference pairs, with $j \leq t$, indicating the cumulative number of comparisons up to that point.

The framework does not impose a fixed limit on the number of user comparisons, constrained only by the finite time horizon T, a constraint that equally applies to the number of arm pulls. The

assumption that $u(\boldsymbol{\mu}, \omega)$ is a linear utility function allows for the estimation of the true weights ω^* using Bayesian logistic regression. This method not only provides a maximum a posteriori estimate of ω^* but also a posterior distribution over these weights, enhancing the precision of the utility function estimation.

In aiming to minimize user regret, the study emphasizes the importance of not overburdening the user with excessive queries, recognizing the potential for such interactions to be time-consuming and potentially intrusive. The proposed algorithms, therefore, strive to reduce both the expected user regret and the number of queries per timestep as the interaction progresses. This objective is pursued through the adoption of two advanced classes of algorithms: the Upper Confidence Bound (UCB) and Thompson Sampling, both of which are traditionally employed in single-objective bandit scenarios but are here adapted to the multi-objective context to enhance user experience and algorithm efficiency.

Utility-MAP Upper Confidence Bound

In the advancement of multi-objective multi-armed bandits, the utility-MAP Upper Confidence Bound (uMAP-UCB) algorithm, as introduced by Roijers et al. [147], represents a significant adaptation of the traditional UCB framework to accommodate multi-objective decision-making contexts. This algorithm is distinctive for its integration of explicit exploration bonuses and its method for deciding when to engage the user through queries. The decision-making process involves computing the maximum a posteriori (MAP) of the utility function to ascertain the optimal arm based on the best mean estimate and on the best mean estimate augmented by an exploration bonus [147]. When these two optimal arms diverge, the uMAP-UCB algorithm solicits user input through pairwise comparisons between the estimated mean reward vectors of the respective arms [147].

The uMAP-UCB algorithm is grounded in the UCB methodology, which is well-established in single-objective multi-armed bandits, where actions are selected based on the arm's estimated mean rewards plus an exploration bonus. This exploration bonus serves to construct an upper confidence bound for the true mean rewards of the arms. Applying this approach to MOMABs introduces several challenges: the necessity to account for a user's linear utility function with an unknown weight parameter ω^*, the selection of actions based on the current MAP estimate of the weight vector $\overline{\omega}$, and the objective to estimate ω^* while concurrently reducing the frequency of user queries over time without significantly impacting user regret.

To estimate w^*, the algorithm employs Bayesian logistic regression, assuming a linear utility function and utilizing pairwise comparisons as the data source. A multi-variate Gaussian prior is set on the weights, leading to the MAP estimate of the weights at each iteration's outset [147]. This estimate is then projected onto the simplex for d objectives, ensuring adherence to simplex constraints, which is crucial for the algorithm's exploration bonuses and regret bounds.

In selecting which arm to play, uMAP-UCB follows the standard UCB schema, incorporating the expected scalarised reward $(\overline{\omega} \cdot \boldsymbol{\mu}_a) \forall a \in \mathcal{A}$ and an exploration bonus. This method presupposes that the weight estimates are independent of the mean reward estimates, allowing for objective pairwise comparisons by the user unaffected by previous actions and comparisons. The exploration bonus is adaptable, reducing to single-objective MAB exploration bonuses when the weight is focused on a single objective, and considering tighter bounds for more evenly distributed weights, although this aspect requires further exploration [147].

uMAP-UCB's querying mechanism is closely tied to its exploration strategy, ensuring continuous user engagement to refine the weight estimates over time. This approach guarantees that the

Algorithm 7: Utility-MAP UCB, from [147]

Input: A MOMAB with K arms, a prior $\pi(\cdot)$ on the distribution of ω, a comparison history $C = \emptyset$, arm pull budget T

$\boldsymbol{\mu}_a \leftarrow$ initialize with single pull $\forall a \in \mathcal{A}$

$n_a^{(t)} \leftarrow 1, \forall a \in \mathcal{A}$

for $t = K + 1$ **to** T **do**

- $\overline{\omega} \leftarrow \text{MAP}(\omega|C)$ `// Update estimate of` ω
- $\bar{a}^* \leftarrow \arg\max_a \overline{\omega} \cdot \boldsymbol{\mu}_a$
- $a^{(t)} \leftarrow \arg\max_a(\bar{\omega} \cdot \boldsymbol{\mu}_a + c(\bar{\omega}, \boldsymbol{\mu}_a, n_a^{(t)}, t))$
- $\mathbf{r}^{(t)} \leftarrow$ play $a^{(t)}$ and observe reward
- Update $\boldsymbol{\mu}_{a^{(t)}}$ with $\mathbf{r}^{(t)}$
- $n_{a^{(t)}}^{(t)} \leftarrow n_{a^{(t)}}^{(t)} + 1$
- **if** $\bar{a}^* \neq a^{(t)}$ **then**
  - $c^{(t)} \leftarrow$ user comparison for $\boldsymbol{\mu}_{\bar{a}^*}$and $\boldsymbol{\mu}_{a^{(t)}}$
  - $C \leftarrow C \cup c^{(t)}$

algorithm does not cease querying, thus avoiding the risk of persistently favoring a suboptimal arm due to inaccurate weight estimates [147]. The algorithm's design aims to diminish the number of user queries over time while ensuring that the exploration mechanism remains active, thereby balancing exploration with the minimization of user burden and regret [147].

Interactive Thompson sampling

In the realm of MOMABs, the work of Roijers et al. [147] introduces a novel approach through their Interactive Thompson Sampling (ITS) algorithm. This algorithm extends the principles of Thompson Sampling, a method known for its efficacy in single-objective multi-armed bandits, to accommodate the complexities inherent in multi-objective optimization and interactive user preference elicitation.

Thompson Sampling, traditionally employed in single-objective MABs, is recognized for its ability to outperform UCB algorithms by starting with a prior distribution on the parameters of each arm's reward distribution, subsequently updating these priors with empirical data to form posterior distributions. The arm with the highest expected reward, based on sampled parameters from these posterior distributions, is then selected for pulling [147].

Building upon this foundational strategy, ITS adapts Thompson Sampling for MOMABs by incorporating an additional layer of complexity: sampling not only from the posteriors of the reward distribution parameters of each arm but also from the posteriors of the user's utility function parameters. This dual-sampling mechanism enables ITS to integrate user preferences into the decision-making process, thereby aligning arm selection with both the anticipated rewards and the user's objective preferences.

At each iteration t, ITS independently draws two sets of samples: one for the reward distribution parameters of each arm, $\tilde{\boldsymbol{\mu}}_1^{(t)}$ and $\tilde{\boldsymbol{\mu}}_2^{(t)}$, and another for the utility function parameters: $\tilde{\boldsymbol{\eta}}_1^{(t)}$ and $\tilde{\boldsymbol{\eta}}_2^{(t)}$ [147]. The first set of samples is utilized to determine the arm to be played by maximizing the expected product of the sampled utility weights and the sampled mean rewards for each arm. This process assumes that the weights and rewards can be independently sampled from their

respective posterior distributions, a crucial assumption that underpins the ITS methodology.

The second set of samples serves a distinct purpose: to assess the necessity and mode of user interaction. By comparing the actions selected by both sets of samples, ITS determines whether a discrepancy exists. If such a discrepancy is present, indicating divergent best-arm selections between the two sample sets, ITS engages the user by requesting a pairwise comparison between the expected rewards of the arms selected by each sample set [147].

Algorithm 8: Interactive Thompson Sampling, from [147]

Input: A MOMAB with K arms, priors $\pi_\omega(\cdot)$ and $\pi_r(\cdot)$ on the distributions of ω and the arms' reward distributions, a comparison history $C = \emptyset$ and reward history $\mathcal{H}^{(0)} = \emptyset$, arm pull budget T

for $t = 1$ **to** T **do**

- $\tilde{\boldsymbol{\mu}}_1^{(t)}, \tilde{\boldsymbol{\mu}}_2^{(t)} \leftarrow$ draw 2 samples from $\pi_u(\cdot|\mathcal{H}^{(t-1)})$
- $\tilde{\boldsymbol{\eta}}_1^{(t)}, \tilde{\boldsymbol{\eta}}_2^{(t)} \leftarrow$ draw 2 samples from $\pi_\omega(\cdot|C)$
- $a_1^{(t)} \leftarrow \arg\max_a \mathbb{E}_{P(\mathbf{r},\omega|a,\tilde{\boldsymbol{\mu}}_1^{(t)},\tilde{\boldsymbol{\eta}}_1^{(t)})}[\omega \cdot \mathbf{r}]$
- $a_2^{(t)} \leftarrow \arg\max_a \mathbb{E}_{P(\mathbf{r},\omega|a,\tilde{\boldsymbol{\mu}}_2^{(t)},\tilde{\boldsymbol{\eta}}_2^{(t)})}[\omega \cdot \mathbf{r}]$
- $\mathbf{r}^{(t)} \leftarrow$ play $a_1^{(t)}$ and observe reward
- $\mathcal{H}^{(t)} \leftarrow \mathcal{H}^{(t-1)} \cup \{a_1^{(t)}, \mathbf{r}^{(t)}\}$
- **if** $a_1^{(t)} \neq a_2^{(t)}$ **then**
  - $\boldsymbol{\mu}_{1,a_1^{(t)}} \leftarrow \mathbb{E}_{P(\mathbf{r}|a_1^{(t)},\tilde{\boldsymbol{\mu}}_1^{(t)})}[\mathbf{r}]$
  - $\boldsymbol{\mu}_{2,a_2^{(t)}} \leftarrow \mathbb{E}_{P(\mathbf{r}|a_2^{(t)},\tilde{\boldsymbol{\mu}}_2^{(t)})}[\mathbf{r}]$
  - $c^{(t)} \leftarrow$ user comparison for $\boldsymbol{\mu}_{1,a_1^{(t)}}$ and $\boldsymbol{\mu}_{2,a_2^{(t)}}$
  - $C \leftarrow C \cup c^{(t)}$

As ITS progresses and the posterior distributions for both rewards and user preferences become more precise, the algorithm experiences a reduction in both suboptimal arm pulls and the necessity for user queries. This reduction is attributed to the increasing alignment between the sampled actions and the true optimal actions, reflective of both the anticipated rewards and the user's utility preferences [147]. Thus, ITS not only offers a sophisticated approach to navigating the multi-objective decision space but also efficiently incorporates user feedback to refine its understanding of the utility landscape, thereby enhancing the decision-making process in MOMAB environments.

8.1.3 Multi-objective Top-Two Thompson Sampling (MOTTTS)

Multi-objective optimization in the context of multi-armed bandit (MAB) problems is a challenging task that necessitates balancing multiple objectives simultaneously. The Multi-objective Top-Two Thompson Sampling (MOTTTS) algorithm extends the single-objective Top-Two Thompson Sampling (TTTS) to the multi-objective setting by learning multivariate belief distributions over the arms.

The MOTTTS algorithm incorporates a utility function, u, which is initially unknown and must be inferred over time through interaction with a decision maker. This utility function helps in determining the best arm to pull by providing a framework to understand and rank the multivariate samples from the arm belief distributions. The interaction with the decision maker

allows for the refinement of the belief distribution over u, which in turn improves the accuracy of the arm selection process [140].

In practice, the utility function is often a human decision maker who finds it challenging to express preferences in absolute numerical terms due to the unnatural and error-prone nature of such expressions. Instead, humans are more consistent in expressing preferences relatively, such as preferring one option over another. This is supported by studies (Tesauro, 1988; Zoghi et al., 2014) that indicate humans find relative feedback easier and more consistent over time [140].

To leverage this relative feedback, MOTTTS translates the process into a binary classification task. Given two propositions, $\mathbf{r}_0$ and $\mathbf{r}_1$, the goal is to predict whether the decision maker prefers $\mathbf{r}_0$ over $\mathbf{r}_1$. Formally, this preference is denoted as $\mathbf{r}_0 \succ \mathbf{r}_1$. Each interaction with the decision maker is thus recorded as a pair $\langle\langle \mathbf{r}_0, \mathbf{r}_1\rangle, \mathbf{r}_0 \succ \mathbf{r}_1\rangle$ and stored in an interaction history $\mathcal{H}^{(t)}$ [140]. This history is crucial for updating the belief distribution over the utility function.

At the start, the algorithm assumes no prior knowledge of the utility function, treating all possible utility functions with equal weight. With each new interaction, the likelihood of each utility function is updated to reflect the decision maker's preferences [140]. Specifically, utility functions that match the decision maker's past answers are assigned higher likelihoods, while those that do not are assigned lower likelihoods. The weight ω of a utility function f given the history $\mathcal{H}^{(t)}$ of relative queries is defined as:

$$\omega = \prod_{h \in \mathcal{H}^{(t)}} \left|(\mathbf{r}_0^h \succ \mathbf{r}_1^h) - \eta\right| (\omega^\top \mathbf{r}_0^h \geq \omega^\top \mathbf{r}_1^h),$$

where η accounts for potential mistakes or changes in preference from the decision maker. When $\eta = 0$, only the utility functions for which all answers match the decision maker's preferences have a non-zero probability of being sampled, ensuring that these utility functions are equally likely to be chosen [140].

In summary, MOTTTS provides an effective framework for multi-objective optimization in MAB problems by incorporating human feedback to update the belief distributions over both arms and the utility function. This approach leverages human consistency in relative preference expression to enhance the accuracy and reliability of the arm selection process.

8.2 Preference-unaware MOMAB algorithms

In this section, four different variants of preference-unaware multi-objective multi-armed bandit (MOMAB) algorithms are presented. All algorithms presented in this section are algorithms for regret minimization, all aiming to minimize their cumulative and unfairness regrets while navigating the exploration-exploitation trade-off in their own unique way: UCB, Thompson sampling, Knowledge gradient, and an Annealing approach.

The DENV MOMAB setting this dissertation aims to propose as one of its main contributions does not align completely with the regret minimization setting that these algorithms were designed for. Instead, it is an instance of the Pareto front identification (PFI) problem for MOMABs. However, their preference-unaware nature does align with the needs of this study. Furthermore, when considering single-objective MAB algorithms, it is often the case that best-arm identification algorithms and algorithms for regret minimization are closely related to one another, where the former has an additional emphasis on exploration. Some examples are the UCB algorithm with a larger value for the exploration-regulating hyperparameter κ, and the

close relation between Thompson sampling and Top-two Thompson Sampling. These facts, combined with the lack of pure MOMAB algorithms for Pareto front identification [140], makes these algorithms very valuable to examine within the context of this dissertation[1]. They might be suitable for adaptation to the MOMAB PFI setting.

Within this context, each of the algorithms presented in this part of this literature review has been reproduced based on the original publications in which they were proposed. Their implementations were verified by examining their cumulative Pareto regrets and cumulative unfairness regrets. This was achieved on a test-bed consisting of a bi-objective Bernoulli bandit with 4 Pareto optimal arms and 16 suboptimal arms. For these experiments, the bandit algorithms presented in this part of the literature review were evaluated over the course of 250,000 arm pulls. This experiment was repeated 100 times to obtain an average regret minimization performance for each of the algorithms. This running experiment with the same test bed was also used to compare the performance of the different algorithms discussed in this section.

8.2.1 Linear scalarized and Pareto Upper Confidence Bound

In this section, the multi-objective multi-armed bandit (MOMAB) algorithms introduced by Drugan and Nowe in 2013 [51] will be delved into, specifically focusing on the scalarized and Pareto upper confidence bound approaches. Their groundbreaking work extends the traditional Upper Confidence Bound (UCB) algorithm, originally designed for single-objective scenarios, to the multi-objective domain. They introduce three innovative adaptations: *Scalarized UCB1*, *Pareto UCB1*, and *Improved scalarized UCB1* [51].

Similar to their single-objective counterparts, these algorithms estimate the mean reward $\boldsymbol{\mu}_a$ for each objective D across all arms $a \in \mathcal{A}$. Additionally, they incorporate a mechanism to account for the uncertainty of these rewards, employing a one-sided confidence interval based on the Chernoff-Hoeffding bounds [51]. This interval is crucial as it represents the confidence in the estimated rewards, thereby guiding the algorithm to explore arms that are potentially more rewarding but have been less frequently selected.

Firstly, the Scalarized UCB1 algorithm, detailed in Algorithm 8 will be looked at. This algorithm operates with two primary inputs: a MOMAB with K arms, and a set of scalarization functions S. To kick-start Scalarized UCB1, it initially considers each scalarization function from S for every arm $a \in \mathcal{A}$ [51]. As a result, at the onset, the total amount of arm pulls $n^{(t)}$ up to time t, is initialized to K, and the count of selections for each arm under every scalarization, $n_a^{(t)}$, starts at 1 [51].

Following this setup phase, the algorithm enters its main loop. At every step, it picks a scalarization function, f^s, from S at random. With f^s selected, the algorithm then computes a scalarized estimate of the mean reward for each arm, incorporating a confidence bound that reflects the number of times both the scalarization function and the arm have been chosen [51]. The arm with the highest aggregate of these computed terms is then selected for the next action, and its resulting reward is recorded [51].

Upon observing the reward, the algorithm updates the mean reward estimate for the selected arm based on this new data. Concurrently, it adjusts the counts to reflect the newly made arm selection [51]. This process dynamically balances exploration and exploitation by updating

[1]It is, however, important to note that within the space of multi-objective optimization problems, Bayesian Optimization algorithms exist that incorporate MABs for their acquisition function [140]. These Bayesian Optimization algorithms can be used to solve MOMAB settings like the one we propose but are beyond the scope of this dissertation.

Algorithm 9: Scalarized UCB1, from [51]

Input: A MOMAB with K arms, set of scalarization functions $S = (f^1, \cdots, f^s, \cdots, f^S)$
$\boldsymbol{\mu}_a \leftarrow$ initialize with single pull $\forall a \in \mathcal{A}$
$n_a^{(t)} \leftarrow 1, \forall a \in \mathcal{A}$
$n^{(t)} \leftarrow K$
for $t \leftarrow K$ **to** $+\infty$ **do**
  Select a function $f^s \in S$ uniformly at random
  for $a \leftarrow 1$ **to** K **do**
    Calculate $f^s(\boldsymbol{\mu}_a) + \sqrt{\frac{2\ln(n^{(t)})}{n_a^{(t)}}}$
  Select the optimal arm $a_s^{(t)}$ that maximizes $f^s(\boldsymbol{\mu}_{a_s^{(t)}}) + \sqrt{\frac{2\ln(n^{(t)})}{n_{a_s^{(t)}}^{(t)}}}$
  Play $a_s^{(t)}$ and observe $\mathbf{r}^{(t)}$
  Update $n^{(t)}$, $n_{a_s^{(t)}}^{(t)}$, and $\boldsymbol{\mu}_{a_s^{(t)}}$

its understanding of each arm's potential based on the latest outcomes and the diversity of scalarization functions considered.

Algorithm 10: Pareto UCB1, from [51]

Input: A MOMAB with K arms
$\boldsymbol{\mu}_a \leftarrow$ initialize with single pull $\forall a \in \mathcal{A}$
$n_a^{(t)} \leftarrow 1, \forall a \in \mathcal{A}$
$n^{(t)} \leftarrow K$
for $t \leftarrow K$ **to** $+\infty$ **do**
  for $a \leftarrow 1$ **to** K **do**
    Calculate $\boldsymbol{\mu}_a + \sqrt{\frac{2\ln(n^{(t)}\sqrt[4]{DK})}{n_a^{(t)}}}$
  Select the Pareto optimal arms $\mathcal{A}^*$ such that $\forall i \in \mathcal{A}^*$ and $\forall j \notin \mathcal{A}^*$,
  $\boldsymbol{\mu}_j + \sqrt{\frac{2\ln(n^{(t)}\sqrt[4]{DK})}{n_j^{(t)}}} \not\succ \boldsymbol{\mu}_i + \sqrt{\frac{2\ln(n^{(t)}\sqrt[4]{DK})}{n_i^{(t)}}}$
  Select $a^{(t)}$ uniformly at random from $\mathcal{A}^*$
  Play $a^{(t)}$ and observe $\mathbf{r}^{(t)}$
  Update $n^{(t)}$, $n_{a^{(t)}}^{(t)}$, and $\boldsymbol{\mu}_{a^{(t)}}$

Next, the Pareto UCB1 algorithm is delved into, another pivotal approach similar to the Scalarized UCB1 but with its unique methodology. Pseudo code for this algorithm can be found in Algorithm 9. This algorithm also only takes a MOMAB with K arms as its input. The initialization process is straightforward: each arm is sampled once [51]. During each cycle of the algorithm, it calculates for every arm the aggregate of its average reward vector and the corresponding confidence interval. From these computations, a set of Pareto optimal arms, denoted as $\mathcal{A}^*$, is derived. This set distinguishes the Pareto optimal arms, meaning that for any arm j not in $\mathcal{A}^*$, there exists at least one arm i within $\mathcal{A}^*$ that outperforms j [51] in terms of both

reward and confidence:

$$\boldsymbol{\mu}_j + \sqrt{\frac{2\ln(n^{(t)}\sqrt[4]{DK})}{n_j^{(t)}}} \not\prec \boldsymbol{\mu}_i + \sqrt{\frac{2\ln(n^{(t)}\sqrt[4]{DK})}{n_i^{(t)}}}$$

From the set of Pareto optimal arms $\mathcal{A}^*$, an arm is chosen at random and executed [51]. This selection mechanism ensures equitable treatment of all Pareto optimal choices. Following this, updates are made to the mean value $\boldsymbol{\mu}_{a^{(t)}}$ and the general counters to reflect this action [51].

As a third algorithm, Drugan and Nowe also propose an improved version of the scalarized UCB1 algorithm. This improved algorithm is designed to mitigate the problem of unfairness between arm pulls in scalarized algorithms, when the shape of the Pareto front is non-convex [51]. A solution to the scalarized multi-objective UCB1 problem is to create an algorithm that retains only the most effective scalarized UCB1 instances. This approach focuses on maintaining a minimal set of the highest-performing scalarized UCB1s, while eliminating any superfluous instances [51]. The criterion for a scalarized UCB1 to be considered essential is its frequency and uniformity in selecting Pareto optimal arms – such a scalarized UCB1 is regarded as both effective and fair. Conversely, a scalarized UCB1 that infrequently or unevenly selects Pareto optimal arms is deemed redundant and is subsequently removed upon reaching a predetermined level of confidence.

Algorithm 11: Improved scalarized UCB1, from [51, 53]

Input: A MOMAB with K arms, set of scalarization functions $S = (f^1, \cdots, f^s, \cdots, f^S)$, arm pull budget T

Set $\widetilde{\Delta}_0 \leftarrow 1$, and $B_0 \leftarrow S$

for *all rounds* $m = 0, 1, \ldots, \lfloor \frac{1}{2}\log_2 \frac{T}{e} \rfloor$ **do**

  if $|B_m| > 1$ **then**

    for *all* $f^j \in B_m$ **do**

      Play the scalarized UCB1 for $n_m \leftarrow \left\lceil \frac{2\cdot\log(T\widetilde{\Delta}_m^2)}{\widetilde{\Delta}_m^2} \right\rceil$ times

      For each arm i, update $\boldsymbol{\mu}_i$

  else

    Choose the only function in B_m until T is reached

  Find the Pareto optimal reward set of round m, $\mathcal{A}_m^*$, using the mean reward vectors $\boldsymbol{\mu}_i$

  for *all* $f^j \in B_m$ **do**

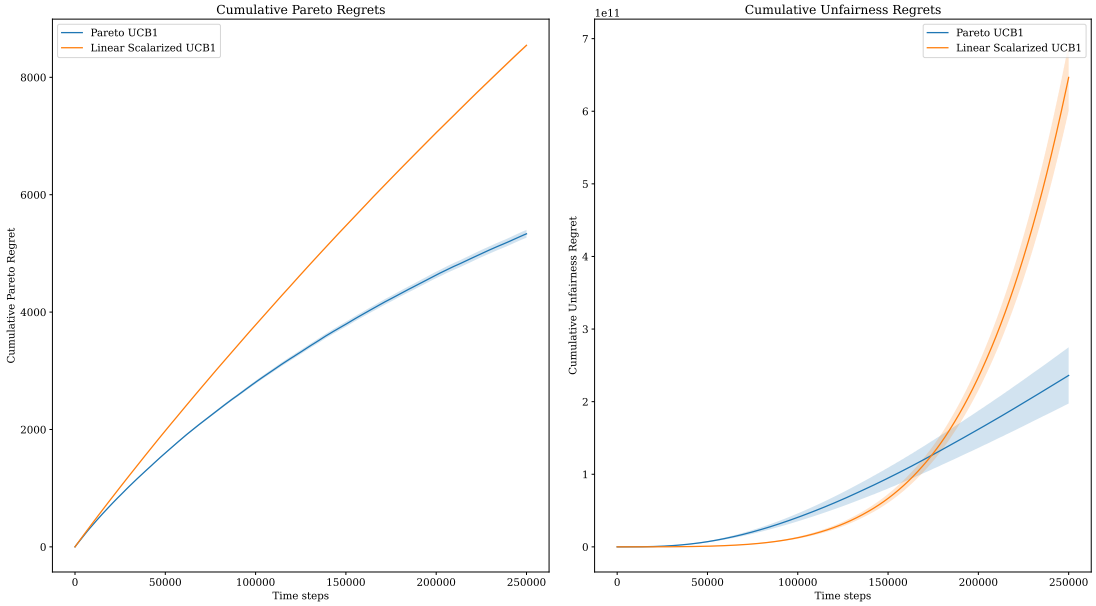
    if $\min_{\ell \in B_m} \phi_m^\ell + \sqrt{\frac{\log(T\widetilde{\Delta}_m^2)}{2n_m}} < \phi_m^j - \sqrt{\frac{\log(T\widetilde{\Delta}_m^2)}{2n_m}}$ **then**

      Delete f^j

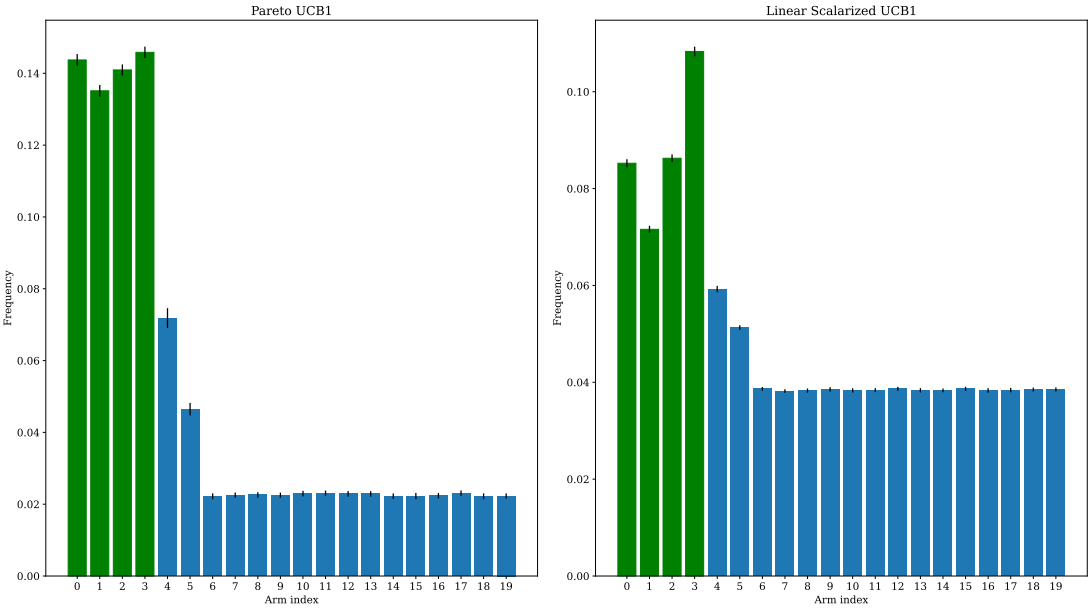
  Update B_{m+1} to the remaining scalarizations

  $\widetilde{\Delta}_{m+1} \leftarrow \frac{\widetilde{\Delta}_m}{2}$

The refined algorithm for the improved scalarized multi-objective UCB1 is outlined in Algorithm 10. The horizon T is presumed to be known beforehand, and the initial set of scalarizations is denoted by $B_0 \leftarrow S$ [51]. A unique scalarization function f^j is linked to each instance of scalarized UCB1 from Algorithm 8. These instances are executed for a fixed number of times n_m. The enhanced algorithm operates over m rounds, with the frequency n_m of each scalarized UCB1 instance's execution incrementally increasing in each round [51]. Upon completion of all



(a) Cumulative Pareto regrets and cumulative unfairness regret for the Pareto UCB1 and Linear scalarized UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean.



(b) Arm pull frequencies for the Pareto UCB1 and Linear scalarized UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean frequency. Pareto optimal arms are shown in green.

Figure 8.2: Comparison of the average Pareto and unfairness regrets, as well as the respective average arm pull frequencies between Pareto UCB1 and Linear scalarized UCB1.

scalarized UCB1 runs, the Pareto optimal set of arms for that round, denoted $\mathcal{A}_m^*$, is determined from the aggregate mean reward vectors $\boldsymbol{\mu}_i$ across all instances. For each scalarized UCB1, the unfairness ϕ_m^j in the current round m is computed [51]. A scalarization function is removed if the difference between its unfairness and the smallest unfairness in the set, minus the confidence interval, is greater than this smallest unfairness plus the confidence interval [51]:

$$\min_{\ell \in B_m} \phi_m^\ell + \sqrt{\frac{\log(T\widetilde{\Delta}_m^2)}{2n_m}} < \phi_m^j - \sqrt{\frac{\log(T\widetilde{\Delta}_m^2)}{2n_m}}$$

This procedure is then iteratively applied after updating the set of remaining scalarizations, B_{m+1}, and the confidence interval factor $\tilde{\Delta}_{m+1}$ [51].

In the study conducted by Drugan, Nowé, and Manderick, explored in [53], an enhanced version of the Upper Confidence Bound algorithm, termed the Improved Pareto Upper Confidence Bound (iPUCB) algorithm, is presented. This variant is designed for multi-objective multi-armed bandit (MOMAB) scenarios, where the challenges include handling multiple, potentially conflicting, objectives. Specifically, the iPUCB algorithm aims to efficiently identify the Pareto optimal set by eliminating suboptimal choices. It accomplishes this by comparing each candidate against the established Pareto front and removing those whose average performance is confidently deemed inferior [53].

An important aspect of their investigation is the introduction of a novel regret measurement derived from the Kullback-Leibler divergence [53]. This metric is tailored to more precisely assess the efficacy of MOMAB algorithms. The results from their experiments suggest a notable advancement in performance with the iPUCB algorithm, particularly in its ability to discern and discard lesser options, thus refining the balance between exploration and exploitation in environments governed by multiple objectives [53].

Their methodology involved evaluating the proposed algorithm against a bi-objective Bernoulli reward distribution, inspired by the practical scenario of optimizing a wet clutch system [53]. Both theoretical analyses and empirical evidence support the conclusion that the iPUCB algorithm stands as a superior alternative to the Pareto UCB1 algorithm, demonstrating efficiency in navigating the complex landscape of multi-objective optimization [53].

8.2.2 Linear scalarized and Pareto Thompson Sampling

In the domain of multi-objective multi-armed bandits (MOMABs), Yahyaa & Manderick [199] introduce an innovative adaptation of Thompson Sampling, a method traditionally applied in single-objective scenarios, to enhance the decision-making process within multi-objective frameworks. This adaptation, manifested through *Linear Scalarization Function Thompson Sampling* (LSF-TS) and *Pareto Thompson Sampling* (PTS), aims to balance the exploration-exploitation trade-off across multiple objectives by assigning a selection probability to each arm within each objective. In the paper presented by Yahyaa & Manderick [199], these probabilities are derived from a Beta distribution, facilitating the selection process based on either a linear scalarization of these probabilities (LSF-TS) or the Pareto dominance relation (PTS) [199]. In the rest of this section, notation has been changed ftom using a Beta prior to a general multivariate prior distribution $\pi(\cdot)$ over the arms' reward distributions.

LSF-TS, shown in Algorithm 11, simplifies the multi-objective problem into a single-objective one by applying a linear scalarization function to the samples for each of the objectives associated with each arm, thus selecting the arm with the highest scalarized function value. The scalarization

Algorithm 12: Linear Scalarization Function Thompson Sampling (LSF-TS), from [199]

Input: A MOMAB with K arms, set of scalarization functions $S = (f^1, \cdots, f^s, \cdots, f^S)$, a prior $\pi(\cdot)$ and history $\mathcal{H}^{(0)} = \emptyset$

for $t \leftarrow 1$ **to** $+\infty$ **do**

- Select a function $f^s \in S$ uniformly at random
- **for** $i \leftarrow 1$ **to** K **do**
  - $\tilde{\boldsymbol{\mu}}_a^{(t)} \sim \pi(\cdot|\mathcal{H}^{(t-1)})$
- Select the optimal arm $a_s^{(t)}$ that maximizes $f^s(\tilde{\boldsymbol{\mu}}_a^{(t)})$
- Play $a_s^{(t)}$ and observe $\mathbf{r}^{(t)}$
- $\mathcal{H}^{(t)} \leftarrow \mathcal{H}^{(t-1)} \cup \{a_s^{(t)}, \mathbf{r}^{(t)}\}$

function effectively transforms the vector of samples obtained for each objective, into a single scalar expected reward. Conversely, PTS employs the Pareto dominance relation to identify a set of non-dominated arms, known as the Pareto front, and selects uniformly at random from this set to ensure optimal exploration while exploiting the best-performing arms [199].

The operational framework of PTS begins with an initial assumption that success and failure rates are equal across all objectives for each of the arms. At each timestep, PTS samples a vector of expected rewards for each arm from the posterior distribution $\pi(\cdot|\mathcal{H}^{(t-1)})$, using these sampled probabilities to select non-dominated arms based on Pareto dominance [199]. A randomly chosen arm from this optimal set is then pulled, and the outcomes are used to update the success and failure rates contributing to the calculation of Pareto and unfairness regrets.

Algorithm 13: Pareto Thompson Sampling (PTS), from [199]

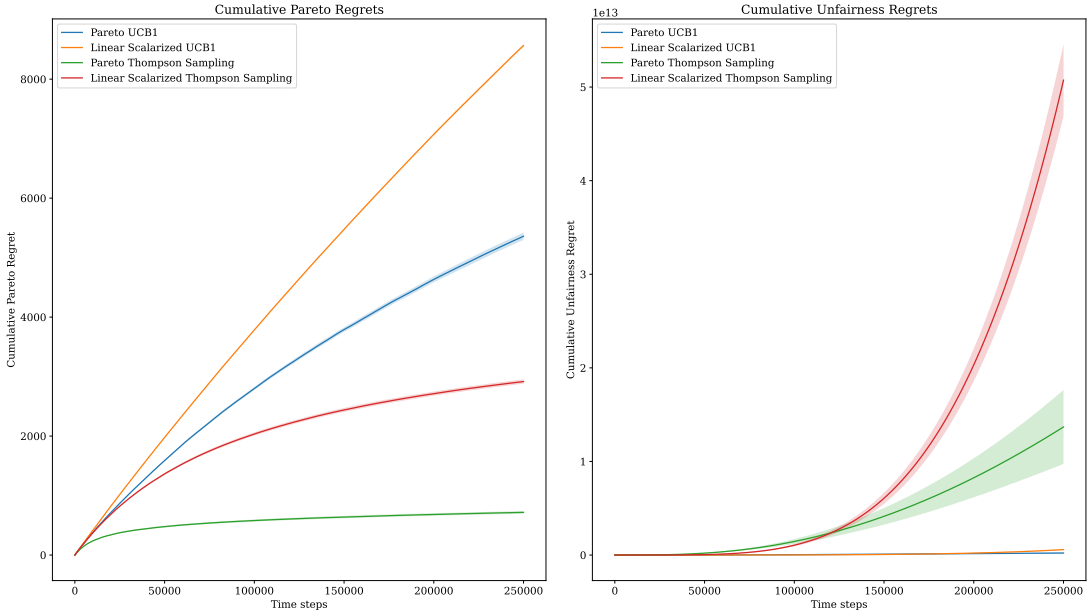
Input: A MOMAB with K arms, a prior $\pi(\cdot)$ and history $\mathcal{H}^{(0)} = \emptyset$

for $t \leftarrow 1$ **to** T **do**

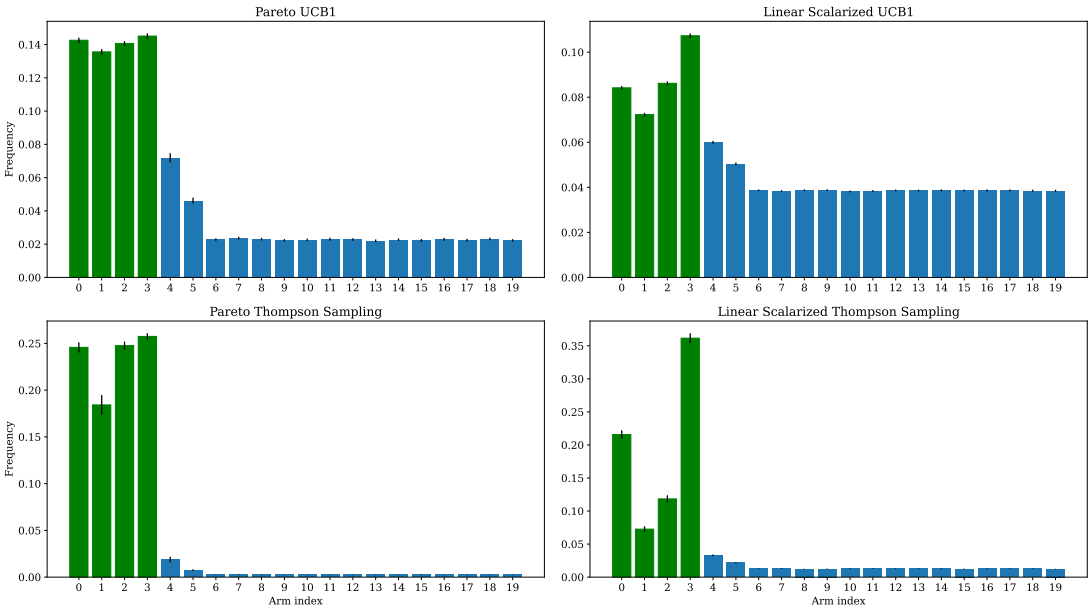
- **for** $i \leftarrow 1$ **to** K **do**
  - $\tilde{\boldsymbol{\mu}}_a^{(t)} \sim \pi(\cdot|\mathcal{H}^{(t-1)})$
- Select the Pareto optimal arms $\mathcal{A}^*$ such that $\forall i \in \mathcal{A}^*$ and $\forall j \notin \mathcal{A}^*$, $\tilde{\boldsymbol{\mu}}_j^{(t)} \not\succ \tilde{\boldsymbol{\mu}}_i^{(t)}$
- Select $a^{(t)}$ uniformly at random from $\mathcal{A}^*$
- Play $a^{(t)}$ and observe $\mathbf{r}^{(t)}$
- $\mathcal{H}^{(t)} \leftarrow \mathcal{H}^{(t-1)} \cup \{a^{(t)}, \mathbf{r}^{(t)}\}$

Similarly, LSF-TS operates under the premise that each arm initially has the same expected reward for each of the objectives. It then selects a scalarization function at random, samples the probability vectors under this function using the posterior distribution $\pi(\cdot|\mathcal{H}^{(t-1)})$, and performs linear scalarization to transform the multi-objective problem into a single-objective one. The optimal arm is selected based on the maximization of this scalarized function, with subsequent pulls updating success and failure rates and contributing to regret calculations [199].

Through the introduction of LSF-TS and PTS within the MOMAB framework, Yahyaa & Manderick [199] extend the utility of Thompson Sampling to multi-objective contexts. Their comparative analysis reveals that PTS exhibits superior performance over LSF-TS in minimizing Pareto and unfairness regrets, thus highlighting the effectiveness of incorporating the Pareto dominance principle in multi-objective decision-making processes. This advancement underscores the poten-



(a) Cumulative Pareto regrets and cumulative unfairness regret for the Pareto Thompson Sampling/UCB1 and Linear scalarized Thompson sampling/UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean.



(b) Arm pull frequencies for the Pareto Thompson Sampling/UCB1 and Linear scalarized Thompson sampling/UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean frequency. Pareto optimal arms are shown in green.

Figure 8.3: Comparison of the average Pareto and unfairness regrets, as well as the respective average arm pull frequencies between Pareto Thompson Sampling/UCB1 and Linear scalarized Thompson Sampling/UCB1.

tial of Thompson Sampling in navigating the complex landscape of multi-objective optimization, offering a robust methodology for achieving balanced and informed decisions across diverse objectives.

