Title

Leveraging Surface Depression-Integrated Hydrologic Modeling to Study the Impact of Irrigation on Malaria Larval Habitats and Transmission

Author

Jiang, Ai-Ling

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Leveraging Surface Depression-Integrated Hydrologic Modeling to Study the Impact of Irrigation on Malaria Larval Habitats and Transmission

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Ai-Ling Jiang

Dissertation Committee:
Professor Kuolin Hsu, Chair
Professor Brett F. Sanders
Distinguished Professor Soroosh Sorooshian
Associate Adjunct Professor Phu Nguyen
Professor Guiyun Yan

2022

DEDICATION

To

my family and friends

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

CLM	Community Land Model
CSI	Critical Success Index
D2P	Depression-Preserved DEM Processing
DEM	Digital Elevation Model
EMOD	Epidemiological MODeling software
ERA5	The Fifth Generation European Centre for Medium-Range Weather Forecasts Reanalysis
FAR	False Alarm Rate
GCEW	Goodwin Creek Experimental Watershed
GLDAS	Global Land Data Assimilation System
GLHYMPS	GLobal HYdrogeology MaPS
GlobeLand30	Global Land Cover Mapping Project
IGBP	International Geosphere-Biosphere Programme
IRS	Indoor Residual Spraying
LLIN	Long-Lasting Insecticidal Net
LSM	Larval Source Management
MERRA-2	The second version of Modern-Era Retrospective analysis for Research and Applications
NAIP	National Agriculture Imagery Program
NWI	National Wetlands Inventory
ParFlow	PARallel FLOW
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System
PERSIANN-CCS-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System-Climate Data Record
POD	Probability of Detection
SRTM	Shuttle Radar Topography Mission
WI	Wetness Index

ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my committee chair, Professor Kuolin Hsu, for his continuous support and guidance throughout my time at UCI. I am thankful for the opportunity to work on the malaria project in Ethiopia, which helped me to gain exposure in a new field. Professor Hsu is patient and encouraging, always finding time to discuss ideas with me. I will always remember we had a meeting on New Year's Day to troubleshoot my modeling results.

I would also like to thank Professor Brett F. Sanders for his valuable advice and brainstorming for research ideas with me. I always feel more assured after talking to Professor Sanders, who is encouraging and enlightening. Thanks to his teaching assistant opportunity in the fluid mechanics class, I was able to challenge myself to enhance my communication and teaching.

Thank you to Professor Phu Ngyuen for making himself available whenever I needed to discuss my research with him and providing useful feedback. He was instrumental in my admission to UCI as I originally reached out to him to enquire about his research and an opportunity to do my PhD at the Center for Hydrometeorology and Remote Sensing (CHRS).

I am grateful to Professor Soroosh Sorooshian for being a father figure and showing care to the whole team in CHRS. There is always something new that I can learn from Professor Sorooshian, who never fails to give insightful comments.

My work on the malaria project would not have been possible without Professor Guiyun Yan’s support, trust and experience in the field of malaria. Coming from an engineering background, I learned so much from him and gained a new perspective from the public health sector. I am motivated by Professor Yan’s passion for his work and leadership and vision in harmonizing field work, modeling and engagement with public agencies.

It is also my pleasure to collaborate with Dr. Ming-Chieh Lee from Public Health, who has provided immense support and guidance on my research. His support extends beyond work as he has given me meaningful advice on my career and life as a friend. Substantial assistance was also provided by Dr. Guofa Zhou, Dr. Daibin Zhong and Dr. Solomon Kibret from Public Health as well as the team from Ethiopia comprising Professor Delenesaw

Yewhalaw, Dr. Dawit Hawaria and Dr. Teshome Degefa. Without their help on the field campaign, the modeling would not have been complete.

Thank you to my colleagues and friends in the CHRS group, past and present: Diane, Dan, Bita, Hoang, Matin, Raied, Baoxiang, Ata, Pooya, Negin, Mohammed, Mojtaba, Eric, Vesta, Vu, Claudia, Jinyang, Nazak and many more. It was my pleasure to work with such a hardworking group of people and I enjoyed spending time with them over happy hours, dinner parties and conferences. Special mention to Hoang who was so kind as to help debug my model even after graduating. Thank you to Mohammed for being a role model and spiritual guide who listened to my problems over therapeutic campus walks, jogs and coffee sessions. Thank you to Eric for lightening up the atmosphere in the office. As batch mates, we always enjoyed sharing our encounters and gossips together. Thank you to my other batch mate, Vesta, for being such a supportive friend and cheering me up whenever I'm down. I'm fortunate to have you by my side as we charge towards the finishing line of this PhD journey. Thank you to Jinyang for being a considerate junior and offering your help whenever I'm busy. I wish you a successful PhD. I also thank my friends outside CHRS, Parakh, Cassendra, Pratiik, Vishakah, Debora, Mahima, Nishtha, Matt, Ari and many more for adding color to my life.

Finally, to my family, without your support, I wouldn't be where I am today. To my parents, thank you for letting me pursue my studies and career overseas and your unwavering care and encouragement. Thank you to my grandparents for your unconditional love. Thank you to my sister and brother-in-law for taking good care of the family so that I can focus on my work from another continent without any worries.

Last but not least, thank you to my fiancé Jianwei for his support and company through my ups and downs.

This work was supported by grants from National Institutes of Health (U19 AI129326), U.S. Department of Energy (HydroWIRES Initiative DE-EE0008943) and California Department of Water Resources (AR Program 4600013361). Chapter 2 of this dissertation is a reprint of the material as it appears in *Predicting distribution of malaria vector larval habitats in Ethiopia by integrating distributed hydrologic modeling with remotely sensed data*. Scientific Reports 11, 10150, used with permission from Nature Publishing Group. The co-authors listed in this publication are Ming-Chieh Lee, Guofa Zhou, Daibin

Zhong, Dawit Hawaria, Solomon Kibret, Delenasaw Yewhalaw, Brett F. Sanders, Guiyun Yan and Kuolin Hsu.

VITA

Ai-Ling Jiang

EDUCATION

Ph.D. in Civil and Environmental Engineering, University of California, Irvine (UCI), CA	2017–2022
M.S. in Environmental Fluid Mechanics & Hydrology, Stanford University, CA	2013–2015
B.S. in Civil Engineering, National Taiwan University, Taiwan	2009–2013
Exchange Student Program, Nanyang Technological University, Singapore	2013

PUBLICATIONS

- Yan, G., Lee, M.C., Zhou, G., Jiang, A.L., Degefa, T., Zhong, D., ... & Kazura, J. W. (2022). Impact of Environmental Modifications on the Ecology, Epidemiology, and Pathogenesis of Plasmodium falciparum and Plasmodium vivax Malaria in East Africa. *The American Journal of Tropical Medicine and Hygiene*, 107(4_Suppl), 5-13.
- Jiang, A.L., Lee, M.C., Zhou, G. et al. (2021). Predicting distribution of malaria vector larval habitats in Ethiopia by integrating distributed hydrologic modeling with remotely sensed data. Scientific Reports 11, 10150.https://doi.org/10.1038/s41598-021-89576-8
- Jiang, A.L., Hsu, K., Sanders, B.F., Sorooshian S. (2022). Topographic hydro-conditioning to resolve depression storage and ponding in a physically based hydrologic model. Manuscript submitted to Advances in Water Resource.
- Jiang, A.L., Lee, M.C., Hsu, K., Yan, G. Investigating the Impact of Irrigation on Malaria Larval Habitats and Transmission using a Hydrology-based Modell. Manuscript in preparation.

CONFERENCE PRESENTATIONS

ORAL PRESENATIONS

- Jiang, A.L., Hsu, K., and Li, J. (2022). Hydrological modeling of the impact of irrigation on malaria vector larval ecology and malaria risk. International Center of Excellence for Malaria Research Scientific Advisory Group Meeting; 2022 May 27.
- Jiang, A.L., Hsu, K., Sanders, B.F., Sorooshian, S. (2021). A new topographic conditioning method to integrate surface depressions in physically based hydrologic modeling. AGU Fall Meeting; 2021 Dec 13-17; New Orleans, LA, USA.
- Jiang, A.L., Hsu, K., Sanders, B.F., and Sorooshian, S. (2020). Downscaling surface water depth to simulate small-scale ponding: a case study in the prairie pothole region of north America. AGU Fall Meeting; 2020 December 1-17; Virtual.
- Jiang, A.L., Hsu, K., and Li, J. (2020). Hydrological modeling of the impact of irrigation on malaria vector larval ecology and malaria risk. International Center of Excellence for Malaria

Research Scientific Advisory Group Meeting; 2020 Oct 20; Virtual.

POSTER PRESENATIONS

- Jiang, A.L., Hsu, K., Sanders, B.F., Sorooshian, S. (2021). Prediction of malaria vector larval habitat distribution and transmission in Ethiopia by integrating distributed hydrologic modeling with remotely sensed Data. AGU Fall Meeting; 2021 Dec 13-17; New Orleans, LA, USA.
- Jiang, A.L., Rahnamay Naeini M, Analui B., Nguyen P., Afzali Gorooh V., Hsu K., Sorooshian, S. (2021). Hydrologic evaluation of PERSIANN precipitation products through streamflow simulation using WRF Hydro. AMS Annual Meeting; 2021 January 10-15; Virtual.
- Li J., Hsu, K., Pan, B., Jiang, A.L. (2020). Applying deep learning models for catchment scale streamflow prediction. AGU Fall Meeting; 2020 December 1-17; Virtual.
- Jiang, A.L., Hsu, K., Lee, MC., Yan, G., Sanders, B.F., Nguyen, P., Sorooshian, S. (2019). Integration of hydrological modeling and remote sensing to identify potential malaria mosquito breeding sites. AGU Fall Meeting; 2019 Dec 9-13; San Francisco, CA, USA.
- Hsu, K., Jiang, A.L., Huang, P., Lee, K. T. (2019). Landslide susceptibility assessment using distributed hydrological modeling and slope stability analysis. AGU Fall Meeting; 2019 Dec 9-13; San Francisco, CA, USA.
- Lee, M., Jiang, A.L., Hsu, K., Yan, G. (2019) Integration of hydrologic modeling with remote sensing data to predict distribution of malaria vector larval habitats in Ethiopia. ASTMH 68th Annual Meeting, 2019 November 20-24; National Harbor, MD USA.
- Jiang, A.L., Sanders, B.F., Nguyen, P., Hsu, K. (2018). What is the role of rainfall and flooding on motor vehicle accidents in Los Angeles? AGU Fall Meeting; 2018 Dec 10-14; Washington, D.C., USA.
- Sanders, B.F., Schubert, J.E., Jiang, A.L. (2018). Towards Real-Time Urban Flood Simulation for Emergency Management Decision-Making: Combining Parallel Computing and Subgrid Models. AGU Fall Meeting; 2018 Dec 10-14; Washington, D.C., USA.

WORK EXPERIENCE

Research Assistant, Center for Hydrometeorology and Remote Sensing, UCI, CA	Sep 2017–Present
Teaching Assistant, Mathematical Methods in Engineering Analysis, UCI, CA	Fall, 2020, 2021
Teaching Assistant, Introduction to Fluid Mechanics, UCI, CA	Fall, 2018, 2019
Water Resource Engineer, RMC, a Woodard and Curran Company, CA	Jul 2015–Jul 2017
Modeling Intern, Risk Management Solutions, Inc, CA	Jun–Aug 2014
Project Management Intern, Sinotech Engineering Consultancy, Ltd, Taiwan	Jul–Oct 2012
Land Surveying Intern, Hong-Yi Engineering Consultancy, Ltd, Taiwan	Jun–Aug 2011

HONORS AND AWARDS

- American Meteorological Society Annual Meeting Poster Presentation Third Place Winner 2021
- Outstanding Engineering Student Awards, Orange County Engineering Council 2020
- Scholarship, the Temasek Foundation LEaRN Program 2013

GRANTS

- American Geophysical Union Annual Meeting Hydrology Section Student Registration Grant 2020
- UCI Associate Graduate Student Travel Grant 2019
- CUAHSI Student Travel Grant 2019

CERTIFICATIONS

- Professional Licensed Civil Engineer in the State of California (License Number: 90458) 2019
- CUAHSI Advanced Short Course: Integrated Simulation of Watershed Systems Using ParFlow 2019
- CUAHSI Short Course: Integrated Simulation of Watershed Systems Using ParFlow 2019
- CUAHSI-NCAR Training Workshop: The Community WRF-Hydro Modeling System 2019

PROFESSIONAL MEMBERSHIPS

- American Geophysical Union
- American Meteorological Society
- American Society of Tropical Medicine and Hygiene

ABSTRACT OF THE DISSERTATION

Leveraging Surface Depression-Integrated Hydrologic Modeling to Study the Impact of Irrigation on Malaria Larval Habitats and Transmission

by

Ai-Ling Jiang

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2022

Professor Kuolin Hsu, Chair

A combination of accelerated population growth and severe droughts has created pressure on food security and driven the development of irrigation schemes across sub-Saharan Africa. Irrigation has been associated with an increase in malaria risk but the underlying mechanism is not well understood. While useful in investigating transmission dynamics, malaria models are limited in irrigated areas as they typically infer mosquito abundance from rainfall. Fundamentally, this assumes rainfall as a proxy for larval habitats and can lead to contradictory transmission results. The availability of water for breeding is governed by hydrological processes which are highly non-linear and spatially variable. The incorporation of hydrologic modeling is therefore essential to understand the effect of irrigation on malaria.

The overarching goal of this dissertation is to study the impact of irrigation on the spatiotemporal distribution of malaria larval habitats and transmission using a hydrology-based malaria model. First, a three-dimensional, distributed hydrologic model was applied to simulate malaria larval habitats at a sugarcane plantation site in Arjo, Ethiopia. The

scarcity of field data was overcome by integrating remotely sensed data with the model. Results suggest that at least half of the irrigated farms had a high probability of larval habitat occurrence. Irrigation dampened and prolonged the seasonality of the larval habitats, with a significant shift from semi-permanent to permanent habitats.

Land surface depressions highly influence the transformation of rainfall to ponding for aquatic larval habitat formation. Hence, I sought to improve the representation of physically meaningful surface depressions in the digital elevation model (DEM) used in the hydrologic model. A new topographic conditioning workflow, Depression-Preserved DEM Processing (D2P), was developed and evaluated through a case study in a pond-dominated watershed in the United States. D2P successfully resolved 86% of the ponds at 10 m DEM resolution. A hydrologic simulation was performed using the D2P processed DEM and demonstrated a more robust characterization of surface water dynamics.

Lastly, I extended the Arjo hydrologic modeling framework with D2P to enhance the larval habitat estimation in an agent-based malaria model and examine transmission dynamics. The inclusion of hydrologic processes increased the variability of larval habitats which resulted in significantly lower transmission. The application of irrigation enabled the development of mosquitoes in dry seasons while stabilizing growth in rainy seasons. The model also revealed that malaria transmission was sustained all-year round and intensified during the main transmission season, with the peak shifted forward by around one month. Finally, I showed how habitat heterogeneity could affect the spatiotemporal dynamics of malaria transmission.

Understanding changes to malaria transmission dynamics by irrigation is important for the development of mitigation strategies. The framework presented in this dissertation particularly helps larval source management as a supplementary vector control by identifying malaria hotspots and prioritizing resources for operational planning.

Chapter 1

Introduction

1.1 Background

1.1.1 Malaria and irrigation in Ethiopia

Malaria is a deadly disease induced by parasites that are transmitted to humans through the bites of infected female *Anopheles* mosquitoes. In 2020, there were a total of 241 million cases of malaria and 627,000 cases of malaria-related deaths worldwide. It is particularly acute in sub-Saharan Africa and remains one of the most pressing public health challenges in the region. 95% of all malaria cases and 96% of deaths were recorded in sub-Saharan Africa alone (World Health Organization, 2021). Human-induced environmental modifications in the form of water resources development projects such as the construction of dams and irrigation projects are known to exacerbate malaria transmission by creating more breeding sites (Keiser et al., 2005).

In Ethiopia, more than 75% of its area is infected by malaria, making it the leading public health challenge in the country (Ayele et al., 2012). Over the years, a combination of severe droughts and the need for socio-economic growth has driven Ethiopia to undergo significant land use change through the implementation of irrigation schemes for agriculture (Hagos et al., 2009; Haile & Kassa, 2015). It has witnessed the fastest growth in irrigation in Africa, with the total irrigated area increasing by more than 50% between 2002 and 2014 (Malabo Montpellier Panel, 2018). While irrigation has helped to alleviate poverty in Ethiopia, it has the potential to increase the abundance and diversity of breeding sites for mosquito (Demissew et al., 2020). Specifically, sites that do not have enough water to sustain breeding in the dry season are converted to habitats that support the continuous production of adult mosquitoes (Fillinger et al., 2004). The higher number and variety of habitats can support the growth of multiple species of mosquitoes, thereby exacerbating the transmission of malaria in the country (Hawaria et al., 2020). The mechanisms underlying this phenomenon remain poorly understood given the complexity of vector ecology, parasite

transmission and environmental interactions. Given the increasing reliance on irrigation in the country, further investigation is necessary to evaluate the impact of irrigation on malaria.

1.1.2 Life cycle of malaria parasite and vector

Malaria parasites belong to the genus *Plasmodium*. The most predominant species in Ethiopia are *Plasmodium falciparum* and *Plasmodium vivax* (Adugna Kendie et al., 2021). The parasite requires two blood meals by the female *Anopheles* adult mosquito to be transmitted from one infected human to another. The infection cycle begins when a mosquito bites an infectious human for egg production (Figure 1.1). During the process, it ingests the parasite in the form of gametocytes from the human bloodstream. The gametocytes within the mosquito develop into sporozoites that subsequently migrate to its salivary glands. At this stage, the mosquito becomes infectious and it injects sporozoites into the new host when taking the next blood meal from another human. The sporozoites infect the liver cells and develop into merozoites in the liver stage. The merozoites subsequently invade the red blood cells and multiply in the blood stage. Whenever the merozoites invade more red blood cells, the human host will experience symptoms. Some of the merozoites will also develop into gametocytes during the same stage. The human host is now infectious and perpetuates the cycle when bitten again by a mosquito (Centers for Disease Control and Prevention, 2020; Vaughan, 2007).

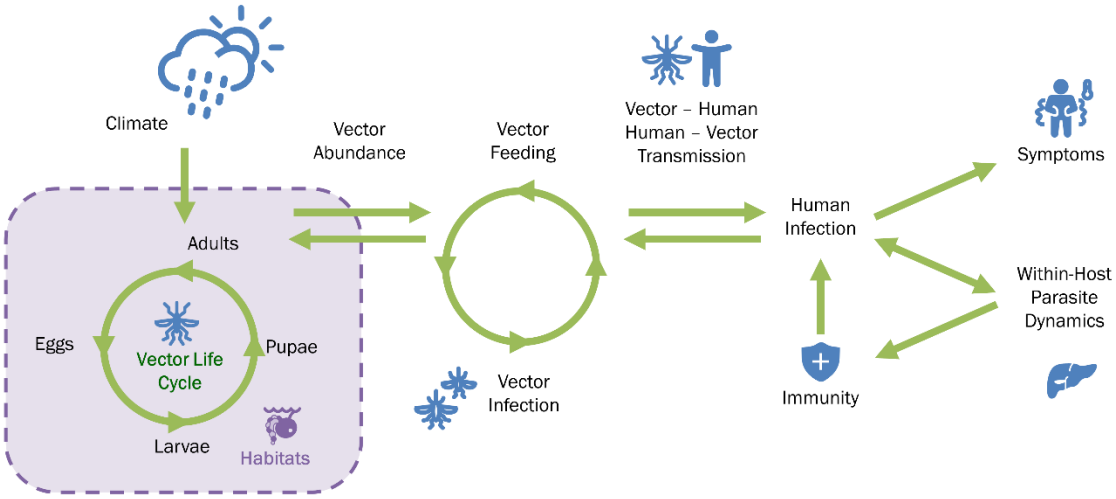


Figure 1.1: Malaria Transmission cycle. Figure modified from Epidemiological MODeling software (EMOD) online manual (Bill & Melinda Gates Foundation, 2022a).

Of the over forty species of Anopheles mosquitoes in Ethiopia, the four species that are vectors of the *Plasmodium* parasite are *Anopheles arabiensis*, *Anopheles funestus*, *Anopheles pharoensis*, and *Anopheles nili*. *Anopheles arabiensis* is the main vector, while *Anopheles funestus* and *Anopheles pharoensis* are the next most important vectors (Adugna Kendie et al., 2021). Typically, the life cycle of an *Anopheles* mosquito comprises four stages, namely egg, larva, pupa and adult. The first three stages are aquatic and the duration from egg to adult varies among species. Mosquitoes can develop from egg to adult in as little as 7 days but usually take 10-14 days in tropical conditions (Centers for Disease Control and Prevention, 2020).

1.1.3 Climate as a driver of malaria

The survival and growth throughout the aquatic stages of the dominant vector species in Ethiopia, the *Anopheles arabiensis*, depend heavily on the water availability and ambient temperature. *Anopheles arabiensis* prefers to breed in temporary, small, sunlit, clear, and shallow fresh-water pools (Sinka et al., 2010). The mosquito is not resistant to drying in the immature stages so the frequency of rainfall is important to ensure that the pool does not desiccate before the adult mosquito emerges. On the other hand, high discharges induced by heavy rainfall can decimate breeding sites by flushing out its larvae so sites characterized by stationary water or very slow flow conditions are preferred (M. W. Smith et al., 2013). Besides water availability, larval growth is also governed by temperature. During the aquatic stages, *Anopheles arabiensis* develops faster with increasing water temperature. The growth rate has previously been studied to be linear up to 32 °C and no survival occurred beyond the lower limit temperature of 15 °C and upper limit temperature of 35 °C (Davies et al., 2016; Lyons et al., 2013).

In the adult mosquito stage, temperature and humidity affect biting frequency, fecundity and longevity. Like most insects, *Anopheles arabiensis* is poikilothermic and can only function normally within a limited temperature range. At higher temperatures, the digestion of blood meals is accelerated which increases the biting frequency (Afrane et al., 2012). There are also more egg batches laid and a higher egg hatchability rate in the summer compared to winter when temperatures are cooler (Maharaj, 2003). On the other hand,

elevated temperatures from 25 °C to 35 °C have been shown to shorten the lifespan of the *Anopheles arabiensis* adult mosquito (Oliver & Brooke, 2017). Like most other insects, mosquitoes are able to survive only for short periods at low humidity, presumably due to susceptibility to water loss through evaporation (Bayoh, 2001). Within the infected mosquito, the parasite development is generally accelerated by higher temperatures. At 30°C, the incubation period for the *Plasmodium vivax* and *Plasmodium falciparum* is around 7 to 9 days. On the other hand, malaria cannot be transmitted when temperatures are too cool to support the development of the sporozoites. The minimum temperature for parasite development of *Plasmodium vivax* and *Plasmodium falciparum* approximates 15°C to 18°C, at which the incubation period stretches to more than 30 days (Patz & Olson, 2006).

Evidently, climate is a key determinant of both the geographic distribution and the seasonality of malaria. Without sufficient rainfall, mosquitoes cannot survive, and if not sufficiently warm, parasites cannot survive in the mosquito. As such, malaria is extremely common in tropical and subtropical climates with hot and humid weather conditions where the parasites can live and are generally active.

1.2 Motivation

1.2.1 Reassessing larval source management

Significant gains have been made in malaria control in the 21st century. In Africa alone, the prevalence of *Plasmodium falciparum* infection reduced by half from 2000 to 2015 (Bhatt et al., 2015). The backbone of the malaria elimination effort comprises long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS), accounting for more than 70% of the reduction (McCann et al., 2017). However, progress has stalled since 2015. The first intermediate milestone of achieving 40% global reduction in malaria by 2020 was already missed, with a higher number of malaria cases in 2020 than in 2015 (The Lancet, 2022). While effective in reducing malaria to a certain extent, LLINs and IRS are inadequate in eliminating malaria due to mosquitoes developing insecticide resistance and behavioral adaptations (Kahamba et al., 2022; Killeen et al., 2017). Hence, there is a need to rethink the current approach to vector control to meet future malaria reduction goals.

One possibility is to reconsider larval source management (LSM) as a supplementary tool (Kahamba et al., 2022). In the early twentieth century, LSM was one of the most effective malaria control methods (Fillinger & Lindsay, 2011; Kahamba et al., 2022). Unlike LLINs and IRS which target the adult vector, LSM focuses on the management or elimination of aquatic larval habitats to prevent the mosquito from developing from the immature stage (i.e., larvae and pupae) into adult vectors. Examples of LSM include habitat modification (e.g., surface water drainage, landscaping), habitat manipulation (e.g., shading or exposing habitats to the sun), biological control (e.g., introduction of predators) and larviciding. LSM has the potential to eliminate residual cases of malaria transmission as it is more effective than LLINs and IRS in cases where vectors have developed resistance to insecticides or where outdoor biting is dominant (Fillinger & Lindsay, 2011). In cases where larval habitats are abundant and difficult to locate, LSM can be difficult to implement and requires significant resources and labor. Hence, a detailed understanding of larval ecology and accurate mapping of larval habitat hotspots are crucial in prioritizing resources and ensuring the successful implementation of LSM.

1.2.2 Hydrology oversimplified in malaria modeling

Rainfall has long been known to affect malaria transmission through the formation of aquatic larval habitats (Patz et al., 1998) but the relationship between the two is not so straightforward. The availability of water for breeding is not only related to rainfall but also dictated by hydrological processes which are complex in nature and spatially variable. In a hydrologic cycle, rainfall is partitioned into infiltration and surface runoff based on the soil type (Yamana & Eltahir, 2010). The resulting surface runoff will accumulate or drain depending on the topography and surrounding vegetation (Atieli et al., 2011). The persistence of the ponded water can also be influenced by evapotranspiration from the vegetation which changes with the land use type. Besides rainfall, breeding sites can develop from groundwater, irrigation and around dams and reservoirs. The heterogenous ecological settings result in habitats of different persistence and productivity. This diversity complicates the pattern of adult mosquito density and malaria transmission intensity (A. J. Hardy et al., 2013; Munga et al., 2006). Relying only on rainfall as a proxy to estimate habitat

availability can lead to contradictory or biased transmission results (M. W. Smith et al., 2020; Yé et al., 2009).

Despite the importance of hydrology, it is often neglected or oversimplified in malaria models where the only driver of mosquito abundance is climate (Ermert et al., 2011; Hoshen & Morse, 2004; Yé et al., 2009). More recently, studies have attempted to incorporate hydrology into malaria transmission modeling to varying levels of complexity and success [35]. Most resort to a simple conceptual water balance model to determine the availability of water for larval habitats (Asare et al., 2016; P. A. Eckhoff, 2011; Montosi et al., 2012; Parham et al., 2012; Patz et al., 1998) while only a few have proposed more sophisticated hydrologic models that further consider canopy processes and subsurface flows (Bomblies et al., 2008; Le et al., 2018).

1.2.3 Complex link between irrigation and malaria dynamics

Discussion on the association between irrigation and malaria transmission is not new (Ijumba & Lindsay, 2001). While past observations tell us that irrigation can increase the adult vector population (Demissew et al., 2020), it remains a challenge to predict where and when breeding will occur (Frake et al., 2020).

Crop production and distribution change seasonally so irrigation varies accordingly. Within a season, the application of irrigation also depends on soil saturation and crop water use. The spatiotemporal heterogeneity in irrigation results in habitats of different persistence and productivity. This diversity complicates the pattern of adult mosquito density and malaria transmission intensity (Frake et al., 2020), which warrants the incorporation of hydrology into malaria modeling. However, a hydrology-based malaria model that can simulate the impact of irrigation is currently lacking. The ability to predict changes to malaria transmission dynamics by irrigation will be important to the development of mitigation strategies.

1.2.4 Small-scale surface depressions overlooked in distributed hydrologic modeling

Larval habitats are associated with land surface depressions which play a central role in the transformation of rainfall to ponding, infiltration and runoff. However, digital

elevation models (DEMs) used by spatially distributed hydrologic models rarely capture land surface depressions at spatial scales relevant to this transformation. This is because DEMs are typically hydro-conditioned (i.e., filling or breaching topographic cells) before use in hydrologic models. The justification is that the precision of the altitude measuring sensors and post-processing techniques can lead to systematic and random errors in the DEMs (Zhu et al., 2013). Hydro-conditioning is therefore implemented to limit variability or noise in land surface slopes for smoother flow routing (i.e., lower chance of computational instability) (Chow & Ben-Zvi, 1973; Rieger, 1998; Zhang & Cundy, 1989). However, this ends up removing physically meaningful land surface depressions.

To improve larval habitat simulation using a hydrologic model, a formalized and reproducible DEM hydro-conditioning tool that resolves fine-scale surface depressions of interest is required.

1.3 Research objectives and dissertation organization

The overarching goal of the dissertation is to incorporate surface depression-integrated hydrologic modeling to study the impact of irrigation on malaria transmission. Specifically, the first aim is to examine the utility of a hydrologic model in simulating the spatiotemporal distribution of larval habitats with the goal of providing guidance for larval source management. The second aim is to resolve surface depressions relevant to larval habitats and smooth small scale, non-physical variability in DEM for depression-integrated and efficient hydrologic modeling. Consequently, the third aim is to integrate the hydrologic model with a malaria model to improve the larval habitat representation of the latter and compare the difference. Finally, the last aim is to apply the proposed framework to study the underlying mechanism of how irrigation affects malaria habitats and transmission dynamics.

Chapter 2 presents an application of the three-dimensional, distributed hydrologic model to predict the location of larval habitats at a sugarcane plantation site in Arjo, Ethiopia. The scarcity of field data was overcome by integrating remotely sensed data with the model. The impact of irrigation on larval habitats was investigated based on the probability of occurrence, stability and temporal patterns of simulated habitats in the study area.

Chapter 3 proposes a new topographic conditioning workflow, Depression-Preserved DEM Processing (D2P), to improve the representation of physically meaningful surface depressions in the DEM for distributed hydrologic modeling. The ability of D2P to identify surface depressions of interest was investigated through a case study in Goodwin Creek Experimental Watershed (GCEW). The hydrologic simulation using D2P-processed DEM was also compared against that of a traditional DEM conditioning (depression filling) method.

In Chapter 4, the hydrologic modeling framework from Chapter 2 was coupled with an agent-based malaria model to enhance the larval habitat estimation and investigate transmission dynamics. The coupled framework utilized D2P proposed in Chapter 3 to process the DEM for hydrologic modeling. The effect of incorporating hydrologic processes to estimate larval habitats on malaria transmission intensity and seasonality was demonstrated. Using the coupled model, the impact of an existing irrigation scheme on the spatiotemporal dynamics of malaria transmission was examined.

Chapter 5 summarizes the key findings of the dissertation and discusses limitations and future work.

Chapter 2

Predicting Distribution of Malaria Vector Larval Habitats in Ethiopia by Integrating Distributed Hydrologic Modeling with Remotely Sensed Data

2.1 Introduction

LLINs and IRS are the key malaria vector control tools (Degefa et al., 2017; Yakob et al., 2011). Scale-up of LLINs and IRS in the past decade has reduced malaria burden in Africa by half (Bhatt et al., 2015), however the progress of malaria control has been stalled in many African countries due to limited efficacy of LLINs and IRS as a result of insecticide resistance and increased outdoor biting behavior (Olalubi, 2016). Thus, there is a recent renewed interest in LSM as an supplementary vector control tool (World Health Organization, 2013). LSM involves larviciding and biological control of malaria vectors, and also modification and manipulation of aquatic habitats (World Health Organization, 2013). LSM has not been widely used in malaria vector control in Africa, partly due to the challenge of widespread and unstable larval sites in many ecosystems. LSM may not be suited to all ecosystems, however LSM would be greatly facilitated if larval habitat distribution under natural climatic conditions can be predicted a priori, so that regions best suited to LSM can be identified. Further, prediction of how environmental modification such as irrigation, canal construction and landscape alteration through engineering approach may change the distribution of transient, semi-permanent and permanent aquatic habitats would greatly help LSM-based malaria vector control program which is much needed in Africa.

The LSM program requires identification of aquatic habitats for malaria vectors. Past studies have attempted to use field-based surveys or harness remotely sensed data for larval habitat identification (A. Hardy et al., 2019; Zou et al., 2006). Field-based surveys involve the use of manual labor or unmanned aerial vehicles with geographic information system to map larval habitats, which can be time consuming (Olalubi, 2016), limited in geographic coverage and weather dependent (Carrasco-Escobar et al., 2019; A. Hardy et al., 2017). Alternatively, researchers have relied on satellite imagery and supervised classification into land use and

land cover maps to delineate potential larval aquatic habitats (Bøgh et al., 2007; Mushinzimana et al., 2006; Mutuku et al., 2009). The type of satellite imagery used is usually optical, which tends to be limited by cloud cover and is unable to identify water bodies hidden by vegetation cover (Catry et al., 2018; Ozesmi & Bauer, 2002).

An alternative approach is required as the aforementioned inadequacies in the existing methods often result in larval habitat mapping of limited coverage or discontinuous frequencies that are unable to support effective LSM. We proposed a novel approach using a physics-based integrated hydrological model that draws on fundamental principles to realistically model potential larval aquatic habitats. Complex hydrologic processes built into the model such as infiltration, evapotranspiration and runoff help to provide a mechanistic understanding of aquatic habitat behavior. The primary inputs of a hydrological model, namely meteorological and topographic datasets, can be acquired globally and are available at high temporal and spatial resolutions respectively through remote sensing (Rabus et al., 2003; Sun et al., 2018). Notably, the larvae of the major malaria vector in Ethiopia, *Anopheles arabiensis* (Nyanjom et al., 2003), have been associated with transient pools (Gimmg et al., 2001) and our approach allows the larval habitats to be resolved down to sub-daily frequencies and tens of meters resolutions necessary to capture the dynamic nature of the habitats. It can also be scaled up in coverage if required.

Several studies in malaria transmission have incorporated hydrologic modeling. Soti et al. combined a simple water balance model with a mosquito population model to predict the abundance of mosquitoes contributing to the transmission of Rift Valley fever in West Africa (Soti et al., 2012). Asare et al. applied another simplified water balance model and parameterized the processes to simulate the fractional water coverage in central Ghana (Asare et al., 2016). The empirical nature of the model and unrealistic assumptions made about infiltration and runoff result in heavy reliance on calibration and can increase model uncertainty substantially. Bomblies et al. used a mechanistic hydrologic model to simulate the surface water area for two villages in Niger (Bomblies et al., 2009). However, the subsurface and surface water components are only coupled one-way such that surface water can only flow to the subsurface but not the other way round. The lack of exfiltration and re-infiltration components precludes the representation of spring-fed pools from groundwater

recharge. Additionally, the model did not account for lateral subsurface flow, which can influence evapotranspiration and redistribute groundwater to low-lying areas, especially at higher spatial resolutions (Ji et al., 2017; Kollet & Maxwell, 2006). To simulate the dynamics of the aquatic larval habitats, the hydrologic model chosen must be able to detail the surface-subsurface interactions and plant processes associated with ponding given the complex interdependence between larval habitats and the environment. In addition, none of the existing hydrology-based malaria models have been used to investigate the impact of irrigation on larval habitats.

In the present study, we aim to examine the potential of integrated hydrological models in predicting the location of mosquito larval habitats by capturing the shallow subsurface dynamics and improve our understanding of the hydrologic processes and environmental modifications that render larval habitats. Specifically, we seek to answer the following: (1) Where are the potential larval habitats located and what is the probability of occurrence? (2) How long can the larval habitats be sustained? (3) Is there a cyclical pattern in the extent of the larval habitats? (4) What is the impact of irrigation on each of the above? The uniqueness of our approach lies in its consideration of irrigation practices and its ability to resolve complex ponding processes that contribute to potential larval habitats such as groundwater-surface water interactions (Jaros et al., 2019). We chose ParFlow (PARallel FLOW) (Ashby & Falgout, 1996a; J. E. Jones & Woodward, 2001; Kollet & Maxwell, 2006, 2008; Maxwell, 2013) for its open-source nature, robust numerical solver (Maxwell, 2013), and compatibility with high-performance computing (Kollet et al., 2010). To take into account irrigation and land cover characteristics, ParFlow was coupled with the Community Land Model (CLM) (Maxwell & Miller, 2005) to simulate soil moisture for the identification of malaria larval habitats in a sugarcane plantation and its vicinity in Arjo, Ethiopia.

2.2 Methods

2.2.1 Model description

ParFlow has been applied in many studies to simulate complex surface-subsurface interactions in heterogeneous environments (Condon & Maxwell, 2014; Ferguson &

Maxwell, 2012). Richards' equation (Richards, 1931), which governs water movement through the unsaturated zone, is used to simulate subsurface flow in three dimensions. The diffusive wave and Manning's equations, which calculate the depth and velocity of the routed water, are used to represent the overland flow in two dimensions (Kollet & Maxwell, 2006). To connect the overland flow and subsurface, the former is imposed as a boundary condition on the latter for natural feedbacks between the two components (Kollet & Maxwell, 2006).

Considering the non-linear nature of the governing equations, ParFlow solves the coupled system implicitly using the Newton–Krylov method for robust convergence to the solution and multigrid preconditioning for parallel scalability. This allows the system to be solved efficiently through parallel computing (Kollet et al., 2010). For details, see Ashby and Falgout (Ashby & Falgout, 1996a), Jones and Woodward (J. E. Jones & Woodward, 2001), and Kollet and Maxwell (Kollet & Maxwell, 2006). Additionally, CLM simulates the land surface water and energy balance which includes evaporation, transpiration, snow processes, heat fluxes, and radiation partitioning (Dai et al., 2003; Maxwell & Miller, 2005). The water fluxes calculated by CLM are incorporated in ParFlow through the source or sink terms in Richards' equation for subsurface flow (Maxwell & Miller, 2005). The two models are coupled over a user-defined number of subsurface layers and this allows ParFlow to take into account the characteristics of the vegetation cover as CLM simulates plant function types corresponding to different vegetation parameters (Kollet & Maxwell, 2008; Maxwell & Miller, 2005).

2.2.2 Study area

The study area is 208 km^2 and comprises Arjo-Didessa sugarcane plantation and its vicinity in the Oromia Region State, western Ethiopia (Figure 2.1). The altitude of the study area is 1,350 meters above sea level and the annual rainfall received is 1,477 mm (Hawaria et al., 2020), with a rainy season between May and October. The area covers most of the Arjo-Didessa sugarcane plantation site, which is characterized by clay and clay loam with low permeability (Ethiopian Corporation, 2020). Due to the slow rate of infiltration, rainwater can accumulate easily and form ponds in the area, which is exacerbated by irrigation of the sugarcane plantation. The widespread and persistent nature of this ponding contributes to the breeding of malaria vector mosquitoes.

