

Modeling and forecasting infectious diseases

by

Chunlin Yi

B.S., Harbin Institute of Technology, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wiegers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Abstract

In recent years, the modeling and forecasting of infectious diseases have received heightened attention, driven by the emergence and global spread of high-profile outbreaks such as the 2014 Ebola epidemic in West Africa and the COVID-19 pandemic. These health crises have highlighted the critical importance of having timely, accurate information about disease transmission dynamics to guide public health decisions and interventions effectively. The role of infectious disease modeling and forecasting extends beyond mere academic inquiry; it serves as a vital tool for public health officials, offering the potential to predict outbreaks, strategize interventions, allocate resources judiciously, and communicate risks to communities. Therefore, this dissertation endeavors to develop and refine systems capable of modeling epidemic dynamics and delivering precise forecasts for a variety of infectious diseases, thereby supporting enhanced public health preparedness and response.

Firstly, we introduce a short-term forecasting system for COVID-19, employing the SEIR compartmental model to simulate the disease's transmission and employing the ensemble Kalman filter (EnKF) to assimilate daily confirmed case data. This system generates three-day forecasts, updated daily as new data becomes available, with forecast accuracy assessed through absolute relative error metrics.

Secondly, our work expands into the vector-borne disease – dengue fever, developing a novel forecasting model that integrates the SEIR-SEI disease model with the EnKF method. Initial twin experiments assimilated synthetic weekly dengue case data, leading to accurate state and parameter estimations and predictions. Subsequently, this model was applied to historical dengue outbreak data from Kaohsiung, Taiwan; Singapore; and Rio de Janeiro, Brazil, yielding forecasts that closely aligned with observed cases and dynamically estimating exposed and infectious human populations. A comparative accuracy assessment of this

model against the SEIR-EnKF and SIR-EnKF models highlighted its superior forecasting capabilities.

Third, we detail a real-time prediction system for West Nile Virus (WNV) that incorporates an adapted compartment model to account for the transmission dynamics among birds, mosquitoes, and humans, including asymptomatic cases and the influence of climatic factors. Using data assimilation techniques, we generate weekly WNV case forecasts for Colorado in 2023, providing valuable insights for public health planning. Comparative analyses underscore the enhanced forecast accuracy achieved by integrating climatic variables into our models.

Fourth, we explore the transmission dynamics of Foot-and-Mouth Disease (FMD) among beef-cattle farms in southwest Kansas through a three-layer network model encompassing direct contacts (cattle movements), indirect contacts (truck visits), and information sharing. This study assesses the impact of these layers on FMD transmission, the effectiveness of information-sharing strategies in mitigating outbreaks, and the economic implications of infection and quarantine measures.

Finally, our work on the development of the PICTUREE—Aedes web application represents a significant advancement in the monitoring and forecasting of dengue fever. By aggregating and visualizing dengue-related data from diverse sources, including temperature, precipitation, mosquito occurrences, and case reports, the application facilitates global risk assessments and forecasts of ongoing outbreaks.

In conclusion, this dissertation presents a comprehensive suite of modeling and forecasting tools for various infectious diseases, offering significant potential to optimize public health responses through improved predictive accuracy and timely information dissemination.

Modeling and forecasting infectious diseases

by

Chunlin Yi

B.S., Harbin Institute of Technology, 2018

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wieggers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Approved by:

Major Professor
Caterina Scoglio

Copyright

© Chunlin Yi 2024.

Abstract

In recent years, the modeling and forecasting of infectious diseases have received heightened attention, driven by the emergence and global spread of high-profile outbreaks such as the 2014 Ebola epidemic in West Africa and the COVID-19 pandemic. These health crises have highlighted the critical importance of having timely, accurate information about disease transmission dynamics to guide public health decisions and interventions effectively. The role of infectious disease modeling and forecasting extends beyond mere academic inquiry; it serves as a vital tool for public health officials, offering the potential to predict outbreaks, strategize interventions, allocate resources judiciously, and communicate risks to communities. Therefore, this dissertation endeavors to develop and refine systems capable of modeling epidemic dynamics and delivering precise forecasts for a variety of infectious diseases, thereby supporting enhanced public health preparedness and response.

Firstly, we introduce a short-term forecasting system for COVID-19, employing the SEIR compartmental model to simulate the disease's transmission and employing the ensemble Kalman filter (EnKF) to assimilate daily confirmed case data. This system generates three-day forecasts, updated daily as new data becomes available, with forecast accuracy assessed through absolute relative error metrics.

Secondly, our work expands into the vector-borne disease – dengue fever, developing a novel forecasting model that integrates the SEIR-SEI disease model with the EnKF method. Initial twin experiments assimilated synthetic weekly dengue case data, leading to accurate state and parameter estimations and predictions. Subsequently, this model was applied to historical dengue outbreak data from Kaohsiung, Taiwan; Singapore; and Rio de Janeiro, Brazil, yielding forecasts that closely aligned with observed cases and dynamically estimating exposed and infectious human populations. A comparative accuracy assessment of this

model against the SEIR-EnKF and SIR-EnKF models highlighted its superior forecasting capabilities.

Third, we detail a real-time prediction system for West Nile Virus (WNV) that incorporates an adapted compartment model to account for the transmission dynamics among birds, mosquitoes, and humans, including asymptomatic cases and the influence of climatic factors. Using data assimilation techniques, we generate weekly WNV case forecasts for Colorado in 2023, providing valuable insights for public health planning. Comparative analyses underscore the enhanced forecast accuracy achieved by integrating climatic variables into our models.

Fourth, we explore the transmission dynamics of Foot-and-Mouth Disease (FMD) among beef-cattle farms in southwest Kansas through a three-layer network model encompassing direct contacts (cattle movements), indirect contacts (truck visits), and information sharing. This study assesses the impact of these layers on FMD transmission, the effectiveness of information-sharing strategies in mitigating outbreaks, and the economic implications of infection and quarantine measures.

Finally, our work on the development of the PICTUREE—Aedes web application represents a significant advancement in the monitoring and forecasting of dengue fever. By aggregating and visualizing dengue-related data from diverse sources, including temperature, precipitation, mosquito occurrences, and case reports, the application facilitates global risk assessments and forecasts of ongoing outbreaks.

In conclusion, this dissertation presents a comprehensive suite of modeling and forecasting tools for various infectious diseases, offering significant potential to optimize public health responses through improved predictive accuracy and timely information dissemination.

Table of Contents

List of Figures	xii
List of Tables	xiv
Acknowledgements	xv
Preface	xvi
1 Introduction	1
1.1 Background and motivation	1
1.2 Epidemic models	4
1.3 Compartmental model	5
1.4 Ensemble Kalman Filtering	7
1.5 Multi-layer network construction	8
1.6 Research contributions	9
1.7 Dissertation structure	11
2 Short-term forecasts for the COVID-19 epidemic in Hubei Province, China	12
2.1 Introduction	12
2.2 Model	13
2.3 Results	16
2.4 Summary	18
3 Estimating and Forecasting Dengue Outbreak Dynamics	20
3.1 Introduction	20

3.2	Method	23
3.2.1	Data	23
3.2.2	Disease model	24
3.2.3	Forecast model	26
3.2.4	Twin experiments	28
3.2.5	Estimations and forecasts of the SEIR-SEI-EnKF model	29
3.2.6	Real-time predictions of different disease models	29
3.3	Results	29
3.3.1	Twin experiments	29
3.3.2	Estimations and forecasts of SEIR-SEI-EnKF model	30
3.3.3	Real-time predictions of different disease models	32
3.4	Discussion	35
4	Real-time forecasting and estimating for West Nile virus	38
4.1	Introduction	38
4.2	Materials and Methods	41
4.2.1	Data	41
4.2.2	Disease model	41
4.2.3	Forecast model	44
4.3	Results	45
4.3.1	Correlation of human WNV cases with local temperature and precipitation	45
4.3.2	Real-time forecast	46
4.3.3	Estimated epidemic dynamics	48
4.4	Discussion	49
5	Multilayer network analysis of FMD transmission and containment among beef cattle farms	53

5.1	Introduction	53
5.2	Materials and methods	55
5.2.1	Data	55
5.2.2	Multilayer network construction	56
5.2.3	Network measurements	57
5.2.4	Epidemic model	59
5.2.5	Impact of cattle movement, vehicle visits, and information sharing on disease spreading	60
5.2.6	Impact of different information-sharing layers and premises participa- tion on disease spreading	61
5.2.7	Sensitivity analysis of the potential economic loss from the epidemic and quarantine	62
5.3	Results	62
5.3.1	Impact of cattle movement, vehicle visits, and information sharing on disease spreading	62
5.3.2	Impact of different information-sharing layers and premises participa- tion on disease spreading	65
5.3.3	Sensitivity analysis of the potential economic loss from the epidemic and quarantine	66
5.4	Conclusion	68
6	PICTUREE—Aedes: A web application for dengue data visualization and case prediction	70
6.1	Introduction	70
6.2	Data	72
6.3	System Architecture	73
6.3.1	Client Application	73
6.3.2	Server Backend	74

6.4	Results	74
6.4.1	Data visualization	74
6.4.2	Simulation results	76
6.5	Conclusion	79
7	Conclusion and future work	81
7.1	Conclusion	81
7.2	Future work	83
	Bibliography	85
A	Estimations and forecasts of dengue outbreak for Singapore and RDJ	103
A.1	Estimations and forecasts of SEIR-SEI-EnKF model for Singapore and RDJ	103
A.1.1	Real-time prediction of different disease models for Singapore and RDJ	106

List of Figures

1.1	Structures used in epidemic modeling	6
1.2	SIR model transition process.	7
1.3	Two-layer network structure.	9
2.1	SEIR compartmental model	14
2.2	Distribution of the short-term prediction	17
2.3	Average number as predictions and the error	18
2.4	Estimated infection rate	19
3.1	SEIR-SEI compartmental model	25
3.2	Estimation and prediction results for twin experiments	31
3.3	Estimation and prediction results for SEIR-SEI-EnKF model	33
3.4	Real-time predictions of total confirmed cases	34
3.5	Real-time predictions of newly confirmed cases	35
3.6	RMS errors of real-time predictions	36
4.1	SEIR-SEI-SEAC compartmental model	43
4.2	Weekly human WNV cases, average temperature, and weekly precipitation	46
4.3	Real-time prediction of the cumulative WNV cases in Colorado	47
4.4	Accuracy evaluation	48
4.5	Estimated epidemic dynamics	50
4.6	Accuracy evaluation of the baseline model	51
5.1	Transmission routes of the FMD virus among different production types of beef-cattle farms	56

5.2	Multilayer network structure of FMD transmission among beef-cattle farms .	57
5.3	Node degree distributions of different networks	58
5.4	SQEIR compartmental model	59
5.5	Number of farms in infected compartments and distribution of epidemic size	63
5.6	The epidemic size and quarantine size versus random participation	65
5.7	The epidemic size and quarantine size versus highest-degree participation . .	66
5.8	Relative total economic losses on networks with different L_{IS} (highest-degree participations)	67
5.9	Range of the total loss under different participations and relative quarantine losses	68
6.1	Components and architecture of the application	73
6.2	Graphical interface of the PICTUREE–Aedes	75
6.3	Weather station data	75
6.4	Occurrence data of dengue, Aedes aegypti, and Aedes albopictus	76
6.5	Google search trends of ”dengue” and ”fever”	76
6.6	Mosquito simulation results visualized spatially	77
6.7	Mosquito simulation results visualized temporally	78
6.8	Forecasts of the dengue cases in Cambodia	78
A.1	Estimation and prediction results for 2020 dengue outbreak in Singapore . .	104
A.2	Estimation and prediction results for 2015 dengue outbreak in RDJ	105
A.3	Real-time predictions for 2020 dengue outbreak in Singapore	106
A.4	Real-time predictions for 2015 dengue outbreak in RDJ	107

List of Tables

3.1	Preset parameters Θ and initial conditions $\bar{x}_0$ of SEIR-SEI model	28
3.2	statistics of the predictions	32
3.3	Peak value of predicted new cases and their absolute relative errors (ARE) .	35
4.1	The correlation coefficients r between human WNV cases and higher orders i of weather variables (temperature T and precipitation P)	51
5.1	Description and value of the parameters in the model	60
6.1	Data sources	73
6.2	Number of predicted cumulative cases and its relative absolute error in Cam- bodia	79

Acknowledgments

First and foremost, my deepest gratitude goes to my advisor, Dr. Caterina Scoglio, for her unwavering support and invaluable guidance throughout my Ph.D. journey. Her patience, enthusiasm, and integrity have not only shaped my academic development but also left a lasting impression on me. The wisdom and lessons I have gained under her mentorship are treasures I will carry with me always.

I extend my heartfelt gratitude to my committee members, Dr. Lee W. Cohnstaedt, Dr. Don Gruenbacher, and Dr. Xiaolong Guo, for their time, insightful feedback, and constructive criticism, which have significantly contributed to my research and professional growth.

My thanks also go out to all my colleagues within our research group for their support, collaboration, and the enriching environment we shared. I am particularly grateful to Dr. Qihui Yang, Dr. Aram Vajdi, and Dr. Tanvir Ferdousi for their contributions to my academic journey.

I cannot express enough gratitude to my parents, whose unconditional love and unwavering support laid the foundation for all my achievements. Their sacrifices and belief in me have been my constant source of motivation and strength.

I am also grateful to Teng Xu, NiuNiu, Tai Stephan, Tierza Stephan, SingSing, Yexi Xu, Zijie Xiao, Yijia Liu, and Autumn Welch for their encouragement and companionship, which have been a great source of comfort and joy to me. Moreover, I owe a debt of gratitude to my other friends, whose presence and shared moments have enriched my life in countless ways.

Preface

This dissertation, "Modeling and forecasting infectious diseases", is submitted for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering at Kansas State University. The research was conducted under the supervision of Professor Caterina Scoglio.

To the best of my knowledge, this work is original, except where acknowledgments and references are indicated. Most of the work has been presented in the following published or submitted peer-reviewed journals:

- Qihui Yang, Chunlin Yi, Aram Vajdi, Lee W Cohnstaedt, Hongyu Wu, Xiaolong Guo, and Caterina M Scoglio. Short-term forecasts and long-term mitigation evaluations for the COVID-19 epidemic in Hubei province, China. *Infectious Disease Modelling*, 5:563-574,2020.
- Chunlin Yi, Lee W. Cohnstaedt, and Caterina M. Scoglio. SEIR-SEI-EnKF: a new model for estimating and forecasting dengue outbreak dynamics. *IEEE Access*, 9:156758–156767, 2021.
- Chunlin Yi, Qihui Yang, and Caterina M. Scoglio. Multilayer network analysis of FMD transmission and containment among beef cattle farms. *Scientific Reports*, 12:15679, 2022.
- Chunlin Yi, Aram Vajdi, Tanvir Ferdousi, Lee W. Cohnstaedt, and Caterina M. Scoglio. Picturee—aedes: A web application for dengue data visualization and case prediction. *Pathogens*, 12:771, 2023.
- Chunlin Yi, Lee W. Cohnstaedt, and Caterina M. Scoglio. A real-time forecasting and estimating system of West Nile virus: a case study of the 2023 WNV outbreak in Colorado, USA. Submitted for journal publication, 2024.

This research is supported by the Department of the Army, U.S. Army Contracting Command, Aberdeen Proving Ground, Natick Contracting Division, Ft Detrick, MD (DWFP grant W911QY-19-1-0004), the National Science Foundation under Grant Award IIS-2027336, the USDA National Institute of Food and Agriculture award number 2022-67015-38059 via the NSF/NIH/USDA/BBSRC/BSF/NSFC Ecology and Evolution of Infectious Diseases Program, and the United States Department of Agriculture ARS under agreement No.:58-3022-1-010.

Chapter 1

Introduction

1.1 Background and motivation

Throughout history, infectious diseases have posed significant threats to human health and well-being, leaving deep scars on society. The Spanish flu pandemic of 1918-1919 serves as a harrowing reminder of this impact, affecting one-third of the global population and resulting in approximately 50 million deaths¹. In more recent times, infectious diseases such as dengue fever and influenza continue to pose serious threats to global health, emphasizing the need for vigilant public health strategies. Dengue fever, a mosquito-borne viral disease, is estimated to infect up to 100 million individuals annually, leading to significant morbidity and mortality². Similarly, influenza, a respiratory illness caused by influenza viruses, accounts for 290,000 to 650,000 deaths worldwide each year³. The emergence of the COVID-19 pandemic has further underscored the devastating impact of infectious diseases, with over 5 million lives lost globally as of April 2023⁴. These outbreaks not only highlight the human cost of infectious diseases but also the critical importance of timely, accurate information on disease transmission to guide public health interventions. Consequently, infectious disease forecasting has become an indispensable tool for public health officials, enabling more effective preparation and response to disease outbreaks.

This dissertation delves into the transmission dynamics of four significant infectious dis-

eases: COVID-19, dengue fever, foot-and-mouth disease (FMD), and West Nile virus (WNV). These diseases have been selected for their varied transmission modes, impacts on human and animal health, and the challenges they present to public health systems globally.

- **COVID-19** The COVID-19 pandemic, caused by the SARS-CoV-2 virus, emerged in Wuhan, China, in late 2019 and quickly became a global crisis. Its rapid spread was facilitated by modern travel and asymptomatic transmission, highlighting vulnerabilities in global health systems. COVID-19's impact has been profound, affecting millions of lives, overwhelming healthcare systems, and causing widespread social and economic disruption⁵⁻⁷. The pandemic underscored the need for robust surveillance systems, effective public health communication, and international cooperation to manage infectious disease threats. As of April 2023, the pandemic had resulted in over 5 million deaths worldwide, although the true toll, directly and indirectly, is likely much higher⁴.
- **Dengue Fever** Dengue fever, transmitted by *Aedes* mosquitoes, is a leading cause of illness and death in tropical and subtropical regions. The disease's symptoms range from mild fever to severe conditions such as dengue hemorrhagic fever and dengue shock syndrome. The global incidence of dengue has grown dramatically in recent decades, with an estimated 100 million to 400 million infections occurring annually. The disease's spread is exacerbated by urbanization, climate change, and globalization, posing significant challenges to disease control and prevention efforts². Large outbreaks have occurred in countries like Brazil, Singapore, and Taiwan, stressing the importance of effective vector control and early detection systems⁸⁻¹¹.
- **West Nile Virus (WNV)** WNV is a mosquito-borne virus that primarily infects birds but can also infect humans and other mammals. In humans, it can cause West Nile fever or severe neurological diseases, such as encephalitis or meningitis. First identified in Uganda in 1937, WNV has since spread globally, with notable outbreaks in the United States, Europe, and the Middle East. The virus's transmission cycle, involving birds and mosquitoes, complicates control efforts, requiring integrated approaches that include mosquito population management and surveillance of bird populations.

The US experienced significant outbreaks in the early 2000s, with thousands of cases and hundreds of deaths, emphasizing the need for ongoing research and public health preparedness^{12–15}.

- **Foot-and-mouth disease (FMD)** FMD is a highly contagious viral disease affecting livestock, with severe economic implications for the agricultural sector. It affects cloven-hoofed animals, leading to reduced productivity and trade restrictions. Outbreaks, such as those in the UK in 2001 and Japan in 2010, have led to the culling of millions of animals and substantial economic losses^{16–18}. Effective FMD management involves vaccination, movement controls, and rapid detection to prevent spread, highlighting the importance of understanding transmission dynamics for outbreak preparedness and response^{19–21}.

This dissertation aims to provide a comprehensive understanding of the transmission dynamics of these diseases, contributing to the development of more effective forecasting models and intervention strategies. This work is driven by the following research questions designed to tackle the multifaceted challenges posed by disease prediction and control:

1. *COVID-19 Prediction Amidst Data Variability*: How can we enhance the accuracy of COVID-19 case predictions amidst significant observational disturbances?
2. *Dengue Fever Transmission Dynamics*: What methodologies can more precisely model the transmission dynamics of dengue fever between humans and mosquitoes?
3. *West Nile Virus (WNV) Dynamics and Climate Effects*: How do we accurately describe the transmission dynamics of WNV across birds, mosquitoes, and humans, incorporating the significant influence of climate variables on disease spread?
4. *Foot-and-Mouth Disease (FMD) Transmission and Control*: How can we identify the differential impacts of various contact types on FMD transmission, and how can we evaluate the effectiveness of the mitigation strategies?

5. *Comprehensive Dengue Risk Assessment*: How can we integrate diverse data sources, along with existing simulation and forecasting tools, to provide a detailed and actionable risk assessment of dengue fever to the public and health authorities?

By tackling these questions, this dissertation aims to bridge the existing gaps in infectious disease research through the application of innovative methodologies. This endeavor not only seeks to improve our predictive capabilities but also to inform more effective response mechanisms to infectious disease outbreaks. Ultimately, the goal is to mitigate the adverse impacts of these diseases on human and animal health, contributing to safer, healthier communities worldwide.

1.2 Epidemic models

The prediction and analysis of the evolution of diseases can be modeled by a spectrum of mathematical methods dedicated to mapping out spreading and diffusion mechanisms. These methodologies have undergone significant evolution, transitioning from straightforward compartmental constructs to more elaborate structured frameworks. Such advancements aim to encapsulate the complexities introduced by the inherent hierarchies and heterogeneities observable at both community and population levels.²²

In Fig. 1.1, we explore four structures that are frequently employed within the realm of epidemic modeling. Here, individuals within the models are represented as circles, with varying shades indicating distinct disease stages. Fig. 1.1a illustrates the concept of homogeneous mixing structure. This model operates under the assumption that interactions among individuals occur in a uniformly random manner, devoid of any preferential patterns or structured contacts. This simplistic approach, while providing initial insights, does not account for the complexity of real-world interactions. In contrast, Fig. 1.1b delves into contact network models. These models map out the intricate network of social interactions between individuals, delineating potential pathways for viral transmission. By acknowledging the diversity in interaction frequencies and intensities, contact network models offer a more detailed understanding of disease spread mechanisms. Fig. 1.1c introduces multi-scale

models, which explore a layered approach to population dynamics. These models differentiate between subpopulations, interconnected through individual movements across spatial or social divides. While assuming homogeneous mixing within each subpopulation, multi-scale models provide insights into how mobility patterns influence disease propagation across broader scales. Lastly, Fig. 1.1d presents agent-based models, which simulate the behaviors, movements, and interactions of each individual entity within a population, with a high degree of detail. Agent-based models are particularly adept at capturing the emergent properties of complex systems, allowing for an in-depth exploration of how individual actions collectively shape epidemic outcomes.

Each of these structured frameworks brings unique perspectives and insights to the study of epidemic dynamics, highlighting the importance of considering diverse modeling approaches to capture the multifaceted nature of disease spread.

1.3 Compartmental model

Compartmental models play a central role in epidemic modeling and can be applied across the various structures outlined in Section 1.2. These models provide a simplified, yet powerful, framework for understanding and predicting the dynamics of infectious diseases within populations. By dividing the population into distinct compartments according to their disease status (e.g., susceptible, infected, recovered), they facilitate a clear understanding of disease progression. The transition of individuals between these compartments, governed by a set of mathematical equations, effectively simulates how an infection disseminates over time.

The Susceptible-Infectious-Recovered (SIR) model is a basic structure of the compartmental models for studying the spread of infectious diseases. The model partitions the population into three compartments: susceptible, infectious, and recovered. Susceptible individuals are those who are susceptible to the disease and can become infected upon contact with an infectious individual. Infectious individuals are those who are currently infected and can transmit the disease to susceptible individuals. Recovered individuals are those

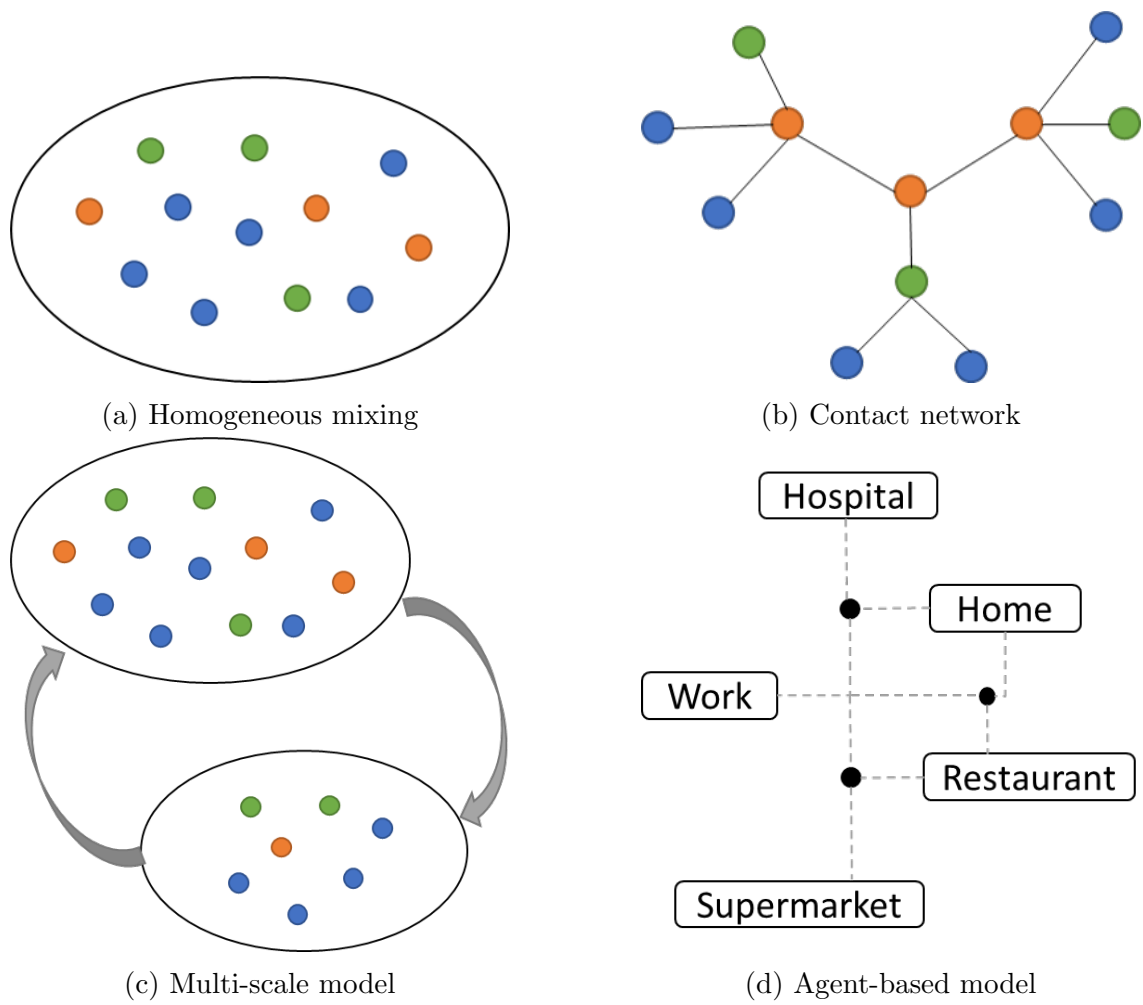


Figure 1.1: Structures used in epidemic modeling

who have recovered from the disease and are no longer susceptible or infectious. The SIR model assumes that the population is homogeneous and that all individuals have an equal likelihood of being infected. It also assumes that the infection rate β is proportional to the product of the infectious and susceptible populations, while the recovery rate δ is constant. The model is characterized by the following differential equations 1.1, as shown in Fig. 1.2.