8.2.3 Linear scalarized and Pareto Knowledge Gradient

In their works from late 2014 [202] and early 2015 [198, 201], Saba Q. Yahyaa, Madalina M. Drugan, and Bernard Manderick propose an extension of the knowledge gradient (KG) policy for the multi-objective multi-armed bandit (MOMAB) problem with the aim to efficiently identify and exploit the Pareto optimal arms. in their papers, they propose three variants of these new KG based MOMAB algorithms: *Pareto Knowledge Gradient*, *Scalarized Knowledge Gradient by arms*, and *Scalarized Knowledge Gradient by objectives* [198, 201, 202].

The Knowledge Gradient policy, introduced by I.O. Ryzhov and Frazier [153], assigns to each arm $i \in \mathcal{A}$, with $|\mathcal{A}| = K$, an index V_i^{KG} as follows:

$$V_i^{KG} = \hat{\bar{\sigma}}_i * \left(- \left| \frac{\boldsymbol{\mu}_i - \max\limits_{j \neq i, j \in K} \boldsymbol{\mu}_j}{\hat{\bar{\sigma}}_i} \right| \right)$$

where $\hat{\bar{\sigma}}_i = \frac{\sigma_i}{n_i}$ is the Root Mean Square Error (RMSE) of the estimated mean of arm i [198, 201]. The auxiliary function $x(\zeta) = \zeta\Phi(\zeta) + \phi(\zeta)$ integrates the standard normal density $\phi(\zeta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta}{2}\right)$ with its cumulative distribution $\Phi(\zeta)$ [198, 201]. The KG policy prefers the arm i with the highest index V_i^{KG}, particularly those arms that have been explored less [198, 201]. These are characterized by larger standard deviations $\hat{\sigma}_i$, implying limited knowledge about their actual means $\boldsymbol{\mu}_i$ [198, 201, 202]. KG thus embodies a trade-off between the exploration of less familiar arms and the exploitation of known performers by selecting the arm $a^{(t)}$ as follows:

$$a^{(t)} = \arg\max_{i \in K} \left(\boldsymbol{\mu}_i + (T - t) V_i^{KG} \right).$$

Here, t represents the time step and T the total number of arm pulls, essentially the experiment's horizon [198, 201, 202].

[196] suggest that the KG policy is a formidable contender in the single-objective multi-armed bandit landscape. Moreover, the KG policy's lack of parameters requiring tuning positions it as an intuitive choice for the MOMAB problem [198, 201, 202].

The Pareto Knowledge Gradient (Pareto-KG) employs the Pareto partial order to organize arms [198, 201, 202]. The pseudocode for Pareto-KG is presented in Algorithm 13. At each time step t, Pareto-KG computes an exploration boundary ExpB for each arm a, denoted as $\text{ExpB}_a = [\text{ExpB}_a^1, \ldots, \text{ExpB}_a^D]$ [198, 201, 202]. This boundary is dependant upon the estimated means and standard deviations of all arms. Specifically, for a given dimension d, the exploration boundary ExpB_a^d is determined by the following calculation:

$$\text{ExpB}_a^d = (T - t) * KD * v_a^d$$

$$v_a^d = \hat{\bar{\sigma}}_a^d * x \left(- \left| \frac{\boldsymbol{\mu}_a^d - \max_{k \neq a, k \in \mathcal{A}} \boldsymbol{\mu}_k^d}{\hat{\bar{\sigma}}_a^d} \right| \right), \quad \forall d \in D$$

Algorithm 14: Pareto Knowledge Gradient, from [198, 201, 202]

Input: A MOMAB with K arms, number of objectives D, arm pull budget T, number of initialization phases P

for $p \leftarrow 1$ **to** P **do**
  $\forall a \in \mathcal{A}$, play arm a and update the estimate for $\boldsymbol{\mu}_a$ and $\boldsymbol{\sigma}_a^2$
for $t \leftarrow P \cdot K$ **to** T **do**
  for $a \leftarrow 1$ **to** K **do**
    Calculate $\boldsymbol{\mu}_a + \text{ExpB}_a$
  Select the Pareto optimal arms $\mathcal{A}^*$ such that $\forall i \in \mathcal{A}^*$ and $\forall j \notin \mathcal{A}^*$,
   $\boldsymbol{\mu}_j + \text{ExpB}_j \not\succ \boldsymbol{\mu}_i + \text{ExpB}_i$
  Select $a^{(t)}$ uniformly at random from $\mathcal{A}^*$
  Play $a^{(t)}$ and observe $\mathbf{r}^{(t)}$
  Update $\boldsymbol{\mu}_{a^{(t)}}$, $\boldsymbol{\sigma}^2_{a^{(t)}}$, and $n_{a^{(t)}}$

where v_a^d is the index of arm a for dimension d, T is the experiment's horizon representing the total number of time steps, K is the total number of arms, D is the number of dimensions, and $\hat{\bar{\sigma}}_a^d$ is the root mean square error of an arm for dimension d which equals $\hat{\sigma}_a^d/\sqrt{n_a}$ [198, 201, 202]. n_a denotes the number of times arm a has been played. Following the calculation of each arm's exploration boundary, Pareto-KG aggregates the exploration bound of arm a with the respective estimated mean [198, 201, 202]. Consequently, Pareto-KG selects the optimal arms i that are not dominated by any other arms. Pareto-KG then chooses uniformly and at random one of the optimal arms $a^{(t)} \in \mathcal{A}^*$, where $\mathcal{A}^*$ encompasses the set of Pareto optimal arms as identified by the KG policy [198, 201, 202]. Subsequent to the selection of the chosen arm $a^{(t)}$, Pareto-KG updates the estimated mean $\boldsymbol{\mu}_{a^{(t)}}$ vector, the estimated standard deviation $\boldsymbol{\sigma}^2_{a^{(t)}}$ vector, and the number of times arm $a^{(t)}$ has been selected $N_{a^{(t)}}$ [198, 201, 202].

The Linear Scalarized-Knowledge Gradient across arms strategy (LS1-KG) immediately transforms the multi-objective estimated mean $\boldsymbol{\mu}_i$ and estimated standard deviation $\boldsymbol{\sigma}_i^2$ of each arm into a single-dimensional format, followed by the calculation of the corresponding exploration boundary ExpB_i [198, 201, 202]. At each step t, LS1-KG assigns weights to both the estimated mean vector and the estimated variance vector, for each arm i, thereby reducing the multi-dimensional vectors to a singular dimension by summing their elements. For each arm, KG formulates an exploration boundary that is contingent on the metrics of all other arms and selects the arm that maximizes the sum of the estimated mean and the exploration bounds [198, 201, 202].

The process as implemented by LS1-KG is as follows:

$$\tilde{\mu}_i = f^j(\boldsymbol{\mu}_i) = w^1\hat{\mu}_i^1 + \ldots + w^D\hat{\mu}_i^D \qquad \forall i$$

$$\tilde{\sigma}_i^2 = f^j(\boldsymbol{\sigma}_i^2) = w^1\hat{\sigma}_i^{2,1} + \ldots + w^D\hat{\sigma}_i^{2,D} \qquad \forall i$$

$$\tilde{\bar{\sigma}}_i^2 = \frac{\hat{\sigma}_i^2}{n_i} \qquad \forall i$$

$$v_i = \tilde{\bar{\sigma}}_i \times x\left(-\left|\frac{\tilde{\mu}_i - \max\limits_{j\neq i, j\in\mathcal{A}} \tilde{\mu}_j}{\tilde{\bar{\sigma}}_i}\right|\right) \qquad \forall i$$

where f^j is a linear scalarization function defined by a set of predetermined weights $(w^1, \ldots, w^D)$.

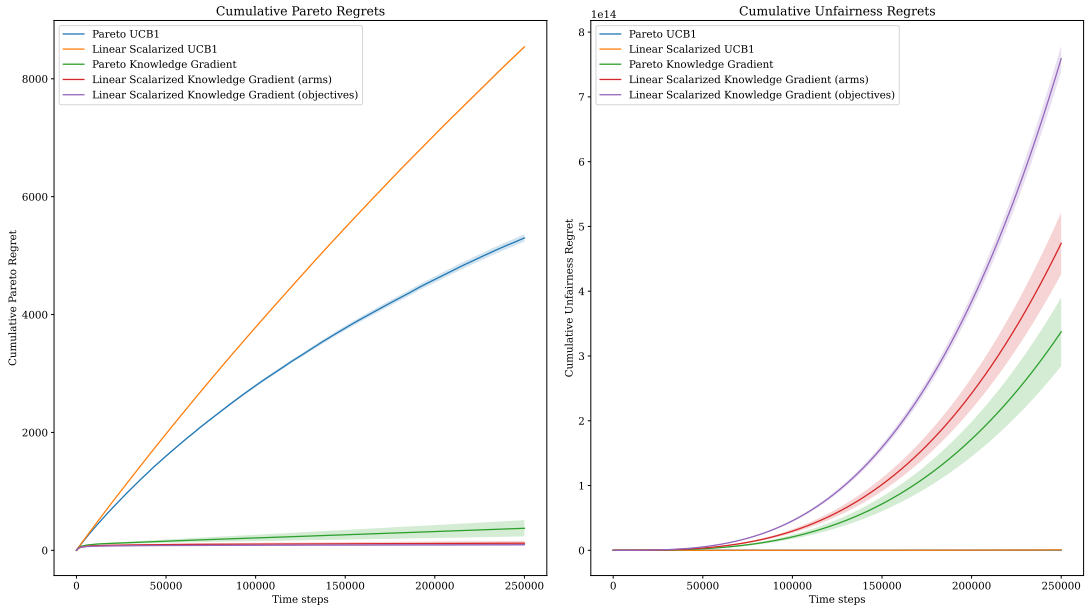


Figure 8.4: Cumulative Pareto regrets and cumulative unfairness regret for the Pareto Knowledge Gradient/UCB1 and Linear scalarized Knowledge gradient/UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean.

The transformed estimated mean and variance for arm i are given by $\tilde{\mu}_i$ and $\tilde{\sigma}_i^2$, respectively. These are one-dimensional values where $\tilde{\tilde{\sigma}}_i^2$ represents the modified RMSE of arm i. The index v_i stands as the KG index for arm i, obtained through the scalarization function [198, 201, 202]. Subsequently, LS1-KG chooses the arm with the highest aggregate of the estimated mean and the exploration bound as the optimal arm i^*_{LS1-KG}:

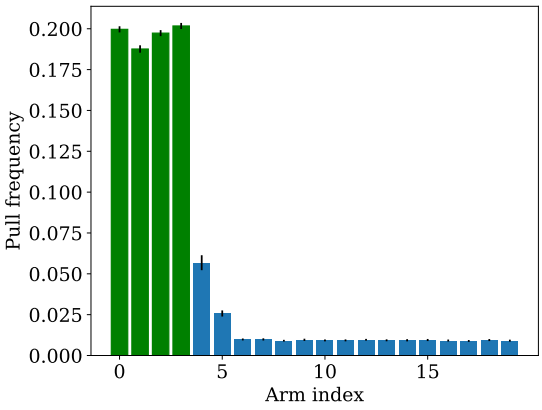
$$\begin{aligned} i^*_{LS1-KG} &= \max_{i=1,\ldots,K} (\tilde{\mu}_i + \text{ExpB}_i) \\ &= \max_{i=1,\ldots,K} (\tilde{\mu}_i + (T-t) \times KD \times v_i) \end{aligned}$$

The Linear Scalarized-Knowledge Gradient across dimensions (LS2-KG) calculates the exploration bound ExpB_i for each arm, that is $\text{ExpB}_i = [\text{ExpB}_i^1, \ldots, \text{ExpB}_i^D]$, and incorporates it into the respective estimated mean vector $\boldsymbol{\mu}_i$. This approach essentially reduces the multi-objective problem into a single dimension. At each iteration t, LS2-KG evaluates the exploration bounds for every dimension of each arm, aggregates the estimated means across dimensions together with their corresponding exploration bounds, and distills the information into a single scalar by summing over the vectors of each arm. The operation of LS2-KG across dimensions is delineated as follows:

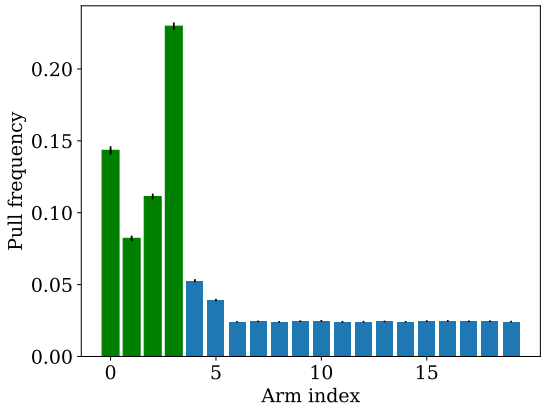
$$f^j(\boldsymbol{\mu}_i) = w^1(\hat{\mu}_i^1 + \text{ExpB}_i^1) + \ldots + w^D(\hat{\mu}_i^D + \text{ExpB}_i^D) \quad \forall i$$

The LS2-KG method selects the optimal arm $a^{(t)}$ that yields the maximum $f^j(\boldsymbol{\mu}_i)$:

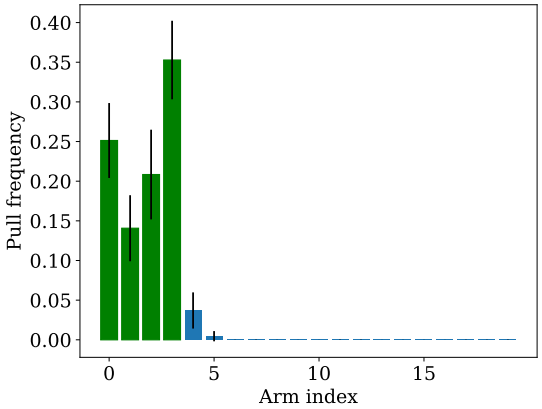
$$i^*_{LS2-KG} = \arg \max_{i=1,\ldots,K} f^j(\boldsymbol{\mu}_i)$$



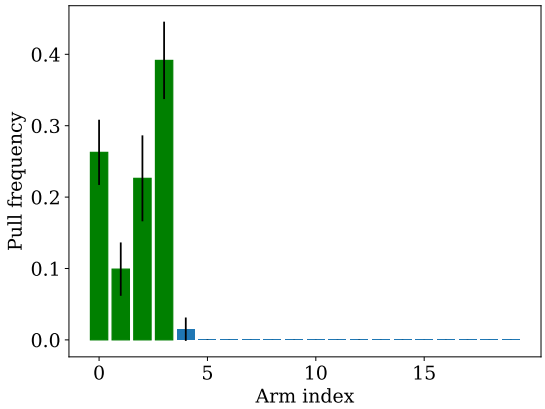
(a) Pareto UCB1 arm pull frequencies



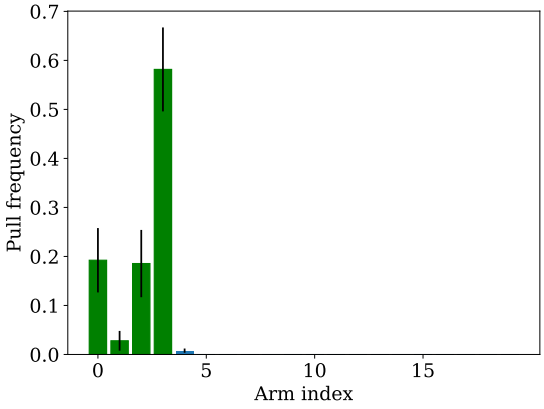
(b) Linear scalarized UCB1 arm pull frequencies



(c) Pareto Knowledge Gradient arm pull frequencies



(d) Linear scalarized Knowledge Gradient across arms arm pull frequencies



(e) Linear scalarized Knowledge Gradient across objectives arm pull frequencies

Figure 8.5: Arm pull frequencies for the Pareto Knowledge Gradient/UCB1 and Linear scalarized Knowledge gradient/UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean frequency. Pareto optimal arms are shown in green.

8.2.4 Annealing Pareto

In their work on the multi-objective multi-armed bandit (MOMAB) problem, Saba Q. Yahyaa, Madalina M. Drugan, and Bernard Manderick propose the Annealing Pareto algorithm as a solution to the exploration-exploitation trade-off inherent in such problems [200, 197]. The MOMAB problem involves an agent selecting one arm from a set of arms at each time step, with each arm providing a reward vector across multiple objectives. The goal is to efficiently explore and exploit the set of Pareto optimal arms.

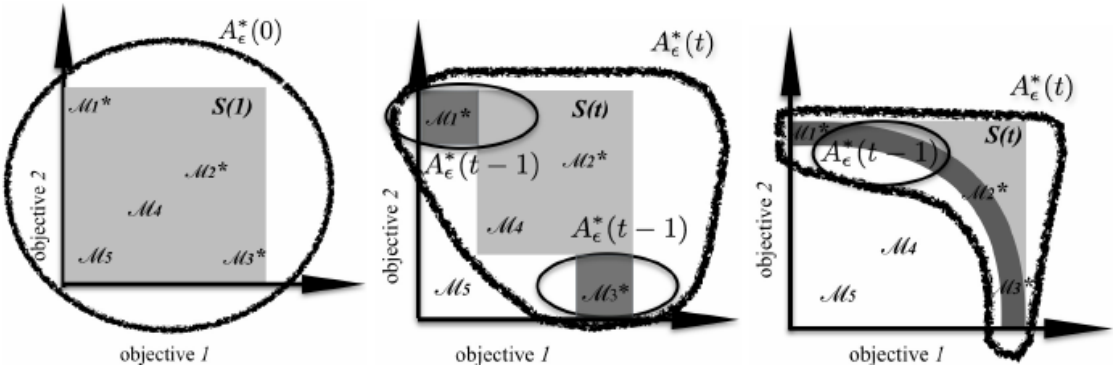


Figure 8.6: The intuition behind the Annealing Pareto algorithm. From [197].

The Annealing Pareto algorithm, shown in Algorithm 14, introduces a decaying parameter, ϵ_t, which balances exploration and exploitation over time. Initially, ϵ_t is high to encourage exploration, but it decreases exponentially to focus more on exploitation as time progresses [200, 197]. This approach ensures that the algorithm starts with a broad exploration of all arms, but gradually narrows down to exploit the most promising arms.

The Annealing Pareto algorithm uses the Pareto dominance relation to track the optimal arms and updates the ϵ-Pareto front at each time step [200, 197]. The intuition behind this is visualized in Figure 8.6. The parameter ϵ_t decreases over time, transitioning the focus from exploration to exploitation. This method ensures a comprehensive initial exploration phase, gradually honing in on the arms with the highest estimated means as the experiment progresses.

In the original study [200] performance of the Annealing Pareto algorithm was compared against other algorithms such as Pareto-UCB1, Pareto-KG, and Pareto Thompson Sampling in a multi-objective Bernoulli distribution setup. The experiments demonstrated that the Annealing Pareto algorithm outperforms the others in terms of both cumulative Pareto regret and unfairness regret. However, these experiments were limited to 1,000 arm pulls. The experiments with 250,000 arm pulls conducted within the context of this dissertation, visualized in Figures 8.7 and 8.8, show that the annealing algorithm is very effective at identifying some Pareto optimal arm and consistently exploiting this arm, thus resulting in a very low Pareto regret. It also becomes clear that the Annealing Pareto algorithm does not exploit all arms within the Pareto front equally, resulting in a significantly higher unfairness regret when compared to the baseline UCB MOMAB algorithms.

The results validate that the Annealing Pareto algorithm effectively manages the exploration-exploitation trade-off by dynamically adjusting the exploration parameter, leading to good performance in multi-objective optimization tasks.

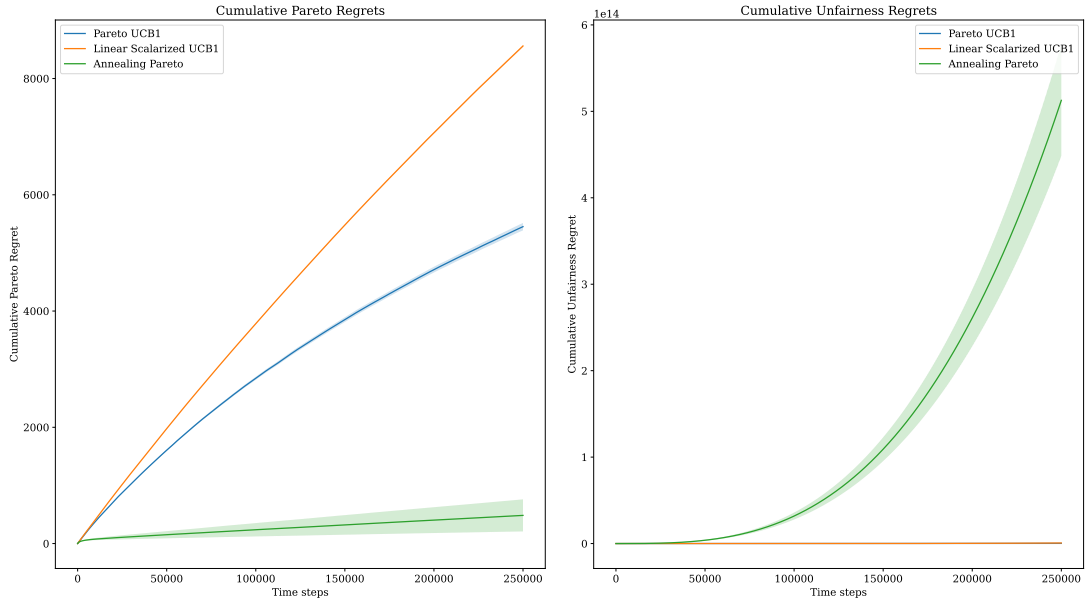
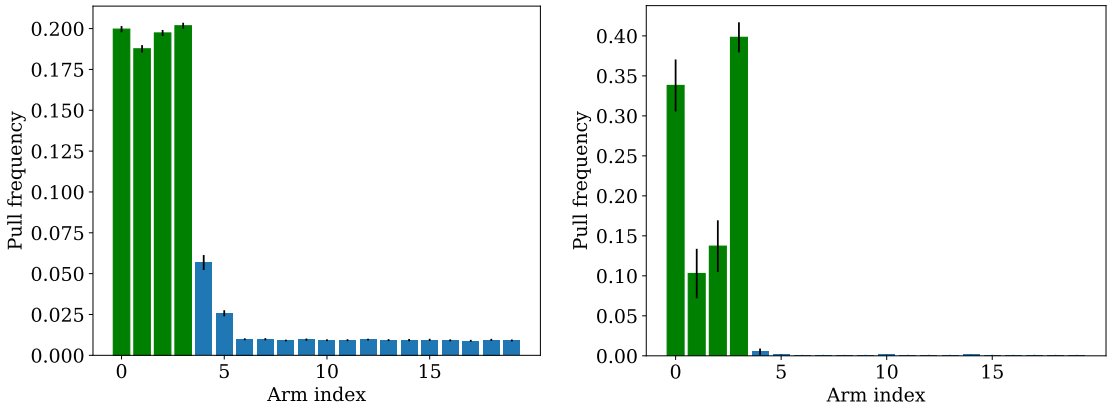


Figure 8.7: Cumulative Pareto regrets and cumulative unfairness regret for the Annealing Pareto, Pareto UCB1 and Linear scalarized UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean.



(a) Pareto UCB1 arm pull frequencies

(b) Annealing Pareto arm pull frequencies

Figure 8.8: Arm pull frequencies for the Annealing Pareto and Pareto UCB1 algorithms over 250,000 arm pulls, averaged over 100 runs of the experiment, together with the 95% confidence interval around the mean frequency. Pareto optimal arms are shown in green.

Algorithm 15: Annealing Pareto, from [200]

Input: A MOMAB with K arms, number of objectives D, an intial ϵ_t and decay parameter $\epsilon_{decay} \in [0, 1]$, number of initialisation phases P

initial ϵ-Pareto front set $\mathcal{A}^*_\epsilon(0) \leftarrow \mathcal{A}$

for $p \leftarrow 1$ **to** P **do**

  $\forall a \in \mathcal{A}$, play arm a and update the estimate for $\boldsymbol{\mu}_a$

for $t \leftarrow P \cdot K$ **to** $+\infty$ **do**

  $\epsilon_t = \epsilon_{decay}^{t/(KD)}$ `// Update decaying parameter`

  for $d \leftarrow 1$ **to** D **do**

    $S^d(t) = \emptyset$

    $\boldsymbol{\mu}^*_d = \max_{1 \le i \le \mathcal{A}} \boldsymbol{\mu}^d_i$

    for $a \leftarrow 1$ **to** K **do**

      if $\boldsymbol{\mu}^d_a \in [\boldsymbol{\mu}^*_d - \epsilon_t, \boldsymbol{\mu}^*_d]$ **then**

        $S^d(t) \leftarrow S^d(t) \cup \{a\}$

  $S(t) \leftarrow S^1(t) \cup S^2(t) \cup \ldots \cup S^D(t)$

  $S_{difference} \leftarrow \mathcal{A}^*_\epsilon(t-1) - S(t)$

  for $a \in S_{difference}$ **do**

    if $\boldsymbol{\mu}_k \not\succ \boldsymbol{\mu}_a, \ \forall k \in \mathcal{A}$ **then**

      $S(t) \leftarrow S(t) \cup \{a\}$

  $\mathcal{A}^*_\epsilon(t) \leftarrow S(t)$

  Select an optimal arm $a^{(t)}$ uniformly, at random from $\mathcal{A}^*_\epsilon(t)$

  Play $a^{(t)}$ and observe $\mathbf{r}^{(t)}$

  Update $\boldsymbol{\mu}_{a^{(t)}}$

8.3 Evolutions and variants of the MOMAB framework

In this final section of the literature review on multi-objective multi-armed bandits, a number of interesting evolutions and variants of the MOMAB framework will be discussed.

One 2018 study by Drugan introduces a novel approach to the MOMAB regret minimization problem [52]. The paper titled "Covariance Matrix Adaptation for Multiobjective Multiarmed Bandits" explores the use of an extended version of the Upper Confidence Bound (UCB) algorithm. The proposed algorithm, Covariance Matrix Adaptation for Pareto UCB (CMA-PUCB), addresses the complexity of handling stochastic reward vectors with correlated objectives.

One of the significant contributions of this paper is the derivation of an upper bound for the regret created by pulling suboptimal arms. The bound is expressed in terms of the logarithmic number of arms K, objectives D, and samples n [52]. Moreover, the paper addresses the challenge of unknown covariance matrices between objectives and provides an upper bound for approximating these matrices. This aspect of the study is crucial for practical implementations where covariance structures are not known a priori [52].

To validate the theoretical findings, Drugan conducts simulations in a three-objective stochastic environment. The results from these simulations demonstrate the applicability and effectiveness of the CMA-PUCB algorithm in real-world multiobjective optimization problems. This empirical evidence supports the theoretical contributions and showcases the potential of the proposed

method in various applications [52].

Part IV

Contributions

Chapter 9

Dengue epidemic modelling

9.1 Reproduction of the 2016 Ferguson et al. model

In this section the partial reproduction of the Dengue virus epidemiological model proposed in 2016 by Ferguson et al. [61] will be discussed. At the time of conducting their research, the first internationally approved dengue vaccine, the Sanofi-Pasteur Dengvaxia [82] vaccine, had been approved in six countries. In their paper titled "Benefits and Risks of the Sanofi-Pasteur Dengue Vaccine: Modeling Optimal Deployment" [61] the authors discuss the development, efficacy, and potential public health impact of the first approved dengue vaccine, Dengvaxia [61]. The authors mention that the development of this vaccine was relatively difficult when compared to vaccines for other flavivirus infections [61, 13]. Many studies agree that challenges arise with the development of highly efficient and safe dengue vaccines due to the interactions between the four different co-circulating dengue virus (DENV) serotypes. These interactions on the immunological level cause immune-mediated enhancement of disease, also often referred to as antibody-dependent enhancement or ADE [61]. Because of these adverse effects, modelling the influence of the introduction of vaccination into the population is crucial. This allows decision makers and public health specialists to make informed decisions about when and how to deploy the available vaccines, while also being informed about the possible negative effect the vaccine might have in some individuals. It is exactly the goal of this dissertation to present the decision makers with the entire range of possible trade-offs of all potentially optimal vaccination strategies.

Dengvaxia is a recombinant chimeric live attenuated DENV vaccine based on a yellow fever 17D vaccine backbone [61]. It was evaluated in two large multicenter phase III trials: one in Southeast Asia involving about 10,000 children aged 2 to 14 years [39], and another in Latin America with about 21,000 children aged 9 to 16 years [28]. The trials reported approximately 60% efficacy against virologically confirmed symptomatic dengue disease and higher efficacy against severe dengue [61, 39, 28]. The efficacy varied by serotype and was significantly higher in seropositive recipients compared to seronegative ones at the time of vaccination.

In order to help future trials and provide additional insights into why these results might be obtained, they developed an advanced model describing the dynamics of dengue fever within a population with support for the simulation of different vaccination strategies. Because of these features, this model was the ideal candidate to reproduce within the framework of this dissertation.

The authors developed mathematical models of DENV transmission to explore the vaccine's action and predict the potential consequences of its routine use. In their paper, they describe how they first attempted to produce a simple model that had waning vaccine protection over time, with a different starting protection/immunity depending on the serostatus of the individual receiving the vaccine [61]. However, this model fit the data produced by the previously conducted trials poorly [61].

This then led them to propose a different, more advanced model: a deterministic vector-host model that consists of a number of compartments that partition the mosquito vector population and the human host population [61]. The main contribution of this model, within the field of modeling vaccination strategies against DENV, is that the immunological effect of vaccination is modeled like a silent infection with one of the four DENV serotypes [61]. This means seronegative recipients of the vaccine gain immunity that is comparable to the levels of immunity observed in individuals who have experienced a single natural infection [103, 154, 155] with one of the four heterologous DENV serotypes [61]. Conversely, vaccinating individuals who have previously experienced a single DENV infection enhances their immunity to levels akin to those observed in individuals with two natural infections. Consequently, their subsequent infection will not exhibit the heightened severity typically associated with natural secondary infections. Instead, they will experience a substantially lower risk of severe disease, akin to the risk seen in tertiary and quaternary infections [61]. This process is illustrated in Figure 9.1.

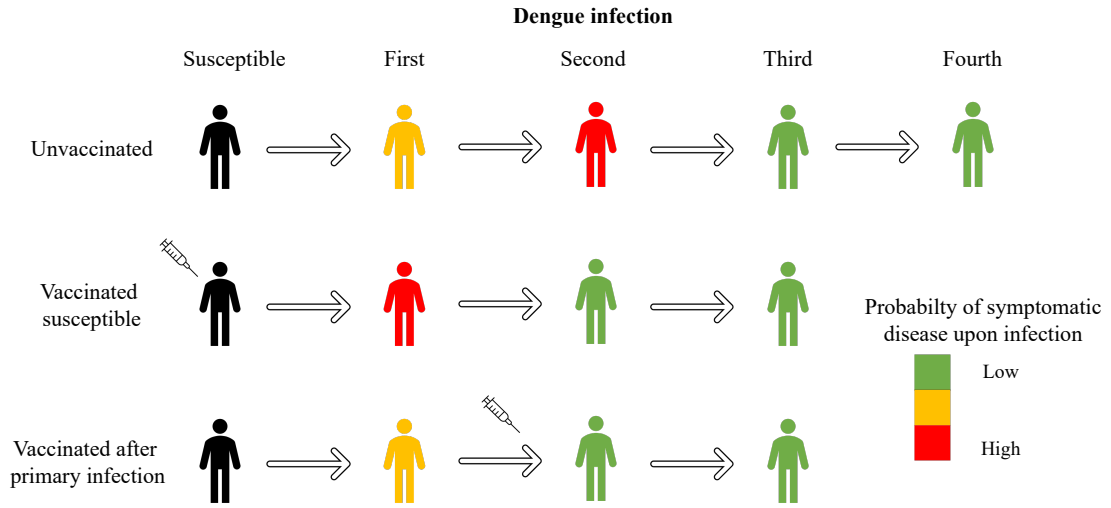


Figure 9.1: Overview of the different scenarios for the time at which people are vaccinated and the severity of their infections. Reproduced from [61].

Apart from their explanation about the representation of vaccination as silent modeling, the authors do not provide any additional information about the nature of their model in their main paper. This information was found in the supplementary materials document [60] that served as a guide for the actual reproduction of the model, but still provides relatively little information about certain aspects of the model such as configuration of the ODE solvers and the initial conditions used for the simulations. The model essentially consists of two major interacting components: (i) a component model describing the dynamics within the mosquito population and (ii) a component model describing the disease dynamics within the human population. The component model describing the disease dynamics within the human population will be discussed

first.

9.1.1 Host population disease dynamics model

As is often the case with epidemiological models, in the model proposed in [61, 60] the host population is compartmentalised into different compartments or groups depending on various factors. The model proposed by Ferguson et al. in 2016 essentially features three levels at which the human host population is stratified.

Firstly, the human host population is stratified based on their age into a series of continuous age groups. This is indicated by the a parameter in the differential equations describing their proposed model [60]. The motivation for this stratification by age group is the relative risks of infection endured by different age groups within the population. This allows for explicit studying of the effects of vaccinating different age groups on the disease dynamics within the entire population [61]. Within each of these continuous age groups, some individuals will be infectious with one of the DENV serotypes to vectors. The mosquitoes that become infected then might infect individuals from other age groups, resulting in an intricate interplay of infectious pressure between age groups. These interactions between age groups can be thought of as an almost fully connected graph where each of the nodes represent an age group and the edges represent infectious pressure between age groups. This can be seen visualized in Figure 9.2.

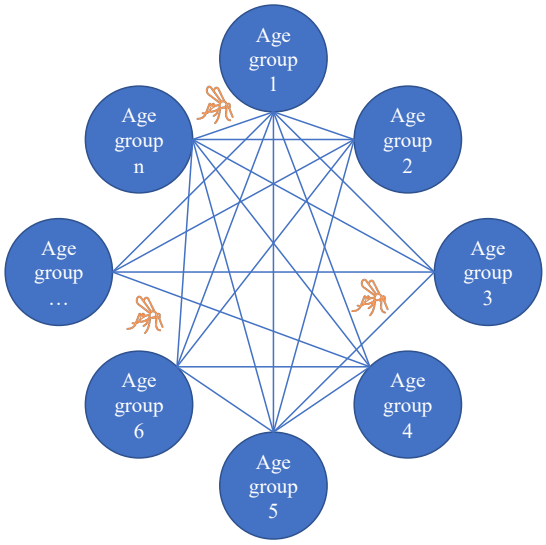


Figure 9.2: Mosquito driven interactions of infectious pressure between the different age groups in which the human host population is compartmentalised.

The next level of stratification takes place within each of the distinct age groups in which the host population was compartmentalised. Within each age group, individuals are assigned a compartment based on their vaccination status v. There are three possible statuses for vaccination an individual might have: 0 or unvaccinated, 1 or vaccinated while seronegative, and 2 or vaccinated while seropositive. The only possible change in vaccination status occurs when an individual is vaccinated. They then move from $v = 0$ to $v = 1$ or $v = 2$ based on their serostatus at the time of partaking in the vaccination program. Once again, the different compartments containing individuals based on their vaccination status will interact with one another through the vectors that

are also present within the environment. These interactions of infectious pressure between different vaccination status-based compartments as well as the composition of the stratification based on vaccination status, within the age-based stratification, can be seen schematised in Figure 9.3.

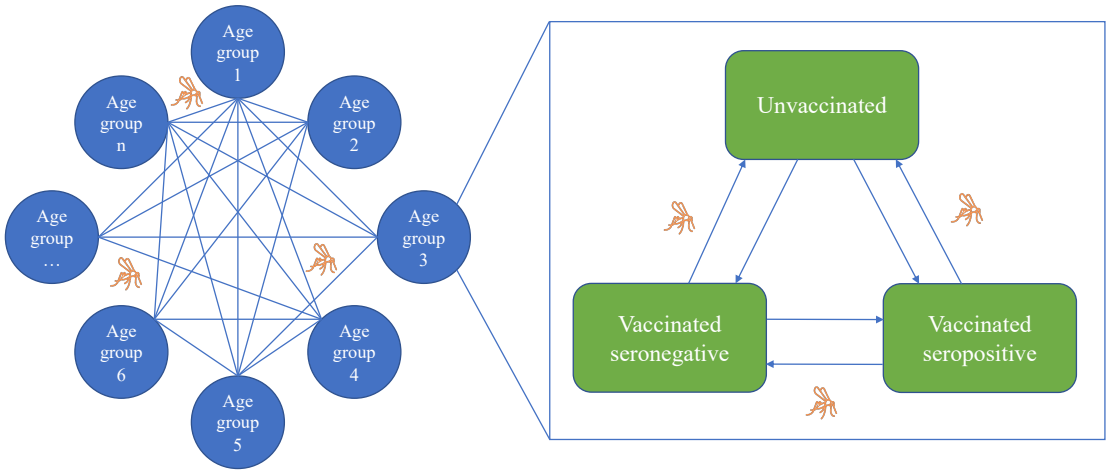


Figure 9.3: Mosquito driven interactions of infectious pressure between the different vaccination status-based groups in which the human host population is compartmentalised. The figure also shows the composition of the vaccination status-based stratification within the age-based stratification.

The third and final level of compartmentalization takes place within each of the vaccination status based compartments. Within each of these groups, the human host population can be structured into compartments based on their infection status. Within the model proposed by Ferguson et al. in 2016 [61], there are two different parameterised infection statuses an individual can have: S_θ and R_θ [60] where θ is a set of DENV serotypes $i \in [1..4]$. S_θ encapsulates all individuals who were previously infected with the serotypes in the set θ and are immune to any infection with them, but are currently susceptible to infection with any DENV serotype $i \notin \theta$ [60]. R_θ encapsulates all individuals who were previously infected with the serotypes in the set θ and are immune to any new infection with one of those serotypes, and are currently temporarily protected against heterologous infection due to a recent infection. As θ is a set containing any combination of DENV serotypes $i \in [1..4]$, the resulting model has huge amount of distinct compartments i which the human host population is stratified.

Finally, combining all three levels of compartmentalisation results in the following human-related state variables for the transmission model:

- $S_\theta^v(t, a)$: "Number of people of age a at time t with vaccine status v who were previously infected and are now immune to serotypes in the set θ, but remain susceptible to infection from all other serotypes $i \notin \theta$." [60]
- $R_\theta^v(t, a)$: "Number of people of age a at time t with vaccine status v who were previously infected and are now immune to serotypes in the set θ and are currently temporarily protected against heterologous infection." [60]

Figure 9.4 shows a diagram depicting the compartment model that is present within each vaccination status-based group v and age group a. In this figure, the compartments of the top three

rows of sequential infections with heterologous DENV serotypes have been collapsed. However, all compartments for the bottom row are depicted as an example. Compartments for the top rows can be represented analogously.

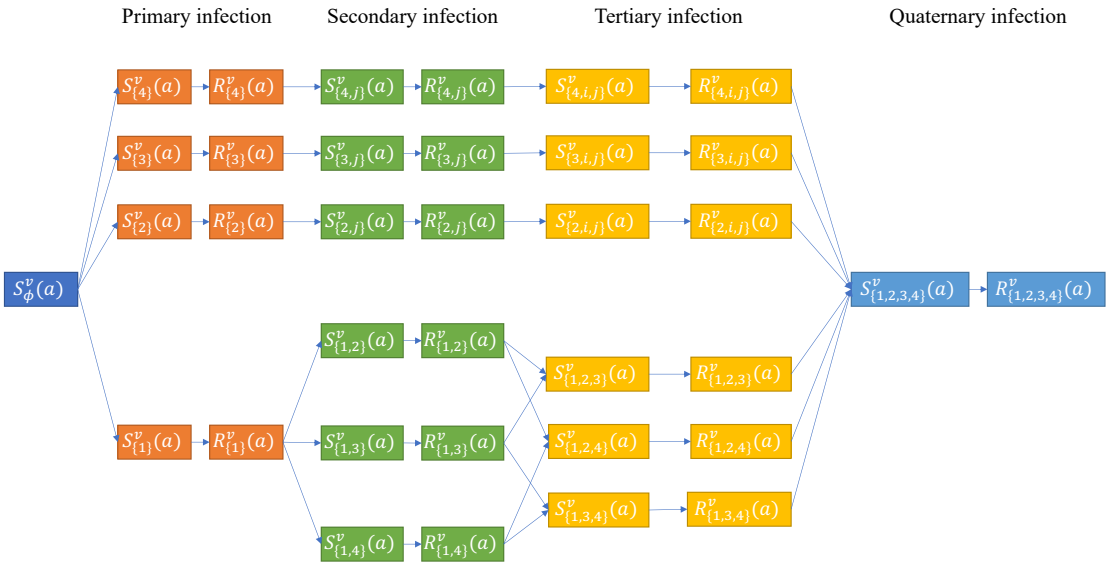


Figure 9.4: Schematic representation of the compartment model proposed by Ferguson et al. in [61]. From left to right, an individual can sequentially experience a total of four infections with each of the DENV serotypes. The model also stratifies the population according to continuous age classes a and a vaccination status v.