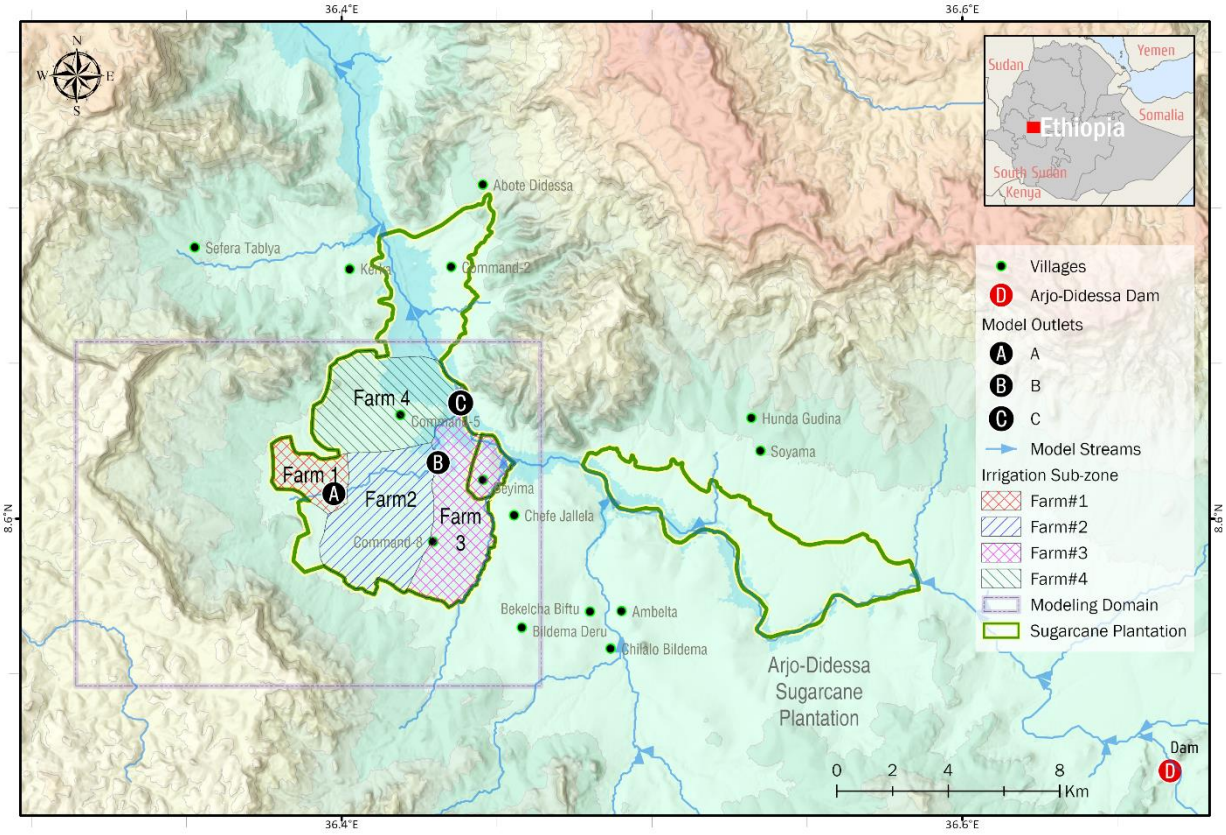


Figure 2.1: Study Area at Arjo-Didessa sugarcane plantation and its vicinity. This area is found in the Oromia Region of western Ethiopia and located 395 km west of the capital, Addis Ababa, at the intersection of the three woredas (districts), Jimma Arjo (East Wollega Zone), Bedele (Buno Badale Zone), and Dabo Hana (Illubabor Zone) at the Didessa River valley. The model area is enclosed by the gray box. The sugarcane plantations in the study area were demarcated by the green lines. To simplify model simulation, the irrigation parcels in the plantation area were further grouped and generalized into four farms, which will be explained in greater detail in the later subsection.

A recent study indicated that malaria is seasonal in this area; transmission mainly followed the rainy season, with the highest cases recorded between September and November (Hawaria et al., 2019). However, some localities experienced persistent malaria due to environmental modifications such as irrigation that support the continuous availability of breeding sites. *Anopheles arabiensis* is the predominant malaria vector species in the area. The major mosquito breeding habitat types included rain pools, stream shoreline, animal foot prints, irrigation canal, hippo trenches, drainage ditches, and puddles in rice cultivation (Hawaria et al., 2020) (Figure A.1).

In the ParFlow-CLM model, the study area was discretized with a resolution of 50 m, resulting in a grid configuration of 332 by 248 cells. The subsurface component was divided into 10 layers and the thickness of each layer varies, depending on the granularity of the data available. In general, the resolution of the subsurface layer increased nearer to the surface to capture the shallow surface processes in greater detail. The layer thicknesses ranged from 0.25 m to 20 m, over a total vertical depth of 100 m.

2.2.3 Input data

As the model domain is a rural area where field data for model construction was scarce, remotely sensed data and global synthetic datasets from published works were used. For example, 1 arc-second DEM from Shuttle Radar Topography Mission (SRTM) (JPL, 2013) was resampled to the 50 m model grid and converted to ground surface slopes as an input to ParFlow. The land cover for each grid cell in CLM was determined by the classification of 30 m resolution Landsat-8 imagery (United States Geological Survey Earth Resources Observation and Science Center, 2013) taken on a cloud-free day in January 2018 into International Geosphere-Biosphere Programme (IGBP) types. To characterize the subsurface, the soil taxonomy distribution (Figure A.2) for the top 2 m from the surface was referenced from the SoilGrids250m TAXOUSDA dataset (Hengl et al., 2017). The saturated hydraulic conductivity of the deeper zone beyond the top 2 m was based on GLHYMPS 2.0 (Gleeson et al., 2014). The depth to bedrock data from SoilGrids250m BDRICM dataset (Hengl et al., 2017) was used to delineate the bedrock zone, which was assigned a very low hydraulic conductivity. For the meteorological forcing, 0.04 degree by 0.04-degree precipitation data from Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System (PERSIANN-CCS) (Hong et al., 2004; Nguyen et al., 2019) was resampled to the model grid using bilinear interpolation. In addition, wind speed from the second version of Modern-Era Retrospective analysis for Research and Applications (MERRA-2) (Gelaro et al., 2017) of 0.5 degrees by 0.625 degrees resolution and air temperature, pressure, specific humidity and radiation data from Global Land Data Assimilation System (GLDAS) (Rodell, 2004) of 0.25 degree by 0.25 degree resolution were averaged for the entire domain. All the forcing data were obtained from

2018 and input to the model hourly. The list of model input data and the relevant details can be found in Table A.1 and Section A.1.

2.2.4 Model scenarios

A 1-year baseline period in 2018 from January 1 to December 31 was simulated with an hourly time step to produce daily soil saturation and groundwater pressure head. Sugarcane is a plant with high water consumption so irrigation during the dry season is essential. Hence, a separate scenario was run for the same period with the implementation of a synthesized irrigation scheme corresponding to the dry season from January to April and November to December. This synthesis was based on the sugarcane plantation irrigation schedule and detailed plans acquired from Arjo-Didessa Sugar Factory in Figure A.3 (Ethiopian Corporation, 2020). Specifically, the sugarcane plantation in the study area was grouped into four irrigation sub-zones as a simplified representation (Figure 2.1). In each irrigation cycle, Farms 1 and 3 are sprinkler-irrigated for 10 days, followed by Farms 2 and 4. Each farm receives a total of 10 mm of irrigation over 22 hours each day during its turn for irrigation.

As an illustration of the hydrological process for the baseline scenario, Figure 2.2 shows a time series of the simulated spatially averaged surface layer soil saturation, along with snapshots of the resulting surface layer soil saturation at five particular time instances in May 2018 when a 7-day rainfall event occurred. The instantaneous snapshots of the soil saturation reflect the spatial distribution of the rainfall. The details of the baseline simulation for the entire year and the irrigation scenario results can be found in Figure A.4, Figure A.5, and Section A.1.2,

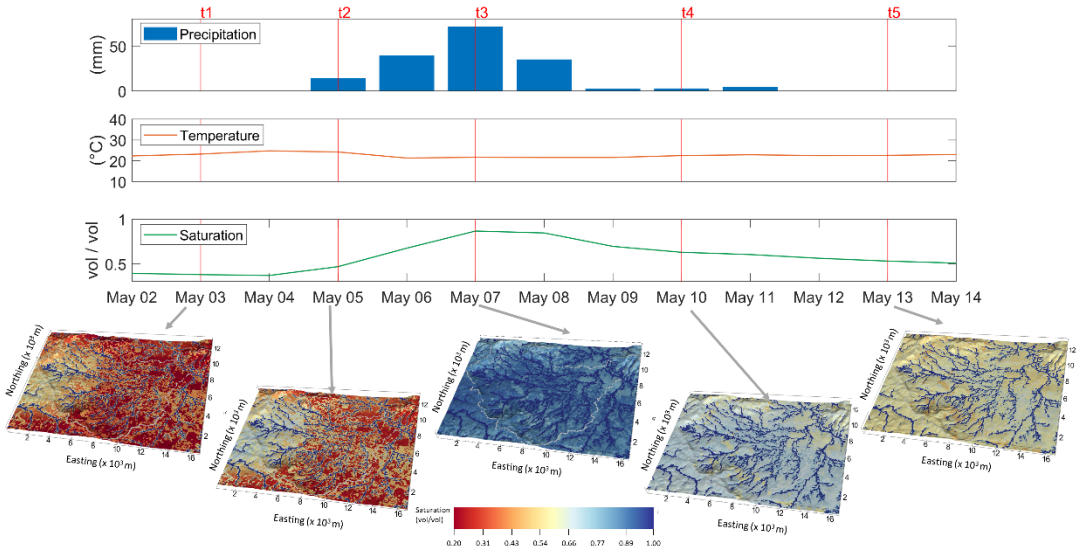


Figure 2.2: The simulated Surface layer (25cm depth) Soil Saturation at Five Time Instances during a Rain Event in May 2018. The snapshots illustrate the close-up views of the soil saturation for the rainfall event between May 5, 2018, and May 11, 2018, along with the time series of the spatially averaged precipitation, temperature, and simulated surface layer. On May 3 (t_1), the surface was generally dry before the onset of the rainfall, except for the mountainous areas on the left. On May 5 (t_2), the rain started to spread from the mountainous areas. By May 7 (t_3), the rain had spread to the entire area. The snapshot at t_4 shows the post-rainfall distribution on May 10, and the snapshot at t_5 shows the area drying up again after the rainfall event on May 13. The soil saturation increased to more than 85% at the peak of the storm across most of the study area and decreased quickly after about 5 days but the streams and the vicinities remain wet. There were no large depressions (e.g., lakes, pools) observed in the simulation.

2.2.5 Wetness index calculation

As mosquito reproduction is successful only if larval habitats remain stable for a period sufficient to complete the aquatic stage (Gianotti et al., 2009; Rejmánková et al., 2013), the viability of the habitat was determined by the persistence of ponding. Hence, we developed a Wetness Index (WI) metric to quantify the persistence of ponding as a basis for potential habitat representation after rain and irrigation. This will be used later to answer our research questions.

As the aquatic habitats at site were typically shallow and the lateral scale was of the order 10m or smaller, it was not feasible to explicitly simulate the surface water depth of the individual habitat. Hence, the simulated soil saturation of the top surface layer as described

in the previous section, coupled with a threshold, was used to assess the availability of the surface water that could contribute to ponding. Since soil saturation measures the extent to which the water content has filled up the voids within the soil, a higher soil saturation means there is a larger volume of water stored in the soil within the 50m grid cell. Hence, a potential occurrence of ponding was assumed if the surface layer soil saturation exceeds the threshold. Otherwise, no ponding occurs. To evaluate the duration of ponding, WI was used and defined as the cumulative number of days of ponding from the start of the simulation year at any grid cell (x, y) and day (t), based on the simulated soil saturation of the top surface layer $S(x, y, t)$ and a soil saturation threshold θ. The computation of the index is as shown below in Equation (2.1):

$$WI(x, y, t) = \begin{cases} WI(x, y, t-1) + 1, & if\ S(x, y, t) \geq \theta \\ 0, & otherwise \end{cases} \quad (2.1)$$

The initial WI for every grid cell was set to 0. The index increases each day if the soil saturation exceeds the threshold. Otherwise, it will reset to zero, implying that the habitats in the grid cell are no longer able to sustain the development of the larvae population.

The soil saturation threshold θ in Equation (2.1) was calibrated based on a field survey of aquatic larval habitats. 134 ponding locations were surveyed for larval growth during the dry (December 2017–February 2018) and rainy (June 2018–September 2018) seasons (Hawaria et al., 2020). For each surveyed location, information regarding whether larval growth was detected, the type of species identified, larval density, habitat dimension, habitat type and land use type were recorded. Regardless of whether larval growth was detected, each survey location served as an indication of ponding for calibration and validation. Some of the surveyed locations such as man-made ponds, tire track puddles, and animal footprints which could not be simulated by the hydrologic model were omitted. In addition, to minimize the influence of dry season irrigation on the parameterization considering that the irrigation in the model was approximated by a simplified scheme, the calibration was only conducted for the rainy season from May to October. In total, 102 of the surveyed locations were used for calibration and validation as shown in Figure 2.3.

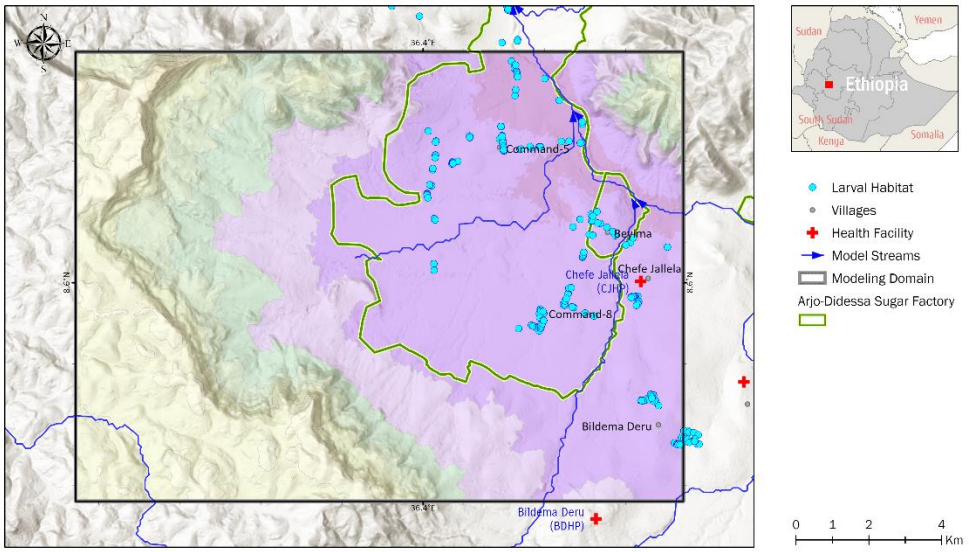


Figure 2.3: Location of the surveyed aquatic habitats. All accessible potential mosquito breeding habitats were surveyed and identified the presence of mosquito larvae during the dry (December 2017–February 2018) and rainy (June 2018–August 2018) seasons.

The objective of the calibration was to maximize the probability of detection (POD), which determines if the model can predict an aquatic habitat successfully. Other measures which can capture overprediction were not chosen as the field data only covered locations with ponding and it was challenging to rule out small puddles within the grid cell using other types of data. To ensure the relevance of the calibrated θ, a bootstrapping method was applied and it was found that the optimal θ was 0.48. In other words, the model would predict the occurrence of ponding for soil saturation above 0.48 at locations in line with the survey. Details of the survey data and calibration method can be found in Section A.1.3.

In summary, the overall schematic of our methodology is shown in Figure 2.4. Using the Wetness Index, we analyzed the potential larval habitats in terms of their spatial distribution, stability and temporal pattern and the results are presented in the next section.

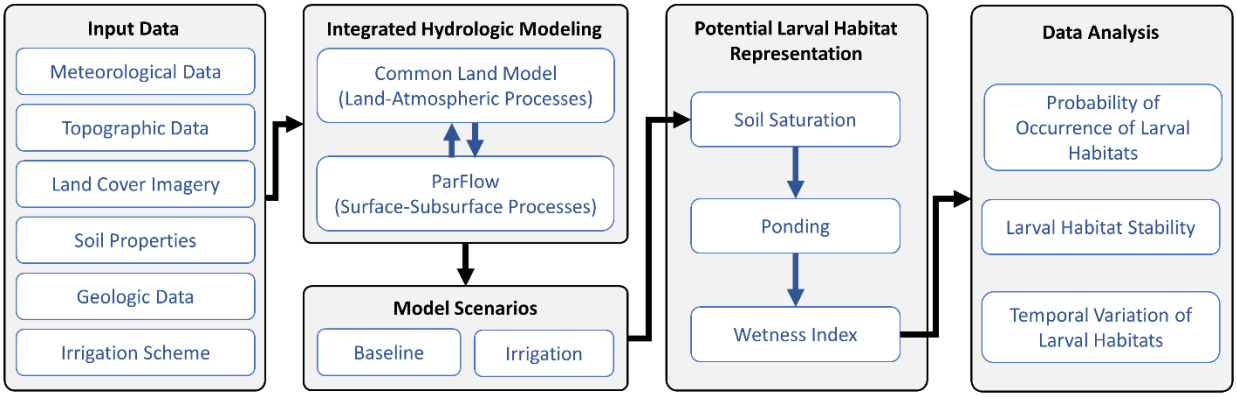


Figure 2.4: Overall schematic of methodology.

2.3 Results

2.3.1 Location of potential larval habitats and probability of occurrence

Generally, *Anopheles arabiensis* mosquito takes around 15 days to develop from egg to adult, but the duration can be as short as 10 days due to selection pressures from the stressed environment such as drought, temperature anomaly, or limited food resources (Afrane et al., 2007, 2012). In this regard, we considered areas with WI exceeding 10 and 15 days to be potential larval habitats under critical and normal conditions, respectively.

To determine the probability of potential larval habitat occurrence, we computed the probability of ponding occurring longer than 10 and 15 days, $P(WI > T)$, as shown in Equation (2.2). $P(WI > T)$ is defined as the ratio of $D(WI(x, y, t) > T)$, the number of cumulated days for which the WI (i.e. persistence of ponding) of a grid cell (x, y) at time t that exceeded T days, to D_{period}, the number of days within a defined period of simulation.

$$P(WI > T) = \frac{D(WI(x,y,t)>T)}{D_{period}}, T \in \{10,15\} \quad (2.2)$$

Figure 2.5 shows the results for the spatial distribution of $P(WI > T)$ over the three periods of simulation, namely the entire year of 2018, the dry season (i.e., January to April and November to December) and the rainy season (i.e., May to October). It can be observed that ponding was persistent throughout the year around the stream edges and the vicinity. $P(WI > 10)$ and $P(WI > 15)$ were consistently close to 1, reflecting a high potential of these areas as larval habitats.

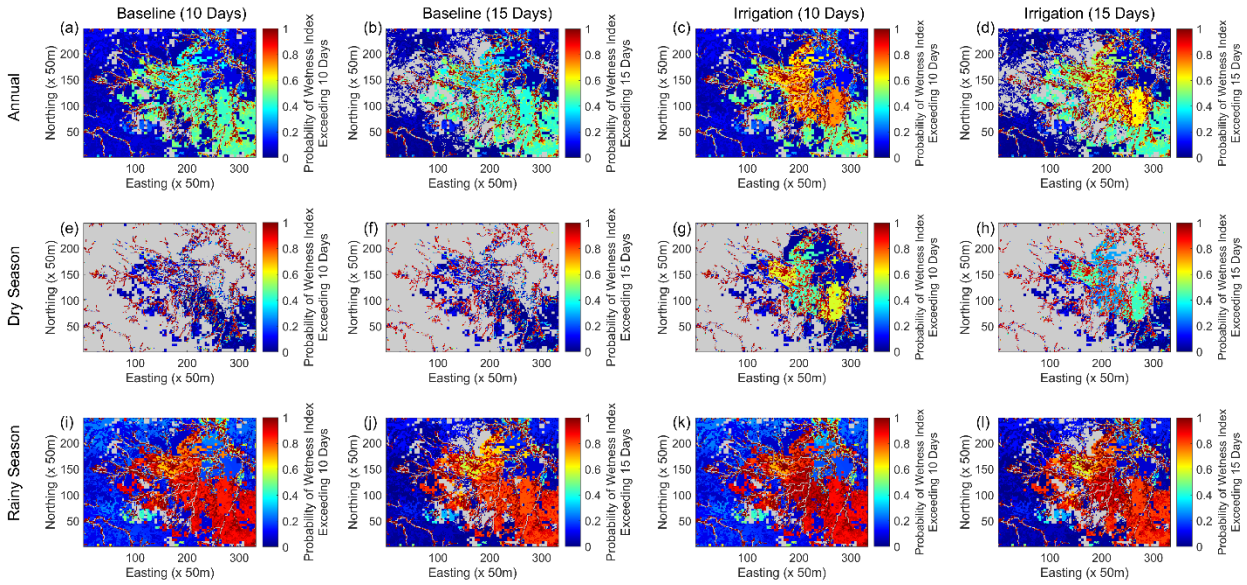


Figure 2.5: Spatial distribution for the probability of potential larval habitat occurrence. The sub-figures 5a-5d represent the probability of the WI exceeding 10 days and 15 days for the baseline scenario and the irrigation scenario for the entire year. The sub-figures 5e-5h represent the probability of the WI exceeding 10 days and 15 days during the dry season, and the sub-figures 5i-5l represent the WI exceeding 10 days and 15 days during the rainy season. Areas where the simulated surface water flowrate exceeded 0.01 m^3/s for 90% of the time in the simulated year were masked out for all the sub-figures since *Anopheles* larvae have a lower chance of surviving in fast-moving water (A. J. Hardy et al., 2013).

For the baseline scenario shown in Figure 2.5a-Figure 2.5b, the $P(WI > T)$ for the areas outside of the streams was predominantly determined by soil type. The areas characterized by Usterts (see Figure A.2) with the lowest hydraulic conductivity in the model domain were the next most at risk, with a $P(WI > T)$ of about 0.4-0.5. In the remaining areas, $P(WI > T)$ was generally 0.2 or less. Comparing Figure 2.5a and Figure 2.5b, the differences were minimal except for the steep areas at the watershed upstream boundary where $P(WI > 15)$ was predominantly zero. The surface water ponding was unable to last more than 15 days due to the terrain gradient.

Figure 2.5c and Figure 2.5d show the results for the irrigation scenario. Compared to the baseline scenario, the year-round persistent ponding around the streams and the vicinity was wider in coverage and more noticeable. Irrigation also increased $P(WI > 10)$ in Figure 2.5c and $P(WI > 15)$ in Figure 2.5d from 0.4-0.5 to about 0.7 and 0.6 respectively for Farm #1, Farm #2, and a significant portion of Farm #3 and Farm #4. The $P(WI > T)$ for the

remaining area within the farms remained relatively unchanged at 0.2 and this could be attributed to the Ustoll soil type which drains easily. The increase in the probability of potential larval habitat occurrence from the baseline was more pronounced for $P(WI > 10)$ than $P(WI > 15)$ since the interval of irrigation was set at 10 days, after which the farm drained without replenishment until the next irrigation cycle.

For the dry season, it can be observed in Figure 2.5e and Figure 2.5f that the stream edges were the only areas with high potential of larval habitat occurrence. In Figure 2.5g and Figure 2.5h, $P(WI > T)$ increased visibly in the farms after irrigation, with a distinct similarity between Farms #1/#3 and Farms #2/#4 that points to the irrigation schedule. While irrigation was alternated evenly between the two groups, Farms#1 and #3 showed a higher $P(WI > T)$ than Farms #2 and #4, possibly due to the timing of the irrigation relative to the rainfall. Irrigation could either coincide with rainfall or function as a supplement when there was no rainfall to augment soil moisture. Noticeably, there was an area to the northeast straddling both Farm #3 and Farm #4 where $P(WI > 10)$ was around 0.1 but $P(WI > 15)$ was almost zero, indicating that irrigation only allowed for larval habitats under critical conditions in that area during the dry season.

For the rainy season, it can be observed in the baseline scenario (Figure 2.5i-Figure 2.5j) that the areas characterized by Ustert exhibited a high potential of larval habitat occurrence, apart from the stream edges. Particularly, there was an area to the north where $P(WI > T)$ was lower than the other parts which could be due to the relatively steeper terrain. In the irrigation scenario (Figure 2.5k-Figure 2.5l), there was no visible difference in $P(WI > T)$ as compared to the baseline scenario, apart from a minor increase around the western part of Farm #4.

As a summary, we present the results in boxplots as shown in Figure 2.6 to illustrate the effect of irrigation in different seasons for the areas inside and outside farms. The relevant statistics can be found in Table 2.1. The $P(WI > T)$ had a highly asymmetrical distribution because it was very low in most of the model domain but could be very high in the remaining areas due to the streams. For the following comparison, we will use the median as it was more representative of the distribution.

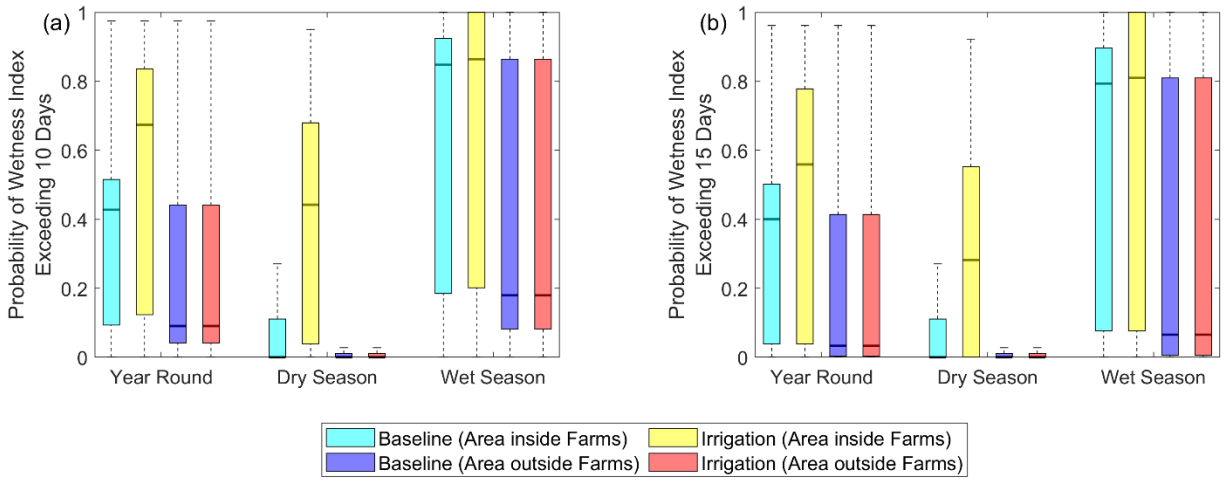


Figure 2.6: Box plots for the probability of potential larval habitat occurrence for the entire year, dry, and rainy season. Probability of the *WI* exceeding (a) 10 days and 15 days (b) for the area inside farms and the area outside farms. The line within each box is the sample median and the top and bottom of each box are the 25th and 75th percentiles. The whiskers were drawn from the two ends of the box and demarcate the observations which were within 1.5 times the interquartile range from the top and bottom of the box.

Table 2.1: Summary statistics of the probability of wetness index for the entire year, dry season, and rainy season. Mean, 25th percentile (P25), median and 75th percentile (P75) of the probability of the *WI* exceeding 10 days and 15 days for the areas (a) inside farms and (b) outside farms. The p-value was derived from the Wilcoxon Rank-Sum test under the null hypothesis that irrigation did not increase the median probability of exceedance.

	(a) Area inside Farms								
	Baseline				**Irrigation**				**p-value**
	Mean	**P25**	**Median**	**P75**	**Mean**	**P25**	**Median**	**P75**	
Wetness Index Exceeding 10 Days									
Dry	0.173	0.000	0.000	0.111	0.424	0.039	0.442	0.680	< 0.01
Rainy	0.607	0.185	0.848	0.924	0.643	0.201	0.864	1.000	< 0.01
Entire Year	0.392	0.093	0.427	0.515	0.534	0.123	0.674	0.836	< 0.01
Wetness Index Exceeding 15 Days									
Dry	0.168	0.000	0.000	0.111	0.347	0.000	0.282	0.553	< 0.01
Rainy	0.553	0.076	0.794	0.897	0.597	0.076	0.810	1.000	< 0.01
Entire Year	0.362	0.038	0.400	0.501	0.473	0.038	0.559	0.778	< 0.01

	(b) Area outside Farms								
	Baseline				Irrigation				p-value
	Mean	P25	Median	P75	Mean	P25	Median	P75	
Wetness Index Exceeding 10 Days									
Dry	0.053	0.000	0.000	0.000	0.053	0.000	0.000	0.000	0.254
Rainy	0.202	0.082	0.120	0.201	0.202	0.082	0.120	0.201	0.437
Entire Year	0.128	0.041	0.060	0.101	0.128	0.041	0.060	0.101	0.430
Wetness Index Exceeding 15 Days									
Dry	0.051	0.000	0.000	0.000	0.051	0.000	0.000	0.000	0.385
Rainy	0.125	0.005	0.038	0.098	0.125	0.005	0.038	0.098	0.440
Entire Year	0.089	0.003	0.019	0.049	0.088	0.003	0.019	0.049	0.443

In the baseline scenario, there was a higher potential for larval habitats to form inside the farms, with a median $P(WI > 10)$ of 0.427 and a median $P(WI > 15)$ of 0.400, than outside the farms, with a median $P(WI > 10)$ of 0.06 and a median $P(WI > 15)$ of 0.019. This is expected because the farms are in an area with relatively flat terrain and a higher concentration of streams. The difference in the median $P(WI > T)$ inside and outside the farms was bigger in the rainy season compared to the dry season, as the higher rainfall received intensified ponding.

Irrigation increased the median $P(WI > T)$ inside the farms drastically in the dry season, with the median $P(WI > 10)$ increasing from 0 to 0.442 and the median $P(WI > 15)$ increasing from 0 to 0.282. Although irrigation was only applied over the dry season, there was also a statistically significant increase in the median $P(WI > T)$ during the rainy season ($p<0.01$). The median $P(WI > 10)$ increased from 0.848 to 0.864 while the median $P(WI > 15)$ increased from 0.794 to 0.810. This was due to irrigation contributing to the antecedent soil moisture before the onset of the rainy season, which shortened the time for the soil to become saturated and ponding to occur. On the other hand, there was no strong evidence outside the farms of an increase in the median $P(WI > T)$ due to irrigation ($p > 0.01$). This applied to both rainy and dry seasons.

2.3.2 Stability of larval habitats

In the previous section, we showed that irrigation did not have a significant impact on areas outside the farms. Here, we evaluated the stability of the potential larval habitats specifically for the areas inside farms based on the distribution of the maximum duration of ponding for each grid cell within the year as shown in the histogram (Figure 2.7a). The total number of cells corresponding to each bin interval of 15 days was expressed as a fraction of the total number of cells in the area inside farms.

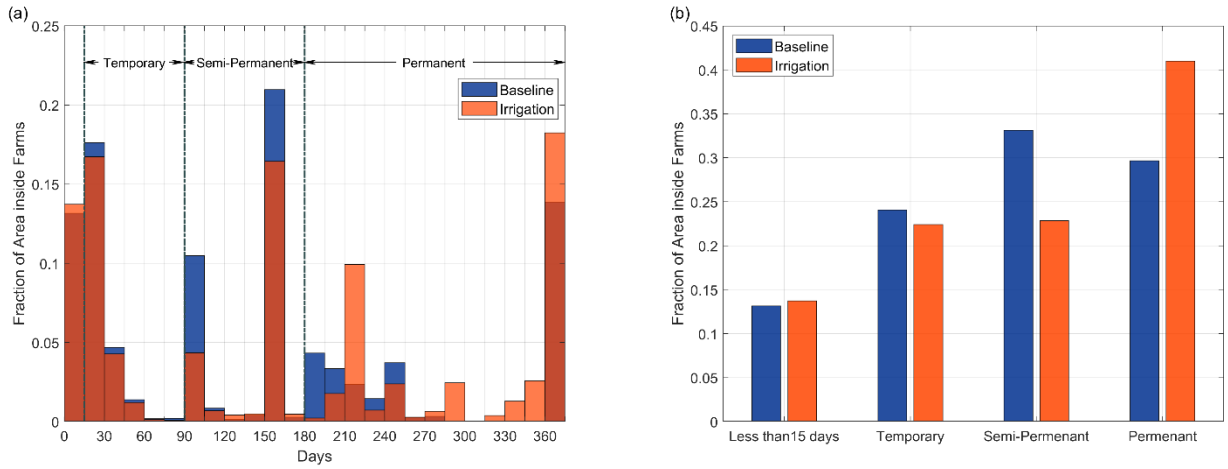


Figure 2.7: The fraction of area inside the irrigated farms for each potential larval habitat types under the baseline and irrigation scenarios. (a) shows the histogram of the maximum duration of ponding within the year for the grid cells in each type of habitats expressed as a fraction of the total area of the farms. The bin size was 15 days. Temporary, semi-permanent, and permanent larval habitats were typically characterized by ponding duration of 15-90 days, 90-180 days, and 180 days and above, respectively. The baseline scenario is represented in blue and the irrigation scenario is represented in orange. The darker orange bin is a result of the two overlapping. (b) shows the comparison of the fractional area occupied by non-habitats (less than 15 days) as well as potential temporary, semi-permanent, and permanent larval habitats inside the farms. Each grid cell within the farm was categorized based on its maximum ponding duration.

From the baseline scenario, 13.2% of the area was not favorable for larval habitats because the maximum duration of ponding in those areas was less than 15 days. The most common maximum ponding duration was between 150-165 days, which accounted for more than 20% of the area. This was followed by 15-30 days and 360 days and above which made up 17.6% and 13.8% of the area respectively. With irrigation, there was a general increase in the maximum ponding durations. The most common maximum ponding duration was 360

days and above, accounting for 18% of the area. Noticeably, the area with maximum ponding duration between 210-225 days increased fourfold to 10%. The remaining increase was for 285 days and above. Counter-intuitively, the area that was not conducive as larval habitats (i.e., maximum ponding duration less than 15 days) also increased slightly by 0.6%. This was because irrigation raised the overland flowrate in these areas, mostly near streams, and made them unfavorable for breeding.

In Figure 2.7b, we grouped the maximum ponding durations into stability periods corresponding to temporary (2 weeks to 3 months), semi-permanent (3-6 months), and permanent (6 months and above) habitats based on field observations from a study at the site (Hawaria et al., 2020). Temporary habitats such as puddles retain water for a short period while permanent habitats such as stream edges and swamps hold water much longer and are more stable. For the baseline scenario, semi-permanent habitats were the most common, occupying 33.1% of the area, while permanent and temporary habitats also accounted for 29.6% and 24.1% of the area, respectively. After irrigation, there was a significant shift from semi-permanent habitats, which reduced to 22.9% of the area, to permanent habitats which increased to 41% of the area. There was also a slight reduction in the extent of temporary habitats to 22.4% of the area.

2.3.3 Temporal pattern of potential larval habitats

To shed light on the temporal patterns, we evaluated $F(WI > T)$, the fractional coverage of potential larval habitats inside farm, for each day throughout the year. We only focused on the area inside farms since irrigation does not have a significant impact on the area outside farms. As shown in Equation (2.3), $F(WI > T)$ is defined as the ratio of $C(WI > T)$, the number of cells for which the WI (i.e. persistence of ponding) exceeded T days, to C_{farm}, the number of cells within the farm area. T is set as 10 days and 15 days, corresponding to critical and normal conditions, respectively.

$$F(WI > T) = \frac{C(WI(x,y,t) \geq T)}{C_{farm}}, T \in \{10,15\} \quad (2.3)$$

In Figure 2.8a, $F(WI > 10)$ increased steeply on January 10th as WI started increasing from 0 at the beginning of the year. For the baseline scenario, the fractional

coverage decreased minimally from 0.18 throughout the dry season despite the sporadic spike in precipitation. At the onset of the rainy season, the peak rainfall event of the year from May 5th to May 11th caused a sharp increase in $F(WI > 10)$ from 0.15 to 0.61 and thereafter, the relentless rainfall maintained the fractional coverage at about 0.6. Throughout the rainy season, there were four recurring peaks at a frequency of about 2 months. Post-rainy season, $F(WI > 10)$ dropped gradually to below 0.2 after the last peak at the end of October.

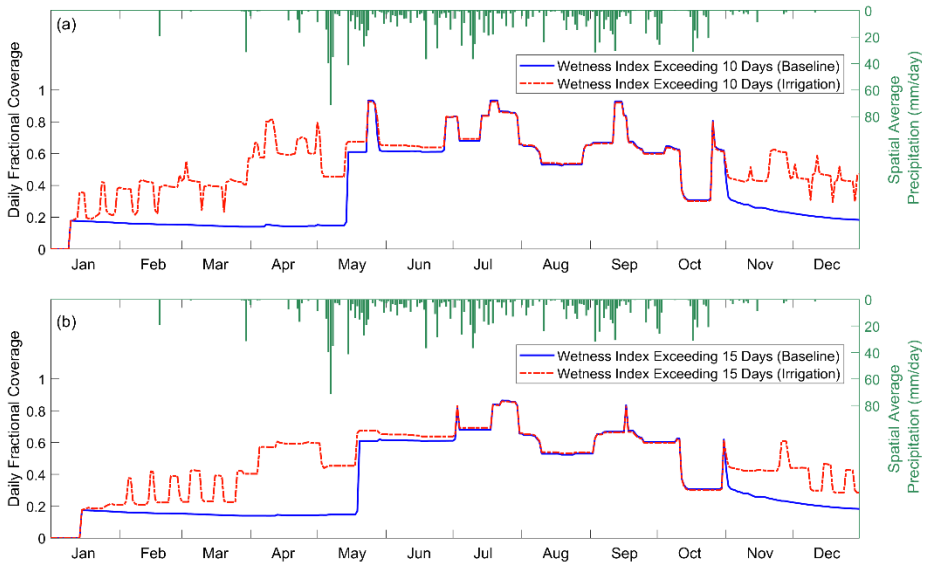


Figure 2.8: Daily variations in the extent of the potential larval habitats for the year. Time series of the fractional coverage of areas with WI exceeding (a) 10 days and (b) 15 days.

For the irrigation scenario, $F(WI > 10)$ increased during the dry season from January to March with visible cyclical variations between 0.2 and 0.4 due to the rotation of irrigation among the four farms. Subsequently, the spike in rainfall at the end of March combined with the higher antecedent soil moisture from irrigation brought forward the step increase in the fractional coverage to April from May in the baseline scenario. As irrigation stopped at the end of April, $F(WI > 10)$ gradually dropped back to the same level as the baseline scenario at the end of June. In the dry season from November to December, the fractional coverage started to deviate from the baseline scenario again with cyclical fluctuations, gradually decreasing towards the end of the year.

In Figure 2.8b, $F(WI > 15)$ remained largely the same for the dry season but the peaks were moderated in the rainy season, compared to $F(WI > 10)$. There was one less peak at the end of May in the early rainy season because the watershed did not accumulate enough rainfall for the persistence of the ponded areas to exceed 15 days. Specifically, for the irrigation scenario, the increase in fractional coverage during the dry season was moderated and less sensitive to the spikes in rainfall. Similarly, irrigation resulted in the early onset of the steep increase in $F(WI > 15)$ in April following the spike in rainfall at the end of March. Also, it took two months after the end of irrigation in April for the fractional coverage to return to the same level as the baseline.

From $F(WI > 10)$ and $F(WI > 15)$, we calculated the corresponding monthly mean, $MF(WI > 10)$, and $MF(WI > 15)$ as well as the 95th confidence interval as shown in Figure 2.9. In Figure 2.9a, $MF(WI > 10)$ in the baseline was the highest for the months between June and September, constituting a four-month window in which at least 50% of the area was conducive for larval habitat formation. Of the four months, the highest monthly mean fractional coverage was in July at 79.9%. Irrigation extended the window to include the months of April and May. The monthly mean fractional coverage increased 4.5 times to 64.3% in April and 1.4 times to 63.7% in May. The $MF(WI > 10)$ for the rest of the months in the window (i.e., June to September) remained one of the highest but the increase due to irrigation was not statistically significant ($p > 0.01$). July remained as the month with the highest monthly mean fractional coverage at 80.0%. In Figure 2.9b, $MF(WI > 15)$ was generally slightly lower than $MF(WI > 10)$ for both the baseline and irrigation scenarios but the general trends were the same.