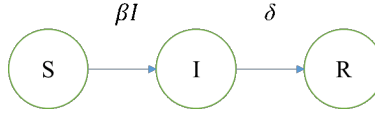


Figure 1.2: SIR model transition process.

$$\begin{aligned}
 S'(t) &= -\beta IS \\
 I'(t) &= \beta IS - \delta I \\
 R'(t) &= \delta I
 \end{aligned} \tag{1.1}$$

1.4 Ensemble Kalman Filtering

The Ensemble Kalman Filter (EnKF) is a data assimilation technique well-suited for time-series infectious disease forecasting with the assumption that the population is homogeneous mixing. In infectious disease modeling, time-series data such as daily or weekly case counts are often used to calibrate and validate disease transmission models. EnKF can improve the accuracy of these models by assimilating real-time data into the model and generating updated forecasts and key epidemiological parameters based on the most recent observations. The EnKF maintains an ensemble of the state variables X_t and the time-varying parameters $\Phi_t = [\beta^{V \rightarrow H} \beta^{H \rightarrow V} \lambda^V]^T$, which can be put together in a matrix $A_t \in R^{10 \times N}$ where N is the number of ensembles

$$A_t = \begin{pmatrix} X_t^1 & X_t^2 & \dots & X_t^N \\ \Phi_t^1 & \Phi_t^2 & \dots & \Phi_t^N \end{pmatrix} \tag{1.2}$$

At each time step, the EnKF first uses the model simulations to predict the prior, which is the expected state of the system

$$A_{t+1}^- = F(A_t) = \begin{pmatrix} f(X_t) \\ g(\Phi_t) \end{pmatrix} + q_t \quad (1.3)$$

where $g(\Phi_t)$ is the difference equation for the parameters, q_t is the model noise. Therefore, the ensemble covariance for the prior is given by

$$\sigma_t^{pri} = \frac{1}{N-1} (A_t^- - \bar{A}_t^-)(A_t^- - \bar{A}_t^-)^T, \quad (1.4)$$

where $\bar{A}_t^-$ is the mean of ensemble members.

To update the estimated state and parameters of the system, the EnKF assimilates new observations y_t into the ensemble $\hat{Y}_t = y_t + \epsilon_t$, where ϵ_t is the perturbation vector. The assimilation step involves comparing the predicted states of the ensemble with the observed data and adjusting the estimated state and parameters of the system to better fit the observations

$$A_t = A_t^- + K_t(\hat{Y}_t - H_t A_t^-) \quad (1.5)$$

where $K_t = \sigma_t^{pri} H_t (H_t \sigma_t^{pri} H_t^T + \sigma_t^{obs})^{-1}$ is the Kalman gain, H_t is the measurement index matrix, $\sigma_t^{obs} = \frac{\epsilon_t \epsilon_t^T}{N-1}$.

Finally, the EnKF generates a new ensemble of model simulations by perturbing the updated estimated state and parameters of the system, and the process repeats for the next time step.

1.5 Multi-layer network construction

Constructing a multilayer network is an essential aspect of infectious disease modeling based on contact network models as it enables the integration of diverse interaction types that can influence disease transmission dynamics. A multilayer network consists of multiple layers or types of networks, representing various interaction types such as social contacts, transporta-

tion networks, or environmental contacts.

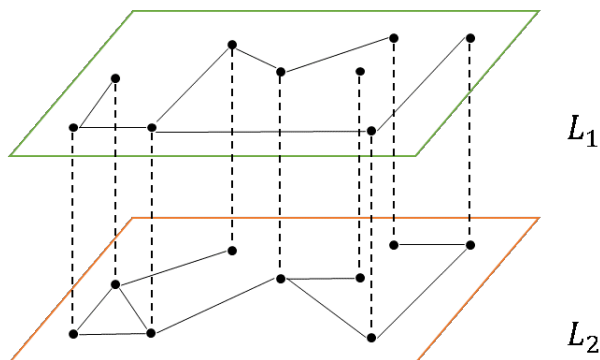


Figure 1.3: Two-layer network structure.

Fig. 1.3 illustrates a two-layer network where the same individuals are linked by a dashed line in different layers, and the edges within the same layer represent the transmission pathways of the pathogen between individuals. Various approaches can be used to construct each layer based on the research question and available data. For example, the social contact layer can be constructed using survey data on the frequency and duration of social interactions, while the transportation network layer can be created using data on traffic routes and volumes between different locations. In cases where data is not available, an Erdos Renyi random network can be used. These layers can be combined using a coupling function that represents the impact of different layers on the probability of infection given contact with an infectious individual. By integrating these diverse interaction types into a single multilayer network, the model can more accurately capture the complex dynamics of disease transmission and provide better guidance for effective public health interventions.

1.6 Research contributions

The dissertation addresses five research questions that delve into the forecasting and modeling of infectious diseases, contributing significantly to the field of epidemiology. Here's a summary of the contributions:

1. *Predicting COVID-19 Cases Amidst Disturbances:* We introduces a short-term forecasting system for COVID-19, utilizing the SEIR model enhanced by the EnKF for

assimilating daily case data, despite its noisiness. This innovative approach allows for the generation of accurate three-day forecasts, marking a significant advancement in real-time infectious disease surveillance and response planning.

2. *Describing Dengue Fever Transmission Dynamics:* A novel SEIR-SEI vector-borne disease model is proposed to accurately describe the transmission dynamics of dengue fever among humans and mosquitoes. This model's precision is validated against simpler models, showcasing its ability to provide detailed insights into dengue transmission mechanisms. This contribution aids in the development of targeted control strategies, potentially reducing the disease's impact.
3. *West Nile Virus (WNV) Transmission Dynamics and Climatic Influences:* We details a real-time prediction system for WNV that incorporates a adapted compartment model SEIR-SEI-SEAC to describe the transmission dynamics among birds, mosquitoes, and humans, including the role of climate variables. This system's ability to generate accurate weekly forecasts represents a substantial advancement in understanding and managing WNV outbreaks.
4. *FMD Transmission Impact and Mitigation Strategies:* The dissertation explores the complex transmission dynamics of FMD through a multi-layer network model, addressing the effects of direct and indirect contacts on disease spread. It particularly highlights the effectiveness of information-sharing strategies in outbreak mitigation, offering a novel approach to managing FMD outbreaks in livestock populations.
5. *Comprehensive Dengue Risk Information Aggregation:* Through the development of the PICTUREE—Aedes web application, We aggregate dengue-related data from multiple sources, including global daily temperature, global daily precipitation, *Aedes* mosquito occurrence, and dengue case occurrence. The PICTUREE—Aedes web application also provides temporal and spatial estimations of the *Aedes* mosquito population, the dengue transmission reproduction number, and the dengue risk globally. This platform employs advanced forecasting algorithms to predict ongoing dengue outbreaks.

In summary, this dissertation’s contributions lie in its development of advanced forecasting models and simulation tools for infectious diseases, enhancing our understanding of disease dynamics and providing valuable insights for public health intervention and policy-making. By addressing these critical research questions, the work advances the field of infectious disease modeling, offering innovative solutions to predict and mitigate the impacts of these diseases on public health.

1.7 Dissertation structure

The dissertation is structured as follows. Chapter 1 provides an overview of the background, motivation, research questions, and contributions of the study. In Chapter 2, we present a short-term forecasting system developed specifically for COVID-19, detailing its methodology and implementation. Chapter 3 introduces a vector-borne disease model, SEIR-SEI, designed to capture the transmission dynamics of dengue fever. To validate its accuracy, this model is compared with a simpler counterpart commonly used in the literature. Chapter 4 elaborates on a real-time prediction system for WNV, detailing an adapted compartment model that accounts for transmission dynamics among birds, mosquitoes, and humans, while considering the influence of climatic factors. In Chapter 5, we delve into the simulation of FMD transmission, examining the role of animal trade and vehicle sharing, and assessing the effectiveness of mitigation strategies, particularly information sharing, across different participation levels. Chapter 6 presents the development of the PICTUREE-Aedes web application, which serves as a comprehensive platform for aggregating dengue-related data, conducting risk simulations, and predicting ongoing dengue outbreaks. Lastly, Chapter 7 offers key conclusions and their limitations of this dissertation, and suggests future research endeavors.

Chapter 2

Short-term forecasts for the COVID-19 epidemic in Hubei Province, China

2.1 Introduction

The 2019 novel coronavirus disease (COVID-19) emerged in Wuhan City, Hubei Province of China, in the late fall of 2019⁵. The highly contagious virus outbreak during the Chinese Spring Festival period made containment an unprecedented challenge. The risks of human-to-human transmission, long persistence on surfaces, and asymptomatic infections increased spread and made case detection difficult^{6;7}. Increasing case reports saturated hospitals in Hubei Province, overwhelming medical resources. Despite the high case reports, the true infection rate was likely under-reported, as infected individuals with only mild symptoms might have recovered without being detected²³.

To stop the viral spread, a series of strict control measures have been implemented in mainland China, including the enforcement of the tourism ban, suspension of the school, and extension of the Spring Festival holiday. From January 23rd to April 8th, 2020, the Chinese government implemented a metropolitan-wide quarantine of Wuhan and nearby

cities, suspending all public transportation and advising residents to stay at home^{24;25}. Since February 8th, a door-to-door check was started in Wuhan to take all confirmed and suspected people into medical care, which ended around February 19th²⁶. Since February 10th, Wuhan City started to close the community.

Understanding the impact of non-pharmaceutical interventions on the epidemic dynamics is crucial for decision-makers to take proper actions to contain the COVID-19. Since the beginning of the outbreak, scientists have been actively retrieving the epidemiological characteristics such as the time from infection to illness onset (incubation period) and the time from illness onset to hospital admission or isolation (infectious period)^{5;27;28}. Researchers have used the reported cases during the early stage of the epidemic and predicted the epidemic trajectory, the basic reproductive number R_0 which ranges from 2.2 to 6.6^{26;29}.

China's aggressive and prompt measures have led to a dramatic decrease in the number of new cases. Those measures saved many lives, although at a high economic cost. As of March 18th, 2020, the cumulative number of confirmed cases in China was 81,186. Compared to the previous day, the number of new cases was only 70, most of which were imported from outside the country. In the meantime, COVID-19 spread further around the world, and parts of Italy, South Korea, the United States, and Iran struggled to contain the epidemic.

In this chapter, we aim to apply mathematical modeling to describe the epidemic dynamics of COVID-19, and use the ensemble Kalman filter for short-term forecasting of the cases reported in Wuhan City. Mathematical modeling can forecast the size and timing of epidemics, so resource allocations and decision-making can be optimized to maximize intervention strategies.

2.2 Model

To describe the epidemiology of COVID-19 pneumonia, we use the Susceptible-Exposed-Infectious-Removed (SEIR) compartmental model^{25;29;30} shown in Fig. 2.1, which divides the population into four compartments susceptible, exposed, infectious, or removed. Susceptible nodes refer to people free of the virus and can become infected due to contact with infectious

nodes. Exposed nodes have been infected by the disease, but have not yet shown symptoms or become infectious. Removed nodes represent those individuals who are recovered, isolated, hospitalized, or dead due to the disease. The equations for the state transitions are shown in equation 2.1.

$$\begin{aligned}
S'(t) &= -\beta IS \\
E'(t) &= \beta IS - \delta E \\
I'(t) &= \delta E - \gamma I \\
R'(t) &= \gamma I
\end{aligned} \tag{2.1}$$

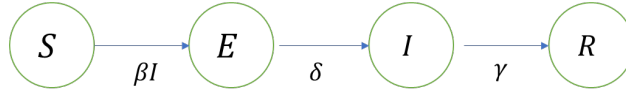


Figure 2.1: SEIR compartmental model

Here β is the infection rate; δ is the transition rate from the exposed compartment to the infectious compartment; γ denotes the removal rate. We assume that the infection process of each infectious individual is Poisson and statistically independent and that the removing process for every infected individual is also Poisson. Considering that the population flow, prevention measures, and self-quarantine could affect the infection rate, we take the infection rate β as a time-varying parameter. The other two rates δ and γ are set as constants 1/3 day-1 and 1/4 day-1³¹.

Short-term prediction is provided for the COVID-19 spreading in Wuhan city using the Kalman filtering, which consists of model state estimation at each observation time based only on the observation up to the present. Considering that there is an unknown time-varying parameter β in the SEIR model which cannot be dealt with by widely used classic Kalman filter³² and extended Kalman filter³³, a dual state-parameter estimation filtering method-ensemble Kalman filter (EnKF) is used for our study^{34;35}. EnKF is based on Monte Carlo, where the approximation of forecast (a priori) state error covariance matrix is made

by propagating an ensemble of model states using the updated states (ensemble members) from the previous time step.

Let the state variables ($SEIR$) be defined as x_k in R^4 , the time-varying parameter β in R . Consider the problem described in equation 2.2,

$$\begin{aligned} x'(t) &= f(x, \beta) + q \\ \beta'(t) &= h(\beta) \\ y(t) &= g(x, \beta) + v \end{aligned} \tag{2.2}$$

where q and v are model and observation noise which are assumed to be normally distributed and with zero mean and the covariance known.

As the true states are generally unknown, we use the ensemble mean as the best estimate of the true state. The error covariance matrices for the predicted and filtered ensemble are given by equation 2.3.

$$\begin{aligned} P^- &= E[(\hat{x}^- - E[\hat{x}^-])(\hat{x}^- - E[\hat{x}^-])^T] \\ P &= E[(\hat{x} - E[\hat{x}])(\hat{x} - E[\hat{x}])^T] \end{aligned} \tag{2.3}$$

where P^- is the error covariance matrix associated with the forecast estimate (prior), P is the one associated with the updated estimate (posterior), $\hat{x}^-$ is the ensemble vector of the forecast estimate from model equations. In the ensemble presentation, the model states x and the time-varying parameter β can be put together in a matrix A in $R^{5 \times N}$, holding N ensemble members.

$$A = \begin{pmatrix} x^1 & x^2 & \dots & x^N \\ \beta^1 & \beta^2 & \dots & \beta^N \end{pmatrix}$$

The ensemble covariance then can be presented as $P = \frac{1}{N-1}(\hat{A} - \bar{\hat{A}})(\hat{A} - \bar{\hat{A}})^T$, $\bar{\hat{A}} = \frac{1}{N}A \times 1_{N \times 1}$. The measurement can also be formulated in an ensemble presentation $Y = [y^1 y^2 \dots y^N]$, $\epsilon = [v^1 v^2 \dots v^N]$, $y^k = y + v^k$, y denotes the true measurement and v^k is the perturbation term.

The measurement covariance matrix can be estimated as $V = \frac{\epsilon \times \epsilon^T}{N-1}$

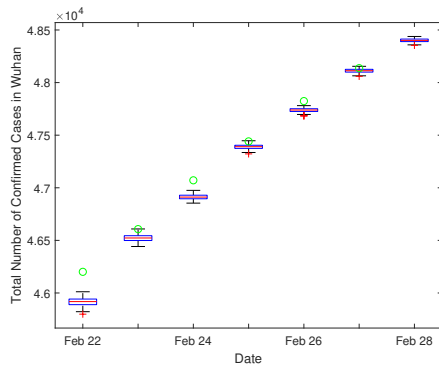
Specify the situation we are dealing with, $A = [SEIR\beta]^T$. The measurement is the total number of confirmed cases reported by the National Health Commission of China, which is compared with the number of removed populations in our model. Therefore, $y = C \times A$ where $C = [00010]$.

As the criteria for counting the confirmed cases were changed on February 12th, 2020, there is a jump in the reported number of confirmed cases. Taking this into account, we separate the whole period into period 1 (from January 26th to February 11th) and period 2 (from February 12th to February 27th) and run simulations in two steps. First, we run simulations for period 1 and estimate the infection rate β . Then, we take the estimated β on February 11th and the reported number of cases on February 12th as the initial β and initial removed population for period 2, respectively, and run simulations. The number of iterations is 250 in every predictive simulation.

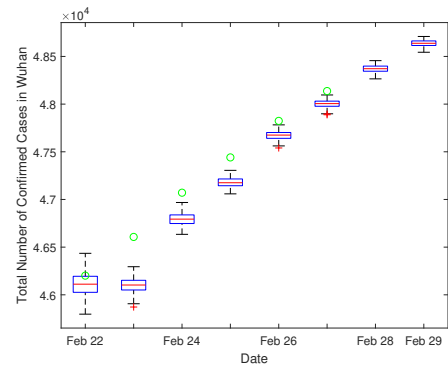
2.3 Results

The predictions of COVID-19 cases in Wuhan city from February 22nd to March 1st are conducted. The model noise and measurement noise are tuned by modifying the approach proposed by³⁶.

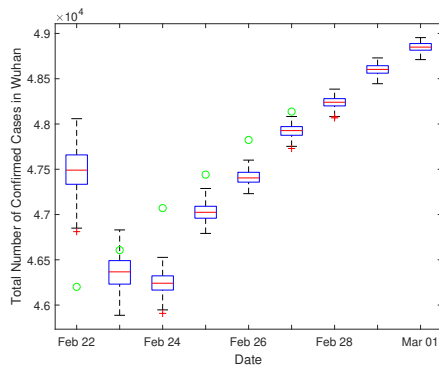
Fig. 2.2 shows the distributions of the predicted numbers of COVID-19 cases for 1-day-ahead, 2-day-ahead, and 3-day-ahead. The solid black lines indicate the maximum and minimum prediction results on each day. The blue box presents the range from the lower quartile to the upper one. The red line inside the box indicates the median of the samples. We can see that most predictions converge to a small range and that most medians are very close to the corresponding real data. Taking the median as our predicted number gives the relative error shown in Fig. 2.2d. The mean absolute percentage error (MAPE) for 1-day-ahead prediction is 0.24%; the MAPE for 2-day-ahead prediction is 0.52%; the one for 3-day-ahead is 1.23%. The performance of 1-day-ahead prediction is the best and the one of 3-day-ahead is the worst. When the n-day-ahead prediction is conducted, the results from



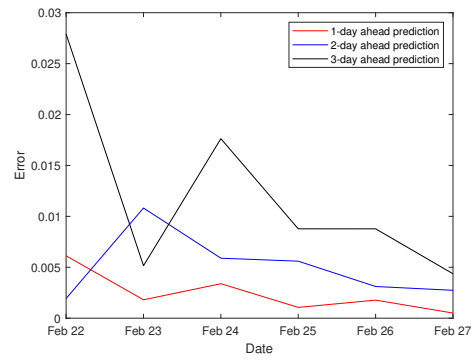
(a) 1-day-ahead prediction



(b) 2-day-ahead prediction



(c) 3-day-ahead prediction



(d) Relative prediction error

Figure 2.2: Distribution of the short-term prediction

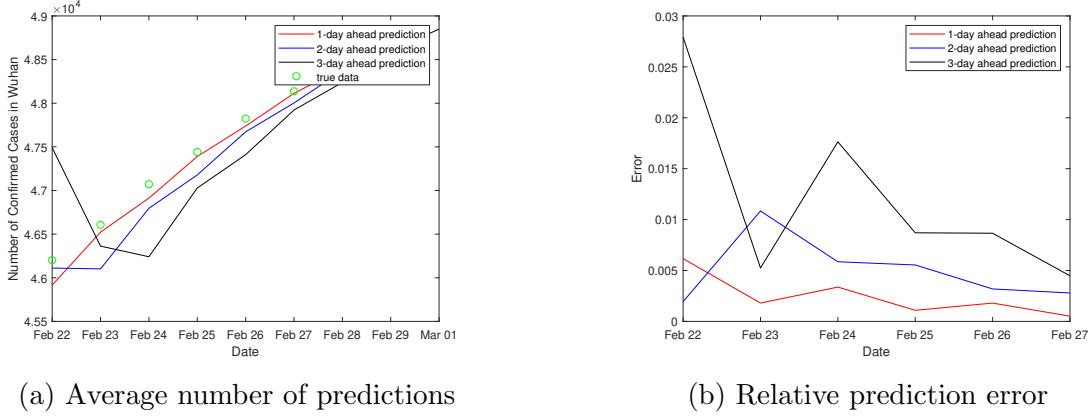


Figure 2.3: Average number as predictions and the error

(n-1)-day-ahead is used, so the errors accumulate.

If we choose the average number of the samples as our predicted number of COVID-19 cases, the results and their relative errors are presented in Fig. 2.3. The MAPE for 1, 2, 3-day-ahead predictions are 0.24%, 0.52% and 1.23%, which are equal to the ones when the medians are predicted numbers.

The self-quarantine policy was carried out on January 23rd in Wuhan and a series of administrative actions were taken to prevent the disease from spreading. Taking the latent period into consideration, we started to estimate the infection rate β from January 26th. On February 12th, Wuhan authorities changed the case definition to include symptomatic patients in addition to tested ones. Therefore, we present estimations of β for period 1 (from January 26th to February 11th) and period 2 (from February 12th to February 27th), which are presented as the blue line and the red line in Fig. 2.4. The orange line in Fig. 2.4 presents the removal rate. From the results, we see that the infection rate β is less than the removal rate γ after February 10th, which means the epidemic had peaked in Wuhan.

2.4 Summary

In this chapter, the ensemble Kalman filter is used to make short-term predictions of the COVID-19 cases in Wuhan City, which resulted in accurate forecasts of daily case reports. The model is able to predict daily cases and the epidemic peak. Knowing the daily cases

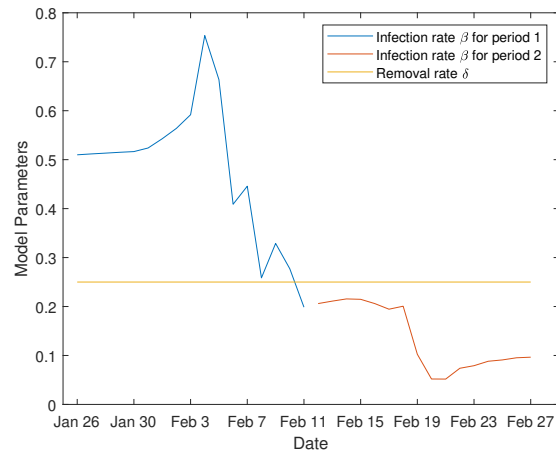


Figure 2.4: Estimated infection rate

from forecasts three days in advance allows for the proper resource allocation. The ensemble Kalman filter also allows parameter estimation, which is extremely useful for modeling purposes.

Chapter 3

Estimating and Forecasting Dengue Outbreak Dynamics

3.1 Introduction

Dengue fever is an acute mosquito-borne viral disease that has spread rapidly in tropical regions over the past decades³⁷. During large dengue outbreaks, public health officials and healthcare facilities often cannot deploy medical personnel, or administer treatment resources and emergency vector control measures in a timely and efficient manner. Therefore, an accurate incidence forecast and advance warning system of dengue outbreaks can reduce morbidity and mortality by proactively positioning resources or targeting mosquito control to high-risk areas forecasted to have increased cases. A study in 2017 showed that dengue is increasing at the highest rate (400% between 2000 and 2013) among communicable disease, which placed a heavy social, economic and health burden globally⁸. Worldwide, there is an estimation of 360 million dengue infections per year³⁸. In Singapore, large dengue outbreaks occurred almost every 5-6 years since the late 1980s despite the public health efforts⁹. Dengue is endemic in Brazil and more than 11 million cases was reported from 2000 to 2016¹⁰. In Taiwan, large dengue outbreaks were recorded in 1981 (about 13000 cases over 15000 residents on Liu-Chiu Islet, Pingtung), 1988 (approximate 5000 cases in Kaohsiung), 2014 (15429 cases

in Kaohsiung) and 2015 (43419 cases in Tainan)^{11;39}.

Recent literature has described a number of approaches to forecast infectious disease outbreaks. One class of methods are based on the mechanistic disease models which can be described by differential equations^{40–44}. Among the mathematical models appropriate for infectious disease, the Susceptible-Infectious-Recovered (SIR) model is the simplest one. In the SIR model, individuals are divided into 3 compartments and each compartment represents a state in the disease progression. Susceptible individuals can become infected and infectious by contact with infectious ones; people who recover from infection are presented by recovered⁴⁵. Based on the SIR model, the SEIR model introduces the new compartment E (exposed) describing the state when individuals are infected but not yet infectious. Other SIR-based models are also commonly used for disease forecast, like S-E-I-H (hospitalized)-F (funeral)-R model in⁴¹ to predict the 2014-2015 Ebola epidemics in Liberia, SIRS model for seasonal influenza prediction in⁴⁶. For vector-borne disease such as dengue fever whose infections occur alternatively between hosts (human) and vectors (mosquito), the SEIR-SEI (host-vector) model was developed and commonly used to simulate the transition dynamics of vector borne disease^{47–50}. However, to our knowledge, there is no previous research that uses the SEIR-SEI model to estimate the model parameters and to forecast the incidence.

Another class of methods based on machine learning techniques are used to predict the number of confirmed cases during epidemics^{51;52}. In order to have an accurate assessment about the epidemic size, health officials also need to have an estimation about the number of the infectious population who are not yet confirmed and the exposed population, which cannot be achieved by machine-learning approaches because machine learning models generally cannot describe the explicit relations between the observation (confirmed cases) and estimations (infectious and exposed population). However, the mechanistic models can be used to estimate all state variables, infer the disease transition rates, describe the outbreak characteristics, and provide understanding of the epidemic dynamics.

The classic Kalman filter (KF) was designed for data assimilation problems of linear system by R.E. Kalman in 1960⁵³. An early adaption of KF is the extended Kaman filter (EKF), which can solve the nonlinear problems by linearizing the nonlinear systems.

However, the usage of EKF on large-scale or severe nonlinear systems always requires high computational cost and closure approximation from neglecting higher order derivative terms of the model. The EnKF was introduced as a parameter and state estimation method by Evensen in 1994 to resolve the closure problems of EKF⁵⁴. In the next decades, the EnKF has been widely applied to many fields, like state and parameter estimation on reservoir models⁵⁵, hydrological models⁵⁶, and doubly fed induction generators⁵⁷. For dengue transmission, the inference of the model parameters is a difficult task because the characteristics and transmission patterns of dengue are influenced by environmental and meteorological factors and human interactions⁵⁸. EnKF is one of the most popular approaches to estimate the key parameters of disease models. The EnKF method in conjunction with an SEIR disease model was used to produce short-term prediction of Covid-19 cases in⁵⁹, and to develop a real-time forecast system of hand-foot-and-mouth disease outbreaks in⁴⁴. In addition, the estimation of the local female mosquito abundance⁶⁰, which plays a very important role on dengue transmission, has not been used for dengue incidence forecast.

The performance of six filtering methods were compared to forecast the seasonal influenza activity in⁶¹. Those methods were the basic particle filter (PF), the maximum likelihood estimation via iterated filtering (MIF), the particle Markov chain Monte Carlo (pMCMC), the ensemble Kalman filter (EnKF), the ensemble adjustment Kalman filter (EAKF), and the rank histogram filter (RHF). The EnKF, the EAKF, the RHF and the PF more accurately forecasted the outbreak than the MIF and the pMCMC because the ensemble filters (EnKF, EAKF, and RHF) and the PF are able to adjust the model parameters prior to each iteration. They concluded that the performance of PF degrades when a larger or more complex system was applied, and that the EnKF performed better than the EAKF because the EnKF is a stochastic filter while EAKF is purely deterministic.