In this model, an individual starts in the dark blue compartments on the far left. They are seronegative and hence θ is equal to the empty set ϕ. Seronegative individuals then have a chance of becoming infected with one of the four different DENV serotypes and recovering from this infection. The compartments corresponding to this primary infection are shown in orange. After their primary infection, the individual can experience up to three more infections, depicted in green, yellow, and blue respectively.

The flow of individuals between the different compartments of the model is described by the following system of partial differential equations from [60]:

$$\frac{\partial S^v_\phi}{\partial t} + \frac{\partial S^v_\phi}{\partial a} = p_V(t)\delta(a - A_V)\left[\delta_{v,1}S^0_\phi - \delta_{v,0}S^v_\phi\right] - \sum_i \Lambda_i(t) f_v(a - A_V)S^v_\phi - \mu(a)S^v_\phi$$

$$\frac{\partial S^v_\theta}{\partial t} + \frac{\partial S^v_\theta}{\partial a} = \sigma R^v_\theta + p_V(t)\delta(a - A_V)\left[\delta_{v,2}S^0_\theta - \delta_{v,0}S^v_\theta\right] - \sum_{i\notin\theta} \Lambda_i(t) f_v(a - A_V)S^v_\theta - \mu(a)S^v_\theta$$

$$\frac{\partial R^v_\theta}{\partial t} + \frac{\partial R^v_\theta}{\partial a} = p_V(t)\delta(a - A_V)\left[\delta_{v,2}R^0_\theta - \delta_{v,0}R^v_\theta\right] + \sum_{i\in\theta} \Lambda_i(t) f_v(a - A_V)S^v_{\theta/i} - [\sigma + \mu(a)]R^v_\theta$$

Within this system, θ/i is the set θ with element i removed, $\delta(x)$ is the Dirac delta function, $\delta_{i,j}$ is the Kronecker delta, and $\mu(a)$ is the hazard of death for an individual of age a [60].

The age at which individuals are vaccinated is determined by A_V. To further allow simulations of possible vaccination strategies, at time t a proportion $p_V(t)$ of individuals can be vaccinated

[60]. This function of time is modeled as a step function where T_V is the time at which the vaccination campaign is commenced:

$$p_V(t) = \begin{cases} 0 & : t < T_V \\ p_{V0} & : t \geq T_V \end{cases}$$

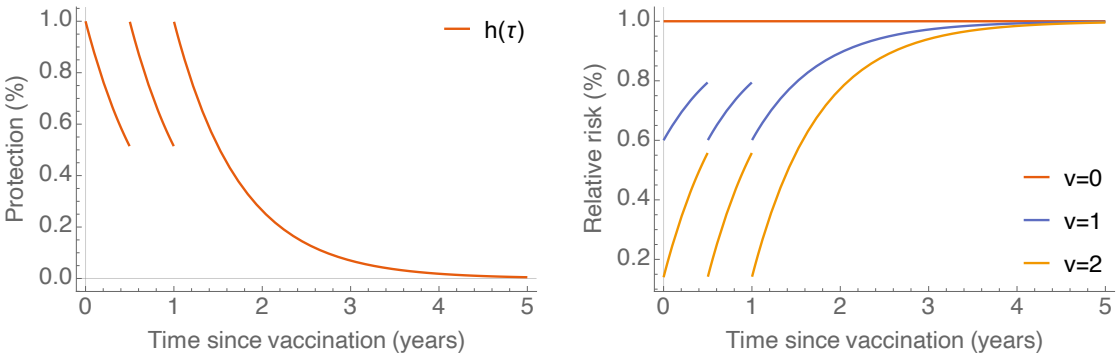
In the model proposed by the authors of the original paper, immunity of the human hosts to natural infection is assumed to have two components: (i) a period after infection with one of the serotypes during which the individual is completely protected against any other infection (R_θ^v), and (ii) permanent protection from all serotypes an individual has already been infected with [60]. The temporary immunity after infection has a duration of $1/\sigma$ [60]. Protection against infection as a result of vaccination is modeled to decay over time and is described in the supplementary materials of the original paper by the relative risk function $f_v(\tau)$ where τ represents the time that has passed since vaccination.

Unvaccinated individuals :	$f_0(\tau) = 1$
Individuals vaccinated while seronegative :	$f_1(\tau) = 1 - VE_- \times h(\tau)$
Individuals vaccinated while seropositive :	$f_2(\tau) = 1 - VE_+ \times h(\tau)$

Here, $h(\tau)$ is the decay function which models an exponential decay of protection after each dose of vaccination with mean duration T_D [60]:

$$h(\tau) = \begin{cases} exp(-\tau/T_D) & : \tau < 0.5 \\ exp(-(\tau - 0.5)/T_D) & : 0.5 < \tau < 1 \\ exp(-(\tau - 1)/T_D) & : \tau > 1 \end{cases}$$

Both the decay of the protection against infection provided by the vaccine $h(\tau)$, as well as the relative risk of infection based on the vaccination status of the individual $f_v(\tau)$, are illustrated in Figure 9.5.



(a) Exponential decay of the protection offered by the vaccine after each of the three administered doses as a function of the time since vaccination.

(b) Evolution of the relative risk of infection an individual experiences based on their vaccination status v and time since vaccination τ.

Figure 9.5: Visualization of waning protection $h(\tau)$ offered by vaccination (left panel) and the associated relative risk of infection $f_v(\tau)$ (right panel).

The final term in the differential equations describing the disease dynamics within the human host population that needs to be discussed is Λ_i. This term represents the force of infection

on humans due to serotype i and is the driving force behind new infections with that specific serotype. In [60] Λ_i is given by:

$$\Lambda_i = \frac{\kappa \beta_{mh}}{M} Y_i.$$

In this equation, κ represents the biting rate per mosquito, β_{mh} is the probability of transmission from mosquito to human, M is the total number of adult mosquitoes and Y_i represents the number of mosquitoes infectious with serotype i [60]. In this equation, the interactions between the host component model and the vector component model become apparent as the infectious pressure on humans is determined by the prevalence of the disease in the mosquito population, and vice versa.

9.1.2 Vector population disease dynamics model

In this section, the component model describing the disease dynamics within the vector population will be discussed. As dengue is a mosquito-borne disease, in [61] the mosquito population is modeled as a Ross-Macdonald type model [60]. In this type of model, the vector population is stratified into four main compartments: larvae (L), adult mosquitoes (A), infected mosquitoes in the incubation phase (H_i^j), and infectious mosquitoes (Y_i). A schematic visualization of the mosquito model presented in [60] can be seen in Figure 9.6.

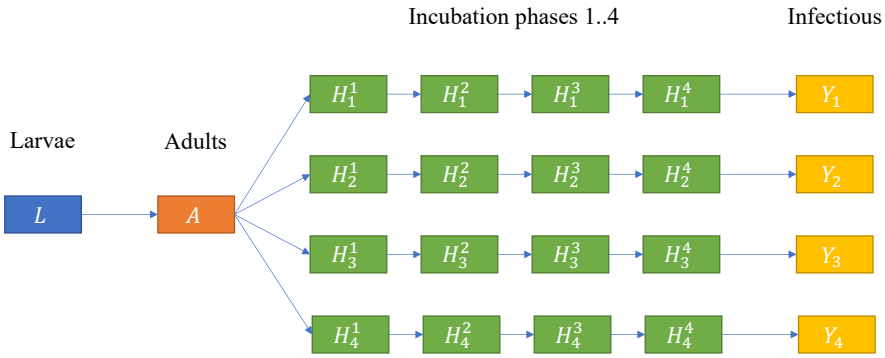


Figure 9.6: Schematic depiction of the Ross-Macdonald type model for describing the disease dynamics within the vector population proposed in [61].

In this model, infections with heterologous serotypes are not considered as the mean life expectancy of the vectors is very low [61, 60]. The flux of vectors between compartments over time is governed by the following system of differential equations from [60]:

$$\begin{aligned}
\frac{dL}{dt} &= bM - \alpha L - \omega L \left[1 + \frac{L}{K(t)}\right] \\
\frac{dA}{dt} &= \alpha L - \sum_i \Psi_i A - \epsilon A \\
\frac{dH_i^j}{dt} &= \delta_{1,j} \Psi_i A + 4\eta(1 - \delta_{1,j}) H_i^{j-1} - (4\eta + \epsilon) H_i^j \quad \text{for } 1 \leq j \leq 4 \\
\frac{dY_i}{dt} &= 4\eta H_i^4 - \epsilon Y_i
\end{aligned}$$

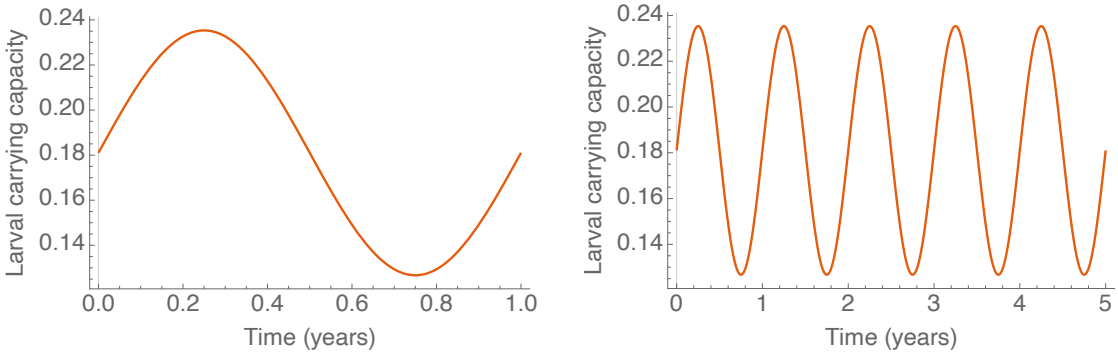
In this system of equations, M represents the total adult female mosquito population size:

$$M = A + \sum_{i,j} H_i^j + \sum_i Y_i,$$

b is the rate at which larvae are produced, $1/\alpha$ is the mean development time of larvae, ω is the mortality rate for larvae and ϵ is the mortality rate for adult mosquitoes [60]. The duration of the incubation period before an infected mosquito becomes infectious is given by η [60]. The model also takes into account the seasonal variability in the size mosquito population as a result of the oscillating larval carrying capacity K throughout the year [60].

$$K(t) = K_0[1 + \Delta_k \sin(2\pi t)]$$

In this equation, K_0 represents the base larval carrying capacity across the entire year, around which the actual K oscillates throughout the seasons with amplitude Δ_k where $0 \leq \Delta_k \leq 1$ [60]. Oscillation of the larval carrying capacity over different time horizons can be seen visualized in Figure 9.7.



(a) Evolution of the larval carrying capacity throughout a single year.

(b) Evolution of the larval carrying capacity over a duration of 5 years.

Figure 9.7: Oscillation of the larval carrying capacity $K(t)$ over a single year and over the course of 5 years. Within the period of each year, the carrying capacity reaches its maximum and minimum value, modeling the influence of seasonality.

In this model, infections within the vector population are driven by Ψ_i, the force of infection on mosquitoes due to serotype i [60]. Ψ_i is a complex expression that depends on the prevalence of the different heterologous serotypes within the human host population:

$$\Psi_i(t) = \frac{\kappa \beta_{hm}}{N} \int_0^\infty \int_{T_{\text{IP}}}^{T_{\text{IP}}+T_{\text{Inf}}} \sum_{\theta \not\supseteq i} \left[cI^v_{i|\theta}(t-\tau, a) + (\Theta - 1) cD^v_{i|\theta}(t-\tau, a) \right] d\tau da$$

The first part of this expression consists of the biting rate per adult mosquito κ, the transmission probability per bite from humans to mosquitoes β_{hm}, and the size of the entire human host population N [60]. Since the human host population was stratified into a series of continuous age groups, an integral over all these age groups is taken to effectively calculate the infinite sum of infectious pressure resulting from the infected individuals in each of these age groups. Humans are infectious as soon as the intrinsic incubation period T_{IP} ends and their infectiousness has a duration of T_{Inf} [60]. Only infectious individuals contribute to the force of infection, so the

integral is taken over this period of infectiousness to vectors. Within this double integral, a summation is computed over the sum of two terms:

- $cI^v_{i|\theta}(t,a)$: Which is defined in [60] as the "Incidence at time t of infection with serotype i in people with past exposure to serotype set θ, age a and vaccine status v.", and
- $cD^v_{i|\theta}(t,a)$: Which is defined in [60] as the "Incidence of symptomatic disease at time t due to infection with serotype i in people with past exposure to serotype set θ, age a and vaccine status v.".

These are exactly the terms representing the individuals that are infectious with serotype i. As individuals experiencing symptomatic disease carry a larger viral load, their influence in the overall force of infection increased by the parameter Θ [60].

9.1.3 Reproduction of the component models

Within the context of this dissertation, the model proposed in [61] looked like the perfect candidate for reproduction and use in the final DENV MOMAB setting. As the model proposed by Ferguson et al. in 2016 [61] is in essence a composition of two interacting component models describing the disease dynamics within the vector and the host population, the choice was made to start with the reproduction of each component model individually.

Starting with the reproduction of the mosquito model, the Wolfram Mathematica language was selected. This language and execution environment feature a powerful syntax for programming mathematical expressions in an elegant way that is close to the written form.

Within this framework, the decay of the protection against infection provided by the vaccine $h(\tau)$, as well as the relative risk of infection based on the vaccination status of the individual and the time passed since vaccination $f_v(\tau)$ were successfully reproduced. Results can be seen visualized in Figure 9.5. Then, the effect of seasonality on the larval carrying capacity was implemented and the result can be seen in Figure 9.7. The force of infection on humans Λ_i due to serotype i was also successfully reproduced and was then used to implement both the incidence of infection $cI^v_{i|\theta}(t,a)$ and incidence of symptomatic disease $cD^v_{i|\theta}(t,a)$. The reproduction of these components could then be used to reproduce the force of infection on mosquitoes due to serotype i, Ψ_i. Since the aim was to first reproduce each component model individually, all interactions with the human model were assumed to be constant temporarily, and Ψ_i was successfully implemented.

The implementation of Ψ_i then allowed for the reproduction of the system of ordinary differential equations governing the disease dynamics within the vector population. It is at this point that some problems arose.

The first problem was due to the integration in Ψ_i over the period within which humans can infect mosquitoes upon being bitten. There, a time-delay is introduced into the system of differential equations. The ODE solver that is present in Wolfram Mathematica is able to solve systems of delay-differential equations like the one that was being reproduced, however, it can only handle discrete delays. As the expression for Ψ_i from [60] requires an integration over the time-delayed terms, the differential equations presented in [60] have a continuous delay as opposed to a discrete delay. After careful consideration, the decision was made to change the expression for Ψ_i in the reproduced model to:

$$\Psi_i(t) = \frac{\kappa\beta_{hm}}{N}\int_0^\infty \sum_{\tau=T_{\mathrm{IP}}}^{T_{\mathrm{IP}}+T_{\mathrm{Inf}}} \sum_{\theta\not\ni i} \left[cI^v_{i|\theta}(t-\tau,a) + (\Theta-1)cD^v_{i|\theta}(t-\tau,a)\right] da$$

in which the integration is changed to a summation, resulting in a discrete delay that could be processed by Wolfram Mathematica's ODE solver.

With this change in place, and after careful review of the parameters used, the results shown in Figure 9.8 were obtained. These graphs show the disease dynamics within the mosquito vector population over a period of 10 years. This period is started 500 years after the beginning of the simulation to allow the system of delay-differential equations to reach its equilibrium state.

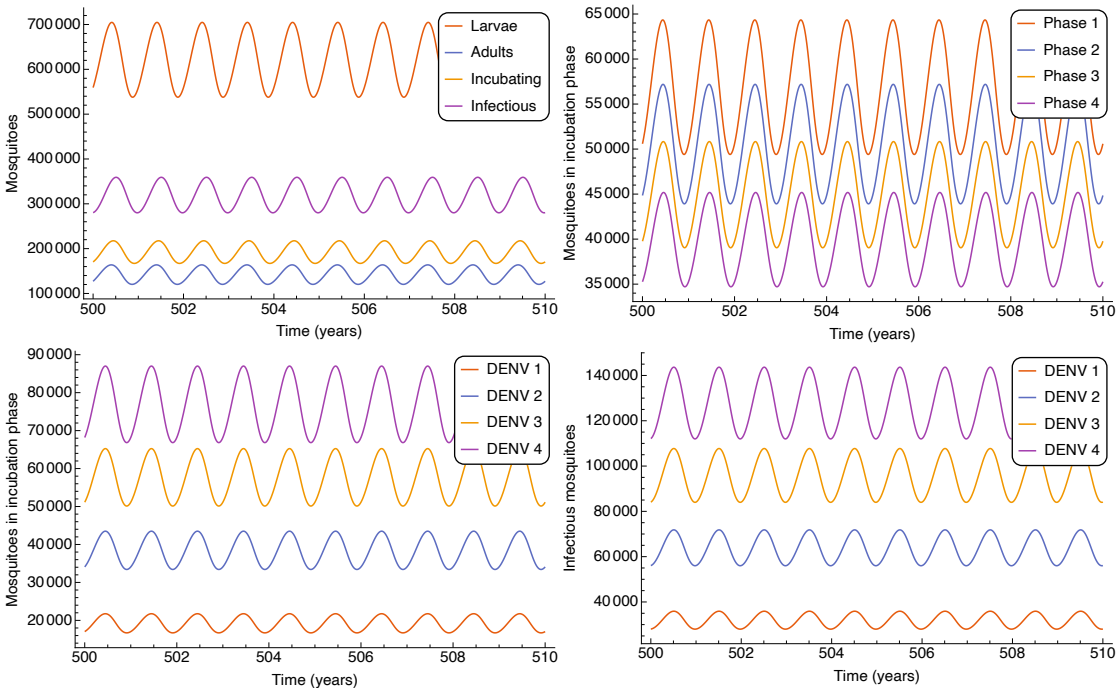


Figure 9.8: Different outputs of the reproduced component model describing the disease dynamics of DENV within the vector population. The top left panel shows the amount of mosquitoes within each of the four main compartments. The top right panel shows the amount of incubating mosquitoes within each of the different incubation phases. The bottom left panel shows the mosquitoes that are in the incubation phase of infection with each of the four DENV serotypes. Finally, the bottom right panel shows the amount of infectious mosquitoes for each of the four DENV serotypes.

After the reproduction of the component model describing the dynamics of dengue virus within the vector population, the next step was to reproduce the component model describing the dynamics of dengue virus within the host population. Due to the three different levels of stratification, this component model has a significantly higher complexity than the previously discussed vector component model, both from the conceptual point of view, as well as with regards to the implementation. Specifically, the stratification into continuous age groups, and the stratification with regards to all possible subsets θ of $\{1, 2, 3, 4\}$ representing all different possible sequential infections individuals might experience with each of the serotypes, represented a significant challenge for the reproduction.

As the Dirac delta function $\delta(x)$ and the Kronecker delta $\delta_{i,j}$ are native to Wolfram Mathematica, all terms for the partial differential equations of the host model, apart from $p_V(t)$, were already

implemented. This meant that after implementing $p_V(t)$, the equations only needed to be initialised and plugged into the existing ODE solver to model the interactions with the reproduced vector model. Unfortunately, this phase was not successfully completed as the initialization of equations across the continuous age groups was not successful. As was the case with the integration over the infectious period in humans, once again the use of continuous variables within time delayed and partial different equations proved difficult.

In this case, no solution was found and the decision was made to dedicate our efforts towards the reproduction of another suitable model. However, this reproduction provided a great amount of learning opportunities about the intricacies of more complex models for dengue virus, and some key insights into the disease dynamics were made. These were later leveraged in the model used in the final DENV MOMAB setting discussed in Section 9.3. This partial reproduction also provides a solid base for future work and continued reproduction efforts. One avenue might be to discretize the age groups and adapt the model. However, this adaptation is non-trivial as the continuous nature of the age groups is crucial for the implementation of certain parts of the model.

9.2 Reproduction of the 2009 Recker et al. model

Long-term studies on the occurrence of dengue fever and dengue haemorrhagic fever show that the rates of these diseases vary over several years and display complex cycles related to the behavior of the four dengue virus serotypes. It s widely accepted that these patterns might result from what is known as antibody-dependent enhancement (ADE). ADE occurs when a secondary infection with a different serotype of the virus leads to increased viral replication. However, some studies published around the same time as [137] have challenged this idea, suggesting that ADE alone does not explain the timing and patterns of dengue cases, pointing instead to the possibility of cross-immunity or external factors influencing these patterns. In their research, Recker et al. demonstrate that ADE, by itself, can indeed create the observed periodic fluctuations and the lack of synchronization in the activity of individual serotypes. The authors achieved this by breaking down ADE into two effects: (i) it increases susceptibility to secondary infections and (ii) it enhances the ability of the virus to spread from people with secondary infections. They show that this approach not only requires a lower level of enhancement for ADE to match real-world disease patterns but also diminishes the risk of the virus dying out due to stochastic effects. Moreover, their analysis uncovers a delayed correlation between the dynamics of the serotypes and the rates of disease, which is crucial for understanding the irregular timing of dengue outbreaks.

The ADE effect makes the development of an effective vaccine extremely difficult. Because of this, understanding and modeling the effects of ADE is of crucial importance when creating a model that will be used to evaluate the effectiveness of different vaccination strategies, as is the case in this study. The model proposed by Recker et al. [137] succeeds in simulating the effects of ADE on the dynamics of dengue fever through the introduction of two well thought-out parameters, as well as relatively limited changes to the equations describing the dynamics of the disease within the population. Because of these characteristics, the model proposed in [137] was the perfect model to reproduce (Section 9.2) and expand (Section 9.3) for the needs of this study, after the partial success of reproducing the extremely complex model proposed by Ferguson in 2016 in Section 9.1.

In their paper [137], the authors introduce the results of a dengue transmission model that utilizes two key parameters to capture the effects of antibody-dependent enhancement (ADE).

The first parameter reflects the increased viral load during secondary infections, and the second considers the heightened susceptibility to infections by different serotypes in individuals who have previously been infected, facilitated by non-neutralizing cross-reactive antibodies [137]. By explicitly incorporating these two aspects, their model successfully reproduces the observed patterns of dengue fever outbreaks and the dynamics of different dengue serotypes, without relying on external influences like seasonal changes or random stochastic events [137].

The authors of the original paper [137] developed a straightforward mathematical model, consisting of several compartments, to understand how four different dengue virus serotypes spread within a population of potential hosts, which is similar to the study done by Cummings et al. in 2005 [46]. Their model assumes that recovering from an infection with a certain dengue virus serotype means you're forever immune to that specific serotype, but it might make future infections with different serotypes worse. This assumption can be made because tertiary and fourth Dengue infections are rare [67].

Under these conditions, the authors break down the population into specific groups: s represents the share of people who haven't caught any of the serotypes and are completely susceptible; y_i is the percentage of those currently infected with serotype i for the first time; r_i indicates the fraction that have recovered from infection with serotype i; y_{ij} is the slice of the population who got a secondary infection with serotype j after already having recovered from an infection with serotype i; and lastly, r is the portion that's fully immune, having recovered from two infections with heterologous serotypes [137]. A visual representation of the model proposed in [137] can be seen in Figure 9.9.

The differential equations describing the model dynamics (from [137]) are given by:

$$\begin{aligned}
\frac{ds}{dt} &= \mu - s\sum_{k=1}^{4}\lambda_k - \mu s,\\
\frac{dy_i}{dt} &= s\lambda_i - (\sigma+\mu)y_i,\\
\frac{dr_i}{dt} &= \sigma y_i - r_i(\mu + \sum_{j\neq i}\gamma_{ij}\lambda_j),\\
\frac{dy_{ij}}{dt} &= r_i\gamma_{ij}\lambda_j - (\sigma+\mu)y_{ij}, i\neq j,\\
\frac{dr}{dt} &= \sigma\sum_{i=1}^{4}\sum_{j\neq i} y_{ij} - \mu r.
\end{aligned}$$

In their paper, the authors make a number of assumptions. The size of the population stays constant throughout the population with the average host life expectancy $1/\mu$ equal to 70 years [137]. The average duration within which a person is infectious $1/\sigma$ is assumed to be 3.65 days [137]. Furthermore, concurrent infections with different serotypes are not taken into account in the model since they are assumed to be extremely rare [137]. The force of infection of serotype i is given by the following expression from [137]:

$$\lambda_i = \beta_i\left(y_i + \sum_{j\neq i}\phi_{ji}y_{ij}\right)$$

In this expression, β_i represents the transmission coefficient for serotype i. This parameter is set to 400 per year in [137]. An important part of the contributions made by the model proposed

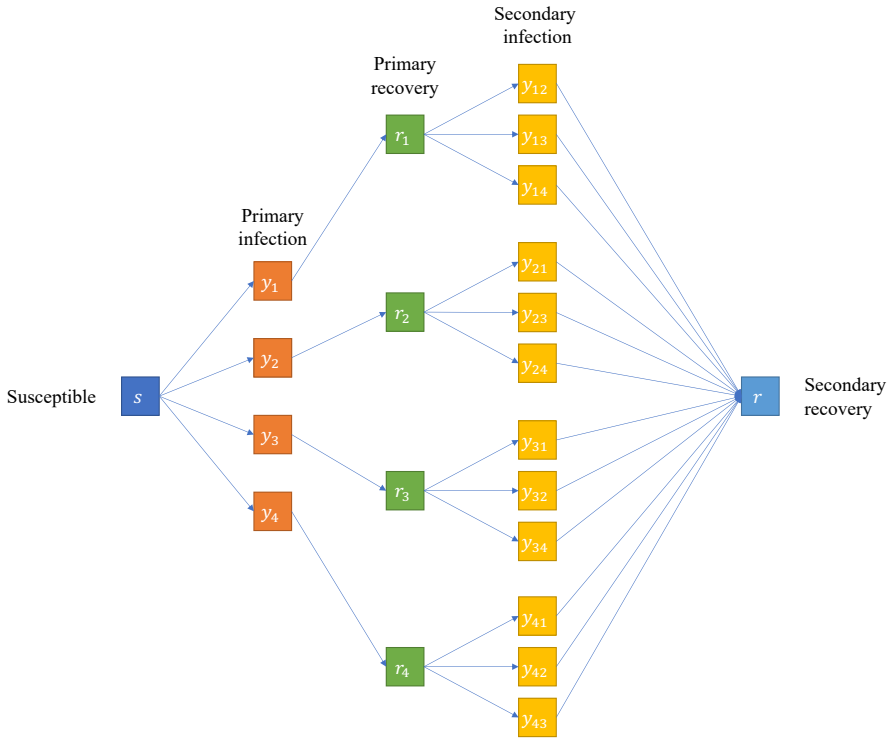


Figure 9.9: Schematic representation of the model presented by Recker et al in [137]. From left to right, an individual within the population starts as susceptible (s) and might become infected with serotype i (y_i). After recovering from the infection with serotype i (r_i), the individual might become infected with a second heterologous serotype j (y_{ij}). After having experienced two consecutive infections, the person is assumed to be fully immune and moves to the r compartment.

in [137] is the representation of ADE through the introduction of two parameters $\gamma_{ij} \geq 1$ and $\phi_{ij} \geq 1$. These parameters represent "the enhancement of susceptibility to secondary infection" and "the increase in transmissibility during secondary infection" respectively [137].

In their paper [137], the authors then continue by conducting experiments in which the ADE related γ and ϕ parameters are varied, while the rest of the parameters are kept constant. This allowed them to study the impact that ADE, and each of its components, might have on the overall transmission and infection dynamics of the four different dengue serotypes [137].

In their experiments, the authors examined and compared different increasing combinations of the proposed enhancement to susceptibility and enhanced transmission during secondary infections, leading to the reproduced results in Figure 9.10. However, the first attempts at reproducing the results obtained by the authors in their original paper, and thus verifying the correctness of the reproduced model, were unsuccessful. Only a subset of the proposed combinations of ϕ and γ yielded sound results, while others caused high degrees of numerical instability in the ODE solver, resulting in results that were not viable for interpretation. Upon contacting the authors of the paper, it became clear that they had failed to mention a so called *importation rate* that causes an additional flux of individuals from the s compartment to the different y_i compartments, i.e. additional primary infections. This adjustment enhances the numerical

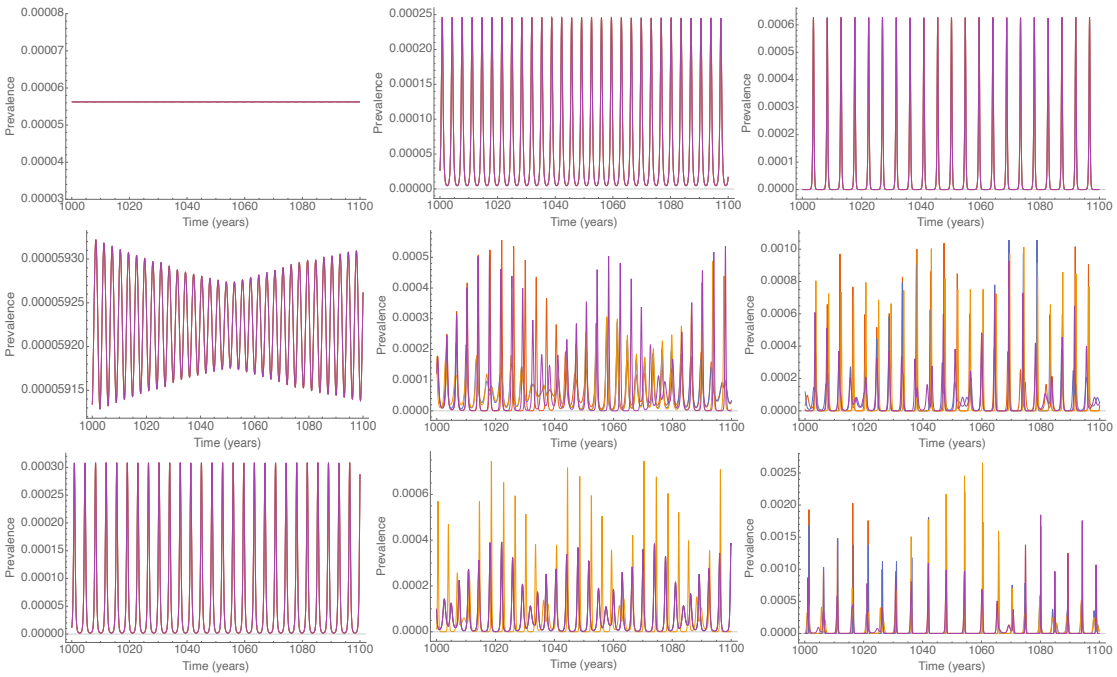


Figure 9.10: Output of the reproduced model. The different graphs show the evolution of the proportion of the population that is infected with each of the four different Dengue virus serotypes. Each color in the figures represents a different serotype. The degree of ADE effects increases throughout the figures as the enhancement in transmissibility ϕ increases from left to right and the enhancement to susceptibility γ varies from top to bottom. These parameters take on the values in $(1, 1.9, 2.4)$. Other parameter values are $\beta = 400$, $\sigma = 100$, and $\mu = 1/70$. Reproduction of Figure 2 from [137].

stability of the simulation. Upon further examination of the importation rate, we hypothesized that it might serve as an additional seeding mechanism necessary for simulating disease dynamics over multiple millennia. After further correspondence with the authors, we concluded that the persistence of dengue during the off-season or in regions not typically considered conducive to dengue transmission remains an open research question. Adding this importation rate χ_i to the differential equations describing the disease dynamics, yields the following adapted equations for the number of susceptible individuals and individuals experiencing a primary infection:

$$\frac{ds}{dt} = \mu - s\left(\sum_{k=1}^{4}\lambda_k + \mu + \sum_{k=1}^{4}\chi_k\right),$$
$$\frac{dy_i}{dt} = s\left(\lambda_i + \chi_i\right) - (\sigma + \mu)y_i.$$

Within the context of this reproduction, these modified equations were implemented in Python and in the Wolfram Mathematica language.

The graphs that can be seen in Figure 9.10 are not identical to those presented in Figure 2 of the original paper [137]. However, without knowing the initial conditions the authors used to solve the system of ODEs, and without knowing the exact value they used for the importation rate χ_i

for each serotype $i \in 1..4$, it is very difficult to obtain graphs that are identical to those presented in the original study. For the initial conditions of this reproduction, it was assumed that 0.01% of the population was infected with each of the different serotypes $i \in [1..4]$. The remaining 99.96% of the population starts out as susceptible. At the start of the simulation, there are no primary recoveries, secondary infections, nor individuals who have gained full immunity. Furthermore, a value of 1×10^{-6} was used for the importation rate χ.

Simulating the dynamics of Dengue fever with these initial conditions and parametrizations over a period of 1100 years, a number of interesting observations can be made from the results shown in Figure 9.10. It can be seen that increasing the enhancement in transmissibilty ϕ as a result of ADE causes the prevalence of the disease to fluctuate over time, introducing epidemic periods. The same is true when increasing the enhanced susceptibility as a result of ADE γ. Unlike the results presented in the original paper, desynchronization between the different serotypes was only achieved by increasing both enhancement parameters, leading to the belief that, in order to obtain realistic results from the model proposed by Recker et al. [137], both ϕ and γ need to be assigned values strictly greater than one.

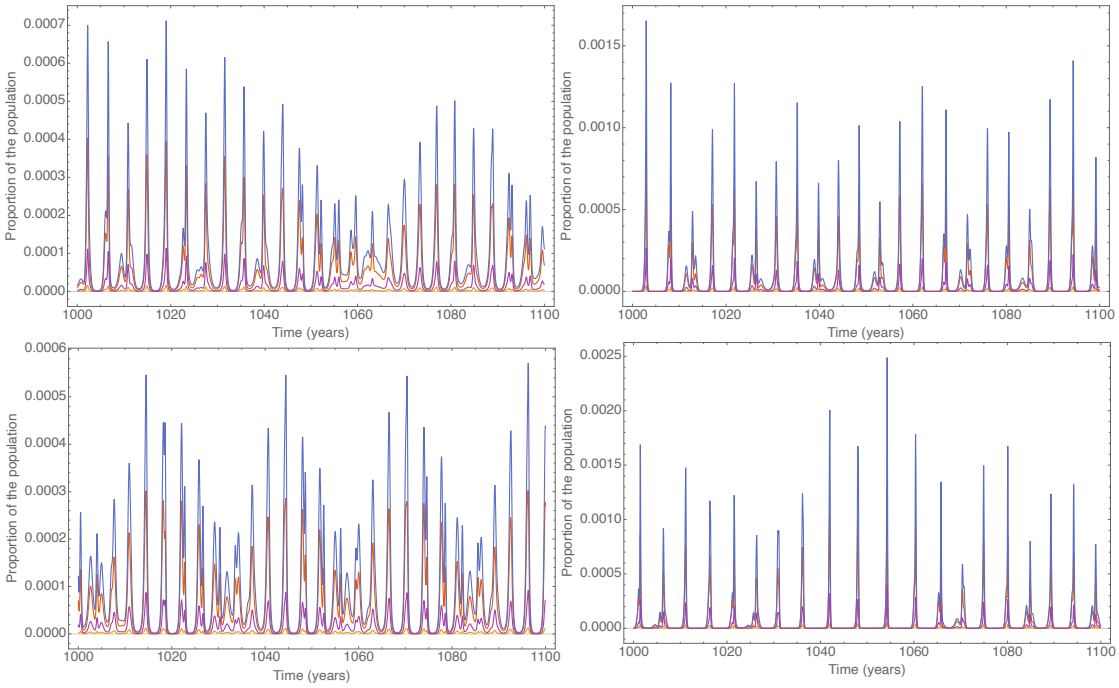


Figure 9.11: Evolution of the proportion of the population that is suffering from symptomatic disease as a result of a primary infection (orange), suffering from symptomatic disease as a result of a secondary infection (blue), is hospitalized as a result of a primary infection (yellow), and is hospitalized as a result of a secondary infection (purple). The values for the ADE parameters ϕ and γ are varied between 1.9 and 2.4, increasing from left to right and top to bottom respectively.

Given the evolution of primary infections y_i and secondary infections y_{ij} over time, the proportions for cases that result in symptomatic disease and hospitalizations used in the model proposed by Ferguson et al. [61], can be used to examine the evolution of symptomatic disease and hospitalizations over the integration period. In the work presented by Ferguson et al. in

2016 [61], it is assumed that 45% of the primary infections will result in symptomatic disease and 4% will result in a hospitalization. When considering secondary infections, due to the ADE effect, these proportions rise to 80% of cases resulting in symptomatic disease and 16% resulting in a hospitalization. Multiplying the proportion of the population that experiences a primary infection y with these proposed proportions, yields the proportion of the population that experiences symptomatic disease and need to be hospitalized upon a first infection. Identically, using y_{ij} and the proportions for secondary infections, the proportion of the population that experiences symptomatic disease and need to be hospitalized upon a secondary infection can be visualized. This evolution can be observed in Figure 9.11. These values can be used as a metric for the medical burden endured by the population as a result of Dengue fever. Therefore, one of the key objectives of the algorithm comparing different vaccination strategies π will be to quickly and efficiently identify the strategies that minimize this medical burden.

9.3 Expansion of the 2009 Recker et al. model

Because of the explicit modeling of the effects of ADE, and the results that are comparable to those of real-world Dengue epidemics, the Dengue epidemic model proposed by Recker et al. in 2009 [137] is a solid starting point to evaluate the efficacy of different vaccination strategies. However, unlike many other studies, it does not include any compartments related to vaccination. Furthermore, in real-world scenarios, vaccination programs often discriminate between different age groups. These point require a dual expansion of the original model with (i) support for vaccination of the population, and (ii) support for different age groups in which the population can be stratified. In this section, the modifications and expansions that were made to facilitate these changes will be discussed, starting with explicit support for vaccination programs.

9.3.1 Support for vaccination

One of the key goals of this study is to identify the set of optimal vaccination strategies among a larger set of possible vaccination strategies. In order to achieve this goal, the model describing the dynamics of Dengue fever within a population of potential hosts proposed in [137] needed to be adapted to allow for the explicit representation of vaccination strategies.

Within the existing framework of compartment models to model infectious disease spread within a population, this is achieved through the addition of compartments that are representative of the partition of the population that has been vaccinated. The implementation used in this study can be seen visualized in Figure 9.12.

In the extended model that is presented here, there are two additional compartments: v_- and v_+. The v_- compartment encapsulates the part of the host population that has been vaccinated against Dengue fever at a time where they were seronegative. This means that up to that point in time, they have never contracted an infection with one of the four heterologous DENV serotypes. In this implementation, this is modeled as a flux of individuals moving from the compartment containing susceptible individuals s, to the compartment containing vaccinated seronegative individuals v_-. The rate ν_1 at which parts of the susceptible population become vaccinated and move to the v_- compartment, is determined by the particular vaccination strategy π that is simulated by the model.

The second compartment that was introduced in this extension of the original model, is the compartment v_+ containing potential hosts that were vaccinated when that had already endured an infection with one of the different DENV serotypes. In the adapted model that is presented

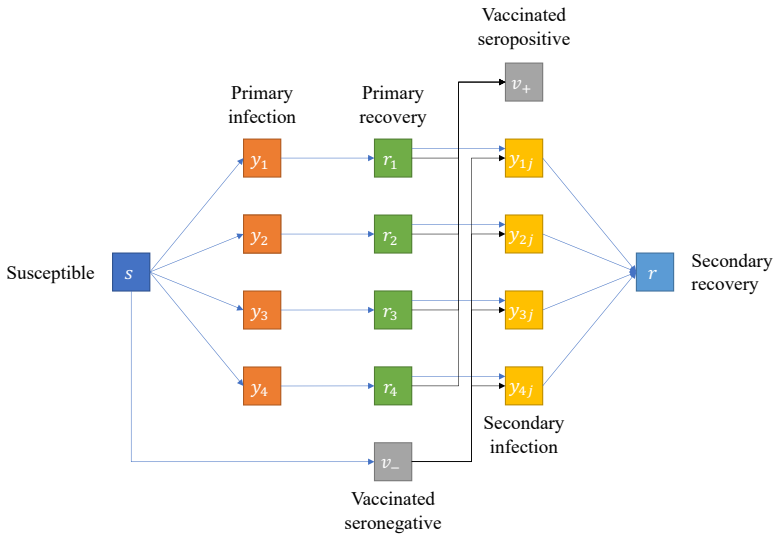


Figure 9.12: Schematic representation of the model proposed in [137], expanded with two addition compartments for individuals who are vaccinated when they are seronegative v_-, and individuals who are vaccinated when they are seropositive v_+. Compartments y_{ij} representing secondary infections have been collapsed.

here, this is modeled as a flow of individuals moving from the different r_i compartments to the newly introduced v_+ compartment. Since the different r_i compartments contain the partition of the population that have recovered from an infection with serotype $i \in [1..4]$, this achieves exactly the intended purpose and definition of the v_+ compartment. In an identical manner to the rate ν_1 at which seronegative individuals are vaccinated, the rate ν_2 at which seropositive individuals are vaccinated is also determined by the vaccination strategy that is being simulated by the model.