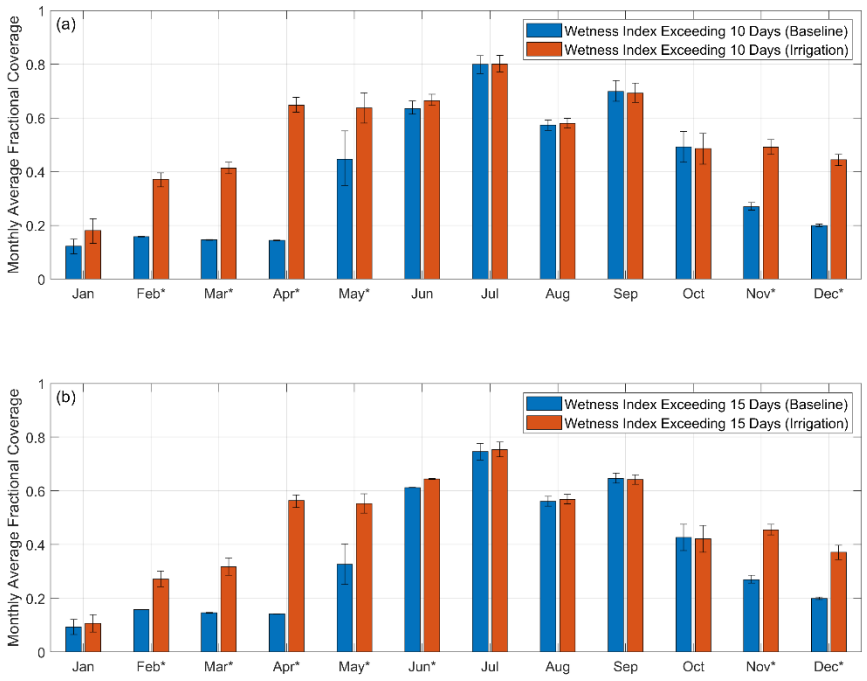


Figure 2.9: Monthly variation in the extent of the potential larval habitats for the year. Monthly mean fractional coverage of areas with a probability of WI exceeding 10 days (a) and 15 days (b). The 95% confidence interval is indicated at the top of each bar chart. The asterisks (*) next to the month on the x-axis indicate that irrigation increased the fractional coverage of the potential larval habitats for the month based on a 2-sample t-test ($p<0.01$).

2.4 Discussion

2.4.1 Impact of irrigation on spatiotemporal distribution of potential larval habitats

This study, for the first time, employed an integrated hydrological model to predict the proliferation of potential larval habitats around an irrigated field during the dry and rainy seasons. Irrigation resulted in a much higher probability to find potential larval habitats in the dry season. Although irrigation was not applied during the rainy season, a slight increase in the probability of potential larval habitat occurrence was also observed due to the higher antecedent soil moisture from the dry season promoting the formation of larval habitats further. Considering that larval habitat availability is one of the direct predictors for vector abundance, our findings tie in with local studies (Demissew et al., 2020; Hawaria et al., 2019, 2020) in the Arjo-Didessa sugarcane plantation which showed a higher occurrence of Anopheline mosquito larval habitats, larval productivity and abundance in the irrigated areas than non-irrigated areas in both dry and wet seasons.

Separately, previous studies have identified irrigation schemes as the cause for prolonging mosquito season and extending the period of malaria transmission (Ijumba & Lindsay, 2001; Kibret et al., 2010). In our study, it was found that the stability of potential larval habitats was similarly prolonged, with a significant shift from semi-permanent (3-6 months) to permanent habitats (6 months and above) in the irrigated farm areas. Originally, the semi-permanent habitats were the most common, occupying a third of the area inside farm without irrigation. In the irrigation scenario, the area of the permanent habitats became the most common, increasing to more than 40% of the area inside farm.

The temporal variations in the extent of the potential larval habitats indicate that rainfall exerts a strong influence. It is common that peak mosquito breeding seasons follow rainfall in the tropics (Kibret et al., 2019). In our study, the occurrence of potential larval habitats was highest for the period from June to September, during which at least 50% of the area inside farms were potential larval habitats. This overlaps with a significant portion of the rainy season. From the irrigation scenario, we observed that irrigation dampened the seasonality of the potential larval habitats by increasing the wetness index in the dry season and extending the peak larval habitat occurrence window to include the months of April and May. Elsewhere in Africa, irrigation is also known to reduce the dependence of larval habitat patterns on rainfall, changing them from seasonal to perennial (Mabaso et al., 2007).

This calls for the need to modify current irrigation strategies and develop tailor-made interventions to mitigate mosquito breeding around irrigated fields in order to combat malaria. Ideally, the aim is to optimize irrigation to minimize larval habitat availability while meeting crop water requirement. To this end, our model has the flexibility to simulate diverse types of irrigation such as flood irrigation, groundwater irrigation, sprinkler irrigation and drip irrigation as well as water allocation strategies. Hence, beyond predicting potential larval habitats, the model can help configure the outline of the irrigation design and sieve out some of the more pertinent and effective strategies. For example, the main method of irrigation in Ethiopia is surface irrigation which has low implementation cost but is known to be inefficient in water use (Eshete et al., 2020) and can aggravate malaria transmission by providing an ideal larval habitat (Bett et al., 2017). For regions where resources are limited, our model can be used to identify irrigated farms with the most serious ponding to prioritize

the installation of higher water efficiency but also costlier irrigation systems such as drip or sprinkler irrigation. It can also be coupled with a water allocation algorithm (Condon & Maxwell, 2013) to investigate the larval habitat distribution under more complex water management operations.

2.4.2 The implication of model assumptions and simplifications on results

In this study, we chose to simulate the surface layer soil saturation at 50 m resolution, coupled with a threshold, to quantify ponding instead of explicitly simulating the surface water depth. Without high resolution and accurate topographic information, it was not feasible to achieve the latter at the scale of the larval habitats surveyed in the study area, most of which measure less than 100 m^2 each. It has been shown that a minimum of 3 model grid cells is required across the width of the land depression for a good balance between accuracy and computational effort when simulating a flood extent (Gallegos et al., 2009). This requires accurate data with a minimum resolution of 3 m that can only be obtained using Lidar, RTK (Real-Time Kinematics), or PPK (Post Processed Kinematics) with aircraft or drone, which is time consuming and expensive. Regardless, the computational efficiency of the model with high-resolution DEM will pose another challenge even if they were available. As such, the model is not intended to pinpoint the exact location of each larval habitat in the study area. Instead, it provides information on the overall likelihood of ponding for each grid cell based on the interactions between system properties and forcing variables at various temporal and spatial scales. The strength of this approach is that the model can afford to run on a regional scale at a fairly high spatial resolution while keeping computational requirements manageable. Furthermore, all the primary data used are freely available for all regions of the world and hence, this framework provides a great opportunity to extend potential larval habitat simulation into other locations without incurring high data acquisition cost.

In terms of performance, the model was able to predict ponding at all the validation points after calibration. As shown in Figure A.6, the POD for both calibration and validation reached 1 (i.e. all the points are detected) at the optimal soil saturation threshold θ of 0.48. However, this does not mean the model is perfect as the calibration and validation did not

account for overprediction at locations without ponding since the survey was only for locations with ponding. This will be resolved in future by ongoing field survey efforts toward compiling a more spatiotemporally comprehensive dataset including locations without ponding. Another limitation was that calibration was only performed on the soil saturation threshold but not the ParFlow model per se although the simulated soil saturation was realistic from its general behavior. This was due to the lack of data to verify surface and subsurface flow rates at a relevant spatial scale. With the collection of more data in future, the key parameters can be fine-tuned to improve model predictions. A more detailed irrigation schedule would also provide insight into the water usage, irrigation management, and surface land cover during sugarcane growth. Such data could serve as a better guide in modeling pond formation that incorporates the effect of irrigation. Unfortunately, records for canal water flow, water usage, and field operations were not available at the time of visit to the Arjo-Didessa Sugar Factory, and only the summary data and annual plan could be obtained.

2.5 Conclusion and future prospects

Using high resolution distributed hydrologic modeling with remotely sensed data, we demonstrated a quantitative assessment of potential malaria vector larval habitats in terms of the spatial distribution and temporal variation. We also evaluated the relative influence of key environmental processes such as rainfall and irrigation on the habitats. Results indicated a higher probability to find potential larval habitats inside the farms, at around 40% of the year, than outside the farms, at less than 10% of the year. Our model also showed that rainfall exerted a strong influence on larval habitat availability based on predictions that at least 50% of the area inside farms were potential larval habitats from June to September during the rainy season. Further, modeling revealed that irrigation increased the probability of finding potential larval habitats inside the farms to 67%. Irrigation also dampened the seasonality of the potential larval habitats such that the peak larval habitat occurrence window during the rainy season was extended. Lastly, the stability of larval habitats was prolonged, with a significant shift from semi-permanent habitats to permanent habitats

lasting beyond 6 months, pointing to the impact of irrigation in creating conducive mosquito habitats throughout most of the year.

Since the effectiveness of major malaria vector control measures is decreasing due to mosquito insecticide resistance and outdoor transmission, the role of LSM as a supplementary vector control tool to reduce malaria transmission becomes more significant. As such, hydrologic modeling with publicly available data, presented herein, constitutes a promising direction in terms of providing a dynamic and systematic approach for the identification and elimination of larval habitats by environmental modification and manipulation. For hydrologic modeling to fulfil its promise in the area, enhanced observational efforts are required in future. Thorough calibration and validation will be critical in evaluating the robustness and quantifying the uncertainty of the model.

Food security will bring economic growth and remains one of the priorities in Africa. To this end, investment in dams and irrigation systems is increasing rapidly in Africa over the past decade. Unfortunately, this might increase the risk of malaria due to environmental modifications and microclimate changes. The broader goal of our research is to harness the hydrological results, along with other epidemiological, entomological and social-economic factors, to translate the knowledge of potential larval habitats to useful information on the spatiotemporal distribution of malaria transmission risks. Remotely sensed data can enable this type of modeling in data-scarce regions where malaria presents a grave threat. This framework has great potential to integrate with malaria epidemiologic modeling such as EMOD (P. A. Eckhoff & Wenger, 2016) to predict malaria risk under different environmental modifications to guide decision-making in water resource management, changes to agricultural practice, and disease prevention.

Chapter 3

Topographic Hydro-conditioning to Resolve Surface Depression Storage and Ponding in a Fully Distributed hydrologic model

3.1 Introduction

Surface depressions have a direct impact on surface and subsurface flows through hydrologic, biochemical and biological exchanges (Cohen et al., 2016). Primarily, runoff can be modulated and delayed by surface depressions through fill-spill dynamics and storage effect (Ameli & Creed, 2017; Brooks et al., 2018). This has been shown to improve downstream water quality and support habitat functions because of longer residence times (Biggs et al., 2017; Cheng & Basu, 2017; Charles Nathan Jones et al., 2018). In addition, large scale surface depressions in the form of wetlands and lakes can be critical for water supply as they replenish aquifers through seepage during dry conditions (Liu et al., 2016).

With a growing emphasis on the impact of surface depressions on hydrologic processes, semi-distributed hydrologic models have been widely used to provide new insights (Evenson et al., 2016; N. Wang et al., 2021). One approach is to aggregate surface depressions within each subbasin into a lumped depression that functions conceptually as a bucket. Water is stored in the bucket and spills as runoff when it exceeds a parameterized threshold volume (Hay et al., 2018; Liu & Schwartz, 2011; Rajib et al., 2020). This type of approach is simple to implement but fails to account for the spatial distribution of the depressions. In reality, some surface depressions may fill and contribute runoff water earlier than the others but in the lumped approach, water is only released to the streams when all the surface depressions (i.e., the aggregated depression) are fully filled. The other approach is to model surface depressions as individual units which better represents wetlands and their hydrologic connectivity. For example, Chu et al. (2013) developed a puddle-to-puddle (P2P) model where the study domains were divided into multiple puddle-based units (PBU). Each PBU contains the highest-level puddle and its contributing area. The PBU can drain to a downstream PBU based on an overflow threshold and the same applies for its embedded

lower-level puddles based on a puddle delineation (PD) algorithm (Chu et al., 2010). Building on the PD algorithm and PBU concept, subsequent improvements were made to account for infiltration and unsaturated flow (Yang & Chu, 2015), improve the channel-puddle cascading mechanism (Nasab et al., 2017) and enhance computational efficiency (N. Wang & Chu, 2020).

Fully distributed hydrologic models have different needs for topographic data compared to semi-distributed models, given that the former is designed to resolve flow paths along and into the land surface (Golden et al., 2014), which in turn supports numerous geophysical applications such as material transport and reactive flows (Golden et al., 2014, 2017; C. Nathan Jones et al., 2019). In particular, fully distributed models require careful attention to spatial distribution of land surface heights as this drives the spatiotemporal variability of land surface storage, infiltration, runoff and soil moisture (Amado et al., 2018; Liu et al., 2016). Hence, fully distributed hydrologic modeling at fine resolution calls for careful attention to the source of topographic data, which is generally supplied by a DEM (Gardner et al., 2018; Jan et al., 2018; Tavares da Costa et al., 2019).

DEMs are developed in many different ways, but fine-resolution DEMs used in hydrologic modeling are typically derived from aerial lidar data or optical data with photogrammetric methods (Lidberg et al., 2017). Raw point cloud data is filtered to differentiate ground surface points from vegetation canopy, and then gridded to produce a preliminary DEM. Subsequently, DEMs are hydro-conditioned (i.e., filling or breaching topographic cells) to limit variability or noise in land surface slopes for smoother flow routing (i.e., lower chance of computational instability) (Chow & Ben-Zvi, 1973; Rieger, 1998; Zhang & Cundy, 1989), yet this process often removes physical land surface depressions that are highly relevant to hydrologic processes at the land surface. Previous studies have worked to identify land surface depressions for the purpose of characterizing the location and size of land surface features such as wetland and karsts (Bertassello et al., 2020; Li et al., 2011; Moreno-Gómez et al., 2019; Wu et al., 2016), but these work did not focus specifically on creating a DEM suited to resolving fine-scale land surface processes with spatially distributed hydrologic models. Hence, the objective of this study is to present a novel topographic processing workflow, referred to as the Depression-Preserved DEM

Processing (D2P) algorithm, with the aim of resolving surface depressions based on the scale of interest and smoothing small scale, non-physical variability in DEMs (e.g., Sanders, 2007) for depression-integrated and efficient modeling of land surface hydrology.

Our approach is designed for hydrologic models that calculate flow across 4 cell faces (i.e., D4) but can be modified easily to handle more flow directions (e.g., D8). Alongside a processed DEM, the algorithm also generates slopes across the cell faces. This adds to the limited number of DEM processing algorithms that can generate slope inputs for D4 routing models using finite difference methods (Condon & Maxwell, 2019). In this study, we detail the workflow of D2P and examine its utility in processing the DEM as an input for the integrated hydrologic model ParFlow (Ashby & Falgout, 1996b; J. E. Jones & Woodward, 2001; Kollet & Maxwell, 2006; Maxwell, 2013). We compare the processed DEM and evaluate the hydrologic simulation against the traditional filling method by Condon and Maxwell (2019) using a case study on GCEW.

3.2 Materials and methods

3.2.1 Study area

GCEW is located in northern Mississippi, US and the area of the watershed is 21.7 km^2, with an elevation ranging from 64.8 m to 129.1 m above mean sea level (Figure 3.1). The land use is mainly idle pasture, followed by forest and cultivated land and there is a high density of small water bodies such as farm ponds (Yasarer et al., 2018). Based on the satellite imagery from the US Fish and Wildlife Service (FWS) National Wetlands Inventory (NWI) (https://www.fws.gov/wetlands/), there are a total of 93 fresh water ponds ranging in size from 500 m^2 to 11,000 m^2.

The DEM used in this study was downloaded from a derived-LiDAR product by the Mississippi Automated Resource Information System. The DEM represents bare earth surface that was created from LiDAR points collected from 2009 to 2010 by the U.S. Army Corps of Engineers (USACE) for the Mississippi Delta Phase 1 project (https://www.maris.state.ms.us/HTML/DATA/Elevation.html#gsc.tab=0) and hydro-flattened. The spatial resolution is 1 m and the vertical accuracy is 0.09 m (Root Mean Square

Error). The DEM was first smoothed by a median filter to remove roughness and aggregated to 10 m resolution to keep the hydrologic simulation computationally manageable.

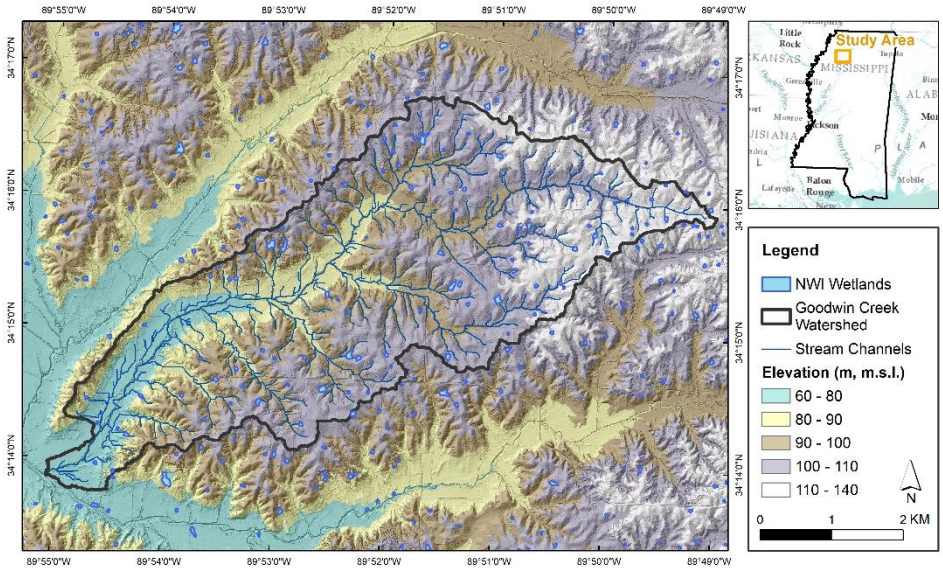


Figure 3.1: GCEW study area.

3.2.2 Overview of Depression-preserved DEM Processing (D2P) algorithm

The D2P algorithm comprises six steps as shown in the flowchart in Figure 3.2. Using DEM and NWI wetland as input data, it identifies surface depressions that are likely to be physically meaningful and generates a processed DEM that is smoothened to reduce fine scale variability while preserving those depressions. Slope data is then derived from the processed DEM. Details of steps 1 to 6 of the flowchart are explained from Sections 3.2.2.1 to 3.2.2.6.

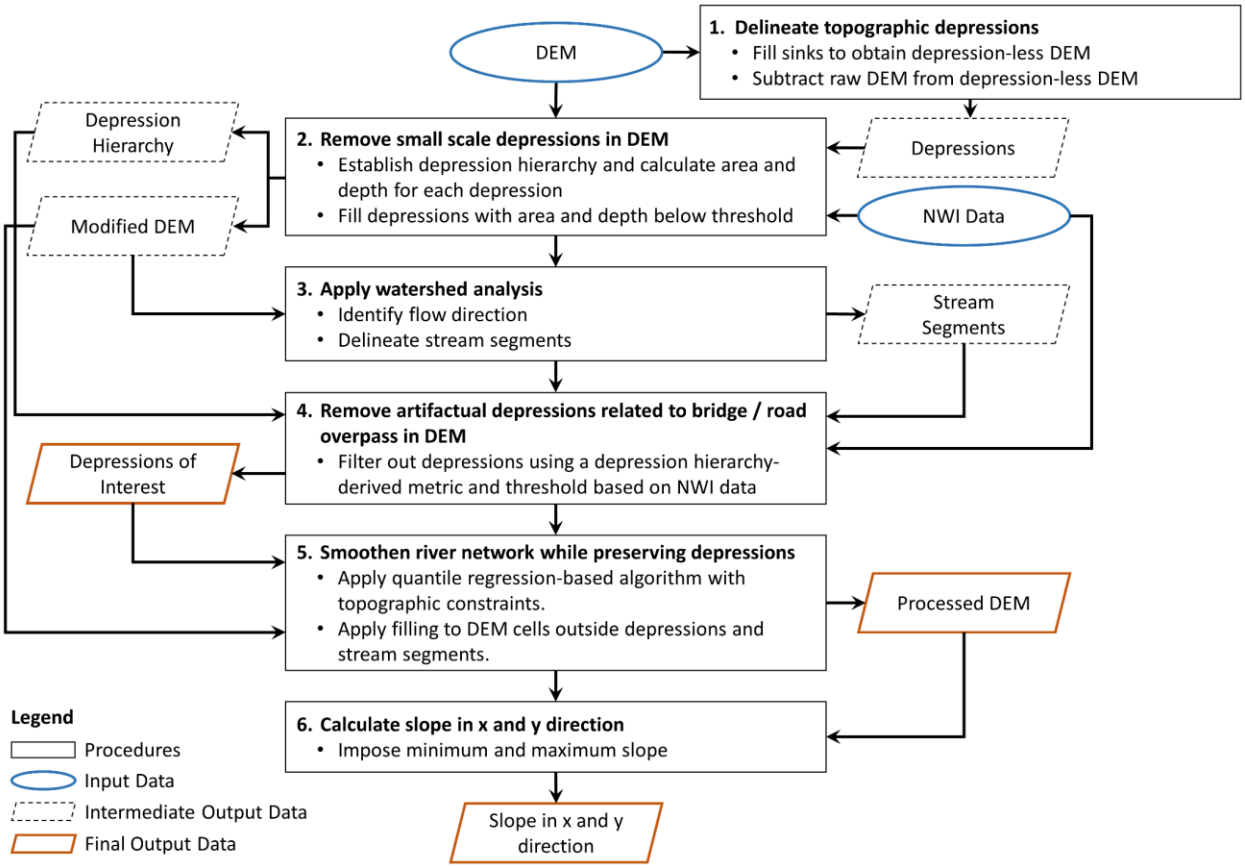


Figure 3.2: Schematic flowchart for the proposed Depression-preserved DEM Processing (D2P) algorithm.

3.2.2.1 Topographic sink extraction

A depression-less DEM was first obtained by filling the sinks using Priority-Flood, a depression-filling tool (R. Barnes et al., 2014). All the topographic depressions comprising real depressions and artifacts were then extracted by subtracting the input DEM from the depression-less DEM. The artifacts are caused by systematic and random errors from the precision of data acquisition instrument and processing techniques (Zhu et al., 2013).

3.2.2.2 Removal of small-scale depressions in DEM

From the extracted topographic depressions (Section 3.2.2.1), an improved adaptation of the level set method (Wu et al., 2019) was applied to characterize the depressions in terms of their geometric properties and hierarchy. A depression hierarchy describes the relationship between each depression whereby depressions can themselves

contain smaller ones. The most-nested depression is referred to as a Level 1 depression and the level increases as the depressions combine to form bigger depressions (Figure B.1).

The level-set method by Wu et al. (2019) establishes the hierarchy by screening each depression from the lowest elevation and moving upwards to determine the spill point at which depressions combine. One disadvantage of this method is that the screening was performed at fixed intervals and could overlook certain depressions (R. Barnes et al., 2019). To avoid this, the screening interval must be reduced but the computation can become inefficient. Therefore. we modified the method to screen at each unique DEM cell elevation within the depression instead of at fixed intervals. This ensures that each depression is assigned the correct level in the hierarchy. The level of each depression will be used in Section 3.2.2.4.

After establishing the hierarchy, the area of each depression was determined and the average depth of each depression, D, was calculated using Equation (3.1).

$$D = \sum_{i=1}^{n}(z_s - z_i)/n \tag{3.1}$$

where n is the number of DEM cells within the depression, z_s is the spill elevation and z_i is the elevation of the i^{th} cell in the depression. Finally, depressions with an average depth and area smaller than their respective predetermined thresholds of 0.0036 m and 900 m2 respectively were assumed to be inconsequential to the overall hydrologic process. These small-scale depressions were removed from the DEM by filling to generate a modified DEM. This is also an effective and computationally simple way to remove artifacts from the DEM (Li et al., 2011). The thresholds were derived from a sensitivity assessment of the final delineated depressions using different threshold values to achieve a balance between matching the wetland data from NWI and excessive false matches (Section B.1.2). The considerations for the setting of the threshold are also discussed in Section 3.4.1.

3.2.2.3 Watershed analysis

The watershed analysis was implemented using the GRASS module, r.watershed (GRASS Development Team, 2020). An A* least-cost search algorithm (Hart et al., 1968; Metz et al., 2011) was applied on the modified DEM from Section 3.2.2.2 to generate flow direction

and river network. This algorithm does not require the DEM to be further altered to determine flow direction and extracts the path of rivers through depressions more accurately compared to traditional methods (Planchon & Darboux, 2001; L. Wang & Liu, 2006) with sink filling (Metz et al., 2011).

3.2.2.4 Classify riverine depressions for targeted preservation

In this section, we focus on riverine depressions after having removed small-scale depressions from the DEM (Section 3.2.2.2). DEMs are especially prone to errors along rivers due to the difficulty of sensors in penetrating thick riparian vegetation and water bodies as well as the resolution limits in resolving the channel bottom elevation (Lindsay, 2016b; Schwanghart & Scherler, 2017). This creates spurious topographic variability along rivers leading to depressions inside the channel width with discontinuous bottom slopes that can cause model instability (Hengl, 2006; Iserles, 2009; Yu et al., 2020). In addition, high resolution LiDAR-derived DEMs can contain false hydrologic barriers such as bridges or roads (Carlson & Danner, 2010) that will form false depressions during the depression extraction process in Section 3.2.2.1. Such false depressions are rarely addressed in previous depression-integrated studies.

Hence, our goal is to filter out the aforementioned spurious depressions within the channel width (Figure 3.3a) and depressions arising from false hydrologic barriers (Figure 3.3b) to finally keep the remaining water bodies that partially coincide with rivers (Figure 3.3c). To achieve this, we developed a new metric that takes advantage of the hierarchy of depression levels. The metric, referred herein as the river to depression ratio (RtD), provides a measure of the contribution of the river to the makeup of the depression. The RtD was calculated for each highest-level depression using the formula in Equation (3.2).

$$RtD = \left(\frac{n_r}{n}\right)\left(\frac{n_{L_1,r}}{n_{L_1}}\right) \tag{3.2}$$

In Equation (3.2), n is the number of DEM cells within the highest-level depression and n_r is a subset number of cells from n that are within the river channel width; while n_{L_1} is the number of DEM cells in all Level 1 depressions (i.e., the most nested depressions) within the highest-level depression and $n_{L_1,r}$ is a subset number of cells from n_{L_1} that are

within the river channel width. The equation is a product of two ratios. The first ratio is a measure of the extent of the depression that coincides with the river. The higher the first ratio, the higher the chance the depression falls within the river channel width (Figure 3.3a). The second ratio was introduced to specifically distinguish between false depressions arising from roads or bridges and actual water bodies that partially coincide with the river. It does so by drawing on the depression hierarchy from Section 3.2.2.2 to gage the geometry of the depression. Ponds and wetlands generally have a bowl-like structure with a flat bottom which manifests itself by the Level 1 cells spreading beyond the river, resulting in a lower $n_{L_1,r}/n_{L_1}$ (Figure 3.3c). On the other hand, in false depressions arising from road or bridge crossings, Level 1 cells are mostly concentrated within the river channel width, resulting in a higher $n_{L_1,r}/n_{L_1}$ (Figure 3.3b).

Using Equation (3.2), depressions with RtD lower than a certain threshold would be selected for preservation in the next section. To calibrate the threshold, the RtD of the depression associated with each wetland from NWI data was calculated. The threshold was set based on the maximum calculated RtD.

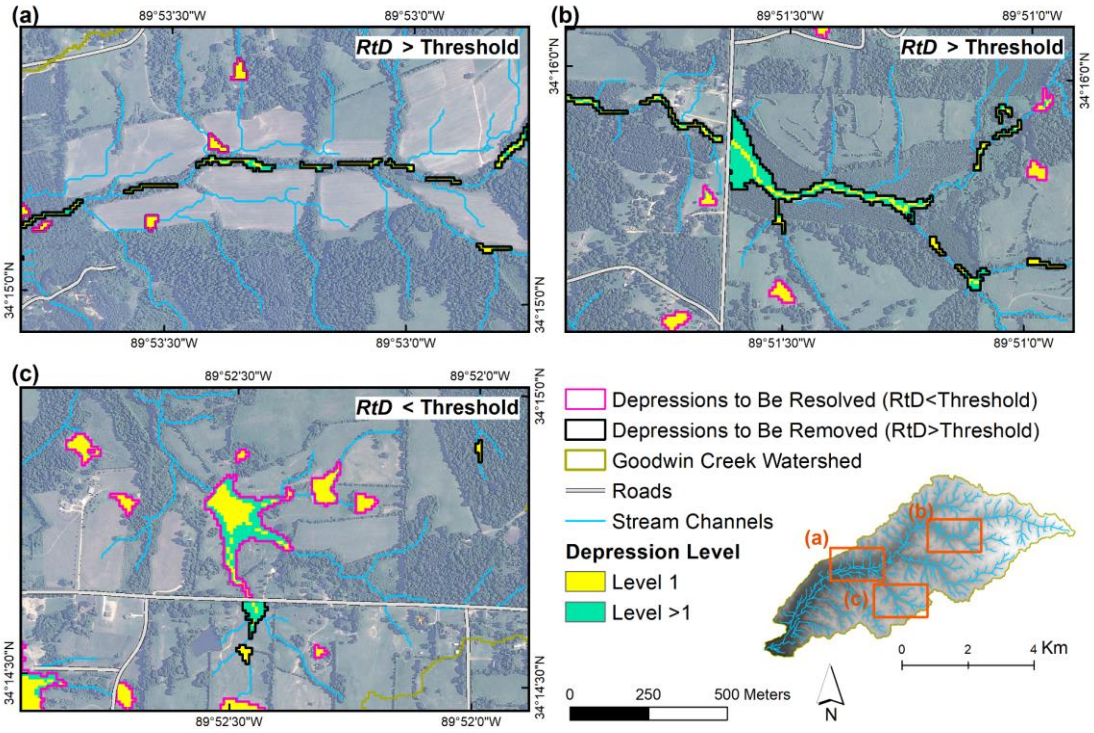


Figure 3.3: Examples of (a) depression within channel width, (b) false depression at road/bridge overpass (c) water bodies partially coinciding with river within GCEW.

By this point, we had identified physically meaningful depressions that we wanted to preserve in the DEM. The preserved depressions compared well against the wetlands in the NWI data (Figure 3.4). The excluded depressions, namely small-scale depressions, depressions within the channel width and depressions associated with false hydrologic barriers were all assumed to be artifacts.

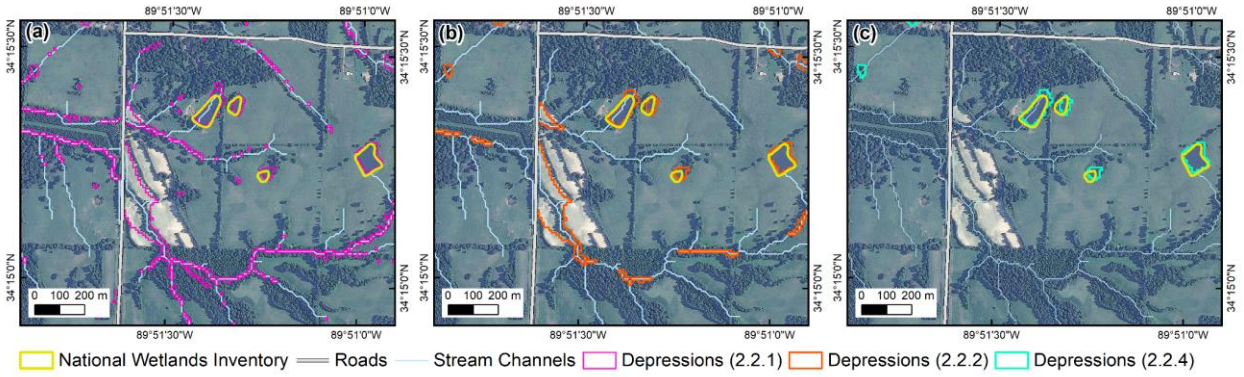


Figure 3.4: Illustration of gradual exclusion of select depressions in GCEW study area. (a) Delineated sinks from Section 3.2.2.1. (b) Post-removal of small-scale depressions from Section 3.2.2.2. (c) Remaining depressions after the removal of depressions within the channel width and depressions associated with false hydrologic barriers from Section 3.2.2.4.

3.2.2.5 River network smoothening

We applied a quantile regression algorithm to smoothen the river while preserving the depressions partially coinciding with the river identified in Section 3.2.2.4. The smoothening aims to maintain the general profile while reducing fine scale variability and adverse slopes. Previous studies smoothen the river profile by modifying the elevation universally based on the average slope of the reach (M. L. Barnes et al., 2016; Condon & Maxwell, 2019). The approach we adopt here offers the flexibility to process the river segment intersecting with depression differently from other segments.

Generally, the algorithm works by minimizing the sum of absolute values of the residuals given by the difference between the smoothened elevation and the raw elevation as well as the roughness. The underlying optimization function of the algorithm, known as

the constrained regularized smoothing (CRS) algorithm, is shown in Equation (3.3) (Schwanghart & Scherler, 2017).

$$\textbf{arg min} \sum_{i=1}^{n}(\rho_\tau(z(x_i) - \mathbf{I}z_\tau(x_i))) + s\int[z_\tau''(x)]^2 \text{ dx}, \tag{3.3}$$

where $z(x_i)$ is the unsmoothed elevation along river profiles x distance upstream of the watershed outlet, n is the number of elevations, $\mathbf{I}$ is the identity matrix, $\tau \in [0,1]$ is the constant chosen according to which quantile needs to be estimated. In this case, τ was set at 0.5. z_τ is the estimated elevation. $\rho_\tau(\cdot)$ is a loss function dependent on the residuals r_i given by

$$\rho_\tau(r_i) = (\tau - \|_{r_i<0})r_i, \tag{3.4}$$

where r_i is defined as $z(x_i) - \mathbf{I}z_\tau(x_i)$ and $\|$ is an indicator function that has a value of 1 if the residual r_i is less than 0 and a value of 0 otherwise. s dictates the degree of smoothening and is defined by

$$s = (\Delta x)^2 K\sqrt{\frac{n}{p}}, \tag{3.5}$$

where Δx is the resolution of the DEM, p is the number of second derivative terms in Equation (3.3), K is the parameter that scales the degree of smoothing.

In Schwanghart & Scherler (2017), the river profile was forced to decrease in the downstream direction and the degree of smoothening s was set to be uniform throughout the entire river network. To avoid distorting the actual depressions that intersect the river path (e.g. Figure 3.3c), we introduced separate conditions to process river segments within depressions and river segments outside of depressions. For river segments outside depressions, the downstream elevation was forced to be lower than the upstream by a minimum difference ε as shown in Equation (3.6)

$$z_\tau(x) \geq z_\tau(x - \delta x) + \varepsilon, \tag{3.6}$$

This was not applied for river segments inside depressions to avoid removing the depressions. In addition, s was set to be higher for river segments outside of depressions but

lower in river segments within depressions to preserve any high curvature structures (e.g., dams) downstream of the depressions.

Lastly, to prevent the river cell being raised inadvertently creating depressions outside the river, we introduced a new condition whereby the elevation of the river segment was globally set not to exceed that of the riverbanks by ε.

$$z_\tau(x) \leq \min\left(z_{\tau,y+\delta y}(x), z_{\tau,y-\delta y}(x)\right) - \varepsilon \tag{3.7}$$

Figure 3.5 illustrates the different constraints that were imposed on the river segments within the depressions and outside the depressions. The modified algorithm will be referred to as the adapted CRS algorithm. Following the implementation of the adapted CRS algorithm, we temporarily masked out the stream segments and preserved depressions and applied filling on any sinks created as a by-product in the DEM (R. Barnes et al., 2014). A final processed DEM was obtained.

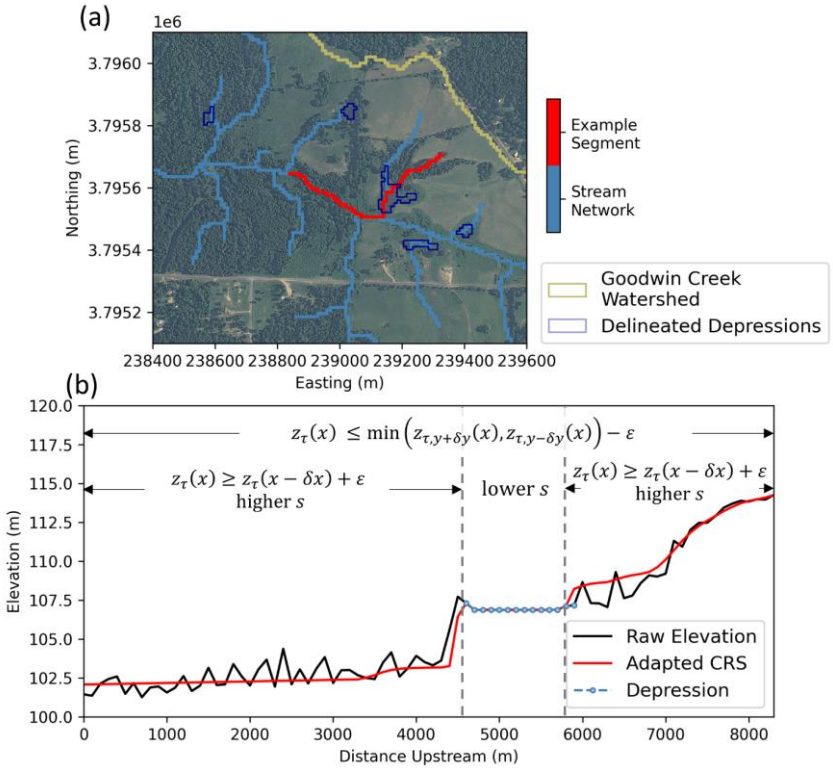


Figure 3.5: Illustration of adapted CRS for smoothening of an example river segment in GCEW study area. (a) Plan view of river segment in red coinciding with a depression delineated in dark blue. (b) Corresponding elevation profile of river segment from (a) before

and after the implementation of the adapted CRS for river smoothening. A downstream slope was enforced on the river except the segment within the depression. The elevation of the rivers segment was globally set not to exceed that of the riverbanks and a lower degree of smoothening was applied on the river segment within the depression.

3.2.2.6 Slope calculation

The outcome of the above steps was a processed DEM with smoothened rivers and retained depressions. In this step, slopes in the north-south and east-west directions were calculated for each cell. The north-south slope is defined at the upper cell face between a cell and the adjacent cell to the north while the east-west slope is defined at the right cell face between a call and the adjacent cell to the east. The calculated slope direction for cells outside of depressions was adjusted to match the flow direction by reversing the sign whenever there was a discrepancy. The slope magnitude remained unchanged but subject to a maximum slope threshold and a minimum slope threshold.

3.2.3 Comparative analysis

To assess the performance of D2P, we evaluated the extent and magnitude of its modification on the raw DEM and the subsequent impact on overland flow. Since ParFlow is used in this study, we chose PriorityFlow (Condon & Maxwell, 2019) , a recent DEM conditioning algorithm developed for ParFlow, to be compared against D2P. The algorithm followed the traditional way of removing all the sinks and the approach is justified when a kinematic wave approximation is used for the overland flow simulation.