The goal of this chapter was to develop a new model to estimate and forecast dengue outbreak dynamics by integrating the SEIR-SEI disease model with the EnKF assimilation method. At first, we designed twin experiments where the SEIR-SEI-EnKF model assimilated the synthetic weekly dengue cases and produced accurate estimations and predictions of states and parameters. Second, the model estimated and forecasted the dengue outbreaks

in Kaohsiung Taiwan, Singapore, and Rio de Janeiro Brazil by assimilating the historical time-series incidence. The estimated and predicted cases corresponded to the observed cases and the model was able to provide estimations of the exposed and infectious human population dynamically. Furthermore, we performed a comparison on the forecast accuracy of the SEIR-SEI-EnKF, SEIR-EnKF, and SIR-EnKF models. The results showed the SEIR-SEI-EnKF model had the most accurate forecasts. In this study, the EnKF and the SEIR-SEI disease model were integrated for the first time to estimate and forecast the epidemic dynamics for vector-borne disease like dengue fever, the model was able to provide parameter and state estimation and dengue case forecast, and the model demonstrated to have more accurate forecasts than the other two commonly used models. This dengue epidemic estimation and forecast model will potentially help healthcare agencies prepare and adapt to the changing situation during dengue outbreaks.

3.2 Method

3.2.1 Data

In this study, weekly dengue incidence data in Kaohsiung Taiwan (population 2,773,533) from epidemic week 27/2015 to 5/2016⁶², Singapore (population 5,913,813) from epidemic week 1/2020 to 50/2020⁶³ and Rio de Janeiro Brazil (population 6,023,699) from epidemic week 7/2015 to 42/2015⁶⁴ were used.

In this chapter, we presented the estimation and forecast results in Kaohsiung Taiwan, and the results in Singapore and Rio de Janeiro were presented in the appendix A. Since the reported dengue cases were the number of people who were diagnosed and hospitalized, we assumed that they were not able to infect others. Therefore, we used the cumulative weekly case count as the observed recovered population in this study.

Dengue is transmitted by the bite of adult female mosquitoes. The regional number of female mosquitoes is greatly influenced by climate conditions like precipitation and temperature. Mosquito population modeling⁶⁵ used the regional environmental data to estimate the

local female mosquito abundance. The estimated female mosquito abundance of Kaohsiung was generated this way.

3.2.2 Disease model

In order to describe the transmission of dengue virus between host and vector population groups, the SEIR-SEI disease model is used. We consider the populations involved in disease transmission, which include humans of all ages and both sexes and adult female mosquitoes. The population of each species is divided into different compartments, which are referred to as state variables. Compartments for the human population are 1) susceptible (S_H), defined as the population that is unexposed to the virus but can be infected by the bites from the infectious mosquitoes; 2) exposed (E_H), defined as the population that is infected but not infectious; 3) infectious (I_H); 4) recovered (R_H), defined as the population that is recovered or died. Compartments for the vector population are 1) susceptible (S_M); 2) exposed (E_M); 3) infectious (I_M). Once the mosquitoes become infectious, they are assumed to remain so until death.

Considering that the transition from the susceptible to the infection for each species depends on the biting rate of the mosquitoes and the transmission probability. The infection rates per susceptible human and susceptible mosquito are given by

$$p_H \frac{N_M}{N_H} b \frac{I_M}{N_M} = p_H b \frac{I_M}{N_H} = \beta_H \bar{I}_M \quad (3.1)$$

$$p_M b \frac{I_H}{N_H} = p_M b \frac{I_H}{N_H} = \beta_M \bar{I}_H \quad (3.2)$$

Where N_H and N_M denote the human and mosquito population size respectively. The biting rate b is the average number of bites per mosquito on humans per day, which is influenced by a number of factors, like climate conditions, mosquito longevity, and so on; the transmission probability p is the probability that the susceptible can be infected per infectious bite. $\beta_H = p_H b, \beta_M = p_M b$ are the infection rates for each species. $\bar{I}_M = \frac{I_M}{N_H}$, $\bar{I}_H = \frac{I_H}{N_H}$ are the normalized values of the infectious population.

Transitions between compartments shown in Fig. 3.1 are described by equation(3.3).

$$\begin{aligned}
\bar{S}'_H(t) &= \lambda_H - \beta_H \bar{S}_H \bar{I}_M - \mu_H \bar{S}_H \\
\bar{E}'_H(t) &= \beta_H \bar{S}_H \bar{I}_M - (\delta_H + \mu_H) \bar{E}_H \\
\bar{I}'_H(t) &= \delta_H \bar{E}_H - (\gamma_H + \mu_H) \bar{I}_H \\
\bar{R}'_H(t) &= \gamma_H \bar{I}_H - \mu_H \bar{R}_H \\
\bar{S}'_M(t) &= \bar{A} - \beta_M \bar{S}_M \bar{I}_H - \mu_M \bar{S}_M \\
\bar{E}'_M(t) &= \beta_M \bar{S}_M \bar{I}_H - (\delta_M + \mu_M) \bar{E}_M \\
\bar{I}'_M(t) &= \delta_M \bar{E}_M - \mu_M \bar{I}_M
\end{aligned} \tag{3.3}$$

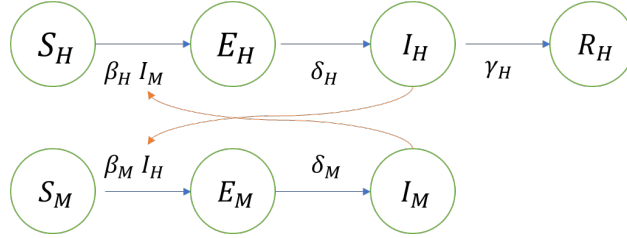


Figure 3.1: SEIR-SEI compartmental model

Where λ_H denotes the birth rate for humans; μ denotes the death rate for both species; δ denotes the transition rate from exposed to infectious; γ denotes the transition rate from infectious to recovered; $\bar{A}$ denotes the normalized mosquito emergence rate. In this study, the birth rate and death rate for humans are approximated as 0, $\lambda_H = 0$, $\mu_H = 0$. A report from WHO shows that the incubation time of dengue on human is 4 – 10 days and the infection duration is 2 – 7 days⁶⁶, so the rates δ_H and γ_H satisfy the following distributions $\frac{1}{\delta_H} \sim N(7, 1.5)$, $\frac{1}{\gamma_H} \sim N(4.5, 1.25)$; the average lifespan of an *Aedes* mosquito in nature is two weeks and the incubation time in the mosquito is about 10 days, so the death rate and incubation rate for mosquito satisfy $\frac{1}{\mu_M} \sim N(14, \sigma)$, $\frac{1}{\gamma_H} \sim N(10, \sigma)$.

The number of new cases at time t is given by $\bar{N}C_H(t) = \bar{R}_H(t) - \bar{R}_H(t-1)$.

The other two simpler disease models – SIR model and SEIR model, are often used for

dengue incidence forecast. The SIR model is expressed in the following equations (3.4).

$$\begin{aligned}
\bar{S}'_H(t) &= \lambda_H - \beta_H \bar{S}_H \bar{I}_H - \mu_H \bar{S}_H \\
\bar{I}'_H(t) &= \beta_H \bar{S}_H \bar{I}_H - (\gamma_H + \mu_H) \bar{I}_H \\
\bar{R}'_H(t) &= \gamma_H \bar{I}_H - \mu_H \bar{R}_H
\end{aligned} \tag{3.4}$$

The equations for the SIER model are shown in (3.5).

$$\begin{aligned}
\bar{S}'_H(t) &= \lambda_H - \beta_H \bar{S}_H \bar{I}_H - \mu_H \bar{S}_H \\
\bar{E}'_H(t) &= \beta_H \bar{S}_H \bar{I}_H - (\delta_H + \mu_H) \bar{E}_H \\
\bar{I}'_H(t) &= \delta_H \bar{E}_H - (\gamma_H + \mu_H) \bar{I}_H \\
\bar{R}'_H(t) &= \gamma_H \bar{I}_H - \mu_H \bar{R}_H
\end{aligned} \tag{3.5}$$

The variables in (3.4) and (3.5) denote the same states or parameters as they do in the SEIR-SEI model. In the transition process of both models, the intervention of mosquitoes is completely ignored.

3.2.3 Forecast model

In order to estimate the unknown parameters and the state variables, we used an efficient data assimilation algorithm – the Ensemble Kalman Filter (EnKF). The EnKF maintains an ensemble of the state variables and the unknown parameters. Let the state variables defined as $x_k \in R^{n \times 1}$, the unknown parameters $\Phi_k \in R^{m \times 1}$, where n is the number of states variables, and m is the number of the unknown parameters. In the ensemble presentation, they can be put together in a matrix $A_k \in R^{(n+m) \times N}$, holding N ensemble member at time k

$$A_k = \begin{bmatrix} x_k^1 & x_k^2 & \dots & x_k^N \\ \phi_k^1 & \phi_k^2 & \dots & \phi_k^N \end{bmatrix} \tag{3.6}$$

The prior at time $k + 1$ $\hat{A}_{k+1}^-$ is given by the equation (3.7)

$$\hat{A}_{k+1}^- = f(\hat{A}_k) + q_k, \quad (3.7)$$

where $\hat{A}_k$ is the posterior at time k , f is the model system, and q_k is the model noise.

In this method, the ensemble mean $\bar{\hat{A}}_k^-$ is the best estimate of the true states. Therefore, the ensemble covariance for prior at time k P_k^- is given by equation (3.8).

$$P_k^- = \frac{1}{N-1}(\hat{A}_k^- - \bar{\hat{A}}_k^-)(\hat{A}_k^- - \bar{\hat{A}}_k^-)^T \quad (3.8)$$

The ensemble presentation of measurements can be formulated as

$$Y_k = \begin{bmatrix} y_k^1 & y_k^2 & \dots & y_k^N \end{bmatrix} = y_k + \epsilon_k, \quad (3.9)$$

where y_k is the true measurement and ϵ_k is the perturbation vector.

$$Y_k = H_k A_k \quad (3.10)$$

H_k is the measurement index matrix at time k . The measurement covariance matrix V_k at time k can be estimated as

$$V_k = \frac{\epsilon_k \epsilon_k^T}{N-1}. \quad (3.11)$$

The Kalman gain is given by

$$G_k = P_k^- H_k^T (H_k P_k^- H_k^T + V_k)^{-1}. \quad (3.12)$$

The posterior and its covariance are then updated by

$$\hat{A}_k = \hat{A}_k^- + G_k(Y_k - H_k \hat{A}_k^-), \quad (3.13)$$

Table 3.1: Preset parameters Θ and initial conditions $\bar{x}_0$ of SEIR-SEI model

Θ	β_H	δ_H	γ_H	$\bar{A}$	β_M	δ_M	μ_M
Values	2.5, 1.5	0.7	0.75	0.2, 0.15	2.5, 1.5	0.5	0.5
$\bar{x}_0$	$\bar{S}_{H0}$	$\bar{E}_{H0}$	$\bar{I}_{H0}$	$\bar{R}_{H0}$	$\bar{S}_{M0}$	$\bar{E}_{M0}$	$\bar{I}_{M0}$
Values	0.99995	0.00002	0.00003	0	0.1	0.0005	0

$$P_k = \frac{1}{N-1}(\hat{A}_k - \bar{\hat{A}}_k)(\hat{A}_k - \bar{\hat{A}}_k)^T. \quad (3.14)$$

In this analysis, the ensemble mean of the prior $\bar{\hat{A}}_k^-$ is our forecast at time k , and the ensemble mean of the posterior $\bar{\hat{A}}_k$ is the estimation of the states and parameters.

3.2.4 Twin experiments

The twin experiments are an effective method to verify the feasibility and evaluate the accuracy of parameter estimation techniques⁶⁷. In the twin experiments, a disease model with a prescribed value of parameters is simulated to produce time-series observations of state variables. Then, we take the parameters of interest as unknown ones and use the “observed states” as measured data to estimate those unknown parameters by EnKF. Therefore, by comparing the estimated value and the prescribed value of the parameters, we can test the correctness and evaluate the accuracy of the estimation method.

In this chapter, synthetic time-series of dengue epidemic dynamics and mosquito abundance are generated by an SEIR-SEI model whose parameters and initial states condition are listed in Table 3.1. In this simulation, the prescribed values for β_H , β_M , and $\bar{A}$ are piece-wise constant, which means their values have a step change at week 26. This action is aim to mimic the dynamic characteristics of these parameters. We add white noise (zero mean and 0.0001 variance) to the synthetic time-series data of the state $\bar{R}_H$ (total confirmed dengue cases) and the sum of $\bar{S}_M$, $\bar{E}_M$ and $\bar{I}_M$ (mosquito abundance). Then, they are input to the EnKF as the observations. The estimated parameters and states by EnKF will be compared with the synthetic model to evaluate the performance of the EnKF.

3.2.5 Estimations and forecasts of the SEIR-SEI-EnKF model

The weekly series of dengue cases (30 weeks in total) and the mosquito abundance in Kaohsiung Taiwan is used to train and test the SEIR-SEI-EnKF model. The training period is from week 1 to week 22. The system assimilates the observations during the training period to estimate the parameters and states. The predictions are produced at the end of the training period for the next 5 weeks, and the reported dengue cases from week 23 to week 27 are used to evaluate the prediction accuracy. The simulation is replicated 200 times, each model state is initialized with a positive random value at the beginning of each simulation. The average estimations and the distribution of the predictions are presented in the results.

3.2.6 Real-time predictions of different disease models

Three disease models – SEIR-SEI, SEIR, and SIR model are applied to the EnKF separately to produce the real-time predictions of the dengue cases in Kaohsiung Taiwan. The prediction and the assimilation occur at the same time i.e. the systems produce predictions for the next 5 weeks at each week after the observation is updated. Thus, 5 steps of predictions are produced in this simulation (predictions in step 1 are the predictions for the next week, predictions in step 2 are the ones in 2 weeks, etc.). Simulations for each model are repeated 200 times. The root mean square (RMS) errors between the predictions and the corresponding observations are calculated.

3.3 Results

3.3.1 Twin experiments

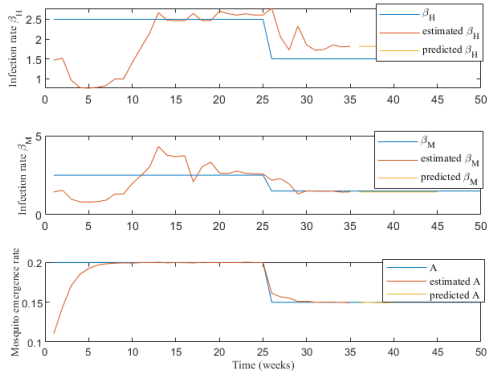
Fig.3.2a shows the estimated value and prescribed value of unknown parameters. In the figure, the estimated infection rate from mosquito to human β_H converges to the prescribed value after week 13 and follows the change at week 26 (1-week delay) and finally becomes stable at 1.8 (20% error); the estimation of β_M has a large overshoot and oscillations but it

converges to the first-stage prescribed value after week 20, the EnKF detects its step decrease at week 26 and the curve converges after week 30; the estimated value of mosquito emergence rate $\bar{A}$ matches exactly with the prescribed one after week 7 and EnKF can detect its step change from the observation immediately.

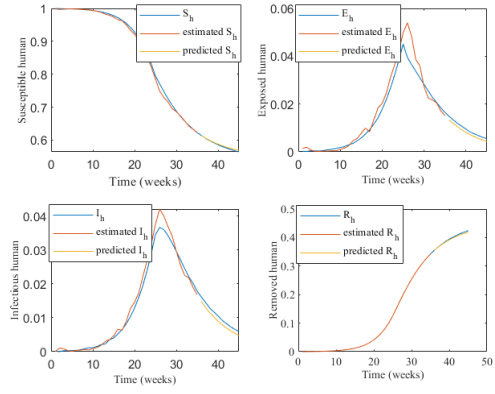
Figs.3.2b, 3.2c show the estimated and predicted epidemic dynamics on both populations. The blue curve in each graph shows the “real” epidemic dynamics (week 1 to week 45); the red curve shows the estimated ones (week 1 to week 35); the orange curve shows the predicted ones (week 36 to week 45). For the human population, the value of $\bar{R}_H$ is the observation and the value of other states is unknown. As shown in Fig.3.2b, the estimated and predicted values of all states are very close to their real value except the peak values of states $\bar{E}_H$ and $\bar{I}_H$. In addition, the estimated epidemic peak week (the week when $\bar{I}_H$ reaches the maximum) matches the real one. The errors for the peak value of states $\bar{E}_H$ and $\bar{I}_H$ are 20% and 13.5%. For the mosquito population, the total number of female mosquitoes is observed. As seen in Fig.3.2c, the estimations and predictions on states $\bar{S}_M$ and $\bar{E}_M$ are almost accurate. For state $\bar{I}_M$, the error is relatively large, but the estimation captures the trend of real data.

3.3.2 Estimations and forecasts of SEIR-SEI-EnKF model

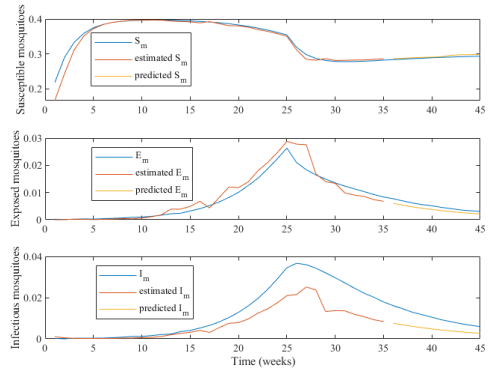
The estimated weekly infection rates of dengue for both species and the mosquito emergence rate are shown in Fig.3.3a. Dengue cases were detected from the first week but the infection rates are very low in the first 10 weeks. During the high-incidence period (week 10 to week 22), the infection rates for both species keep increasing to their peak value at week 21 (1.52 for β_H and 5.65 for β_M); the mosquito emergence rate A has two “steady states” – around 3.3×10^{-4} before peak weeks and around 1.8×10^{-4} after. The value of infection rate β_M is much greater than β_H , which means mosquitoes are more likely to be infected by infectious humans. The estimations and the forecasts of the model states are presented in Fig.3.3b, 3.3c. On the graph, the dots are estimated values of each state; the boxes are the distribution of the predicted value over 200 repeated simulations (the red line inside the box is the median, the upper and lower edges are the quantiles); the solid line is the observed recovered



(a) Parameter estimation



(b) Epidemic dynamics on human population



(c) Epidemic dynamics on mosquito population

Figure 3.2: Estimation and prediction results for twin experiments

Table 3.2: statistics of the predictions

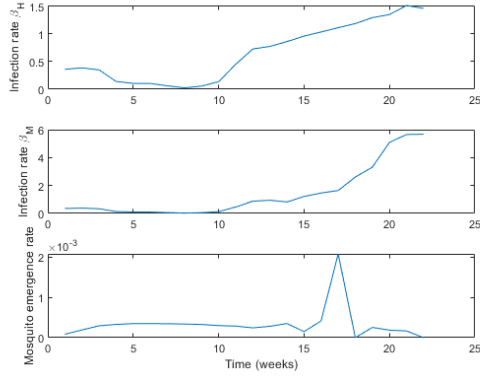
Predictions	Week 23	Week 24	Week 25	Week 26	Week 27
Upper quantile	17550	18426	18926	19231	19407
Median	17469	18217	18641	18876	18978
Lower quantile	17374	18040	18373	18514	18067
Real value	17542	18439	19066	19428	19682

human population. In Fig.3.3b, the estimated recovered human population perfectly matches the observed data. The statistics of the predictions are listed in Table 3.2. Taking the median as our best forecast, the root mean square (RMS) error is 455 which is over 40 times smaller than the real values. The estimated peak week for the exposed human population and infectious human population is week 21. The estimated normalized magnitude of the infectious population is 9.8×10^{-4} which means the estimated maximum infectious population is 2715. The epidemic dynamics of the mosquito population are shown in Fig.3.3c. The estimated number of infectious mosquitoes is very high from week 10 to week 22, which corresponds to the high-incidence period for the human population.

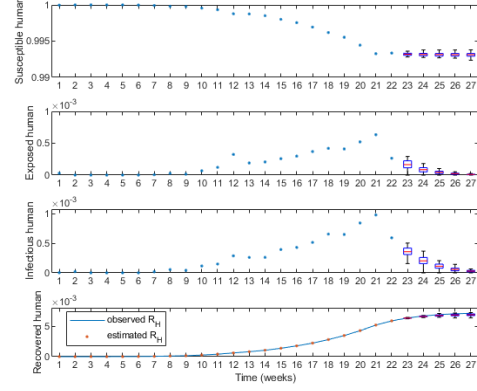
3.3.3 Real-time predictions of different disease models

We further test the prediction performance of the SEIR-SEI-EnKF model by comparing its real-time prediction results with the SEIR-EnKF model and the SIR-EnKF model. Each model assimilates observations each week and produces real-time predictions in the next 5 weeks.

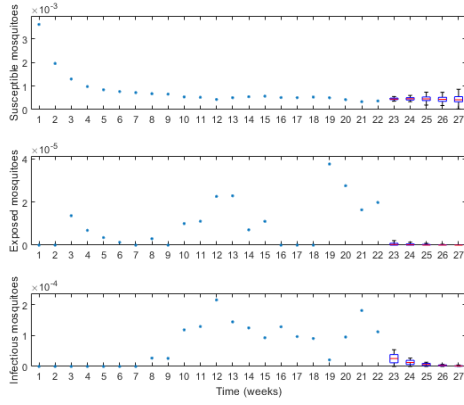
Fig.3.4 shows the prediction results of the total confirmed cases, which is the predicted recovered human population. The 1-step prediction refers to the predictions for the next week of the observation; the 2-step prediction is the predicted cases in 2 weeks of the observation week, etc. In the figure, dots present observations, and the curves present the predictions of different models. Overall, the first 3 predictions of all 3 models correspond well with the observations. In the step 4 prediction and the step 5 prediction, the SEIR-SEI-EnKF model underestimates the cases while the SEIR-EnKF model and SIR-EnKF model overestimate that.



(a) Parameter estimation



(b) Epidemic dynamics on human population



(c) Epidemic dynamics on mosquito population

Figure 3.3: Estimation and prediction results for SEIR-SEI-EnKF model

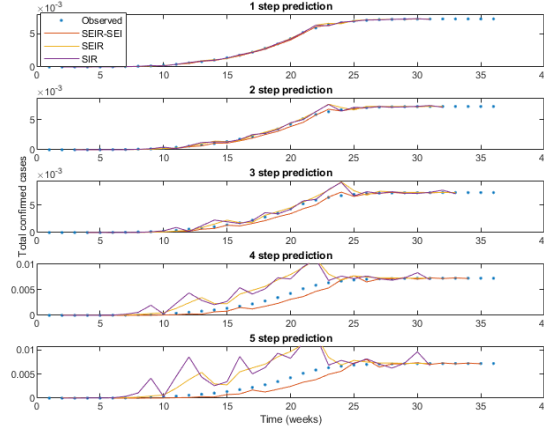


Figure 3.4: Real-time predictions of total confirmed cases

The real-time predictions of the weekly new confirmed cases are presented in Fig.3.5. The new confirmed cases are given by $NC_H^i = R_H^i - R_H^{i-1}$, which is not a state of the disease model. Meanwhile, the reported weekly cases are not the model observation either. The relative errors of the weekly case predictions are greater than that of the confirmed case because the EnKF adjusts the model parameters and calculates the Kalman gain by the error between the prior and the corresponding observation which is the total confirmed cases. The step 4 and the step 5 predictions of each model are not able to reflect the weekly new confirmed cases. The peak week of the reported weekly confirmed cases is week 21; the predicted peak weeks of the SEIR-SEI-EnKF model are week 22 for step 1, week 22 for step 2, and week 23 for step 3; those of the SEIR-EnKF model and SIR-EnKF model are week 22 for step 1, week 23 for step 2, and week 24 for step 3. The peak value of the reported new confirmed cases is 2569 and Table 3.3 lists the predicted peak values of new confirmed cases for each model and their absolute relative errors (ARE) with the reported one. The results show that the models can detect the peak of the newly confirmed cases by observing the total confirmed cases and that the SEIR-SEI-EnKF model has a more accurate performance in predicting the peak time and peak value of the weekly confirmed cases.

To evaluate the prediction performance more objectively, we calculate the RMS error of each prediction. Fig.3.6 shows the RMS errors over each prediction step for each model and there are three evident patterns: (1) the 1-step prediction errors for all 3 models are very

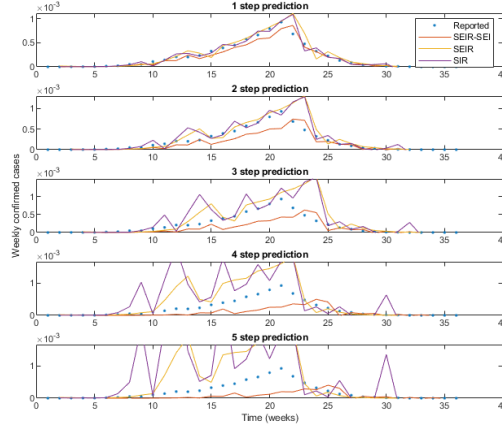


Figure 3.5: Real-time predictions of newly confirmed cases

Table 3.3: Peak value of predicted new cases and their absolute relative errors (ARE)

Peak/ARE	1 step	2 step	3 step	Average ARE
SEIR-SEI	2382/7.3%	2054/20%	1725/32.9%	20.1%
SEIR	3070/19.5%	3624/41.1%	4299/67.3%	42.7%
SIR	3016/17.4%	3550/38.2%	4200/63.5%	39.7%

close, which means that all 3I models can have very good performance when we only forecast the cases in the next week; (2) for each model, the errors over the earlier prediction steps are smaller; (3) the prediction errors for SEIR-SEI-EnKF model are smaller than those for SEIR-EnKF model and SIR-EnKF model. Moreover, the errors for the SIR-EnKF model increase exponentially with the prediction weeks. Therefore, when we try to forecast the dengue cases in the long term, the SIR model is not robust and reliable. The prediction errors from the SEIR-EnKF model are smaller than the SIR model but greater than the SEIR-SEI model, and they increase faster than the SEIR-SEI-EnKF model.

3.4 Discussion

In this chapter, we developed a new SEIR-SEI-EnKF model for estimating and forecasting dengue outbreak dynamics. The model estimated the model parameters and states and dynamically produced accurate short-term forecasts by assimilating the observations.