Naturally, the introduction of additional compartments and their interaction with the existing compartment necessitates updates to the existing system of differential equations describing the model. The modified system of ordinary differential equations that describes the dynamics of this extended model is given by Equation (8.1) to Equation (8.7). A differential equation describing the incoming and outgoing fluxes of individuals has been added for each of the new compartments, increasing the number of equations in the system from the original 26 to 28. Changes have also been made to some of the original equations. These include incorporation of the different vaccination rates ν_1 and ν_2 into the equations for the number of susceptible individuals s, and the number of people who have recovered from a primary infection r_i with

one of the four serotypes $i \in [1..4]$.

$$\frac{ds}{dt} = \mu - s\left(\nu_1 + \sum_{k=1}^{4}\lambda_k + \sum_{k=1}^{4}\chi_k + \mu\right), \tag{9.1}$$

$$\frac{dy_i}{dt} = s\left(\lambda_i + \chi_i\right) - y_i(\sigma + \mu), \tag{9.2}$$

$$\frac{dr_i}{dt} = \sigma y_i - r_i(\mu + \frac{\nu_2}{4} + \sum_{j\neq i}\gamma_{ij}\lambda_j), \tag{9.3}$$

$$\frac{dy_{ij}}{dt} = r_i\gamma_{ij}\lambda_j + \frac{v_-\lambda_j}{3} - y_{ij}(\sigma + \mu), i \neq j, \tag{9.4}$$

$$\frac{dr}{dt} = \sigma\sum_{i=1}^{4}\sum_{j\neq i} y_{ij} - \mu r, \tag{9.5}$$

$$\frac{dv_-}{dt} = s\nu_1 - v_-\left(\mu + \sum_{k=1}^{4}\lambda_k\right), \tag{9.6}$$

$$\frac{dv_+}{dt} = \frac{\nu_2}{4}\sum_{k=1}^{4} r_k - v_+\mu. \tag{9.7}$$

Since vaccination is modeled as a silent infection, like previously explored in [61] and Section 9.1, and in this modeled version of reality individuals can be infected with Dengue virus a total of two times before gaining full immunity, seronegative individuals who experience a silent infection through vaccination can experience a secondary infection. This process is visualized by the blue arrow from the v_- compartment to the different y_{ij} compartments in Figure 9.12 and required inclusion of a term $\frac{v_-\lambda_j}{3}$ in the equations for the different y_{ij} compartments. Noticing the collapsed nature of the y_{ij} compartments in Figure 9.12, the division by 3 is needed to ensure an equal outgoing rate from v_- and incoming rate into y_{ij}.

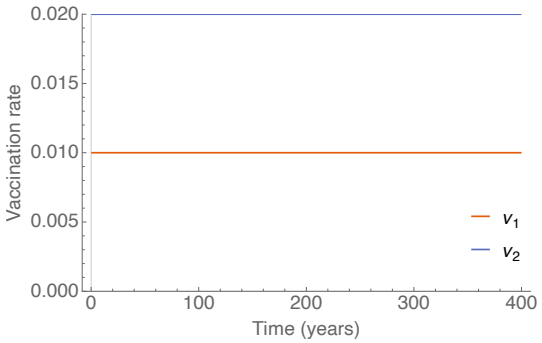
It is at this point that one of the key difficulties with vaccination against dengue virus arises. The ADE effect causes individuals experiencing a secondary infection with a heterologous serotype to have a significantly increased likelihood of developing symptomatic disease and requiring hospitalization. This causes an increase in transmission since they will carry an increased viral load, increasing the spread of the disease in the population, and also induces a very large strain on the public health system. This negative effect also arises when the first infection an individual encounters is a silent one experienced due to vaccination. This means that the vaccination of seronegative individuals in some cases might give rise to more individuals suffering from the consequences of ADE and developing severe, in some cases even life-threatening, symptoms. This process is clarified further in Figure 9.1 reproduced from [61] that was previously explored in Section 9.1.

Individuals who have not been vaccinated (as shown in the top row of Figure 9.1) typically endure a moderate initial infection. Subsequent infections increase in severity, with the second being more severe, but the third and fourth infections tend to be milder. Seronegative individuals who are vaccinated while still fully susceptible to dengue (illustrated in the middle row) initially receive transient protection against all four dengue serotypes, similar to the natural immunity observed after the first infection [155, 139]. However, as antibody levels decline, this protection wanes and may even facilitate the virus, increasing the likelihood of symptomatic and severe disease upon a primary breakthrough infection [79, 35]. Therefore, our model posits that for

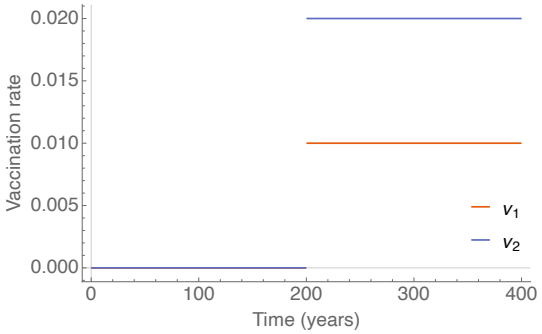
these vaccinated, seronegative individuals, a primary breakthrough infection is likely to lead to symptomatic or severe disease at a rate comparable to a secondary infection in an unvaccinated individual.

In contrast, individuals who have been vaccinated after one or more dengue infections (depicted in the bottom row of Figure 9.1) achieve immunity levels akin to those who have had multiple infections. As a result, any subsequent infection post-vaccination behaves like a tertiary infection in unvaccinated individuals, characterized by a lower probability of resulting in symptomatic or severe disease.

Because of these complications with the vaccination against Dengue virus, policy makers might want to chose between vaccination of either seronegative individuals or seropositive individuals, or a mixture of both. It is for this exact reason that our proposed expansion incorporates distinct compartments for both vaccinated seropositive individuals v_+ and individuals who were vaccinated while seronegative v_-, and the two rates ν_1 and ν_2 at which either seronegative or seropositive individuals are vaccinated. This allows for a plethora of different vaccination strategies $\pi = (\nu_1, \nu_2)$ that might be considered by policy makers.



(a) Evolution of vaccination rates ν_i over time when vaccination is introduced immediately at the start of the simulation.



(b) Evolution of vaccination rates ν_i over time when vaccination is introduced a certain number of years $\tau = 200$ after the start of the simulation.

Figure 9.13: Visualization of immediate introduction of vaccination after the start of the simulation (left panel) and introduction of vaccination after a predetermined number of years τ (right panel).

One more factor that needs to be taken into account when expanding the model proposed in [137] with explicit support for vaccination, is the fact that in real life situations vaccination will not be available as soon as the disease emerges (Figure 9.13a). This can be incorporated into the model proposed in this study by making both vaccination rates ν_1 and ν_2 functions of time t, both remaining at zero until a specified time τ at which the vaccination program starts (Figure 9.13b). This yields the following expressions:

$$\nu_1(t) = \begin{cases} 0 & : t < \tau \\ \nu_1 & : t \geq \tau \end{cases}$$

$$\nu_2(t) = \begin{cases} 0 & : t < \tau \\ \nu_2 & : t \geq \tau \end{cases}$$

Another element that requires careful consideration when evaluating different vaccination strategies π, is their associated cost κ. This is needed because in real-life situations the policy makers

will always have to operate within the framework of the available financial resources. The cost for the vaccination of a single individual is denoted κ_1. Taking into account the ADE effect that negatively affects those who are vaccinated when they have not yet experienced a natural infection, it might seem like a good idea to vaccinate all seropositive individuals. However, following this vaccination strategy requires that all individuals be tested for antibodies against one of the four circulating DENV serotypes, inducing an additional testing cost κ_2 for each individual that is vaccinated. The total cost for the vaccination of a seropositive individual then becomes equal to the sum of the cost of the vaccine and the cost of the test: $\kappa = \kappa_1 + \kappa_2$, while for a seronegative individual the total cost of vaccination $\kappa = \kappa_1$.

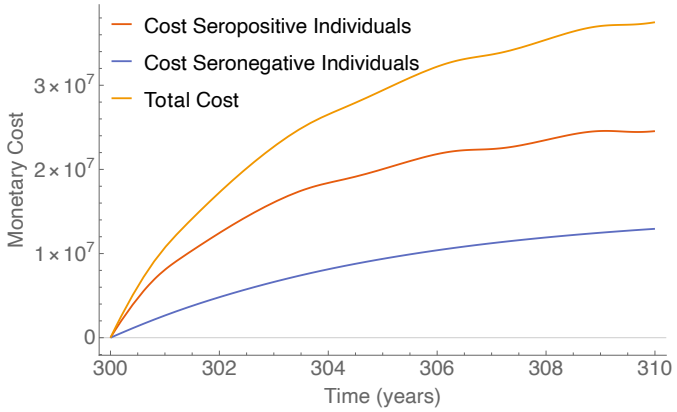


Figure 9.14: The cost of vaccination of seronegative individuals, seropositive individuals, and the total cost of a specific vaccination strategy as a function of time. Vaccination starts 300 years after the beginning of the simulation. In this figure $\nu_1(t) = \nu_2(t) = 0.5$, meaning every year 50% of the eligible population is vaccinated, $\kappa_1 = 10$, and $\kappa_2 = 20$

This cost κ is an interesting secondary objective in addition to the previously discussed medical burden of the disease that needs to be minimized. Considering two vaccination strategies π_1 and π_2 that have the same decrease in medical burden of the disease on the population, π_1 can be considered the preferable, more optimal strategy if the cost associated with this strategy is lower than the cost associated with the other strategy π_2.

9.3.2 Support for age-heterogeneity

In real-world epidemiological scenarios, it is almost always necessary to consider individuals according to their age group. This is due to the fact that the disease and prevention measures might function differently in individuals belonging to different age groups and this needs to be taken into consideration when developing and evaluating preventive strategies. The model proposed by Recker et al. in [137] provides clear insights into the effects of the ADE on the population level and is able to fit real-world data from regions in which dengue is endemic [137]. In the previous section, this model was successfully extended with explicit support for simulating vaccination strategies. In this section, the expansions to the model that allow support for age-heterogeneity will be discussed.

To integrate age-dependent mixing into the proposed model with support for vaccination from the previous section, the population is divided into 2 distinct age categories, children and adults, establishing a separate model for each group. These age-specific models are subsequently inter-

connected to simulate the age-dependent mixing among the various age cohorts. The approach taken for this extension is very similar to the age-heterogeneous SIR models discussed in Section 4.1.3. A schematic representation of the proposed extension to two heterogeneous age groups can be seen in Figure 9.15.

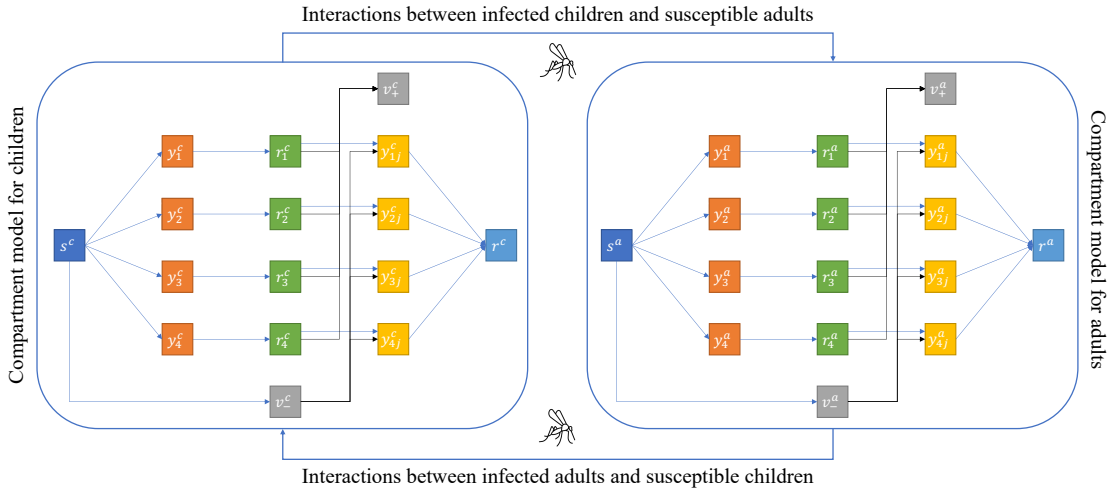


Figure 9.15: Schematic representation of the proposed extension of the model from [137], extended with support for vaccination, to support age-heterogeneity.

To accomplish support for different age groups, the differential equations governing the disease dynamics within the population need to be parametrized with a parameter g representing the age group. This essentially allows for the creation of identical sets of compartments $s^g, y_i^g, r_i^g, y_{ij}^g$, and r^g for each age group g that requires representation within the model. The modified system of ordinary differential equations then becomes:

$$
\begin{aligned}
\frac{ds^g}{dt} &= \mu - s^g \left(\nu_1^g + \sum_{k=1}^{4} \lambda_k^g + \sum_{k=1}^{4} \chi_k + \mu \right), \\
\frac{dy_i^g}{dt} &= s^g \left(\lambda_i^g + \chi_i \right) - y_i^g (\sigma + \mu), \\
\frac{dr_i^g}{dt} &= \sigma y_i^g - r_i^g (\mu + \frac{\nu_2^g}{4} + \sum_{j \neq i} \gamma_{ij} \lambda_j^g), \\
\frac{dy_{ij}^g}{dt} &= r_i \gamma_{ij} \lambda_j^g + \frac{v_-^g - \lambda_j^g}{3} - y_{ij}^g (\sigma + \mu), \quad i \neq j, \\
\frac{dr^g}{dt} &= \sigma \sum_{i=1}^{4} \sum_{j \neq i} y_{ij}^g - \mu r^g, \\
\frac{dv_-^g}{dt} &= s^g \nu_1^g - v_-^g \left(\mu + \sum_{k=1}^{4} \lambda_k^g \right), \\
\frac{dv_+^g}{dt} &= \frac{\nu_2^g}{4} \sum_{k=1}^{4} r_k^g - v_+^g \mu.
\end{aligned}
$$

Figure 9.15 shows such compartments for two age groups: children c on the left, and adults a on the right.

Apart from parametrization based on the age groups, allowing for the creation of a model per age group, the interactions between the different age groups also need to be incorporated into the model. Infectious individuals from one age group might be stung by a mosquito which might be infected from this interaction with an infectious individual. Through another bite, this infectious vector might then infect an individual from a different age group. These are essentially vector-driven interactions of infectious pressure between the different age groups, represented in Figure 9.15 by the blue arrows between the two age groups' models. In the original model proposed by Recker et al. in [137], these indirect interactions between susceptible and infectious individuals are encapsulated within the serotype-specific force of infection λ_i which was equal to:

$$\lambda_i = \beta_i \left(y_i + \sum_{j \neq i} \phi_{ji} y_{ij} \right)$$

Analogously to the approach taken for the age-heterogeneous SIR models in Section 4.1.3, this expression can be made age-specific to model the interactions between individuals of different age groups by introducing a contact matrix C. This matrix has a row and a column for each age group and can be used to weigh the interactions in λ_i^g according to the amount of contacts occurring between individuals of different age groups. Incorporating the contact matrix C gives the following new expression for λ_i^g:

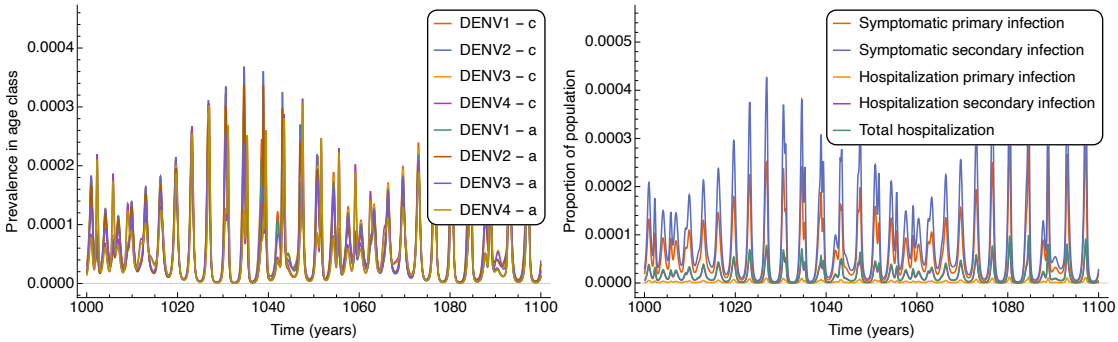
$$\lambda_i^g = \beta_i \left(\sum_{k=0}^{n} \frac{y_i^k \times C_{gk}}{|N_k|} + \sum_{j \neq i} \left(\phi_{ji} \sum_{k=0}^{n} \frac{y_{ij}^k \times C_{gk}}{|N_k|} \right) \right)$$

In this force of infection on age group g due to serotype i, $|N_k|$ represents the size of age group k and C_{gk} is the entry in the contact matrix C describing the number of contacts between susceptible individuals of age group g and infectious individuals of age group k. Note that such contact matrix based approaches are most suited to modeling diseases in which the virus is directly spread from individual to individual without the need for a vector intermediary. However, after careful consideration it was decided that such a model is also sound for simulating dengue epidemics depending on the values chosen for C. The reasoning behind this is that wherever individuals meet, they can infect others via the mosquito vectors that are present in the area. The values in C essentially weigh the importance of contacts between individuals in the force of infection. Hence, the decision was made to use values for the entries of C where the contacts between age groups are of relatively low importance but still play a role in the overall disease dynamics.

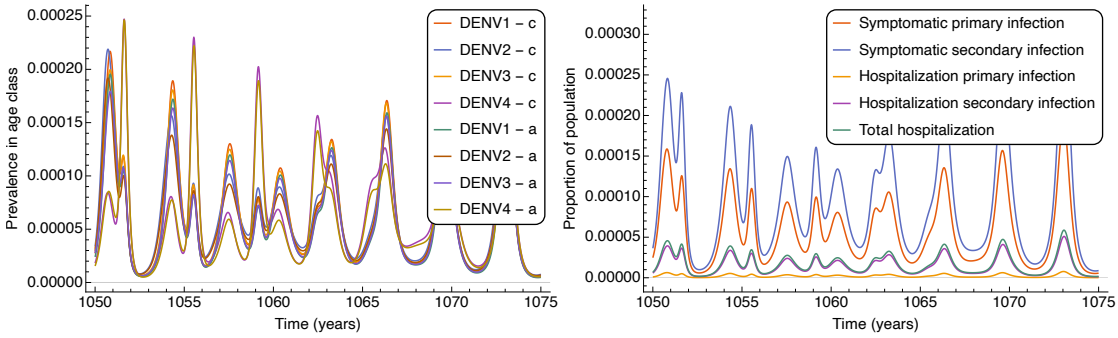
The extended model proposed in this dissertation uses two age groups: children c and adults a. As visualized in Figure 9.15, each age group has their own compartment model with support for vaccination introduced in the previous section. Each of these component models is governed by a system of 28 differential equations, resulting in a total of 56 ODEs for the final model presented here. Actually simulating the dynamics of DENV within the age-heterogeneous population requires initial conditions for each of these differential equations. Within each age group, a single seeding case was used as initial infection with each of the four serotypes:

$$y_i^g(0) = 1, \quad \forall i \in [1..4], \quad \forall g \in \{c, a\}.$$

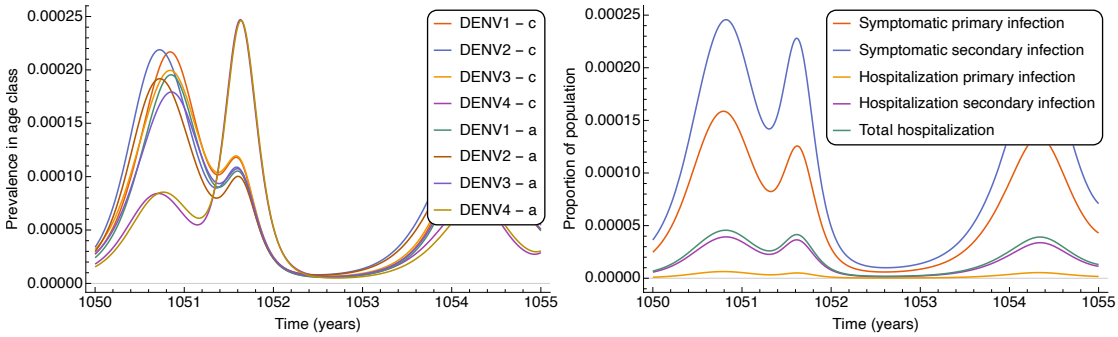
All other individuals of the host population start as susceptible. Initially, there are no individuals who have recovered from a primary infection, are experiencing a secondary infection, have recovered from a secondary infection and are fully immune, or have been vaccinated.



(a) Simulated disease dynamics within the two age classes over a period of 100 years.



(b) Simulated disease dynamics within the two age classes over a period of 25 years.



(c) Simulated disease dynamics within the two age classes over a period of 5 years.

Figure 9.16: Output of proposed model that extends the model presented in [137] with support for age-heterogeneity. The model presented here has two age classes: children c and adults a. The left panels show the prevalence of the different DENV serotypes within each age class. The right panels show the proportion of the population experiencing symptomatic disease or requiring hospitalization.

Figure 9.16 shows the output of the proposed extended model incorporating age-heterogeneity over periods of 100, 25 and 5 years. The left panels show the disease dynamics within each age class by visualizing the prevalence of each of the four DENV serotypes within the c and a classes. The right panels give an overview of the resulting dynamics on the level of the entire population, presenting the proportion of the population suffering from symptomatic disease or requiring hospitalization. In the next section the combination of both presented extensions will be discussed: simulating vaccination strategies over the age-based stratified population.

9.4 Simulating vaccination strategies

The stratification of the population into different age groups allows for the representation of age-specific rates of vaccination in addition to the serostatus-specific vaccination introduced in Section 9.3.1. Extending and generalising the expressions for the vaccination rates $\nu_1(t)$ and $\nu_2(t)$ from Section 9.3.1 yields:

$$\nu_s^g(t) = \begin{cases} 0 & : t < \tau \\ \nu_s^g & : t \geq \tau \end{cases} \qquad 0 \leq \nu_s^g \leq 1$$

as the rate of vaccination over time t for individuals with serostatus s belonging to age group g where vaccination starts τ years after the beginning of the simulation. Applying this general expression to the specific extended model proposed in this dissertation distinguishing between seropositive and seronegative individuals, as well as distinguishing between children and adult individuals, yields the four vaccination rates in Table 9.1.

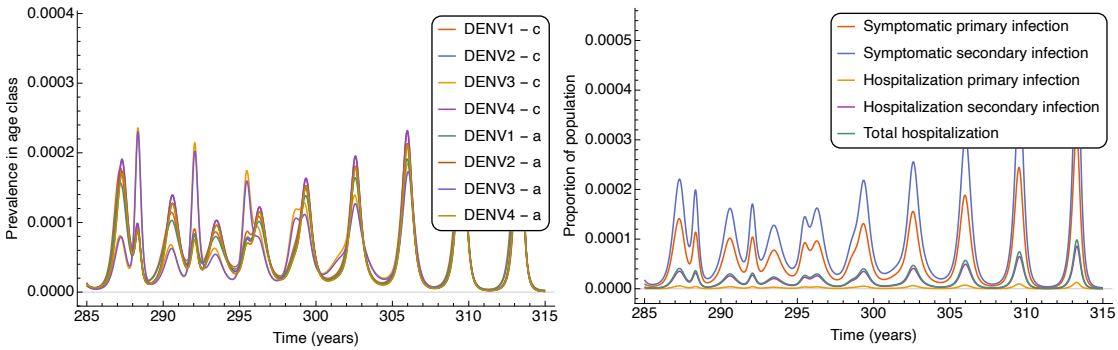
Table 9.1: Different vaccination rates ν_s^g based on the stratification of the population based on serostatus s and age class g.

ν_s^g	**Children**	**Adults**
Seronegative	ν_-^c	ν_-^a
Seropositive	ν_+^c	ν_+^a

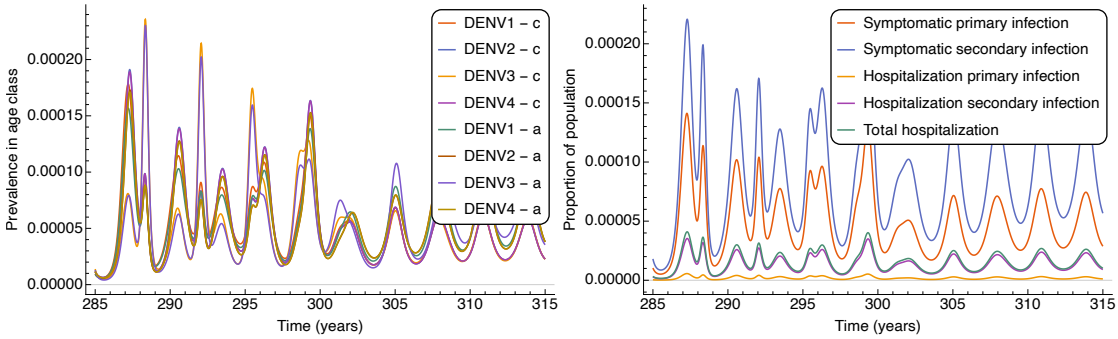
Therefore, within the framework of the proposed extended model, a vaccination program or policy π can be represented as a tuple $(\nu_-^c, \nu_-^a, \nu_+^c, \nu_+^a)$. These values can be assigned according to the vaccination strategy that is under evaluation. They are integrated into the model through the mechanisms discussed in Section 9.3.1, regulating the flow of individuals to the vaccination-specific compartments.

Figure 9.17 shows the output of the model without vaccination (Figure 9.17a) and when simulating 3 different vaccination strategies (Figures 9.17b to 9.17d). In the first simulated vaccination program, every year 50% of eligible children are vaccinated, regardless of their serostatus. In the second simulated vaccination program, every year 50% of eligible adults are vaccinated, regardless of their serostatus. Finally, in the third simulated vaccination program, every year 50% of the entire eligible population is vaccinated, regardless of their serostatus or age class. Comparing the prevalence of the different serotypes within each age class before and after the introduction of vaccination, it can be seen that vaccination results in lower prevalence. When looking at the population level dynamics of symptomatic disease and hospitalizations, the introduction of vaccination also clearly has a positive effect.

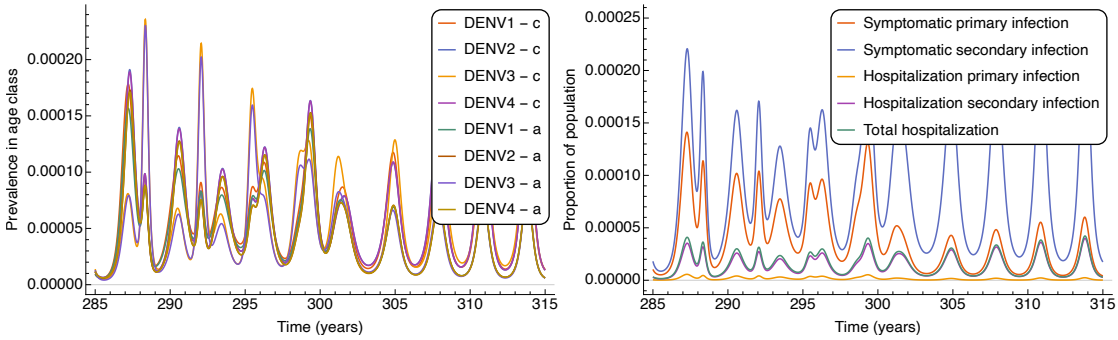
The vaccination strategies shown in Figure 9.17 are in no way intended to be optimal. They only serve as an indicator of the soundness of the proposed extended model. However, in Chapter 11,



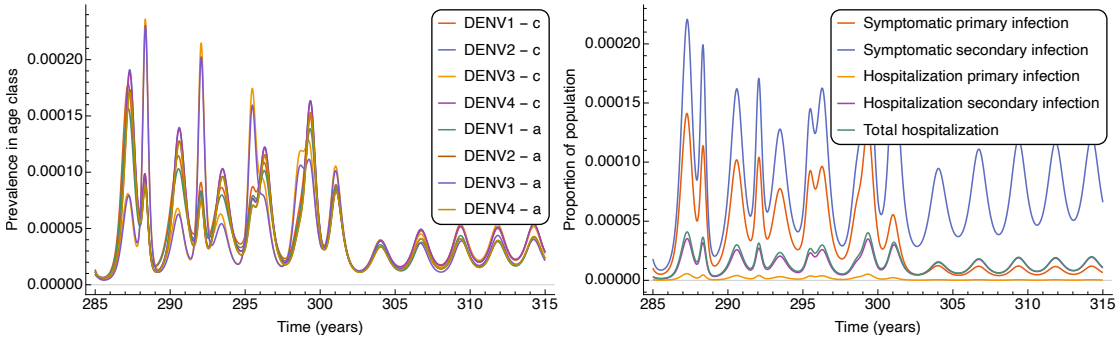
(a) Simulated disease dynamics without vaccination.



(b) Simulated disease dynamics when vaccinating 50% of all eligible children each year.



(c) Simulated disease dynamics when vaccinating 50% of all eligible adults each year.



(d) Simulated disease dynamics when vaccinating 50% of the eligible population each year.

Figure 9.17: Output of the proposed model with support for vaccination and age-heterogeneity when simulating different vaccination strategies π. Vaccination is introduced 300 years after the beginning of the simulation. $\pi_a = (0, 0, 0, 0)$, $\pi_b = (0.5, 0, 0.5, 0)$, $\pi_c = (0, 0.5, 0, 0.5)$, $\pi_d = (0.5, 0.5, 0.5, 0.5)$.

the extended model will be used to identify the subset of optimal vaccination strategies $\boldsymbol{\pi}^*$ in combination with the MOMAB algorithms that are discussed in the next chapter of this dissertation.

Chapter 10

Multi-objective multi-armed bandits

One of the main goals of this dissertation is to identify the subset of optimal vaccination strategies $\boldsymbol{\pi}^*$ within a larger set of considered vaccination strategies. The vaccination strategies are evaluated within an epidemiological model for dengue virus. As disease transmission dynamics are inherently stochastic, when developing such models, the decision can be made to incorporate stochasticity into the model to mimic disease transmission dynamics in a more realistic manner. Hence, the evaluation of a vaccination strategy in such simulators can be seen as pulling an arm in a multi-objective multi-armed bandit (MOMAB) setting, where the stochastic output of the model corresponds to the rewards obtained for pulling that specific arm. Identifying the optimal vaccination strategies then corresponds to the multi-objective extension of the best-arm identification problem for MABs: Pareto front identification (PFI).

In this chapter, three metrics will first be introduced to evaluate the quality of the recommendations made by the PFI MOMAB algorithms. Secondly, the reproduction of a number of MOMAB algorithms for regret minimization from literature and their adaptation to the PFI setting will be discussed. Finally, a completely novel PFI MOMAB algorithm is proposed.

10.1 Performance metrics for MOMAB Pareto front identification

In this section, three metrics for the performance of Pareto front identification (PFI) multi-objective multi-armed bandit (MOMAB) algorithms will be discussed: the Bernoulli metric, the Jaccard similarity metric, and the hypervolume metric. The goal of these metrics is to evaluate the quality of the recommendations made by a PFI MOMAB algorithm with respect to the actual set of Pareto optimal arms $\mathcal{A}^*$. Within the context of these algorithms, a recommendation can be formalised as a subset $\mathcal{R}_t \subseteq \mathcal{A}$ of the set of all arms of the MOMAB, consisting of the arms $a_r \in \mathcal{R}_t$ that are considered to be Pareto optimal by the algorithm at time t.

10.1.1 Bernoulli metric

The first way to evaluate the quality of a recommendation $\mathcal{R}_t$ with respect to the actual set of Pareto optimal arms $\mathcal{A}^*$ is to consider the recommendations as a *Bernoulli* trial after each arm pull t. The trial is considered a success if the recommendation $\mathcal{R}_t$ is exactly equal to the actual set of Pareto optimal arms $\mathcal{A}^*$. Otherwise, it is considered a failure. A score of 1 is associated with a success, and a score of zero is associated with a failure.

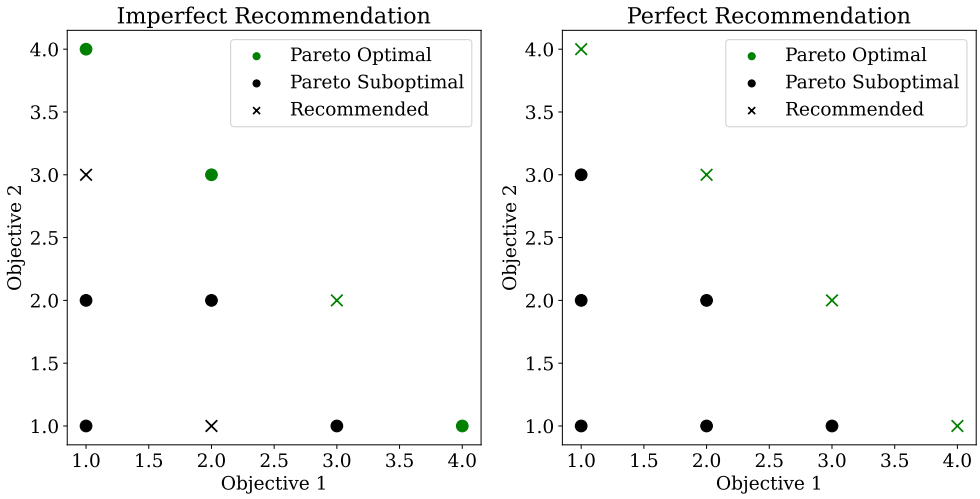


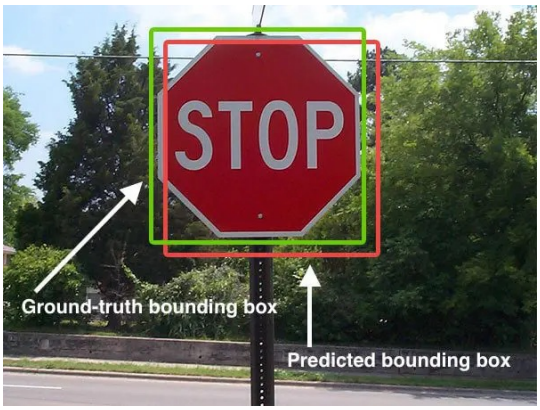
Figure 10.1: Visualization of a perfect and an imperfect recommendation. Points corresponding to Pareto-optimal arms are plotted in green, recommended arms are plotted as crosses.

These scores for successes and failures can be averaged over multiple runs of the algorithm to obtain the averaged empirical success rate of the algorithm at time t. This yields a proportion that gives a good indication of the algorithm's ability to give perfect recommendations.

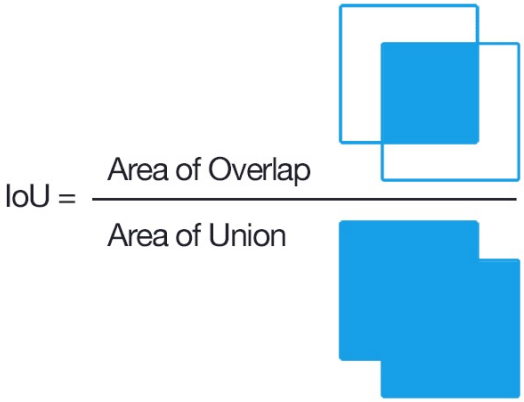
The Bernoulli metric presented here is a very strict metric: only perfect recommendations are recompensed. This means that if the recommendation at time t contains all but one of the Pareto optimal arms, the Bernoulli metric will be equal to 0. Similarly, if the recommendation contains all Pareto optimal arms and some additional suboptimal arm, the Bernoulli metric will also be equal to 0. From an intuitive point of view, recommendations that are far from perfect receive the same score as recommendations that are nearly perfect. Consequently, the Bernoulli metric does not provide that much insight into the acquisition process of the Pareto front by the PFI algorithm. However, we argue that the Bernoulli metric is a strong indicator of the algorithms ability to identify the Pareto optimal arms, that is particularly useful when the goal is to perfectly identify the Pareto front.

10.1.2 Jaccard similarity metric

Another way to evaluate the quality of the recommendations $\mathcal{R}_t$ made by the PFI MOMAB algorithms, is by using a metric inspired by the *Jaccard similarity* coefficient. This coefficient was developed independently by Grove Karl Gilbert [129], Paul Jaccard [89], and T. T. Tanimoto [168] under different names. Within the Computer Science community, it is most often referred to as the Jaccard similarity.



(a) Example of a ground-truth and predicted bounding box for a stop sign. From [148].



(b) Visualization of the intersection over union calculation that is used to calculate the Jaccard similarity of two sets. From [148].

Figure 10.2: The Jaccard similarity applied to a computer vision task. To evaluate the quality of he predicted bounding boxes, the Jaccard similarity of the corresponding sets is calculated. Images from [148].

Figure 10.2 shows an example use case of the Jaccard similarity where it is applied within a computer vision task [148]. The quality of the predicted bounding box can be evaluated with respect to the ground truth bounding box by representing both boxes as a set and calculating the Jaccard similarity.

As the Jaccard similarity is used as a measure for the similarity between two finite sets, it is an ideal candidate to evaluate the quality of the recommendations $\mathcal{R}_t$ made by the PFI algorithm with respect to the true Pareto front $\mathcal{A}^*$. It is computed as the size of the intersection of the two sets, divided by the size of the union of the two sets:

$$J(\mathcal{R}_t, \mathcal{A}^*) = \frac{|\mathcal{R}_t \cap \mathcal{A}^*|}{|\mathcal{R}_t \cup \mathcal{A}^*|}.$$

By design, the value of the Jaccard similarity metric will be $0 \leq J(\mathcal{R}_t, \mathcal{A}^*) \leq 1$ where greater values indicate a higher degree of similarity between the two sets. Once again, this metric can be calculated after each recommendation and averaged over multiple runs of the algorithm. Unlike the Bernoulli metric, this metric provides insights into the evolution of the recommendations from sets of arms that might be very dissimilar to the Pareto front, to increasingly similar sets, reaching a value of 1 for perfect recommendations.

10.1.3 Hypervolume metric

The *hypervolume* metric is one of the most prevalent measures in the literature for evaluating the performance of coverage sets [118, 173, 179, 203, 205]. This metric assesses the quality of the set of recommended arms $\mathcal{R}_t$ by calculating its volume with respect to a predetermined reference point p [140]. Conceptually, it represents the union of boxes defined by the reference point and all arms a_r within $\mathcal{R}_t$:

$$H(\mathcal{R}_t) = \Lambda\left(\bigcup_{a_r \in \mathcal{R}_t} [p, a_r]\right)$$

where Λ denotes the Lebesgue measure (also known as the n-dimensional volume) and $[p, a_r] = \{q \in \mathbb{R}^n \mid q \geq p \wedge q \leq a_r\}$ is the box bounded below by the reference point p and above by the recommended arm a_r [140].