DEM processing algorithms are known to introduce errors since it is difficult to ascertain the origin of each depression so the algorithm should ideally achieve a reasonable flow path with minimal modification to the DEM (Lindsay & Creed, 2005). Hence, we quantified the change in elevation of the processed DEM relative to the source DEM as a measure of the change to the landscape. In addition, we performed a rainfall recession simulation in ParFlow using the processed DEMs and the raw DEM by applying a rainfall rate of 5 mm/hour for 10 hours followed by 20 hours of recession, similar to the test used by Barnes et al. (2016) and Condon and Maxwell (2019) The simulation was performed on an impervious surface to focus on surface water processes. Diffusive wave approximation was used for overland flow.

Table 3.1 summarizes the five DEMs that were used in the comparative study. The raw DEM, the processed DEM from the conventional algorithm and the processed DEM from D2P are represented by Case 0, Case 1 and Case 4 respectively. Case 1 fills the depressions first and smoothens the river by enforcing a constant average slope between two ends of each reach. Due to the fundamental differences between the conventional algorithm and D2P in terms of the treatment of the depression and river smoothening, two cases were added to isolate the effect of each change on the DEM from Case 1 to Case 4. Case 2 changes the river smoothening in Case 1 to the use of the adapted CRS with the uniform constraint of elevation decreasing downstream (Equation (3.6)). Case 3 uses the adapted CRS with uniform constraint of elevation decreasing downstream before filling. Case 3 is the closest to Case 4 in terms of the processed DEM except that select depressions are preserved during the adapted CRS step and the filling step in Case 4.

In summary, Cases 1 to 3 condition DEMs using different combination of depression treatment and stream smoothening methods without preserving any depression, while Case 4 incorporates select depressions. An example of a processed river segment from Case 0 to Case 4 is illustrated in Figure B.6.

Table 3.1: Summary of DEM Processing Methods used for the comparison.

	Description	Depression Treatment	River Smoothening	Depression Preserved
Case 0	**Raw DEM**	None	No	Yes
Case 1	**Conventional algorithm** (Condon et al., 2019)	Filling	Yes (Constant Slope)	No
Case 2	Case 1 + improved stream smoothing	Filling	Yes (Adapted CRS – uniform decreasing elevation downstream)	No
Case 3	Case 2 + reduced modification to DEM	Breaching/filling during river smoothening + post-filling	Yes (Adapted CRS - uniform decreasing elevation downstream)	No

Case 4	**Proposed algorithm (D2P)**	Selective filling of small-scale depressions + Breaching/filling of depressions within channel width and false depressions during river smoothening	Yes (Adapted CRS – uniform decreasing elevation downstream except at preserved depressions)	Yes (Select Depressions)

3.3 Results

3.3.1 Evaluation of identified depressions

To evaluate the accuracy of the delineated depressions by D2P algorithm, 1-m resolution aerial imagery from the National Agriculture Imagery Program (NAIP) (https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/) was used to manually delineate the ponds as a benchmark for comparison. The imagery was chosen from July 2009 to match the date of the LiDAR DEM used in this study. Alongside the DEM resampled to 10-m resolution, the comparison was also done at the original LiDAR DEM resolution of 1 m and two other resampled resolutions of 5 m and 20 m. The area and depth thresholds are set the same for all the resolutions to solely investigate the impact of DEM resolution on depression identification.

From Figure 3.6a-Figure 3.6d, the D2P algorithm was able to identify at least 85% of the delineated ponds from the NAIP imagery (56 out of 66) at the DEM resolution of 1 m, 5 m and 10 m but the performance dropped to 35% (23 out of 66) at the coarsest resolution of 20 m. This indicates the inability of the 20-m resolution DEM to resolve the scale and geometry of the depressions in the study area. Most of the NAIP ponds had an area of the same order of magnitude (~400 m^2 to 4,000 m^2) as the 20 m by 20 m grid cell so the depressions in the DEM could easily be lost during the resampling of the DEM. We compared the area of the depressions identified by D2P against the NAIP ponds and found a close agreement, with a coefficient of determination (R^2) of at least 0.75 across all the resolutions. In general, there is an overestimation in the area of the depressions delineated by D2P as shown in Figure 3.6e. This is because the D2P depression area is determined by filling the

depressions in the raw DEM up to the spill point elevation, which makes it the maximum possible area. The manually delineated areas from NAIP reflect the state of the pond at a specific time but may not be the maximum given the tendency to fluctuate depending on the climate. Moreover, vegetation and shadows in the NAIP imagery may mask the true boundary of the ponds during manual delineation and result in underestimation of the pond extent (Yasarer et al., 2018). Notably, the positive bias tends to increase with the DEM resolution because the DEM at higher resolution is better able to represent the boundary of the maximum depression area more closely. On the other hand, a coarser resolution DEM may fail to capture the irregular boundary of the depression and underestimate the area, as is the case for the 20 m DEM especially for the smaller ponds.

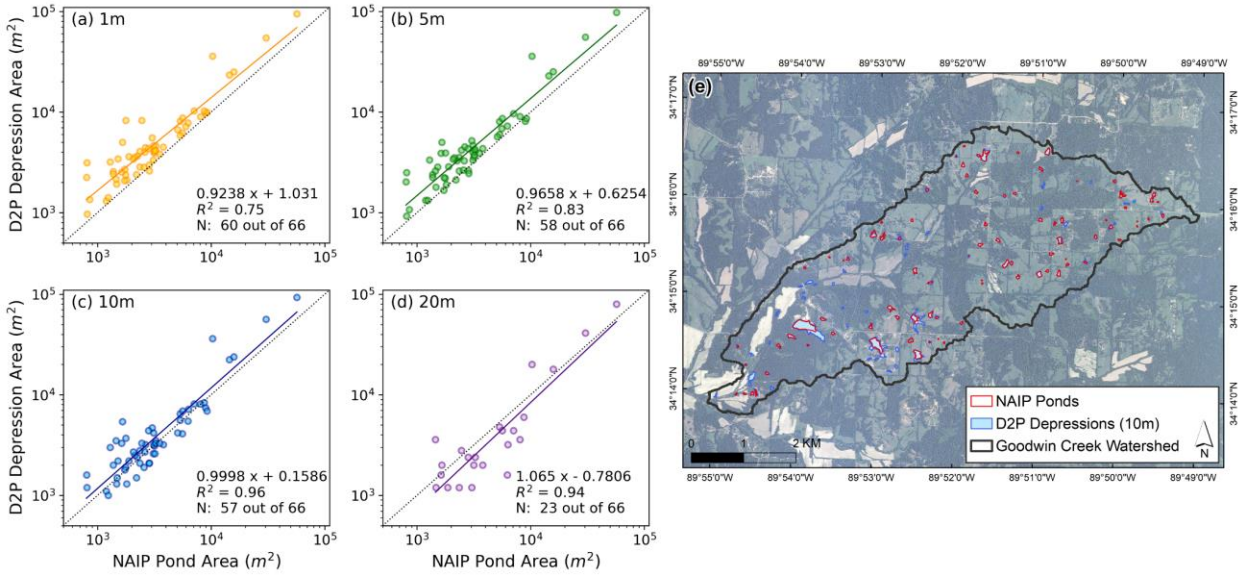


Figure 3.6: Scatter plot of area of depressions identified by D2P against that of the manually delineated ponds from NAIP imagery at DEM resolutions of (a) 1 m, (b) 5 m, (c) 10 m and (d) 20 m. (e) Comparison of spatial distribution of NAIP ponds and D2P depressions using DEM resolution of 10 m. N is the number of D2P depressions coinciding with NAIP ponds (out of a total of 66 ponds).

Another statistical analysis in the form of the binary pattern measures shown in Table 3.2 was conducted to further examine the degree to which the D2P depressions agree with the NAIP ponds in terms of location and extent. The results affirmed the performance of D2P as the *POD* indicated that more than 85% of the depressions matched the ponds for DEM resolutions 10m and lower. D2P produced false alarms across all four DEM resolutions, with

the false alarm rate (FAR) ranging from 44% to 53%. This is to be expected due to the tendency to overestimate the depression area as explained previously and the fact that not all depressions may result in the formation of a pond. In general, as the DEM resolution increases, the POD increases but this is offset by an increase in the FAR, so the resulting critical success index (CSI) is similar for the DEM resolutions of 1 m, 5 m and 10 m at about 50%. However, at the DEM resolution of 20 m, we observe that the POD drops drastically but the FAR does not decrease in tandem and this causes the CSI to decrease significantly to 32%.

Table 3.2: Statistical binary comparison of D2P depressions against rasterized NAIP ponds across the different DEM resolutions. The definition of POD, FAR and CSI can be found in Section B.1.3.

Resolution	POD	FAR	CSI
1 m	0.92	0.47	0.50
5 m	0.89	0.46	0.50
10 m	0.86	0.44	0.52
20 m	0.49	0.53	0.32

From the above comparisons, we have shown that D2P is able to reasonably identify depressions provided that the DEM resolution is able to resolve the scale of the depressions of interest. While D2P will perform better in capturing both the location and geometry of the known depressions from existing datasets (e.g., NWI, NAIP) with a higher resolution DEM, it also generates a higher number of unknown depressions, which can be hard to validate without ground truth data.

3.3.2 Effect of processing methods on DEM

The extent and magnitude of the impact to the raw DEM by the different methods decreased gradually from Case 1 to Case 4, as shown by the percentage of modified cells and mean absolute elevation offset in Table 3.3. This agrees with previous research (Lindsay, 2016a) which showed filling (Case 1 and Case 2) to have a bigger impact on the DEM than a hybrid of breaching and filling (Case 3 and Case 4). Between Case 4 and Case 1, D2P reduced the percentage of modified cells for the entire watershed from 11.3% to 5.58% and the mean absolute elevation offset by 5-fold from 0.0745 m to 0.0145 m compared to the conventional algorithm.

Focusing on the river network which was more likely to have changed due to the additional smoothing step, the elevation of practically all the cells in Case 1 were modified compared to only 30.6% of the cells in Case 4. In addition, the magnitude of modification was also smaller in Case 4, with a mean absolute elevation offset of 0.223 m compared to 0.648 m in Case 1. From the boxplot in Figure 3.7, there was a predominant increase in elevation in Case 1 and Case 2 but not Case 3 and 4 due to the nature of the filling algorithm. At the same time, the elevation offsets were highly variable in Case 1 (-1.5 m to +2 m) and Case 2 (-0.6 m to +1 m) compared to Case 3 (-0.4 m to +0.4 m) and Case 4 (-0.4 m to +0.4 m). Overall, the differences between the river profile from the processed DEM and the raw DEM were the highest in Case 1 followed by Case 2 and finally Cases and 3 and 4 which were similar.

Table 3.3: Statistics of modified cells and absolute elevation offset compared to the raw DEM based on all grid cells in the watershed for the four cases. The values in the bracket are the statistics for only the grid cells in the river network.

DEM Conditioning Method	Elevation	
	% Of Modified Cells	Mean Absolute Elevation Offset (m)
Case 1	11.3 (97.1)	0.0745 (0.648)
Case 2	8.40 (34.0)	0.0394 (0.328)
Case 3	6.35 (33.3)	0.0160 (0.240)
Case 4	5.58 (30.6)	0.0145 (0.223)

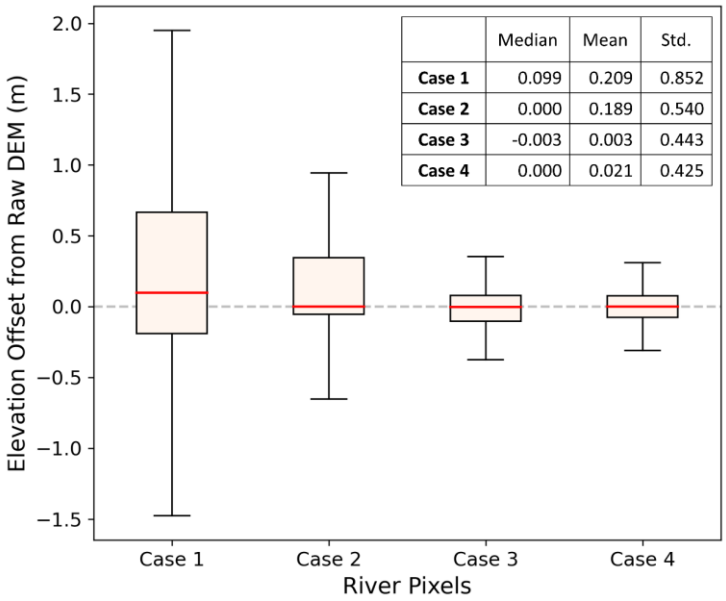


Figure 3.7: Boxplot of elevation offset from raw DEM of the grid cells in the river network for the four cases. The median is represented by the red line in the box. Outliers have been removed.

3.3.3 Effect of processing methods on hydrologic simulation

To understand the impact of the processing methods on hydrologic simulation, we compared the distribution of the simulated surface water depth and the time series of the surface water storage and streamflow. Figure 3.8 shows the instantaneous surface water depth distribution for two selected locations in the watershed with a high density of ponds. 10 hours into the simulation at the end of the heavy rainfall, the location and extent of the ponds in both Case 0 and Case 4 generally matched that of the ponds extracted from NAIP. However, artifactual depressions could be observed in Case 0 such as the one in the black circle in Location 2 as a result of using the raw DEM. This is a classic example of a false depression arising when a road crosses over the river. In Cases 1, 2 and 3, the depression removal process created flat terrain in the NAIP pond areas, resulting in minimal accumulation of water less than 0.1m in depth. Notably, Case 1 resulted in a larger flooded area compared to Cases 2 and 3 as the elevation of the river profile was raised gradually upstream to maintain a constant slope, thereby reducing the capacity of the channels. 30 hours into the simulation at the end of the recession, the residual water in flat areas in Cases 1 to 3 had already dried out. On the other hand, the depressions in Case 4 functioned as storage by trapping the water and formed isolated ponds. The result was not available for Case 0 as the model stopped running at the 11th hour due to the numerical instability from using the raw DEM. Overall, the DEM in Case 4 allowed ponding to be modeled realistically and the choice of DEM processing resulted in significant difference in the spatial distribution of ponding.

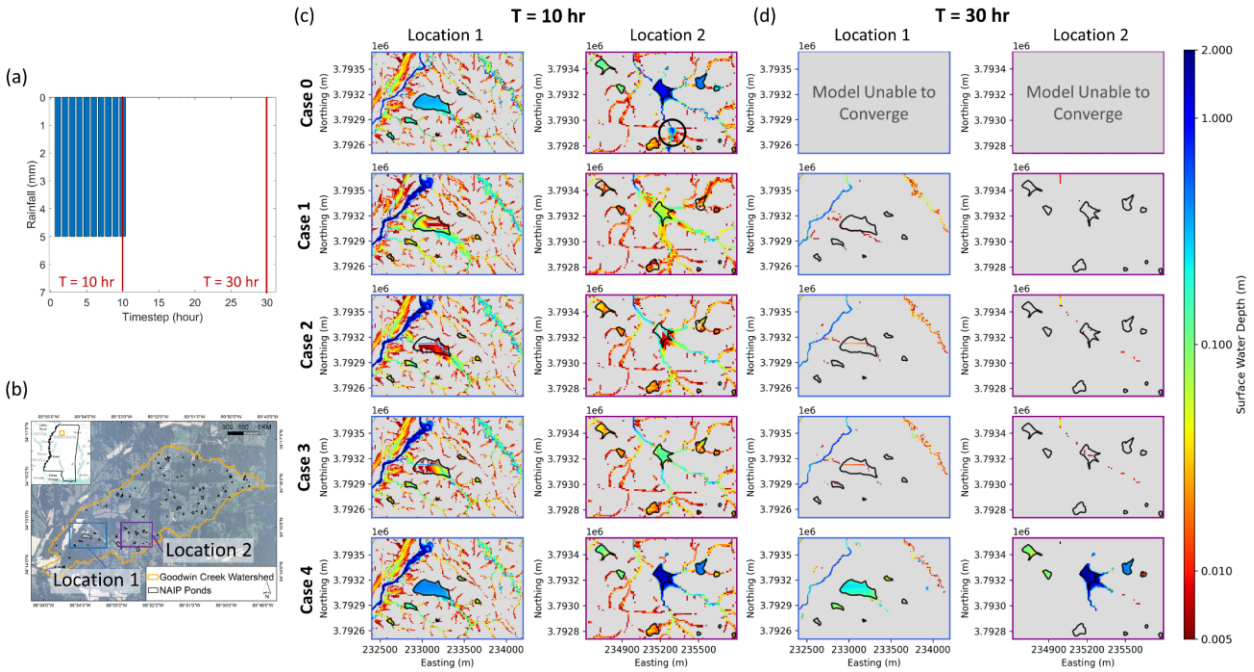


Figure 3.8: Example surface water depth maps for two-time instants at 2 locations for the simulations using the raw DEM (Case 0) and four DEM processing approaches (Case 1-4). (a) Rainfall hyetograph in ParFlow simulation (b) Distribution of benchmark depressions extracted from NAIP and Locations 1 and 2 (c) Surface water depth map 10 hours into the simulation (d) Surface water depth map 30 hours into the simulation.

Next, we analyzed the surface water storage time series for all cases. The total surface water storage (Figure 3.9a) was broken down into a river storage component (Figure 3.9b) and a non-river storage component (Figure 3.9c).

In Figure 3.9a, the total surface water storage was highest in Case 0 and Case 4 at the 10th hour. Subsequently, the total surface water storage in Case 0 remained the highest until the model terminated prematurely. This is because the drainage was poorest in Case 0 due to the lack of a smoothened river network that could direct water to flow between cell faces. The water from the rainfall became trapped in sinks found mostly along rivers and could barely reach the outlet of the watershed. Of the other cases, the total surface water storage was highest in Case 4 due to retention of water by the preserved depressions. Although the total surface water storage was similar across all depression-less simulations (i.e., Cases 1, 2 and 3), the choice of river smoothening mechanism and depression treatment resulted in a large difference in the distribution of storage between the river and non-river components. In Figure 3.9b, the river storage was lower in Cases 1 and 2 than Case 3. This can be attributed

to the shallower channel depth (See Table 3.4) from the use of filling to remove depressions in Cases 1 and 2. Conversely, the shallower channel depth makes the river more prone to overflowing and results in a higher non-river surface water storage in Case 1 and 2 compared to Case 3 (Figure 3.9c).

To examine the effect of preserving depressions on the hydrologic simulation, Case 4 was compared to Case 3. The river storage in Case 4 was lower (Figure 3.9b) because surface runoff was intercepted as non-river storage in surface depressions. After the rain had stopped, the non-river storage in Case 4 decreased much slower due to the retention effect of the depressions.

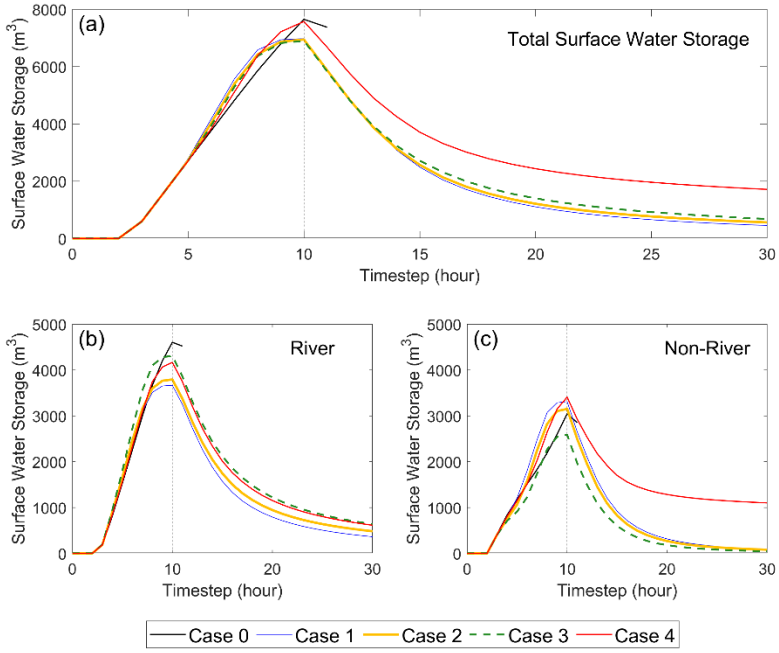


Figure 3.9: Time series of the surface water storage for the simulations using raw DEM (Case 0) and four DEM processing approaches (Case 1-4). (a) Total surface water storage (b) Surface water storage in river cells (c) Surface water storage in non-river cells.

Table 3.4: Statistics of channel slope and channel depth of GCEW calculated from the raw DEM (Case 0) and conditioned DEMs (Case 1 to 4).

DEM Conditioning Method	Channel Slope (-)		Channel Depth (m)	
	Mean	Std.	Mean	Std.
Case 0	-0.016	0.051	0.50	0.72
Case 1	-0.014	0.010	0.39	0.71
Case 2	-0.014	0.015	0.38	0.49

Case 3	-0.015	0.016	0.50	0.63
Case 4	-0.015	0.019	0.49	0.63

Figure 3.10 shows the hydrographs at the outlet of the watershed. Without any DEM processing, the peak flow was substantially lower in Case 0 as it was difficult for the water to flow past the rugged terrain to the outlet of the watershed. The streamflow took about the same time to peak for Cases 1, 2 and 3. However, the peak streamflow in Case 3 was visibly higher than Case 1 and Case 2 because of the higher flow depth at the outlet. Comparing Case 4 and Case 3, an attenuation of the peak streamflow was observed in the former due to the preserved depressions. The peak stream flow in Case 4 was reduced by 10% compared to Case 3 and the hydrograph for Case 4 lagged that of Case 3 by 0.4 hour.

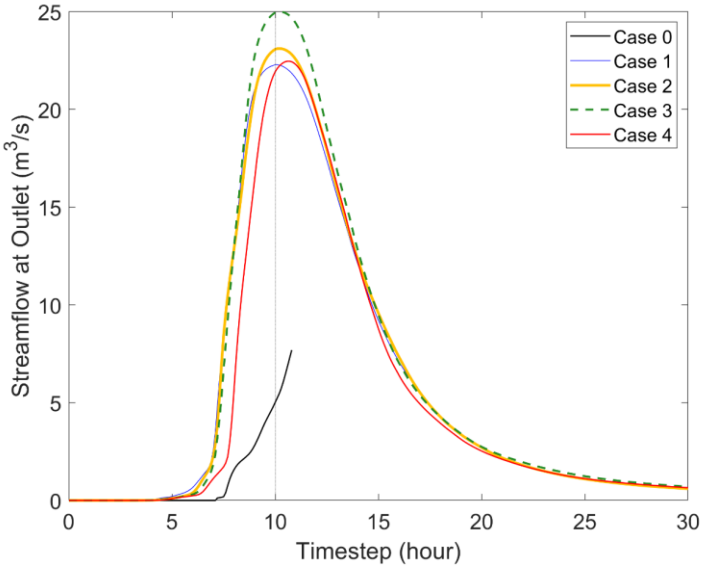


Figure 3.10: Hydrograph at the outlet of GCEW for the simulations using the raw DEM (Case 0) and four DEM processing approaches (Case 1-4).

3.4 Discussion

3.4.1 Threshold for removal of artifacts

For this study, to distinguish farm ponds from depressions too small to be significant to the hydrologic simulation, we set a minimum threshold depression depth of 0.0036 m and area of 900 m^2 based on NWI data. There is not a universal value that can be applied to all cases and the threshold should ideally be determined based on the scale of interest,

knowledge of the existing site, DEM resolution and computational efficiency of the hydrologic simulation.

For example, to investigate malaria vector habitat dynamics in Africa, the hydrologic model needs to consider smaller (~100 m^2) and shallower depressions (< 0.5 m) as the dominant malaria vectors prefer transient pools for breeding (Minakawa et al., 1999). On the other hand, the threshold will be very different for a reservoir simulation considering that the scale of the water body of interest is much bigger. Any existing site information will also be helpful in refining the thresholds. Guided by the Minnesota Karst Feature Database (KFDB), Wu et al., 2016 used a minimum depression area of 100 m^2 and minimum depression depth of 0.5 m to identify natural sinkholes. In a separate experiment conducted on the Prairie Pothole Region of North Dakota, a different threshold of a minimum depression size of 2,000 m^2 and minimum depression depth of 0.3 m was chosen based on spatial data from US Geological Survey (Wu et al., 2019).

Area and depth thresholds can also vary with DEM resolution. From the sensitivity analysis in Section B.1.2, the performance of the tested thresholds in the successful identification of wetlands from NWI data by D2P generally increased with DEM resolution. Higher resolution DEMs typically require larger area threshold as there is a higher probability of small artifactual depressions that need to be removed (Li et al., 2011). However, the same trend was not observed in our study. Instead, the optimal area threshold across the resolutions of 1 m, 5 m, 10 m were all around 800 m^2 to 900 m^2. This could be due to the quality and nature of the NWI data which is a record of wetlands and not all depressions in the study area. On the other hand, a lower depth threshold may be needed for lower resolution DEMs obtained by aggregating a higher resolution DEM. Aggregating a higher resolution DEM has the effect of averaging out the topographic variability, resulting in shallower depressions.

Another factor to consider is the trade-off between the realism of the topographic representation and computational efficiency. While setting a low threshold would allow the DEM to capture most of the depressions, it will also increase the variability in the topography which could inadvertently introduce slope discontinuities and develop oscillation and

instability issues for some models. For instance, setting the threshold area too low increases the chances of including deep yet small depressions that are anomalous and can greatly hurt the numerical performance.

Surface depressions within the channel width and depressions associated with false hydrologic barriers were removed based on the calculated RtD. The depressions were removed if their RtD exceeded the threshold that was calculated for known wetland locations from NWI. In general, the RtD threshold decreases as the resolution increases. The calculated RtD thresholds were 0.8, 0.54, 0.33 and 0.076 for DEM resolutions of 20 m, 10 m, 5 m and 1 m respectively. This is due to the overestimation of the areal representation of the river at lower resolutions. The performance of the RtD threshold in distinguishing false depressions associated with roads or bridges from water bodies that coincide with the river drops as the resolution decreases. This is due to the inability of the DEM to fully resolve the depression hierarchy at lower resolutions. At 20-m DEM resolution, the FAR (Table 3.2) was particularly high which can be attributed to the inclusion of false depressions.

3.4.2 Impact of filling vs hybrid algorithms on hydrologic simulation

Many studies have compared the influence of different DEM processing algorithms on the DEM in terms of elevation, slope and hydrological attributes (Callow et al., 2007; Lindsay, 2016a; Lindsay & Creed, 2005; Woodrow et al., 2016). In line with those studies, our comparison across the four cases shows that the filling based algorithms in Case 1 and Case 2 produced a greater impact on the DEM quantified by the number of modified cells and elevation offset in Table 3.3 and Figure 3.7. However, the impact of the DEM processing algorithms on the hydrological simulation is rarely discussed. From the rainfall recession simulation, we observed local differences in the pattern of the ponding between filling and hybrid algorithms, especially around false depressions arising from the crossing of a road over a river. In Case 1 and Case 2, filling raised the elevation of the upstream river valley to at least match the level of the road crossing, and this resulted in inundation over a large area upstream of the road crossing with relatively similar elevation. In extreme cases, the flat inundated area could create an obstruction to the incoming river and cause the flow to back up, increasing the computational demands. On the other hand, the hybrid algorithm in Case

3 breached the road crossing and allowed water to flow through the river channel relatively easily. In addition, both the filling and hybrid algorithms resulted in roughly similar mean channel slopes, but the former reduced the channel depth by 22% to 25% compared to the latter. This could cause overestimation of fluvial flooding as alluded to by the higher non-river storage in Figure 3.9c and potentially result in overly conservative flood management policies.

3.4.3 Impact of depression integration on hydrologic simulation

We showed in our rainfall recession simulation that the inclusion of depression modified the hydrologic response of the watershed. We observed an attenuation in streamflow at the watershed outlet and increase in lag time of the hydrograph as the depressions stored some of the rainfall and delayed the arrival of runoff to the rivers. This mechanism is widely acknowledged by other similar studies (Nasab & Chu, 2020; Yasarer et al., 2018; Zeng et al., 2020). At the same time, by explicitly representing the depressions in the DEM, we presented a viable means to capture the hydrologic connectivity between the depressions which is rarely achieved in the existing depression-integrated hydrologic models. The connectivity between ponds, headwater streams and downstream waters was evident in Case 4 at the end of the rainfall event in Figure 3.8c. Long after the rainfall event when the supply of water had stopped, the connecting streams that were ephemeral in nature gradually dried up, resulting in geographically isolated ponds in Figure 3.8d. This process was absent in Case 3 as there was no storage function with a depression-less DEM. By accounting for depression storage in model simulations, it is possible to improve the accuracy of flood hazard studies and allow surface depressions to be leveraged as part of a holistic approach in mitigating flood risk and preventing overpredictions. Beyond flood hazard studies, depression integrated hydrologic modeling has the potential to augment water resources management by enhancing the spatial delineation of water across the watershed and providing a more effective guide for local irrigation and crop rotation strategies (Rajib et al., 2020). Lastly, it is feasible to explicitly represent the connectivity between depressions and streams in studies to understand the effects of the transport of materials and biological fluxes within aquatic ecosystems (Leibowitz et al., 2019).

3.4.4 Limitation of the algorithm

There are three main limitations to D2P. Firstly, the artifactual riverine depressions were removed if the RtD was below the threshold, which requires knowledge of locations of wetlands in the study area especially those that intersect with the river. In our case, we derived the RtD threshold from NWI data, which is only available within the US. For future applications outside of the US, there is a need to check the local database for similar information or resort to other means to determine the RtD threshold such as through field survey. If the above is still not feasible, the threshold can be estimated iteratively by a sensitivity analysis.

Secondly, a D2P-processed DEM can result in higher computational demands in distributed hydrological modeling than depression-free DEMs. By comparing the number of solver iterations among the five cases in Figure 3.11, Case 4 generally had a higher computational demand than Cases 1 to 3. The number of iterations increased sharply near the 5^{th} time step where the water was just starting to fill the depression. This process was harder to solve in ParFlow at the point when the overland flow component was activated. From the 7th to 10th timestep, Case 1 required the highest number of iterations than Cases 2 to 4. One possible reason was that the flat areas created by the filling process in Case 1 obstructed the incoming flow from the river and caused it to back up, increasing the computational demands. After the rain stopped, the surface water flow in Cases 1 to 3 decreased and became gradually easier to solve compared to Case 4 where water was retained by the depressions. Notably, the number of iterations required in Case 0 was substantially higher than the other cases, indicating that some degree of DEM processing was necessary for modeling purpose.

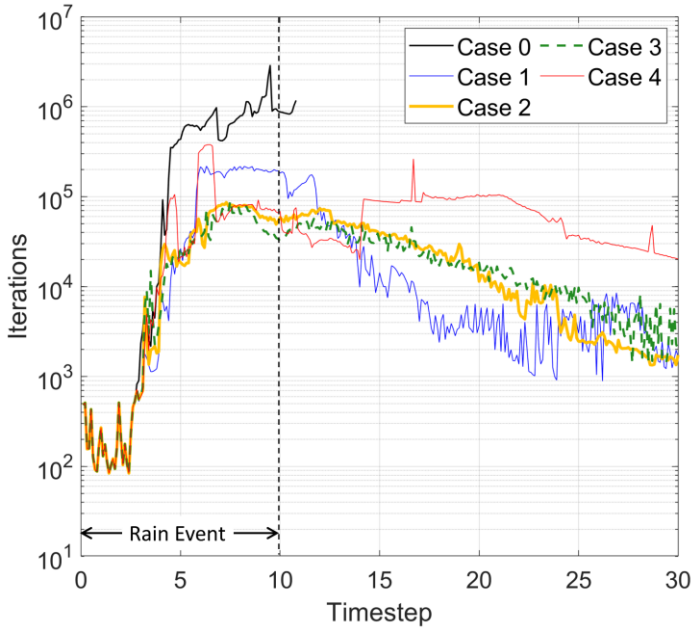


Figure 3.11: Comparison of number of solver iterations in ParFlow for each time step across the cases.

Finally, we note that hydro-conditioned DEMs produced by the D2P workflow are optimized for hydrologic routing methodologies that rely on topographic slopes, such as ParFlow. Consequently, D2P DEMs may not necessarily be optimal for hydrologic models that rely directly on topographic heights or in flood hazard simulation where the extent and depth of ponded areas are sensitive to certain topographic features along the river (Hodges, 2015; Sanders & Schubert, 2019). In Section 3.2.2.4 and 3.2.2.5, we removed all depressions within the river channel width as shown in Figure 3.3a as we could not distinguish artificial depressions from real features such as riffles and pools. However, the D2P workflow could be adapted to meet slightly different needs with respect to removing some features in the DEM and retaining others. For example, the RtD threshold can be adjusted higher to retain more stream depressions and exempt them from the enforcement of the downslope condition in the adapted CRS algorithm.

3.5 Summary and conclusions

Despite the growing emphasis on the impact of surface depressions on hydrologic processes, there is a lack of a formalized DEM processing tool that can resolve surface

depressions of interest for fully distributed hydrologic modeling. To address this gap, we present the Depression-Preserved DEM Processing (D2P) algorithm, a novel automated workflow that provides a representation of the land surface consistent with hydrologic processes being resolved while removing small-scale, non-physical variability in the DEM for computational efficiency. To achieve the above, the D2P algorithm includes several features, namely (1) improved delineation of surface depressions by screening depressions at variable instead of fixed intervals, (2) adoption of a new metric, RtD, leveraging depression hierarchy to filter out artifactual riverine depressions and (3) adaptation of a hybrid approach to smoothen the river in a way that enforces the general downslope profile while preserving surface depressions that intersect the river path.

The application of the proposed algorithm in a case study in GCEW minimizes the modification to the original DEM and the identified depressions match well with the delineated ponds from NAIP imagery. The performance of D2P in identifying depressions was evaluated over four different resolutions of 1 m, 5 m, 10 m and 20 m. In addition, the use of the D2P-processed DEM in a distributed hydrologic model also highlights the difference that the incorporation of depressions makes to the simulated ponding, water storage and streamflow. Following are the conclusions drawn from the study:

(1) For the selected study area in GCEW, the D2P algorithm was able to identify at least 85% of the delineated ponds from NAIP satellite imagery at the DEM resolution of 1 m, 5 m and 10 m but not at 20 m. The performance of D2P did not vary much as the DEM resolution decreased from 1 m to 10 m.

(2) Compared to a conventional DEM processing algorithm which removes all the depressions, the D2P algorithm has a significantly lower impact on the raw DEM by reducing the percentage of modified cells for the entire watershed by 51% and the mean absolute elevation offset by 81%.

(3) The hydrologic simulation for the rainfall recession test showcases the ability of D2P to capture the spatial distribution of physically meaningful depressions and their storage function. The depression-integrated simulation results in higher total

surface water storage as well as an attenuated and delayed peak streamflow due to retention of water by the preserved depressions.

Although LiDAR DEMs are generally of high accuracy and can provide a good representation of the landscape, the direct use of the raw DEM in hydrologic models using topographic slopes to route flow results in significant computational demands. The proposed D2P algorithm provides a systematic means to balance the need to process DEM for ease of computation and the preservation of surface depressions of interest. While we have only demonstrated the application of D2P for small water bodies in GCEW, the algorithm can also be used to process DEMs in the context of larger water bodies such as Prairie Potholes. The choice of hydrologic model would have to be tailored to the nature and geometry of the water bodies of interest.

Our case study of a test rainfall event clearly suggests that the treatment of surface depressions and the choice of method for smoothening topographic variability (e.g., filling and breaching) have a significant influence on the hydrologic simulation results. D2P adopts a minimum impact approach to avoid distorting terrain attributes unnecessarily while offering the user flexibility in altering depressions under different settings. Future work involves testing and evaluating the D2P algorithms across different landscapes. D2P facilitates depression-integrated studies using more fully distributed hydrologic models like ParFlow. This can provide new insights into complex hydrologic systems and help improve water resources management and environmental sustainability decisions.

Chapter 4

Investigating the Impact of Irrigation on Malaria Larval Habitats and Transmission using a Hydrology-based Model

4.1 Introduction

Malaria is a deadly disease induced by parasites that are transmitted to humans through the bites of infected female *Anopheles* mosquitoes. It is particularly acute in sub-Saharan Africa and remains one of the most pressing public health challenges in the region. 95% of malaria cases and 96% of deaths in 2020 were recorded in sub-Saharan Africa alone (World Health Organization, 2021). A combination of accelerated population growth and arid conditions worsened by climate change has inevitably created pressure on food security (Ward et al., 2016). This drives the development of several irrigation schemes across the region, which has been associated with increased malaria risk (Mangani et al., 2022). Most studies assess the relationship between irrigation and transmission based on field observations at a limited number of discrete points in space and time (Haileselassie et al., 2021; Kibret et al., 2014; Ondeto et al., 2022). It remains a challenge to understand the underlying mechanism and develop counter measures to mitigate the impact of irrigation.

Malaria modeling has been used over a century to understand transmission dynamics and guide intervention strategies (Ross, 1908). Originating from the basic Ross-Macdonald model, many compartmental models have inherited its simplifying assumptions such as homogeneous biting and well-mixing of hosts and vectors, which is a shortcoming of representing the vectors and hosts as a population group rather than individuals (Reiner et al., 2013). In cases where spatial heterogeneity and stochasticity are important such as in a low-transmission setting, agent-based models (ABMs) can provide explicit representation of individual actions and responses (N. R. Smith et al., 2018). Examples of advanced ABMs include EMOD (P. A. Eckhoff, 2011), OpenMalaria (T. Smith et al., 2006) and a model developed at Imperial College (Griffin et al., 2010). These models are widely used in malaria intervention studies (Galactionova et al., 2021) but they cannot be applied directly in regions

with irrigation as they typically rely only on rainfall as a source of water to quantify larval habitats (P. A. Eckhoff, 2011).

Essentially, ABMs tend to oversimplify the link between vector ecology and hydrology. The availability of water for breeding is affected by hydrological processes which are highly non-linear and spatially variable. In a hydrologic cycle, rainfall is partitioned into infiltration and surface runoff based on the soil type. The resulting surface runoff will accumulate or drain depending on the topography and surrounding vegetation. The persistence of the ponded water can also be influenced by evapotranspiration which varies with land use type. Besides rainfall, breeding sites can develop from groundwater, irrigation and around dams and reservoirs. In irrigated settings, crop production and distribution change seasonally so irrigation varies accordingly. Within a season, the application of irrigation also depends on soil saturation and crop water use. The spatiotemporal heterogeneity in irrigation results in habitats of different persistence and productivity. This diversity complicates the pattern of adult mosquito density and malaria transmission intensity (Frake et al., 2020), which warrants the incorporation of hydrology into ABMs.

In this study, we couple an integrated physical based hydrologic model, ParFlow-CLM (Ashby & Falgout, 1996a; J. E. Jones & Woodward, 2001; Kollet & Maxwell, 2006; Maxwell, 2013; Maxwell & Kollet, 2008), with EMOD for a test site in Ethiopia. EMOD was chosen because it is open source and can be easily modified to assimilate inputs from an external hydrologic model. Our aim is to demonstrate how incorporating hydrologic processes to estimate larval habitats can affect malaria transmission intensity and seasonality. Using the coupled model, we then examine the impact of an existing irrigation scheme on the spatiotemporal dynamics of malaria transmission.