Dengue fever is a mosquito-borne viral disease that is dependent on mosquito and hu-

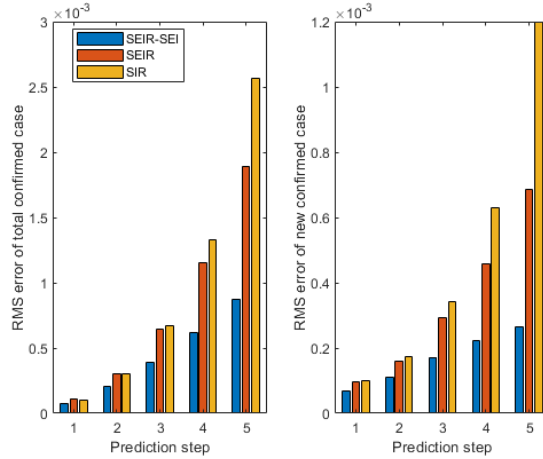


Figure 3.6: RMS errors of real-time predictions

man interactions. Consequently, high mosquito populations are correlated with elevated viral transmission and infection rates. In disease transmission, the infection rates and the mosquito abundance are essential factors that have a great influence on the overall epidemic size. Weather conditions, such as warmer temperatures and increased precipitation, are correlated to increased viral transmission, whereas human interventions (mosquito management) reduce contact with mosquitoes which reduces transmission. In our disease model, the three estimated parameters (two infection rates β_H, β_M , and the mosquito emergence rate A) were taken as small positive random constants initially and adjusted during assimilation by EnKF. The dependence of these parameters on weather added error to the estimations if local weather data was not available.

The SEIR-SEI model generated more accurate forecasts because it accounted for the mosquito population when forecasting dengue transmission. Another reason for the increased accuracy was that the SEIR-SEI model had a higher degree of freedom than the other two models; thus, the SEIR-SEI model had more flexibility to fit the data. Higher degree systems become over-fitted when more observations are fed into the systems. Our forecasts in three cities demonstrate SEIR-SEI model is the better-fitted one than the SIR and SEIR models.

Based on the simulations of the synthetic data and the simulations of the historical time-serial data, the SEIR-SEI-EnKF model is reliable at estimating and predicting dengue outbreak dynamics. During an ongoing dengue outbreak, this model can produce incidence

predictions and estimate the current exposed and infectious populations, which is required for public health agencies to optimize the treatment resources and take control measures in advance.

There are several limitations in our study. First, the collected dengue case data did not include the infected population who did not seek medical assistance, which rendered the reported dengue cases the lower bound of the actual case. Further, the estimated parameters (two infection rates β_H, β_M and the mosquito emergence rate A) might be better estimated using available weather data and this will be one of our limitations and future work. This study did not compare all possible adaptations of disease models; there might be an adaptation of the SEIR-SEI model which has a more accurate forecast than the SEIR-SEI model.

Chapter 4

Real-time forecasting and estimating for West Nile virus

4.1 Introduction

West Nile virus (WNV) is a mosquito-borne arbovirus first identified in the West Nile District of Uganda in 1937. Since then, WNV has spread to other parts of the world, including Europe, the Middle East, North America, and West Asia, and received widespread attention⁶⁸. The United States has experienced seasonal outbreaks of WNV since its initial detection in New York City in 1999¹². Some years have seen a particularly high number of reported cases. For example, in 2003, there were over 9,800 reported cases of WNV in the US¹³. The year 2012 was marked by the highest number of severe cases and deaths, with over 5,600 total infections, 2,873 neuroinvasive cases, and 243 deaths¹⁴. Recent data from the Centers for Disease Control and Prevention (CDC)¹⁵ indicate that in 2023, there were 2,406 detected WNV cases in the US, with Colorado reporting the highest number of cases (631) among states. As WNV outbreaks occur annually in the US, long-term predictions of their timing, magnitude, and duration are crucial for public health officials in resource allocation and implementing targeted prevention and control measures. Additionally, accurate parameter estimation and short-term real-time prediction of ongoing outbreaks are essential for imme-

mediate response and intervention, resource optimization, and predictive model refinement.

Recently, forecasting the outbreaks and severity of WNV and other infectious diseases has become a critical area of research. To achieve this, a diverse range of approaches has been developed. These include statistical methods such as time series analyses⁶⁹, logistic regression models⁷⁰, and machine-learning approaches⁷¹, that focus on fitting historical observations to identify risk factors and predict outbreak likelihood. Additionally, mechanistic models that utilize the susceptible–infected–recovered (SIR) framework cooperating with data assimilation methods are also employed to estimate and forecast the transmission dynamics of disease outbreaks^{72;73}. The CDC’s WNV Forecasting Challenge is an example of improving public health response to WNV through predictive models. In this challenge, participating teams are provided with annual West Nile Neuroinvasive Disease (WNND) cases from previous years to forecast annual WNND cases for each county in the contiguous United States using a variety of modeling approaches^{71;74;75}. However, the task of year-long forecasting is complicated by global warming, which leads to significant yearly variations in local temperature and precipitation. Numerous studies have underscored how these environmental factors influence WNV transmission^{76–78}. This variability presents challenges in accurately capturing the outbreak dynamics in regions experiencing erratic weather conditions. To address this issue, the development of a real-time forecasting system could provide a solution, offering more immediate and adaptive predictions to guide public health interventions.

For the real-time forecasting of infectious disease, a model-inference framework that consists of an epidemiological model, up-to-date surveillance data and a data assimilation method has been widely used^{79;80}. In this framework, the epidemiological model is used for understanding the spread of infectious diseases within a population, the SIR (Susceptible - Infectious - Recovered) model is a fundamental one⁸¹. Given its simplicity, the SIR model has been adapted and extended in various ways to better capture the complexity of real-world disease transmission. For instance, the SEIR model adds an ‘Exposed’ category to account for incubation periods, offering insights into COVID-19 and Ebola outbreaks^{82;83}. Similarly, the SEI-SEIR model adapts to vector-borne diseases like dengue, illustrating the intricate interplay between human and mosquito populations⁸⁴. The transmission cycle of

West Nile Virus (WNV), involving multiple hosts and vectors—birds, mosquitoes, and humans—necessitates a more complex modeling approach. A specialized compartmental model, denoted as $S_b E_b I_b R_b - S_m E_m I_m - S_h E_h I_h R_h$, is proposed to capture the unique epidemiological characteristics of WNV in⁸⁵. The data assimilation method is used to integrate new surveillance data into the epidemiological model to update predictions and reduce uncertainties. Techniques such as the ensemble Kalman filter (EnKF)^{84;86}, particle filter (PF)^{87;88} are commonly used for this effort. The EnKF, by updating predictions with each new piece of information, significantly refines the accuracy of short-term forecasts and contributes to a more comprehensive understanding of disease transmission dynamics.

This chapter enhances the existing methodology for real-time prediction of WNV. We first introduce an adapted compartment model that describes the transmission dynamics of WNV among birds, mosquitoes, and humans, incorporating both the infected asymptomatic human population and the significant impact of climatic factors such as temperature and precipitation on disease spread. We further assimilate reported human WNV cases from Colorado in 2023 into our model using the EnKF for robust data assimilation. This integration enables us to generate weekly forecasts of human WNV cases throughout the 2023 season. Such forecasts are invaluable for public health planning and intervention strategies. Moreover, our forecasting system estimates the epidemic dynamics of each involved population and calculates infection rates, providing critical insights like the estimated numbers of asymptomatic individuals and infectious mosquitoes. The precision of our forecasts is rigorously evaluated using the absolute root mean square error of the average forecasts and the logarithmic score of the forecasted new cases. By conducting a comparative analysis with models that omit climatic variables, we demonstrate that incorporating temperature and precipitation significantly enhances the accuracy of our forecasts. This finding underscores the crucial role of environmental factors in predicting WNV transmission and highlights the potential for improved public health responses through the integration of weather data into epidemiological modeling.

4.2 Materials and Methods

4.2.1 Data

The data utilized in our analysis is sourced from the following references: for Colorado West Nile Virus cases, we relied on data provided by the Colorado Department of Public Health and Environment, available at⁸⁹; for United States weather data, we accessed comprehensive weather information from⁹⁰. These reliable sources ensure the accuracy and credibility of the data used in our analysis.

4.2.2 Disease model

The primary transmission cycle of WNV is between birds and mosquitoes. Many bird species are reservoir hosts for the virus. When a mosquito feeds on an infected bird, it can acquire the virus. Humans, along with other mammals like horses, are considered incidental hosts. When an infected mosquito bites a human, the virus can be transmitted. Modeling of West Nile virus is complicated by a variety of factors: WNV has been found in 138 bird species within the United States, with susceptibility, transmissibility, and infectious period varying between species; multiple mosquito vectors may be responsible for transmission. A compartment model equation 4.1 containing five basic elements: house sparrows (as the reservoir bird species), crows (as a sentinel bird species), ornithophilic and generalist (opportunistic) mosquitoes, and humans is proposed in⁸⁵. The populations are subdivided into S, E, I, and R corresponding to susceptible, exposed, infectious, and recovered, and use the subscripts S, C, O, G, and H, to refer to sparrows, crows, ornithophilic mosquitoes, generalist mosquitoes,

and humans:

$$\begin{aligned}
\bar{S}'_B(t) &= v_B - (r_O\beta_{OB}\bar{I}_O + r_G\beta_{GB}\bar{I}_G)\bar{S}_B - \mu_B\bar{S}_B \\
\bar{E}'_B(t) &= (r_O\beta_{OB}\bar{I}_O + r_G\beta_{GB}\bar{I}_G)\bar{S}_B - (\delta_B + \mu_B)\bar{E}_B \\
\bar{I}'_B(t) &= \delta_B\bar{E}_B - (\gamma_B + \mu_B)\bar{I}_B \\
\bar{R}'_B(t) &= \gamma_B\bar{I}_B - \mu_B\bar{R}_B \\
\bar{S}'_M(t) &= v_M - (r_m\beta_{SM}\bar{I}_S + r_m\beta_{CM}\bar{I}_C)\bar{S}_M - \mu_M\bar{S}_M \\
\bar{E}'_M(t) &= (r_m\beta_{SM}\bar{I}_S + r_m\beta_{CM}\bar{I}_C)\bar{S}_M - (\delta_M + \mu_M)\bar{E}_M \\
\bar{I}'_M(t) &= \delta_M\bar{E}_M - \mu_M\bar{I}_M \\
\bar{S}'_H(t) &= v_H - \beta_{MH}\bar{S}_H\bar{I}_M - \mu_H\bar{S}_H \\
\bar{E}'_H(t) &= \beta_{MH}\bar{S}_H\bar{I}_M - (\delta_H + \mu_H)\bar{E}_H \\
\bar{I}'_H(t) &= \delta_H\bar{E}_H - (\gamma_H + \mu_H)\bar{I}_H \\
\bar{R}'_H(t) &= \gamma_H\bar{I}_H - \mu_H\bar{R}_H
\end{aligned} \tag{4.1}$$

where v and μ denote the birth rate and death rate, $B \in \{S, C\}$, $M \in \{O, G\}$, $r_O = \frac{b_O}{N_s + N_C}$, $r_G = \frac{b_G}{N_s + N_C + N_H}$, b denotes the female mosquito biting rate, N denotes the species population, β denotes the infection rate, δ and γ represent the transition rate from state E to I and from I to R.

Given the complexity of modeling interactions across five groups, the task of parameterization presents challenges, especially considering the difficulty in gathering data on the two bird species (sparrows and crows) and the two mosquito types. Given our objective to forecast human cases and understand the epidemic dynamics within the human population, we simplify the model by consolidating the bird species into a single group and the mosquito vectors into another. Since infected humans do not transmit the virus to others, the compartment designated for infectious humans (I_H) is unnecessary and thus eliminated. It is noteworthy that approximately 80% of individuals infected with WNV remain asymptomatic⁹¹. Therefore, we introduce a compartment A_H to represent the asymptomatic pop-

ulation and C_H to represent the confirmed human population, refining the model to more accurately reflect the observed disease dynamics.

Transitions between compartments of the refined model are described by the following differential equations 4.2 and the transition diagram is presented in Fig. 4.1

$$\begin{aligned}
\bar{S}'_B(t) &= v_B - \beta_{MB} \bar{I}_M \bar{S}_B - \mu_B \bar{S}_B \\
\bar{E}'_B(t) &= \beta_{MB} \bar{I}_M \bar{S}_B - (\delta_B + \mu_B) \bar{E}_B \\
\bar{I}'_B(t) &= \delta_B \bar{E}_B - (\gamma_B + \mu_B) \bar{I}_B \\
\bar{R}'_B(t) &= \gamma_B \bar{I}_B - \mu_B \bar{R}_B \\
\bar{S}'_M(t) &= v_M - \beta_{BM} \bar{I}_B \bar{S}_M - \mu_M \bar{S}_M \\
\bar{E}'_M(t) &= \beta_{BM} \bar{I}_B \bar{S}_M - (\delta_M + \mu_M) \bar{E}_M \\
\bar{I}'_M(t) &= \delta_M \bar{E}_M - \mu_M \bar{I}_M \\
\bar{S}'_H(t) &= v_H - \beta_{MH} \bar{S}_H \bar{I}_M - \mu_H \bar{S}_H \\
\bar{E}'_H(t) &= \beta_{MH} \bar{S}_H \bar{I}_M - (\delta_H + \mu_H) \bar{E}_H \\
\bar{A}'_H(t) &= \alpha \delta_H \bar{E}_H - \mu_H \bar{A}_H \\
\bar{C}'_H(t) &= (1 - \alpha) \delta_H \bar{E}_H - \mu_H \bar{C}_H
\end{aligned} \tag{4.2}$$

where α denotes the fraction of the infected asymptomatic human population.

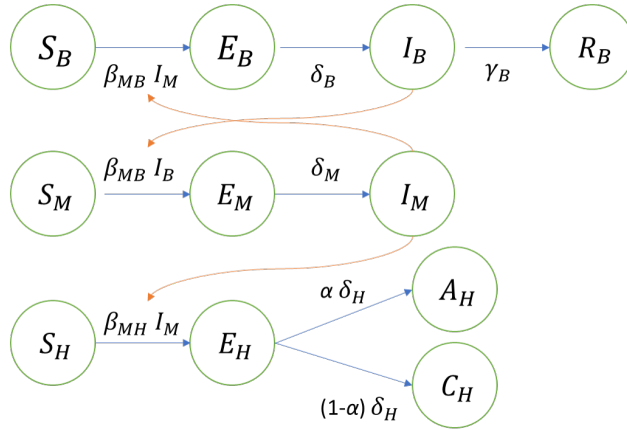


Figure 4.1: SEIR-SEI-SEAC compartmental model

In our research, the incubation periods for birds, mosquitoes, and humans are estimated to be within the ranges of 3-6 days, 8-10 days, and 2-6 days respectively, i.e. $\frac{1}{\delta_B} \sim N(4.5, 0.5)$ day, $\frac{1}{\delta_M} \sim N(9, 0.3)$ day, $\frac{1}{\delta_H} \sim N(4, 1)$ day; for birds, the recruitment rate is set at 0.015, the natural mortality rate falls between 0.001 and 0.002 per day, and the removal rate for infected birds ranges from 0.5 to 1 per day, i.e. $v_B \sim N(0.015, 0.001)$, $\mu_B \sim N(0.0015, 0.0002)$, $\gamma_B \sim N(0.75, 0.1)$; for mosquitoes, both the recruitment and mortality rates are influenced by temperature and precipitation, with the magnitude of recruitment rate established at 0.15 per day and natural mortality rate for mosquitoes varies between 0.02 and 0.07 per day⁹²; the birth and death rates of humans are considered negligible over this time frame, the fraction of the human asymptomatic population is approximated as 0.8, i.e. $\alpha \sim N(0.8, 0.03)$. The recruitment rate of mosquitoes shows a nearly linear increase with temperature within the range of 16°C to 31°C, while their mortality rate tends to decrease as temperatures rise within this same range⁹³, we assume that precipitation also influences the mosquito recruitment rate in a linear fashion. Since the highest average temperature in Colorado in 2023 did not go above 31°C, we proposed the following model 4.4 without considering the situation above 31°C,

$$\begin{aligned} v_M(T, P) &= \begin{cases} 0.15 \times 0.5 \times \left(\frac{T-16}{T_{max}-16} + \frac{P}{P_{max}} \right) & 16 < T < 31 \\ 0.01 & \text{otherwise} \end{cases} \\ \mu_M(T) &= \begin{cases} -\frac{T}{300} + 0.12 & 16 < T < 31 \\ 0.07 & \text{otherwise} \end{cases} \end{aligned} \quad (4.3)$$

The unknown parameters $\beta_{MB}, \beta_{BM}, \beta_{MH}$ are estimated by EnKF. The fitted β_{MB}, β_{BM} are very close in a research⁹². Therefore, we assume $\beta_{MB} = \beta_{BM}$ in this study.

4.2.3 Forecast model

To accurately estimate the unknown parameters and the state variables, we employ an efficient data assimilation technique – the Ensemble Kalman Filter (EnKF)⁸⁴. This method

is grounded in Bayesian statistics, specifically leveraging Bayes' rule to update the estimated state of the system at specific time intervals. It does so by integrating the most recent observations with a priori estimates of the state, thereby refining the system's current state understanding. The EnKF operates by generating an ensemble of system state forecasts, which accounts for the uncertainties in both the system's initial state and the model itself. When new observational data become available, the EnKF assimilates this information, updating each member of the ensemble to more accurately reflect the observed reality. This process not only adjusts the observed components of the system's state but also infers changes to unobserved state variables and parameters, enhancing the model's overall fidelity and predictive capability. Critically, the EnKF facilitates a continuous cycle of prediction and updating, allowing for real-time improvements in forecasting the system's future states. This dynamic adjustment is particularly valuable in complex systems where direct observation of all variables and parameters is challenging or impossible.

4.3 Results

4.3.1 Correlation of human WNV cases with local temperature and precipitation

Fig. 4.2 shows the weekly reported human WNV cases, the weekly average temperature, and the weekly cumulative precipitation. From the figures, we can observe that the weekly reported cases are lag-correlated with the temperature and the precipitation. The correlation coefficient between WNV cases and average temperature is 0.69 with a lag of three weeks. When calculating the coefficient between WNV cases with precipitation, we use the moving average with window size three of the precipitation to reduce the noise of the observation. The correlation coefficient between WNV cases and the moving average of precipitation is 0.57 with a one-week lag. This finding indicates that the temperature and precipitation have the potential to inform reliable prediction of human WNV cases.

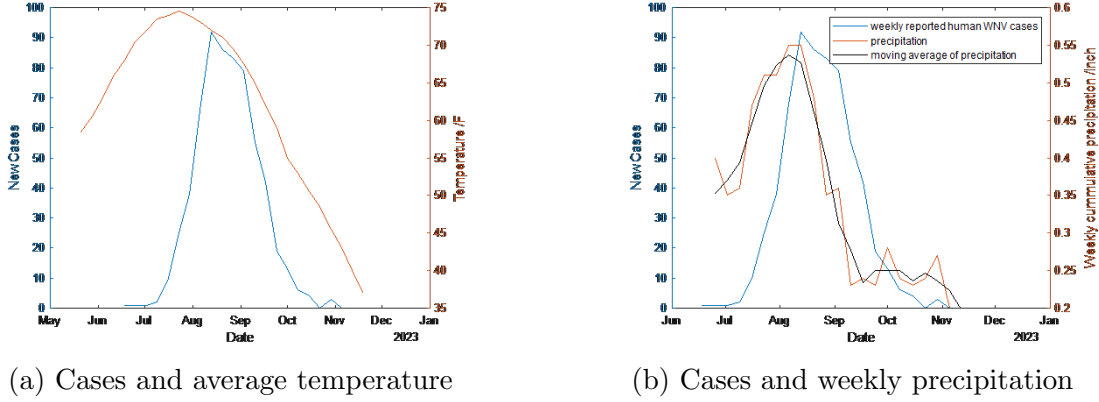
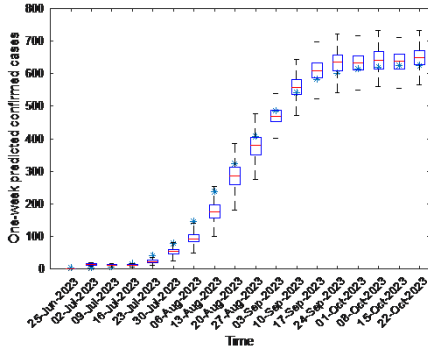


Figure 4.2: Weekly human WNV cases, average temperature, and weekly precipitation

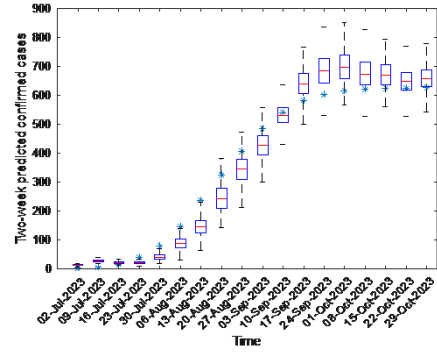
4.3.2 Real-time forecast

The disease model SEIR-SIE-SEAC is applied to the EnKF to generate real-time forecasting of human WNV cases in Colorado, USA, for the year 2023. The forecasting process starts at the onset of the outbreak on June 4th, utilizing the cumulative confirmed human WNV cases C_H as the observation. This data, in combination with average temperature readings and weekly cumulative precipitation figures, is assimilated weekly into the EnKF. As a result, the system is able to generate forecasts for the forthcoming three weeks at each assimilation point. This approach yields three sets of forecast leads during the simulation (where the one-week lead forecasts pertain to the following week, the two-week lead forecasts to the week after next, and so on). The simulation process was iterated 500 times to ensure robustness. The prediction results of the cumulative confirmed cases are shown in Fig. 4.3, where Fig. 4.3a, Fig. 4.3b, and Fig. 4.3c delineate the distribution of the prediction across the three weeks. In these figures, observed case counts are denoted by stars, whereas the forecast distributions are illustrated with box plots: the extremities of the distributions are marked by black lines, the interquartile ranges by blue boxes, and the medians by red lines. Fig. 4.3d presents the average forecast trajectories. It is observed that the actual case numbers predominantly fell within the forecasted range, with shorter lead times yielding more precise predictions. The average forecasts capture the general trend of the epidemic's growth.

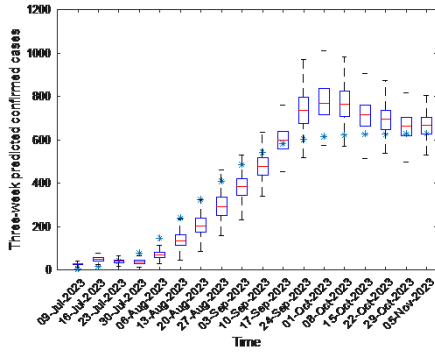
The effectiveness of the forecasting model is assessed by the Absolute Root Mean Square



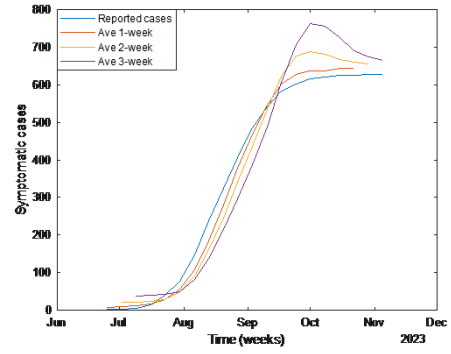
(a) One-week forecast



(b) Two-week forecasts



(c) Three-week forecasts



(d) Average forecasts

Figure 4.3: Real-time prediction of the cumulative WNV cases in Colorado

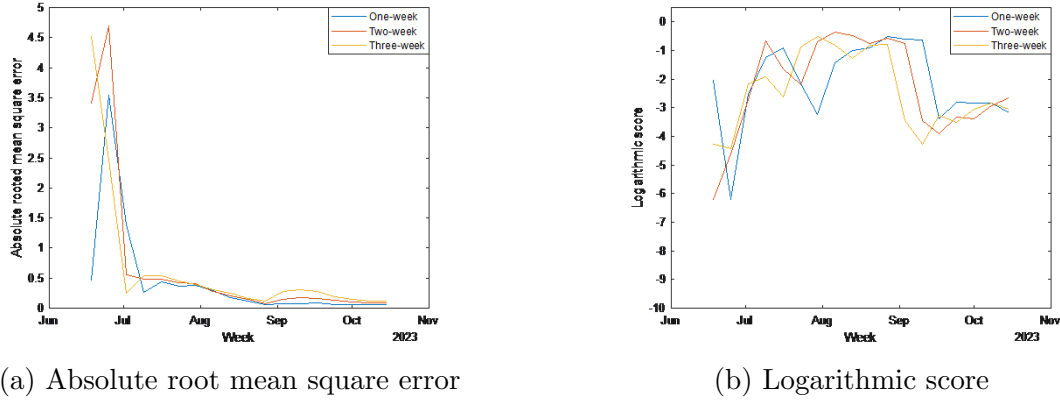


Figure 4.4: Accuracy evaluation

Error (ARMSE) and the Logarithmic Score (LS). The ARMSE provides a measure of the forecast’s precision by quantifying the average magnitude of the differences between predicted values and actual observations. This calculation involves the data in the predicted distribution against the recorded cases. On the other hand, the LS offers insights into the probabilistic accuracy of the forecasts. To compute the LS, we first deduce the probability distribution of new case occurrences from the cumulative case predictions. Then, we classify case counts into predetermined bins (0, 1–5, 6–10, $\dots$, to 46–50, 51–100, 101–150, 151–200, and ≥ 200), aligning with the categorization scheme utilized in the CDC challenge⁷⁵. The outcomes are depicted in Fig. 4.4, where the ARMSE values for each forecast lead after July 9th (the fourth data point) are maintained below 0.5 as shown in Fig. 4.4a, signifying a high level of precision in the model’s predictions relative to the observed data; and the LS remains robust, generally higher than -3 during the epidemic’s peak phases. This indicates that, after an initial period of model adjustment and alignment with the actual epidemiological conditions, the forecasting system demonstrates a remarkable capability to generate real-time predictions with significant accuracy.

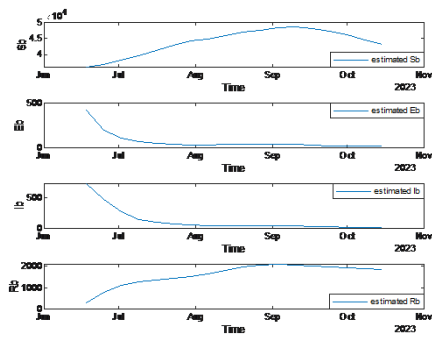
4.3.3 Estimated epidemic dynamics

Our system can also estimate both unobserved states and unknown parameters by assimilating limited observational data. Fig. 4.5 shows the estimated epidemic dynamics across

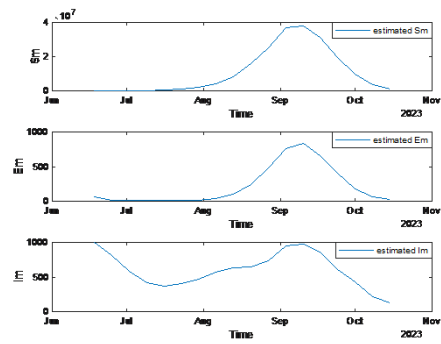
different species, representing the progression of the outbreak. To ensure the model reflects an outbreak scenario, we initiate it with a substantially large number of initially infected birds and mosquitoes. The EnKF will align the model’s state trajectories close to actual conditions as the outbreak starts. From the Fig. 4.5a and Fig. 4.5b, we observe that the population of infectious mosquitoes greatly exceeds that of infectious birds throughout the epidemic, which can be attributed to the longer infectious period of mosquitoes, who remain carriers of the virus until their death, while birds exhibit a high removal rate due to recovery or death, which significantly shortens the duration the virus persists within their system. As shown in Fig. 4.5d, the infection rate between mosquitoes and birds surpasses that from mosquitoes to humans, indicating a higher susceptibility of birds to the WNV. These observations highlight the critical role of mosquitoes in spreading the virus and birds as essential reservoirs for the virus, and further emphasize the need for targeted vector control strategies to mitigate the outbreak. In Fig. 4.5c, the reported number of human cases in Colorado reached 621 by October 8th. However, our model estimates indicate the presence of approximately 2,518 asymptomatic individuals, alongside 100 exposed individuals who may still develop symptoms.