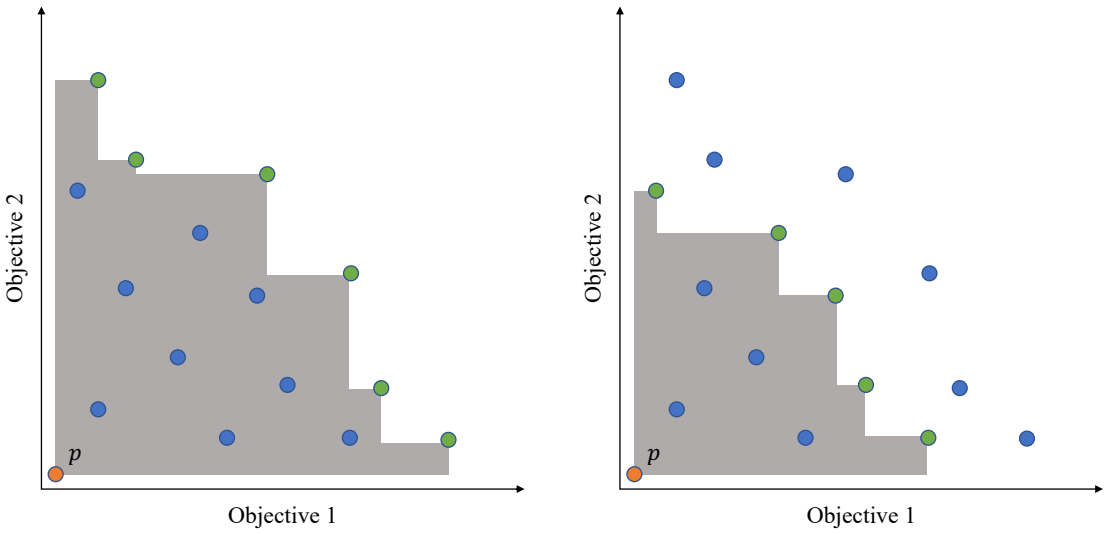


Figure 10.3: Visualization of the hypervolume metric in the MOMAB bandit setting where larger values for the objectives are associated with better arms. Recommended arms are plotted in green, the reference point p is plotted in orange. The perfect recommendation in the left panel has a larger hypervolume than the imperfect recommendation in the right panel.

The reference point is chosen as a lower bound on achievable returns to ensure that the volumes are always positive [140]. Consequently, the hypervolume encompasses all possible values that are dominated by the coverage set, with more dominating coverage sets yielding a larger hypervolume. By definition, the hypervolume is maximized for the Pareto front, as no other possible solution can increase its volume due to their dominance [140]. Although the hypervolume metric is extensively used and provides a measure of a solution set's coverage, its interpretation can be challenging [140]. The implications of changes in hypervolume are not immediately evident to end users and do not necessarily correlate with significant changes in expected utility. In high-dimensional objective spaces, the addition or removal of a single point can result in substantial variations in hypervolume, particularly if the point is extreme [140].

10.2 Reproductions and adaptations to the Pareto front identification setting

In this section the reproductions of a number of MOMAB algorithms for regret minimization from literature, and their adaptation to the PFI setting will be discussed.

In the literature review on MOMAB algorithms in Chapter 8, the lack of literature on pure bandit algorithms for preference-unaware Pareto front identification, was identified. The PFI setting is essentially the multi-objective extension of the best-arm identification setting for single-objective MABs discussed in Section 5.2. Within the context of this dissertation, the preference-unaware PFI setting is of particular interest since it corresponds exactly to the goal of providing the

decision maker with the complete set of possible trade-offs between different optimal vaccination strategies, without making any assumptions about their preferences or attempting to learn them.

In Section 8.2, four different variants of preference-unaware MOMAB algorithms were presented. All algorithms presented in Section 8.2 are algorithms for regret minimization, all aiming to minimize their cumulative and unfairness regrets while navigating the exploration-exploitation trade-off in their own unique way: UCB, Thompson sampling, Knowledge gradient, and an Annealing approach.

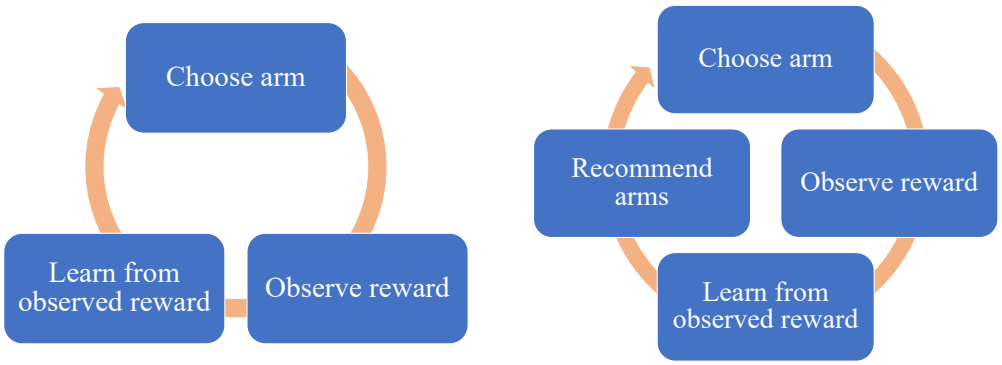


Figure 10.4: Comparison between the algorithmic loop for the regret minimization setting and the Pareto front identification setting.

The first main goal of this dissertation does not align completely with the regret minimization setting that these algorithms were designed for. Instead, it is an instance of the PFI setting for MOMABs. The difference between these two setting is illustrated in Figure 10.4. However, the preference-unaware nature of the MOMAB algorithms presented in Section 8.2 does align with the needs of this study.

Furthermore, when considering single-objective MAB algorithms, it is regularly the case that best-arm identification algorithms and regret minimization algorithms are closely related to one another, where the former has an additional emphasis on exploration. Some examples are the UCB algorithm with a larger value for the exploration-regulating hyperparameter κ, and the close relation between Thompson sampling and Top-two Thompson Sampling. These facts, combined with the lack of pure MOMAB algorithms for Pareto front identification[1], identified in Chapter 8 and confirmed by Reymond in [140], makes the algorithms presented in Section 8.2 very valuable to examine within the context of this dissertation: They might be suitable for adaptation to the MOMAB Pareto front identification setting.

Within this context, each of the algorithms presented in Section 8.2 has been reproduced in Python based on the original publications in which they were proposed. Their implementation was verified by examining their cumulative Pareto regrets and cumulative unfairness regrets. More information about the conducted experiments, and detailed descriptions of the mechanisms used in these MOMAB algorithms can be found in Section 8.2.

The Python implementation of the reproduced MOMAB algorithms is available upon request.

[1]It is important to note that Bayesian Optimization algorithms exist which can solve the kind of MOMAB PFI setting presented in this study. Some of these algorithms leverage MABs for their acquisition function.

Each algorithm is implemented as a class, all of them supporting an identical interface that corresponds exactly to the algorithmic loops shown in Figure 10.4. Figure 10.4 also shows that MOMAB algorithms for regret minimization can be extended to the PFI setting by adding a method for recommending arms to their interface. Figure 10.5 shows a complete overview of the MOMAB algorithms that were reproduced and extended to the PFI setting.

Algorithm	Variant	Version	Reproduced	Extended to PFI
UCB	Pareto		✅	✅
	Scalarized	Linear	✅	
		Chebyshev	✅	
Thomson sampling	Pareto		✅	✅
	Scalarized	Linear	✅	
		Chebyshev	❌	
Knowledge gradient	Pareto		✅	✅
	Scalarized across arms	Linear	✅	
		Chebyshev	❌	
	Scalarized across objectives	Linear	✅	
		Chebyshev	❌	
Annealing	Pareto		✅	✅

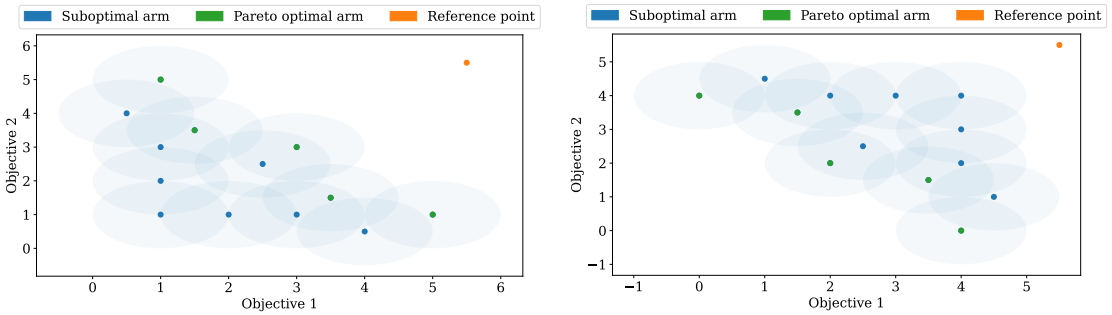
Figure 10.5: Overview of the MOMAB algorithms that were reproduced and extended to the PFI setting.

As can be seen in Figure 10.5, only the Pareto variants of each of the algorithms were extended to the PFI setting. By design, the scalarized variants of the MOMAB algorithms do not explore the multi-objective space directly. Instead, they reduce the multi-objective space to single scalar value for each arm based on a scalarization function. From this scalar value for each of the arms, a relative order between the arms can always be established, where a single arm is considered to be the optimal arm. Hence, the scalarized variants of the MOMAB algorithms will always recommend a single arm when extended to the PFI setting. Within the context of this dissertation, this corresponds to the algorithm only identifying a single vaccination strategy which is optimal under the function used for scalarization. This clearly goes against the goal of presenting the decision makers with all possible trade-offs. Furthermore, when using such scalarized algorithms, the scalarization function(s) employed by the algorithm need to be defined a priori, thus necessitating assumptions to be made about the possible preferences of the decision maker. In this dissertation, the goal is to avoid making assumptions about the utility of the decision maker. Due to these reasons, the scalarized variants of the MOMAB algorithms were not suitable for extension to the PFI setting within the context of this study.

The algorithms that were extended to the PFI setting were evaluated with respect to the three metrics proposed in Section 10.1. This was achieved through an experiment using the arms

visualized in Figure 10.6. A multivariate Gaussian distribution is used for each of the arm's reward distribution. The standard deviation for each of the dimensions of the reward distribution was set to 1. This results in overlapping areas between the arms' reward distributions, making it more challenging for the PFI MOMAB algorithms to distinguish which arms are the optimal ones.

Figure 10.6 also shows the reference point used for the calculation of the hypervolume metric. Note that this point is associated with rewards for each of the objectives that are larger than the mean rewards of the optimal arms. This is exactly the inverse case of what is shown in Section 10.1.3 and fig. 10.3, and is due to the implementation of the hypervolume metric in the *pymoo*[2] library. To ensure that a perfect recommendation of the optimal arms by the algorithm is associated with the largest hypervolume of all possible recommendations, an inverted setup is used for the calculation of the hypervolume metric. This inverted setup is shown in Figure 10.6b.



(a) Arms within the PFI experiment.

(b) Inverted arms within the PFI experiment.

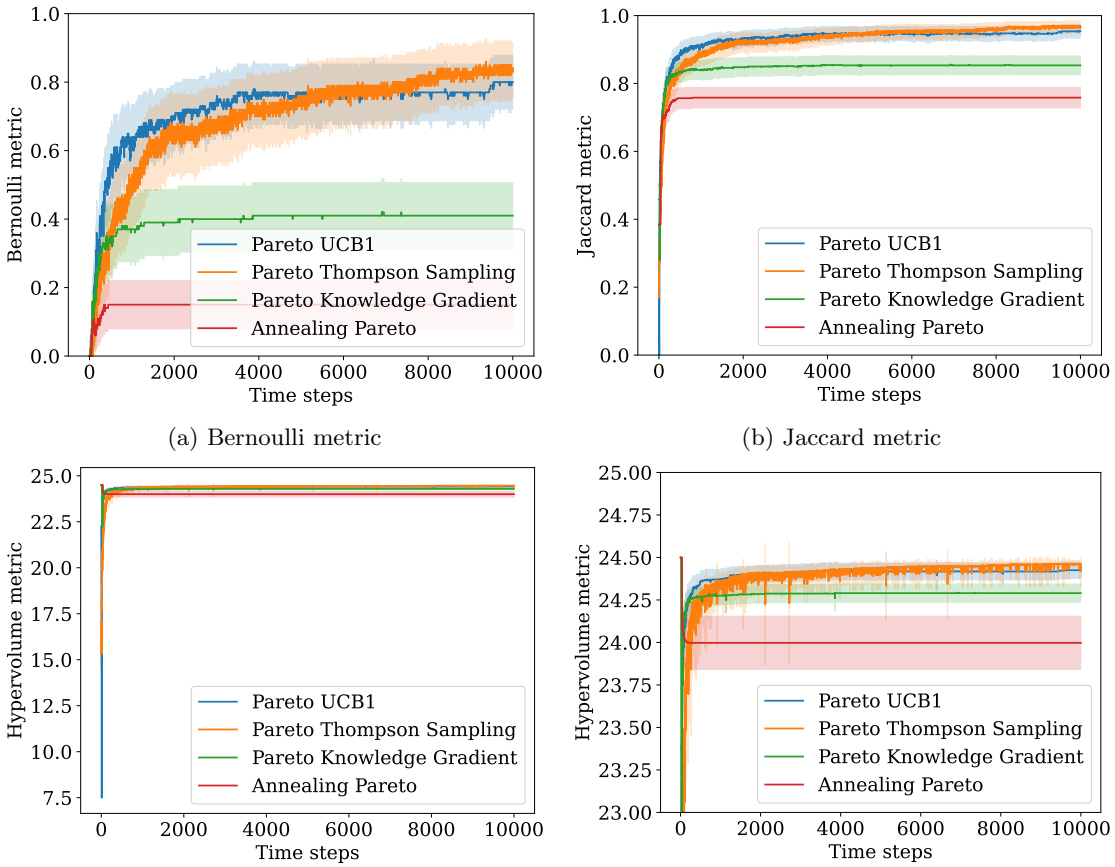
Figure 10.6: Overview of the arms within experimental setup for the Pareto-front identification experiment. There are 5 optimal arms plotted in green and 8 suboptimal arms plotted in blue, resulting in a total of 13 arms. The standard deviations for the arms' reward distributions are all equal to 1 and are plotted as a shaded area around the means.

The results of the PFI experiment are shown in Figure 10.7. The results show the average performance of the algorithms over 100 runs, each with a budget of 10,000 arm pulls. For the Pareto Thompson Sampling algorithm, a multivariate normal-inverse-gamma distribution is used as a prior for the arms' multivariate Gaussian reward distributions as the normal-inverse-gamma distribution is the conjugate prior for a normal distribution with unknown mean and variance.

The Bernoulli metric results, illustrated in Figure 10.7a, indicate that the algorithms' recommendations gradually correspond exactly to the true Pareto front more frequently. However, none of the algorithms consistently achieved perfect recommendations throughout the experiment. The Pareto Thompson Sampling and Pareto UCB1 algorithms exhibited higher success rates compared to the Pareto Knowledge Gradient and Annealing Pareto algorithms. Initially, the Pareto UCB1 algorithm performs best with respect to the Bernoulli metric. However, its average empirical success rate stagnates at around 0.8, while the Pareto Thompson Sampling algorithm keeps improving its rate of perfect recommendations. The results across all algorithms suggest that achieving an exact match with the Pareto front is challenging under the given experimental conditions.

Figure 10.7b presents the performance of the algorithms according to the Jaccard similarity

[2] `https://pymoo.org`



(a) Bernoulli metric

(b) Jaccard metric

(c) Hypervolume metric

(d) Hypervolume metric - magnified

Figure 10.7: Results of the extended MOMAB algorithms in PFI experiment with respect to the Bernoulli, Jaccard, and hypervolume metrics. Results are averaged over 100 runs of the algorithm with a budget of 10,000 arm pulls. The shaded area shows the 95% confidence interval around the mean.

metric. All algorithms demonstrated a steady increase in Jaccard similarity over time, with the Pareto Thompson Sampling and Pareto UCB1 algorithms again showing superior performance. The Pareto Thompson Sampling algorithm achieved the highest Jaccard similarity, closely followed by the Pareto UCB1 algorithm, indicating that these algorithms were more effective in approximating the true Pareto front as the number of arm pulls increased. The Jaccard similarity metric of the Pareto Thompson Sampling and Pareto UCB1 algorithms is very close to 1, indicating that these algorithms were extremely close to identifying the entire Pareto front.

The hypervolume metric results, shown in Figure 10.7c, further support the findings from the other metrics. The hypervolume indicates that the Pareto Thompson Sampling and Pareto UCB1 algorithms covered a larger volume compared to the Pareto Knowledge Gradient and Annealing Pareto algorithms. This suggests that the former algorithms not only identified sets that were closer to the true Pareto front but also provided a broader coverage of the objective space. The hypervolume metric values for Pareto Thompson Sampling and Pareto UCB1 approached the

maximum achievable values, demonstrating their effectiveness in the PFI setting.

The experimental results indicate that the Pareto Thompson Sampling and Pareto UCB1 algorithms are most effective for the PFI setting, with Pareto Thompson Sampling performing slightly better. The higher performance in terms of Jaccard similarity and hypervolume metrics suggests that these algorithms are better at approximating and covering the true Pareto front than the other algorithms. The strict nature of the Bernoulli metric, however, underscores the difficulty of achieving perfect recommendations.

These findings highlight the potential of adapting existing MOMAB algorithms to the PFI setting. Future work could explore further adaptations and enhancements to these algorithms to improve their performance. Additionally, investigating the impact of different experimental setups, such as varying the standard deviation of the reward distributions could provide deeper insights into the robustness and applicability of these algorithms in real-world scenarios.

10.3 Top-two Pareto Fronts Thompson Sampling

In the previous section, it was established that it is often the case that best-arm identification algorithms and algorithms for regret minimization are closely related to one another, where the former has an additional emphasis on exploration. Notable examples are the UCB algorithm with a larger value for the exploration-regulating hyperparameter κ, and the close relation between Thompson sampling and Top-two Thompson Sampling.

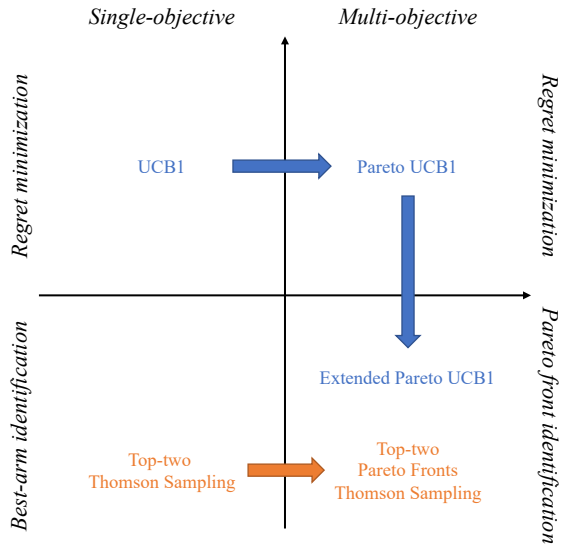


Figure 10.8: Relationships between the single-objective and multi-objective MABs (horizontal axis) for the regret minimization and BAI/PFI settings (vertical axis). UCB1 was first adapted for the multi-objective regret minimization setting in [51] and then further extended in this dissertation for the PFI setting. Best-arm identification algorithms like Top-two Thompson Sampling can be adapted to reach the PFI setting directly.

In the literature review on preference-unaware MOMAB algorithms in Section 8.2, it was observed that the UCB and Thompson Sampling strategies had already been extended to the multi-objective regret minimization setting. This allowed for the adaptation of these existing algorithms

to the PFI setting, inspired by their single-objective counterparts. These extensions to the PFI setting were discussed in Section 10.2. Based on the observations about the relationships between these different types of algorithms in Figure 10.8, it became clear that it also might be possible to directly extend existing best-arm identification MAB algorithms to the PFI setting.

Algorithm 16: Top-two Pareto Fronts Thompson Sampling

Input: A MOMAB with K arms, a probability ρ, a prior $\pi(\cdot)$ and history $\mathcal{H}^{(0)} = \emptyset$, arm pull budget T

for $t \leftarrow 1$ **to** T **do**
- **for** $a \leftarrow 1$ **to** K **do**
  - $\tilde{\boldsymbol{\mu}}_a^{(t)} \sim \pi(\cdot|\mathcal{H}^{(t-1)})$
- Compute the Pareto optimal arms $\mathcal{A}^*$ such that $\forall i \in \mathcal{A}^*$ and $\forall j \notin \mathcal{A}^*$, $\tilde{\boldsymbol{\mu}}_j^{(t)} \not\succ \tilde{\boldsymbol{\mu}}_i^{(t)}$
- $b \sim \mathcal{B}(\rho)$
- **if** $b = 1$ **then**
  - Select arm $a^{(t)}$ uniformly at random from $\mathcal{A}^*$
- **else**
  - Compute $\mathcal{A}' = \mathcal{A} - \mathcal{A}^*$
  - Compute the Pareto optimal arms $\mathcal{A}'^* \subseteq \mathcal{A}'$ such that $\forall i \in \mathcal{A}'^*$ and $\forall j \notin A'^*$, $\boldsymbol{\mu}_j^t \not\succ \boldsymbol{\mu}_i^t$
  - Select arm $a^{(t)}$ uniformly at random from $\mathcal{A}'^*$
- Play $a^{(t)}$ and observe $\mathbf{r}^{(t)}$
- $\mathcal{H}^{(t)} \leftarrow \mathcal{H}^{(t-1)} \cup \{a^{(t)}, \mathbf{r}^{(t)}\}$
- Compute and recommend the arms currently considered optimal based on the means of the posteriors

After further investigating the literature on MOMABs, a multi-objective extension of Top-two Thompson Sampling proposed by Reymond was identified in [140]. The multi-objective Top-two Thompson Sampling (MOTTTS) algorithm proposed in [140] learns a multivariate belief distributions over the arms to account for the multiple objectives. One of the crucial aspects of the single-objective Top-two Thompson Sampling algorithm is the ranking operator that is used to provide a relative ordering between the arms. In the single-objective case, this relative ordering can always be established. Hence, the algorithm can differentiate between the top-two arms when choosing which arm to pull next. In the multi-objective case however, if the utility of the user is unknown or unconsidered, at each time there will be multiple Pareto optimal arms. This is explained in detail in Section 6.1. As a consequence, in the preference-unaware multi-objective case, it is impossible for the algorithm to determine the top-two arms when choosing which arm to pull next. The implementation of MOTTTS proposed in [140] solves this problem by also imposing a belief distribution over the utility of the decision maker. As the algorithm progresses, the user is queried about their preferences and thus, the algorithm learns the utility of the user. Based on this learned utility, the algorithm is able to establish a relative order between the arms and determine the top two best performing arms with respect to the user's utility. This is an elegant solution to the challenges imposed by the multi-objective setting, but due to its preference-incorporating nature, MOTTTS as presented in [140] was not suitable for use within this study.

Although MOTTTS was not applicable to the goals of this dissertation, the preference-incorporating solution proposed by [140] prompted consideration of how Top-two Thompson Sampling might

be extended in a preference-unaware context. Specifically, it raised questions regarding how the problem of ranking arms could be addressed without incorporating user preferences. After thorough contemplation on how to address this problem, ultimately we devised the *Top-two Pareto Fronts Thompson Sampling* (TTPFTS) algorithm as illustrated in Algorithm 15.

When following the TTPFTS strategy to decide which arm to pull, the initial step taken by the algorithm is identical to that of Pareto Thompson Sampling: it samples from the multivariate posterior distributions it has built for each of the arms $a \in \mathcal{A}$. This first step is illustrated in panel (a) of Figure 10.9. With a probability of ρ, the algorithm proceeds further like Pareto Thompson Sampling and computes the set of optimal arms $\mathcal{A}^*$ based on the samples that were just obtained, as shown in panel (b) of Figure 10.9. From this set of optimal arms, an arm $a^{(t)}$ is selected uniformly at random to be pulled. With a probability of $1 - \rho$, the set of optimal arms $\mathcal{A}^*$ is computed and these optimal samples are removed, resulting in $\mathcal{A}' = \mathcal{A} - \mathcal{A}^*$. This new set is depicted in panel (c) of Figure 10.9.

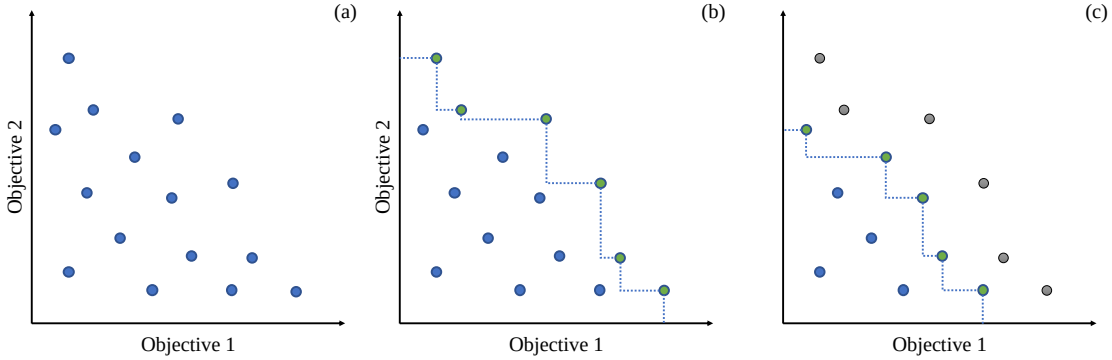


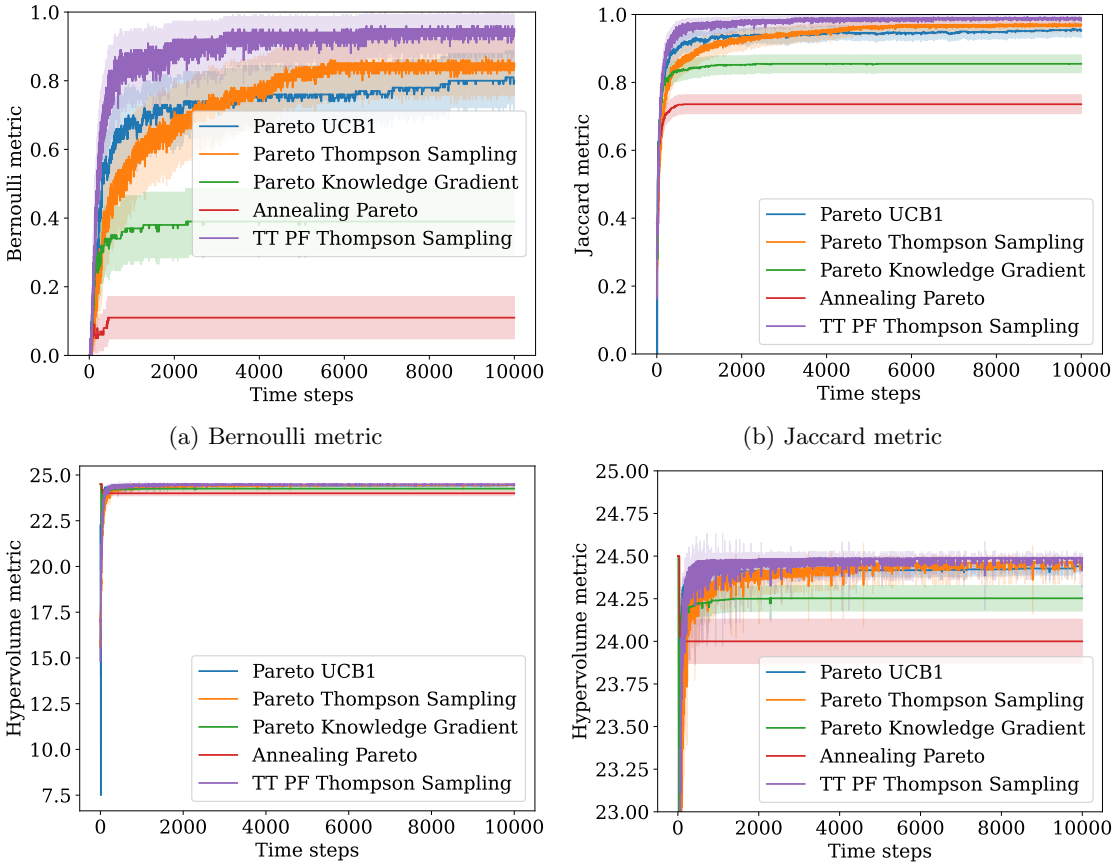
Figure 10.9: Visualization of the Top-two Pareto Fronts strategy. Panel (a) shows the samples obtained by the algorithm by sampling from the posterior distributions of each of the arms. With probability of ρ, the optimal arms are computed and one of them is chosen at random (b). With a probability of $1 - \rho$, the optimal arms are removed and one of the optimal arms of the remaining set is chosen (c).

Using the samples in this reduced set $\mathcal{A}'$, the set of optimal arms based on the reduced set $\mathcal{A}'^*$ can be calculated. This is essentially the subset of optimal arms within the subset of suboptimal arms. In other words, it is the set of the most optimal suboptimal arms. Finally, in this case one of the arms $a^{(t)}$ from this set $\mathcal{A}'^*$ is chosen at random to be pulled. Then, the reward vector $\mathbf{r}^{(t)}$ associated with pulling arm $a^{(t)}$ is observed and the posterior distribution for arm $a^{(t)}$ is updated using this newly observed data. Finally, after the algorithm has learned from the newly observed reward, the arms that are currently considered optimal by the algorithm are computed and recommended to the user.

Within the context of this study, TTPFTS as presented in Algorithm 15 was implemented in Python using the same unified MOMAB API that was developed when reproducing the preference-unaware MOMAB algorithms in Sections 8.2 and 10.2. The implementation of the algorithm is available upon request. Currently, there is support for using either a Beta distribution or Normal-inverse-gamma distribution as a prior. This means that in its current state, TTPFTS supports settings where the arms' reward distributions are either multivariate Bernoulli distributions or multivariate normal distributions with unknown mean and variance.

To evaluate the performance of the novel TTPFTS algorithm, it was compared to the algorithms

that were extended to the PFI setting in Section 10.2. In this comparison, the algorithms were evaluated with respect to the three metrics proposed in Section 10.1. This was achieved through an experiment using the arms visualized in Figure 10.6. A multivariate Gaussian distribution is used for each of the arm's reward distribution. The standard deviation for each of the dimensions of the reward distribution was set to 1. This results in overlapping areas between the arms' reward distributions, making it more challenging for the PFI MOMAB algorithms to distinguish which arms are the optimal ones. Each of the algorithms was given a budget of 10,000 arm pulls, and the experiment was repeated 100 times to achieve the average performance shown in Figure 10.10.



(a) Bernoulli metric

(b) Jaccard metric

(c) Hypervolume metric

(d) Hypervolume metric - magnified

Figure 10.10: Results of the extended MOMAB algorithms including TTPFTS in the PFI experiment with respect to the Bernoulli, Jaccard, and hypervolume metrics. Results are averaged over 100 runs of the algorithm with a budget of 10,000 arm pulls. The shaded area shows the 95% confidence interval around the mean.

This exact experiment was already used to evaluated the MOMAB algorithms extended to the PFI setting in Section 10.2. There, the results without TTPFTS indicated that while no algorithm consistently achieved perfect recommendations, all algorithms improved over time according to Bernoulli, Jaccard similarity, and hypervolume metrics. Specifically, the Pareto Thompson Sampling and Pareto UCB1 algorithms demonstrated superior performance across all metrics, with Pareto Thompson Sampling generally slightly outperforming Pareto UCB1. These algo-

rithms consistently approached or achieved high values in Jaccard similarity and hypervolume metrics, indicating their effectiveness in approximating and covering the true Pareto front in the objective space. However, the Bernoulli metric highlighted the challenge of achieving perfect recommendations under the experimental conditions. These results are confirmed by this second repetition of the experiment and the results shown in Figure 10.10.

When introducing the results obtained for the TTPFTS algorithm, it can be observed that it significantly outperforms other algorithms within this specific experimental setup. Focusing on the Bernoulli performance metric shown in Figure 10.10a, the TTPFTS algorithm's performance increases rapidly, consistently providing perfect recommendations of the true set of Pareto optimal arms more than 90% of the time after only 3000 arm pulls. This represents a substantial improvement over the 80% success rate achieved by the best-performing alternative algorithm, Pareto Thompson Sampling, after 5000 arm pulls. Furthermore, the TTPFTS algorithm also surpasses the other tested algorithms concerning the Jaccard similarity metric and the hypervolume metric. These results suggest that in this setting, the TTPFTS algorithm is more effective and efficient at identifying the Pareto front than the other algorithms.

However, it is important to note that these findings are based on a single experimental setting. To make more robust claims about the TTPFTS algorithm's performance, additional experiments should be conducted with varying numbers of arms, objectives, relative distances between arms, and different reward distributions and variabilities for the arms. Nonetheless, these results are highly encouraging for the potential performance of the novel TTPFTS algorithm proposed in this dissertation.

Chapter 11

Dengue virus MOMAB setting

Apart from the meticulous study and extension of the MOMAB framework in Chapter 10, another primary research question of this dissertation was whether multi-objective multi-armed bandit algorithms could be used to identify optimal vaccination strategies for the mitigation of dengue epidemics, in a more sample-efficient manner than the currently employed Uniform Sampling algorithm. More formally, given a set $\boldsymbol{\pi}$ of possible vaccination strategies π_i that are evaluated with respect to different objectives in a model for dengue epidemics, can the MOMAB efficiently identify the Pareto front, consisting of the subset $\boldsymbol{\pi}^*$ of Pareto optimal vaccination strategies. This set of Pareto optimal vaccination strategies and the trade-offs between them, can then be provided to the decision maker. In the first section of this chapter, the methodology behind the proposed experimental setup is discussed in detail. The second section of this chapter details the experiments conducted using the described setup, and in the final section of this chapter the obtained results are visualized and analyzed.

11.1 Composing the DENV MOMAB setting

In this section, we will discuss the composition of various previously discussed elements of this study into the experimental DENV MOMAB setting. This section represents a major methodological contribution in which all previously made insights and results, into the modelling of dengue epidemics and the MOMAB framework applied to the PFI setting, are combined into a single experimental setup. In this section, the reader will first be presented with a general overview of the proposed experimental setup. Afterwards, each of the components of the proposed setup will be discussed in detail.

The general idea behind the proposed experimental setup is visualized in Figure 11.1 by following the pathway from the upper left corner to the bottom right corner. The first step is to use a deterministic compartmental model to evaluate the performance of a set $\boldsymbol{\pi}$ of considered vaccination strategies π_i. The simulations that are conducted using this model yield a single scalar value for each of the considered objectives. Afterwards, the results obtained from the deterministic simulation of each of the considered vaccination strategies is mapped to a PFI MOMAB setting. To achieve this mapping the observed deterministic rewards, in each of the objectives, for each of the considered vaccination strategies is transformed to a multivariate reward distribution. This reward distribution has a dimension for each of the considered objectives and hence, a MOMAB setting is created where each arm corresponds to a considered vaccination strategy. Concretely,

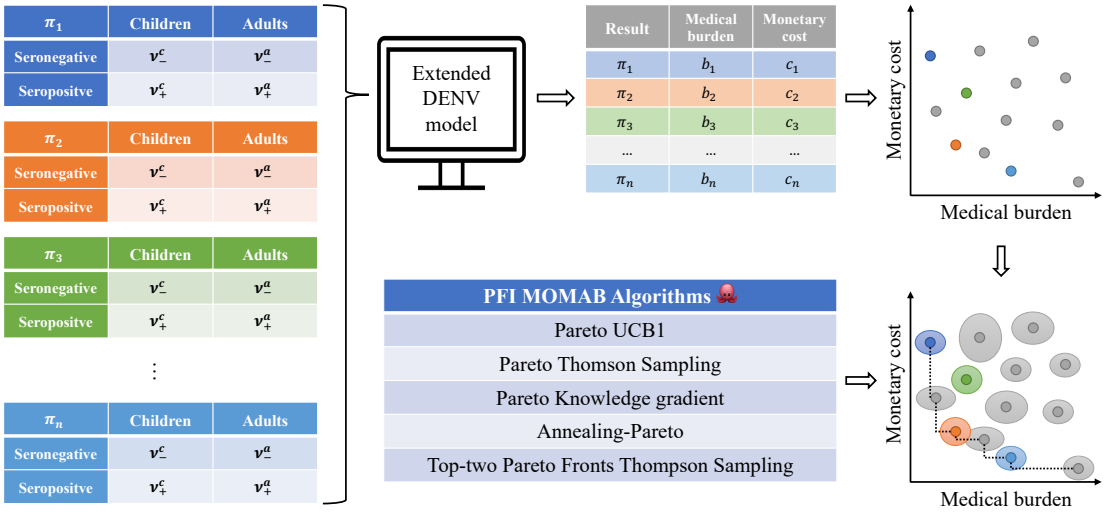


Figure 11.1: Visualization of the proposed DENV MOMAB setting. A set $\boldsymbol{\pi}$ of vaccination strategies π_i is simulated. Each of these simulations yields a medical burden b_i and monetary cost c_i for the simulated vaccination strategy. These are used as the means for the multivariate reward distribution of the arm associated with the vaccination strategy. Within this MOMAB setting, the ability of the PFI algorithms to efficiently identify the Pareto front is evaluated.

the observed deterministic performances, with respect to to each of the objectives, of each of the simulated vaccination strategies are used as the means for the multivariate reward distributions. To this mean, a standard deviation is added in each dimension, achieving a multivariate Gaussian reward distribution. This process is illustrated on the right side of Figure 11.1.

As can be observed from Figure 11.1, this process results in a MOMAB PFI setting with probabilistic rewards based on the deterministic simulations. Within this study, this serves as a basic proxy for possibly more complex and computationally expensive stochastic models that might be used in future work. This pseudo-stochastic output of the epidemiological model is not the most realistic, but since the true means are known, the ground truth for the Pareto front is also known and the performance of the PFI MOMAB algorithms can be evaluated.

In more detail, the experimental setup described in this section is essentially a combination of all aspects of this dissertation that have been discussed in the previous chapters. It consists of four main components:

1. A finite set $\boldsymbol{\pi}$ of vaccination strategies π_i from which the set of Pareto optimal vaccination strategies $\boldsymbol{\pi}^*$ will be determined.
2. An epidemic model that simulates dengue epidemics with support for incorporating the effects of the vaccination strategies π_i from $\boldsymbol{\pi}$ into the simulated disease dynamics.
3. Multiple (possibly conflicting) objectives for evaluating the performance of the vaccination strategies.
4. MOMAB algorithms for the PFI setting, used for the identification of the Pareto optimal vaccination strategies.

Starting with the first component, in Section 9.4 the nature of the vaccination strategies that

can be simulated within the proposed DENV model was discussed. Within the framework of the proposed extended model, a vaccination program or policy π_i can be represented as a tuple $(\nu_-^c, \nu_-^a, \nu_+^c, \nu_+^a)$. These values can be assigned according to the vaccination strategy that is under evaluation. They are integrated into the model through the mechanisms discussed in Section 9.3.1, regulating the flow of individuals to the vaccination-specific compartments. Within the experimental setup of the experiments discussed here, $\boldsymbol{\pi}$ consists of 53 unique vaccination strategies, An exhaustive list of all considered vaccination strategies can be found in Figure 11.3. These strategies can be split into several larger groups based on the vaccination rates.

Firstly, a control strategy without vaccination is considered:

$$\pi_0 = (0, 0, 0, 0)$$

Note that this strategy will always be one of the Pareto optimal strategies as it is associated with the lowest possible monetary cost of 0. When following this strategy, no individuals are ever vaccinated. Hence this strategy also provides a clear baseline for the impact of vaccination.

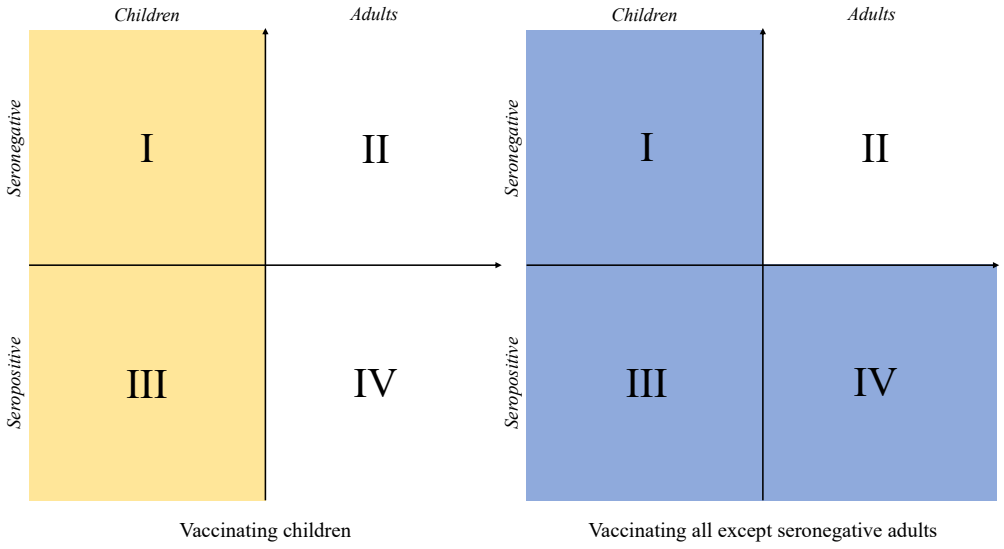


Figure 11.2: Quadrant-based representation of two vaccination strategies.