4.2 Materials and methods

4.2.1 Study site

We conducted the study in the Arjo-Didessa sugarcane plantation and its surrounding area in the Oromia Region State, western Ethiopia (Figure 4.1). The study site includes a commercial sugarcane plantation of 5,000 hectares that rely on many migrant workers for

planting and harvesting. The plantation area will be expanded to 80,000 hectares when the irrigation channel construction is completed. The site elevation ranges from 1275 to 2105 m above sea level with a mean annual rainfall of 1,560 mm from 1994 to 2020 (Figure C.1). The rainy season is between May and October and the dry season occurs in the rest of the year. Monthly average relative humidity varies widely from around 40% to 80% and follows the rainfall pattern while monthly average temperature ranges from 19 °C to 24 °C and is lower in rainy season than dry season (Figure C.2). Malaria prevalence in this area is less than 3% and transmission is seasonal, with cases peaking between September to December (Hawaria et al., 2019). *Anopheles arabiensis* is the major vector in Arjo. The two main parasites are Plasmodium falciparum and Plasmodium vivax with the latter being the predominant species.

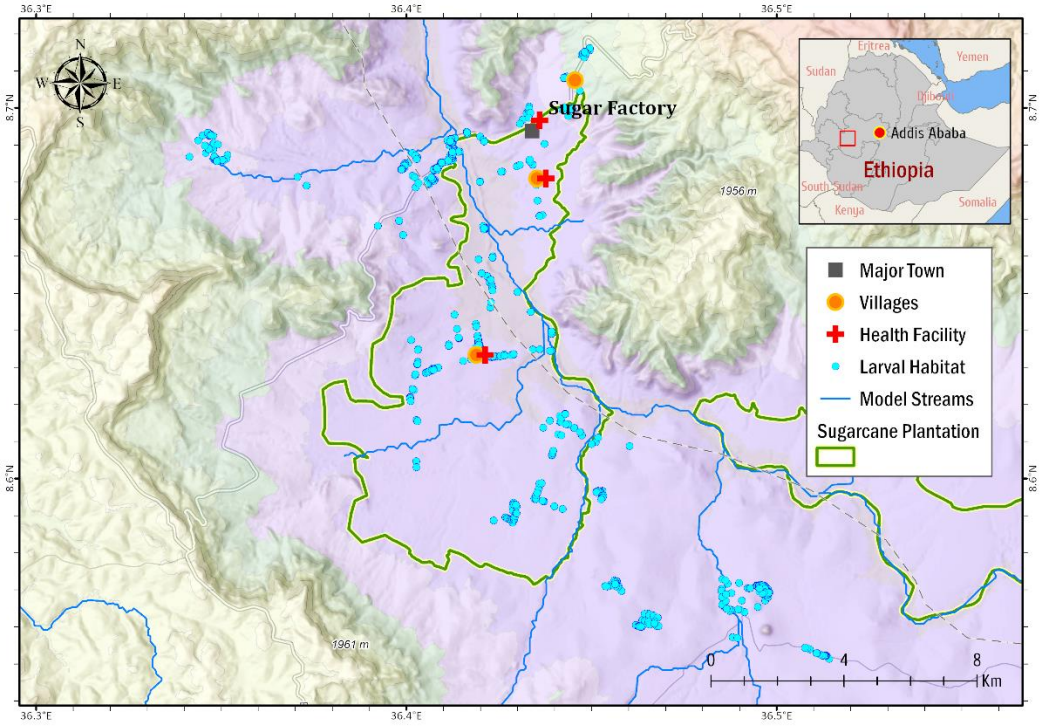


Figure 4.1: Study site and model domain. The hydrologic model domain demarcated in black contains part of the Arjo-Didessa sugarcane plantation. The EMOD model domain demarcated in purple only focuses on the irrigated farms. The surveyed larval habitats were used to estimate larval density and those within the hydrologic model boundary were used for hydrologic model calibration. The three health facilities are frequented by the plantation workers and provide clinical data for EMOD calibration.

4.2.2 Data collection

ParFlow-CLM requires climate data including rainfall, short-wave radiation, long-wave radiation, air temperature, surface pressure, specific humidity and wind speeds to drive the hydrologic processes. The data were obtained from Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System-Climate Data Record (PERSIANN-CCS-CDR) (Sadeghi et al., 2021) and the Fifth Generation European Centre for Medium-Range Weather Forecasts Reanalysis (ERA5) (Hersbach et al., 2018). Air temperature, humidity and long-wave radiation were adjusted to account for the mismatch in surface elevation between ERA5 and the study area, following the method by Cosgrove et al. (2003). For processes at the land surface, topography was obtained from a commercial global digital elevation model, ALOS World 3D (Takaku & Tadono, 2017), and land covers information was extracted from Global Land Cover Mapping Project which is based on Landsat and Chinese HJ-1 satellite images (Chen et al., 2015). The DEM was processed using the D2P algorithm from Chapter 2. To characterize the subsurface, soil properties were referenced from SoilGrids250m TAXOUSDA dataset (Hengl et al., 2017) for the top 2 m from the surface and the parameters for the deeper zone beyond the top 2 m were defined using GLHYMPS 2.0 (Gleeson et al., 2014). The bottom of the deeper zone was delineated based on the depth to bedrock data from SoilGrids250m BDRICM (Hengl et al., 2017). In EMOD, the input air temperature, land temperature and humidity were also obtained from ERA5 while rainfall was from PERSIANN-CCS-CDR. The list of model input data can be found in Table C.1.

Besides the above data, field survey was conducted. Larval habitat data were collected over several campaigns from 2017 to 2021, with a total of 769 sampled points. The survey mainly targeted villages and the surveyed larval habitat locations are shown in Figure 4.1. Mosquito larvae were sampled following the WHO standard larval survey procedure using a standard dipper (350 ml). Larvae were identified morphologically and sorted by genus as *Anopheles* or *Culex* in the field. Larval density, habitat type and habitat dimension were recorded. Habitat locations within the hydrologic modeling domain were used to validate the predicted aquatic habitats simulated in ParFlow-CLM. The measured larval density was used to determine the larval carrying capacity of the habitats in EMOD. To

validate EMOD, population data, malaria prevalence rates and clinical cases were obtained from local health facilities (Figure 4.1).

4.2.3 Model approach

4.2.3.1 Model background

We used ParFlow-CLM, a process-based gridded model, to simulate the hydrologic processes in the Arjo study site in Ethiopia. ParFlow solves the variably saturated subsurface flow and overland flow while CLM calculates the canopy water balance and terrestrial energy balance influenced by land cover characteristics. Due to its ability to simulate complex surface-subsurface interactions, ParFlow-CLM can resolve a diverse range of water bodies driven by heterogenous hydrological and geomorphological processes which result in different breeding habitats such as rain-fed pools, flood basins and spring-fed ponds (M. W. Smith et al., 2013).

EMOD was used to simulate malaria transmission in the study area. The modeled region in EMOD can be represented as a single node or divided into multiple nodes. As a stochastic ABM, it simulates the simultaneous interactions between humans and mosquitos within each node, using decision rules based on individual agent properties and with randomness (Bill & Melinda Gates Foundation, 2022b). The properties are defined by user inputs on demographic, climate, mosquito, parasite and intervention parameters. The model simulates vector population dynamics (e.g. vector life cycle, vector survival and feeding), human population dynamics, human immunity, within-host parasite dynamics as well as effects of interventions such as antimalarial drugs and vaccines (P. Eckhoff, 2013).

4.2.3.2 Linking habitat representation in EMOD with ParFlow-CLM

In EMOD, natural larval habitats commonly comprise temporary, semi-permanent, permanent (constant) habitats and each habitat type is calculated based on a different equation (P. A. Eckhoff, 2011). Temporary habitats are driven mainly by rainfall and decays at a rate proportional to the evaporation rate which is a function of temperature and humidity. The area of temporary habitats in each node at time t, H_{temp}^{t}, is calculated by:

$$H_{temp}^{t} = H_{temp}^{t-1} + \lambda_{temp} P^{t} {D_{cell}}^{2} - H_{temp}^{t-1} \tau_{temp}^{t} \Delta t \quad (4.1)$$

and

$$\tau_{temp}^{t} = 5.1 \times 10^{11} e^{\frac{-5628.1}{T^t}} k_{temp} \sqrt{\frac{0.018}{2\pi R T^t}} (1 - RH^t), \quad (4.2)$$

where λ_{temp} is a scaling factor, P^t is rainfall at time t, τ_{temp}^{t} is a decay rate at time t, D_{cell} is the grid size in degree, Δt is the time interval, T^t is temperature in Kelvin at time t, k_{temp} is a decay factor, R is the universal gas constant 8.314 J/mol/K and RH^t is the relative humidity at time t.

Similar to temporary habitats, semi-permanent habitats are also driven by rainfall but the decay rate is a constant independent of temperature and humidity. Semi-permanent habitats are configured to decay slower than temporary habitats. Using a scaling factor λ_{semi} and a decay rate τ_{semi}, the area of semi-permanent habitats H_{semi}^{t} is calculated as:

$$H_{semi}^{t} = H_{semi}^{t-1} + \lambda_{semi} P^t {D_{cell}}^2 - H_{semi}^{t-1} \tau_{semi} \Delta t, \quad (4.3)$$

Lastly, permanent habitats are assumed to be independent of rainfall, temperature and humidity. The area, H_{perm}^{t}, remains the same over time and is determined by a constant λ_{perm}:

$$H_{perm}^{t} = \lambda_{perm} {D_{cell}}^2, \quad (4.4)$$

We incorporated hydrology into EMOD by replacing the habitat calculation with the output from ParFlow-CLM. In ParFlow-CLM, we simulated the spatiotemporal distribution of surface soil saturation, which was used to determine the availability of surface water that could contribute to ponding. At each grid cell, ponding is assumed to occur if the soil saturation exceeds a threshold, θ, calibrated based on field larval habitat observations. Each cell is classified into temporary (15-90 days), semi-permanent (90-180 days) or permanent habitat (more than 180 days) depending on the duration of ponding. Rivers with high flowrates were not considered since *Anopheles* larvae have a lower chance of surviving in fast-moving water (A. J. Hardy et al., 2013). Details of the concept of the hydrologic simulation and larval habitat identification can be referenced in Jiang et al. (2021). For each time step, the fraction of the study area covered by each habitat type (i.e., F_{temp}^{t}, F_{semi}^{t}, and

F_{perm}^{t}) was calculated and input into EMOD. Finally, the area for each habitat type was obtained after multiplying the fractional area coverage by the nodal area as follows:

$$H_{temp}^{t} = F_{temp}^{t} {D_{cell}}^{2}, \tag{4.5}$$

$$H_{semi}^{t} = F_{semi}^{t} {D_{cell}}^{2}, \tag{4.6}$$

and

$$H_{perm}^{t} = F_{perm}^{t} {D_{cell}}^{2}, \tag{4.7}$$

where F_{temp}^{t}, F_{semi}^{t}, F_{perm}^{t} are the fractional area coverage of temporary, semi-permanent and permanent habitats, respectively.

4.2.3.3 Habitat larval capacity

EMOD requires the user to define a larval capacity per unit area (LC) for each habitat type, which represents the hypothetical number of larvae that can co-exist within a 1 degree by 1 degree habitat area. The habitat larval capacity per unit area is multiplied by the variation in habitat area derived from the previous section to determine the habitat larval capacity time series in the model domain.

LC was estimated using field survey data for each habitat type (Section C.1.1). In Table 4.1, LD_{dip} represents the larval density in number of larvae per dip. We then converted LD_{dip} to an equivalent number of larvae per unit degree squared (LD) based on the opening area of the standard dipper which is 13 cm in diameter. To obtain LC, we adjusted LD by a scaling factor (s) during the calibration of EMOD. The adjustment is necessary because using LD directly will overestimate the larval capacity as surveyors tend to dip at locations with a higher density of larvae within a sampled habitat.

Table 4.1: Laval density derived from field survey and calibrated larval capacity per unit area for each habitat type.

Habitat Type	Larval Density		Larval Capacity per Unit Area $LC = s \times LD$ (#/degree^2)
	LD_{dip} (#/dip)	LD (#/degree^2)	
Temporary	0.167	1.97×10^{11}	2.28×10^{7}

Semi-Permanent	0.089	1.05×10^{11}	1.21×10^{7}
Permanent	0.440	5.18×10^{11}	6.00×10^{7}

In summary, we identified potential larval habitats in ParFlow-CLM and classified them into temporary, semi-permanent and permanent habitats as an input to the vector cycle simulation in EMOD. The overall schematic of our modeling approach is shown in Figure 4.2.

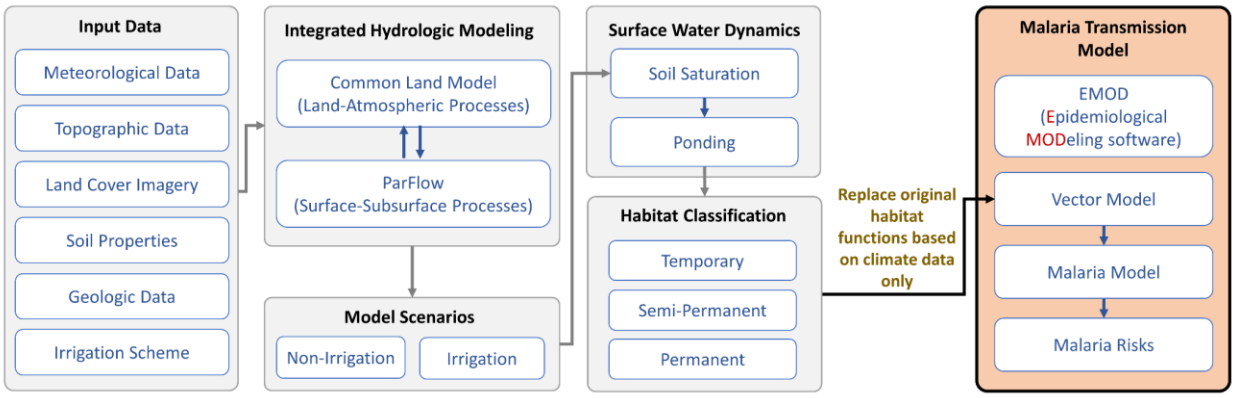


Figure 4.2. Framework for integrating ParFlow-CLM with EMOD.

4.2.4 Model configuration

4.2.4.1 Model domain

In a previous study, ParFlow-CLM was successfully applied in Arjo for larval habitat identification over a one-year period through 2018 (Jiang et al., 2021). Here, we expanded the hydrologic simulation to a 20-year period from 2000 to 2020. The model domain area was 208 km2 (Figure 4.1). To keep the computational time manageable, we decreased the spatial resolution from 50 m to 100 m and reduced the number of subsurface layers from 10 layers to 5 layers. The layer thickness from ground surface to bedrock was 0.1 m, 0.3 m, 0.6 m, 1 m and 78 m respectively, over a total vertical depth of 100 m.

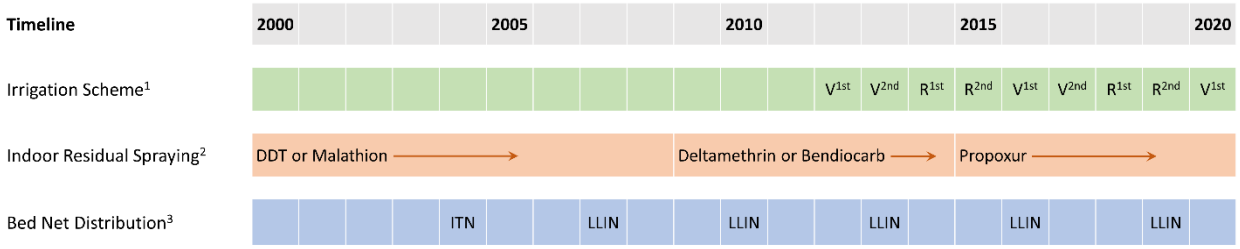
For malaria transmission modeling in EMOD, we focused on the farmland areas within the hydrologic modeling domain (Figure 4.1). The EMOD domain was 10 km by 10 km and configured as a single node.

4.2.4.2 Model scenarios

Three model scenarios were developed for this study. In the first scenario, herein referred to as "Default EMOD", malaria transmission was simulated in EMOD based on the default larval habitat equations (Equations (4.1) to (4.4)) which are functions of rainfall, temperature and relative humidity. The second scenario, "Integrated EMOD-NonIrr", replaced the default habitat equations in EMOD with the habitat simulation from ParFlow-CLM using Equations (4.5) to (4.7). No irrigation was applied in this scenario. The third scenario, "Integrated EMOD-Irr", is similar to "Integrated EMOD-NonIrr" but irrigation was applied.

Irrigation started in 2012 and a 4-year sugarcane planting cycle was adopted in the model as shown in Figure 4.3. The cycle includes a 2-year cycle for virgin and two 1-year cycles for ratoons. The irrigation scheme was designed based on the planting cycle and the months with irrigation are shown in Figure C.4. During the irrigation season, irrigation was applied every 10 days for the first 3 days. Each day, 5.3 mm/hour of sprinkler water was applied for 22 hours. The derivation of the irrigation rate can be found in Section C.1.2.

In the three scenarios, malaria intervention including indoor residual spraying and the use of bed nets was modeled per the schedule shown in Figure 4.3 based on communication with the local district health office.



[1] V^{1st} and V^{2nd}: First and second year of virgin planting; R^{1st} and R^{2nd}: First and second year of ratoon planting.
[2] DDT: Dichlorodiphenyltrichloroethane.
[3] ITN: Insecticide-treated net (permethrin); LLIN: Long-lasting insecticidal net.

Figure 4.3: Configuration of intervention and irrigation in the coupled model.

As we do not have information on the total area of the habitats in the study area and to allow a fair comparison, the scaling factors (λ_{temp}, λ_{semi} and λ_{perm}) in the "Default EMOD" scenario were set individually such that the average area of each habitat type over the

simulation period was aligned with that of the "Integrated EMOD-NonIrr" scenario. In addition, the decay parameters k_{temp} and τ_{semi} in the "Default EMOD" scenario were also adjusted to match the variability of the "Integrated EMOD-NonIrr" scenario (Figure C.5).

4.2.4.3 Model calibration

We chose the "Integrated EMOD-Irr" scenario to calibrate ParFlow-CLM and EMOD and used the same calibrated parameters for the "Default EMOD" and "Integrated EMOD-NonIrr" scenarios. This is because observed data was only available for the period after irrigation was implemented in the study area. This also prevented the effect of incorporating hydrology or irrigation from being obscured by using different parameters in each scenario.

For ParFlow-CLM, as the spatial resolution was modified from the previous study, we recalibrated the saturation threshold, θ. The calibration was to ensure that the model will predict the occurrence of ponding at locations in line with the field surveyed larval habitats for soil saturation above the selected threshold (Section C.1.3) (Jiang et al., 2021). In EMOD, we identified 10 key parameters (Table C.2) for calibration after a sensitivity analysis. The calibration was performed such that the simulated prevalence rate and pattern of clinical cases could match local data (Section C.1.3). The rest of the parameters were either set based on default values in EMOD or referenced from Gerardin et al. (2015) and Selvaraj et al. (2018).

4.2.4.4 Spatial realization of transmission through heterogeneity of habitats

To demonstrate how the heterogeneity of habitats can affect malaria transmission, we conducted an additional analysis by discretizing the EMOD domain into a 10 by 10 grid. Each grid cell was 1 km by 1 km and modeled as an individual node in EMOD. All the nodes were assigned the same calibrated parameters from the single node model but the input habitat was specific to the area covered by each grid cell. This was only applicable to the "Integrated EMOD-NonIrr" and "Integrated EMOD-Irr" scenarios. The habitat simulation in the "Default EMOD" was unable to reflect spatial heterogeneity due to the scale of the climate data used to derive the habitats.

4.3 Results

4.3.1 Effect of hydrology on larval habitats and malaria transmission

Figure 4.4 shows a comparison of the larval habitat area as a fraction of the study area between the "Default EMOD" scenario and the "Integrated EMOD-NonIrr" scenario. In both scenarios, the variation in the total larval habitat area was in tandem with rainfall and largely seasonal (Figure 4.4a). The mean total larval habitat area was 0.27 of the study area in both scenarios but the "Integrated EMOD-NonIrr" scenario exhibited a wider range. The use of the default habitat equations in EMOD caused high frequency fluctuations that are a direct result of rainfall. In contrast, the larval habitats derived from the hydrologic model do not exhibit such quick response to rainfall. Overall, the habitats in the "Integrated EMOD-NonIrr" scenario have a larger intra-annual variability (Figure 4.4b) and inter-annual variability (Figure 4.4c). This is due to the fundamental difference in the underlying equations leading to the derivation of habitats between the two scenarios, which will be discussed further in Section 4.4.1.

Figure 4.4d, Figure 4.4e and Figure 4.4f show the breakdown for each habitat type. Of the total larval habitat area, permanent habitats were the most dominant, accounting for 0.16 of the study area, followed by semi-permanent habitats (0.084) and temporary habitats (0.033). In the "Integrated EMOD-NonIrr" scenario, the fluctuation in habitat area gradually becomes smoother from temporary, to semi-permanent and finally permanent habitats which corresponds with the increasing stability of the habitats. In each year, the distribution between the habitat types can vary significantly depending on the magnitude and duration of rainfall in that year. For example, there were more temporary habitats relative to semi-permanent habitats in 2012, and vice-versa in 2015, due to a difference in rainfall pattern. In contrast, the difference in stability and dynamic distribution between temporary and semi-permanent habitats was less apparent in the "Default EMOD" scenario. Notably, the area of permanent habitats remained constant throughout the years, a key difference from the "Integrated EMOD-NonIrr" scenario, in which permanent habitats were defined as habitats with more than 180 days of ponding and subject to temporal variations.

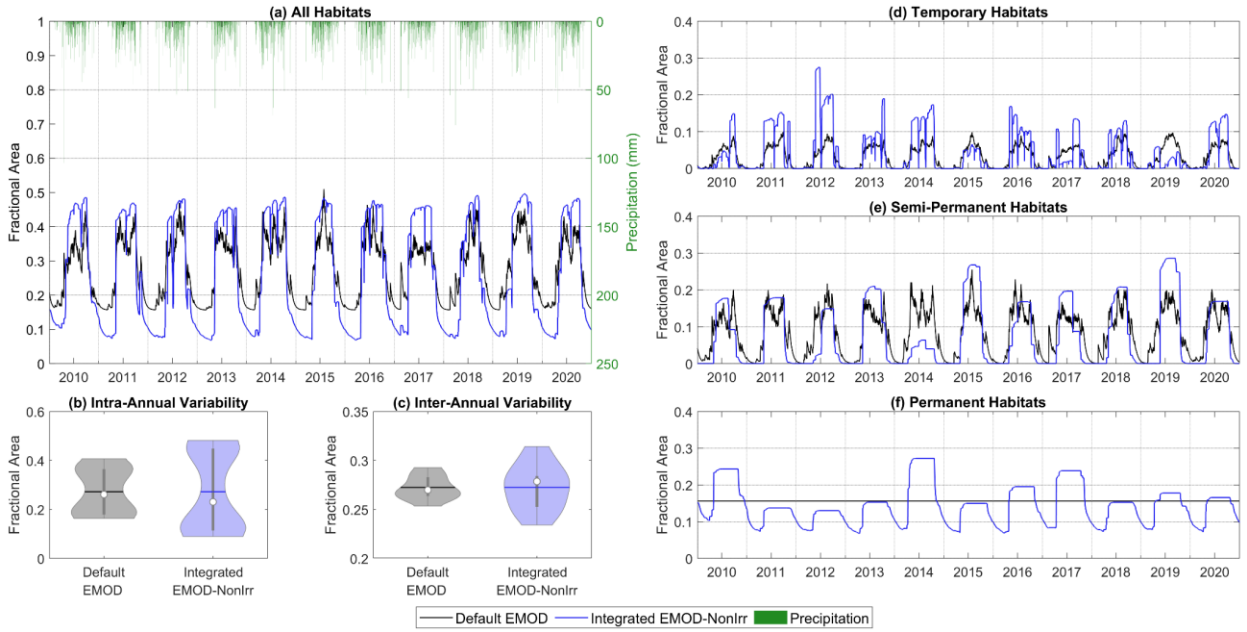


Figure 4.4: Comparison of daily simulated larval habitat area between "Default EMOD" scenario and "Integrated EMOD-NonIrr" scenario from 2010 to 2020. (a) Total habitat area is broken down into (d) temporary habitats (e) semi-permanent habitats and (f) permanent habitats. The total habitat area was used to plot (b) intra-annual variability and (c) inter-annual variability. Intra-annual variability represents the distribution of 20-year average habitat area for each day of a year. Inter-annual variability represents the distribution of annual average habitat area for each year. The simulation was performed for a 20-year period from 2000 to 2020 but here we only show the results from 2010 to 2020 for simplicity.

Figure 4.5 compares the annual average results of habitat larval capacity (Figure 4.6a), total adult vectors (Figure 4.6b), fraction of vector infected (Figure 4.6c) and parasite prevalence rate (Figure 4.6d) between the "Default EMOD" scenario and the "Integrated EMOD-NonIrr" scenario. The average habitat larval capacity was identical in both scenarios since the average total larval habitat area was the same. However, the number of adult vectors was slightly higher (by 3%) in the "Default EMOD" scenario. The resulting difference was further amplified to 5.3 times for average fraction of vectors infected and 4.7 times for average prevalence rate. Given that all other input data and parameters in EMOD were the same, this can be traced back to the discrepancy in temporal pattern of the daily habitat larval capacity (Figure C.9c), with a visibly lower variability in the "Default EMOD" scenario providing a stable environment for the vector to thrive in throughout the year. The results suggest the temporal variation in larval habitat from incorporating hydrology can significantly alter malaria transmission dynamics.

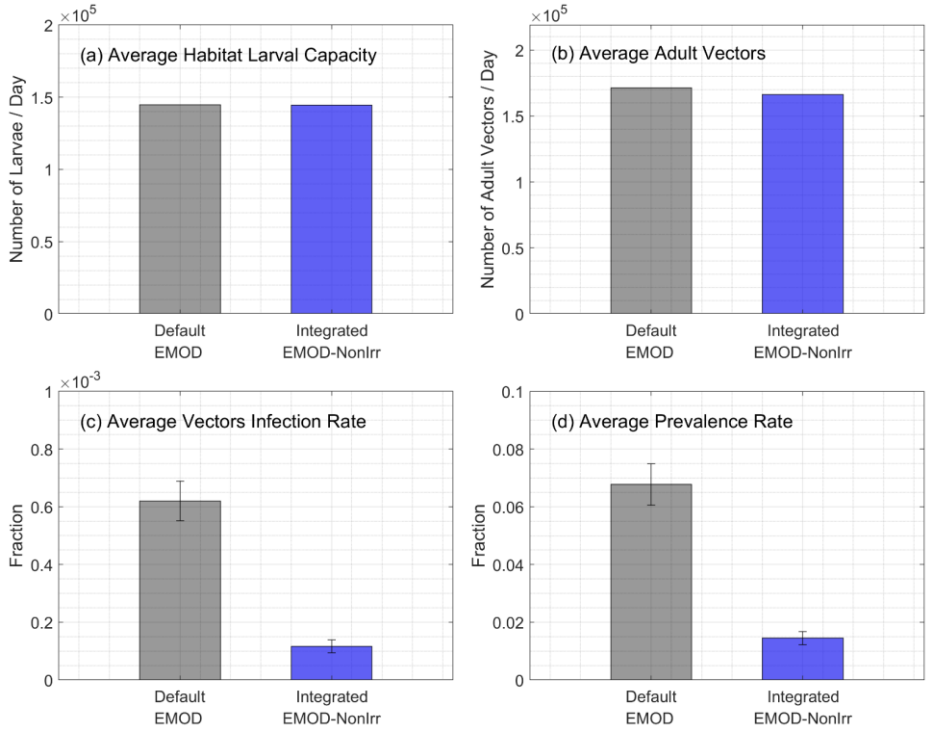


Figure 4.5: Comparison of annual average malaria transmission indictors between "Default EMOD" scenario and "Integrated EMOD-NonIrr" scenario from 2000 to 2020. The indicators include (a) habitat larval capacity (b) adult vector abundance (c) adult vector infection rate and (d) parasite prevalence rate.

4.3.2 Effect of irrigation on larval habitats and malaria transmission

Figure 4.6 compares the dry season and rainy season larval habitat area for each habitat type between the "Integrated EMOD-NonIrr" scenario and the "Integrated EMOD-Irr" scenario over the irrigated period from 2012 to 2020. Overall, irrigation increased the habitat area in both seasons. In the dry season (Figure 4.6a), irrigation increased the larval habitat area across all the habitat types. Although irrigation was not applied in the rainy season, irrigation before and after the rainy season prolonged the stability of temporary and semi-permanent habitats and converted them to permanent habitats (Figure 4.6b). The results indicate that irrigation may enable the development of vectors in the dry season while stabilizing the growth in rainy seasons.

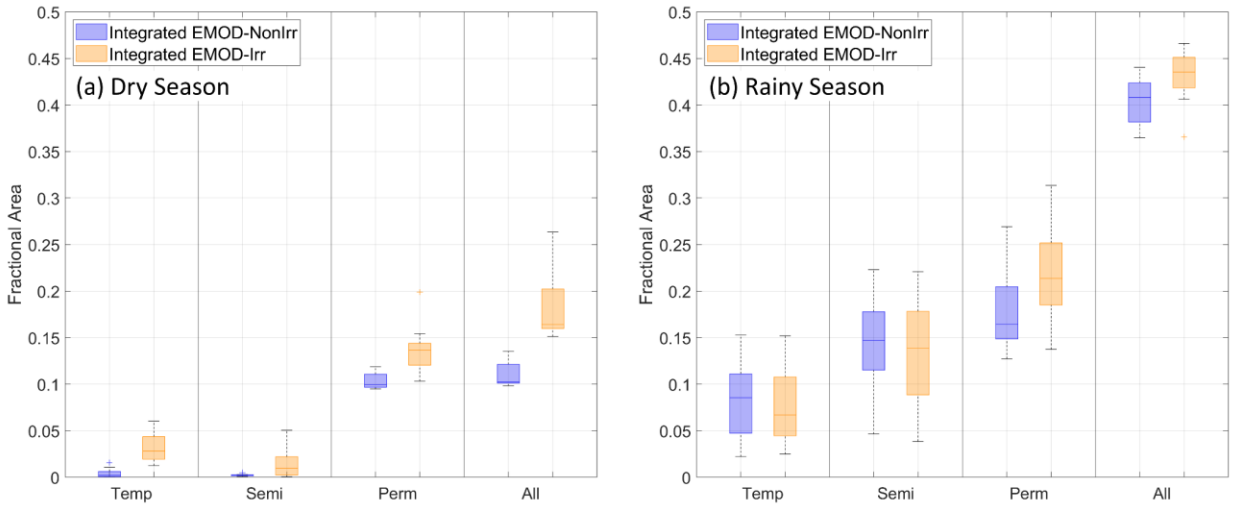


Figure 4.6: Comparison of larval habitat area during (a) dry season and (b) rainy season between the "Integrated EMOD-NonIrr" scenario and the "Integrated EMOD-Irr" scenario. Habitats are further classified into temporary, semi-permanent and permanent types. The simulation was performed for a 20-year period from 2000 to 2020 but here we only show the results from 2012 onwards when irrigation began.

Figure 4.7 compares the daily times series of habitat area (Figure 4.7b), larval capacity (Figure 4.7c), adult vector population (Figure 4.7d), entomological inoculation rate (EIR) (Figure 4.7e) and parasite prevalence (Figure 4.7f) between the "Integrated EMOD-NonIrr" scenario and the "Integrated EMOD-Irr" scenario over the irrigated period from 2012 to 2020. The differences in the larval density for each habitat type introduce more variability to larval capacity as the relative abundance of each habitat type is dynamic. The pattern of the adult vector population generally follows that of habitat larval capacity. However, the EIR cycle lags adult vector population cycle by 2 months and the parasite prevalence cycle lags the EIR cycle by another 1 month.

The increase in habitat area arising from the applied irrigation contributed to an increase in adult vector population beyond the irrigation periods as well as EIR and parasite prevalence. The simulated daily EIR hit a maximum of 0.0061 in September 2017, a 4.7-fold increase compared to the "Integrated EMOD-NonIrr" scenario. In the same year, the maximum prevalence occurred around October with a 3-fold increase to 0.033. It was also found that the EIR and prevalence peaks, which typically occurred around

October/November and November/December respectively, were both shifted forward by about one month after irrigation was applied.

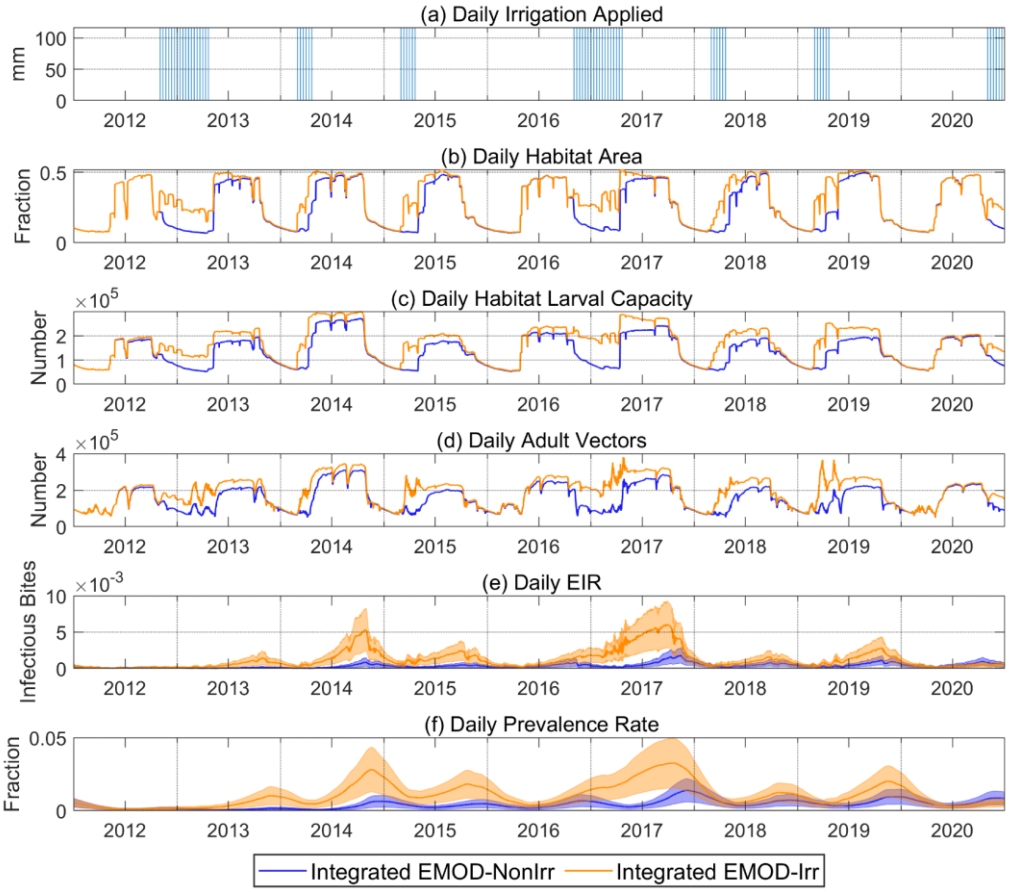


Figure 4.7: Time series of daily applied irrigation in the "Integrated EMOD-Irr" scenario and comparison of simulated daily malaria transmission results between "Integrated EMOD-NonIrr" scenario and "Integrated EMOD-Irr" scenario. Malaria transmission results include (b) habitat area (c) habitat larval capacity (d) adult vector abundance (e) entomological inoculation rate and (f) parasite prevalence rate. The simulation was performed for a 20-year period from 2000 to 2020 but here we only show the results from 2012 onwards when irrigation began.

4.3.3 Spatial variation of malaria transmission

Figure 4.8 compares the spatial distribution of larval habitats, adult vector abundance and parasite prevalence between the "Integrated EMOD-NonIrr" scenario and the "Integrated EMOD-Irr" scenario for the dry season from November 2016 to April 2017 and rainy season from May 2017 to October 2017. This period was selected because the effect of irrigation was the most pronounced (Figure 4.7). Overall, larval habitats formed more easily

in the southwestern region which is characterized by clay-rich soil with low permeability (Figure C.10). Besides soil type, the distribution of the habitat types within the study area varied substantially with hydrologic processes depending on the local topography, land use and irrigation. In both the dry and rainy seasons, irrigation expanded the area covered by habitats and increased the stability of existing habitats from the "Integrated EMOD-NonIrr" scenario.

Like habitat area, the adult vector hotspots in both seasons were enlarged and intensified by irrigation. The increase in vector population was larger in the dry season, due to the creation of more habitats in the northern region by irrigation. The adult vector hotspots were mostly concentrated around permanent habitats, which were configured with the highest larval capacity based on field data.

In general, irrigation increases the difference in prevalence rate between the rainy season and dry season. Without irrigation, the parasite prevalence cycle peaked in November and December of the dry season, lagging the peak adult vector population from the preceding rainy season by about three months. Hence, the average prevalence in the rainy season was not necessarily higher than the dry season. In 2017, irrigation was applied from November 2016 to April 2017, connecting two rainy seasons and creating favorable conditions for breeding over more than a year. As a result, the prevalence in the second rainy season was visibly higher than the preceding dry season (also see Table C.3). This reiterates that the effect of irrigation was not just limited to the dry season and attention should also be given to the rainy season.

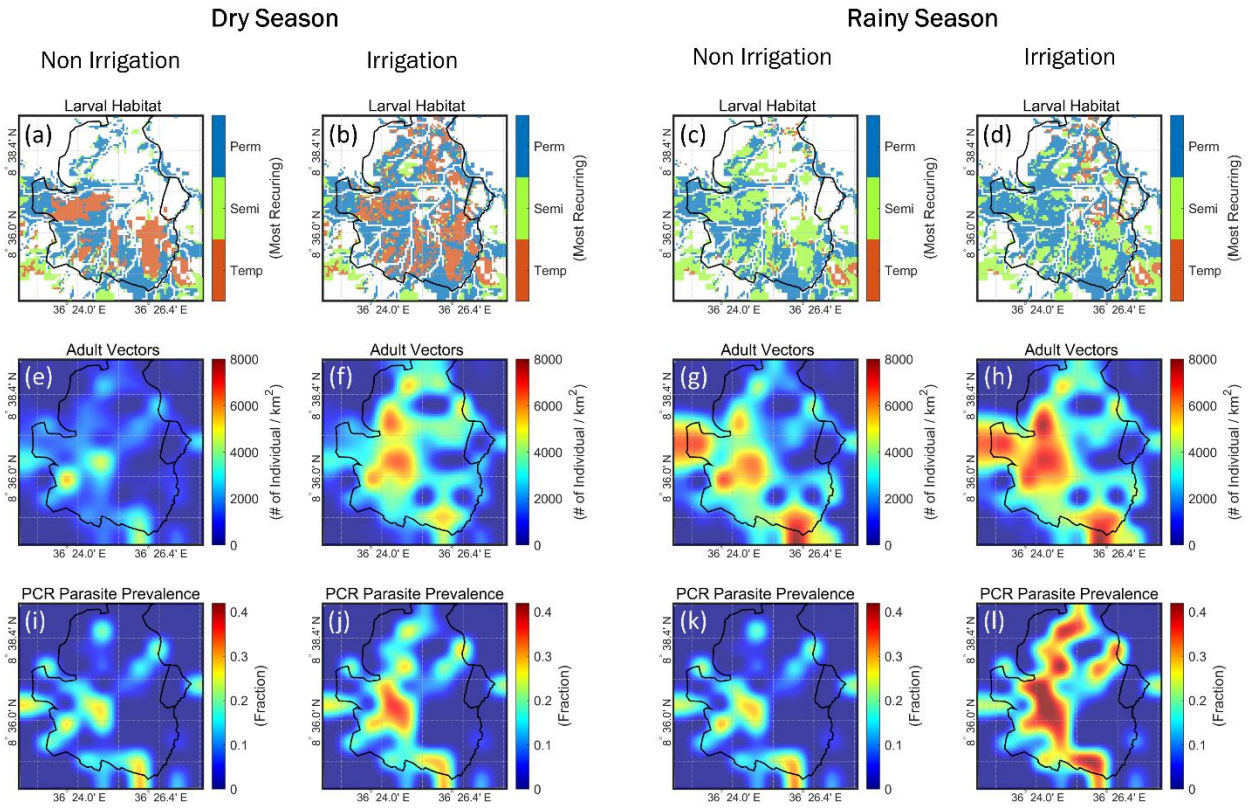


Figure 4.8: Spatial distribution of daily average larval habitats, adult vectors and PCR parasite prevalence in the dry season (November 2016-April 2017) and rainy season (May 2017-October 2017). The simulated larval habitats, adult vectors and PCR parasite prevalence from the "Integrated EMOD-NonIrr" scenario are presented in (a), (e) and (i) for the dry season and (c), (g) and (k) for the rainy season. Similarly, the simulated larval habitats, adult vectors and PCR parasite prevalence from the "Integrated EMOD-Irr" scenario are presented in (b), (f) and (j) for the dry season and (d), (h) and (l) for the rainy season.