4.4 Discussion

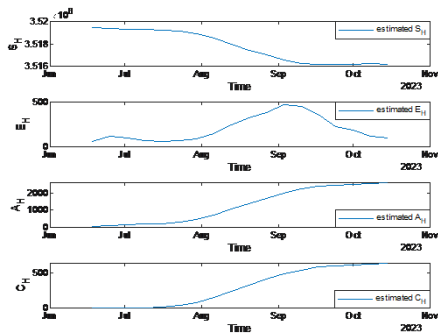
Our analysis reveals a significant correlation between the reported weekly human WNV cases and the weather factors (temperature and precipitation). Intriguingly, the correlation strengthens with higher powers of these weather variables. Specifically, the temperature raised to the ninth power exhibits the highest correlation coefficient of 0.88 with the human cases, whereas for precipitation, a power of six yields a correlation coefficient of 0.65, as detailed in Table 4.1. This suggests that the relationship between weather factors and WNV transmission may be more complex than linear associations, and points to the need for high-dimensional systems to depict the intricate dynamics between weather factors and WNV transmission. Our SEIR-SEI-SEAC model incorporates 11 states, allowing for the exploration of parameters with high orders.



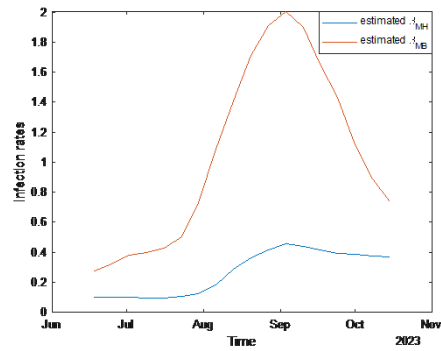
(a) Birds



(b) Mosquitoes



(c) Humans

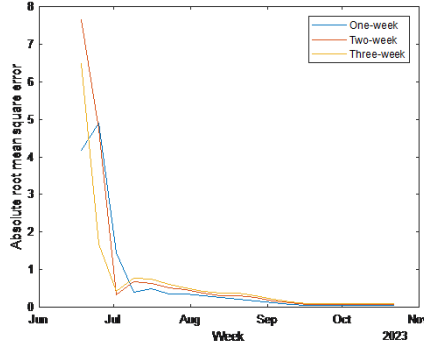


(d) Infection rates

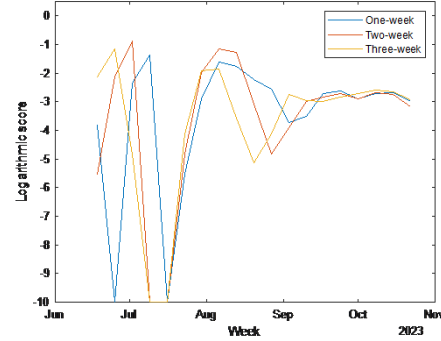
Figure 4.5: Estimated epidemic dynamics

Table 4.1: The correlation coefficients r between human WNV cases and higher orders i of weather variables (temperature T and precipitation P)

i	2	3	4	5	6	7	8	9	10
r_{T^i}	0.73	0.77	0.79	0.82	0.84	0.85	0.87	0.88	0.88
r_{P^i}	0.60	0.62	0.63	0.64	0.65	0.65	0.65	0.65	0.65



(a) Absolute root mean square error



(b) Logarithmic score

Figure 4.6: Accuracy evaluation of the baseline model

Our results reveal that integrating weekly human WNV case data and weather variables into our forecasting model enables the production of accurate real-time forecasts for human WNV cases. We evaluate the effect of the wether variables on the precision of our forecasts against a baseline model that disregards weather influences, maintaining constant mosquito recruitment ($v_M \sim N(0.15, 0.01), \mu_M \sim N(0.07, 0.01)$). Fig. 4.6 presents the ARMSE and LS of the forecasts from the baseline model. Comparing with Fig. 4.4, we observe that the model incorporating weather variables consistently outperforms the baseline. Specifically, the average ARMSE for the baseline model are 0.19, 0.27, and 0.31 for successive leads after July 9th, whereas the wether-incorporated model achieves improved averages of 0.17, 0.23, and 0.28, respectively; similarly, the average LS for the baseline model is -3.2, -3.8, and -3.9 for each forecast lead, in contrast to the more favorable scores of -1.85, -1.85, and -2.2 achieved with the integration of weather variables. Therefore, our findings underscore the substantial benefits of incorporating weather variables into epidemic forecasting models, highlighting their critical role in enhancing forecast accuracy for human WNV cases.

Our forecasting system offers not only precise real-time predictions of WNV cases with

considerable lead time, but also provides a detailed estimation of the epidemic dynamics across different populations. The forecasts along with the estimated data such as the estimated numbers of infectious birds and mosquitoes, the count of exposed humans, and the prevailing infection rates hold significant potential for aiding public health officials in several key areas. Firstly, it enables the strategic allocation of medical and preventive resources. Secondly, it supports targeted mosquito control initiatives, by identifying periods of high transmission risk with estimated mosquito or bird infection rates, thereby allowing for more effective vector control measures. Lastly, it provides a foundation for ongoing surveillance and response planning, ensuring that public health strategies can be dynamically adjusted in response to evolving epidemic patterns.

Chapter 5

Multilayer network analysis of FMD transmission and containment among beef cattle farms

5.1 Introduction

Foot-and-mouth disease (FMD) is a highly contagious livestock viral disease, which can lead to destructive economic losses¹⁶. Several major outbreaks have occurred in the UK (2001)¹⁷, Netherlands (2001)¹⁹, Japan (2010)¹⁸, Uganda (2006)²⁰ etc. The estimated losses of the FMD outbreak in the UK, in 2001 amounted to about Pound Sterling 3.1 billion¹⁷; about 290,000 animals were culled during the FMD outbreak occurred in Japan, 2010¹⁸; a study estimated that one FMD outbreak constrained to Kansas US would result in an economic loss varying from US Dollar 43 to 706 million²¹. Therefore, it is important to understand the impacts of different FMD virus transmission routes on the outbreak and design control measures to prevent the spread of the disease in the livestock industries.

Moving infected animals and sharing contaminated equipment are considered two of the most common routes for between-farm infectious disease transmission^{94;95}. A number of researches have been conducted to explore the impacts of the different transmission routes

on the spread of infectious diseases by building variant between-farm contact networks. Generally, the between-farm transmission routes can be described by two types of contact networks – direct contact networks (DCN) which are formed by the movement of livestock between farms, and indirect contact networks (ICN) which are formed by the personnel visits, sharing transport vehicle and tools, etc.^{96–99}. Instead of focusing on a single contact network, many studies combined the DCN and ICN into a two-layer network (D&ICN) to analyze the roles each transmission route plays in the disease spreading^{95;100–103}. Study¹⁰⁰ showed that the porcine reproductive and respiratory syndrome virus transmitted by the D&ICN resulted in a larger outbreak compared with the situation when the virus was transmitted by DCN. Sharing haulage vehicles and animal product transport vehicles were found able to increase the contacts between farms by more than 50% and accelerate disease transmission⁹⁵. A study¹⁰¹ concluded that indirect contact through vehicles and operators was crucial to accurate predictions of the epidemic size. These results all highlighted the impacts of indirect contact on the disease spread. Thanks to the rapidly developed communication technology, mitigation strategies based on information sharing¹⁰³ have become increasingly promising. Information sharing was primarily used to improve supply chain resilience to disruptions in livestock production^{104;105}. Recently, many scholars have used it to control the progressive epidemic^{103;106–108}. In a study¹⁰³, an agent-based model was developed, where the FMD virus was spread through the DICN, the infected farms would inform their partners through the information-sharing network (ISN) and the informed farms would implement preventive measures to suppress the disease spreading. In their study, they assumed that all farms participated in the information-sharing system. However, Farmers’ willingness to participate in the information-sharing system depends on complicated factors such as risk attitudes, privacy, and transparency issues^{109;110}. Therefore, efforts are still needed to investigate how participation levels will affect the effectiveness of information-sharing mitigation strategies. In this chapter, we generated a three-layer network including the direct-contact layer (cattle movements), the indirect-contact layer (truck visits), and the information-sharing layer to simulate the FMD virus transmission dynamics among beef-cattle farms in southwest Kansas (SW KS) and the effectiveness of the mitigation strategies based on the information-sharing

method. Our goal is to investigate the potential impacts of the virus transmitted by cattle movement (direct contacts) and vehicle visits (indirect contacts) on the FMD outbreaks and to assess how the information-sharing networks and the fraction of participation farms can affect the epidemic size. To our knowledge, there is no previous study constructing a three-layer network analysis on FMD involving information-sharing techniques. In addition, the influence of different participations in information sharing is explored unprecedentedly. In this study, hypothetical FMD outbreaks were first simulated on a single-layer DCN (cattle-movement network), a two-layer D&ICN (cattle-movement layer and truck-visit layer), and a three-layer DIC&ISN (cattle-movement layer, truck-visit layer, and information-sharing layer) to assess the effects of each layer on the transmission of FMD virus among beef-cattle farms. We also investigated the impacts of different types of information-sharing layers and participation levels on the epidemic size and quarantine size. Furthermore, a sensitive analysis was conducted to explore the potential economic losses caused by the infection and quarantine to optimize the economic loss and control costs.

5.2 Materials and methods

5.2.1 Data

We extracted the weekly cattle movement records and the vehicle sharing records between July 1st, 2019, and December 31st, 2019 from the database¹¹¹. The database provides comprehensive cattle trading records between premises and premises information in SW KS. Cattle trading records include the identifiers for departure premises and destination premises, truck identifiers for every movement, headcount of cattle carried by truck, and transport date. The premises information includes the premises type (ranches, stockers, and feedlots) and the capacity which is the maximum number of cattle one can hold. In this study, 301 premises in different production types are involved, among which there are 18 ranches, 50 stockers, and 233 feedlots.

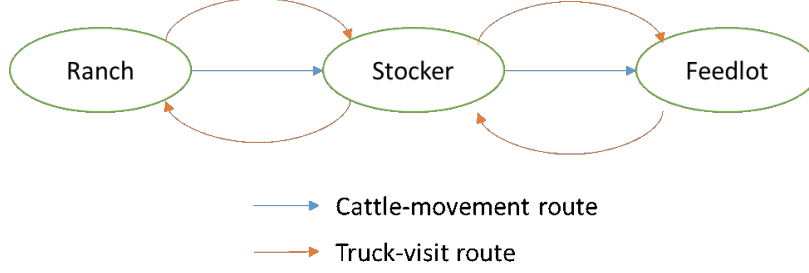


Figure 5.1: Transmission routes of the FMD virus among different production types of beef-cattle farms

5.2.2 Multilayer network construction

To describe the FMD transmission routes by cattle movements and truck visits shown in Fig. 5.1, we construct a weekly series of weighted two-layer contact networks based on the movement data. On top of the contact network, we add an information-sharing layer to simulate the influence of different information-sharing networks and the participation level in suppressing the epidemic. Thus, each network was formed by three independent layers as shown in Fig. 5.2. The nodes with the same identifier on each layer represent the same premise; the edges within the layer represent the virus (contact layers) or information (information-sharing layer) transmission routes between premises. Direct-contact layer, indirect-contact layer, and information-sharing layer are presented as $L_{DC} = (V, E_{DC})$, $L_{IC} = (V, E_{IC})$, and $L_{IS} = (V, E_{IS})$, where V represents the set of nodes, and E_i represents the set of edges within layer i . In the direct-contact layer L_{DC} , every cattle movement is generated as a directed and weighted edge from the departure premise (node) pointing to the destination one, with the weight as the headcount of transported cattle. For the indirect-contact layer L_{IC} , the premises are connected if they are visited by the same truck in a week. Because the trucks usually return to the departure premises after transporting cattle, the edges within L_{IC} are undirected. The weights on E_{IC} are the number of visits between premises.

We design three information-sharing layers L_{IS} with different network structures but the same network density. The first type is an Erdos Renyi random network¹¹² with the same network density (0.0068) as the L_{IC} ; the second one is the contact-based network, where the premises share their infection status with their business partners, so $L_{IC}^{(i-1)}$ at week

$i - 1$ is used as the L_{IS}^i at week i ; the third one is based on the geographic distance between premises, where the premises whose distance is shorter than 450 kilometers are connected (with 450 km threshold distance, the distance-based information-sharing network has very similar network density with the other ones). Finally, the infection status of each node on every layer is aggregated into the multilayer network. We construct 26 time-series multilayer networks in total. Each network represents the contact activities in one week, as one week (7 days) is a range of farm-level infection periods preceding detection⁹⁵.

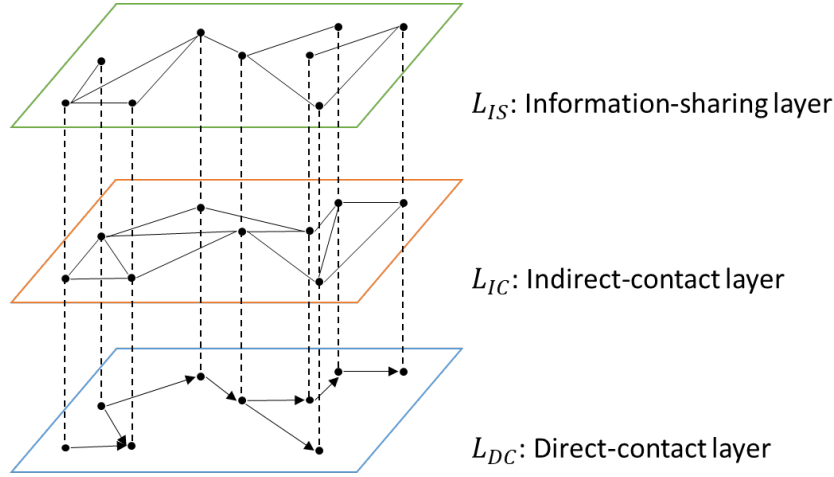
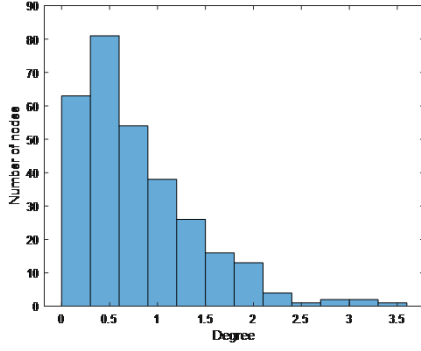


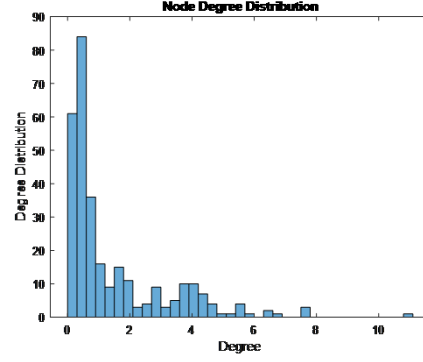
Figure 5.2: Multilayer network structure of FMD transmission among beef-cattle farms

5.2.3 Network measurements

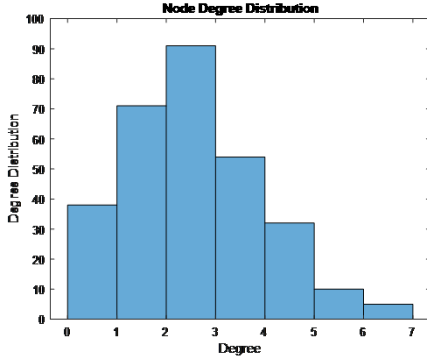
To have an insight into the network topology, we compute some basic measurements that are commonly used to describe the structure of the network. Those measurements include network density which is defined as the ratio of the number of actual edges to the number of potential edges, distribution of node degrees which is defined as the distribution of the number of edges pointing to each node, and the average node degree which is defined as the mean value of the node degrees. Considering that we have 26 homogeneous networks for the cattle movements and the truck visits, we present the mean value of all those measurements over 26 networks. For cattle movement networks, the mean value of the network density is 0.0026, the distribution of the node degrees is presented in Fig. 5.3a, and the mean value



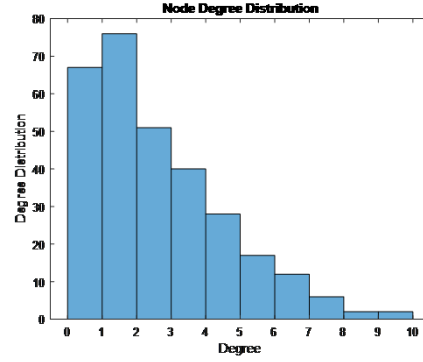
(a) Cattle movement networks



(b) Truck visit networks



(c) Erdos Renyi random network



(d) Distance-based network

Figure 5.3: Node degree distributions of different networks

of the average node degree is 0.79. For the truck visits networks, the mean value of the network density is 0.0068, the distribution of the node degrees is shown in Fig. 5.3b, and the mean value of the average node degree is 2.04. The network densities and the average node degrees of both types of networks are very small, which means both cattle-movement networks and truck-visit networks are sparse. The node degree distributions for cattle-movement networks and the truck-visit networks indicate that there is a small group of nodes with high node degrees. For the generated information-sharing networks, the network densities are set close to 0.0068 to keep the sparse characteristics (random network 0.0068 and distance-based network 0.0071); the degree distributions of the Erdos Renyi random network and the distance-based network are shown in Fig. 5.3c and Fig. 5.3d.

5.2.4 Epidemic model

To describe the transition process of infection status on each premise, we propose a susceptible-quarantined-exposed-infectious-removed (SQEIR) model. In this model, each premise (node) can only be in one of the five states. A susceptible (S) node can become exposed (E) if it is connected to infectious nodes on L_{DC} and L_{IC} . The transition rate from S to E is $\beta_{DC} \times I_{DC} + \beta_{IC} \times I_{IC}$, where β_{DC} and I_{DC} denote the infection rate and the number of contacts to infectious nodes on L_{DC} , β_{IC} and I_{IC} denote the ones on L_{IC} . An exposed (E) node is not yet infectious, but it will transit to the infectious (I) state with rate δ . Finally, an infectious node will be removed (R) with a removing rate γ , which means the cattle in that farm are culled. A susceptible (S) node can become quarantined (Q) with rate $\alpha \times I_{IS}$ if it receives notification from infectious neighboring farms on L_{IS} . The quarantined (Q) node can also become susceptible (S) at a rate μ . The state transition process is expressed in equation 5.1 and shown in Fig. 5.4.

$$\begin{aligned}
 S'(t) &= -(\beta_{DC}I_{DC} + \beta_{IC}I_{IC})S - \alpha I_{IS} + \mu Q \\
 Q'(t) &= \alpha I_{IS} - \mu Q \\
 E'(t) &= (\beta_{DC}I_{DC} + \beta_{IC}I_{IC})S - \delta E \\
 I'(t) &= \delta E - \gamma I \\
 R'(t) &= \gamma I
 \end{aligned} \tag{5.1}$$

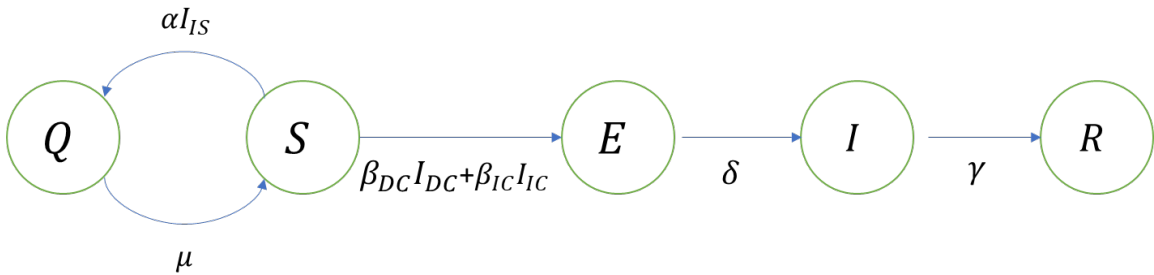


Figure 5.4: SQEIR compartmental model

The parameters used for this model are specified in Table 5.1 1; parameters μ and μ are

Table 5.1: Description and value of the parameters in the model

Parameter	Description	Value	Reference
β_{DC}	Infection rate of direct contact	0.95	study ¹⁰³
β_{IC}	Infection rate of indirect contact	0.15	study ¹⁰³
δ	Incubation rate	0.83	study ¹⁰³
γ	removing rate	0.12	study ¹⁰³
α	Quarantine rate	1	Assumed
μ	Releasing rate	0	Assumed

related to farm owners' precautionary awareness of FMD. For simplicity, we assume that $\alpha = 1$ and $\mu = 0$, which means the informed premises will be quarantined and remain so till the epidemic ends.

The SQEIR model embedded in the multilayer network is simulated in the Generalized Epidemic Modelling Framework (GEMF), which was developed by the Network Science and Engineering group at Kansas State University. GEMF can numerically simulate spreading processes on static multilayer networks¹¹³ and is available in MATLAB, R, Python, and C programming languages. In this chapter, we further adapt the GEMF toolbox in Matlab for simulating disease spreading on time-series multilayer networks.

5.2.5 Impact of cattle movement, vehicle visits, and information sharing on disease spreading

The hypothetical FMD spreading is simulated on a weekly series of DCN, D&ICN, and multilayer networks with contacts layers and information layer (DIC&ISN). Those different networks are formed by the L_{DC} , L_{IC} , and L_{IS} we constructed in the previous section. For this simulation, the Erdos Renyi random network is used as the L_{IS} , and all premises participate in the information-sharing system. The FMD virus is seeded in three randomly selected premises in a random week among the first four weeks. The simulation is repeated 5000 times. The numbers of the premises in states E, I, and R versus time and the distributions of the final epidemic size (the fraction of the removed premises at the end of the epidemic) of the spread of FMD (at least one farm is infected by the seeded premises) are

presented and evaluated.

5.2.6 Impact of different information-sharing layers and premises participation on disease spreading

The idea behind the information-sharing network is the following. Suppose that one farm is discovered to have cattle infected with FMD. This farm will immediately go to isolation, following the state of Kansas procedure. At this point, this farm will inform its neighboring farms (the farms connected in the network are defined as neighboring farms) of its infection status. The neighboring farms will, in turn, go into a quarantine state, halting all movements of animals and trucks. Cattle in infected farms will be culled, while cattle in farms protected by quarantine, will return to normal conditions once the epidemic is contained. Three different types of the information-sharing layer are added individually to the contact networks. The hypothetical FMD spreading is simulated on the weekly series of DIC&ISNs with different L_{IS} –the Erdos Renyi random information-sharing layer, the contact-based information-sharing layer, and the distance-based information-sharing layer. Different premises participations are also considered. We randomly activate 10% to 100% (with an interval 10%) participation nodes and deactivate the rest in each L_{IS} , and simulate the hypothetical FMD spreading. Considering there are hubs (nodes with higher degrees) in both contact networks, we rank the nodes by their node degrees and activate 10% to 100% (with an interval 10%) participation of nodes with the highest degree, and simulate the FMD spread again. All the simulations are repeated 5000 times. The distributions of the final epidemic size and the quarantine size of the different participations are presented and evaluated.

5.2.7 Sensitivity analysis of the potential economic loss from the epidemic and quarantine

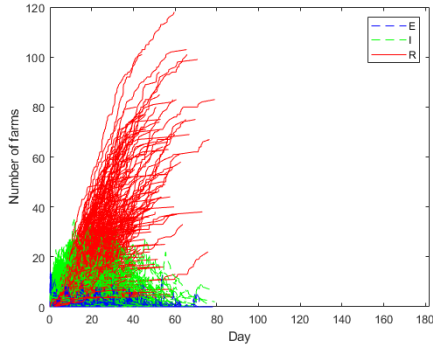
According to study¹⁶, the FMD infection can cause losses by reducing milk production, suppressing the growth rate of livestock, and culling the infected cattle. Considering that quarantined farms stop contact with other farms, which leads to economic losses as well, we proposed a loss function $l = xR + yQ$ ($x > y$), where x is the average loss for one removed farm, y is the average loss for one quarantined farm, R is the number of the removed farms, and Q is the number of the quarantined farms. To qualitatively analyze the potential loss from the epidemic and the quarantine, we rewrite the loss function as $\bar{l} = R + \bar{y}Q$, where the relative loss per quarantined farm $\bar{y} = y/x$, and the relative total loss $\bar{l} = l/x$. We vary the value of $\bar{y}$ from 0 to 0.5 with step increment 0.05, and calculate $\bar{l}$ of each L_{IS} with different participation levels.

5.3 Results

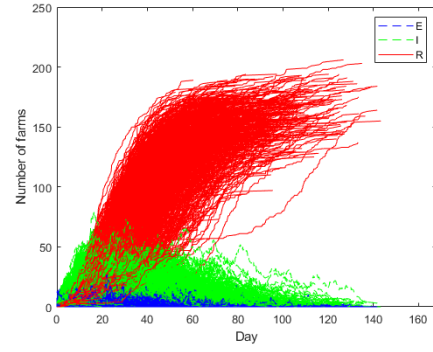
5.3.1 Impact of cattle movement, vehicle visits, and information sharing on disease spreading

The numbers of premises in different infection compartments and the distribution of the epidemic size based on three networks with different layers are shown in Fig. 5.5. The blue dash line, green dash line, and red solid line represent the number of premises in state E, I, and R versus time separately; the boxplots shows the distributions of the final epidemic size, where the red line in the box represents the median; the upper and lower edges of the box represent the 75% and 25% quartiles; the short lines outside the box are the adjacent (maximum and minimum without outliers); red plus signs are the outliers.