For the other groups of vaccination strategies, consider Figure 11.2. In this figure, two vaccination strategies are represented visually using the four quadrants of an orthogonal two-dimensional coordinate system. For easy reference, the colours used are the same as in Figure 11.3. Within the proposed extended DENV model from Section 9.3, the host population is stratified according to their serostatus s and their age class g. The serostatus can either be positive or negative, the age group can either be c for children or a for adults. This gives rise to the four different combinations of stratification that correspond exactly to the quadrants in Figure 11.2. This quadrant-based approach was used to ensure that a large and diverse set of possible vaccination strategies were considered. The first four groups in Figure 11.3 represent the vaccination strategies where individuals from one of the quadrants are eligible for vaccination. The next four groups represent strategies where individuals from two of the quadrants are eligible for vaccination. Thirdly, the next four groups represent strategies where individuals from three of the quadrants are eligible for vaccination. Finally, the last group encapsulates the strategies where individuals from all

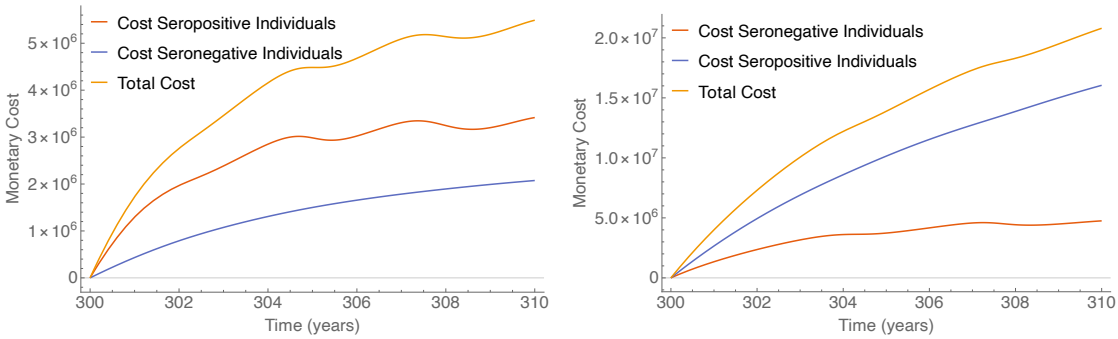
Children Seronegative	Children Seropositive	Adults Seronegative	Adults Seropositive	Medical Burden	Monetary Cost	Description
0.0	0.0	0.0	0.0	0.000178212	0.0	No vaccination
0.2	0.0	0.0	0.0	0.000177384	151.118	Vaccinating seronegative children
0.4	0.0	0.0	0.0	0.000175363	223.344	
0.6	0.0	0.0	0.0	0.000178122	264.34	
0.8	0.0	0.0	0.0	0.000179258	290.265	
0.0	0.2	0.0	0.0	0.000171716	89.4062	Vaccinating seropositive children
0.0	0.4	0.0	0.0	0.0000612023	159.024	
0.0	0.6	0.0	0.0	0.000100764	210.062	
0.0	0.8	0.0	0.0	0.0000479323	242.876	
0.0	0.0	0.2	0.0	0.000215534	872.916	Vaccinating seronegative adults
0.0	0.0	0.4	0.0	0.000207216	1219.95	
0.0	0.0	0.6	0.0	0.000229086	1396.45	
0.0	0.0	0.8	0.0	0.00023056	1501.35	
0.0	0.0	0.0	0.2	0.000114814	438.905	Vaccinating seropositive adults
0.0	0.0	0.0	0.4	0.0000375001	767.09	
0.0	0.0	0.0	0.6	0.0000912115	995.132	
0.0	0.0	0.0	0.8	0.0000241216	1157.46	
0.2	0.2	0.0	0.0	0.000154743	238.203	Vaccinating children
0.4	0.4	0.0	0.0	0.000152824	374.318	
0.6	0.6	0.0	0.0	0.00014888	460.615	
0.8	0.8	0.0	0.0	0.00014711	520.502	
0.0	0.0	0.2	0.2	0.000170809	1344.03	Vaccinating adults
0.0	0.0	0.4	0.4	0.000157591	2034.91	
0.0	0.0	0.6	0.6	0.000150665	2461.16	
0.0	0.0	0.8	0.8	0.000146629	2754.15	
0.2	0.0	0.2	0.0	0.000193672	1030.78	Vaccinating seronegative individuals
0.4	0.0	0.4	0.0	0.000232698	1456.25	
0.6	0.0	0.6	0.0	0.000212928	1673.57	
0.8	0.0	0.8	0.0	0.000188332	1802.22	
0.0	0.2	0.0	0.2	0.000244826	520.393	Vaccinating seropositive individuals
0.0	0.4	0.0	0.4	0.0000130795	833.552	
0.0	0.6	0.0	0.6	0.000012716	1087.78	
0.0	0.8	0.0	0.8	0.0000126325	1276.81	
0.2	0.2	0.2	0.0	0.00018908	1138.94	Vaccinating all except seropositive adults
0.4	0.4	0.4	0.0	0.000204887	1646.38	
0.6	0.6	0.6	0.0	0.000205712	1916.92	
0.8	0.8	0.8	0.0	0.000199903	2085.50	
0.2	0.2	0.0	0.2	0.000123913	705.175	Vaccinating all except seronegative adults
0.4	0.4	0.0	0.4	0.000074736	1204.60	
0.6	0.6	0.0	0.6	0.0000575192	1547.13	
0.8	0.8	0.0	0.8	0.0000413498	1789.70	
0.2	0.0	0.2	0.2	0.000168138	1511.94	Vaccinating all except seropositive children
0.4	0.0	0.4	0.4	0.000169054	2290.00	
0.6	0.0	0.6	0.6	0.00016065	2756.86	
0.8	0.0	0.8	0.8	0.000156271	3073.46	
0.0	0.2	0.2	0.2	0.00015572	1479.58	Vaccinating all except seronegative children
0.0	0.4	0.4	0.4	0.000130456	2344.22	
0.0	0.6	0.6	0.6	0.0000998519	2914.99	
0.0	0.8	0.8	0.8	0.0000919958	3313.40	
0.2	0.2	0.2	0.2	0.00016067	1638.18	Vaccinating everyone
0.4	0.4	0.4	0.4	0.000141263	2518.22	
0.6	0.6	0.6	0.6	0.000136102	3075.19	
0.8	0.8	0.8	0.8	0.000127594	3461.97	

Figure 11.3: Table of the vaccination strategies that were considered within the experiments, together with the corresponding rates of vaccination ν_s^g and the resulting medical burden and monetary cost.

four quadrants are eligible for vaccination. The quadrants are numbered according to the indices of the rate of vaccination ν_s^g in π_i. Within each group four different vaccination strategies were considered where respectively 20%, 40%, 60%, and 80% of eligible individuals, for that group of vaccination strategies, are vaccinated each year.

The second major component of the experimental setup is the epidemic model for dengue epidemics that is able to simulate the effects of the different vaccination strategies. Within the context of this dissertation, two dengue epidemic models were reproduced. The reproduction of the first model by Ferguson et al. from 2016 [61, 60] can be found in Section 9.1. The main idea behind this model was that vaccination could be modeled as a silent infection [61]. Although the reproduction of this model was not completed due to technical restrictions, the ideas behind it were then used to expand the model proposed by Recker et al. in 2009 [137]. The reproduction for this model can be found in Section 9.2. Subsequently, this reproduced model was expanded with support for vaccination strategies and age-heterogeneity in Section 9.3. This expanded model, visualized in Figure 9.15, suited the needs of this study perfectly and is thus used to evaluate the vaccination strategies.

The objectives used for evaluating the performance of the vaccination strategies are the third component of the experimental setup. It was established in Chapter 9 that the medical burden endured by the host population as a result of the prevalence of the disease in the population, and the monetary cost associated with the vaccination strategy, are interesting conflicting objectives to examine. It was decided that infections requiring hospitalization would be used as the value for the medical burden endured by the population. In their paper, Ferguson et al. propose proportions for the number of infections that result in symptomatic disease, and the number of cases with symptomatic disease that require hospitalization [60]. These proportions are defined for primary, secondary, tertiary, and quaternary infections. Hence, the proposed values for the primary and secondary infections can be used together with the simulated primary and secondary infections data from the DENV model to calculate the total hospitalizations within a certain time frame used for evaluation. An example of this hospitalization data can be seen in Figure 9.17.



(a) Monetary cost associated with vaccinating children.

(b) Monetary cost associated with vaccinating all except seronegative adults.

Figure 11.4: Visualization of the monetary cost associated with two distinct vaccination strategies, evaluated over a ten year period after the start of the vaccination program.

As the model has explicit compartments which individuals pass through as they are vaccinated, the cost associated with a vaccination strategy can be determined based on the number of individuals in these compartments. As discussed in Section 9.3.1, it is very important to make a distinction between the cost of vaccinating a seronegative individual and the cost of vaccinating

a seropositive individual as these differ. The latter might have come into contact with dengue before but be unaware of this due to the possible silent nature of the infection. There is not that much information to be found on the approximate true cost of vaccinating a single individual against DENV or testing them for antibodies. Using the values used in other studies [204, 167], in this study the cost of vaccination κ_1 is set to 20 and the cost of screening for antibodies κ_2 is set to 10.

The fourth and final major component of the experimental setup consists of the MOMAB PFI algorithms. The primary objective of the experiments is to evaluate whether these MOMAB PFI algorithms can be used to efficiently identify the Pareto optimal vaccination strategies. The goal is to identify the entire subset of Pareto optimal vaccination strategies in a sample-efficient manner[1] and be able to present this set of trade-offs to the decision makers, without making assumptions about their possible preferences.

In Section 10.2, the reproductions of nine distinct MOMAB algorithms for regret minimization were discussed. There, the differences and relations between the regret minimization setting for which the algorithms were designed, and the PFI setting of this dissertation are discussed. Section 10.2 also deals with the valuable insight that might be gained from the reproduction of these regret minimization MOMAB algorithms, even though their purpose does not correspond exactly to the MOMAB PFI setting presented in this dissertation. Finally, in Section 10.2 the extension of four of the regret minimization MOMAB algorithms to the PFI setting was discussed: Pareto UCB1, Pareto Thompson Sampling, Pareto Knowledge Gradient, and Annealing Pareto. It was also motivated why the remaining five variants were not suited for extension to the PFI setting. A complete overview of the reproduced and extended algorithms can be found in Figure 10.5. The performance of the four extended algorithms was then evaluated with respect to the PFI metrics (Bernoulli, Jaccard similarity, and hypervolume) proposed by this dissertation in Section 10.1, using the experiment visualized in Figure 10.6. From these experiments it became clear that all extended algorithms were well-suited to the PFI setting, with the Pareto UCB1 and Pareto Thompson Sampling algorithms performing better than the Pareto Knowledge Gradient and Annealing Pareto algorithms, especially with respect to efficiently identifying the entire set of optimal arms.

Last but definitely not least, in addition to the four MOMAB PFI algorithms discussed in Section 10.2, this study also proposed a completely novel preference-unaware MOMAB PFI algorithm in Section 10.3: Top-two Pareto Fronts Thompson sampling (TTPFTS). Intuitively, this algorithm is an extension of the Top-two Thompson Sampling algorithm for the single-objective best-arm identification setting. The mechanisms employed by TTPFTS to decide which arms to pull and to recommend to the user are discussed in detail in Section 10.3. The performance of the novel TTPFTS algorithm was also compared against the other four MOMAB PFI algorithms with respect to the PFI metrics proposed in Section 10.1. The results from those experiments can be found in Figure 10.10.The results showed that the TTPFTS algorithm clearly outperformed the four other PFI MOMAB algorithms. Hence, the four extended algorithms from Section 10.2 and the novel TTPFTS algorithm from Section 10.3 are used in the final experiments. To assess whether the MOMAB PFI algorithms represent an improvement in sample-efficiency, Uniform Sampling was also included in the final experiments as a performance baseline.

With that, all major components of the proposed experimental setup have been discussed. In the next section, the actual experiments that were conducted within the context of this study using the presented methodological contribution will be looked at in detail.

[1] I.e. more sample-efficient than the Uniform Sampling algorithm currently employed in most studies evaluating the effects of mitigation strategies for various epidemics.

11.2 Experiments

In this section, the experiments conducted using the proposed experimental setup from the previous section will be discussed. The goal of these experiments was to verify whether Pareto front identification (PFI) multi-objective multi-armed bandit (MOMAB) algorithms can be used to identify the subset $\boldsymbol{\pi}^*$ of optimal vaccination strategies from a larger set $\boldsymbol{\pi}$ of vaccination strategies under consideration in sample-efficient manner.

Starting with the set $\boldsymbol{\pi}$ of vaccination strategies under consideration, Figure 11.3 shows a complete overview of the 53 vaccination strategies that were considered for this study's final experiments. A rich and diverse set of different vaccination strategies was ensured using the quadrant-based approach described in Section 11.1. The effects of each of the vaccination strategies were then simulated using the model proposed by Recker et al. [137] extended with support for vaccination and age-heterogeneity as described in Sections 9.2 and 9.3. Based on these simulations, each of the vaccination strategies is associated with a value for both the conflicting objectives: the medical burden endured by the population as a result of the prevalence of DENV, and the monetary cost of the vaccination program. The values for both objectives for each of the simulated vaccination strategies are shown in Figure 11.3 in the *Medical Burden* and *Monetary Cost* columns. The far right column of Figure 11.3 provides some information about the nature of the vaccination strategies within each group.

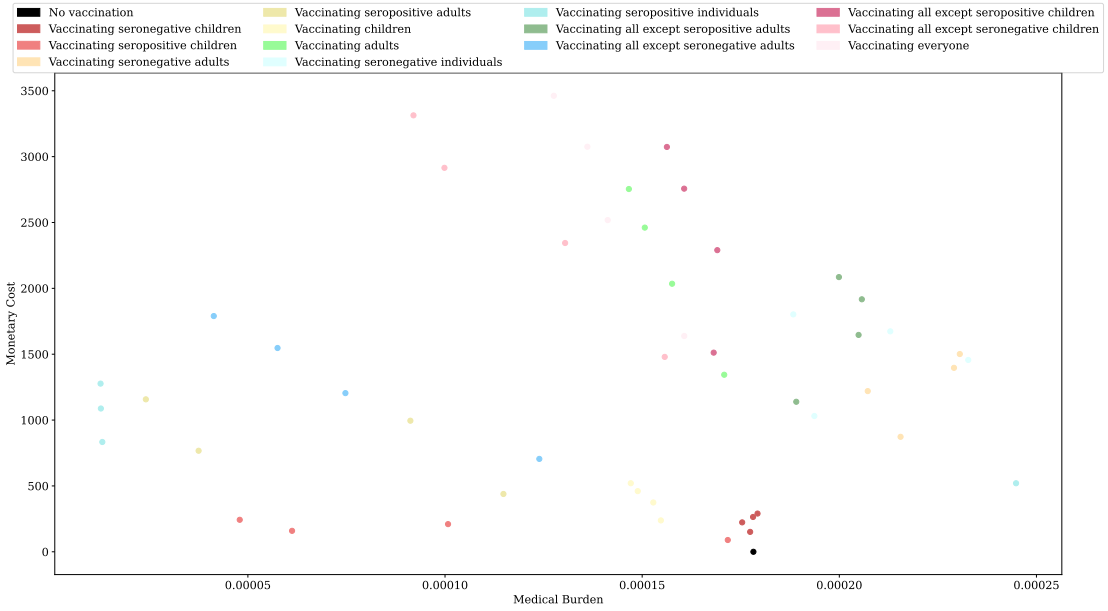


Figure 11.5: Scatter plot of the obtained simulated medical burden and monetary cost over a ten year period associated with each of the 53 vaccination strategies.

The results with respect to the two objectives for each of the vaccination strategies are then plotted as shown in Figure 11.5. In this figure, the results with respect to the two objectives for each of the vaccination strategies are shown as a scatter plot where the horizontal axis represents the *Medical Burden* objective and the vertical axis represents the *Monetary Cost* objective. For easier interpretation, the points have been colored based on their group according to the colors used in Figure 11.3.

Optimal vaccination strategies are associated with a low medical burden and a low medical cost. As the goal is to minimize these two conflicting objectives, the plot in Figure 11.5 can be modified, yielding the annotated scatter plot shown in Figure 11.6.

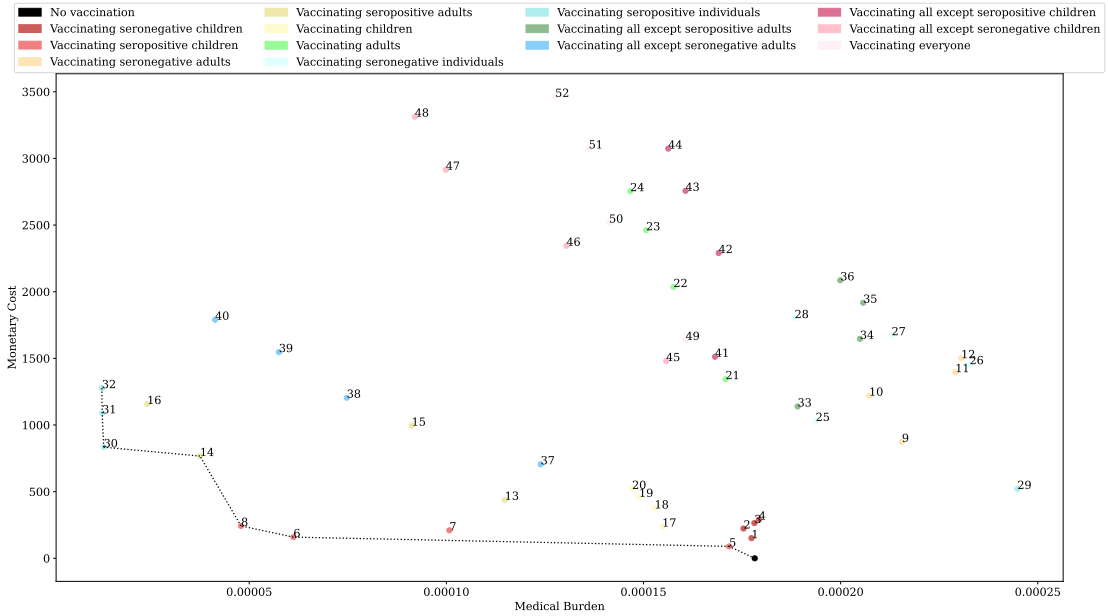


Figure 11.6: Scatter plot of the obtained simulated medical burden and monetary cost over a ten year period associated with each of the 53 vaccination strategies. Strategies are annotated with their index from Figure 11.3. Optimal strategies are connected with a dotted line.

This annotated version shows the index of each of the strategies based on their position in Figure 11.3. The optimal strategies are connected with a dotted line for easier identification. The dotted line does not correspond directly with the true Pareto front for this DENV MOMAB setting. However, it gives some insights into the nature of the Pareto front. The first observation is the strong non-convex nature of the Pareto front. This is visible between strategies 30, 14, and 8. Secondly, it can be observed that some of the optimal strategies are associated with values for the objectives that are extremely close to one another. A good example of this are strategies 30, 31, and 32 which are very closely related with respect to the *Medical Burden* objective. Both this non-convex nature, as well as the close proximity of some of the optimal strategies result in an interesting DENV MOMAB setting that is challenging for the PFI algorithms.

As can be observed from Figures 11.3, 11.5 and 11.6, by nature, the units used for the medical burden objective and the monetary cost differ, resulting in a very large difference in range between both objectives. To further standardize the experiments proposed in this study, it was decided to normalize the data with respect to both objectives. The result of this normalization can be seen in figure Figure 11.7. Both objectives were scaled to have values between 0 and 1, while keeping the relative distance between points identical. This normalization also ensures that the hypervolume metric will have a value between 0 and 1, in correspondence to the Bernoulli and Jaccard metrics.

The final way in which the data was transformed in preparation for the experiments was through an inversion shown in Figure 11.8. The PFI MOMAB algorithms assess the optimality of an arm,

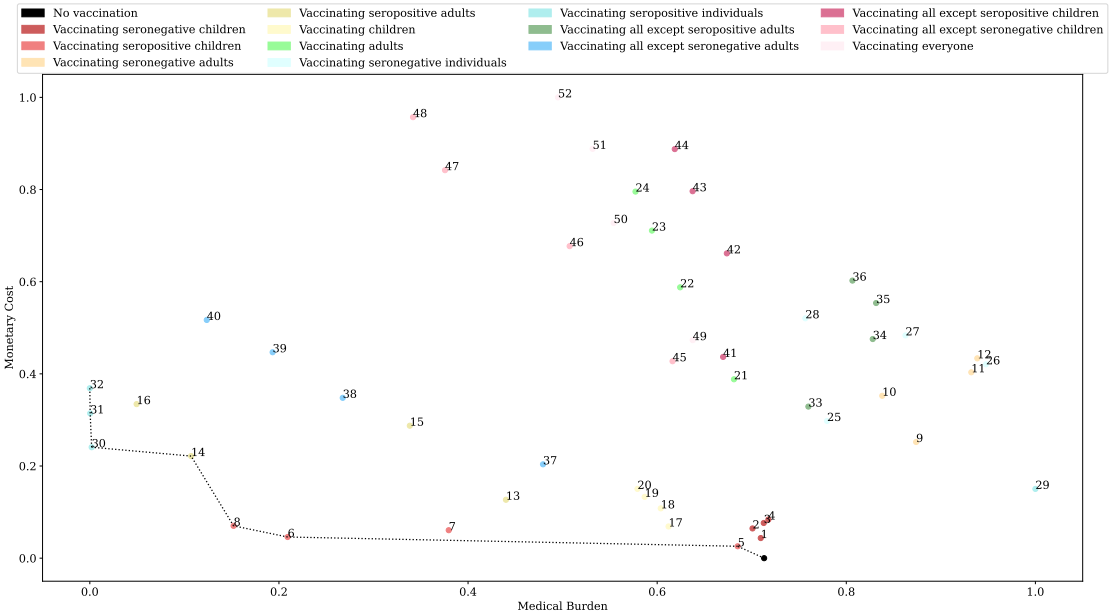


Figure 11.7: Scatter plot of the normalized obtained simulated medical burden and monetary cost over a ten year period associated with each of the 53 vaccination strategies. Strategies are annotated with their index from Figure 11.3. Optimal strategies are connected with a dotted line.

or in this case vaccination strategy, based on the observed rewards with respect to the different objectives, where larger rewards indicate a more performant arm. This inversion essentially transforms the minimization setting into a maximization setting within which the PFI MOMAB algorithms can be employed. The points corresponding to the optimal vaccination strategies that were previously found in the bottom left, are now located in the top right corner of the plot, connected via a dotted line.

Up to this point, the data for which the transformation process has been discussed, was the output of the deterministic DENV model proposed in Section 9.3. However, the discussed PFI MOMAB algorithms work with stochastic rewards. Furthermore, one of the main goals of this dissertation was to examine whether PFI MOMAB algorithms could be used to identify the subset of optimal vaccination strategies based on the output of a computationally expensive stochastic model in a sample-efficient manner. The idea now is to use the deterministic output of the DENV model, together with a fixed standard deviation, to define a multivariate Gaussian reward distribution for each of the arms. These probabilistic rewards based on the deterministic simulations then serve as a basic proxy for possibly more complex and computationally expensive stochastic models that might be used in future work. This pseudo-stochastic output of the epidemiological model is not the most realistic, but since the true means are known, the ground truth for the Pareto front is also known and the performance of the PFI MOMAB algorithms can be evaluated. In the experiments presented here, a standard deviation of 0.1 is used consistently across all arms and objectives.

To evaluate the ability of the PFI MOMAB algorithms to identify the entire subset of optimal vaccination strategies based on the created multivariate reward distributions, each algorithm was

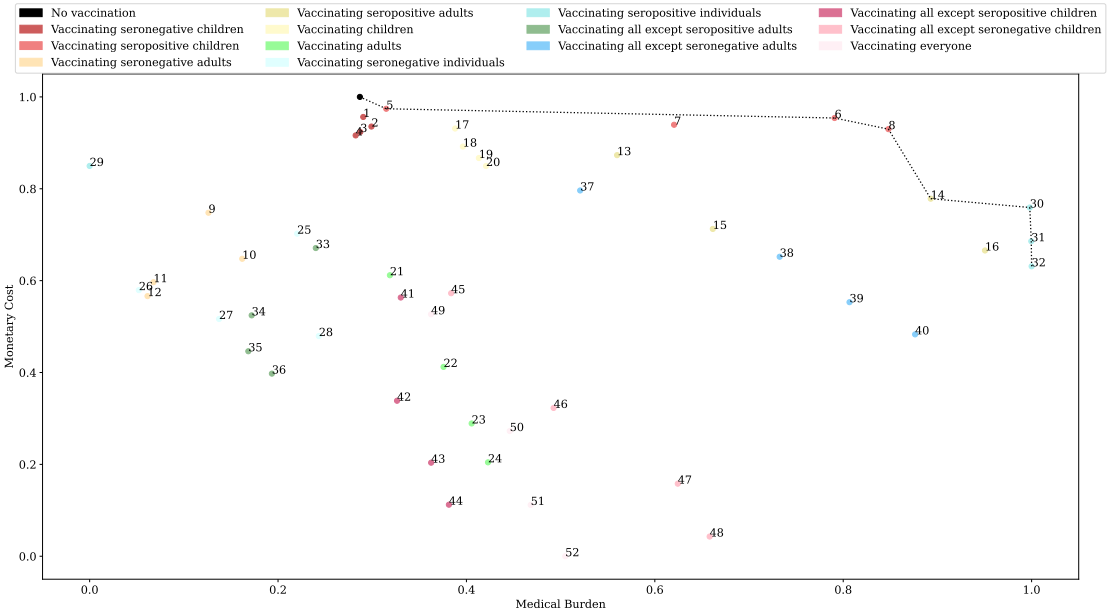


Figure 11.8: Scatter plot of the normalized and inverted obtained simulated medical burden and monetary cost over a ten year period associated with each of the 53 vaccination strategies. Strategies are annotated with their index from Figure 11.3. Optimal strategies are connected with a dotted line.

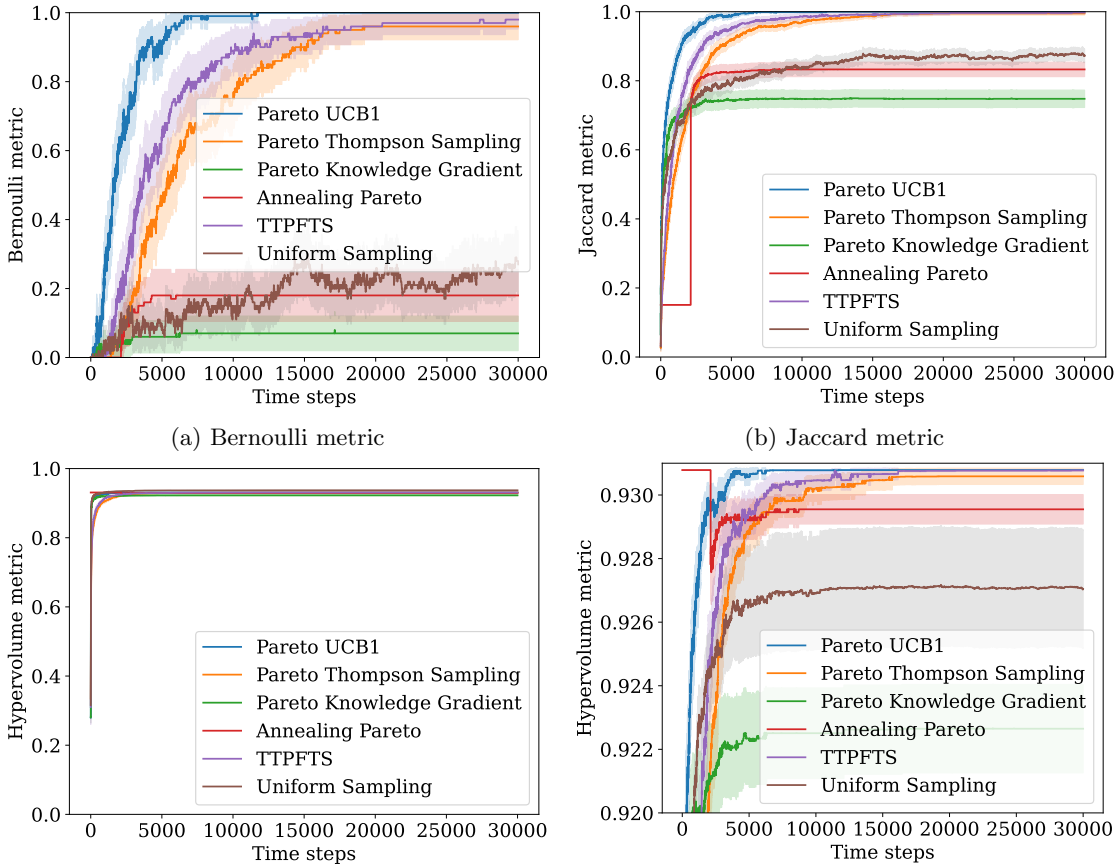
ran with a budget of 30,000 arm pulls. After each arm pull t, the set of arms $\mathcal{R}_t$ recommended by the PFI MOMAB algorithms was logged. This experiment was repeated 100 times for each algorithm. With that, the experimental setup and the way it was used to examine the research question have been discussed in detail. In the next section, the results of the experiments will be discussed.

11.3 Results

In this final section on the DENV MOMAB setting proposed within this dissertation, the results of the PFI experiments will be discussed. As was discussed in the previous section, each of the PFI MOMAB algorithms was given a budget of 30,000 arm pulls to identify the complete subset $\boldsymbol{\pi}^*$ of optimal vaccination strategies. After each arm pull t, the PFI MOMAB algorithm was asked for its set of recommended arms $\mathcal{R}_t$. This information was logged and the experiment was repeated 100 times for each of the algorithms.

Based on the logged recommendations made by each of the algorithms over the different runs of the experiment, the average quality of the recommendations made by the PFI algorithms was evaluated using the performance metrics presented in Section 10.1. These metrics were specifically developed to quantify the quality of the recommendations $\mathcal{R}_t$ made by a PFI MOMAB algorithm with respect to the true Pareto front $\mathcal{A}^*$. The results of this analysis can be found in Figure 11.9.

Starting with the Bernoulli metric in Figure 11.9a, it can be seen that all presented PFI MOMAB algorithms are able to give perfect recommendations to a certain extend, but some algorithms



(a) Bernoulli metric

(b) Jaccard metric

(c) Hypervolume metric

(d) Hypervolume metric - magnified

Figure 11.9: Results of the extended MOMAB algorithms including TTPFTS in the final DENV MOMAB setting experiment with respect to the Bernoulli, Jaccard, and hypervolume metrics. Results are averaged over 100 runs of the algorithm with a budget of 30,000 arm pulls. The shaded area shows the 95% confidence interval around the mean.

perform noticeably better than others. Starting with the worst performing algorithm, it can be observed that Pareto Knowledge Gradient gives perfect recommendations approximately 10% of the time after 7500 arm pulls. Improving on this, the Annealing Pareto algorithm gives perfect recommendations approximately 20% of the time after only 5000 arm pulls. This indicates that within this setting, the Annealing Pareto algorithm is more sample-efficient than the Pareto Knowledge Gradient algorithm, while also being better at learning the true Pareto front. Although initially, the Bernoulli metric for the Annealing Pareto algorithm shows a rapid increase, it plateaus around 0.2, being surpassed in performance by the remaining three other PFI MOMAB algorithms. It is also valuable to examine the performance of Pareto Knowledge Gradient and Annealing Pareto with respect to the Uniform Sampling baseline. In the proposed experimental DENV MOMAB setting, Pareto Knowledge Gradient is consistently outperformed with regards to the Bernoulli metric by the Uniform Sampling baseline. However, when observing the Bernoulli metric curves for the Uniform Sampling baseline and Annealing Pareto, it can be deduced that Annealing Pareto is better at giving perfect recommendations when the budget

is limited to approximately 12.500 arm pulls. If the budget is chosen to be larger, the Bernoulli metric performance of Uniform Sampling is greater than that of Annealing Pareto. This result indicates both a strength and a weakness of the Annealing Pareto algorithm in the DENV MOMAB setting when compared to the baseline, as it learn to give perfect recommendations more often using less arm pulls, yet stagnates being overtaken by the baseline.

When looking at the Bernoulli metric curve for Pareto Thompson Sampling, it can be seen in Figure 11.9a that this algorithm shows a rapid increase in perfect recommendation frequency after around 2000 arm pulls. This initial increase indicates that, as more arms are pulled, Pareto Thompson Sampling quickly learns to make better recommendations that are identical to the true Pareto front. This increase slows after approximately 7500 arm pulls where a value of 0.75 is reached for the Bernoulli metric. After this point, the value for the Bernoulli metric steadily keeps increasing until approximately 20,000 arm pulls of the budget have been used. From 20,000 arm pulls onwards, Pareto Thompson Sampling gives perfect recommendations approximately 95% of the time, with the shaded confidence interval reaching the theoretical maximum value of 1 for the Bernoulli metric. The next-best performing algorithm with respect to the Bernoulli metric is the novel Top-two Pareto Fronts Thompson sampling (TTPFTS) algorithm that was presented in Section 10.3. Its learning curve is very similar in nature to that of Pareto Thompson Sampling, with the advantage of achieving similar great performance in less arm pulls. Finally, of the five tested algorithms, Pareto UCB1 showed the best performance with respect to the Bernoulli metric in the DENV MOMAB setting. The learning curve of this algorithm is characterised by an almost immediate steep increase in Bernoulli metric. This steep increase indicates that the algorithm very quickly identifies and recommends the entire set of Pareto optimal arms. After approximately 5000 arm pulls, the increase slows. However, at this point, the Pareto UCB1 algorithm already makes perfect recommendations 90% of the time. After the 5000'th arm pull, the algorithm's performance increases further, reaching the theoretical maximum value of 1 for the Bernoulli metric after approximately 12.000 arm pulls. This indicates that from that point onward, the algorithm will always perfectly recommend the complete subset of optimal arms.

Comparing the performance of Pareto Thompson Sampling, TTPFTS, and Pareto UCB1 with respect to the Bernoulli metric with the performance of the Uniform Sampling baseline, it can be observed that these three top algorithms perform significantly better. After 17.500 arm pulls Pareto Thompson Sampling, TTPFTS, and Pareto UCB1 almost always perfectly recommend the Pareto optimal arms. When given the same budget, the baseline manages perfect recommendations approximately 30% of the time. This means that, when employing a conservative approach to calculating the performance increase, these three MOMAB algorithms represent a threefold improvement in performance relative to the baseline with respect to the Bernoulli metric.

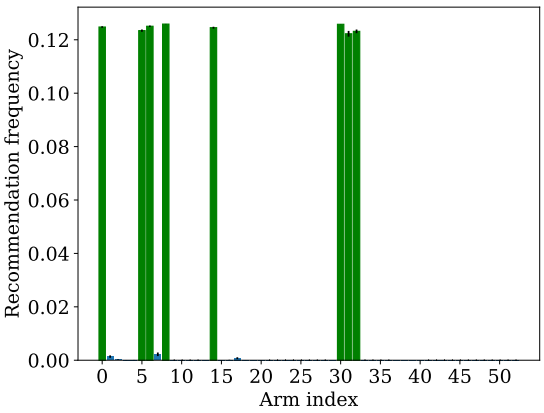
Moving on to the Jaccard similarity metric, the relative performance between algorithms that was established with respect to the Bernoulli metric remains unchanged. From Figure 11.9b, it can be observed that all algorithms show a very fast increase in Jaccard metric as they spend the initial arm pulls of their budget. This indicates that the set of arms recommended by the algorithms and the true set of Pareto optimal arms quickly become more similar, which implies the ability of the algorithms to learn the subset of Pareto optimal arms. Figure 11.9b shows that after approximately 5000 arm pulls, the performance of Pareto Knowledge Gradient and Annealing Pareto stagnates at values of 0.75 and 0.85 respectively. This is an indication that both these algorithms managed to learn a significant portion of the true subset op optimal arms, but did not manage to identify all optimal arms. From the data shown in Figure 11.9b, it can also be observed that the other algorithms did not stagnate, each reaching values for the Jaccard

metric close or equal to the theoretical maximum value 1. Similarly to the performance with respect to the Bernoulli metric, the learning curves for Pareto Thompson Sampling and TTPFTS share the same shape, reaching extremely good Jaccard similarity after 15,000 and 10,000 arm pull respectively. Once again, TTPFTS outperforms Pareto Thompson Sampling. Pareto UCB1 reaches very high values with respect to the Jaccard metric after only 5000 arm pulls, indicating the algorithm's ability to efficiently learn the true set of Pareto optimal arms. Interestingly, the point at which Pareto UCB1, Pareto Thompson Sampling, and TTPFTS reach values for the Jaccard metric that are close to the theoretical maximum of 1, corresponds to the point at which their steep increase in Bernoulli metric starts slowing down. This is an indication that some arms in the Pareto front are particularly challenging to efficiently identify as being optimal. This will be verified later in this section using the arm recommendation frequencies shown in Figure 11.10.

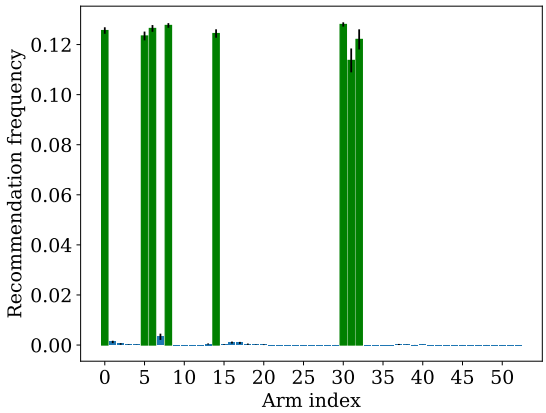
When comparing the performance of the PFI MOMAB algorithms with the Uniform Sampling baseline with respect to the Jaccard metric, the same trends that emerged when studying the Bernoulli metric are visible. Once again, Pareto Knowledge Gradient is outperformed by the baseline, and Annealing Pareto is more sample-efficient than Uniform Sampling up to approximately 12.500 arm pulls. It can also be deduced from Figure 11.9b that Pareto UCB1, Pareto Thompson Sampling, and TTPFTS are more efficient than the baseline at recommending sets of arms more similar to the true Pareto optimal set. After approximately 17.500 arm pulls, all three of these MOMAB algorithms reach the theoretical maximum value of 1 for the Jaccard metric, while the Uniform Sampling baseline reaches 85% Jaccard similarity. Combining these observations with those of the Bernoulli metric curves in Figure 11.9a already provides strong evidence for improved sample-efficiency MOMAB PFI algorithms can offer.

The final metric that needs to be discussed is the hypervolume metric. Once again, the relative performance between algorithms that was established when observing the previous two metrics remains identical. However, from this metric shown in Figure 11.9d, it can be seen that the quality of the recommendations made by the Annealing Pareto algorithm is significantly better than those made by Pareto Knowledge Gradient. It is even the case that the hypervolume of the recommendations made by the Annealing Pareto algorithm is close to the hypervolume of the recommendations made by Pareto Thompson Sampling. This result is interesting as the Pareto Thompson Sampling algorithm clearly outperformed Annealing Pareto with respect to the other two metrics by quite a large margin. From this, it can be deduced that even though Annealing Pareto fails to identify the entire Pareto front, it still makes high quality recommendations that dominate close to optimally large parts of the objective space. The same interesting trend continues when also considering the Uniform Sampling baseline. Even though the performance of the baseline algorithm with respect to the previous two metrics surpassed that of Annealing Pareto for budgets larger than 12.500 arm pulls, the hypervolume of the recommendations made by Annealing Pareto remain larger than those of the baseline. This is another indication that although Annealing Pareto fails to identify the entire Pareto front, the recommendations it makes are of a high quality. This also highlights that, within this setting, it is very useful to have multiple performance metrics, as only a single metric might not tell the entire story.