4.4 Discussion

4.4.1 Role of hydrology in habitat seasonality and implications on transmission

The average vector infection and prevalence rates in the "Default EMOD" scenario were significantly higher than the "Integrated EMOD-NonIrr" scenario, which was attributed to the larval habitat input. From Figure 4.4a, one of the most noticeable differences between the two scenarios was the degree of seasonality of the larval habitat. In this case, the mean habitat area was the same in both scenarios but the magnitude of the seasonal variation was larger when hydrology was incorporated. The first reason is the default habitat function in EMOD assumed that the permanent habitat area was at equilibrium and remained constant

throughout the simulation. In reality, permanent habitats such as those in river edges are likely to vary in area with climate conditions over a multi-year simulation. Secondly, there was no infiltration mechanism for the other two habitat types (Equation (4.1) and Equation (4.3)) so new ponds started forming immediately at the onset of the rainy season and continued forming towards the end of the rainy season whenever there was rainfall. The result was an earlier rising limb and a delayed falling limb in the time series compared to the "Integrated EMOD-NonIrr" scenario (see Figure C.11for an example). As the mean area in both scenarios were the same, the "Default EMOD" scenario time series naturally ended up with a wider but flatter crest.

Due to high larval capacity per unit area for permanent habitats, the degree of seasonality in the habitat larval capacity time series is even more pronounced (Figure C.9c). To evaluate the effect of the degree of seasonality in the larval habitat on malaria transmission, we conducted a sensitivity analysis using a synthetic sinusoidal time series for larval habitat fractional area with the same mean but different amplitudes:

$$F_{\alpha}^{t} = \alpha \cos(365.25t + 125.2) + \bar{F}, \tag{4.6}$$

where α is amplitude of fractional area, F_{α}^{t} is fractional area at time t, $\bar{F}$ is the mean fractional area specific to the study derived from the hydrologic model.

The results of the sensitivity analysis can be found in Figure C.12 and are summarized in Table 4.2. By reducing the seasonal amplitude from 0.2 to 0.1, adult vector population remained relatively unchanged but the vector infection and prevalence rates tripled. For the extreme case when α was reduced to 0, the vector infection and prevalence rates increased further by 4.29 times and 4.80 times respectively. This finding agrees with the higher simulated malaria transmission in the "Default EMOD" scenario compared to the "Integrated EMOD-NonIrr" scenario (Figure C.9e-g). One possible explanation is that the consistent adult vector population in the case where α was 0 resulted in a stable parasite transmission throughout the year. As α increased, the disparity between the high and low vector abundance seasons increased. In the low vector abundance season, transmission was minimal. In the high vector abundance season, transmission increased but was still limited by the human population. This resulted in an overall lower average vector infection rate and

prevalence. Therefore, a nuanced approach considering not only the general trend but also the seasonality of larval habitats is required to accurately predict malaria transmission.

Table 4.2: Average simulated adult vector abundance, adult vector infection rate and parasite prevalence rate for different amplitudes of larval habitat seasonality, α.

α	Adult Vectors	Vectors Infection Rate	Parasite Prevalence Rate
0.2	1.00	1.00	1.00
0.15	1.00	2.10	2.22
0.1	1.01	2.96	3.23
0.05	1.01	3.76	4.18
0	1.01	4.29	4.80

4.4.2 Insights provided by modeling on the effect of irrigation

By coupling hydrologic modeling with EMOD, we were able to investigate the effect of irrigation on malaria by comparing two scenarios whereby irrigation was the only difference. This allows us to isolate the effect of other environmental and social variables such as temperature, rainfall, topography and demography from the relationship between irrigation and malaria transmission. The significance of our approach is that it supplements past field comparative studies whereby the effect of irrigation could have been obscured by different field settings (Ijumba & Lindsay, 2001). The approach also has the potential to explore hypothetical scenarios to better guide decision making in water resource management.

Our modeling elucidates a few ways in which irrigation affected malaria transmission dynamics through larval habitats. First, all three habitat types increased in the dry season while temporary and semi-permanent habitats converted to permanent habitats in the rainy season. During the dry season, the increase was greater in the temporary and semi-permanent habitats compared to permanent habitats which were the dominant habitat (Figure 4.6a). The result was an increase in diversity of the habitats and agrees with field observations (Hawaria et al., 2019). On the other hand, permanent habitats became even more dominant in the rainy season. The change in relative abundance and stability of the habitats may favor the growth and survival of one vector species over the other, shifting the

predominant vector species in the extreme case (Bamou et al., 2018; Chaves et al., 2021; Naranjo-Díaz et al., 2020).

Next, irrigation not only creates transmission all-year round but also intensifies the main transmission period during the rainy season in terms of EIR and prevalence rate (Figure 4.7e-f). Studies have shown that irrigation can extend malaria transmission throughout the year due to the availability of water for breeding in the dry season (Kibret et al., 2014). Our results also show that irrigation during the dry season can also increase the stability of the habitats in the rainy season by creating high soil moisture conditions favorable for ponding prior to the onset of the rainy season. As habitat stability is linked to adult vector density (Ndenga et al., 2011), the dominance of permanent habitats during the rainy season resulted in a proliferation of adult vectors compared to the non-irrigation scenario which had a higher proportion of semi-permanent habitats. Besides a higher adult vector population in the main transmission season, there could also be carryover of parasites in the human population from the dry season (Midekisa et al., 2015), resulting in a higher vector infection rate. This ripple effect has been observed in past studies (Midekisa et al., 2015; Pascual et al., 2008) investigating the link between malaria transmission season and preceding rainfall. Our results suggest that irrigation can also produce the same cascading effect.

Thirdly, the modeling revealed that peak malaria transmission was shifted forward by around one month in the irrigation scenario (Figure 4.7e-f). Studies in East Africa have shown that rainfall is significantly correlated with malaria transmission with a lag time of 1 to 2 months (Loevinsohn, 1994; Zhou et al., 2004). The lag can be attributed to the time for infiltration to occur and runoff to accumulate low-lying areas as well as the development time for parasite growth. Past observations have proven that irrigation plays the same function as rainfall in providing larval habitats to support vector growth (Herrel et al., 2001; Ohta & Kaga, 2014). Hence, irrigation in the dry season in our study created a pseudo early rainfall season, which causes earlier onset of mosquito breeding and peak in transmission.

Lastly, we demonstrated the effect of irrigation on the spatiotemporal distribution of malaria transmission by considering the heterogeneity of larval habitats (Figure 4.8).

Discussion on the association between irrigation and malaria transmission is not new (Ijumba & Lindsay, 2001). While past observations tell us that irrigation can increase the adult vector population (Demissew et al., 2020), it remains a challenge to predict where and when breeding will occur (Frake et al., 2020). Integrating local irrigation practice and environmental characteristics such as land use, topography and soil properties, the model provided new insights into the breeding hotspots which were broken down into temporary, semi-permanent and permanent habitat types. This information can help LSM as a supplementary vector control by prioritizing resources for operational planning. LSM is known to be efficient where habitats are findable, few and fixed (Djamouko-Djonkam et al., 2019; Stanton et al., 2021). Based on the results, we can identify the location of habitats, period with manageable habitat abundance and single out semi-permanent and permanent habitats for targeted larviciding. Comparing the non-irrigation to irrigation scenario also allows us to distinguish habitats hotspots induced by irrigation from those already present without irrigation. Other means such as water resource management can be considered to control the former.

4.5 Conclusion

Malaria transmission is intrinsically related to larval habitats, which cannot be characterized by climate alone. By coupling a hydrologic model with an agent-based malaria model, the variability of larval habitats increased and resulted in a significantly lower malaria transmission as opposed to modeling habitats based on a simplified function of climate factors. We also demonstrated how habitat heterogeneity based on hydrologic processes could affect the spatiotemporal dynamics of malaria transmission.

The hydrology-integrated framework enabled us to investigate the effect of irrigation on malaria transmission through changes to larval habitats which were broken down into temporary, semi-permanent and permanent types. Results indicated that all three habitat types increased in the dry season while temporary and semi-permanent habitats converted to permanent habitats in the rainy season. This influenced the transmission dynamics significantly as transmission was sustained all-year round and intensified during the main season. Lastly, the malaria transmission peak was shifted forward by around one month.

These findings can help guide malaria intervention strategies to mitigate the effect of irrigation.

This study presents a novel generalizable framework that simulates the spatiotemporal dynamics of malaria transmission under the influence of irrigation by integrating hydrologic modeling with an agent-based model. The model can be used to simulate different water resource management practices with the objective of developing tailor-made intervention strategies.

Chapter 5

Conclusion

5.1 Summary

Irrigation has been associated with an increase in malaria risk but the underlying mechanism is not well understood. Irrigation can increase the availability of water for larval habitat formation but the governing hydrologic processes are complex and spatially variable. The incorporation of hydrologic modeling is therefore essential to understand the effect of irrigation on malaria. A framework to improve the larval habitat representation in an agent-based model by coupling with a distributed hydrologic model was presented. A new topographic conditioning workflow was developed to resolve physically meaningful land surface depressions relevant to larval habitats in the hydrologic model. The framework was used to investigate the effect of irrigation on transmission dynamics. An irrigated sugarcane plantation in Arjo, Ethiopia was selected as the study area.

In Chapter 2, using high resolution (50 m) distributed hydrologic modeling with remotely sensed data, I examined its ability to simulate potential malaria vector larval habitats and analyzed the influence of irrigation in the year of 2018. The model was calibrated using field observations of larval habitats to successfully predict ponding at all surveyed locations from the validation dataset. Results show that without irrigation, the area inside farms was inherently more prone to ponding than the area outside farms. With irrigation, the probability of finding potential larval habitats inside the farms in the year increased from 40% to 67%. Irrigation also dampened the seasonality of the potential larval habitats such that the peak larval habitat occurrence window during the rainy season was extended. Lastly, the stability of larval habitats was prolonged, with a significant shift from semi-permanent habitats to permanent habitats lasting beyond 6 months, pointing to the impact of irrigation in creating conducive mosquito habitats throughout most of the year.

Despite a growing emphasis on the impact of surface depressions on hydrologic processes, there is a lack of a formalized DEM processing tool that can resolve surface

depressions of interest for fully distributed hydrologic modeling. In Chapter 3, a new topographic conditioning workflow, Depression-Preserved DEM Processing (D2P) algorithm, was presented with the following key features: (1) an adaptive screening interval for delineation of depressions, (2) the ability to filter out anthropogenic land surface features (e.g., bridges), (3) the ability to blend river smoothing (e.g., a general downslope profile) and depression resolving functionality. The proposed algorithm was applied in a case study in GCEW.

For the selected study area, the D2P algorithm performed well in identifying delineated ponds with area of 900 m^2 or larger from NAIP satellite imagery at the DEM resolution of 1 m, 5 m and 10 m but not at 20 m. By comparing D2P and a traditional DEM filling method, D2P had a significantly lower impact on the raw DEM as it avoids distorting terrain attributes unnecessarily by offering user flexibility in conditioning depressions. Furthermore, hydrologic simulation using the D2P processed DEM resulted in a more robust characterization on surface water dynamics based on higher surface water storage as well as an attenuated and delayed peak streamflow. The case study also suggests that the treatment of surface depressions and the choice of method for smoothening topographic variability (e.g., filling and breaching) have a significant influence on the hydrologic simulation results.

Building on Chapter 2 and Chapter 3, Chapter 4 details the coupling of the hydrologic modeling framework with an agent-based malaria model (EMOD). The habitat function in EMOD was replaced by larval habitats simulated by the hydrologic model, which were classified into temporary, semi-permanent and permanent habitats. By including detailed hydrologic process, the inter- and intra-annual variabilities of larval habitats increased and resulted in a significantly lower malaria transmission. The study further revealed that the higher the amplitude of seasonality, the lower the transmission. This is due to the lack of a consistent adult vector population to sustain parasite transmission throughout the year.

Using the coupled model to investigate the effect of irrigation on malaria transmission, it was found that all three habitat types increased in the dry season while temporary and semi-permanent habitats converted to permanent habitats in the rainy

season. This indicates that irrigation may enable the development of vectors in the dry season while stabilizing the growth in rainy seasons. As a result, transmission was sustained all-year round and intensified during the main season. The latter can also be attributed to the carryover of parasites in the human population from the dry season. Lastly, the malaria transmission peak was found to be shifted forward by around one month. Presumably, irrigation in the dry season in the study created a pseudo early rainfall season, which caused an earlier onset of mosquito breeding and peak in transmission.

Through developing a hydrology-based malaria model and improving the representation of surface depressions in the DEM used by the model, insights were gained into the impact of irrigation on the spatiotemporal distribution of larval habitats and malaria transmission. The framework presented in this dissertation particularly helps larval source management as a supplementary vector control by identifying malaria hotspots and prioritizing resources for operational planning.

5.2 Limitations and future work

This dissertation presents a novel generalizable framework that seeks to explicitly simulate the spatiotemporal distribution of malaria larval habitats and study transmission dynamics under the influence of irrigation by integrating hydrologic modeling with an agent-based model. Some limitations remain which require more future work.

Firstly, I chose to simulate the surface layer soil saturation at 50 m resolution in Chapter 2 and 100 m in Chapter 4, coupled with a threshold, to quantify ponding instead of explicitly simulating the surface water depth. It is not feasible to simulate surface water depth at the scale of the larval habitats because this requires higher model resolution which will result in an unmanageable run time. An integrated hydrologic model like ParFlow details the nonlinear interactions and connections between groundwater levels, soil moisture and land energy. Even with parallel computing resources, running multi-year simulations at resolutions lower than 50 m will still be a challenge. With ongoing advances in accelerator architectures such as graphic processing units (GPUs), modifying ParFlow to adopt these architectures may alleviate this constraint (Hokkanen et al., 2021). In addition, developing a

scaling relationship from low to high resolution for soil saturation or water level can also be explored. Applying this relationship to the simulated results can allow the model to run at lower resolution with low computational cost while producing detailed spatial variability and patterns of ponding.

Secondly, the artificial habitats (i.e., formation independent of topography) are not considered in this framework. Examples of artificial habitats include water containers, discarded tires and water tanks which are more related to human activities. Although these habitats are less likely to be observed in the rural setting like Arjo, it is not the case in the urban setting and hence limits the applicability of this framework. Future work should incorporate aerial imagery and socio-economic data on household, street and landmarks and explore other means to estimate artificial larval habitats.

REFERENCES

Adugna Kendie, F., Hailegebriel, T. W., Nibret Semegn, E., & Wale Ferede, M. (2021). Prevalence of Malaria among Adults in Ethiopia: A Systematic Review and Meta-Analysis. *Journal of Tropical Medicine*, *2021*. https://doi.org/10.1155/2021/8863002

Afrane, Y. A., Zhou, G., Lawson, B. W., Githeko, A. K., & Yan, G. (2007). Life-table analysis of Anopheles arabiensis in western Kenya highlands: effects of land covers on larval and adult survivorship. *The American Journal of Tropical Medicine and Hygiene*, *77*(4), 660–666.

Afrane, Y. A., Githeko, A. K., & Yan, G. (2012). The ecology of Anopheles mosquitoes under climate change: case studies from the effects of deforestation in East African highlands. *Annals of the New York Academy of Sciences*, *1249*, 204–210. https://doi.org/10.1111/j.1749-6632.2011.06432.x

Amado, A. A., Politano, M., Schilling, K., & Weber, L. (2018). Investigating Hydrologic Connectivity of a Drained Prairie Pothole Region Wetland Complex using a Fully Integrated, Physically-Based Model. *Wetlands*, *38*(2), 233–245. https://doi.org/10.1007/s13157-016-0800-5

Ameli, A. A., & Creed, I. F. (2017). Quantifying hydrologic connectivity of wetlands to surface water systems. *Hydrology and Earth System Sciences*, *21*(3), 1791–1808. https://doi.org/10.5194/hess-21-1791-2017

Asare, E. O., Tompkins, A. M., & Bomblies, A. (2016). A Regional Model for Malaria Vector Developmental Habitats Evaluated Using Explicit, Pond-Resolving Surface Hydrology Simulations. *PLOS ONE*, *11*(3), e0150626. https://doi.org/10.1371/journal.pone.0150626

Ashby, S. F., & Falgout, R. D. (1996a). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear Science and Engineering*, *124*(1), 145–159. https://doi.org/10.13182/NSE96-A24230

Ashby, S. F., & Falgout, R. D. (1996b). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear Science and Engineering*, *124*(1), 145–159. https://doi.org/10.13182/NSE96-A24230

Atieli, H. E., Zhou, G., Lee, M. C., Kweka, E. J., Afrane, Y., Mwanzo, I., et al. (2011). Topography as a modifier of breeding habitats and concurrent vulnerability to malaria risk in the western Kenya highlands. *Parasites and Vectors*, *4*(1), 1–12. https://doi.org/10.1186/1756-3305-4-241

Ayele, D. G., Zewotir, T. T., & Mwambi, H. G. (2012). Prevalence and risk factors of malaria in Ethiopia. *Malaria Journal*, *11*, 1–9. https://doi.org/10.1186/1475-2875-11-195

Bamou, R., Mbakop, L. R., Kopya, E., Ndo, C., Awono-Ambene, P., Tchuinkam, T., et al. (2018). Changes in malaria vector bionomics and transmission patterns in the equatorial forest region of Cameroon between 2000 and 2017. *Parasites and Vectors*, *11*(1), 1–13. https://doi.org/10.1186/s13071-018-3049-4

Barnes, M. L., Welty, C., & Miller, A. J. (2016). Global Topographic Slope Enforcement to Ensure Connectivity and Drainage in an Urban Terrain. *Journal of Hydrologic Engineering*, *21*(4), 06015017. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001306

Barnes, R., Lehman, C., & Mulla, D. (2014). Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences*, *62*, 117–127. https://doi.org/10.1016/J.CAGEO.2013.04.024

Barnes, R., Callaghan, K. L., & Wickert, A. D. (2019). Computing water flow through complex landscapes, Part 2: Finding hierarchies in depressions and morphological segmentations. *Earth Surface Dynamics*, 1–19. https://doi.org/10.5194/esurf-2019-34

Bayoh, M. N. (2001). *Studies on the development and survival of anopheles gambiae sensu stricto at various temperatures and relative humidities*. University of Durham. Retrieved from http://etheses.dur.ac.uk/4952/

Bertassello, L. E., Rao, P. S. C., Jawitz, J. W., Aubeneau, A. F., & Botter, G. (2020). Wetlandscape hydrologic dynamics driven by shallow groundwater and landscape topography. *Hydrological Processes*, *34*(6), 1460–1474. https://doi.org/10.1002/hyp.13661

Bett, B., Said, M. Y., Sang, R., Bukachi, S., Wanyoike, S., Kifugo, S. C., et al. (2017). Effects of flood irrigation on the risk of selected zoonotic pathogens in an arid and semi-arid area in the eastern Kenya. *PLoS ONE*, *12*(5), 1–15. https://doi.org/10.1371/journal.pone.0172626

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., et al. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature*, *526*(7572), 207–211. https://doi.org/10.1038/nature15535

Biggs, J., von Fumetti, S., & Kelly-Quinn, M. (2017). The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. *Hydrobiologia*, *793*(1), 3–39. https://doi.org/10.1007/s10750-016-3007-0

Bill & Melinda Gates Foundation. (2022a). Malaria model. Retrieved November 8, 2022, from https://docs.idmod.org/projects/emod-malaria/en/latest/malaria-model-overview.html

Bill & Melinda Gates Foundation. (2022b). Overview of EMOD software. Retrieved May 11, 2022, from https://docs.idmod.org/projects/emod-malaria/en/2.20_a/software-overview.html

Bøgh, C., Lindsay, S. W., Clarke, S. E., Dean, A., Jawara, M., Pinder, M., & Thomas, C. J. (2007). High spatial resolution mapping of malaria transmission risk in the Gambia, west Africa, using LANDSAT TM satellite imagery. *The American Journal of Tropical Medicine and Hygiene*, *76*(5), 875–881.

Bomblies, A., Duchemin, J.-B., & Eltahir, E. A. B. (2008). Hydrology of malaria: Model development and application to a Sahelian village. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2008WR006917

Bomblies, A., Duchemin, J.-B., & Eltahir, E. A. (2009). A mechanistic approach for accurate simulation of village scale malaria transmission. *Malaria Journal*, *8*(1), 223. https://doi.org/10.1186/1475-2875-8-223

Brooks, J. R., Mushet, D. M., Vanderhoof, M. K., Leibowitz, S. G., Christensen, J. R., Neff, B. P., et al. (2018). Estimating Wetland Connectivity to Streams in the Prairie Pothole Region: An Isotopic and Remote Sensing Approach. *Water Resources Research*, *54*(2), 955–977. https://doi.org/10.1002/2017WR021016

Callow, J. N., Van Niel, K. P., & Boggs, G. S. (2007). How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis? *Journal of Hydrology*, *332*(1–2), 30–39. https://doi.org/10.1016/J.JHYDROL.2006.06.020

Carlson, R., & Danner, A. (2010). Bridge detection in grid terrains and improved drainage enforcement. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 250–259). https://doi.org/10.1145/1869790.1869827

Carrasco-Escobar, G. I., Manrique, E. I., Ruiz-Cabrejos, J. I., Saavedra, M., Alava, F., Bickersmith, S., et al. (2019). High-accuracy detection of malaria vector larval habitats using drone-based multispectral imagery. *PLoS Neglected Tropical Diseases*, *13*(1), e0007105. https://doi.org/10.1371/journal.pntd.0007105

Catry, T., Li, Z., Roux, E., Herbreteau, V., Gurgel, H., Mangeas, M., et al. (2018). Wetlands and malaria in the amazon: Guidelines for the use of synthetic aperture radar remote-sensing. *International Journal of Environmental Research and Public Health*, *15*(3), 468.

https://doi.org/10.3390/ijerph15030468

Centers for Disease Control and Prevention. (2020). Biology. Retrieved November 8, 2022, from https://www.cdc.gov/malaria/about/biology/index.html

Chaves, L. S. M., Bergo, E. S., Conn, J. E., Laporta, G. Z., Prist, P. R., & Sallum, M. A. M. (2021). Anthropogenic landscape decreases mosquito biodiversity and drives malaria vector proliferation in the Amazon rainforest. *PLoS ONE*, *16*(1), e0245087. https://doi.org/10.1371/journal.pone.0245087

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., et al. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 7–27. https://doi.org/10.1016/j.isprsjprs.2014.09.002

Cheng, F., & Basu, N. B. (2017). Biogeochemical hotspots: Role of small water bodies in landscape nutrient processing. *Water Resources Research*, *53*, 1–19. https://doi.org/https://doi.org/10.1002/2016WR020102

Chow, V. Te, & Ben-Zvi, A. (1973). Hydrodynamic Modeling of Two-Dimensional Watershed Flow. *Journal of the Hydraulics Division*, *99*(11), 2023–2040. https://doi.org/10.1061/JYCEAJ.0003790

Chu, X., Zhang, J., Chi, Y., & Yang, J. (2010). An Improved Method for Watershed Delineation and Computation of Surface Depression Storage. In *Watershed Management 2010: Innovations in watershed management under land use and climate change* (pp. 1113–1122). American Society of Civil Engineer. Retrieved from https://doi.org/10.1061/41143(394)100

Chu, X., Yang, J., Chi, Y., & Zhang, J. (2013). Dynamic puddle delineation and modeling of puddle-to-puddle filling-spilling-merging-splitting overland flow processes. *Water Resources Research*, *49*(6), 3825–3829. https://doi.org/10.1002/wrcr.20286

Cohen, M. J., Creed, I. F., Alexander, L., Basu, N. B., Calhoun, A. J. K., Craft, C., et al. (2016). Do geographically isolated wetlands influence landscape functions? *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), 1978–1986. https://doi.org/10.1073/pnas.1512650113

Condon, L. E., & Maxwell, R. M. (2013). Implementation of a linear optimization water allocation algorithm into a fully integrated physical hydrology model. *Advances in Water Resources*, *60*, 135–147. https://doi.org/10.1016/j.advwatres.2013.07.012

Condon, L. E., & Maxwell, R. M. (2014). Groundwater-fed irrigation impacts spatially

distributed temporal scaling behavior of the natural system: A spatio-temporal framework for understanding water management impacts. *Environmental Research Letters*, *9*(3). https://doi.org/10.1088/1748-9326/9/3/034009

Condon, L. E., & Maxwell, R. M. (2019). Modified priority flood and global slope enforcement algorithm for topographic processing in physically based hydrologic modeling applications. *Computers & Geosciences*, *126*, 73–83. https://doi.org/10.1016/J.CAGEO.2019.01.020

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., et al. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(22). https://doi.org/10.1029/2002jd003118

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). THE COMMON LAND MODEL. *Bulletin of the American Meteorological Society*, *84*(8), 1013–1023. https://doi.org/10.1175/BAMS-84-8-1013

Davies, C., Coetzee, M., & Lyons, C. L. (2016). Effect of stable and fluctuating temperatures on the life history traits of Anopheles arabiensis and An. quadriannulatus under conditions of inter- and intra-specific competition. *Parasites and Vectors*, *9*(1), 1–9. https://doi.org/10.1186/s13071-016-1630-2

Degefa, T., Yewhalaw, D., Zhou, G., Lee, M.-C., Atieli, H., Githeko, A. K., & Yan, G. (2017). Indoor and outdoor malaria vector surveillance in western Kenya: Implications for better understanding of residual transmission. *Malaria Journal*, *16*(1), 1–13. https://doi.org/10.1186/s12936-017-2098-z

Demissew, A., Hawaria, D., Kibret, S., Animut, A., Tsegaye, A., Lee, M. C., et al. (2020). Impact of sugarcane irrigation on malaria vector Anopheles mosquito fauna, abundance and seasonality in Arjo-Didessa, Ethiopia. *Malaria Journal*, *19*(1), 1–8. https://doi.org/10.1186/s12936-020-03416-0

Djamouko-Djonkam, L., Mounchili-Ndam, S., Kala-Chouakeu, N., Nana-Ndjangwo, S. M., Kopya, E., Sonhafouo-Chiana, N., et al. (2019). Spatial distribution of Anopheles gambiae sensu lato larvae in the urban environment of Yaoundé, Cameroon. *Infectious Diseases of Poverty*, *8*(1), 1–15. https://doi.org/10.1186/s40249-019-0597-6

Eckhoff, P. (2013). Mathematical models of within-host and transmission dynamics to determine effects of malaria interventions in a variety of transmission settings. *American Journal of Tropical Medicine and Hygiene*, *88*(5), 817–827. https://doi.org/10.4269/ajtmh.12-0007

Eckhoff, P. A. (2011). A malaria transmission-directed model of mosquito life cycle and ecology. *Malaria Journal*, *10*(1), 1–17. https://doi.org/10.1186/1475-2875-10-303

Eckhoff, P. A., & Wenger, E. A. (2016). The EMOD Individual-Based Model. In S. M. Niaz Arifin, G. R. Madey, & F. H. Collins (Eds.), *Spatial Agent-Based Simulation Modeling in Public Health* (1st ed., pp. 185–208). John Wiley and Sons. https://doi.org/10.1002/9781118964385.ch11

Ermert, V., Fink, A. H., Jones, A. E., & Morse, A. P. (2011). Development of a new version of the Liverpool malaria model. II. Calibration and validation for West Africa. *Malaria Journal*, *10*, 1–19. https://doi.org/10.1186/1475-2875-10-62

Eshete, D. G., Sinshaw, B. G., & Legese, K. G. (2020). Critical review on improving irrigation water use efficiency: Advances, challenges, and opportunities in the Ethiopia context. *Water-Energy Nexus*, *3*, 143–154. https://doi.org/10.1016/j.wen.2020.09.001

ESRI. (2020). ArcGIS Pro. Redlands, CA: Environmental Systems Research Institute.

Ethiopian Corporation. (2020). Arjo Didessa Sugar Factory. Retrieved April 5, 2020, from https://www.ethiopiansugar.com/blog/arjo-dediessa-sugar-factory/

Evenson, G. R., Golden, H. E., Lane, C. R., & D'Amico, E. (2016). An improved representation of geographically isolated wetlands in a watershed-scale hydrologic model. *Hydrological Processes*, *30*(22), 4168–4184. https://doi.org/10.1002/hyp.10930

Ferguson, I. M., & Maxwell, R. M. (2012). Human impacts on terrestrial hydrology: Climate change versus pumping and irrigation. *Environmental Research Letters*, *7*(4). https://doi.org/10.1088/1748-9326/7/4/044022

Fillinger, U., & Lindsay, S. W. (2011, December 13). Larval source management for malaria control in Africa: Myths and reality. *Malaria Journal*. BioMed Central. https://doi.org/10.1186/1475-2875-10-353

Fillinger, U., Sonye, G., Killeen, G. F., Knols, B. G. J., & Becker, N. (2004). The practical importance of permanent and semipermanent habitats for controlling aquatic stages of Anopheles gambiae sensu lato mosquitoes: Operational observations from a rural town in western Kenya. *Tropical Medicine and International Health*, *9*(12), 1274–1289. https://doi.org/10.1111/j.1365-3156.2004.01335.x

Frake, A. N., Namaona, W., Walker, E. D., & Messina, J. P. (2020). Estimating spatio-temporal distributions of mosquito breeding pools in irrigated agricultural schemes: A case study at the Bwanje Valley Irrigation Scheme. *Malaria Journal*, *19*(1), 1–21. https://doi.org/10.1186/s12936-020-3113-3

Galactionova, K., Smith, T. A., & Penny, M. A. (2021). Insights from modelling malaria vaccines for policy decisions: the focus on RTS,S. *Malaria Journal*, *20*(1), 1–8. https://doi.org/10.1186/s12936-021-03973-y

Gallegos, H. A., Schubert, J. E., & Sanders, B. F. (2009). Two-dimensional, high-resolution modeling of urban dam-break flooding: A case study of Baldwin Hills, California. *Advances in Water Resources*, *32*(8), 1323–1335. https://doi.org/10.1016/J.ADVWATRES.2009.05.008

Gardner, M. A., Morton, C. G., Huntington, J. L., Niswonger, R. G., & Henson, W. R. (2018). Input data processing tools for the integrated hydrologic model GSFLOW. *Environmental Modelling and Software*, *109*, 41–53. https://doi.org/10.1016/j.envsoft.2018.07.020

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gerardin, J., Ouédraogo, A. L., McCarthy, K. A., Eckhoff, P. A., & Wenger, E. A. (2015). Characterization of the infectious reservoir of malaria with an agent-based model calibrated to age-stratified parasite densities and infectiousness. *Malaria Journal*, *14*(1), 1–13. https://doi.org/10.1186/S12936-015-0751-Y

Gianotti, R. L., Bomblies, A., & Eltahir, E. A. B. (2009). Hydrologic modeling to screen potential environmental management methods for malaria vector control in Niger, *45*(8), 8438. https://doi.org/10.1029/2008WR007567

Gimmg, J. E., Ombok, M., Kamau, L., & Hawley, W. A. (2001). Characteristics of larval anopheline (Diptera: Culicidae) habitats in western Kenya. *Journal of Medical Entomology*, *38*(2), 282–288. https://doi.org/10.1603/0022-2585-38.2.282

Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, *41*(11), 3891–3898. https://doi.org/10.1002/2014GL059856

Golden, H. E., Lane, C. R., Amatya, D. M., Bandilla, K. W., Raanan Kiperwas, H., Knightes, C. D., & Ssegane, H. (2014). Hydrologic connectivity between geographically isolated wetlands and surface water systems: A review of select modeling methods. *Environmental Modelling & Software*, *53*, 190–206. https://doi.org/10.1016/J.ENVSOFT.2013.12.004

Golden, H. E., Creed, I. F., Ali, G., Basu, N. B., Neff, B. P., Rains, M. C., et al. (2017). Integrating

geographically isolated wetlands into land management decisions. *Frontiers in Ecology and the Environment*, *15*(6), 319–327. https://doi.org/10.1002/fee.1504

GRASS Development Team. (2020). Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. Retrieved from https://grass.osgeo.org

Griffin, J. T., Hollingsworth, T. D., Okell, L. C., Churcher, T. S., White, M., Hinsley, W., et al. (2010). Reducing Plasmodium falciparum malaria transmission in Africa: A model-based evaluation of intervention strategies. *PLoS Medicine*, *7*(8), e1000324. https://doi.org/10.1371/journal.pmed.1000324

Hagos, F., Makombe, G., Namara, R. E., & Awulachew, S. B. (2009). *Importance of Irrigated Agriculture to the Ethiopian Economy: Capturing the Direct Net Benefits of Irrigation* (Vol. 128). Colombo, Sri Lanka: International Water Management Institute. https://doi.org/10.22004/ag.econ.61100

Haile, G. G., & Kassa, A. K. (2015). Irrigation in Ethiopia, a review. *Journal of Environment and Earth Science*, *3*(10), 264–269.

Haileselassie, W., Zemene, E., Lee, M. C., Zhong, D., Zhou, G., Taye, B., et al. (2021). The effect of irrigation on malaria vector bionomics and transmission intensity in western Ethiopia. *Parasites and Vectors*, *14*(1), 1–11. https://doi.org/10.1186/s13071-021-04993-y

Hardy, A., Makame, M., Cross, D., Majambere, S., & Msellem, M. (2017). Using low-cost drones to map malaria vector habitats. *Parasites and Vectors*, *10*(1), 29. https://doi.org/10.1186/s13071-017-1973-3

Hardy, A., Ettritch, G., Cross, D. E., Bunting, P., Liywalii, F., Sakala, J., et al. (2019). Automatic detection of open and vegetated water bodies using Sentinel 1 to map African malaria vector mosquito breeding habitats. *Remote Sensing*, *11*(5), 593. https://doi.org/10.3390/rs11050593

Hardy, A. J., Gamarra, J. G. P., Cross, D. E., Macklin, M. G., Smith, M. W., Kihonda, J., et al. (2013). Habitat Hydrology and Geomorphology Control the Distribution of Malaria Vector Larvae in Rural Africa. *PLoS ONE*, *8*(12), e81931. https://doi.org/10.1371/journal.pone.0081931

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107. https://doi.org/10.1109/TSSC.1968.300136

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., & Borriello, G. (2010). Open Data Kit: Tools to Build Information Services for Developing Regions. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. New York, United States: Association for Computing Machinery. https://doi.org/10.1145/2369220.2369236

Hawaria, D., Getachew, H., Zhong, G., Demissew, A., Habitamu, K., Raya, B., et al. (2019). Ten years malaria trend at Arjo-Didessa sugar development site and its vicinity, Southwest Ethiopia: A retrospective study. *Malaria Journal*, *18*(1), 1–11. https://doi.org/10.1186/s12936-019-2777-z

Hawaria, D., Demissew, A., Kibret, S., Lee, M. C., Yewhalaw, D., & Yan, G. (2020). Effects of environmental modification on the diversity and positivity of anopheline mosquito aquatic habitats at Arjo-Dedessa irrigation development site, Southwest Ethiopia. *Infectious Diseases of Poverty*, *9*(1), 1–11. https://doi.org/10.1186/s40249-019-0620-y

Hay, L., Norton, P., Viger, R., Markstrom, S., Steven Regan, R., & Vanderhoof, M. (2018). Modelling surface-water depression storage in a Prairie Pothole Region. *Hydrological Processes*, *32*(4), 462–479. https://doi.org/10.1002/hyp.11416

Hengl, T. (2006). Finding the right pixel size. *Computers & Geosciences*, *32*(9), 1283–1298. https://doi.org/10.1016/J.CAGEO.2005.11.008

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Herrel, N., Amerasinghe, F. P., Ensink, J., Mukhtar, M., Van Der Hoek, W., & Konradsen, F. (2001). Breeding of Anopheles mosquitoes in irrigated areas of South Punjab, Pakistan. *Medical and Veterinary Entomology*, *15*(3), 236–248. https://doi.org/10.1046/j.0269-283X.2001.00312.x

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., et al. (2018). ERA5 hourly data on single levels from 1959 to present. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hinne, I. A., Attah, S. K., Mensah, B. A., Forson, A. O., & Afrane, Y. A. (2021). Larval habitat

diversity and Anopheles mosquito species distribution in different ecological zones in Ghana. *Parasites and Vectors*, *14*(1), 1–14. https://doi.org/10.1186/s13071-021-04701-w

Hodges, B. R. (2015). Representing hydrodynamically important blocking features in coastal or riverine lidar topography. *Natural Hazards and Earth System Sciences*, *15*(5), 1011–1023. https://doi.org/10.5194/nhess-15-1011-2015

Hokkanen, J., Kollet, S., Kraus, J., Herten, A., Hrywniak, M., & Pleiter, D. (2021). Leveraging HPC accelerator architectures with modern techniques — hydrologic modeling on GPUs with ParFlow. *Computational Geosciences*, *25*(5), 1579–1590. https://doi.org/10.1007/s10596-021-10051-4

Hong, Y., Hsu, K.-L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, *43*(12), 1834–1853. https://doi.org/10.1175/jam2173.1

Hoshen, M. B., & Morse, A. P. (2004). A weather-driven model of malaria transmission. *Malaria Journal*, *3*, 1–14. https://doi.org/10.1186/1475-2875-3-32

Ijumba, J. N., & Lindsay, S. W. (2001). Impact of irrigation on malaria in Africa: Paddies paradox. *Medical and Veterinary Entomology*, *15*(1), 1–11. https://doi.org/10.1046/j.1365-2915.2001.00279.x

Iserles, A. (2009). *A First Course in the Numerical Analysis of Differential Equations* (2nd ed., Vol. 148). New York: Cambridge University Press.