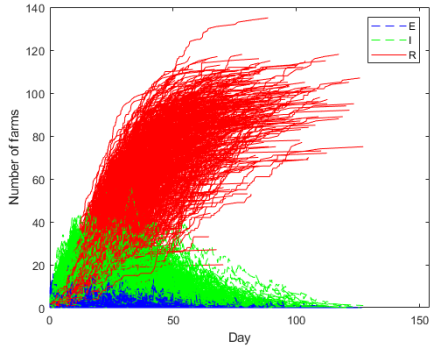
For the hypothetical FMD spreading simulations, the number of simulations producing additional infections on different networks are 1628 (32.56%) on DCN, 1799 (35.98%) on D&ICN, and 1770 (35.40%) on DIC&ISN. The reason for this low outbreak probability is



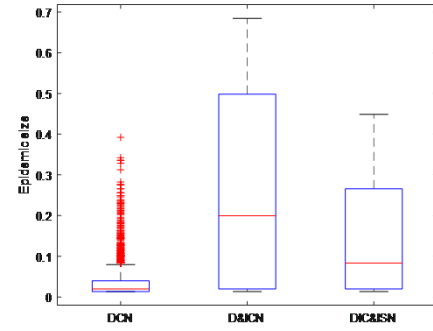
(a) Number of infected farms on DCN



(b) Number of infected farms on D&ICN



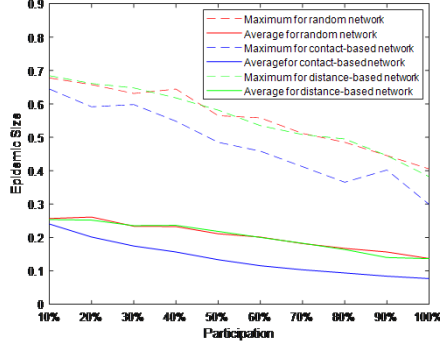
(c) Number of infected farms on DIC&ISN



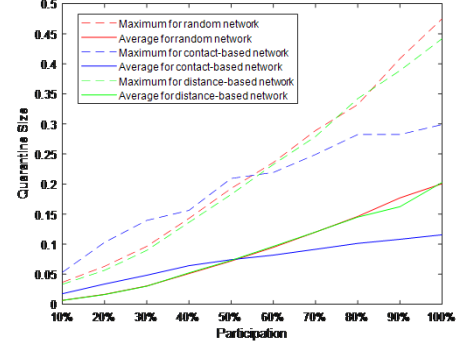
(d) Distribution of epidemic size on three networks

Figure 5.5: Number of farms in infected compartments and distribution of epidemic size

the sparsity of the contact layers, in which there are often no links to or from the initial infected farms during simulation. Surprisingly, the increment of the outbreak probability (the fraction of simulations producing additional infections) on D&ICN from DCN is only 171 (3.42%) even if the average network density of truck-visit network is more than twice of the cattle-movement network. Two facts should be considered here. On the one hand, the indirect infection rate of the truck visit layer is much smaller than the direct infection rate. On the other hand, the number of truck visits is smaller than the number of cattle transported (normally several cattle transported by one truck), which means the weight of the truck visit layer is smaller than that of the cattle movement layer. Therefore, the pathogens are more likely to be transmitted by the cattle movements when the number of infections is small. For those spread FMD, the average epidemic duration is 4.44 weeks on DCN (minimum 2 weeks and maximum 12 weeks), 8.29 weeks on D&ICN (minimum 2 weeks and maximum 22 weeks), and 6.89 weeks on DIC&ISN (minimum 2 weeks and maximum 18 weeks). As shown in Fig.4d, the maximum and average epidemic size of spread FMD are 0.392 (118/301) and 0.040 (12/301) on DCN, 0.684(206/301) and 0.261 (78.5/301) on D&ICN, and 0.449 (135/301) and 0.138(41.6/301) on DIC&ISN. Our results show that the truck-visit layer not only prolong the average epidemic duration by 86%, but also enlarge the overall epidemic size by 554% on average and 74% on the maximum. Apparently, the truck visits greatly boost the disease transmission. Truck visits become increasingly effective for FMD transmission as the number of infected farms is increasing, although the infection rate and the weight for it are relatively small. The other important characteristics of the truck visit network like higher density and undirected links also contribute to spreading the disease to more farms. The random information-sharing layer has apparent impact on mitigating the epidemics of D&ICN with 17% shorter average duration and 47% smaller average epidemic size.



(a) Epidemic size with different L_{IS}



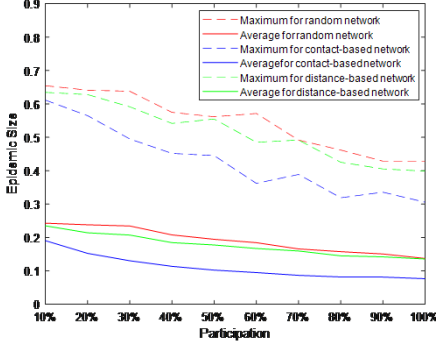
(b) Quarantine size with different L_{IS}

Figure 5.6: The epidemic size and quarantine size versus random participation

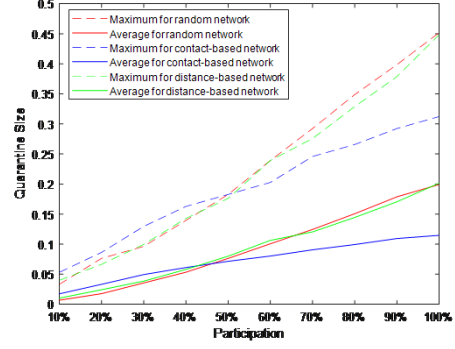
5.3.2 Impact of different information-sharing layers and premises participation on disease spreading

Fig. 5.6 shows the maximums and average of the epidemic size and quarantine size versus participations when different L_{IS} are applied. In this scenario, the participated farms are randomly selected. The solid lines in the figure represent the average while the dashed lines are the maximums; the red, blue, and green lines present the data from random L_{IS} , contact-based L_{IS} , and distance-based L_{IS} separately. In general, the epidemic size decreases and the quarantine size increases when more farms participated in the information-sharing networks. The maximum and average epidemic size of the contact-based L_{IS} is the smallest almost under any participation level, which means the contact-based L_{IS} has the best performance on containing the disease spread. The quarantine size of the contact-based L_{IS} is the largest when the participation is less than 50%, and the smallest when participation is greater than 60%. Based on the equation 5.1, infected farms is promoting the growth of quarantine while the quarantined farms is prohibiting the growth of infection in every three-layer networks. From the results, we can deduct that the contact-based L_{IS} is able to make more farms quarantined under lower participation level to prohibit the infection more effectively than other two L_{IS} , and that the smaller epidemic size of contact-based L_{IS} under higher participation level makes the quarantine size small.

In order to more effectively prohibit the disease spreading on the contact networks, we



(a) Epidemic size with different L_{IS}



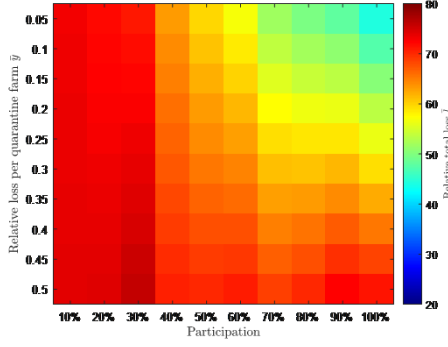
(b) Quarantine size with different L_{IS}

Figure 5.7: The epidemic size and quarantine size versus highest-degree participation

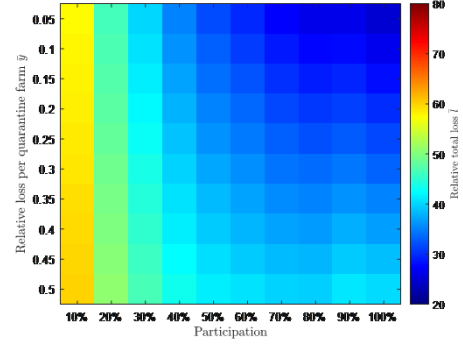
primarily include the farms with the highest node degrees into the information-sharing networks. The epidemic size and the quarantine size are shown in Fig. 5.7. Compare with the random participation, the performance of contact-based L_{IS} on containing epidemic size has an obvious improvement. Under 10% participation, the average epidemic size is 0.19, which is 20.8% decrease from 0.24 with random participation. The average epidemic size decrease to less than 0.1 when participation level is higher than 50%. However, different types of participations have little effect on epidemic size of the other two L_{IS} and the quarantine size of all three L_{IS} .

5.3.3 Sensitivity analysis of the potential economic loss from the epidemic and quarantine

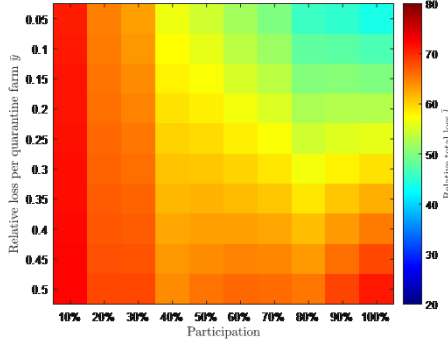
The potential losses caused by the infection and the quarantine are displayed in Fig. 5.8. The x-axis represents the participation level ranging from 10% to 100%; y-axis represents the relative loss per quarantine farm $\bar{y}$; the scaled color in each pixel represents the relative total loss $\bar{l}$. Comparing the loss color map of three networks, the relative total loss $\bar{l}$ with the contact-based L_{IS} are mostly the smallest under any participation and quarantine loss $\bar{y}$ pair. The smallest economic loss is always achieved in the upper right corner, which suggests that the more farms participated in the information-sharing network, the lower economic losses can be achieved, especially when the quarantine cost is much lower than infection (when $\bar{y}$



(a) Erdos Renyi random L_{IS}



(b) Contact-based L_{IS}



(c) Distance-based L_{IS}

Figure 5.8: Relative total economic losses on networks with different L_{IS} (highest-degree participations)

is negligible).

The results in Fig. 5.9 show the range of the total loss under different participation or different quarantine loss $\bar{y}$. The range of the total loss is taken as a measurement of the sensitivity between the total loss and the other variables. The ranges of the total loss under 10% participations are less than 3, which means the total loss is not sensitive to the relative quarantine loss $\bar{y}$ under 10% participations. The range increases with the participation level, and decreases with the relative quarantine loss, which means the loss becomes more sensitive to quarantine loss $\bar{y}$ under greater participations, and less sensitive to the participations when the relative quarantine loss $\bar{y}$ grows. From Fig. 5.7, we know that quarantine size Q is much smaller than epidemic size R when the participation is low, so that R is the dominant term in the loss calculation. The increase rate of the quarantine size is greater than the decrease rate of the epidemic size with participation level based on Fig. 5.7. The loss decreased from

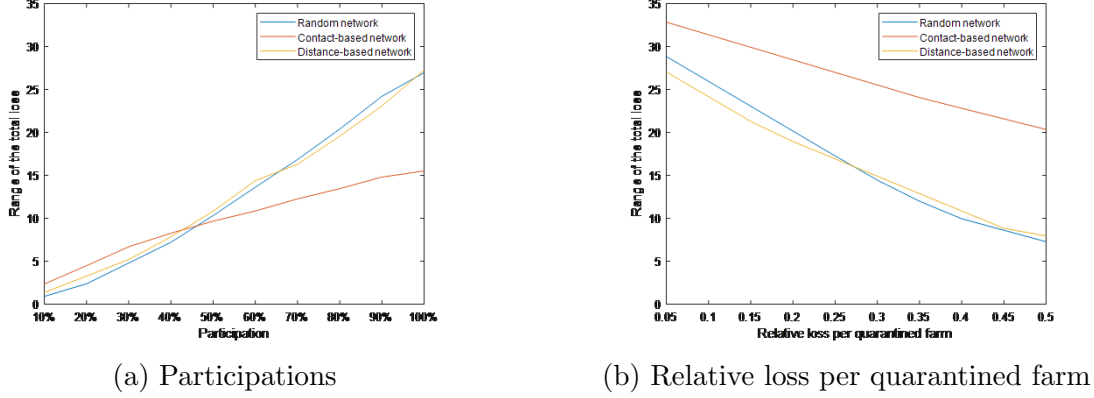


Figure 5.9: Range of the total loss under different participations and relative quarantine losses

R can be compensated by the increase in Q especially when the relative quarantine loss $\bar{y}$ is large.

5.4 Conclusion

In this chapter, we firstly conducted hypothetical FMD spreads on a single-layer DCN, a two-layer D&ICN, and a three-layer DIC&ISN, where the results show that the FMD has larger epidemic size and long duration on the two-layer D&ICN than the single-layer DCN, and that the information-sharing layer is able to mitigate the spread of the disease. These results re-addressed the significant influence of the truck visits on the between-farm transmission of FMD virus, and revealed that truck-visit routes can greatly enlarge the outbreak epidemic size, but have limited influence on increasing the outbreak probability. Therefore, the regional animal disease control agencies should take actions to reduce direct-contact rate and control the cattle movements in the early stage of the epidemic to prevent the outbreak. For existing outbreaks, increasing truck sterilization frequency and limiting truck visits are useful policies to mitigate the epidemic size. The information-sharing network is tested effective for epidemic mitigation. We compared the mitigating performance of three information-sharing networks with different topologies and participation levels. The contact-based network has the best performance on suppressing the epidemic size while maintaining

low quarantine size especially when the farms with the highest degrees are participated. For all three information-sharing networks, increased participations result in decreased epidemic sizes but increased quarantine sizes. For contact-based ISN, the epidemic size decreases fastest, while the quarantine size increases slowest. A sensitive analysis is conducted to explore the potential economic losses caused by the infection and quarantine. We found that the losses with contact-based ISN is the smallest under any participation and relative quarantine loss pair, and that an increased participation level generally leads to a decreased economic loss when the quarantine costs are negligible ($\bar{y} \leq 0.3$). The regional agencies could work on calling more farms to participate in the contact-based information-sharing network to ensure that the infection status of the farms is accessible to their business partners. In our results, an 80% participation level is able to display fairly good performance on reducing the economic losses. We focused on the beef-cattle industry in SW KS, but the implement of information-sharing network can be used by the whole livestock industry. In this chapter, there are several limitations. First, the other indirect contacts, like personnel visits and air-borne are not considered. Another limit is we did not consider the auction market. Removing these limitations can be the subject of the future work, together with the determination of the optimal participation level.

Chapter 6

PICTUREE—Aedes: A web application for dengue data visualization and case prediction

6.1 Introduction

The PICTUREE—Aedes web application is an innovative system that integrates tools for displaying dengue-related data, visualizing simulated results, and forecasting future cases for an ongoing outbreak. This application is a product of the Predicting Insect Contact and Transmission Using histoRical Entomological and Environmental data (PICTUREE) project, which was initiated in 2019.

Dengue fever is an *Aedes* mosquito-borne viral infection that threatens the health of over 3 billion people in numerous tropical and subtropical regions¹¹⁴. *Aedes aegypti* mosquito is the primary carrier of dengue viruses, with the ability to bite multiple individuals in a brief period, and its immature phase can be found in water containers near human dwellings¹¹⁵. The proliferation, biting rates, infective rate, and survivability rate of *Aedes* mosquitoes are substantially affected by climatic conditions like rainfall, temperature, and relative humidity, which, in turn, have a significant impact on the spread of the dengue virus^{116–118}. The

breeding time and maturity period of *Aedes* mosquitoes are shortened due to global warming, resulting in the accelerated growth of the vector population¹¹⁹. Furthermore, increasing human mobility and high population density have facilitated the geographical spread of dengue viruses over long distances¹²⁰. Therefore, an integrated system is needed to aggregate collected data, analyze them, estimate parameters, and forecast future risks and incidence to proactively combat the spread of the virus.

Numerous research studies have been conducted to identify and quantify the temperature-dependent variables that impact the life cycle of the mosquito vector, due to the susceptibility of the mosquito's life cycle to temperature^{121–124}. Mathematical equations can be used to model the dynamics of the mosquito population, and estimate the parameters such as mortality, transition rate, and oviposition rate at different temperatures¹¹⁸. Another dengue transmission model, which took into account various temperatures, survival rates, between-bite intervals, and extrinsic incubation periods, discovered that an increase in mosquito longevity and temperature led to an augmentation in dengue transmission¹²⁵.

There has been an increasing number of approaches to predicting the incidence of infectious diseases, including the use of statistical models^{126;127}, state-space compartmental models^{128–131}, deep learning methods^{132–134}, etc. The state-space models coupled with data assimilation algorithms can generate real-time forecasts and estimate key parameters. The estimated model parameters can then be utilized to determine the basic reproduction number of the disease, the carrying capacity of the vector population, and other relevant metrics. A state-space compartmental disease model combined with an ensemble Kalman filter (EnKF) was employed to estimate and predict future dengue incidence, infection rates, and mosquito emergence rates in¹²⁹. Deep learning approaches, especially the long-short-term memory (LSTM) recurrent neural network, are also widely used to generate reliable epidemic forecasts¹³⁴. Machine learning approaches have the advantage of being able to handle non-linearity and complex data structures. Another emerging technique is the ensemble forecast that enhances the forecasting accuracy of individual models^{135;136}. A superensemble forecast was developed by combining three models - susceptible-infectious-recovered-ensemble adjustment Kalman filter, Bayesian weighted outbreaks, and historical likelihood - using the

weighted average method¹³⁶. The superensemble was used to predict dengue cases in San Juan, and it was found to provide more accurate forecasts compared to individual models. The use of ensemble forecasts, which combine multiple forecasting models or techniques, can help reduce errors and biases that may be present in individual forecasts.

The PICTUREE—Aedes web application aggregates dengue-related data from multiple sources, including global daily temperature, global daily precipitation, *Aedes* mosquito occurrence, and dengue case occurrence. These data are stored in an SQL database and can be readily visualized through user-friendly displays. The PICTUREE—Aedes web application also provides estimations of the *Aedes* mosquito population, the dengue transmission reproduction number R_t , and the dengue risk globally, which can be viewed by users in both space and time. Furthermore, the application employs several algorithms, including the ensemble Kalman filter, the recurrent neural network, the particle filter, and the super ensemble, to forecast dengue cases for an ongoing outbreak. To generate forecasts, users must provide information about the cumulative reported cases in recent weeks, the outbreak location, and the local population size. After processing this data, the application outputs the statistics of the three-week forecasted cases.

6.2 Data

Table 6.1 presents a comprehensive list of the data sources. The daily global temperature and precipitation data are obtained from the National Oceanic and Atmospheric Administration (NOAA) weather stations¹³⁷. The ERA5 hourly data provides hourly estimates of atmospheric variables, such as temperature, pressure, and humidity, which are used for simulation purposes¹³⁸. The terrestrial ecoregions data are utilized to select areas with similar climate conditions^{139–141}. The *Aedes* mosquito occurrence dataset and dengue occurrence dataset are geo-positioned occurrence records spanning from 1960 to 2012^{142;143}. Country-wise Google trend data for the keywords "dengue" and "fever" are also collected. The aforementioned data are either directly displayed on the map or used for the simulations and forecasts¹⁴⁴.

Table 6.1: Data sources

Item	Source	Remark
Daily weather observation	study ¹³⁷	Updated weekly
ERA5 hourly data	study ¹³⁸	2018 - present
Terrestrial ecoregions	study ¹³⁹⁻¹⁴¹	As of December 14, 2009
<i>Aedes</i> mosquito occurrence	study ¹⁴²	1960 - 2014
Dengue occurrence	study ¹⁴³	1960 - 2012
Country-wise Google trends	study ¹⁴⁴	2015 - 2022

6.3 System Architecture

The web application consists of several components as shown in Fig. 6.1, which we broadly categorize into (1) the frontend client application and (2) the server backend.

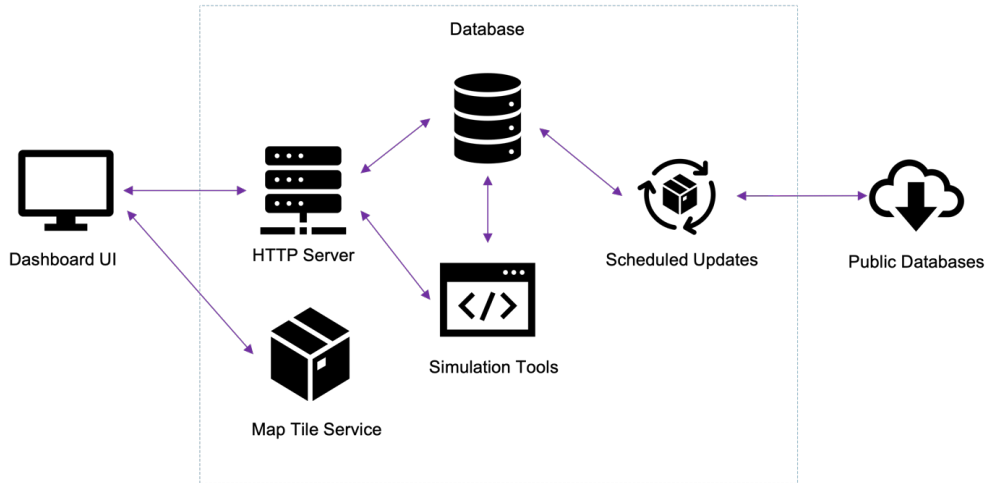


Figure 6.1: Components and architecture of the application

6.3.1 Client Application

The web-based client is built on top of HTML and is accessible using any web browser. The application’s graphical interface consists of an interactive map viewer and tools for data selection, filtering, parameter inputs, and submission of on-demand analyses. The map is displayed using the open-source JavaScript library leaflet.js¹⁴⁵. The application communicates directly with the server backend to load maps, results, and data.

6.3.2 Server Backend

- HTTP Server: The backend system exposes multiple API (application programming interface) endpoints to the client via an HTTP Server. This HTTP server runs in Node.js environment¹⁴⁶;
- Database: The database service consists of PostgreSQL and PostGIS¹⁴⁷ extensions to handle spatial queries. The database contains multiple tables to store data collected from various sources and simulation results;
- Map Tile Service: When the client requests the home page of the map tool, the HTTP server loads map tiles from a separate map tile service called MapTiler¹⁴⁸. This service helps with memory management by loading only the parts and zoom levels required by the client. The service takes input from large tile files collected from OpenStreetMap¹⁴⁹;
- Simulation Tools: The system can deploy and run executable code to perform on-demand analysis and simulations on selected data. The HTTP service directly communicates with the executables to pass arguments and collect results;
- Scheduled Updates: The system contains scripts that run regularly to update the database tables with new weather data. The script connects to the NOAA-NCEI¹³⁷ servers for daily weather observation data.

6.4 Results

6.4.1 Data visualization

Fig. 6.2 displays the graphical interface of the Map Tool, where the 2D projection map of the Earth is centrally located, and users can pan and zoom to their areas of interest. The interface's left side features various tools that assist with data selection and visualization. A

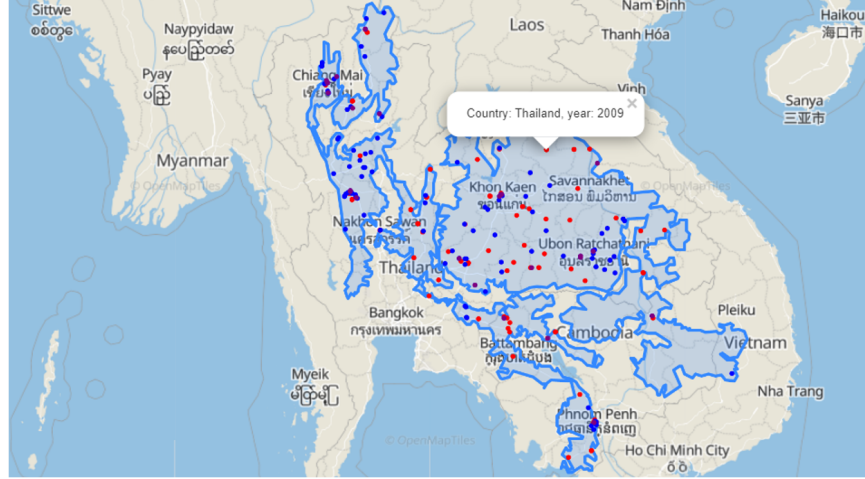


Figure 6.4: Occurrence data of dengue, *Aedes aegypti*, and *Aedes albopictus*

Additionally, the PICTUREE–Aedes supports the display of information on Google Trends. Fig. 6.5 illustrates the trend of "dengue" and "fever" in Cambodia during January 2018.

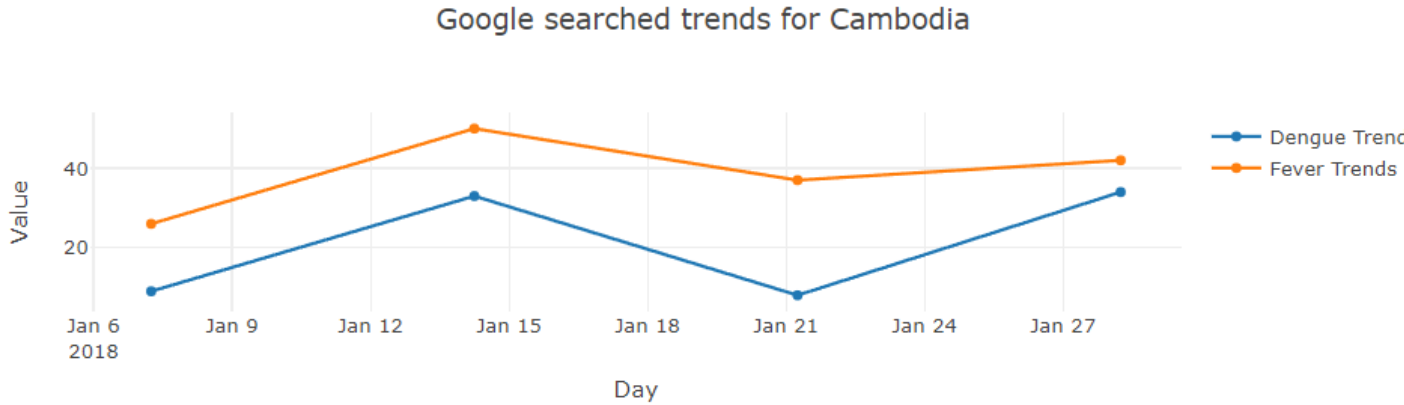


Figure 6.5: Google search trends of "dengue" and "fever"

6.4.2 Simulation results

The PICTUREE–Aedes offers four types of mosquito simulation results, including daily average temperature, vector host ratio, reproduction number, and dengue risk, which can be visualized spatially and temporally. To view the data spatially, users must select a specific

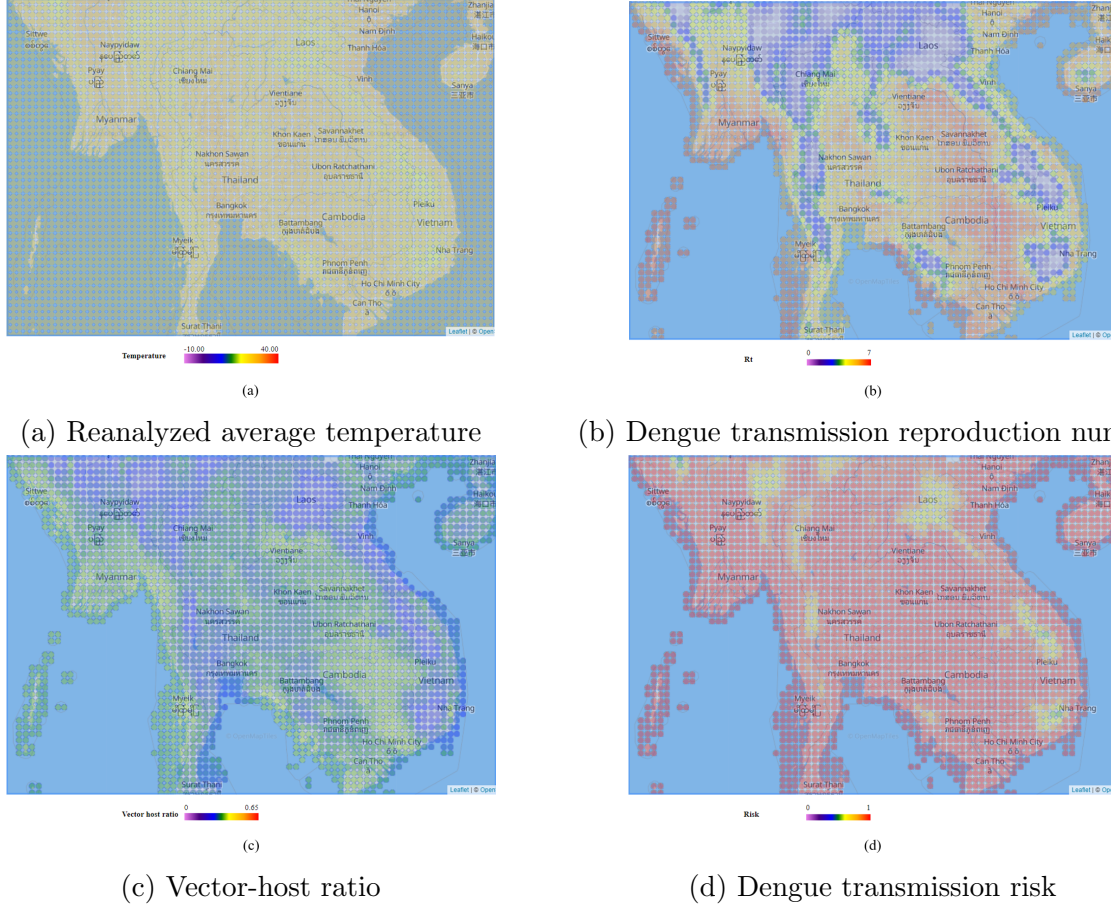


Figure 6.6: Mosquito simulation results visualized spatially

day and a heat map of the selected area will display the selected data value. Fig. 6.6 displays the average temperature (a), dengue transmission reproduction number (b), vector-host ratio(c), and dengue transmission risk (d) of the Indochina peninsula on 09/15/2022. Visualizing data beyond the next six months from the current time is limited.