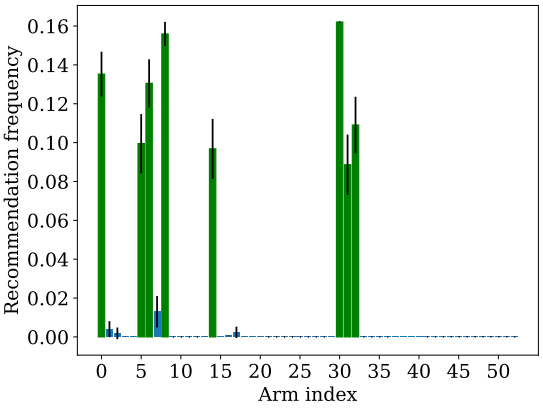
To gain further insights into the results obtained with respect to the different performance metrics in Figure 11.9, the frequency with which the different algorithms recommend the different arms was also examined. This statistic can be seen visualized in Figure 11.10. In this figure, the average recommendation frequency for each arm over the 100 experiments is plotted as a bar, with the 95% confidence interval around the mean being plotted as an error bar. The bars corresponding to the optimal arms are plotted in green, while the suboptimal arms are plotted in blue. Starting with the data for Pareto Knowledge Gradient in Figure 11.10c, it becomes



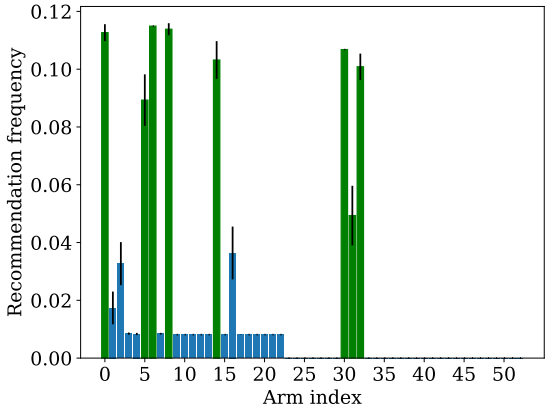
(a) Pareto UCB1



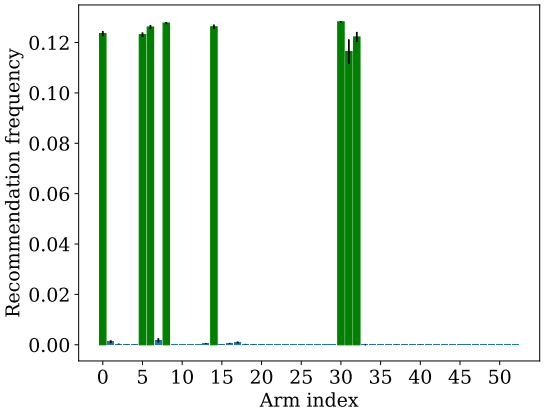
(b) Pareto Thompson Sampling



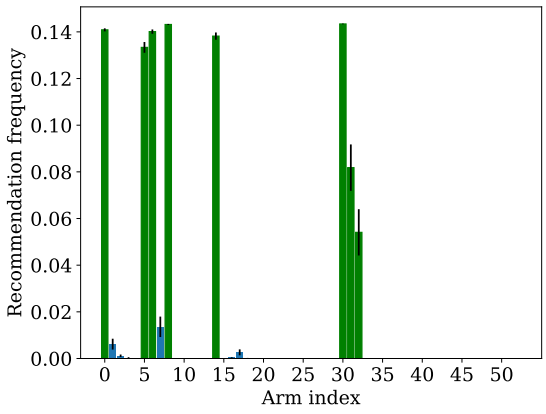
(c) Pareto Knowledge Gradient



(d) Annealing Pareto



(e) Top-two Pareto Fronts Thompson Sampling



(f) Uniform Sampling

Figure 11.10: Average arm recommendation frequencies of the PFI MOMAB algorithms in the final DENV MOMAB setting experiment. Results are averaged over 100 runs of the algorithm. The error bars show the 95% confidence interval around the mean. Optimal arms are plotted in green.

clear why this algorithm had the worst performance in the experiments. The frequency with which the optimal arms are recommended varies greatly between optimal arms, indicating that some arms were identified as optimal and recommended much more often than other arms. The confidence intervals for individual optimal arms are also relatively large when compared to those of the other algorithms. This indicates a large variability in whether that specific optimal arm is identified as being optimal and subsequently recommended to the user. Pareto Knowledge Gradient also suffers from recommending arms that are close to the Pareto front but suboptimal, such as arms 1, 2, and 7.

Annealing Pareto was the fourth best performing algorithm, showing interesting results when comparing its scores for the different performance metrics in Figure 11.9. Figure 11.10d shows the average frequency with which Annealing Pareto recommended each of the arms. This visualization corresponds neatly to the intuition behind the Annealing Pareto algorithm which gradually tightens the bound for the arms it considers to be optimal. Because of this, the arms close to the Pareto front are recommended often until the bound becomes strict enough for them not to be considered optimal anymore. This is especially clear for suboptimal arms 1, 2, and 7 which are closest to the true Pareto front. Note that the intuition behind the mechanisms employed by Annealing Pareto, visible in Figure 11.10d, also provides a plausible explanation as to why its performance with respect to the hypervolume metric was superior to the performance with regards to the other metrics. By design and due to the choice of the reference point, the highest attainable hypervolume is associated with recommendations that include the entire subset of optimal arms. During its initialization phase, Annealing Pareto does not tighten the threshold and hence all arms, including the true subset of optimal arms, are considered to be Pareto optimal. This results in the maximum attainable hypervolume metric during approximately the first 2500 arm pulls shown in Figure 11.9d. After the initialization phase, Annealing Pareto starts updating the set of arms it considers to be Pareto optimal. Gradually, arms that are suboptimal are removed depending on how suboptimal the algorithm estimates the arms to be, where more suboptimal arms are removed earlier. Consequently, the set of arms considered to be optimal by Annealing Pareto might not correspond perfectly to the true subset of Pareto optimal arms, but will always include arms that are (close to) optimal. The algorithm then recommends this set which will always be associated with a large hypervolume, explaining why Annealing Pareto performs so good with regards to this metric. The visualization in Figure 11.10d indicates that the improvement in performance over Pareto Knowledge Gradient stems mainly from the improved consistency with which optimal arms are identified. Where Pareto Knowledge Gradient only consistently identified arm 30, Annealing Pareto steadily recommends arms 0, 6, 8, and 30.

Moving to the best performing algorithms, it can be observed that their recommendation frequencies look very similar to one another. Pareto UCB1 performed best with respect to all proposed performance metrics. When looking at the frequency with which Pareto UCB1 recommends each of the arms, we can see that it almost never recommends suboptimal arms, and whenever a suboptimal arm is recommended, it is a suboptimal arm that is very close to the true Pareto front. Furthermore, it can also be observed that it identifies and recommends all optimal arms with the same frequency. The absence of noticeable error bars for the bars corresponding to the optimal arms also shows the stability with which Pareto UCB1 identifies the subset of optimal arms. The other two algorithms that performed very good in the experiments are Pareto Thompson Sampling and TTPFTS. Their average recommendation frequencies plotted in Figure 11.10b and Figure 11.10e respectively look very similar, both consistently identifying the entire subset of optimal arms. The largest variability in recommendation frequency is present for arms 31 and 32 for both algorithms. As this variability is not present in the data from Pareto UCB1, this can be used to explain the performance deficit with respect to Pareto UCB1: Pareto Thompson

Sampling and TTPFTS were challenged by arms 30, 31, and 32 which are in extremely close proximity of one another yet are all Pareto optimal.

In conclusion, this section has provided a detailed visualization and analysis of the experimental results obtained from the novel DENV MOMAB setting investigated in this dissertation. Through extensive evaluation using multiple performance metrics, the effectiveness and efficiency of various PFI MOMAB algorithms has been thoroughly examined. Each algorithm's ability to identify optimal vaccination strategies within a defined budget of arm pulls was assessed across 100 experimental runs, shedding light on their respective strengths and weaknesses, while also comparing their performance with the Uniform Sampling baseline. With these results meticulously analysed and discussed, the subsequent parts of this dissertation deal with the final discussion and conclusion.

Part V

Discussion

Throughout this dissertation, an extensive number of methodological contributions were proposed, interesting results were presented and analyzed, assumptions were made, and sometimes strong statements were posited. This second to last part of the study is dedicated to the observations and results which require some additional discussion. Furthermore, throughout this section, we aim to reflect critically on the conducted research as this is of great importance to any scientific work, resulting in new insights. Finally, we ruminate about possible future work and how this might build upon the research conducted in this study.

Throughout this Discussion section, an identical structure to the rest of the dissertation will be used, starting with discussing dengue epidemiological modelling, then moving on to the MOMAB framework, and finally arriving at the newly proposed DENV MOMAB setting.

Starting with the creation of an epidemiological model simulating dengue epidemics, during the initial phases of the study, it quickly became clear that a model needed three properties in order to be suitable: (i) it needed to allow the simulation of the effects of various vaccination strategies, (ii) it needed to include age-heterogeneity, and (iii) its output needed to be stochastic for use in a MOMAB setting. To this end, the initial plan was to reproduce the state-of-the-art, yet deterministic compartmental model proposed in 2016 by Ferguson et al, as this model excellently incorporated needs (i) and (ii).

Following this reproduction, the plan was then to extend an existing stochastic individual based model for dengue epidemics with the features from the 2016 Ferguson et al. model, resulting in a model that suited all needs. However, due to various factors, the reproduction of the 2016 Ferguson et al. model was only partially completed. Two main reasons why the reproduction was only a partial success were identified: technical limitations, and lack of certain information aiding reproducibility provided throughout the original publications. Starting with the technical limitations, Wolfram Mathematica was used to reproduce compartmental models as these models are governed by systems of differential equations. While the simplest models are governed by systems of ordinary differential equations, the 2016 Ferguson et al model was governed by continuous time-delayed partial differential equations. The use of such complex equations was warranted by the modeled effects of the incubation period on infection and the stratification of the host population into continuous age classes. The ODE solver that is present in Wolfram Mathematica is able to solve systems of delay-differential equations like the one that was being reproduced, however, it can only handle discrete delays. Furthermore, the initialization of the equations across a continuous number of age classes was unable to be completed due to Wolfram Mathematica's limitations with regards to continuous variables. In their original publication, Ferguson et al. did not include their own implementation of the model, neither the ODE solver they employed, nor the initial conditions used to obtain their results. The combination of these elements made it so that the reproduction effort had to be abandoned. However, this partial reproduction provided a great amount of learning opportunities about the intricacies of more complex models for dengue virus, and the key insight that vaccination could be modeled as a silent infection was made. These were later leveraged in the model used in the final DENV MOMAB. This partial reproduction also provides a solid base for future work and continued reproduction efforts. One avenue might be to discretize the age groups and adapt the model. However, this adaptation is non-trivial as the continuous nature of the age groups is crucial for the implementation of certain parts of the model.

After the partial success of the reproduction of the 2016 Ferguson et al. model, the decision was made to reproduce and extend the 2009 Recker et al. model instead. However, the first attempts at reproducing the results obtained by the authors in their original paper, and thus verifying the correctness of the reproduced model, were unsuccessful. Only a subset of the proposed

combinations of parameter values yielded sound results, while others caused high degrees of numerical instability in the ODE solver, resulting in results that were not viable for interpretation. Upon contacting the authors of the paper, it became clear that they had not mentioned a so called importation rate that causes an additional flux of individuals from the susceptible compartment to the primary infections compartments. This adjustment enhances the numerical stability of the simulation. Upon further examination of the importation rate, we hypothesized that it might serve as an additional seeding mechanism necessary for simulating disease dynamics over multiple millennia. After further interesting correspondence with the authors, we concluded that the persistence of dengue during the off-season or in regions not typically considered conducive to dengue transmission remains an open research question. After including the importation rate, similar results to those obtained in the original study were obtained, and consequently the now verified reproduced model was extended to suit all needs imposed by this study. However, it is interesting to note that for the reproduction of both models, information that could drastically increase the reproducibility of the models was omitted. This increases the difficulty of the verification process and the development of novel methods building upon these models. It is however important to disclose that the authors of the 2009 Recker et al. model were very open to discussion about their model, and that, due to already having started the reproduction of the 2009 Recker et al. model, we did not contact the authors of the 2016 Ferguson et al. model. We argue that, in the spirit of open science, public or on-demand availability of technical artefacts, such as code and software, accompanying published research could be of great benefit to the scientific community, increasing the ease of reproducibility and verification and allowing researchers to more easily build upon existing work.

To suit the needs of this study, the 2009 Recker et al. model was extended with explicit compartments for vaccinated individuals. This allowed for the evaluation of the impact of various vaccination strategies, but also the cost of implementing those strategies. There is not that much information to be found on the approximate true cost of vaccinating a single individual against DENV or testing them for antibodies. Using the values used in other studies [204, 167], in this study the cost of vaccination is set to 20 and the cost of screening for antibodies is set to 10. However, it is important to note that these costs will differ between settings depending on the vaccine that is administered, the test that is used, the existing medical infrastructure, the availability of medical personnel, and various other factors. The 2009 Recker et al. model was also extended with support for age-heterogeneity. Within this study, this was achieved using a contact matrix. Such a matrix is incorporated into the model and weighs the relative influence of contacts between individuals of various age groups on the overall disease dynamics. It is important to disclose that such contact matrix based approaches are most suited to modeling diseases in which the virus is directly spread from individual to individual without the need for a vector intermediary. However, after careful consideration it was decided that such a model is also sound for simulating dengue epidemics depending on the values chosen for the entries of the matrix. The reasoning behind this is that wherever individuals meet, they can infect others via the mosquito vectors that are present in the area. The values in the matrix essentially weigh the importance of contacts between individuals in the force of infection. Hence, the decision was made to use values for the entries where the contacts between age groups are of relatively low importance but still play a role in the overall disease dynamics. Based on the considerations made by this study and the survey of relevant literature on vaccination against DENV, we hypothesize that the age at which individuals are vaccinated and age-related ADE factors will be of greater importance in the overall disease dynamics than the contacts between age groups. This is an interesting avenue for future work that can be studied using minimal adaptations to the model presented in this study.

Moving on, one of this dissertation's main objectives was to study and gain new insights into the MOMAB framework, specifically within the context of Pareto-front identification (PFI), and to contribute to its development by proposing a completely novel PFI MOMAB algorithm. This objective arose due to the adaptability of the MOMAB framework for identifying optimal prevention strategies for various infectious diseases. This independence of the underlying process highlights one of the MOMAB framework's key strengths: general applicability. When modeling infectious diseases, and when creating models in general, many setting-specific assumptions about the underlying real-world processes are made. While these assumptions are necessary, as models cannot be developed without them, they also serve the important function of making our assumptions explicit, thereby encouraging rigorous reasoning. In the introduction, we claimed that the bandit framework works independently of the assumptions underlying the models, relying solely on the stochastic outputs to learn the optimal strategies (arms). This characteristic enables the MOMAB framework to identify optimal policies across a wide range of models, making it a highly valuable area for study and the generation of new insights. However, it is important to note that, whenever Bayesian MOMAB algorithms are employed, the underlying process might influence the prior that is selected. In general, an uninformative prior can be selected to conserve the model-independent nature of the MOMAB framework. Whenever a prior is chosen which incorporates knowledge about the underlying process, this model-independence property is violated.

In the process of studying the MOMAB framework, specifically with respect to the PFI setting, we proposed a number of performance metrics to quantify the quality of the recommendations made by PFI MOMAB algorithms. After using these metrics throughout various experiments, the Bernoulli and hypervolume metrics warrant some further discussion. Firstly, it was observed that the Bernoulli metric is a very strict metric: only perfect recommendations are recompensed. This means that if a recommendation contains all but one of the Pareto optimal arms, the Bernoulli metric will be equal to 0. Similarly, if the recommendation contains all Pareto optimal arms and some additional suboptimal arm, the Bernoulli metric will also be equal to 0. From an intuitive point of view, recommendations that are far from perfect receive the same score as recommendations that are nearly perfect. Consequently, the Bernoulli metric does not provide that much insight into the acquisition process of the Pareto front by the PFI algorithm. However, we argue that the Bernoulli metric is a strong indicator of the algorithms ability to identify the Pareto optimal arms, that is particularly useful when the goal is to perfectly identify the Pareto front. During the processing of the results from the experiments conducted using the final DENV MOMAB setting, a minor flaw with the design of the hypervolume metric was observed. By design and due to the choice of the reference point, the highest attainable hypervolume is associated with recommendations that include the entire subset of optimal arms. This means that very broad recommendations including large numbers of arms have a very high chance of achieving a high value for the hypervolume metric, while the quality of these recommendations might be questionable. Hence, in isolation, the hypervolume metric might not be as suitable for the evaluation of MOMAB PFI algorithms that start by considering all arms as Pareto optimal like the Annealing Pareto algorithm. This also highlights that, within the MOMAB PFI setting, it is very useful to have multiple performance metrics, as only a single metric might not tell the entire story. Further studies into quality metrics for MOMAB algorithms might prove an interesting avenue for future work.

An interesting observation that came from this dissertation's literature review, and is confirmed by [140], is the lack of literature on pure MOMAB algorithms for Pareto front identification (PFI) where the preference of the user is not taken into account. It is, however, important to note that within the space of multi-objective optimization problems, Bayesian Optimization algorithms

exist that incorporate MABs for their acquisition function [140]. These Bayesian Optimization algorithms can be used to solve MOMAB settings like the one we propose but are beyond the scope of this dissertation. The PFI setting is essentially the multi-objective extension of the best-arm identification (BAI) setting for single-objective MABs discussed in Section 5.2. This setting is of particular interest since it corresponds exactly to the objective of this dissertation: to provide the decision maker with the complete set of possible trade-offs between different optimal vaccination strategies, without making any assumptions about their preferences.

Following this observation, the decision was made to study the relationships between MAB algorithms for the single-objective setting and the multi-objective setting, as well as MAB algorithms for the regret minimization setting and the best-arm identification setting. In Section 8.2, four different variants of preference-unaware MOMAB algorithms were presented. All algorithms presented in Section 8.2 are algorithms for regret minimization, all aiming to minimize their cumulative and unfairness regrets while navigating the exploration-exploitation trade-off in their own unique way: UCB, Thompson sampling, Knowledge gradient, and an Annealing approach. The first main goal of this dissertation does not align completely with the regret minimization setting that these algorithms were designed for. Instead, it is an instance of the PFI setting for MOMABs. However, the preference-unaware nature of the MOMAB algorithms presented in Section 8.2 does align with the needs of this study. Furthermore, when considering single-objective MAB algorithms, it is regularly the case that best-arm identification algorithms and regret minimization algorithms are closely related to one another, where the former has an additional emphasis on exploration. Some examples are the UCB algorithm with a larger value for the exploration-regulating hyperparameter κ, and the close relation between Thompson sampling and Top-two Thompson Sampling. These facts, combined with the lack of pure MOMAB algorithms for Pareto front identification, identified in Chapter 8 and confirmed by Reymond in [140], made the algorithms presented in Section 8.2 very valuable to examine within the context of this dissertation: They might be suitable for adaptation to the MOMAB Pareto front identification setting.

Only the Pareto variants of each of the algorithms were extended to the PFI setting. By design, the scalarized variants of the MOMAB algorithms do not explore the multi-objective space directly. Instead, they reduce the multi-objective space to single scalar value for each arm based on a scalarization function. From this scalar value for each of the arms, a relative order between the arms can always be established, where a single arm is considered to be the optimal arm. Hence, the scalarized variants of the MOMAB algorithms will always recommend a single arm when extended to the PFI setting. Within the context of this dissertation, this corresponds to the algorithm only identifying a single vaccination strategy which is optimal under the function used for scalarization. This clearly goes against the goal of presenting the decision makers with all possible trade-offs. Furthermore, when using such scalarized algorithms, the scalarization function(s) employed by the algorithm need to be defined a priori, thus necessitating assumptions to be made about the possible preferences of the decision maker. In this dissertation, the goal is to avoid making assumptions about the utility of the decision maker. Due to these reasons, the scalarized variants of the MOMAB algorithms were not suitable for extension to the PFI setting within the context of this study. Following the extension of these algorithms to the PFI setting, experiments were conducted for which the findings highlighted the potential of adapting existing MOMAB algorithms to the PFI setting. Future work could explore further adaptations and enhancements to these algorithms to improve their performance. Additionally, investigating the impact of different experimental setups, such as varying the standard deviation of the reward distributions could provide deeper insights into the robustness and applicability of these algorithms in real-world scenarios. Furthermore, the study of relationships between various

(MO)MAB algorithms for both the regret minimization and the best-arm identification/PFI setting provides us with useful insights into how existing MOMAB algorithms can be extended to fit different settings and which algorithms might be suitable for use within the PFI setting. Interesting future research might include adapting other MOMAB algorithms to the PFI setting or extending existing MAB algorithms for the BAI setting, like Successive Rejects, to the PFI setting.

One of the most major contributions made by this dissertation is the proposal of a completely novel PFI MOMAB algorithm: Top-two Pareto Fronts Thompson Sampling (TTPFTS). This algorithm is essentially a preference-unaware MO extension of the Top-two Thompson Sampling algorithm for the BAI setting. When following the TTPFTS strategy, the algorithm either pulls an arm from the subset of arms it considers to be Pareto-optimal, or it temporarily ignores these arms and pulls an arm from the subset of the most optimal suboptimal arms. We argue that, even though this is a relatively simple algorithm it seems to capture the complexity of the problem. When introducing the results obtained for the TTPFTS algorithm, it can be observed that it significantly outperforms other algorithms within the single specific experimental setup used in this dissertation. Focusing on the Bernoulli performance metric shown in Figure 10.10a, the TTPFTS algorithm's performance increases rapidly, consistently providing perfect recommendations of the true set of Pareto optimal arms more than 90% of the time after only 3000 arm pulls. This represents a substantial improvement over the 80% success rate achieved by the best-performing alternative algorithm, Pareto Thompson Sampling, after 5000 arm pulls. Furthermore, the TTPFTS algorithm also surpasses the other tested algorithms concerning the Jaccard similarity metric and the hypervolume metric. These results suggest that in this setting, the TTPFTS algorithm is more effective and efficient at identifying the Pareto front than the other algorithms.

However, it is important to note that these findings are based on a single experimental setting. To make more robust claims about the TTPFTS algorithm's performance, additional experiments should be conducted with varying numbers of arms, objectives, relative distances between arms, and different reward distributions and variabilities for the arms. Nonetheless, these results are highly encouraging for the potential performance of the novel TTPFTS algorithm proposed in this dissertation.

We also argue that a number of interesting opportunities for future work present themselves based on the proposal of the TTPFTS algorithm. For example, it can be studied whether the binary notion of either pulling an arm from the estimated Pareto front or the most optimal suboptimal arms, could be made more subtle by for example not considering only part of the estimated Pareto front. Another interesting avenue for future work would be to look into potential guarantees for the performance of TTPFTS.

Currently, there is support for using either a Beta distribution or Normal-inverse-gamma distribution as a prior in the Python implementation of the TTPFTS algorithm. The reason for this is that the code for sampling from the posterior distribution and updating the posteriors upon observing a new reward is different based on which prior is selected. This means that in its current state, TTPFTS supports settings where the arms' reward distributions are either multivariate Bernoulli distributions or multivariate normal distributions with unknown mean and variance. However, in the future, the Python implementation can be changed to allow a third party to implement these methods based on their prior of choice.

Finally, a number of aspects related to the final DENV MOMAB setting require additional discussion. The first thing to note is the difference in relative performance between the different

PFI algorithms in the experiments conducted to compare them to TTPFTS in Section 10.3, and the final experiments using the DENV MOMAB setting. This might be due to the different amounts of hyperparamter tuning that was employed for both of these experiments. For the experiments in Section 10.3, no hyperparameter tuning was conducted. The exploration-regulating hyperparamter for Pareto UCB1 was left at 1 to be in line with its original publication, while the ρ parameter of TTPFTS was left at 0,5 to be in line with the value that is most often employed in Top-two Thompson Sampling. Meanwhile, for the final experiments using the novel DENV MOMAB setting, extensive hyperparameter tuning was employed to ensure the optimal performance for each of the algorithms. It is important to note that in real studies examining the optimality of various preventive strategies based on a computationally expensive stochastic model, such extensive hyperparamter tuning is not feasible. Hence, it is important for studies like this one to experiment with various hyperparameters and for future studies to establish baseline values for the various hyperparameters of the MOMAB PFI algorithms. It can also be valuable for a future study to explore how the PFI MOMAB algorithms perform in different settings over different ranges of hyperparamter values.

The DENV MOMAB setting presented in Chapter 11 uses probabilistic rewards based on deterministic simulations by using the outcome of the simulation as a mean and adding a standard deviation. Within this study, this serves as a basic proxy for possibly more complex and computationally expensive stochastic models that might be used in future work. It is important to disclose that this pseudo-stochastic output of the epidemiological model is not the most realistic, but since the true means are known, the ground truth for the Pareto front is also known and the performance of the PFI MOMAB algorithms can be evaluated. One way in which future studies could improve upon the research presented here, is by replacing the dengue epidemic model we used with an inherently stochastic individual-based model.

Part VI

Conclusion

One of the primary objectives of this dissertation was to explore the efficacy of multi-objective multi-armed bandit (MOMAB) algorithms for the Pareto-front identification (PFI) setting in identifying optimal vaccine allocation strategies for mitigating dengue epidemics.

To this end, a central research question was addressed: "Can MOMAB algorithms for the PFI setting be used to identify the subset of optimal vaccination strategies for the mitigation of dengue epidemics, and the trade-offs between them, in a sample efficient manner within the allocated budget, based on the output of computationally expensive stochastic simulations?". This question guided significant parts of the research process, from the development of epidemiological models to the adaptation of MOMAB algorithms.

To tackle this question, we began by delving into the existing literature on dengue epidemiology and modeling, as well as multi-objective reinforcement learning and optimization. This foundational knowledge led to the reproduction and extension of two dengue epidemic models, specifically the 2016 Ferguson et al. model and the 2009 Recker et al. model. Using the insights gained from the more recent Ferguson et al. model, the Recker et al. model was enhanced to support vaccination strategies and age-heterogeneity, based on the state-of-the-art idea that vaccination can be modeled as a silent infection.

With the extended model in place, we introduced Gaussian noise to emulate stochastic behavior, aligning the model outputs with the stochastic reward functions used by MOMABs. Subsequently, in line with the second major objective, we reproduced and adapted several MOMAB algorithms to the PFI setting and proposed three performance metrics: the Bernoulli metric, the Jaccard similarity metric, and the Hypervolume metric. These metrics were essential for evaluating the quality of the recommendations made by the PFI MOMAB algorithms.

The experimental phase involved testing a total of 53 vaccination strategies using the extended Recker et al. model. By combining the use of stochastic simulations and multi-objective multi-armed bandits, we aimed to efficiently pinpoint the subset of vaccination strategies that balance minimizing both the medical burden and monetary costs. Each strategy's performance was assessed with respect to those two objectives and used to create multivariate reward distributions. Five MOMAB algorithms, including a completely novel Top-two Pareto Fronts Thompson Sampling (TTPFTS) algorithm, were evaluated with a limited budget of 30,000 arm pulls each, across 100 experimental repetitions. The Uniform Sampling algorithm, which is currently used in literature for the evaluation of preventive strategies, was used as a performance baseline, also being granted a budget of 30,000 arm pulls.

The results revealed that among the algorithms tested, Pareto UCB1 consistently performed best in terms of efficiency and stability in identifying the subset of optimal vaccination strategies. Pareto Thompson Sampling and TTPFTS also performed excellently, even though they exhibited slight variability in distinguishing closely situated optimal solutions. This highlighted the necessity of utilizing multiple metrics for a comprehensive evaluation of algorithm performance. We argue that this issue is not as significant, as the trade-offs between closely situated optimal solutions are minimal, and their differences in effects are also very small. However, it is still preferable for the algorithms to identify these closely related optimal solutions, as one may be preferred by public health officials and decision-makers. Most importantly, the results also showed that Pareto UCB1, Pareto Thompson Sampling, and TTPFTS outperformed Uniform Sampling by a large margin with respect to all three considered performance metrics. This indicates the ability of MOMAB algorithms for the PFI setting to identify the subset of optimal vaccination strategies for the mitigation of dengue epidemics in a more sample-efficient manner than is currently used in most studies. As minimizing the number of required model evaluations

significantly reduces the total time needed to assess a set of preventive strategies, we argue that the use of PFI MOMAB algorithms can accelerate the decision making process of selecting optimal vaccination strategies for the mitigation of dengue epidemics. Given their efficient use of the evaluation budget, PFI MOMAB algorithms can also make the use of individual-based models feasible in studies where it might otherwise be computationally impractical, and free up computational resources in studies that already utilize individual-based models, allowing researchers to explore a broader set of model scenarios.

In previous studies [112, 111, 110], single-objective multi-armed bandits have already been used to efficiently identify optimal preventive strategies for the mitigation of influenza epidemics. Both the extension to multi-objective setting and the new application to dengue epidemics presented in this dissertation open intriguing avenues for new insights. Firstly, in real-life scenarios, decision-makers typically need to consider several potentially conflicting objectives. Therefore, it is a very encouraging result that in this dissertation, we were able to apply the multi-armed bandit framework in a multi-objective epidemiological setting. Although this dissertation focuses on two objectives, the proposed framework operates independently of the number of objectives, allowing for evaluation with respect to additional objectives as necessary. This could be a promising avenue for future work, as the scalability of this technique to more than two objectives has not yet been investigated. Furthermore, the fact that we were able to directly apply the (multi-objective) multi-armed bandit framework to the unexplored epidemiological setting of dengue epidemics highlights one of its key strengths: general applicability. When modeling infectious diseases like dengue, and when creating models in general, many setting-specific assumptions about the underlying real-world processes are made. While these assumptions are necessary, as models cannot be developed without them, they also serve the important function of making our assumptions explicit, thereby encouraging rigorous reasoning. However, the bandit framework works independently of the assumptions underlying the models, relying solely on the stochastic outputs to learn the optimal strategies. this dissertation underlines this characteristic of the (MO)MAB framework, further indicating its potential to identify optimal policies across a wide range of models and settings.

Due to the inherent general applicability of the (MO)MAB framework, a secondary objective of this dissertation was to gain new insights into this framework and contribute to its development. This objective was achieved after an observation was made from carefully examining the existing literature on PFI MOMAB algorithms. From this literature review, we identified that there exist very little pure PFI MOMAB algorithms where the preference of the user is not taken into account. This observation was also confirmed by Reymond in [140]. It is, however, important to note that within the space of multi-objective optimization problems, Bayesian Optimization algorithms exist that incorporate MABs for their acquisition function [140]. These Bayesian Optimization algorithms can be used to solve MOMAB settings like the one we propose but are beyond the scope of this dissertation. However, this setting is of particular interest since it corresponds exactly to the objective of this dissertation: to provide the decision maker with the complete set of possible trade-offs between different optimal vaccination strategies, without making any assumptions about their preferences. This prompted us to study the relations between MAB algorithms for the single-objective setting and the multi-objective setting, and the relations between the regret minimization setting and the best-arm identification setting, for which the PFI setting is the multi-objective extension. After meticulously examining the relations between these settings, we identified the possibility of extending existing best-arm identification algorithms to the PFI setting. To further prove this concept, we introduced a completely novel preference-unaware PFI MOMAB algorithm inspired by Top-two Thompson Sampling: Top-two Pareto Fronts Thompson Sampling (TTPFTS). The performance of TTPFTS was eval-

uated through two separate experiments, both demonstrating its ability to efficiently identify the complete subset of optimal arms. However, to make more robust claims about the TTPFTS algorithm's performance, additional experiments should be conducted with varying numbers of arms, objectives, relative distances between arms, and different reward distributions and variabilities for the arms. Nonetheless, the observed results are highly encouraging for the potential performance of the novel TTPFTS algorithm proposed in this dissertation and we argue that this algorithm is a very valuable contribution to the field.

In summary, this dissertation succeeded in both its primary objectives. We demonstrated that MOMAB algorithms adapted to the PFI setting, are effective in identifying optimal vaccination strategies for dengue epidemic mitigation, while also introducing and evaluating a completely novel PFI MOMAB algorithm. The insights gained from this research contribute to optimizing vaccination strategies, offering a robust approach to decision-making in the face of computationally expensive simulations and several conflicting objectives. By efficiently using the evaluation budget and accurately identifying the true set of optimal strategies, this study provides a valuable framework for addressing complex public health challenges. The findings underscore the potential of MOMAB algorithms to enhance the strategic deployment of vaccination programs, ultimately contributing to better management and control of dengue epidemics.

Bibliography

[1] The bugs book: A practical introduction to bayesian analysis. *CHANCE*, 26:56 – 59, 2013.

[2] Nicole L. Achee, Fred Gould, T. Alex Perkins, Robert C. Reiner, Jr., Amy C. Morrison, Scott A. Ritchie, Duane J. Gubler, Remy Teyssou, and Thomas W. Scott. A critical assessment of vector control for dengue prevention. *PLOS Neglected Tropical Diseases*, 9(5):1–19, 05 2015.

[3] B. Adams, E. C. Holmes, C. Zhang, M. P. Mammen, S. Nimmannitya, S. Kalayanarooj, and M. Boots. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in bangkok. *Proceedings of the National Academy of Sciences*, 103(38):14234–14239, 2006.

[4] Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem, 2012.

[5] Maíra Aguiar, Vizda Anam, Konstantin B. Blyuss, Carlo Delfin S. Estadilla, Bruno V. Guerrero, Damián Knopoff, Bob W. Kooi, Akhil Kumar Srivastav, Vanessa Steindorf, and Nico Stollenwerk. Mathematical models for dengue fever epidemiology: A 10-year systematic review. *Physics of Life Reviews*, 40:65 – 92, 2022.

[6] Maira Aguiar and Nico Stollenwerk. Mathematical models of dengue fever epidemiology: multi-strain dynamics, immunological aspects associated to disease severity and vaccines. *Communication in Biomathematical Sciences*, 1(1):1–12, Dec. 2017.

[7] Maíra Aguiar, Sebastien Ballesteros, Bob W. Kooi, and Nico Stollenwerk. 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. *Journal of Theoretical Biology*, 289:181–196, 2011.

[8] Maíra Aguiar, Bob W. Kooi, Filipe Rocha, Peyman Ghaffari, and Nico Stollenwerk. How much complexity is needed to describe the fluctuations observed in dengue hemorrhagic fever incidence data? *Ecological Complexity*, 16:31–40, 2013. Modelling ecological processes: proceedings of MATE 2011.

[9] Maíra Aguiar and Nico Stollenwerk. Dengvaxia Efficacy Dependency on Serostatus: A Closer Look at More Recent Data. *Clinical Infectious Diseases*, 66(4):641–642, 10 2017.

[10] Maíra Aguiar and Nico Stollenwerk. Dengvaxia: age as surrogate for serostatus. *The Lancet Infectious Diseases*, 18(3):245, 2018.

[11] Maíra Aguiar and Nico Stollenwerk. The impact of serotype cross-protection on vaccine trials: Denvax as a case study. *Vaccines*, 8(4), 2020.

[12] Maíra Aguiar, Nico Stollenwerk, and Scott B. Halstead. The impact of the newly licensed dengue vaccine in endemic countries. *PLOS Neglected Tropical Diseases*, 10(12):1–23, 12 2016.

[13] Maíra Aguiar, Nico Stollenwerk, and Scott B. Halstead. The risks behind dengvaxia recommendation. *The Lancet Infectious Diseases*, 16(8):882–883, 2016.

[14] Aguiar, M., Kooi, B.W., Martins, J., and Stollenwerk, N. Scaling of stochasticity in dengue hemorrhagic fever epidemics. *Math. Model. Nat. Phenom.*, 7(3):1–11, 2012.

[15] Aguiar, Maíra, Kooi, Bob, and Stollenwerk, Nico. Epidemiology of dengue fever: A model with temporary cross-immunity and possible secondary infection shows bifurcations and chaotic behaviour in wide parameter regions. *Math. Model. Nat. Phenom.*, 3(4):48–70, 2008.

[16] Edward J. Allen, Linda J. S. Allen, Armando Arciniega, and Priscilla E. Greenwood. Construction of equivalent stochastic differential equation models. *Stochastic Analysis and Applications*, 26:274 – 297, 2008.

[17] L. Alonso-Palomares, M. Moreno-García, H. Lanz-Mendoza, and M. Salazar. Molecular basis for arbovirus transmission by aedes aegypti mosquitoes. *Intervirology*, 61:255 – 264, 2019.

[18] David F. Anderson. Incorporating postleap checks in tau-leaping. *The Journal of chemical physics*, 128 5:054103, 2007.

[19] David F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of chemical physics*, 127 21:214107, 2007.

[20] N. Anggriani, H. Tasman, M.Z. Ndii, A.K. Supriatna, E. Soewono, and E Siregar. The effect of reinfection with the same serotype on dengue transmission dynamics. *Applied Mathematics and Computation*, 349:62–80, 2019.

[21] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *Annual Conference Computational Learning Theory*, 2010.

[22] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002.

[23] Frank G. Ball, Tom Britton, Thomas A. House, Valerie Isham, Denis Mollison, Lorenzo Pellis, and Gianpaolo Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–7, 2015.

[24] Shahera Banu, A. Choudhury, and S. Tong. Dengue: Emergence, determinants and climate change. pages 237–248, 2016.

[25] Nicole E. Basta, Dennis L. Chao, M. Elizabeth Halloran, Laura Matrajt, and Ira M. Longini. Strategies for pandemic and seasonal influenza vaccination of schoolchildren in the united states. *American Journal of Epidemiology*, 170:679 – 686, 2009.

[26] Lora Billings, Ira B. Schwartz, Leah B. Shaw, Marie McCrary, Donald S. Burke, and Derek A.T. Cummings. Instabilities in multiserotype disease models with antibody-dependent enhancement. *Journal of Theoretical Biology*, 246(1):18–27, 2007.

[27] Shibadas Biswal et al. Efficacy of a tetravalent dengue vaccine in healthy children aged 4–16 years: a randomised, placebo-controlled, phase 3 trial. *The Lancet*, 395(10234):1423–1433, 2020.

[28] Shibadas Biswal, Humberto Reynales, Xavier Saez-Llorens, Pio Lopez, Charissa Borja-Tabora, Pope Kosalaraksa, Chukiat Sirivichayakul, Veerachai Watanaveeradej, Luis Rivera, Felix Espinoza, LakKumar Fernando, Reynaldo Dietze, Kleber Luz, Rivaldo Venâncio da Cunha, José Jimeno, Eduardo López-Medina, Astrid Borkowski, Manja Brose, Martina Rauscher, Inge LeFevre, Svetlana Bizjajeva, Lulu Bravo, and Derek Wallace. Efficacy of a tetravalent dengue vaccine in healthy children and adolescents. *New England Journal of Medicine*, 381(21):2009–2019, 2019. PMID: 31693803.

[29] Shibadas Biswal, Humberto Reynales, Xavier Saez-Llorens, Pio Lopez, Charissa Borja-Tabora, Pope Kosalaraksa, Chukiat Sirivichayakul, Veerachai Watanaveeradej, Luis Rivera, Felix Espinoza, LakKumar Fernando, Reynaldo Dietze, Kleber Luz, Rivaldo Venâncio da Cunha, José Jimeno, Eduardo López-Medina, Astrid Borkowski, Manja Brose, Martina Rauscher, Inge LeFevre, Svetlana Bizjajeva, Lulu Bravo, and Derek Wallace. Efficacy of a tetravalent dengue vaccine in healthy children and adolescents. *New England Journal of Medicine*, 381(21):2009–2019, 2019. PMID: 31693803.

[30] Wolfgang Bock and Yashika Jayathunga. Optimal control of a multi-patch dengue model under the influence of wolbachia bacterium. *Mathematical Biosciences*, 315:108219, 2019.

[31] Kobporn Boonnak, Kaitlyn M. Dambach, Gina C. Donofrio, Boonrat Tassaneetrithep, and Mary A. Marovich. Cell type specificity and host genetic polymorphisms influence antibody-dependent enhancement of dengue virus infection. *Journal of Virology*, 85(4):1671–1683, 2011.

[32] Oliver J. Brady, Peter W. Gething, Samir Bhatt, Jane P. Messina, John S. Brownstein, Anne G. Hoen, Catherine L. Moyes, Andrew W. Farlow, Thomas W. Scott, and Simon I. Hay. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLOS Neglected Tropical Diseases*, 6(8):1–15, 08 2012.

[33] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International Conference on Algorithmic Learning Theory*, 2009.

[34] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theor. Comput. Sci.*, 412:1832–1852, 2011.

[35] D. S. Burke, A. Nisalak, D. E. Johnson, and R. M. Scott. A prospective study of dengue infections in bangkok. *The American Journal of Tropical Medicine and Hygiene*, 38(1):172–180, Jan 1988.

[36] Anne Tuiskunen Bäck and Åke Lundkvist. Dengue viruses – an overview. *Infection Ecology & Epidemiology*, 3(1):19839, 2013. PMID: 24003364.

[37] Xiaodong Cai. Exact stochastic simulation of coupled chemical reactions with delays. *The Journal of chemical physics*, 126 12:124108, 2007.

[38] Yang Cao, Daniel T. Gillespie, and Linda R. Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of chemical physics*, 124 4:044109, 2006.