Jan, A., Coon, E. T., Graham, J. D., & Painter, S. L. (2018). A Subgrid Approach for Modeling Microtopography Effects on Overland Flow. *Water Resources Research*, *54*(9), 6153–6167. https://doi.org/10.1029/2017WR021898

Jaros, A., Rossi, P. M., Ronkanen, A. K., & Kløve, B. (2019). Parameterisation of an integrated groundwater-surface water model for hydrological analysis of boreal aapa mire wetlands. *Journal of Hydrology*, *575*, 175–191. https://doi.org/10.1016/j.jhydrol.2019.04.094

Ji, P., Yuan, X., & Liang, X.-Z. (2017). Do Lateral Flows Matter for the Hyperresolution Land Surface Modeling? *Journal of Geophysical Research: Atmospheres*, *122*(22), 12–077. https://doi.org/10.1002/2017JD027366

Jiang, A.-L., Lee, M.-C., Zhou, G., Zhong, D., Hawaria Dawit, Kibret, S., et al. (2021). Predicting Distribution of Malaria Vector Larval Habitats in Ethiopia by Integrating Distributed Hydrologic Modeling with Remotely Sensed Data. *Scientific Reports*, *11*(1), 1–14.

https://doi.org/https://doi.org/10.1038/s41598-021-89576-8

Jones, C. Nathan, Ameli, A., Neff, B. P., Evenson, G. R., McLaughlin, D. L., Golden, H. E., & Lane, C. R. (2019). Modeling Connectivity of Non-floodplain Wetlands: Insights, Approaches, and Recommendations. *Journal of the American Water Resources Association*, *55*(3), 559–577. https://doi.org/10.1111/1752-1688.12735

Jones, Charles Nathan, Evenson, G. R., McLaughlin, D. L., Vanderhoof, M. K., Lang, M. W., McCarty, G. W., et al. (2018). Estimating restorable wetland water storage at landscape scales. *Hydrological Processes*, *32*(2), 305–313. https://doi.org/10.1002/hyp.11405

Jones, J. E., & Woodward, C. S. (2001). Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, *24*(7), 763–774. https://doi.org/10.1016/S0309-1708(00)00075-0

JPL, N. (2013). NASA Shuttle Radar Topography Mission Global 1 arc second. NASA EOSDIS Land Processes DAAC. https://doi.org/https:doi.org/10.5066/F7PR7TFT

Kahamba, N. F., Finda, M., Ngowo, H. S., Msugupakulya, B. J., Baldini, F., Koekemoer, L. L., et al. (2022). Using ecological observations to improve malaria control in areas where Anopheles funestus is the dominant vector. *Malaria Journal*, *21*(1), 1–15. https://doi.org/10.1186/s12936-022-04198-3

Keiser, J., De Castro, M. C., Maltese, M. F., Bos, R., Tanner, M., Singer, B. H., & Utzinger, J. (2005). Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *American Journal of Tropical Medicine and Hygiene*, *72*(4), 392–406. https://doi.org/10.4269/ajtmh.2005.72.392

Kibret, S., Alemu, Y., Boelee, E., Tekie, H., Alemu, D., & Petros, B. (2010). The impact of a small-scale irrigation scheme on malaria transmission in Ziway area, Central Ethiopia. *Tropical Medicine and International Health*, *15*(1), 41–50. https://doi.org/10.1111/j.1365-3156.2009.02423.x

Kibret, S., Wilson, G. G., Tekie, H., & Petros, B. (2014). Increased malaria transmission around irrigation schemes in Ethiopia and the potential of canal water management for malaria vector control. *Malaria Journal*, *13*(1), 1–12. https://doi.org/10.1186/1475-2875-13-360

Kibret, S., Glenn Wilson, G., Ryder, D., Tekie, H., & Petros, B. (2019). Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in Ethiopia. *Malaria Journal*, *18*(1), 54. https://doi.org/10.1186/s12936-019-2689-y

Killeen, G. F., Tatarsky, A., Diabate, A., Chaccour, C. J., Marshall, J. M., Okumu, F. O., et al. (2017). Developing an expanded vector control toolbox for malaria elimination. *BMJ Global Health*, *2*(2), 1–8. https://doi.org/10.1136/bmjgh-2016-000211

Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, *29*(7), 945–958. https://doi.org/10.1016/J.ADVWATRES.2005.08.006

Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, *44*(2). https://doi.org/10.1029/2007WR006004

Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., & Simmer, C. (2010). Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research*, *46*(4), 1–7. https://doi.org/10.1029/2009WR008730

Kweka, E. J., Zhou, G., Munga, S., Lee, M. C., Atieli, H. E., Nyindo, M., et al. (2012). Anopheline Larval Habitats Seasonality and Species Distribution: A Prerequisite for Effective Targeted Larval Habitats Control Programmes. *PLoS ONE*, *7*(12), e52084. https://doi.org/10.1371/journal.pone.0052084

Le, P. V. V., Kumar, P., & Ruiz, M. O. (2018). Stochastic lattice-based modelling of malaria dynamics. *Malaria Journal*, *17*(1), 1–17. https://doi.org/10.1186/s12936-018-2397-z

Leibowitz, S. G., Jr, P. J. W., Schofield, K. A., Alexander, L. C., Vanderhoof, K., & Golden, H. E. (2019). Connectivity of Streams and Wetlands to Downstream Waters : an Integrated Systems Framework. *Journal of the American Water Resources Association*, *54*(2), 1–42. https://doi.org/10.1111/1752-1688.12631

Li, S., MacMillan, R. A., Lobb, D. A., McConkey, B. G., Moulin, A., & Fraser, W. R. (2011). Lidar DEM error analyses and topographic depression identification in a hummocky landscape in the prairie region of Canada. *Geomorphology*, *129*(3–4), 263–275. https://doi.org/10.1016/j.geomorph.2011.02.020

Lidberg, W., Nilsson, M., Lundmark, T., & Ågren, A. M. (2017). Evaluating preprocessing methods of digital elevation models for hydrological modelling. *Hydrological Processes*, *31*(26), 4660–4668. https://doi.org/10.1002/hyp.11385

Lindsay, J. B. (2016a). Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. *Hydrological Processes*, *30*(6), 846–857. https://doi.org/10.1002/hyp.10648

Lindsay, J. B. (2016b). The practice of DEM stream burning revisited. *Earth Surface Processes and Landforms*, *41*(5), 658–668. https://doi.org/10.1002/esp.3888

Lindsay, J. B., & Creed, I. F. (2005). Removal of artifact depressions from digital elevation models: Towards a minimum impact approach. *Hydrological Processes*, *19*(16), 3113–3126. https://doi.org/10.1002/hyp.5835

Liu, G., & Schwartz, F. W. (2011). An integrated observational and model-based analysis of the hydrologic response of prairie pothole systems to variability in climate. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2010WR009084

Liu, G., Schwartz, F. W., Wright, C. K., & McIntyre, N. E. (2016). Characterizing the Climate-Driven Collapses and Expansions of Wetland Habitats with a Fully Integrated Surface–Subsurface Hydrologic Model. *Wetlands*, *36*, 287–297. https://doi.org/10.1007/s13157-016-0817-9

Loevinsohn, M. E. (1994). Climatic warming and increased malaria incidence in Rwanda. *The Lancet*, *343*(8899), 714–718. https://doi.org/10.1016/S0140-6736(94)91586-5

Lyons, C. L., Coetzee, M., & Chown, S. L. (2013). Stable and fluctuating temperature effects on the development rate and survival of two malaria vectors, Anopheles arabiensis and Anopheles funestus. *Parasites and Vectors*, *6*(1), 1–9. https://doi.org/10.1186/1756-3305-6-104

Mabaso, M. L. H., Craig, M., Ross, A., & Smith, T. (2007). Environmental predictors of the seasonality of malaria transmission in Africa: The challenge. *American Journal of Tropical Medicine and Hygiene*, *76*(1), 33–38. https://doi.org/10.4269/ajtmh.2007.76.33

Maharaj, R. (2003). Life Table Characteristics of Anopheles arabiensis (Diptera: Culicidae) Under Simulated Seasonal Conditions. *Journal of Medical Entomology*, *40*(6), 737–742. https://doi.org/10.1603/0022-2585-40.6.737

Malabo Montpellier Panel. (2018). *WATER-WISE Smart Irrigation Strategies for Africa*.

Mangani, C., Frake, A. N., Chipula, G., Mkwaila, W., Kakota, T., Mambo, I., et al. (2022). Proximity of Residence to Irrigation Determines Malaria Risk and Anopheles Abundance at an Irrigated Agroecosystem in Malawi. *American Journal of Tropical Medicine and Hygiene*, *106*(1), 283–292. https://doi.org/10.4269/ajtmh.21-0390

Maxwell, R. M. (2013). A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Advances in Water Resources*, *53*, 109–117. https://doi.org/10.1016/J.ADVWATRES.2012.10.001

Maxwell, R. M., & Kollet, S. J. (2008). Interdependence of groundwater dynamics and land-energy feedbacks under climate change. *Nature Geoscience*, *1*(10), 665–669. https://doi.org/10.1038/ngeo315

Maxwell, R. M., & Miller, N. L. (2005). Development of a Coupled Land Surface and Groundwater Model. *Journal of Hydrometeorology*, *6*(3), 233–247. https://doi.org/10.1175/JHM422.1

McCann, R. S., van den Berg, H., Diggle, P. J., van Vugt, M., Terlouw, D. J., Phiri, K. S., et al. (2017). Assessment of the effect of larval source management and house improvement on malaria transmission when added to standard malaria control strategies in southern Malawi: Study protocol for a cluster-randomised controlled trial. *BMC Infectious Diseases*, *17*(1), 1–15. https://doi.org/10.1186/s12879-017-2749-2

Metz, M., Mitasova, H., & Harmon, R. S. (2011). Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrology and Earth System Sciences*, *15*(2), 667–678. https://doi.org/10.5194/hess-15-667-2011

Midekisa, A., Beyene, B., Mihretie, A., Bayabil, E., & Wimberly, M. C. (2015). Seasonal associations of climatic drivers and malaria in the highlands of Ethiopia. *Parasites and Vectors*, *8*(1), 1–11. https://doi.org/10.1186/s13071-015-0954-7

Minakawa, N., Mutero, C. M., Githure, J. I., Beier, J. C., & Yan, G. (1999). Spatial distribution and habitat characterization of anopheline mosquito larvae in western Kenya. *American Journal of Tropical Medicine and Hygiene*, *61*(6), 1010–1016. https://doi.org/10.4269/ajtmh.1999.61.1010

Montosi, E., Manzoni, S., Porporato, A., & Montanari, A. (2012). An ecohydrological model of malaria outbreaks. *Hydrology and Earth System Sciences*, *16*(8), 2759–2769. https://doi.org/10.5194/hess-16-2759-2012

Moreno-Gómez, M., Liedl, R., & Stefan, C. (2019). A New GIS-Based Model for Karst Dolines Mapping Using LiDAR; Application of a Multidepth Threshold Approach in the Yucatan Karst, Mexico. *Remote Sensing*, *11*(10), 1147. https://doi.org/10.3390/rs11101147

Munga, S., Minakawa, N., Zhou, G., Mushinzimana, E., Barrack, O. O. J., Githeko, A. K., & Yan, G. (2006). Association between land cover and habitat productivity of malaria vectors in western Kenyan highlands. *American Journal of Tropical Medicine and Hygiene*, *74*(1), 69–75. https://doi.org/10.4269/ajtmh.2006.74.69

Mushinzimana, E., Munga, S., Minakawa, N., Li, L., Feng, C., Bian, L., et al. (2006). Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malaria Journal*, *5*(1), 1–11. https://doi.org/10.1186/1475-

2875-5-13

Mutuku, F. M., Bayoh, M. N., Hightower, A. W., Vulule, J. M., Gimnig, J. E., Mueke, J. M., et al. (2009). A supervised land cover classification of a western Kenya lowland endemic for human malaria: associations of land cover with larval Anopheles habitats. *International Journal of Health Geographics*, *8*(1), 1–13. https://doi.org/10.1186/1476-072X-8-19

Naranjo-Díaz, N., Hernandez-Valencia, J. C., Marín, A., & Correa, M. M. (2020). Relationship between land cover and Anophelinae species abundance, composition and diversity in NW Colombia. *Infection, Genetics and Evolution*, *78*, 104114. https://doi.org/10.1016/j.meegid.2019.104114

Nasab, M. T., & Chu, X. (2020). Macro-HyProS: A new macro-scale hydrologic processes simulator for depression-dominated cold climate regions. *Journal of Hydrology*, *580*, 124366. https://doi.org/10.1016/j.jhydrol.2019.124366

Nasab, M. T., Zhang, J., & Chu, X. (2017). A new depression-dominated delineation (D-cubed) method for improved watershed modelling. *Hydrological Processes*, *31*(19), 3364–3378. https://doi.org/10.1002/hyp.11261

Ndenga, B. A., Simbauni, J. A., Mbugi, J. P., Githeko, A. K., & Fillinger, U. (2011). Productivity of malaria vectors from different habitat types in the western kenya highlands. *PLoS ONE*, *6*(4), e19473. https://doi.org/10.1371/journal.pone.0019473

Nguyen, P., Shearer, E. J., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., et al. (2019). The CHRS data portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Scientific Data*, *6*(1), 1–10. https://doi.org/10.1038/sdata.2018.296

Nyanjom, S. R. G., Chen, H., Gebre-Michael, T., Bekele, E., Shililu, J., Githure, J., et al. (2003). Population Genetic Structure of Anopheles arabiensis Mosquitoes in Ethiopia and Eritrea. *Journal of Heredity*, *94*(6), 457–463. https://doi.org/10.1093/jhered/esg100

Ohta, S., & Kaga, T. (2014). Effect of irrigation systems on temporal distribution of malaria vectors in semi-arid regions. *International Journal of Biometeorology*, *58*(3), 349–359. https://doi.org/10.1007/s00484-012-0630-y

Olalubi, O. A. (2016). Promoting Larval Source Management as a Vital Supplemental Addendum and More Likely Cost-Effective Approach for Malaria Vector Control in Nigeria. *Journal of Prevention and Infection Control*, *2*, 2. https://doi.org/10.21767/2471-9668.100015

Oliver, S. V., & Brooke, B. D. (2017). The effect of elevated temperatures on the life history

and insecticide resistance phenotype of the major malaria vector Anopheles arabiensis (Diptera: Culicidae). *Malaria Journal*, *16*(1), 1–13. https://doi.org/10.1186/s12936-017-1720-4

Ondeto, B. M., Wang, X., Atieli, H., Orondo, P. W., Ochwedo, K. O., Omondi, C. J., et al. (2022). Malaria vector bionomics and transmission in irrigated and non-irrigated sites in western Kenya. *Parasitology Research*, 1–17. https://doi.org/10.1007/s00436-022-07678-2

Ozesmi, S. L., & Bauer, M. E. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, *10*(5), 381–402. https://doi.org/10.1023/A:1020908432489

Parham, P. E., Pople, D., Christiansen-Jucht, C., Lindsay, S., Hinsley, W., & Michael, E. (2012). Modeling the role of environmental variables on the population dynamics of the malaria vector Anopheles gambiae sensu stricto. *Malaria Journal*, *11*(1), 1–13. https://doi.org/10.1186/1475-2875-11-271

Pascual, M., Cazelles, B., Bouma, M. J., Chaves, L. F., & Koelle, K. (2008). Shifting patterns: Malaria dynamics and rainfall variability in an African highland. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1631), 123–132. https://doi.org/10.1098/rspb.2007.1068

Patz, J. A., & Olson, S. H. (2006). Malaria risk and temperature: Influences from global climate change and local land use practices. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(15), 5635–5636. https://doi.org/10.1073/pnas.0601493103

Patz, J. A., Strzepek, K., Lele, S., Hedden, M., Greene, S., Noden, B., et al. (1998). Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya. *Tropical Medicine and International Health*, *3*(10), 818–827. https://doi.org/10.1046/j.1365-3156.1998.00309.x

Planchon, O., & Darboux, F. (2001). A fast , simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, *46*(2–3), 159–176.

Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission - A new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(4), 241–262. https://doi.org/10.1016/S0924-2716(02)00124-7

Rajib, A., Golden, H. E., Lane, C. R., & Wu, Q. (2020). Surface depression and wetland water storage improves major river basin hydrologic predictions. *Water Resources Research*, *56*(7), e2019WR026561. https://doi.org/10.1029/2019wr026561

Reiner, R. C., Perkins, T. A., Barker, C. M., Niu, T., Chaves, L. F., Ellis, A. M., et al. (2013). A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. *Journal of the Royal Society Interface*, *10*(81). https://doi.org/10.1098/rsif.2012.0921

Rejmánková, E., Grieco, J., Achee, N., & Roberts, D. R. (2013). Ecology of Larval Habitats. In S. Manguin (Ed.), *Anopheles mosquitoes—New insights into malaria vectors* (pp. 397–446). IntechOpen. https://doi.org/http://dx.doi.org/10.5772/55229

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010

Rieger, W. (1998). A phenomenon-based approach to upslope contributing area and depressions in DEMs. *Hydrological Processes*, *12*(6), 857–872. https://doi.org/10.1002/(SICI)1099-1085(199805)12:6

Rodell, M. (2004). Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophysical Research Letters*, *31*(20), L20504. https://doi.org/10.1029/2004GL020873

Ross, R. (1908). *Report on the Prevention of Malaria in Mauritius*. *Waterlow & Sons Ltd*. London. Retrieved from https://wellcomecollection.org/works/rjj6m9qm

Sadeghi, M., Nguyen, P., Naeini, M. R., Hsu, K., Braithwaite, D., & Sorooshian, S. (2021). PERSIANN-CCS-CDR, a 3-hourly 0.04° global precipitation climate data record for heavy precipitation studies. *Scientific Data*, *8*(1), 1–11. https://doi.org/10.1038/s41597-021-00940-9

Sanders, B. F. (2007). Evaluation of on-line DEMs for flood inundation modeling. *Advances in Water Resources*, *30*(8), 1831–1843. https://doi.org/10.1016/j.advwatres.2007.02.005

Sanders, B. F., & Schubert, J. E. (2019). PRIMo: Parallel raster inundation model. *Advances in Water Resources*, *126*, 79–95. https://doi.org/10.1016/j.advwatres.2019.02.007

Schwanghart, W., & Scherler, D. (2017). Bumps in river profiles: Uncertainty assessment and smoothing using quantile regression techniques. *Earth Surface Dynamics*, *5*(4), 821–839. https://doi.org/10.5194/esurf-5-821-2017

Selvaraj, P., Wenger, E. A., & Gerardin, J. (2018). Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: A modeling study. *BMC Infectious Diseases*, *18*(1), 1–14. https://doi.org/10.1186/s12879-018-3319-y

Sinka, M. E., Bangs, M. J., Manguin, S., Coetzee, M., Mbogo, C. M., Hemingway, J., et al. (2010). The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasites & Vectors*, *3*(1), 1–34. https://doi.org/10.1186/1756-3305-3-117

Smith, M. W., Macklin, M. G., & Thomas, C. J. (2013). Hydrological and geomorphological controls of malaria transmission. *Earth-Science Reviews*, *116*(1), 109–127. https://doi.org/10.1016/j.earscirev.2012.11.004

Smith, M. W., Willis, T., Alfieri, L., James, W. H. M., Trigg, M. A., Yamazaki, D., et al. (2020). Incorporating hydrology into climate suitability models changes projections of malaria transmission in Africa. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-18239-5

Smith, N. R., Trauer, J. M., Gambhir, M., Richards, J. S., Maude, R. J., Keith, J. M., & Flegg, J. A. (2018). Agent-based models of malaria transmission: A systematic review. *Malaria Journal*, *17*(1), 1–16. https://doi.org/10.1186/s12936-018-2442-y

Smith, T., Killeen, G. F., Maire, N., Ross, A., Molineaux, L., Tediosi, F., et al. (2006). Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of Plasmodium falciparum malaria: Overview. *The American Journal of Tropical Medicine and Hygiene*, *75*(Suppl 2), 1–10.

Soti, V., Tran, A., Degenne, P., Chevalier, V., Lo Seen, D., Thiongane, Y., et al. (2012). Combining Hydrology and Mosquito Population Models to Identify the Drivers of Rift Valley Fever Emergence in Semi-Arid Regions of West Africa. *PLoS Neglected Tropical Diseases*, *6*(8), e1795. https://doi.org/10.1371/journal.pntd.0001795

Stanton, M. C., Kalonde, P., Zembere, K., Hoek Spaans, R., & Jones, C. M. (2021). The application of drones for mosquito larval habitat identification in rural environments: a practical approach for malaria control? *Malaria Journal*, *20*(1), 1–17. https://doi.org/10.1186/s12936-021-03759-2

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Takaku, J., & Tadono, T. (2017). Quality updates of 'AW3D' global DSM generated from ALOS PRISM. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 5666–5669). IEEE. https://doi.org/10.1109/IGARSS.2017.8128293

Takaku, J., Tadono, T., Tsutsui, K., & Ichikawa, M. (2016). Validation of 'AW3D' Global DSM Generated from ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and*

Spatial Information Sciences, *III–4*, 25–31. https://doi.org/10.5194/isprsannals-III-4-25-2016

Tavares da Costa, R., Mazzoli, P., & Bagli, S. (2019). Limitations posed by free DEMs in watershed studies: The case of river Tanaro in Italy. *Frontiers in Earth Science*, *7*, 141. https://doi.org/10.3389/feart.2019.00141

The Lancet. (2022). Malaria in 2022: a year of opportunity. *Lancet (London, England)*, *399*(10335), 1573. https://doi.org/10.1016/S0140-6736(22)00729-2

United States Department of Agriculture. Soil Science Division Staff. (2017). *Soil Survey Manual No. 18*. https://doi.org/10.2307/1233734

United States Geological Survey Earth Resources Observation and Science Center. (2013). LANDSAT 8 OLI/TIRS Collection 1. https://doi.org/https:doi.org/10.5066/F71835S6

Vaughan, J. A. (2007). Population dynamics of Plasmodium sporogony. *Trends in Parasitology*, *23*(2), 63–70. https://doi.org/10.1016/j.pt.2006.12.009

Wang, L., & Liu, H. (2006). An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science*, *20*(2), 193–213. https://doi.org/10.1080/13658810500433453

Wang, N., & Chu, X. (2020). A new algorithm for delineation of surface depressions and channels. *Water*, *12*(1), 7. https://doi.org/10.3390/w12010007

Wang, N., Chu, X., & Zhang, X. (2021). Functionalities of surface depressions in runoff routing and hydrologic connectivity modeling. *Journal of Hydrology*, *593*, 125870. https://doi.org/10.1016/j.jhydrol.2020.125870

Ward, C., Torquebiau, R., & Xie, H. (2016). *Improved Agricultural Water Management for Africa's Drylands*. *World Bank Publications*. https://doi.org/10.1596/978-1-4648-0832-6

Woodrow, K., Lindsay, J. B., & Berg, A. A. (2016). Evaluating DEM conditioning techniques, elevation source data, and grid resolution for field-scale hydrological parameter extraction. *Journal of Hydrology*, *540*, 1022–1029. https://doi.org/10.1016/J.JHYDROL.2016.07.018

World Health Organization. (2013). Larval source management – a supplementary measure for malaria vector control. An operational manual. Geneva, Switzerland: World Health Organization. https://doi.org/10.1564/v25_feb_13

World Health Organization. (2021). *Word Malaria Report 2021*. Geneva, Switzerland.

Wu, Q., Deng, C., & Chen, Z. (2016). Automated delineation of karst sinkholes from LiDAR-derived digital elevation models. *Geomorphology*, *266*, 1–10. https://doi.org/10.1016/J.GEOMORPH.2016.05.006

Wu, Q., Lane, C. R., Wang, L., Vanderhoof, M. K., Christensen, J. R., & Liu, H. (2019). Efficient Delineation of Nested Depression Hierarchy in Digital Elevation Models for Hydrological Analysis Using Level-Set Method. *Journal of the American Water Resources Association*, *55*(2), 354–368. https://doi.org/10.1111/1752-1688.12689

Yakob, L., Dunning, R., & Yan, G. (2011). Indoor residual spray and insecticide-treated bednets for malaria control: theoretical synergisms and antagonisms. *Journal of the Royal Society, Interface*, *8*(59), 799–806. https://doi.org/10.1098/rsif.2010.0537

Yamana, T. K., & Eltahir, E. A. (2010). Early warnings of the potential for malaria transmission in rural Africa using the hydrology, entomology and malaria transmission simulator (HYDREMATS). *Malaria Journal*, *9*(1), 1–10. https://doi.org/10.1186/1475-2875-9-323

Yang, J., & Chu, X. (2015). A new modeling approach for simulating microtopography-dominated , discontinuous overland flow on infiltrating surfaces. *Advances in Water Resources*, *78*, 80–93. https://doi.org/10.1016/j.advwatres.2015.02.004

Yasarer, L. M. W., Bingner, R. L., & Momm, H. G. (2018). Characterizing ponds in a watershed simulation and evaluating their influence on streamflow in a Mississippi watershed. *Hydrological Sciences Journal*, *63*(2), 302–311. https://doi.org/10.1080/02626667.2018.1425954

Yé, Y., Hoshen, M., Kyobutungi, C., Louis, V. R., & Sauerborn, R. (2009). Local scale prediction of plasmodium falciparum malaria transmission in an endemic region using temperature and rainfall. *Global Health Action*, *2*(1), 1923. https://doi.org/10.3402/gha.v2i0.1923

Yu, C. W., Hodges, B. R., & Liu, F. (2020). A new form of the Saint-Venant equations for variable topography. *Hydrology and Earth System Sciences*, *24*(8), 4001–4024. https://doi.org/10.5194/hess-24-4001-2020

Zeng, L., Shao, J., & Chu, X. (2020). Improved hydrologic modeling for depression-dominated areas. *Journal of Hydrology*, *590*, 125269. https://doi.org/10.1016/j.jhydrol.2020.125269

Zhang, W., & Cundy, T. W. (1989). Modeling of two-dimensional overland flow. *Water*

Resources Research, *25*(9), 2019–2035. https://doi.org/10.1029/WR025I009P02019

Zhou, G., Minakawa, N., Githeko, A. K., & Yan, G. (2004). Association between climate variability and malaria epidemics in the East African highlands. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(8), 2375–2380. https://doi.org/10.1073/pnas.0308714100

Zhu, D., Ren, Q., Xuan, Y., Chen, Y., & Cluckie, I. D. (2013). An effective depression filling algorithm for DEM-based 2-D surface flow modelling. *Hydrology and Earth System Sciences*, *17*(2), 495–505. https://doi.org/10.5194/hess-17-495-2013

Zou, L., Miller, S. N., & Schmidtmann, E. T. (2006). Mosquito larval habitat mapping using remote sensing and GIS: Implications of coalbed methane development and West Nile virus. *Journal of Medical Entomology*, *43*(5), 1034–1041. https://doi.org/10.1603/0022-2585(2006)43[1034:MLHMUR]2.0.CO;2

Appendix A

Supporting Information for Chapter 2

A.1 Supplementary text

A.1.1 Input data

Using GRASS 7.6, the 1 arc-second DEM from SRTM was resampled to 50 m grid resolution. Subsequently, the resampled DEM data was hydro-conditioned to ensure that the drainage networks are connected using the global slope enforcement approach (M. L. Barnes et al., 2016) and converted to ground surface slopes as an input to ParFlow. The land cover type data used in CLM was determined by 30 m resolution Landsat-8 data, which was first filtered by cloud cover and processed with an unsupervised classification approach using k-mean clustering into ten classes in ENVI. The ten classes were resampled to 50 m resolution and then grouped into the IGBP land cover types and checked against Google Earth images for consistency. The dominant IGBP types for the study area are cropland and grassland.

To characterize the subsurface, the soil taxonomy distribution (Figure A.2) for the top 2 m from the surface was referenced from the SoilGrids250m TAXOUSDA dataset (Hengl et al., 2017) and ranked by their relative permeability. Each soil type was then assigned a saturated hydraulic conductivity within the range of 0.0015-0.015 m/hour (United States Department of Agriculture. Soil Science Division Staff, 2017) characteristic of either clay or clay loam. The saturated hydraulic conductivity of the deeper zone beyond the top 2 m was assigned an averaged value of 0.11 m/h based on GLHYMPS 2.0. The depth to bedrock data from SoilGrids250m BDRICM dataset (Hengl et al., 2017) was used to delineate the bedrock zone and the hydraulic conductivity was set to 0.00001 m/h to render the corresponding grid cells impermeable.

For the climate forcing, 0.04 degree by 0.04-degree precipitation data was resampled to the model grid using bilinear interpolation while the rest of the forcing data fields were relatively coarser at either 0.25 degree by 0.25 degree or 0.5 degrees by 0.625 degrees so an

average value was used for the entire domain. All the forcing data were pre-processed in NCAR Command Language (NCL) and input to the model hourly.

The list of model input data can be found in Table A.1.

A.1.2 Model simulation

In this study, a baseline was generated over the year of 2018 followed by a scenario with the implementation of a synthetic irrigation scheme during the dry season. The watershed model spans an area of 208 km^2 and has a high resolution, both spatially and temporally. The model resolution is 50 m and there are 332 by 248 cells. The model was divided into 10 subsurface layers and the thickness of each layer varies, decreasing gradually towards the surface. For the irrigation scenario, the sugarcane plantation in the study area was grouped into four farms. The soil textural classes of the study area are clay and heavy clay soils and the peak admissible infiltration rate of these soils was 4mm/hour. The efficiency of sprinkler irrigation was estimated at 75%. Hence, a 10 mm/day irrigation operational module using the hose-moving sprinkler system was applied on alternate farms. The daily irrigation time was 22 hours per day and the irrigation interval of 10 days proposed by sugar factory agronomists was adopted.

The model was initialized in two stages. In the first stage, the CLM component was not activated and the initial water table was set at the surface without meteorological forcing. The simulation was run until the average water table reached 2 m below the land surface. During this process, the water drains from the mountainous areas to valleys, forming rivers and streams naturally. Next, the CLM component was activated and the transient meteorological forcing from the year of 2017 was repeated three years for the model to reach a sufficient state of equilibrium.

All model simulations were conducted on the High-Performance Computing Cluster at the University of California, Irvine. A 1-year simulation required 88 CPU cores and about 12 days to complete.

Figure A.4 shows the time series of precipitation, temperature, and simulated surface layers soil saturation for the baseline simulation and streamflow at points A, B, and C (see Figure 2.1).

Figure A.5 compares the time series of the simulated surface layers soil saturation for the baseline and irrigation simulations. In general, the surface layer soil saturation fluctuated in tandem with precipitation. Irrigation during the dry season increased the surface layer soil saturation and the pattern of fluctuation was influenced heavily by the alternating irrigation scheme. Also, the surface layer soil saturation varied depending on the location. For instance, Point B which is in Farm 2 had the highest soil saturation and was near saturation point most of the time from August to September because it is located close to a small stream. The baseline results show that the surface soil saturation of the four locations was relatively low in the dry months (January to April), ranging from 5% to 15% but could go up to as high as 100% in the rainy season (May to October). In the scenario with irrigation, the soil saturation increased two-fold to about 50 in the dry months. This contributes to surface water detention, which can potentially generate more breeding sites. As the irrigation ended in April with the onset of the rainy season, the increase in the saturation of the irrigation scenario over the baseline gradually reduced until they were both the same after more than a month.

A.1.3 Soil saturation threshold θ parameterization

A.1.1.1 Aquatic habitat survey

The aquatic habitats were surveyed during the dry (December 2017–February 2018) and rainy (June 2018–August 2018) seasons (Hawaria et al., 2020). Mosquito larvae were sampled following the WHO standard larval survey procedure using a standard dipper (350 ml). Larvae were identified morphologically and sorted by genus as Anopheles and Culex in the field. Environmental variables such as habitat type, habitat dimension and land cover type were recorded using the Android-based tablet PC with ODK Collect application (Hartung et al., 2010) while the coordinates of each surveyed aquatic habitat were captured with the built-in Geographic Positioning System (GPS) sensor. All the data were eventually

uploaded onto the ODK Aggregate Server (Hartung et al., 2010) on the cloud-based Amazon Web Services (AWS).

Post survey, all the aquatic habitats and larvae survey metadata were pulled out from the ODK Aggregate (Hartung et al., 2010) MySQL database (version 5.7, Oracle Corporation, California, USA) and analyzed with JMP (version 14, SAS Institute, North Carolina, USA) and Microsoft Excel (Version 2019, Microsoft Corporation, Washington, USA). Spatial data aggregation, analysis, and visualization were produced with ArcGIS Pro 2.5 (ESRI, 2020) The locations of the surveyed aquatic habitats are shown in Figure 2.3.

A.1.1.2 Calibration and validation

The soil saturation threshold θ for the rainy season from May to October of the baseline scenario was calibrated using the survey data in the same period to minimize the influence of dry season irrigation on the parameterization. This is because irrigation was not accounted for in the baseline scenario and was only approximated by a simplified scheme in the irrigation scenario. Of the 134 samples for the year 2018, some of the surveyed aquatic habitats such as man-made ponds, tire track puddles, and animal footprints which could not be simulated by the hydrologic model were omitted, leaving 102 samples for calibration and validation as shown in Figure 2.3.

The objective of the calibration was to maximize the POD, which determines if the model can predict an aquatic habitat successfully. Other measures which can capture overprediction were not chosen here as the field data only cover locations with ponding and it is challenging to definitively rule out small puddles within the grid cell using satellite data. As shown in Equation (A.1), the POD was calculated based on the ratio of the number of successful predictions or hits, H, to the total number of samples, S. In a successful prediction, the WI is at least 1 day given that an aquatic habitat is sampled at the same time and location. The POD ranges from 0 to 1, with 1 corresponding to perfect detection.

$$POD = \frac{H}{S} \tag{A.4}$$

As the same dataset was used for both calibration and validation, 75% of the samples were randomly selected for the former and 25% for the latter. To ensure the relevance of the

calibrated saturation threshold, bootstrapping was performed by randomly resampling the same dataset based on the 75:25 ratio with replacement to derive 1000 combinations as a representation of all the possible combinations. Each combination was used to calculate the POD corresponding to the saturation threshold within a predefined range of 0.4 and 1, at intervals of 0.02. To determine and assess the reliability of the optimal threshold, the average POD, and the associated 95% confidence interval values were computed for each saturation threshold.

Figure A.6 shows the results from the calibration and validation of the bootstrap samples. It can be observed that the average POD of the validation was consistent with that of the calibration for each saturation threshold. The confidence interval for validation was wider than that of calibration due to the smaller number of survey samples used. The average POD for both the calibration and validation samples first reached 1 when the saturation threshold value decreased to 0.48. At the same time, the confidence intervals decreased to close to zero, indicating high reliability. Hence, the optimal saturation threshold was set as 0.48 to ensure that all the observed aquatic habitats were captured in the model.

A.2 Supplementary figures

Figure A.1: Typical rain-fed ponding in the study area in Arjo. The study site was in the valley bottom which is characterized by clay and clay loam with low permeability soil with high annual precipitation around 1500 mm which mainly distributed in the rainy season from May to October. This rainfall distribution and irrigation during the dry season created many long-term surface water storages that provide potential mosquito habitats.

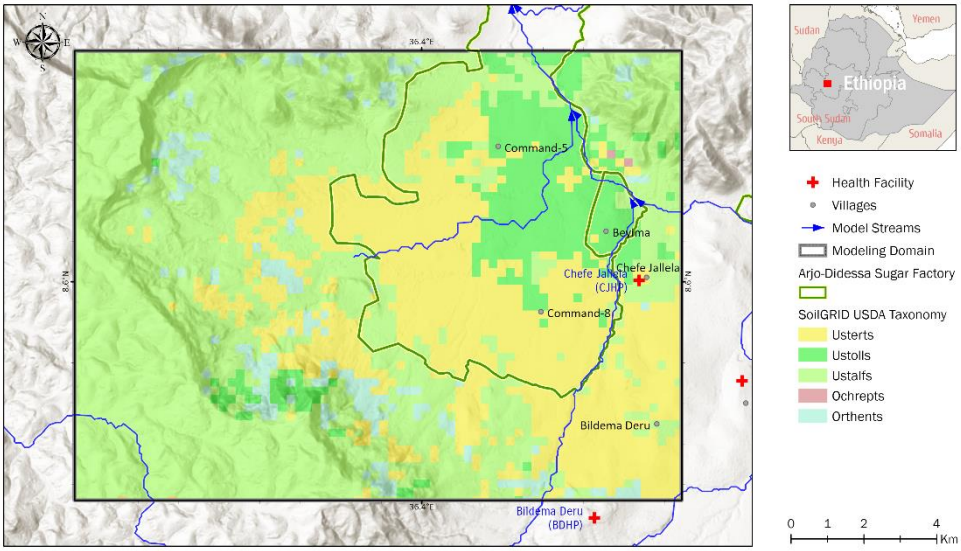


Figure A.2: The distribution of the top 2mm soil types in USDA soil taxonomy from SoilGrids250m TAXOUSDA dataset. Most soil types in this area are characterized as clay or clay loam with low permeability ranging from 0.0015 to 0.015 m/hour.

(a)	1st Year												2nd Year												1st Ratoon			
Mn	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4
Sym.	MA	H	H	LW	P	P	P	P	RF	RF	IR	IR	IR	IR	IR	IR	RF	RF	RF	RF	RF	RF	MA	MA	MA	H	IR	IR

(b)	1st Ratoon												2nd Ratoon												Virgin Planting			
Mn	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4
Sym.	MA	H	IR	IR	RF	RF	RF	RF	RF	RF	MA	MA	MA	H	IR	IR	RF	RF	RF	RF	RF	RF	MA	MA	MA	H	H	LW

Figure A.3: Arjo-Didessa Sugar Factory sugarcane plantation irrigation schedule. A normal sugar planting schedule (a) includes a 2-year cycle for virgin and (b) 1-year cycle for following (2~8) ratoons. The symbol for each stage: MA: Maturity/water withdraw; H: Harvesting; LW: Land work; P: Planting; RF: Rainfed; IR: Irrigation.

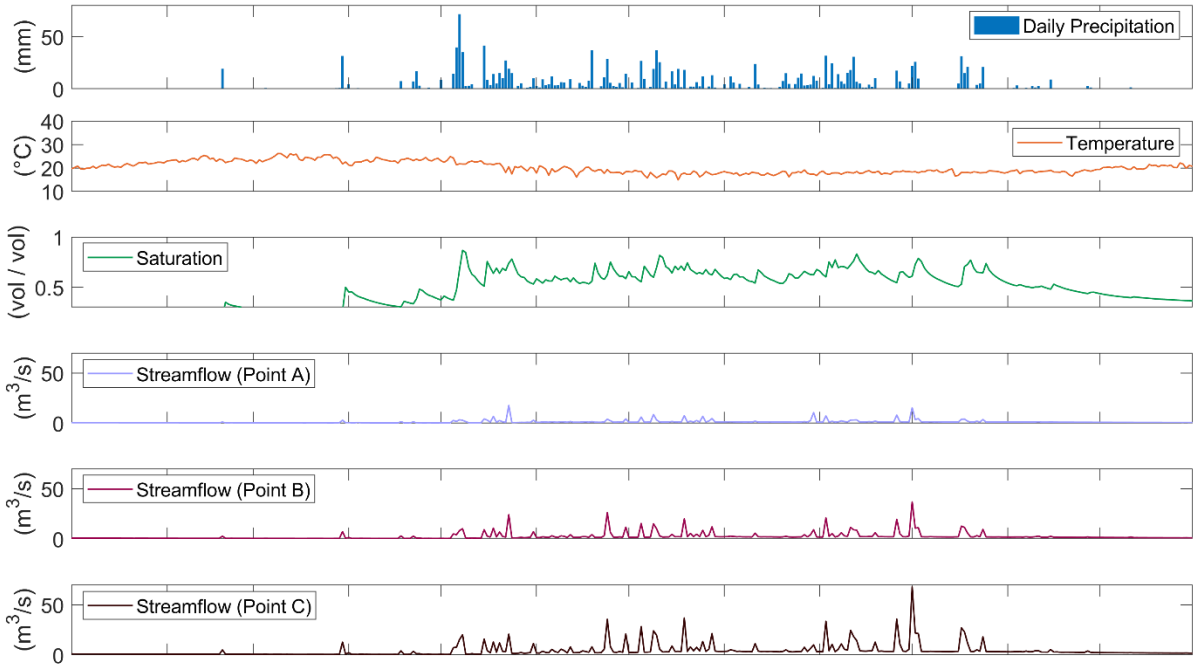


Figure A.4: Time series of precipitation, temperature, and simulated surface layers soil saturation for the baseline simulation and streamflow at point A, B, and C (see Figure 2.1) All the time series were spatially-averaged, except streamflow.