To visualize data temporally, users must select a location and a time range. Fig. 6.7 displays mosquito simulation data from 01/01/2022 to 11/06/2022 at the user-specified location. Please note that the query time frame is limited to a six-month horizon from the current date.

The PICTUREE–Aedes provides three types of predictive tools and their ensemble forecast. Fig. 6.8 illustrates the predicting results of the dengue outbreak in Cambodia during 2022. The first 13 data points (before April 27th) represent weekly dengue cases, while the rest points (including/after April 27th) denote the predictions. Fig. 6.8a displays the dis-

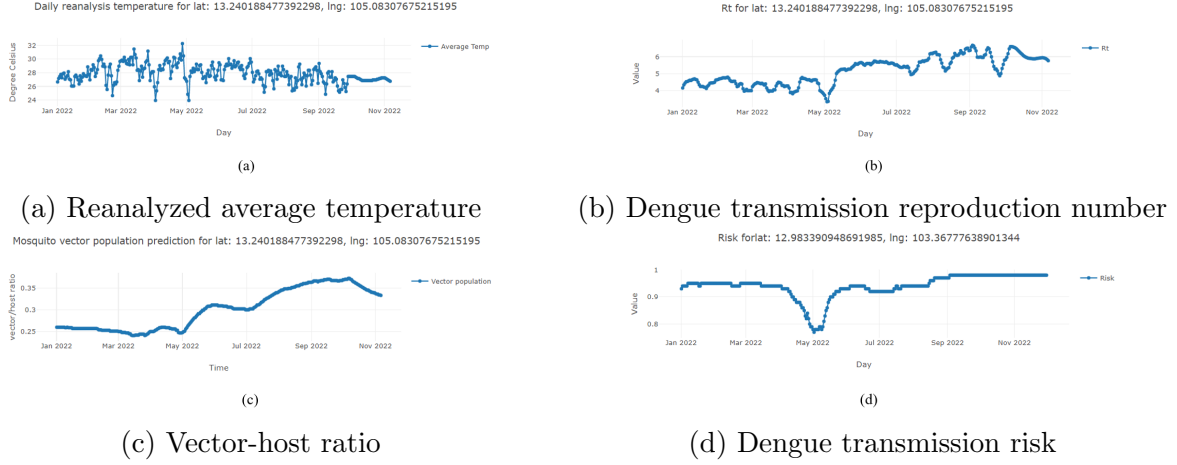


Figure 6.7: Mosquito simulation results visualized temporally

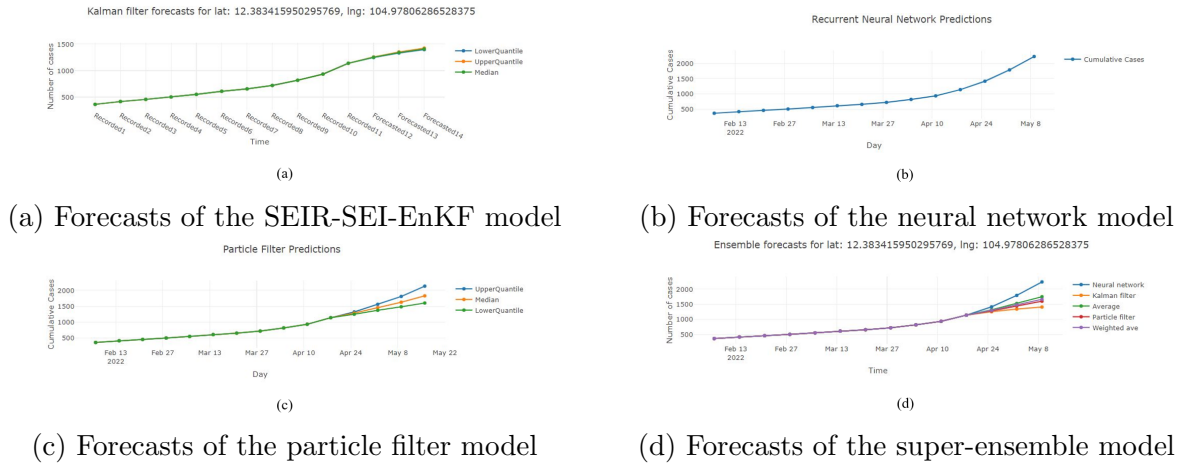


Figure 6.8: Forecasts of the dengue cases in Cambodia

tribution of SEIR-SEI-EnKF predictions for the next three weeks, including quantiles and median. Fig. 6.8b illustrates the neural network prediction for the next three weeks. Fig. 6.8c shows the results of the particle filter, presenting the distribution for the next four weeks. Fig. 6.8d displays the result of the super-ensemble, showcasing the median prediction from each member, the average, and the weighted average of the members.

Table 6.2 displays the outcome of the prediction, as well as the reported cases. All of the reported cases in the following weeks are situated within the prediction range, which confirms the precision of the prediction tools.

Table 6.2: Number of predicted cumulative cases and its relative absolute error in Cambodia

	April 27th 2022	May 4th 2022	May 11th 2022
Reported cases	1379	1685	1995
Kalman filter	1229/10.9%	1293/23.3%	1338/32.9%
Neural network	1414/2.5%	1795/6.5%	2243/12.4%
Particle filter	1298/5.9%	1492/11.5%	1686/15.5%
Super ensemble	1302/5.6%	1500/11.0%	1703/14.6%

6.5 Conclusion

The PICTUREE-Aedes is an innovative and unique system that serves as a valuable resource for both researchers and health officials in their efforts to advance scientific research and effectively manage dengue outbreaks. By collecting and aggregating a comprehensive set of critical factors related to dengue transmission, such as reanalyzed temperature, precipitation, occurrences of *Aedes* mosquitoes, and dengue cases, the PICTUREE-Aedes provides researchers with a rich dataset for their investigations. This data-driven approach allows researchers to gain deeper insights into the complex dynamics of dengue transmission and aids in the development of more effective prevention and control strategies. In addition to its research-oriented features, the PICTUREE-Aedes offers essential information to health officials, enabling them to proactively prepare and adapt to the changing situation during dengue outbreaks. The system employs sophisticated algorithms to estimate *Aedes* mosquito abundance, dengue transmission reproduction number, and dengue risk, providing crucial insights into the intensity and geographical spread of the disease. Armed with this knowledge, health officials can make informed decisions about when and where to implement targeted vector control measures, effectively minimizing the risk of dengue transmission and safeguarding public health. Moreover, the PICTUREE-Aedes goes beyond monitoring and estimation by offering valuable and accurate forecasting capabilities. By utilizing its advanced modeling techniques, the system can predict future dengue case counts, empowering health officials and healthcare facilities to deploy medical personnel and allocate treatment resources more efficiently. This forecasting functionality is instrumental in anticipating the potential impact of dengue outbreaks, enabling proactive measures to be taken in terms of resource allocation,

capacity planning, and public health awareness campaigns.

The selection of a 6-month forecast horizon was made without a specific rationale. In fact, the forecast horizon relies on the availability of dependable environmental variables for the future. Presently, we employ the average temperatures from previous years for the corresponding timeframe in the next six months. However, the current values of the calculated variables are adjusted based on the most recent available environmental data. It's important to note that the current levels will influence the future levels of the calculated variable. However, for the dengue case number forecast horizon we chose 3 weeks because human intervention can change the underlying physical transmission process in the long term.

Chapter 7

Conclusion and future work

7.1 Conclusion

This dissertation has explored various methodologies for modeling and forecasting several infectious diseases, highlighting the crucial role of infectious disease forecasting in mitigating the significant harm these diseases inflict on human life and well-being. With this background, the study focuses on four infectious diseases, tackling five research questions designed to yield accurate forecasts and in-depth analyses of epidemic dynamics. The insights from addressing these research questions are instrumental for public health officials, empowering them to prepare for and respond more effectively to disease outbreaks. The key contributions of this dissertation are summarized as follows:

Chapter 2 unveils a short-term forecasting system for COVID-19 that utilizes the SEIR compartmental model to simulate the disease’s spread, combined with the EnKF to integrate noisy daily confirmed case data. This innovative system generates three-day forecasts, refreshed daily with incoming data, and employs a cone of uncertainty to increase its forecast accuracy.

In Chapter 3, we broaden our focus to dengue fever, a vector-borne disease, by developing a vector-borne forecasting model that integrates the SEIR-SEI disease model with the EnKF method. Through twin experiments, we validate that the model can produce accurate

state and parameter estimation and forecasting. Applying this model to historical dengue outbreak data from diverse locales—Kaohsiung, Taiwan; Singapore; and Rio de Janeiro, Brazil—resulted in forecasts that closely mirrored actual case counts and provided dynamic estimates of exposed and infectious human populations. A comparative evaluation against the SEIR and SIR models demonstrated its enhanced forecasting precision.

Chapter 4 details a real-time prediction system for WNV that features an adapted compartment model SEIR-SEI-SEAC to describe the transmission dynamics among birds, mosquitoes, and humans, considering both symptomatic and asymptomatic cases and the impact of climatic factors. Employing data assimilation techniques, the system provides accurate weekly WNV case forecasts for Colorado in 2023, with comparative analyses underscoring the heightened forecast accuracy achieved through the integration of climatic variables into our model.

Chapter 5 investigates the transmission dynamics of Foot-and-Mouth Disease (FMD) across beef-cattle farms in southwest Kansas using a three-layer network model that includes direct contacts (cattle movements), indirect contacts (truck visits), and information sharing. The study reveals the substantial role of indirect contact in amplifying disease spread and underscores the efficacy of information-sharing strategies in outbreak mitigation and in minimizing the economic fallout from infections and quarantine measures.

Lastly, Chapter 6 presents the PICTUREE—Aedes web application, a robust platform aggregating dengue-related data from multiple sources, including daily temperature and precipitation, *Aedes* mosquito occurrences, and dengue case reports. This application provides estimations of *Aedes* mosquito populations, the dengue transmission reproduction number (R_t), and global dengue risk, all accessible to users spatially and temporally. It also employs several forecasting algorithms—ensemble Kalman filter, recurrent neural network, particle filter, and super ensemble—to predict ongoing dengue outbreak cases.

From the predictions generated by the above methods, specific mitigation strategies can be made to effectively manage and reduce the impact of infectious disease outbreaks. For example, predictions can identify populations at greatest risk or areas where an outbreak is likely to spread. Vaccination efforts can be concentrated in these areas to prevent the

disease from taking hold or to mitigate an ongoing outbreak. Predictive models can also guide the strategic distribution of limited resources such as medical personnel, testing kits, and hospital beds. What’s more, insights from forecasts can inform the timing and nature of public health interventions, such as mask mandates, social distancing requirements, and travel restrictions. By integrating these specific mitigation strategies based on predictive modeling, public health authorities can enact more proactive, precise, and effective responses to infectious disease threats.

7.2 Future work

This section outlines recommendations for future research, building upon the limitations identified in this dissertation. Addressing these limitations presents a pathway to enhancing the accuracy and applicability of forecasting models for infectious diseases.

The forecasting models presented in Chapter 2, Chapter 3, and Chapter 4 could be significantly improved by incorporating the effects of human interventions and public health measures implemented during outbreaks. Local health agencies have carried out various actions, such as lockdowns, mask mandates, and vaccination drives, which have substantial impacts on disease transmission dynamics. Future work should focus on integrating these human intervention factors into the forecasting model to enhance its predictive accuracy and usefulness for public health planning. In addition, given the diversity of host species and virus serotypes, there is a clear need for more detailed data collection. Future work should aim to gather comprehensive information on the transmission dynamics of different virus serotypes across different host species and integrate this data into the model, potentially improving the accuracy and specificity of disease forecasts.

In Chapter 5, the model’s consideration of indirect contacts was limited to truck visits, neglecting other significant transmission vectors such as personnel visits and airborne transmission. The exclusion of auction markets from the analysis also represents a gap in understanding the full scope of FMD transmission. Future studies should aim to incorporate these additional indirect contact pathways and examine the role of auction markets in

disease spread.

The PICTUREE—Aedes web application, discussed in Chapter 6, relies on temperatures recorded at weather stations, which may not accurately reflect micro-climatic conditions in nearby areas. Moreover, the model’s consideration of water sources for *Aedes* mosquito breeding is limited to precipitation, omitting other critical sources such as wastewater and artificial water bodies. Future developments should focus on integrating local temperature data and a broader range of mosquito breeding sites into the model to refine risk assessments and forecasts.

In addition, the future integration of artificial intelligence (AI) into infectious disease forecasting will lead to a dramatic improvement in our capacity to predict, manage, and mitigate the impacts of epidemics and pandemics^{150;151}. AI, with its unparalleled computational power and learning capabilities, keeps revolutionizing our approach to disease prediction by analyzing vast datasets ranging from real-time health data to social media trends and environmental changes, adapting to new information, and refining its predictions in real-time as outbreaks evolve.

Across all chapters, a recurring theme is the need for more detailed data collection and the incorporation of a wider range of environmental and human factors into disease transmission models. Future research should also explore the development of more sophisticated data assimilation techniques and algorithms to enhance model accuracy and predictive capabilities.

Bibliography

- [1] Jeffery K. Taubenberger and David M. Morens. 1918 influenza: the mother of all pandemics. *Revista Biomedica*, 17(1):69–79, 2006.
- [2] World Health Organization. Dengue and severe dengue. World Health Organization, 2021. URL <https://www.who.int/news-room/q-adeTAIL/dengue-and-severe-dengue>.
- [3] World Health Organization. Influenza (seasonal). World Health Organization, 2021. URL [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)).
- [4] World Health Organization. Weekly epidemiological update on covid-19 - 27 april 2023. World Health Organization, 2023. URL <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-1927-april-2023>.
- [5] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, and et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: A descriptive study. *The Lancet*, 395:507–513, 2020. URL [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
- [6] J. F. W. Chan, S. Yuan, K. H. Kok, K. K. W. To, H. Chu, J. Yang, and et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*, 395:514–523, 2020. URL [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9).
- [7] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, and et al. Transmission of 2019-ncov infection from an asymptomatic contact in germany. *New*

- England Journal of Medicine*, 382:970–971, 2020. URL <https://doi.org/10.1056/NEJMc2001468>.
- [8] C. Fitzmaurice et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA Oncology*, 3(4):524–548, 2017.
 - [9] J. Rajarethinam et al. Dengue in singapore from 2004 to 2016: cyclical epidemic patterns dominated by serotypes 1 and 2. *The American Journal of Tropical Medicine and Hygiene*, 99(1):204, 2018.
 - [10] E. J. Luna et al. A cohort study to assess the incidence of dengue, brazil, 2014–2018. *Acta Tropica*, 204:105313, 2020.
 - [11] W.-J. Chen. Dengue outbreaks and the geographic distribution of dengue vectors in taiwan: A 20-year epidemiological analysis. *Biomedical Journal*, 41(5):283–289, 2018.
 - [12] R.S. Lanciotti, J.T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K.E. Volpe, M.B. Crabtree, J.H. Scherret, R.A. Hall, J.S. MacKenzie, C.B. Cropp, B. Panigrahy, E. Ostlund, B. Schmitt, M. Malkinson, C. Banet, J. Weissman, N. Komar, H.M. Savage, W. Stone, T. McNamara, and D.J. Gubler. Origin of the west nile virus responsible for an outbreak of encephalitis in the northeastern united states. *Science*, 286:2333–2337, 1999. URL <https://doi.org/10.1126/science.286.5448.2333>.
 - [13] P. Revesz and S. Wu. Spatiotemporal reasoning about epidemiological data. *Artificial Intelligence in Medicine*, 38:157–170, 2006.
 - [14] J. C. Guth, S. A. Futterer, T. A. Hijaz, E. M. Liotta, N. F. Rosenberg, A. M. Naidech, and M. B. Maas. Pearls & oy-sters: bilateral thalamic involvement in west nile virus encephalitis. *Neurology*, 83:e16–e17, 2014.
 - [15] Current season west nile virus (wnv) infection statistics and maps. CDC (Centers

- for Disease Control and Prevention), 2024. URL <https://www.cdc.gov/westnile/statsmaps/current-season-data.html>.
- [16] T. J. Knight-Jones and J. Rushton. The economic impacts of foot and mouth disease—what are they, how big are they and where do they occur? *Preventive Veterinary Medicine*, 112(3–4):161–173, 2013.
 - [17] D. Thompson et al. Economic costs of the foot and mouth disease outbreak in the united kingdom in 2001. *Revue scientifique et technique (International Office of Epizootics)*, 21(3):675–687, 2002.
 - [18] N. Muroga, Y. Hayama, T. Yamamoto, A. Kurogi, T. Tsuda, and T. Tsutsui. The 2010 foot-and-mouth disease epidemic in japan. *Journal of Veterinary Medical Science*, 74(4):399–404, 2012.
 - [19] A. Bouma et al. The foot-and-mouth disease epidemic in the netherlands in 2001. *Preventive Veterinary Medicine*, 57(3):155–166, 2003.
 - [20] S. N. Balinda et al. Prevalence estimates of antibodies towards foot-and-mouth disease virus in small ruminants in uganda. *Transboundary and Emerging Diseases*, 56(9–10):362–371, 2009.
 - [21] D. L. Pendell, J. Leatherman, T. C. Schroeder, and G. S. Alward. The economic impacts of a foot-and-mouth disease outbreak: a regional analysis. *Journal of Agricultural and Applied Economics*, 39(s1):19–33, 2007.
 - [22] A. Barrat, M. Barthelemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.
 - [23] H. Nishiura, S. Jung, N. M. Linton, R. Kinoshita, Y. Yang, K. Hayashi, and et al. The extent of transmission of novel coronavirus in wuhan, china, 2020. *Journal of Clinical Medicine*, 9:330, 2020. URL <https://doi.org/10.3390/jcm9020330>.

- [24] Z. Du, L. Wang, S. Cauchemez, X. Xu, X. Wang, B. J. Cowling, and L. A. Meyers. Early release—risk for transportation of 2019 novel coronavirus disease from wuhan to other cities in china. *Emerging Infectious Diseases Journal - CDC*, 26:5, 2020. URL <https://doi.org/10.3201/eid2605.200146>.
- [25] J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: A modelling study. *The Lancet*, 2020. URL [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
- [26] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, and et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*, 92:214–217, 2020. URL <https://doi.org/10.1016/j.ijid.2020.01.050>.
- [27] W. Guan, Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, and et al. Clinical characteristics of coronavirus disease 2019 in china. *New England Journal of Medicine*, 2020. URL <https://doi.org/10.1056/NEJMoa2002032>.
- [28] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, and et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395: 497–506, 2020. URL [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- [29] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell. Novel coronavirus 2019-ncov (covid-19): early estimation of epidemiological parameters and epidemic size estimates. *Philosophical Transactions of the Royal Society B*, 376:20200265, 2021. URL <https://doi.org/10.1098/rstb.2020.0265>.
- [30] X. Li, X. Zhao, and Y. Sun. The lockdown of hubei province causing different transmission dynamics of the novel coronavirus (2019-ncov) in wuhan and beijing. *MedRxiv*, 2020. URL <https://doi.org/10.1101/2020.02.09.20021477>.

- [31] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S. S. Musa, and et al. A conceptual model for the coronavirus disease 2019 (covid-19) outbreak in wuhan, china with individual reaction and governmental action. *International Journal of Infectious Diseases*, 93: 211–216, 2020. URL <https://doi.org/10.1016/j.ijid.2020.02.058>.
- [32] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960. URL <https://doi.org/10.1115/1.3662552>.
- [33] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Courier Corporation, 2007.
- [34] J. P. Jensen. Ensemble kalman filtering for state and parameter estimation on a reservoir model. NTNU Open, 2007. URL <https://ntnuopen.ntnu.no/ntnu-xmli/handle/11250/259624>.
- [35] H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources*, 28:135–147, 2005. URL <https://doi.org/10.1016/j.advwatres.2004.09.002>.
- [36] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368:489–493, 2020. URL <https://doi.org/10.1126/science.abb3221>.
- [37] M. G. Guzman. A new moment for facing dengue? *Pathogens and Global Health*, 109: 1–2, 2015.
- [38] S. Bhatt et al. The global distribution and burden of dengue. *Nature*, 496(7446): 504–507, 2013.
- [39] J.-J. Tsai et al. Seroprevalence of dengue virus in two districts of kaohsiung city after the largest dengue outbreak in taiwan since world war ii. *PLoS Neglected Tropical Diseases*, 12(10):e0006879, 2018.

- [40] T. K. Yamana, S. Kandula, and J. Shaman. Superensemble forecasts of dengue outbreaks. *Journal of The Royal Society Interface*, 13(123):20160410, 2016.
- [41] S. Venkatramanan, B. Lewis, J. Chen, D. Higdon, A. Vullikanti, and M. Marathe. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics*, 22:43–49, 2018.
- [42] C. Viboud et al. The rapid ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, 2018.
- [43] P. Narula, V. Piratla, A. Bansal, S. Azad, and P. Lio. Parameter estimation of tuberculosis transmission model using ensemble kalman filter across indian states and union territories. *Infection, Disease & Health*, 21(4):184–191, 2016.
- [44] Z. Zhan, W. Dong, Y. Lu, P. Yang, Q. Wang, and P. Jia. Real-time forecasting of hand-foot-and-mouth disease outbreaks using the integrating compartment model and assimilation filtering. *Scientific Reports*, 9(1):1–9, 2019.
- [45] R. M. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1992.
- [46] J. Shaman and A. Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.
- [47] L. Esteva and C. Vargas. Analysis of a dengue disease transmission model. *Mathematical Biosciences*, 150(2):131–151, 1998.
- [48] M. Andraud, N. Hens, C. Marais, and P. Beutels. Dynamic epidemiological models for dengue transmission: a systematic review of structural approaches. *PloS one*, 7(11): e49085, 2012.
- [49] M. Otero, D. H. Barmak, C. O. Dorso, H. G. Solari, and M. A. Natiello. Modeling dengue outbreaks. *Mathematical Biosciences*, 232(2):87–95, 2011.

- [50] R. A. Erickson, S. M. Presley, L. J. Allen, K. R. Long, and S. B. Cox. A dengue model with a dynamic aedes albopictus vector population. *Ecological Modelling*, 221(24):2899–2908, 2010.
- [51] F. Liang, P. Guan, W. Wu, and D. Huang. Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in liaoning, from 2011 to 2015. *PeerJ*, 6:e5134, 2018.
- [52] H. Hu, H. Wang, F. Wang, D. Langley, A. Avram, and M. Liu. Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. *Scientific reports*, 8(1):1–8, 2018.
- [53] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.
- [54] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- [55] J. P. Jensen. Ensemble kalman filtering for state and parameter estimation on a reservoir model. Master’s thesis, Institutt for teknisk kybernetikk, 2007.
- [56] H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources*, 28(2):135–147, 2005.
- [57] R. Fan, Z. Huang, S. Wang, R. Diao, and D. Meng. Dynamic state estimation and parameter calibration of a dfig using the ensemble kalman filter. In *2015 IEEE Power & Energy Society General Meeting*, pages 1–5, 2015.
- [58] C. W. Morin, A. J. Monaghan, M. H. Hayden, R. Barrera, and K. Ernst. Meteorologically driven simulations of dengue epidemics in san juan, pr. *PLoS Neglected Tropical Diseases*, 9(8):e0004002, 2015.

- [59] Q. Yang et al. Short-term forecasts and long-term mitigation evaluations for the covid-19 epidemic in hubei province, china. *medRxiv*, 2020.
- [60] Yang H. M., M. L. G. Macoris, K. C. Galvani, M. T. M. Andrighetti, and D. M. V. Wanderley. Assessing the effects of temperature on dengue transmission. *Epidemiology & Infection*, 137(8):1179–1187, 2009.
- [61] W. Yang, A. Karspeck, and J. Shaman. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS Computational Biology*, 10(4):e1003583, 2014.
- [62] Taiwan national infectious disease statistics system. <https://nidss.cdc.gov.tw/en/nidss/disease?id=061>, 2020. Accessed: 2020-12-12.
- [63] Singapore national environment agency. <https://www.nea.gov.sg/dengue-zika/dengue/dengue-cases>, 2020. Accessed: 2020-12-12.
- [64] E. Mussumeci and F. C. Coelho. Large-scale multivariate forecasting models for dengue-lstm versus random forest regression. *Spatial and Spatio-temporal Epidemiology*, 35:100372, 2020.
- [65] A. Vajdi. Personal communication, Feb 2021.
- [66] World Health Organization. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*. World Health Organization, Geneva, 2009.
- [67] K. Sasaki, A. Yamanaka, S. Ito, and H. Nagao. Data assimilation for phase-field models based on the ensemble kalman filter. *Computational Materials Science*, 141:141–152, 2018.
- [68] J.T. Roehrig. West nile virus in the united states—a historical perspective. *Viruses*, 5:3088–3108, 2013.