[39] Maria Rosario Capeding, Ngoc Huu Tran, Sri Rezeki S Hadinegoro, Hussain Imam HJ Muhammad Ismail, Tawee Chotpitayasunondh, Mary Noreen Chua, Chan Quang Luong, Kusnandi Rusmil, Dewa Nyoman Wirawan, Revathy Nallusamy, Punnee Pitisut-

tithum, Usa Thisyakorn, In-Kyu Yoon, Diane van der Vliet, Edith Langevin, Thelma Laot, Yanee Hutagalung, Carina Frago, Mark Boaz, T Anh Wartel, Nadia G Tornieporth, Melanie Saville, and Alain Bouckenooghe. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *The Lancet*, 384(9951):1358–1365, 2014.

[40] Clara Champagne and Bernard Cazelles. Comparison of stochastic and deterministic frameworks in dengue modelling. *Mathematical Biosciences*, 310:1–12, 2019.

[41] Dennis L. Chao, M. Elizabeth Halloran, Valerie Obenchain, Ira M. Longini, and Angela R. McLean. Flute, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6, 2010.

[42] Dennis L. Chao, Scott B. Halstead, M. Elizabeth Halloran, and Ira M. Longini. Controlling dengue with vaccines in thailand. *PLoS Neglected Tropical Diseases*, 6, 2012.

[43] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In *Neural Information Processing Systems*, 2011.

[44] Hannah E. Clapham, Derek A. T. Cummings, and Michael A. Johansson. Immune status alters the probability of apparent illness due to dengue virus infection: Evidence from a pooled analysis across multiple cohort and cluster studies. *PLOS Neglected Tropical Diseases*, 11(9):1–12, 09 2017.

[45] Laurent Coudeville and Geoff P. Garnett. Transmission dynamics of the four dengue serotypes in southern vietnam and the potential impact of vaccination. *PLOS ONE*, 7(12):1–11, 12 2012.

[46] Derek Cummings, Ira Schwartz, Lora Billings, Leah Shaw, and Samuel Burke. Dynamic effects of antibody-dependent enhancement on the fitness of viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 102:15259–64, 11 2005.

[47] Gabriel de Oliveira Ramos, Bruno C. da Silva, Roxana Radulescu, Ana L. C. Bazzan, and Ann Nowé. Toll-based reinforcement learning for efficient equilibria in route choice. *Knowl. Eng. Rev.*, 35:e8, 2020.

[48] Wanwisa Dejnirattisai, Amonrat Jumnainsong, Naruthai Onsirisakul, Patricia Fitton, Sirijitt Vasanawathana, Wannee Limpitikul, Chunya Puttikhunt, Carolyn Edwards, Thaneeya Duangchinda, Sunpetchuda Supasa, Kriangkrai Chawansuntati, Prida Malasit, Juthathip Mongkolsapaya, and Gavin Screaton. Cross-reacting antibodies enhance dengue virus infection in humans. *Science*, 328(5979):745–748, 2010.

[49] Sheng-Qun Deng, Xian Yang, Yong Wei, Jia-Ting Chen, Xiao-Jun Wang, and Hong-Juan Peng. A review on dengue vaccine development. *Vaccines*, 8(1), 2020.

[50] Odo Diekmann, Hans Heesterbeek, and Tom Britton. Mathematical tools for understanding infectious disease dynamics. 2012.

[51] Madalina M. Drugan and Ann Nowe. Designing multi-objective multi-armed bandits algorithms: A study. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2013.

[52] Mădălina M. Drugan. Covariance matrix adaptation for multiobjective multiarmed bandits. *IEEE Transactions on Neural Networks and Learning Systems*, 30(8):2493–2502, 2019.

[53] Mădălina M Drugan, Ann Nowé, and Bernard Manderick. Pareto upper confidence bounds algorithms: An empirical study. In *2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pages 1–8, 2014.

[54] Gabriele Eichfelder. Adaptive scalarization methods in multiobjective optimization. In *Vector Optimization*, 2008.

[55] Stephen Eubank, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, and Nan Wang. Structure of social contact networks and their impact on epidemics. In *Discrete Methods in Epidemiology*, 2004.

[56] Eyal Even-Dar, Shie Mannor, and Y. Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *J. Mach. Learn. Res.*, 7:1079–1105, 2006.

[57] Jorge A. Falcón-Lezama, Ruth A. Martínez-Vega, Pablo A. Kuri-Morales, José Ramos-Castañeda, and Ben Adams. Day-to-day population movement and the management of dengue epidemics. *Bulletin of Mathematical Biology*, 78(10):2011–2033, Oct 2016.

[58] Neil Ferguson, Roy Anderson, and Sunetra Gupta. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proceedings of the National Academy of Sciences*, 96(2):790–794, 1999.

[59] Neil M. Ferguson, Derek A. T. Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Cooper Meeyai, Sopon Iamsirithaworn, and Donald S. Burke. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437:209–214, 2005.

[60] Neil M. Ferguson, Isabel Rodríguez-Barraquer, Ilaria Dorigatti, Luis Mier-y Teran-Romero, Daniel J. Laydon, and Derek A. T. Cummings. Supplementary materials for benefits and risks of the sanofi-pasteur dengue vaccine: Modeling optimal deployment, September 2016. Science 353, 1033 (2016). DOI: 10.1126/science.aaf9590.

[61] Neil M. Ferguson, Isabel Rodríguez-Barraquer, Ilaria Dorigatti, Luis Mier y Teran-Romero, Daniel J. Laydon, and Derek A. T. Cummings. Benefits and risks of the sanofi-pasteur dengue vaccine: Modeling optimal deployment. *Science*, 353(6303):1033–1036, 2016.

[62] Max Souza Filipe Rocha, Maíra Aguiar and Nico Stollenwerk. Time-scale separation and centre manifold analysis describing vector-borne disease dynamics. *International Journal of Computer Mathematics*, 90(10):2105–2125, 2013.

[63] Diana B. Fischer, Yale Arbo, and Scott B. Halstead. Observations related to pathogenesis of dengue hemorrhagic fever. v. examination of agspecific sequential infection rates using a mathematical model. *The Yale Journal of Biology and Medicine*, 42:329 – 349, 1970.

[64] Laura Fumanelli, Marco Ajelli, Stefano Merler, Neil M. Ferguson, and Simon Cauchemez. Model-based comprehensive analysis of school closure policies for mitigating influenza epidemics and pandemics. *PLoS Computational Biology*, 12, 2016.

[65] Timothy C. Germann, Kai Kadau, Ira M. Longini, and Catherine Macken. Mitigation strategies for pandemic influenza in the united states. *Proceedings of the National Academy of Sciences of the United States of America*, 103 15:5935–40, 2006.

[66] Jayanta Kumar Ghosh, Uttam Ghosh, and Susmita Sarkar. Qualitative analysis and optimal control of a two-strain dengue model with its co-infections. *International Journal of Applied and Computational Mathematics*, 6(6):161, Oct 2020.

[67] Robert Gibbons, Siripen Kalanarooj, Richard Jarman, Ananda Nisalak, David Vaughn, Timothy Endy, Mammen Mammen, and Anon Srikiatkhachorn. Analysis of repeat hospital admissions for dengue to estimate the frequency of third or fourth dengue infections resulting in admissions and dengue hemorrhagic fever, and serotype sequences. *The American journal of tropical medicine and hygiene*, 77:910–3, 12 2007.

[68] Michael A. Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104:1876–1889, 2000.

[69] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977.

[70] Robert J. Glass, Laura M. Glass, Walter E. Beyeler, and H. Jason Min. Targeted social distancing designs for pandemic influenza. *Emerging Infectious Diseases*, 12:1671 – 1681, 2006.

[71] N.L. González Morales, M. Núñez-López, J. Ramos-Castañeda, and J.X. Velasco-Hernández. Transmission dynamics of two dengue serotypes with vaccination scenarios. *Mathematical Biosciences*, 287:54–71, 2017. 50th Anniversary Issue.

[72] Aditya Gopalan, Shie Mannor, and Y. Mansour. Thompson sampling for complex online problems. In *International Conference on Machine Learning*, 2013.

[73] Adriana O Guilarde, Marilia D Turchi, Joao Bosco Siqueira Jr, Valeria C. R Feres, Benigno Rocha, Jose E Levi, Vanda A. U. F Souza, Lucy Santos Vilas Boas, Claudio S Pannuti, and Celina M. T Martelli. Dengue and Dengue Hemorrhagic Fever among Adults: Clinical Outcomes Related to Viremia, Serotypes, and Antibody Response. *The Journal of Infectious Diseases*, 197(6):817–824, 03 2008.

[74] Maria G. Guzman, Duane J. Gubler, Alienys Izquierdo, Eric Martinez, and Scott B. Halstead. Dengue infection. *Nature Reviews Disease Primers*, 2(1):16055, 2016.

[75] Maria G. Guzman, Scott B. Halstead, Harvey Artsob, Philippe Buchy, Jeremy Farrar, Duane J. Gubler, Elizabeth Hunsperger, Axel Kroeger, Harold S. Margolis, Eric Martínez, Michael B. Nathan, Jose Luis Pelegrino, Cameron Simmons, Sutee Yoksan, and Rosanna W. Peeling. Dengue: a continuing global threat. *Nature Reviews Microbiology*, 8(12):S7–S16, 2010.

[76] Maria G. Guzman and Eva Harris. Dengue. *The Lancet*, 385(9966):453–465, 2015.

[77] Maria G. Guzmán and Gustavo Kouri. Dengue: an update. *The Lancet Infectious Diseases*, 2(1):33–42, 2002.

[78] Sri Rezeki Hadinegoro, Jose Luis Arredondo-García, Maria Rosario Capeding, Carmen Deseda, Tawee Chotpitayasunondh, Reynaldo Dietze, H.I. Hj Muhammad Ismail, Humberto Reynales, Kriengsak Limkittikul, Doris Maribel Rivera-Medina, Huu Ngoc Tran, Alain Bouckenooghe, Danaya Chansinghakul, Margarita Cortés, Karen Fanouillere, Remi Forrat, Carina Frago, Sophia Gailhardou, Nicholas Jackson, Fernando Noriega, Eric Plennevaux, T. Anh Wartel, Betzana Zambrano, and Melanie Saville. Efficacy and long-term safety of a dengue vaccine in regions of endemic disease. *New England Journal of Medicine*, 373(13):1195–1206, 2015. PMID: 26214039.

[79] S. B. Halstead. Immune enhancement of viral infection. *Progress in Allergy*, 31:301–364, 1982.

[80] Scott B Halstead. Neutralization and antibody-dependent enhancement of dengue viruses. volume 60 of *Advances in Virus Research*, pages 421–467. Academic Press, 2003.

[81] Scott B. Halstead. Critique of World Health Organization Recommendation of a Dengue Vaccine. *The Journal of Infectious Diseases*, 214(12):1793–1795, 08 2016.

[82] Scott B. Halstead, Leah C. Katzelnick, Philip K. Russell, Lewis Markoff, Maira Aguiar, Leonila R. Dans, and Antonio L. Dans. Ethics of a partially effective dengue vaccine: Lessons from the philippines. *Vaccine*, 38(35):5572–5576, 2020.

[83] llkka Hanski. *Metapopulation Ecology*. Oxford University Press, 03 1999.

[84] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[85] Ross-William S. Hendron and Michael B. Bonsall. The interplay of vaccination and vector control on small dengue networks. *Journal of Theoretical Biology*, 407:349–361, 2016.

[86] Joshua T. Herbeck, John E. Mittler, Geoffrey S. Gottlieb, and James I. Mullins. An hiv epidemic model based on viral load dynamics: Value in assessing empirical trends in hiv virulence and community viral load. *PLoS Computational Biology*, 10, 2014.

[87] Hippocrates. *Hippocrates*. Heinemann; Putnam, London; New York, 1923. With an English translation by W.H.S. Jones and Paul Potter.

[88] K. Hu, C. Thoens, S. Bianco, S. Edlund, M. Davis, J. Douglas, and J.H. Kaufman. The effect of antibody-dependent enhancement, cross immunity, and vector population on the dynamics of dengue fever. *Journal of Theoretical Biology*, 319:62–74, 2013.

[89] Paul Jaccard. Etude de la distribution florale dans une portion des alpes et du jura. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, 37:547–579, 01 1901.

[90] Kevin G. Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. *ArXiv*, abs/1502.07943, 2015.

[91] Junchen Jin and Xiaoliang Ma. A multi-objective agent-based control approach with application in intelligent traffic signal system. *IEEE Trans. Intell. Transp. Syst.*, 20(10):3900–3912, 2019.

[92] K.M. Ariful Kabir and Jun Tanimoto. Cost-efficiency analysis of voluntary vaccination against n-serovar diseases using antibody-dependent enhancement: A game approach. *Journal of Theoretical Biology*, 503:110379, 2020.

[93] Anmol Kagrecha, Jayakrishnan Nair, and Krishna Jagannathan. Constrained regret minimization for multi-criterion multi-armed bandits, 2023.

[94] Parastu Kasaie, Stephen A. Berry, Maunank S. Shah, Eli S. Rosenberg, Karen W. Hoover, Thomas L. Gift, Harrell Chesson, Jeff Pennington, Danielle German, Colin P. Flynn, Chris Beyrer, and David W. Dowdy. Impact of providing preexposure prophylaxis for human immunodeficiency virus at clinics for sexually transmitted infections in baltimore city: An agent-based model. *Sexually Transmitted Diseases*, 45(12), 2018.

[95] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, 2008.

[96] Alexander Kensert, Pieter Libin, Gert Desmet, and Deirdre Cabooter. Deep reinforcement learning for the direct optimization of gradient separations in liquid chromatography. *Journal of Chromatography A*, 1720:464768, 2024.

[97] William Ogilvie Kermack and À. G. Mckendrick. A contribution to the mathematical theory of epidemics. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, 115:700–721, 1927.

[98] Aaron A. King, Matthieu Domenech de Cellès, Felicia M. G. Magpantay, and Pejman Rohani. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806):20150347, May 2015.

[99] Srisakul C. Kliks, Ananda Nisalak, Walter E. Brandt, Larry Wahl, and Donald S. Burke. Antibody-dependent enhancement of dengue virus growth in human monocytes as a risk factor for dengue hemorrhagic fever. *The American Journal of Tropical Medicine and Hygiene*, 40(4):444 – 451, 1989.

[100] Gerhart Knerer, Christine S. M. Currie, and Sally C. Brailsford. Impact of combined vector-control and vaccination strategies on transmission dynamics of dengue fever: a model-based analysis. *Health Care Management Science*, 18(2):205–217, Jun 2015.

[101] Gerhart Knerer, Christine S. M. Currie, and Sally C. Brailsford. The economic impact and cost-effectiveness of combined vector-control and dengue vaccination strategies in thailand: results from a dynamic transmission model. *PLOS Neglected Tropical Diseases*, 14(10):1–32, 10 2020.

[102] Diána Knipl and Seyed M. Moghadas. The potential impact of vaccination on the dynamics of dengue infections. *Bulletin of Mathematical Biology*, 77(12):2212–2230, Dec 2015.

[103] Tadeusz J. Kochel, Douglas M. Watts, Alfonso S. Gozalo, Daniel F. Ewing, Kevin R. Porter, and Kevin L. Russell. Cross-serotype neutralization of dengue virus in aotus nancymae monkeys. *The Journal of Infectious Diseases*, 191(6):1000–1004, March 2005.

[104] Bob W. Kooi, Maíra Aguiar, and Nico Stollenwerk. Bifurcation analysis of a family of multi-strain epidemiology models. *Journal of Computational and Applied Mathematics*, 252:148–158, 2013. Selected papers on Computational and Mathematical Methods in Science and Engineering (CMMSE).

[105] Bob W. Kooi, Maíra Aguiar, and Nico Stollenwerk. Analysis of an asymmetric two-strain dengue model. *Mathematical Biosciences*, 248:128–139, 2014.

[106] Elena Krasheninnikova, Javier García, Roberto Maestre, and Fernando Fernández. Reinforcement learning for pricing strategy optimization in the insurance industry. *Eng. Appl. Artif. Intell.*, 80:8–19, 2019.

[107] Volodymyr Kuleshov and Doina Precup. Algorithms for multi-armed bandit problems. *arXiv preprint arXiv:1402.6028*, 2014.

[108] B. Lefebvre, Rojina Karki, Renaud Misslin, K. Nakhapakorn, É. Daudé, and R. Paul. Importance of public transport networks for reconciling the spatial distribution of dengue and the association of socio-economic factors with dengue risk in bangkok, thailand. *International Journal of Environmental Research and Public Health*, 19, 2022.

[109] Katerina Lepenioti, Minas Pertselakis, Alexandros Bousdekis, Andreas Louca, Fenareti Lampathaki, Dimitris Apostolou, Gregoris Mentzas, and Stathis Anastasiou. Machine learning for predictive and prescriptive analytics of operational data in smart manufacturing. In Sophie Dupuy-Chessa and Henderik A. Proper, editors, *Advanced Information Systems Engineering Workshops - CAiSE 2020 International Workshops, Grenoble, France, June 8-12, 2020, Proceedings*, volume 382 of *Lecture Notes in Business Information Processing*, pages 5–16. Springer, 2020.

[110] Pieter Libin. *Guiding the mitigation of epidemics with reinforcement learning.* PhD thesis, 2020.

[111] Pieter Libin, Timothy Verstraeten, Diederik M. Roijers, Jelena Grujic, Kristof Theys, Philippe Lemey, and Ann Nowé. Bayesian best-arm identification for selecting influenza mitigation strategies, 2018.

[112] Pieter Libin, Timothy Verstraeten, Kristof Theys, Diederik M. Roijers, Peter Vrancx, and Ann Nowé. Efficient evaluation of influenza mitigation strategies using preventive bandits. In Gita Sukthankar and Juan A. Rodríguez-Aguilar, editors, *Autonomous Agents and Multiagent Systems - AAMAS 2017 Workshops, Visionary Papers, São Paulo, Brazil, May 8-12, 2017, Revised Selected Papers*, volume 10643 of *Lecture Notes in Computer Science*, pages 67–85. Springer, 2017.

[113] Pieter J. K. Libin, Arno Moonens, Timothy Verstraeten, Fabian Perez-Sanjines, Niel Hens, Philippe Lemey, and Ann Nowé. Deep reinforcement learning for large-scale epidemic control. In Yuxiao Dong, Georgiana Ifrim, Dunja Mladenic, Craig Saunders, and Sofie Van Hoecke, editors, *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track - European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part V*, volume 12461 of *Lecture Notes in Computer Science*, pages 155–170. Springer, 2020.

[114] Fredrik Liljeros, Christofer R Edling, Luis A. Nunes Amaral, Harry Eugene Stanley, and Yvonne Åberg. The web of human sexual contacts. *Nature*, 411:907–908, 2001.

[115] José Lourenço and Mario Recker. Viral and epidemiological determinants of the invasion dynamics of novel dengue genotypes. *PLOS Neglected Tropical Diseases*, 4(11):1–12, 11 2010.

[116] José Lourenço and Mario Recker. Natural, persistent oscillations in a spatial multi-strain disease system with application to dengue. *PLOS Computational Biology*, 9(10):1–11, 10 2013.

[117] Sandra B. Maier, Xiao Huang, Eduardo Massad, Marcos Amaku, Marcelo N. Burattini, and David Greenhalgh. Analysis of the optimal vaccination age for dengue in brazil with a tetravalent dengue vaccine. *Mathematical Biosciences*, 294:15–32, 2017.

[118] Patrick Mannion, Karl Mason, Sam Devlin, Jim Duggan, and Enda Howley. Multi-objective dynamic dispatch optimisation using multi-agent reinforcement learning: (extended abstract). In *Adaptive Agents and Multi-Agent Systems*, 2016.

[119] Maia Martcheva. *An Introduction to Mathematical Epidemiology.* Texts in Applied Mathematics. Springer New York, NY, New York, NY, 1 edition, 2015.

[120] Renaud Marti, Zhichao Li, T. Catry, E. Roux, M. Mangeas, P. Handschumacher, J. Gaudart, A. Tran, L. Demagistri, J. Faure, J. J. Carvajal, Bruna Drumond, Lei Xu, V. Her-

breteau, H. Gurgel, N. Dessay, and P. Gong. A mapping review on urban landscape factors of dengue retrieved from earth observation data, gis techniques, and survey questionnaires. *Remote. Sens.*, 12:932, 2020.

[121] Nico Stollenwerk Maíra Aguiar and Bob W. Kooi. Torus bifurcations, isolas and chaotic attractors in a simple dengue fever model with ade and temporary cross immunity. *International Journal of Computer Mathematics*, 86(10-11):1867–1877, 2009.

[122] Kaisa Miettinen. Nonlinear multiobjective optimization. In *International Series in Operations Research and Management Science*, 1998.

[123] Arti Mishra and Sunita Gakkhar. Non-linear dynamics of two-patch model incorporating secondary dengue infection. *International Journal of Applied and Computational Mathematics*, 4(1):19, Nov 2017.

[124] Arti Mishra and Sunita Gakkhar. Modeling of dengue with impact of asymptomatic infection and ade factor. *Differential Equations and Dynamical Systems*, 28(3):745–761, Jul 2020.

[125] Kristof Van Moffaert, Madalina M. Drugan, and Ann Nowé. Scalarized multi-objective reinforcement learning: Novel design techniques. In *Proceedings of the 2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning, ADPRL 2013, IEEE Symposium Series on Computational Intelligence (SSCI), 16-19 April 2013, Singapore*, pages 191–199. IEEE, 2013.

[126] Amy C Morrison, Emily Zielinski-Gutierrez, Thomas W Scott, and Ronald Rosenberg. Defining challenges and proposing solutions for control of the virus vector aedes aegypti. *PLOS Medicine*, 5(3):1–5, 03 2008.

[127] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, 5(3):1–1, 03 2008.

[128] David Murillo, Susan A. Holechek, Anarina L. Murillo, Fabio Sanchez, and Carlos Castillo-Chavez. Vertical transmission in a two-strain model of dengue fever. *Letters in Biomathematics*, 1(2):249–271, January 2014.

[129] Allan H. Murphy. The finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, 11(1):3 – 20, 1996.

[130] M. Z. NDII, D. ALLINGHAM, R. I. HICKSON, and K. GLASS. The effect of wolbachia on dengue dynamics in the presence of two serotypes of dengue: symmetric and asymmetric epidemiological characteristics. *Epidemiology and Infection*, 144(13):2874–2882, 2016.

[131] Meksianis Z. Ndii, R.I. Hickson, David Allingham, and G.N. Mercer. Modelling the transmission dynamics of dengue in the presence of wolbachia. *Mathematical Biosciences*, 262:157–166, 2015.

[132] Ian Osband, Daniel Russo, and Benjamin Van Roy. (more) efficient reinforcement learning via posterior sampling. In *Neural Information Processing Systems*, 2013.

[133] A. PANDEY and J. MEDLOCK. The introduction of dengue vaccine may temporarily cause large spikes in prevalence. *Epidemiology and Infection*, 143(6):1276–1286, 2015.

[134] Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory*. Studies in Managerial Economics. Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1961.

[135] Peter Rashkov and Bob W. Kooi. Complexity of host-vector dynamics in a two-strain dengue model. *Journal of Biological Dynamics*, 15(1):35–72, 2021. PMID: 33357025.

[136] David A. Rasmussen, Oliver Ratmann, and Katia Koelle. Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Computational Biology*, 7, 2011.

[137] Mario Recker, Konstantin Blyuss, Cameron Simmons, Hien Tinh Tran, Bridget Wills, Jeremy Farrar, and Sunetra Gupta. Immunological serotype interactions and their effect on the epidemiological pattern of dengue. *Proceedings. Biological sciences / The Royal Society*, 276:2541–8, 04 2009.

[138] Nicholas G. Reich, Sourya Shrestha, Aaron A. King, Pejman Rohani, Justin Lessler, Siripen Kalayanarooj, In-Kyu Yoon, Robert V. Gibbons, Donald S. Burke, and Derek A. T. Cummings. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of The Royal Society Interface*, 10(86):20130414, 2013.

[139] Nicholas G. Reich, Sourya Shrestha, Aaron A. King, Pejman Rohani, Justin Lessler, Siripen Kalayanarooj, In-Kyu Yoon, Robert V. Gibbons, Donald S. Burke, and Derek A. T. Cummings. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of the Royal Society Interface*, 10(86):20130414, Sep 2013.

[140] Mathieu Reymond. *On incorporating prior knowledge about the decision maker in multi-objective reinforcement learning*. PhD thesis, Vrije Universiteit Brussel, 2023.

[141] Mathieu Reymond, Conor F. Hayes, Lander Willem, Roxana Radulescu, Steven Abrams, Diederik M. Roijers, Enda Howley, Patrick Mannion, Niel Hens, Ann Nowé, and Pieter Libin. Exploring the pareto front of multi-objective COVID-19 mitigation policies using reinforcement learning. *Expert Syst. Appl.*, 249:123686, 2024.

[142] Luis Rivera, Shibadas Biswal, Xavier Sáez-Llorens, Humberto Reynales, Eduardo López-Medina, Charissa Borja-Tabora, Lulu Bravo, Chukiat Sirivichayakul, Pope Kosalaraksa, Luis Martinez Vargas, Delia Yu, Veerachai Watanaveeradej, Felix Espinoza, Reynaldo Dietze, LakKumar Fernando, Pujitha Wickramasinghe, Edson Duarte MoreiraJr, Asvini D Fernando, Dulanie Gunasekera, Kleber Luz, Rivaldo Venâncioda Cunha, Martina Rauscher, Olaf Zent, Mengya Liu, Elaine Hoffman, Inge LeFevre, Vianney Tricou, Derek Wallace, MariaTheresa Alera, and for the TIDES study group Borkowski, Astrid. Three-year Efficacy and Safety of Takeda's Dengue Vaccine Candidate (TAK-003). *Clinical Infectious Diseases*, 75(1):107–117, 10 2021.

[143] Luis Rivera, Shibadas Biswal, Xavier Sáez-Llorens, Humberto Reynales, Eduardo López-Medina, Charissa Borja-Tabora, Lulu Bravo, Chukiat Sirivichayakul, Pope Kosalaraksa, Luis Martinez Vargas, Delia Yu, Veerachai Watanaveeradej, Felix Espinoza, Reynaldo Dietze, LakKumar Fernando, Pujitha Wickramasinghe, Edson Duarte MoreiraJr, Asvini D Fernando, Dulanie Gunasekera, Kleber Luz, Rivaldo Venâncioda Cunha, Martina Rauscher, Olaf Zent, Mengya Liu, Elaine Hoffman, Inge LeFevre, Vianney Tricou, Derek Wallace, MariaTheresa Alera, and for the TIDES study group Borkowski, Astrid. Three-year Efficacy and Safety of Takeda's Dengue Vaccine Candidate (TAK-003). *Clinical Infectious Diseases*, 75(1):107–117, 10 2021.

[144] Christian P. Robert. The bayesian choice : from decision-theoretic foundations to computational implementation. 2007.

[145] Filipe Rocha, Maíra Aguiar, Max Souza, and Nico Stollenwerk. Understanding the effect of vector dynamics in epidemic models using center manifold analysis. *AIP Conference Proceedings*, 1479(1):1319–1322, 09 2012.

[146] Filipe Rocha, Luís Mateus, Urszula Skwara, Maíra Aguiar, and Nico Stollenwerk. Understanding dengue fever dynamics: a study of seasonality in vector-borne disease models. *International Journal of Computer Mathematics*, 93(8):1405–1422, 2016.

[147] Diederik Roijers, Luisa Zintgraf, and Ann Nowe. Interactive thompson sampling for multi-objective multi-armed bandits. pages 18–34, 09 2017.

[148] Adrian Rosebrock. Intersection over union (iou) for object detection, 2016. Accessed: 2024-06-17.

[149] Alan L. Rothman. Dengue: defining protective versus pathologic immunity. *The Journal of Clinical Investigation*, 113(7):946–951, 4 2004.

[150] Alan L. Rothman. Cellular immunology of sequential dengue virus infection and its role in disease pathogenesis. *Current Topics in Microbiology and Immunology*, 338(1):83 – 98, 2009. Cited by: 98.

[151] Daniel Russo. Simple bayesian algorithms for best arm identification. In *Annual Conference Computational Learning Theory*, 2016.

[152] Daniel Russo and Benjamin Van Roy. Eluder dimension and the sample complexity of optimistic exploration. In *Neural Information Processing Systems*, 2013.

[153] Ilya O. Ryzhov, Warren B. Powell, and Peter I. Frazier. The knowledge gradient algorithm for a general class of online learning problems. *Operations Research*, 60(1):180–195, 2012.

[154] A B SABIN. The dengue group of viruses and its family relationships. *Bacteriological Reviews*, 14(3):225–232, September 1950.

[155] A B Sabin. Research on dengue during world war ii. *The American Journal of Tropical Medicine and Hygiene*, 1(1):30–50, Jan 1952.

[156] Nadhirat Sangkawibha, Suntharee Rojanasuphot, Sompop Ahandrik, Sukho Viriyapongse, Sujarti Jatanasen, Viraj Salitul, Boonluan Phanthumachinda, and Scott B. Halstead. Risk factors in dengue shock syndrome: A prospective epidemiologic study in rayong, thailand: I. the 1980 outbreak. *American Journal of Epidemiology*, 120(5):653–669, 11 1984.

[157] Boris V. Schmid and Mirjam E E Kretzschmar. Determinants of sexual network structure and their impact on cumulative network measures. *PLoS Computational Biology*, 8, 2012.

[158] Steven L. Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26:639–658, 2010.

[159] Afrina Andriani Sebayang, Hilda Fahlena, Vizda Anam, Damián Knopoff, Nico Stollenwerk, Maíra Aguiar, and Edy Soewono. Modeling dengue immune responses mediated by antibodies: A qualitative study. *Biology*, 10(9), 2021.

[160] Eunha Shim. Optimal dengue vaccination strategies of seropositive individuals. *Mathematical Biosciences and Engineering*, 16(3):1171–1189, 2019.

[161] Beatriz Sierra, Ana B. Perez, Katrin Vogt, Gissel Garcia, Kathrin Schmolke, Eglys Aguirre, Mayling Alvarez, Florian Kern, Gustavo Kourí, Hans-Dieter Volk, and Maria G. Guzman. Secondary heterologous dengue infection risk: Disequilibrium between immune regulation and inflammation. *Cellular Immunology*, 262(2):134–140, 2010.

[162] Alexander Slepoy, Aidan P. Thompson, and Steven J. Plimpton. A constant-time kinetic monte carlo algorithm for simulation of large biochemical reaction networks. *The Journal of chemical physics*, 128 20:205101, 2008.

[163] Aleksandrs Slivkins et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[164] Saranya Sridhar, Alexander Luedtke, Edith Langevin, Ming Zhu, Matthew Bonaparte, Tifany Machabert, Stephen Savarino, Betzana Zambrano, Annick Moureau, Alena Khromava, Zoe Moodie, Ted Westling, Cesar Mascareñas, Carina Frago, Margarita Cortés, Danaya Chansinghakul, Fernando Noriega, Alain Bouckenooghe, Josh Chen, Su-Peing Ng, Peter B. Gilbert, Sanjay Gurunathan, and Carlos A. DiazGranados. Effect of dengue serostatus on dengue vaccine safety and efficacy. *New England Journal of Medicine*, 379(4):327–340, 2018. PMID: 29897841.

[165] Ashley L. St. John and Abhay P. S. Rathore. Adaptive immune responses to primary and secondary dengue virus infections. *Nature Reviews Immunology*, 19(4):218–230, 2019.

[166] N. Stollenwerk, M. Aguiar, S. Ballesteros, J. Boto, B. Kooi, and L. Mateus. Dynamic noise, chaos and parameter estimation in population biology. *Interface Focus*, 2(2):156–169, 2012.

[167] Auliya Abdurrohim Suwantika, Woro Supadmi, Mohammad Ali, and Rizky Abdulah. Cost-effectiveness and budget impact analyses of dengue vaccination in indonesia. *PLoS Neglected Tropical Diseases*, 15(8):e0009664, 2021.

[168] T. T. Tanimoto. *An Elementary Mathematical Theory of Classification and Prediction*. International Business Machines Corporation, 1958. Accessed: 2024-06-18.

[169] Quirine A. ten Bosch, Brajendra K. Singh, Muhammad R. A. Hassan, Dave D. Chadee, and Edwin Michael. The role of serotype interactions and seasonality in dengue model selection and control: Insights from a pattern matching approach. *PLOS Neglected Tropical Diseases*, 10(5):1–25, 05 2016.

[170] William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933.

[171] Sherry Towers. Sir infectious disease model with age classes, 2012. URL: http://sherrytowers.com/2012/12/11/sir-model-with-age-classes/ (Access Date: Access Date).

[172] T. Tsheten, A. Clements, D. Gray, R. Adhikary, and K. Wangdi. Clinical features and outcomes of covid-19 and dengue co-infection: a systematic review. *BMC Infectious Diseases*, 21, 2021.

[173] Peter Vamplew, Richard Dazeley, Adam Berry, Rustam Issabekov, and Evan Dekker. Empirical evaluation methods for multiobjective reinforcement learning algorithms. *Machine Learning*, 84:51–80, 2011.

[174] David W. Vaughn, Sharone Green, Siripen Kalayanarooj, Bruce L. Innis, Suchitra Nimmannitya, Saroj Suntayakorn, Timothy P. Endy, Boonyos Raengsakulrach, Alan L. Roth-

man, Francis A. Ennis, and Ananda Nisalak. Dengue Viremia Titer, Antibody Response Pattern, and Virus Serotype Correlate with Disease Severity. *The Journal of Infectious Diseases*, 181(1):2–9, 01 2000.

[175] Joannes Vermorel and Mehryar Mohri. Multi-armed bandit algorithms and empirical evaluation. In *European conference on machine learning*, pages 437–448. Springer, 2005.

[176] Luis Villar, Gustavo Horacio Dayan, José Luis Arredondo-García, Doris Maribel Rivera, Rivaldo Cunha, Carmen Deseda, Humberto Reynales, Maria Selma Costa, Javier Osvaldo Morales-Ramírez, Gabriel Carrasquilla, Luis Carlos Rey, Reynaldo Dietze, Kleber Luz, Enrique Rivas, Maria Consuelo Miranda Montoya, Margarita Cortés Supelano, Betzana Zambrano, Edith Langevin, Mark Boaz, Nadia Tornieporth, Melanie Saville, and Fernando Noriega. Efficacy of a tetravalent dengue vaccine in children in latin america. *New England Journal of Medicine*, 372(2):113–123, 2015. PMID: 25365753.

[177] Eryu Wang, Haolin Ni, Renling Xu, Alan D. T. Barrett, Stanley J. Watowich, Duane J. Gubler, and Scott C. Weaver. Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *Journal of Virology*, 74(7):3227–3234, 2000.

[178] Wei-Kung Wang, Day-Yu Chao, Chuan-Liang Kao, Han-Chung Wu, Yung-Ching Liu, Chien-Ming Li, Shih-Chung Lin, Shih-Ting Ho, Jyh-Hsiung Huang, and Chwan-Chuen King. High levels of plasma dengue viral load during defervescence in patients with dengue hemorrhagic fever: Implications for pathogenesis. *Virology*, 305(2):330–338, 2003.

[179] Weijia Wang and Michèle Sebag. Hypervolume indicator and dominance reward based multi-objective monte-carlo tree search. *Machine Learning*, 92:403 – 429, 2013.

[180] Daniela Weiskopf and Alessandro Sette. T-cell immunity to infection with dengue virus in humans. *Frontiers in Immunology*, 5, 2014.

[181] E.G. Westaway, M.A. Brinton, S.Ya. Gaidamovich, M.C. Horzinek, A. Igarashi, L. Kääriäinen, D.K. Lvov, J.S. Porterfield, P.K. Russell, and D.W. Trent. Flaviviridae. *Intervirology*, 24(4):183–192, 07 2008.

[182] Paul S. Wikramaratna, Cameron P. Simmons, Sunetra Gupta, and Mario Recker. The effects of tertiary and quaternary infections on the epidemiology of dengue. *PLOS ONE*, 5(8):1–8, 08 2010.

[183] Paul S. Wikramaratna, Cameron P. Simmons, Sunetra Gupta, and Mario Recker. The effects of tertiary and quaternary infections on the epidemiology of dengue. *PLOS ONE*, 5(8):1–8, 08 2010.

[184] Annelies Wilder-Smith. Dengue vaccine development by the year 2020: challenges and prospects. *Current Opinion in Virology*, 43:71–78, 2020. Viral elimination * Special Section: Defeat Dengue.

[185] Lander Willem, Sean Stijven, Ekaterina Vladislavleva, Jan Broeckhove, Philippe Beutels, and Niel Hens. Active learning to understand infectious disease models and improve policy making. *PLoS Computational Biology*, 10, 2014.

[186] Lander Willem, Frederik Verelst, Joke Bilcke, Niel Hens, and Philippe Beutels. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). *BMC Infectious Diseases*, 17, 2017.

[187] Hannah Woodall and Ben Adams. Partial cross-enhancement in models for dengue epidemiology. *Journal of Theoretical Biology*, 351:67–73, 2014.

[188] World Health Organization. Meeting of the strategic advisory group of experts on immunization, april 2016: conclusions and recommendations. *Weekly Epidemiological Record*, 91(21):266–284, 5 2016. Accessed: 2024-01-24.

[189] World Health Organization. Dengue vaccine: Who position paper, july 2016 – recommendations. *Vaccine*, 35(9):1200–1201, 2017.

[190] World Health Organization. Dengue vaccines: Who position paper – september 2018. *Weekly Epidemiological Record*, 93(35):457–476, 9 2018. Accessed: 2024-01-24.

[191] World Health Organization. Dengue and severe dengue, March 2023. Accessed: 2024-01-24.

[192] World Health Organization. Dengue - global situation, May 2024. [Accessed: 2024-07-21].

[193] Joseph T. Wu, Joseph T. Wu, Steven Riley, Christophe Fraser, and Gabriel M. Leung. Reducing the impact of the next influenza pandemic using household-based public health interventions. *Hong Kong medical journal = Xianggang yi xue za zhi*, 15 Suppl 9:38–41, 2006.

[194] Ling Xue, Hongyu Zhang, Wei Sun, and Caterina Scoglio. Transmission dynamics of multi-strain dengue virus with cross-immunity. *Applied Mathematics and Computation*, 392:125742, 2021.

[195] Luis Mier y Teran-Romero, Ira B. Schwartz, and Derek A.T. Cummings. Breaking the symmetry: Immune enhancement increases persistence of dengue viruses in the presence of asymmetric transmission rates. *Journal of Theoretical Biology*, 332:203–210, 2013.

[196] Saba Yahyaa. The exploration vs exploitation trade-off in bandit problems: An empirical study. 04 2012.

[197] Saba Yahyaa, Madalina Drugan, and Bernard Manderick. Multivariate normal distribution based multi-armed bandits pareto algorithm. 01 2014.

[198] Saba Yahyaa, Madalina M. Drugan, and Bernard Manderick. *Scalarized and Pareto Knowledge Gradient for Multi-objective Multi-armed Bandits*, pages 99–116. Springer International Publishing, Cham, 2015.

[199] Saba Yahyaa and Bernard Manderick. Thompson sampling for multi-objective multi-armed bandits problem. 04 2015.

[200] Saba Q. Yahyaa, Madalina M. Drugan, and Bernard Manderick. Annealing-pareto multi-objective multi-armed bandit algorithm. In *2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pages 1–8, 2014.

[201] Saba Q. Yahyaa, Madalina M. Drugan, and Bernard Manderick. Knowledge gradient for multi-objective multi-armed bandit algorithms. In Béatrice Duval, H. Jaap van den Herik, Stéphane Loiseau, and Joaquim Filipe, editors, *ICAART 2014 - Proceedings of the 6th International Conference on Agents and Artificial Intelligence, Volume 1, ESEO, Angers, Loire Valley, France, 6-8 March, 2014*, pages 74–83. SciTePress, 2014.

[202] Saba Q. Yahyaa, Madalina M. Drugan, and Bernard Manderick. Linear scalarized knowledge gradient in the multi-objective multi-armed bandits problem. In *22th European Sym-*

posium on Artificial Neural Networks, ESANN 2014, Bruges, Belgium, April 23-25, 2014, 2014.

[203] Logan Michael Yliniemi and Kagan Tumer. Multi-objective multiagent credit assignment in reinforcement learning and nsga-ii. *Soft Computing*, 20:3869 – 3887, 2016.

[204] Wu Zeng, Yara A Halasa-Rappel, Nicolas Baurin, Laurent Coudeville, and Donald S Shepard. Cost-effectiveness of dengue vaccination in ten endemic countries. *Vaccine*, 36(3):413–420, 2018.

[205] Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Trans. Evol. Comput.*, 3:257–271, 1999.

[206] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and Viviane Grunert da Fonseca. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans. Evol. Comput.*, 7:117–132, 2003.

[207] Masrour Zoghi, Shimon Whiteson, Remi Munos, and Maarten de Rijke. Relative upper confidence bound for the k-armed dueling bandit problem, 2013.