- [69] M. Maleki, M.R. Mahmoudi, D. Wraith, and K.-H. Pho. Time series modelling to forecast the confirmed and recovered cases of covid-19. *Travel medicine and infectious disease*, 37:101742, 2020.
- [70] W. Anggraeni, S. Sumpeno, E. M. Yuniarno, R. F. Rachmadi, A. B. Gumelar, and M. H. Purnomo. Prediction of dengue fever outbreak based on climate factors using fuzzy-logistic regression. *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 199–204, 2020.
- [71] K. M. Holcomb et al. Multi-model prediction of west nile virus neuroinvasive disease with machine learning for identification of important regional climatic drivers. *Geo-Health*, 7:e2023GH000906, 2023. URL <https://doi.org/10.1029/2023GH000906>.
- [72] C. Yi, Q. Yang, and C.M. Scoglio. Multilayer network analysis of fmd transmission and containment among beef cattle farms. *Scientific Reports*, 12:15679, 2022.
- [73] N. B. DeFelice, E. Little, S. R. Campbell, and J. Shaman. Ensemble forecast of human west nile virus cases and mosquito infection rates. *Nature Communications*, 8:14592, 2017.
- [74] M.S. McCarter, S. Self, K.C. Dye-Braumuller, C. Lee, H. Li, M. Nolan, and M.S. Nolan. The utility of a bayesian predictive model to forecast neuroinvasive west nile virus disease in the united states of america, 2022. *Plos one*, 18:e0290873, 2023.
- [75] K. M. Holcomb et al. Evaluation of an open forecasting challenge to assess skill of west nile virus neuroinvasive disease prediction. *Parasites & Vectors*, 16:11, 2023. URL <https://doi.org/10.1186/s13071-022-05630-y>.
- [76] D. M. Hartley, C. M. Barker, A. Le Menach, T. Niu, H. D. Gaff, and W. K. Reisen. Effects of temperature on emergence and seasonality of west nile virus in california. *The American Journal of Tropical Medicine and Hygiene*, 86:884, 2012.
- [77] K.H. Smith, A.J. Tyre, J. Hamik, M.J. Hayes, Y. Zhou, and L. Dai. Using climate

- to explain and predict west nile virus risk in nebraska. *GeoHealth*, 4:e2020GH000244, 2020. URL <https://doi.org/10.1029/2020GH000244>.
- [78] J.E. Soverow, G.A. Wellenius, D.N. Fisman, and M.A. Mittleman. Infectious disease in a warming world: How weather influenced west nile virus in the united states (2001–2005). *Environ Health Perspect*, 117:1049–1052, 2009. URL <https://doi.org/10.1289/ehp.0800487>.
- [79] S. Funk, A. Camacho, A. J. Kucharski, R. M. Eggo, and W. J. Edmunds. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics*, 22:56–61, 2018.
- [80] Z. Zhan, W. Dong, Y. Lu, P. Yang, Q. Wang, and P. Jia. Real-time forecasting of hand-foot-and-mouth disease outbreaks using the integrating compartment model and assimilation filtering. *Scientific reports*, 9:1–9, 2019.
- [81] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A*, 115:700–721, 1927. URL <https://doi.org/10.1098/rspa.1927.0118>.
- [82] Q. Yang, C. Yi, A. Vajdi, L.W. Cohnstaedt, H. Wu, X. Guo, and C.M. Scoglio. Short-term forecasts and long-term mitigation evaluations for the covid-19 epidemic in hubei province, china. *Infectious Disease Modelling*, 5:563–574, 2020.
- [83] P.E. Lekone and B.F. Finkenstädt. Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62:1170–1177, 2006. URL <https://doi.org/10.1111/j.1541-0420.2006.00609.x>.
- [84] C. Yi, L.W. Cohnstaedt, and C.M. Scoglio. Seir-sei-enkf: A new model for estimating and forecasting dengue outbreak dynamics. *IEEE Access*, 9:156758–156767, 2021.
- [85] Matthew James Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, NJ, 2008. Illustrated, Hardcover.

- [86] R. Lal, W. Huang, and Z. Li. An application of the ensemble kalman filter in epidemiological modelling. *Plos one*, 16:e0256227, 2021.
- [87] C. Yi, A. Vajdi, T. Ferdousi, L.W. Cohnstaedt, and C. Scoglio. Picturee—aedes: A web application for dengue data visualization and case prediction. *Pathogens*, 12:771, 2023.
- [88] A. Safarishahrbijari, A. Teyhouee, C. Waldner, J. Liu, and N.D. Osgood. Predictive accuracy of particle filtering in dynamic models supporting outbreak projections. *BMC Infect Dis*, 17:648, 2017. URL <https://doi.org/10.1186/s12879-017-2726-9>.
- [89] West nile virus data. Colorado Department of Public Health & Environment (CDPHE), 2024. URL <https://cdphe.colorado.gov/animal-related-diseases/west-nile-virus/west-nile-virus-data>.
- [90] Climate data for the united states. US Climate Data, 2024. URL <https://www.usclimatedata.com/climate/united-states/us>.
- [91] S. Zou, G.A. Foster, R.Y. Dodd, L.R. Petersen, and S.L. Stramer. West nile fever characteristics among viremic persons identified through blood donor screening. *The Journal of infectious diseases*, 202:1354–1361, 2010.
- [92] K.A. Pawelek, P. Niehaus, C. Salmeron, E.J. Hager, and G.J. Hunt. Modeling dynamics of culex pipiens complex populations and assessing abatement strategies for west nile virus. *PloS one*, 9:e108452, 2014.
- [93] J.E. Ruybal, L.D. Kramer, and A.M. Kilpatrick. Geographic variation in the response of culex pipiens life history traits to temperature. *Parasites & Vectors*, 9:116, 2016. URL <https://doi.org/10.1186/s13071-016-1402-z>.
- [94] E. M. Fèvre, B. M. de C. Bronsvoort, K. A. Hamilton, and S. Cleaveland. Animal movements and the spread of infectious diseases. *Trends in Microbiology*, 14(3): 125–131, 2006.

- [95] T. Porphyre, B. M. de C. Bronsvoort, G. J. Gunn, and C. Correia-Gomes. Multilayer network analysis unravels haulage vehicles as a hidden threat to the british swine industry. *Transboundary and Emerging Diseases*, 67(3):1231–1246, 2020.
- [96] K. L. VanderWaal et al. Network analysis of cattle movements in uruguay: quantifying heterogeneity for risk-based disease surveillance and control. *Preventive Veterinary Medicine*, 123:12–22, 2016.
- [97] G. Rossi, R. L. Smith, S. Pongolini, and L. Bolzoni. Modelling farm-to-farm disease transmission through personnel movements: from visits to contacts, and back. *Scientific Reports*, 7(1):1–11, 2017.
- [98] M. Nöremark, J. Frössling, and S. S. Lewerin. A survey of visitors on swedish livestock farms with reference to the spread of animal diseases. *BMC Veterinary Research*, 9(1):1–10, 2013.
- [99] T. W. Bates, M. C. Thurmond, and T. E. Carpenter. Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three california counties, with reference to control of potential foot-and-mouth disease transmission. *American Journal of Veterinary Research*, 62(7):1121–1129, 2001.
- [100] K. K. Thakur, J. Sanchez, D. Hurnik, Z. Poljak, S. Opps, and C. W. Revie. Development of a network based model to simulate the between-farm transmission of the porcine reproductive and respiratory syndrome virus. *Veterinary Microbiology*, 180(3–4):212–222, 2015.
- [101] A. Bernini, L. Bolzoni, and R. Casagrandi. When resolution does matter: Modelling indirect contacts in dairy farms at different levels of detail. *PloS ONE*, 14(10):e0223652, 2019.
- [102] G. Rossi et al. The potential role of direct and indirect contacts on infection spread in dairy farm networks. *PLoS Computational Biology*, 13(1):e1005301, 2017.

- [103] Q. Yang et al. Impact of truck contamination and information sharing on foot-and-mouth disease spreading in beef cattle production systems. *PloS ONE*, 15(10): e0240819, 2020.
- [104] N. Q. Viet, B. Behdani, and J. Bloemhof. The value of information in supply chain decisions: A review of the literature and research agenda. *Computers & Industrial Engineering*, 120:68–82, 2018.
- [105] M. Kim and S. Chai. The impact of supplier innovativeness, information sharing and strategic sourcing on improving supply chain agility: Global supply chain perspective. *International Journal of Production Economics*, 187:42–52, 2017.
- [106] C. Granell, S. Gómez, and A. Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical Review Letters*, 111(12):128701, 2013.
- [107] C. Zheng, C. Xia, Q. Guo, and M. Dehmer. Interplay between sir-based disease spreading and awareness diffusion on multiplex networks. *Journal of Parallel and Distributed Computing*, 115:20–28, 2018.
- [108] F. Sahneh, F. Chowdhury, G. Brase, and C. Scoglio. Individual-based information dissemination in multilayer epidemic modeling. *Mathematical Modelling of Natural Phenomena*, 9(2):136–152, 2014.
- [109] I. C. Pappa, C. Iliopoulos, and T. Massouras. What determines the acceptance and use of electronic traceability systems in agri-food supply chains? *Journal of Rural Studies*, 58:123–135, 2018.
- [110] L. Wiseman, J. Sanderson, A. Zhang, and E. Jakku. Farmers and their data: An examination of farmers’ reluctance to share their data through the lens of the laws impacting smart farming. *NJAS-Wageningen Journal of Life Sciences*, 90:100301, 2019.

- [111] Q. Yang, C. Scoglio, and D. Gruenbacher. Eager: Ssdim: Data generation for the coupled system composed of the beef cattle production infrastructure and the transportation services infrastructure in southwestern kansas. *Designsafe-CI*, 2019. doi: 10.17603/DS2-3FT2-0441.
- [112] P. Erdos, A. Renyi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.
- [113] F. D. Sahneh, C. Scoglio, and P. Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.
- [114] A. Wilder-Smith, E.-E. Ooi, O. Horstick, and B. Wills. Dengue. *The Lancet*, 393(10169):350–363, 2019.
- [115] J. Rocklöv, M. B. Quam, B. Sudre, M. German, M. U. G. Kraemer, O. Brady, and I. I. Bogoch et al. Assessing seasonal risks for the introduction and mosquito-borne spread of zika virus in europe. *EBioMedicine*, 9:250–256, 2016.
- [116] S. Polwiang. The time series seasonal patterns of dengue fever and associated weather variables in bangkok (2003-2017). *BMC infectious diseases*, 20(1):1–10, 2020.
- [117] P.-C. Wu, H.-R. Guo, S.-C. Lung, C.-Y. Lin, and H.-J. Su. Weather as an effective predictor for occurrence of dengue fever in taiwan. *Acta tropica*, 103(1):50–57, 2007.
- [118] H. M. Yang, M. D. L. D. G. Macoris, K. C. Galvani, M. T. M. Andrighetti, and D. M. V. Wanderley. Assessing the effects of temperature on the population of aedes aegypti, the vector of dengue. *Epidemiology & Infection*, 137(8):1188–1202, 2009.
- [119] D. M. Watts, D. S. Burke, B. A. Harrison, R. E. Whitmire, and A. Nisalak. Effect of temperature on the vector efficiency of aedes aegypti for dengue 2 virus. *ARMY MEDICAL RESEARCH INST OF INFECTIOUS DISEASES FORT DETRICK MD*, 1986.

- [120] C. J. Struchiner, J. Rocklöv, A. Wilder-Smith, and E. Massad. Increasing dengue incidence in singapore over the past 40 years: population growth, climate and mobility. *PloS one*, 10(8):e0136286, 2015.
- [121] D. Ae Focks, D. G. Haile, E. Daniels, and G. Ae Mount. Dynamic life table model for aedes aegypti (diptera: Culicidae): analysis of the literature and model development. *Journal of medical entomology*, 30(6):1003–1017, 1993.
- [122] M. H. Riad, L. W. Cohnstaedt, and C. M. Scoglio. Risk assessment of dengue transmission in bangladesh using a spatiotemporal network model and climate data. *The American journal of tropical medicine and hygiene*, 104(4):1444, 2021.
- [123] L. Eisen, A. J. Monaghan, S. Lozano-Fuentes, D. F. Steinhoff, M. H. Hayden, and P. E. Bieringer. The impact of temperature on the bionomics of aedes (stegomyia) aegypti, with special reference to the cool geographic range margins. *Journal of medical entomology*, 51(3):496–516, 2014.
- [124] Jannelle Couret, Ellen Dotson, and Mark Q. Benedict. Temperature, larval diet, and density effects on development rate and survival of aedes aegypti (diptera: Culicidae). *PloS one*, 9(2):e87468, 2014.
- [125] Philippe Barbazan, M. Guiserix, W. Boonyuan, W. Tuntaprasart, D. Pontier, and J-P. Gonzalez. Modelling the effect of temperature on transmission of dengue. *Medical and veterinary entomology*, 24(1):66–73, 2010.
- [126] Edson Zangiacomi Martinez, Elisângela Aparecida Soares da Silva, and Amaury Lelis Dal Fabbro. A sarima forecasting model to predict the number of cases of dengue in campinas, state of são paulo, brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, 44:436–440, 2011.
- [127] Logan C. Brooks, David C. Farrow, Sangwon Hyun, Ryan J. Tibshirani, and Roni Rosenfeld. Flexible modeling of epidemics with an empirical bayes framework. *PLoS computational biology*, 11(8):e1004382, 2015.

- [128] Yuliang Chen, Tao Liu, Xiaolin Yu, Qinghui Zeng, Zixi Cai, Haisheng Wu, and Qingying Zhang et al. An ensemble forecast system for tracking dynamics of dengue outbreaks and its validation in china. *PLoS Computational Biology*, 18(6):e1010218, 2022.
- [129] Chunlin Yi, Lee W. Cohnstaedt, and Caterina M. Scoglio. Seir-sei-enkf: a new model for estimating and forecasting dengue outbreak dynamics. *IEEE Access*, 9:156758–156767, 2021.
- [130] Nicholas B. DeFelice, Eliza Little, Scott R. Campbell, and Jeffrey Shaman. Ensemble forecast of human west nile virus cases and mosquito infection rates. *Nature Communications*, 8(1):14592, 2017.
- [131] M. Javvad ur Rehman, Sarat Chandra Dass, Vijanth Sagayan Asirvadam, and Ahmed Adly. Parameter estimation for nonlinear disease dynamical system using particle filter. In *2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pages 143–147. IEEE, 2015.
- [132] Xinxing Zhao, Kainan Li, Candice Ke En Ang, and Kang Hao Cheong. A deep learning based hybrid architecture for weekly dengue incidences forecasting. *Chaos, Solitons & Fractals*, 168:113170, 2023.
- [133] Tanvir Ferdousi, Lee W. Cohnstaedt, and Caterina M. Scoglio. A windowed correlation-based feature selection method to improve time series prediction of dengue fever cases. *IEEE Access*, 9:141210–141222, 2021.
- [134] Jiucheng Xu, Keqiang Xu, Zhichao Li, Fengxia Meng, Taotian Tu, Lei Xu, and Qiyong Liu. Forecast of dengue cases in 20 chinese cities based on the deep learning method. *International journal of environmental research and public health*, 17(2):453, 2020.
- [135] Soudeep Deb and Sougata Deb. An ensemble method for early prediction of dengue outbreak. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(1):84–101, 2022.

- [136] Teresa K. Yamana, Sasikiran Kandula, and Jeffrey Shaman. Superensemble forecasts of dengue outbreaks. *Journal of The Royal Society Interface*, 13(123):20160410, 2016.
- [137] NOAA National Centers for Environmental Information. Global surface summary of the day - gsod. NOAA National Centers for Environmental Information, 2001. Available online: <https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00516> (accessed on 3rd April 2023).
- [138] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J-N. Thépaut. Era5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2023. DOI: 10.24381/cds.adbb2d47 (Accessed on 03-Apr-2023).
- [139] David M. Olson and Eric Dinerstein. The global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical garden*, pages 199–224, 2002.
- [140] USDA Forest Service The Nature Conservancy and U.S. Geological Survey. Description of the ecoregions of the united states, 1995. Based on Bailey, Robert G. 1995. Misc. Pub. No. 1391, Map scale 1:7,500,000. USDA Forest Service. 108pp.
- [141] E. B. Wiken. Terrestrial ecozones of canada. ecological land classification series no. 19. Environment Canada, Hull Quebec, 1986.
- [142] Moritz UG Kraemer, Marianne E. Sinka, Kirsten A. Duda, Adrian Mylne, Freya M. Shearer, Oliver J. Brady, and Jane P. Messina et al. The global compendium of aedes aegypti and ae. albopictus occurrence. *Scientific data*, 2(1):1–8, 2015.
- [143] Jane P. Messina, Oliver J. Brady, David M. Pigott, John S. Brownstein, Anne G. Hoen, and Simon I. Hay. A global compendium of human dengue virus occurrence. *Scientific Data*, 1(1):1–6, 2014.

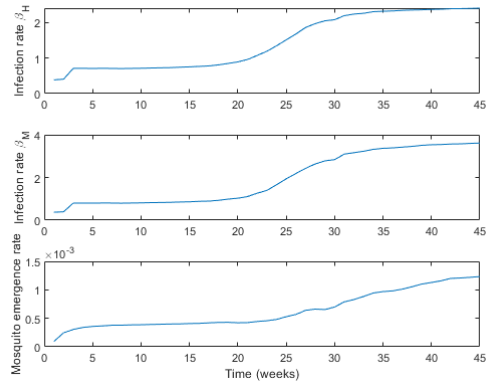
- [144] dengue fever - explore - google trends. Google Trends, 2023. Available online: <https://trends.google.com/trends/explore?date=now%201d&geo=US&q=dengue,fever&hl=en> (accessed on 3rd April 2023).
- [145] Dennis Edler and Mark Vetter. The simplicity of modern audiovisual web cartography: An example with the open-source javascript library leaflet.js. *KN-Journal of Cartography and Geographic Information*, 69(1):51–62, 2019.
- [146] Stefan Tilkov and Steve Vinoski. Node.js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80–83, 2010.
- [147] Esteban Zimányi, Mahmoud Sakr, and Arthur Lesuisse. Mobilitydb: A mobility database based on postgresql and postgis. *ACM Transactions on Database Systems (TODS)*, 45(4):1–42, 2020.
- [148] Jesús-Daniel Cascón-Katchadourian and Antonio Ángel Ruiz-Rodríguez. Description and evaluation of maptiler software: From scanned map to interactive layer for web publishing, 2016.
- [149] Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008.
- [150] Z. Farooq, J. Rocklöv, J. Wallin, N. Abiri, M.O. Sewe, H. Sjödin, and J.C. Semenza. Artificial intelligence to predict west nile virus outbreaks with eco-climatic drivers. *The Lancet Regional Health–Europe*, 17, 2022.
- [151] S. Agrebi and A. Larbi. Use of artificial intelligence in infectious diseases. *Artificial intelligence in precision health*, pages 415–438, 2020.

Appendix A

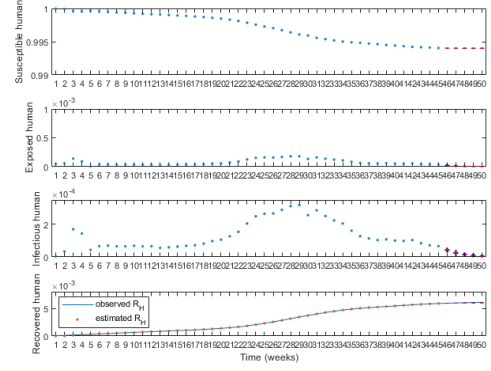
Estimations and forecasts of dengue outbreak for Singapore and RDJ

A.1 Estimations and forecasts of SEIR-SEI-EnKF model for Singapore and RDJ

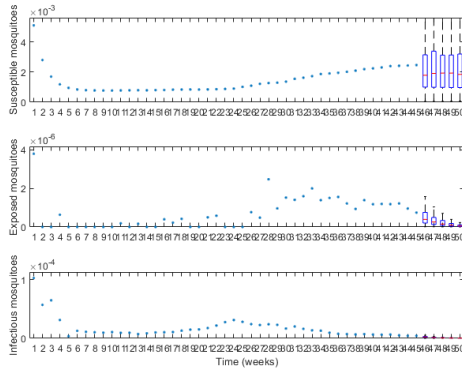
Fig.A.1 and Fig.A.2 show the parameter estimations, estimated and predicted dengue outbreak dynamics in Singapore and Rio de Janeiro Brazil based on the SEIR-SEI-EnKF model. The dengue case data from epidemic week 1/2020 to 50/2020 in Singapore and the case data from epidemic week 7/2015 to 42/2015 were used as the observations. The simulation results are similar to the ones in Kaohsiung Taiwan. The curves for the infection rates are similar and the value of β_{M} is larger; the estimated and predicted recovered population corresponds to the actual data. The epidemic peak time is estimated at week 29 in Singapore and week 13 in RDJ.



(a) Parameter estimation

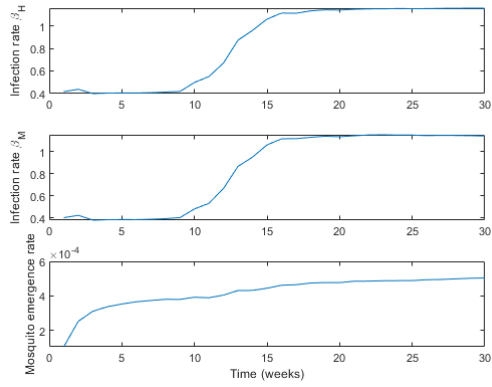


(b) Epidemic dynamics on human population

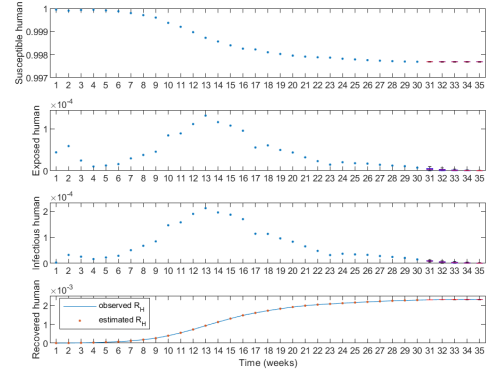


(c) Epidemic dynamics on mosquito population

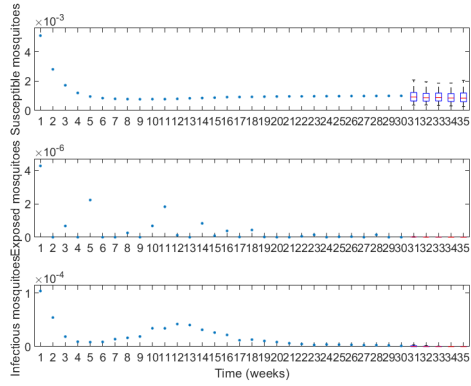
Figure A.1: Estimation and prediction results for 2020 dengue outbreak in Singapore



(a) Parameter estimation

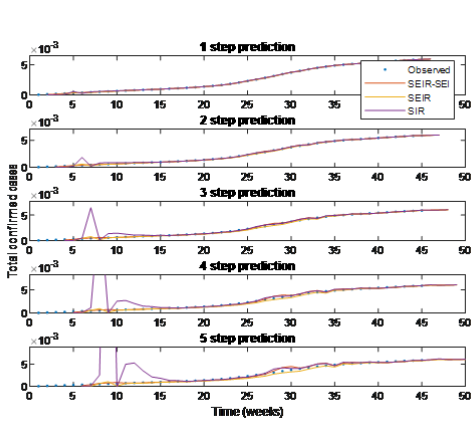


(b) Epidemic dynamics on human population

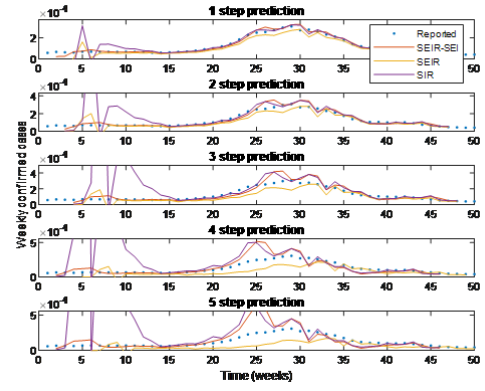


(c) Epidemic dynamics on mosquito population

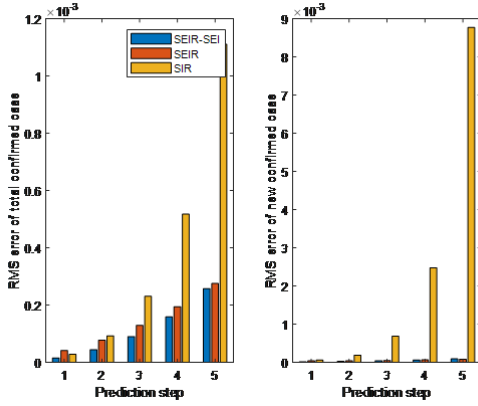
Figure A.2: Estimation and prediction results for 2015 dengue outbreak in RDJ



(a) Real-time predictions for total confirmed cases



(b) Real-time predictions for newly confirmed cases

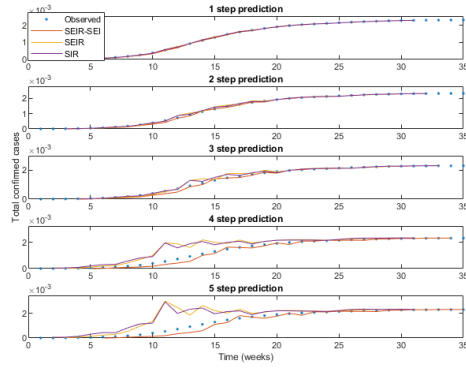


(c) RMS errors of different models

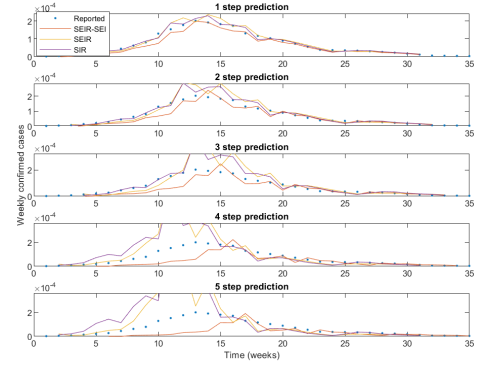
Figure A.3: Real-time predictions for 2020 dengue outbreak in Singapore

A.1.1 Real-time prediction of different disease models for Singapore and RDJ

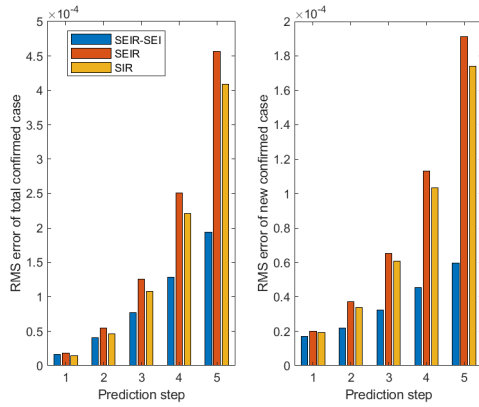
Fig.A.3 and Fig.A.4 show the real-time predictions of the dengue cases in Singapore and RDJ from three models. The results are also similar to the ones in Kaohsiung. The 1-step predictions for each model are the most accurate among all 5-step ones, and the predictions from the SEIR-SEI model are more accurate than the other two models.



(a) Real-time predictions for total confirmed cases



(b) Real-time predictions for newly confirmed cases



(c) RMS errors of different models

Figure A.4: Real-time predictions for 2015 dengue outbreak in RDJ