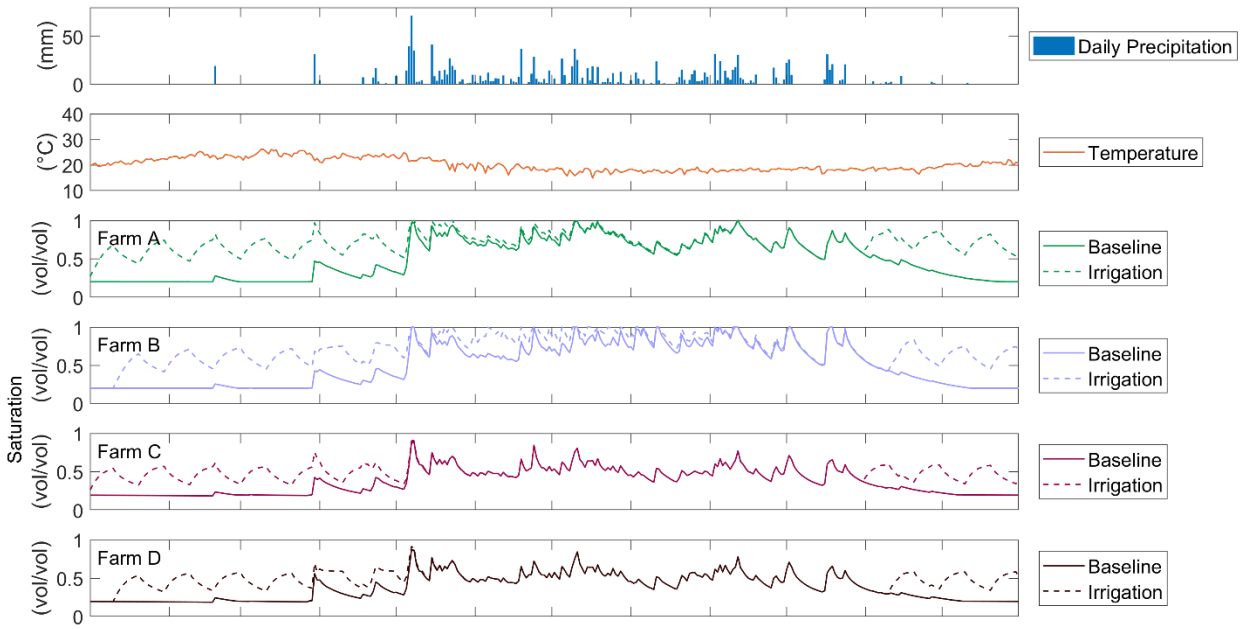


Figure A.5: Time series of precipitation, temperature, and simulated surface layers soil saturation for the irrigation scenario at a specific point of each of the four farms (see Fig 1). The precipitation and temperature profiles were spatially averaged. The soil saturation profile reflects the 2-cycle rotational irrigation schedule whereby Farm #1 and Farm #3 were irrigated first followed by Farm #2 and Farm #4.

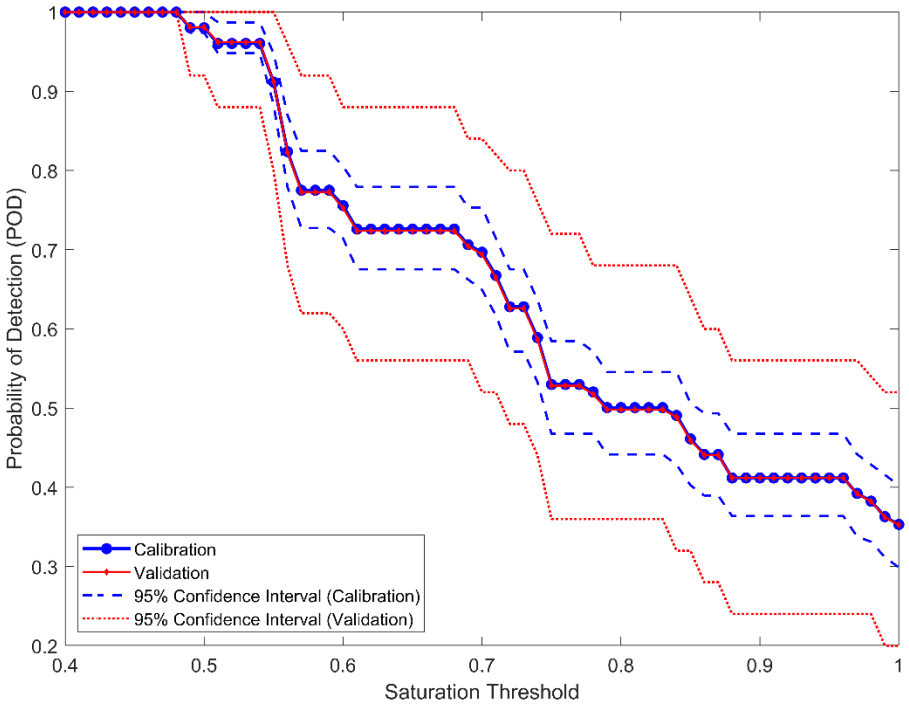


Figure A.6: Model performance from the calibration and validation of the saturation threshold. The 75% of the samples were selected for calibration and the other 25% for validation and the process was repeated 1000 times. The mean POD values from the calibration and validation are represented by the blue line with a circle marker and red line with a cross marker, respectively. The corresponding 95% confidence intervals are indicated by the dotted blue and red lines, respectively.

A.3 Supplementary tables

Table A.1: ParFlow-CLM model input data and their properties.

Variable	Resolution	Input Type	Latency	Source
Topography	30-meter	Distributed	-	SRTM Version 3.0 (JPL, 2013)
Precipitation	0.04°×0.04°, 1-hourly	Distributed	~1 hour	PERSIANN-CCS (Hong et al., 2004; Nguyen et al., 2019)
Short-Wave radiation	0.25°×0.25°, 3-hourly	Distributed	4.5-5.5 month	GLDAS (Rodell, 2004)
Long-Wave radiation	0.25°×0.25°, 3-hourly	Distributed	4.5-5.5 month	GLDAS
Air Temperature	0.25°×0.25°, 3-hourly	Distributed	4.5-5.5 month	GLDAS
Atmospheric Pressure	0.25°×0.25°, 3-hourly	Distributed	4.5-5.5 month	GLDAS
Water-vapor specific humidity	0.25°×0.25°, 3-hourly	Distributed	4.5-5.5 month	GLDAS
North-to-South Component of Wind Speed	0.5°× 0.625°, 1-hourly	Distributed	4.5-5.5 month	MERRA-2 (Gelaro et al., 2017)
East-to-West Component of Wind Speed	0.5°× 0.625°, 1-hourly	Distributed	4.5-5.5 month	MERRA-2 (Gelaro et al., 2017)
Land use	30-meter 16-day	Distributed	~12 hours	Landsat 8 (United States Geological Survey Earth Resources Observation and Science Center, 2013)

Soil type	250-meter	Distributed	-	SoilGrids250m, TAXOUSDA (Hengl et al., 2017)
Depth to Bedrock	250-meter	Distributed	-	SoilGrids250m, BDRICM (Hengl et al., 2017)
Near Surface Permeability (< 100 m)	Regional Scale	Distributed	-	GLHYMPS, 2.0 (Gleeson et al., 2014)

Appendix B

Supporting Information for Chapter 3

B.1 Supplementary text

B.1.1 Depression hierarchy

An example of depression hierarchy is illustrated by the binary tree representation in Figure B.1. Each node represents a nested depression which aggregates to form a compound depression. Nodes 1 and 2 from Level 1 fill the Node 3 to form a larger Level 2 depression. When Node 3 overflows along with Node 4, they begin to fill Node 5 up to Level 3. There can be other depressions like Node 6 that do not contain nested depressions.

B.1.2 Sensitivity analysis on threshold selection

The vertical accuracy of the original 1-m resolution LiDAR DEM is 0.09 m in terms of the root mean square error which means that the vertical error of the DEM is about 0.18 m at 95% confidence interval. Resampling the DEM to a lower resolution has the effect of reducing the vertical area by the scaling factor. For example, by error propagation analysis, the vertical error of the resampled 10-m resolution DEM is 0.018 m. In theory, the most conservative approach is to set the depth threshold to be greater than the vertical error. However, the depressions are characterized by an average depth (Equation (3.1)) so the vertical error will result in an underestimation of the number of depressions if set directly as the depth threshold. At best, it can only serve as a rough guide on the order of magnitude for the threshold if there is no other information.

As an alternative, we use the NWI data to evaluate the performance of using a combination of different thresholds for depth and area in a sensitivity analysis. Similar to Li et al. (2011), the performance was quantified by a sensitivity metric and a specificity metric as shown below. The higher the sensitivity and specificity, the better the performance of the tested threshold. Ideally, an appropriate threshold should be set low enough so that D2P can

match a good number of depressions but not too low as to compromise the ability of D2P to distinguish depressions likely to be true from artifacts.

$$\text{Sensitivity} = \frac{number\ of\ matched\ depressions}{number\ of\ depressions\ identified\ by\ the\ D2P\ algorithm} \times 100\%$$

$$\text{Specificity} = \frac{number\ of\ matched\ depressions}{number\ of\ depressions\ in\ the\ NWI\ dataset} \times 100\%$$

Figure B.2 shows the sensitivity and specificity matrices for different depth and area thresholds applied on the 10-m resampled DEM. We narrow down our selection to a threshold combination ranging from 800 m^2 to 1200 m^2 in area and 0 m to 0.0036 m in depth whereby both sensitivity and specificity exceed 45%. Generally, sensitivity is lower compared to specificity across all the combinations which is reasonable because D2P identifies both wet and dry depressions whereas the benchmark NWI data only records ponds or wetlands. Hence, we prioritized specificity over sensitivity in selecting the threshold when the combined value is the same. Finally, we chose an area threshold of 900 m2 to maximize specificity. For the selected area threshold, both sensitivity and specific were indifferent to the depth threshold between 0 m and 0.0036 m so the upper limit was selected since the depressions with average depth below 0.0036 m may not be of significance to the hydrologic simulation.

In addition, we also evaluated the sensitivity and specificity of the threshold for the original DEM resolution of 1 m and the other resampled resolutions of 5 m and 20 m in Figure B.3 to Figure B.5. For the same corresponding thresholds, the sensitivity and specificity generally increased with resolution. Notably, both the sensitivity and specificity were able to exceed 55% for both the 1-m and 5-m resolutions. This occurred in the 1-m resolution for 5 combinations of area and depth thresholds namely, 1200 m^2, 0 m and 0.0009 m; 900 m^2, 0.09 m; 800 m^2, 0.09 m and 0.18 m. For the 5-m resolution DEM, the optimal threshold was 800 m^2 and 0.09 m. At 20-m resolution, the performance of all the threshold combinations was poor, with a sensitivity less than 25% and specificity less than 20%. This indicated that the 20-m resolution DEM was not suitable for identifying depressions in the study area.

Li et al., 2011 stated that higher resolution DEMs typically require larger area threshold as there is a higher probability of small artifactual depressions that need to be removed. However, the same trend was not observed in our study. Based on this analysis, the optimal area threshold across the resolutions of 1 m, 5 m, 10 m were all around 800 m2 to 900 m^2. A possible reason could be that the NWI information may not be up to date, thereby distorting the number of matched depressions in both sensitivity and specificity. On the other hand, the study by Li et al. identified wetlands using a combination of aerial photography and field survey.

B.1.3 Metrics for evaluation of identified surface depressions

Each cell in the study area was classified as M_1 if it belongs to a depression identified by D2P and M_0 otherwise. Similarly, each cell in the same study area was classified as B_1 if it belongs to a pond delineated using the benchmark NAIP imagery and B_0 otherwise.

$$POD = \frac{M_1 \cap B_1}{M_1 \cap B_1 + M_0 \cap B_1}$$

$$FAR = \frac{M_1 \cap B_0}{M_1 \cap B_0 + M_0 \cap B_0}$$

$$CSI = \frac{M_1 \cap B_1}{M_1 \cap B_1 + M_1 \cap B_0 + +M_0 \cap B_1}$$

M_0 & M_1 : Binary classification of depressions identified by D2P.

B_0 & B_1 : Binary classification of depressions identified using NAIP as benchmark.

B.1.4 Comparison of DEM processing method from Case 1 to Case 4

To highlight the differences across the cases, Figure B.6 shows a comparison of an example river network and river elevation profile for a river segment overlapping with a depression in the GCEW study area. Due to the difference in DEM processing methods, there is a slight difference in the river paths (Figure B.6a, Figure B.6c, Figure B.6e and Figure B.6g) but the river density was set to be the same in all cases for consistency. From the river profile elevation in Figure B.6b, the slope within each reach is defined by the elevation at the upstream and downstream ends in Case 1 after the depressions are filled. In Case 2 (Figure

B.6d), the depressions are first filled, similar to Case 1, but the slope varies locally within each reach to follow the depression-less profile more closely. In Case 3 (Figure B.6f), the adapted CRS algorithm is applied directly before filling, and this has the effect of removing the depressions overlapping with the river by breaching. Case 4 implements the adapted CRS algorithm with different constraints for depressions and non-depression segments, so the depressions are preserved along with the downstream structure such as the dam as shown in Figure B.6h.

B.2 Supplementary figures

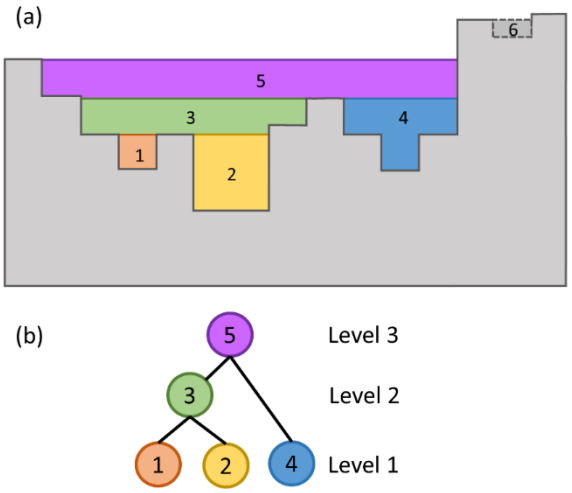


Figure B.1: Illustration of (a) depression hierarchy and (b) corresponding binary tree representation.

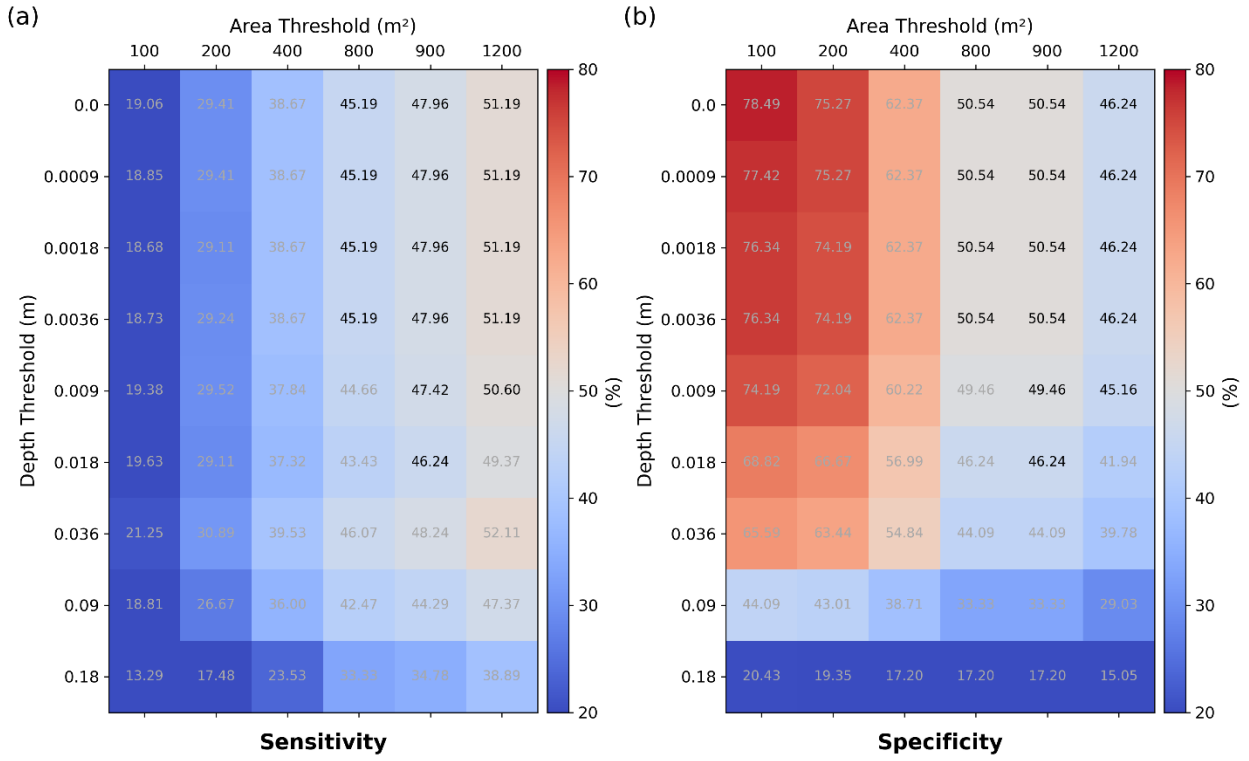


Figure B.2: Specificity and sensitivity matrices for identified surface depressions based on 10-meter resolution DEM. Values greater than 45% in both matrices are in black.

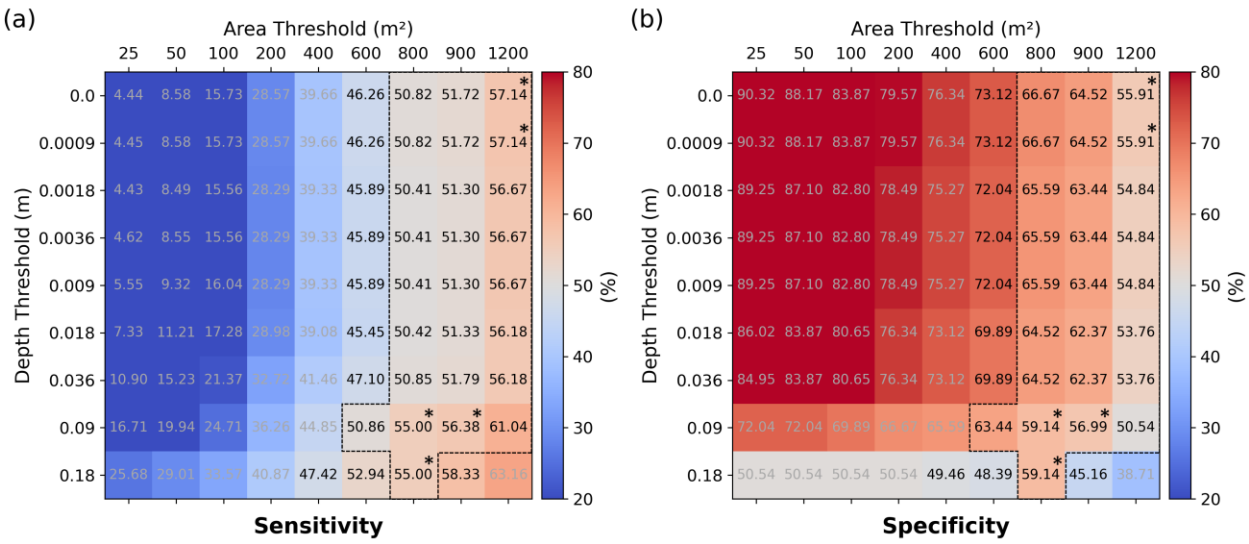


Figure B.3: Specificity and sensitivity matrices for identified surface depressions based on 1-meter resolution DEM. Values greater than 45% in both matrices are in black, values greater than 50% are demarcated by black dashed line, and values greater than 55% are denoted by asterisk.

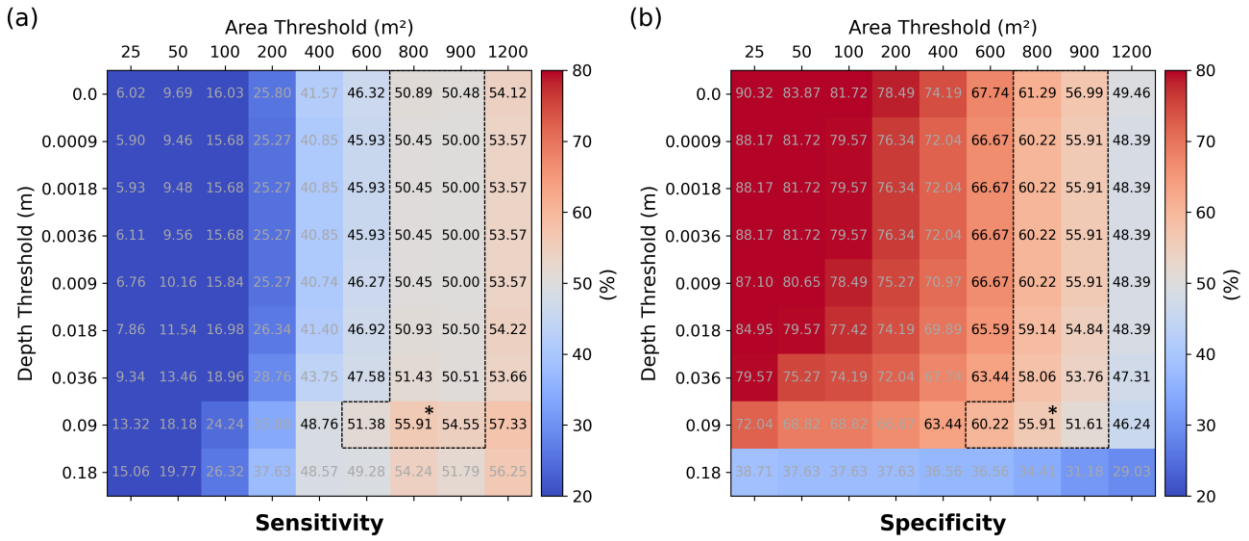


Figure B.4: Specificity and sensitivity matrices for identified surface depressions based on 5-meter resolution DEM. Values greater than 45% in both matrices are in black, values greater than 50% are demarcated by black dashed line, and values greater than 55% are denoted by asterisk.

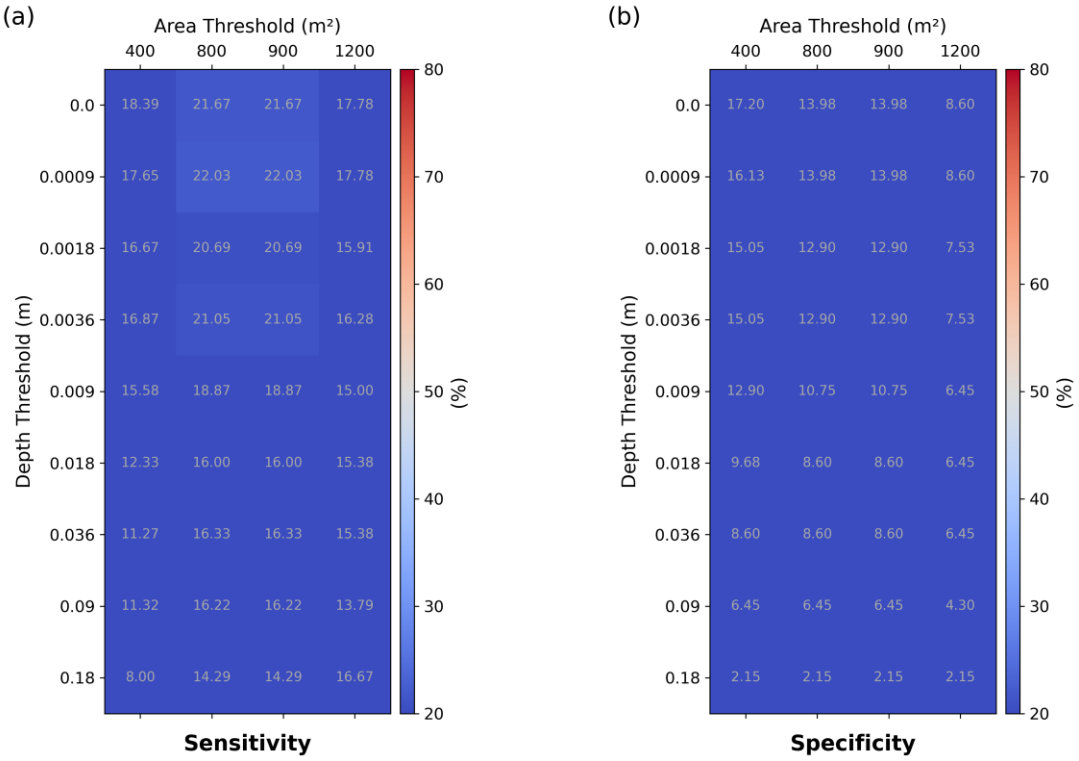


Figure B.5: Specificity and sensitivity matrices for identified surface depressions based on 20-meter resolution DEM. Values greater than 45% in both matrices are in black, values greater than 50% are demarcated by black dashed line, and values greater than 55% are denoted by asterisk.

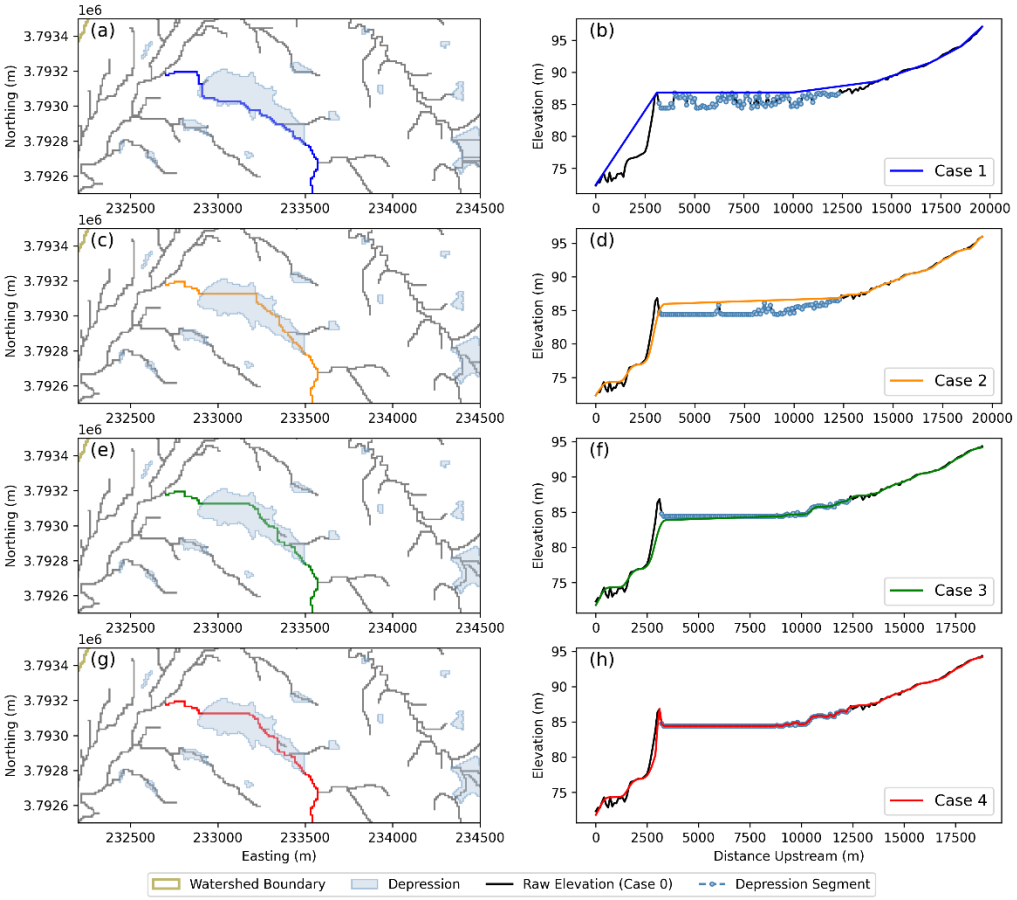


Figure B.6: Comparison of the river network (left column) and river elevation profile (right column) for a river segment overlapping with a depression for (a)(b) Case 1, (c)(d) Case 2, (e)(f) Case 3 and (g)(h) Case 4.

Appendix C

Supporting Information for Chapter 4

C.1 Supplementary text

C.1.1 Larval density estimation

In the study area, the surveyed larval habitats include drainage ditch, river edge/reservoir shoreline, swamp/marsh, rice puddle, animal footprint, tire track/road puddle, man-made pond, natural pond/rain pool, rock pool, water container, irrigation canal, and brick pit. The larval habitats were classified as temporary, semi-permanent or permanent habitats. Since larval density can be significantly different in the dry and rainy season (Hinne et al., 2021; Kweka et al., 2012) and the timing and duration of the survey periods were inconsistent, we sorted the measured larval densities from the 769 sample points (Figure 4.1) into dry season (January to April; November to December) and rainy season (May to October). We then calculated the average larval density for each season as shown in Figure C.3.

In the surveyed area, the larval density for temporary habitats was higher in rainy season than dry season during which the habitats are less stable. On the other hand, the larval density for semi-permanent and permanent habitats were higher in the dry season. Most of the semi-permanent and permanent were associated with river edges and swamps whereby the larvae are prone to flushing in the rainy season. Finally, the larval density in Table 4.1 was calculated based on the average of the dry season and rainy season densities.

C.1.2 Irrigation scheme design

Figure C.4 shows the monthly irrigation schedule obtained from the Arjo-Didessa Sugar Factory which is tailored to the sugarcane planting cycle.

To model irrigation in ParFlow-CLM, the irrigation interval and rate are required user inputs. For the irrigation interval, a report on the design of the local irrigation system recommended an interval of 8-12 days so we set the interval as 10 days.

To determine the irrigation rate, we first calculated the irrigation depth, which is defined as the amount of water that needs to be applied when soil water content is depleted to the wilting point. The irrigation depth ($IrrD$) was calculated as

$$IrrD = (FC - WP) \times Depth_{soil}, \quad \text{(C.1)}$$

where FC is the field capacity, WP is the permanent wilting point and $Depth_{soil}$ is the soil depth.

A soil depth of 2 m was assumed. The study area is characterized by clay and clay loam with low permeability. Based on resources by the Northeast Region Certified Crop Advisor (https://nrcca.cals.cornell.edu/soil/CA2/CA0212.1-3.php), the field capacity volumetric soil moisture content of clay was set as 50% and the wilting point volumetric soil moisture content was set as 15%. An irrigation depth of 700 mm was obtained using Equation (C.1). As we set the irrigation to be applied when 50% of the irrigation depth was depleted, the actual irrigation depth to be applied over the 10-day irrigation interval was 350 mm. Assuming an intermittent irrigation strategy, we set the irrigation to be applied for 22 hours in a day over a 3-day period within the 10-day cycle. The irrigation rate was then calculated to be 5.3 mm/hour.

C.1.3 Model calibration

At each grid cell, ponding is assumed to occur if the soil saturation exceeds the threshold, θ. The threshold was calibrated to ensure that the model will predict the occurrence of ponding at locations in line with the field surveyed larval habitats for soil saturation above θ. The value for θ was obtained based on a sensitivity analysis by altering the threshold and noting the corresponding change in the POD. The POD determines if the model can predict an aquatic habitat successfully and was calculated based on the ratio of the number of successful predictions or hits, H, to the total number of samples, S:

$$POD = H/S, \quad \text{(C.2)}$$

Figure C.6 shows the results of the sensitivity analysis. In general, the POD curve is higher for the simulation excluding dry season. This is because irrigation was only approximated by a simplified scheme in the dry season and may not resolve the localized irrigation dynamics. As the threshold was lowered, POD increased because ponding occurred across a larger area in the model. The ponding was highly governed by soil type and the influence of topography was less apparent. On the other hand, when the threshold was increased, less ponding was predicted resulting in a lower POD but the topographic variability was better represented. Since the survey only focused on areas with ponding, we do not have definitive information on locations without ponding in the study area. Hence, we selected a threshold of 0.75 for a reasonable POD of 0.66 (excluding dry season) without obscuring topographic variability.

In EMOD, we calibrated 10 key parameters identified from a sensitivity analysis and Table C.2 presents the calibrated values. Using the calibrated parameters, we compared the simulated prevalence rate against field data for January 2018 and October 2018 (Figure C.7). The results are within the same order of magnitude. We also compared the simulated monthly number of clinical cases with the recorded malaria cases from April 2018 to May 2020 (Figure C.8). Apart from the two peaks that were missed in October 2018 and November 2019, the simulated malaria cases compare reasonably well with observation both in terms of magnitude and dynamics. As the clinical malaria cases were sourced from major hospitals within the study area, the two peaks in recorded cases could be anomalous due to an influx of patients from outside seeking treatment at the hospitals. Overall, the model shows a good agreement with the field observation.

C.2 Supplementary figures

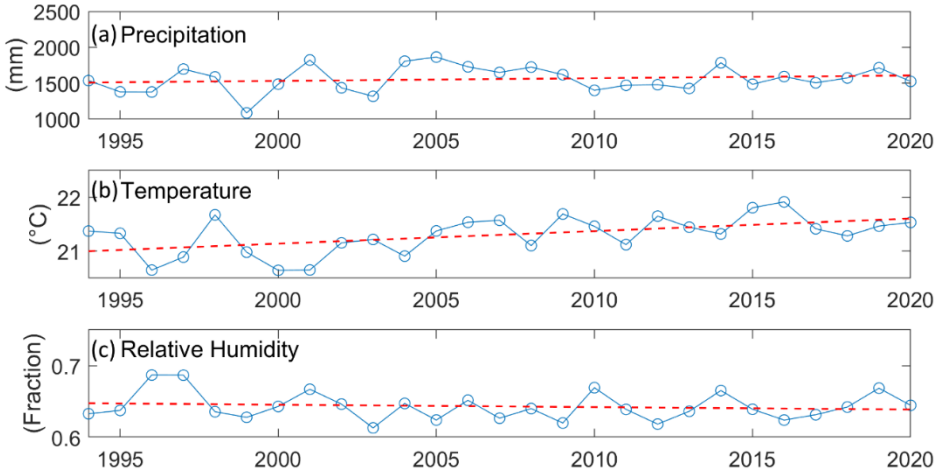


Figure C.1: Annual climate data from PERSIANN-CCS-CDR and ERA5 in the study area in Arjo. (a) total precipitation, (b) average temperature and (c) average relative humidity.

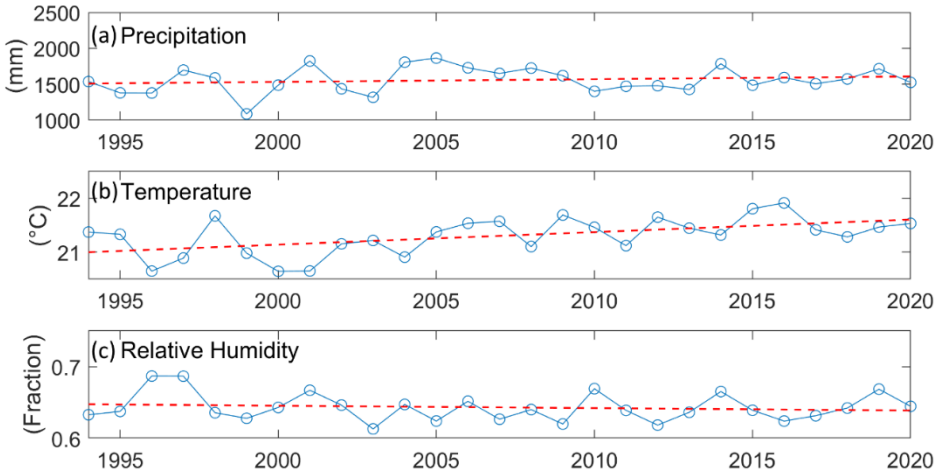


Figure C.2: Monthly climatology derived from PERSIANN-CCS-CDR and ERA5 climate data in the study area in Arjo. (a) total precipitation, (b) average temperature and (c) average relative humidity.

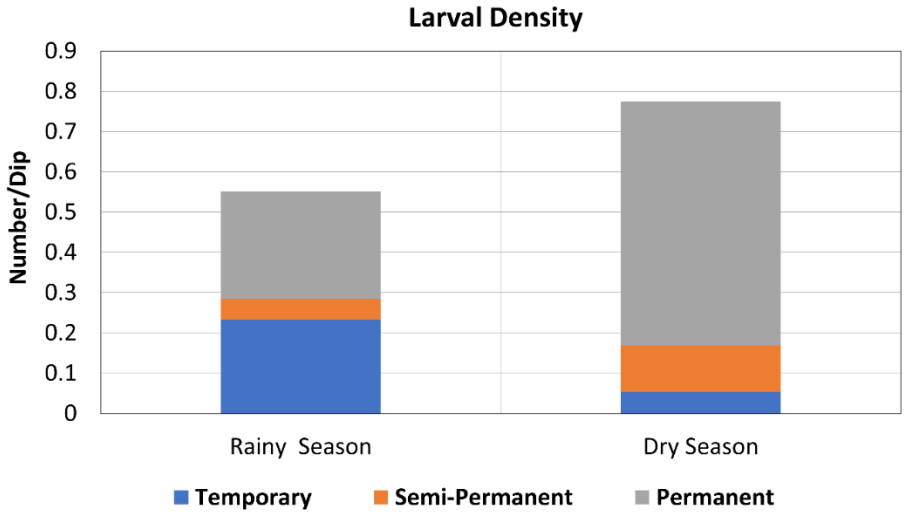


Figure C.3: Average larval densities for temporary, semi-permanent, and permanent habitats during rainy and dry seasons based on field survey.

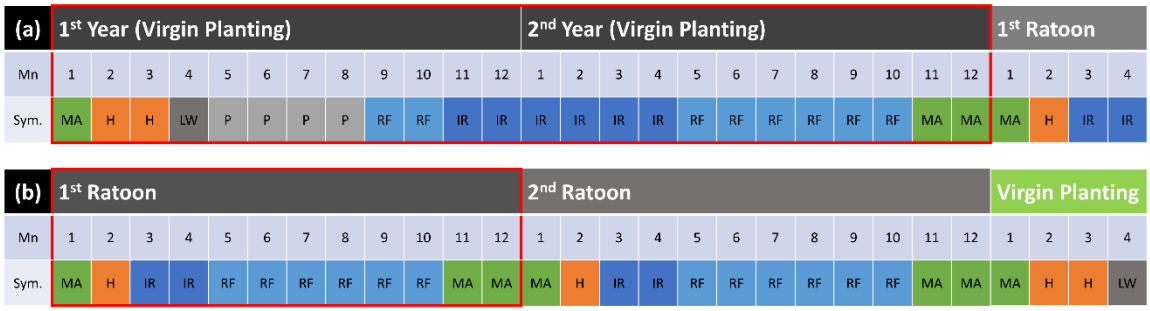


Figure C.4: Arjo-Didessa Sugar Factory sugarcane plantation irrigation schedule. A normal sugar planting schedule includes (a) a 2-year cycle for virgin and (b) a 1-year cycle for following 2 ratoons. MA: Maturity/water withdraw; H: Harvesting; LW: Land work; P: Planting; RF: Rainfed; IR: Irrigation.

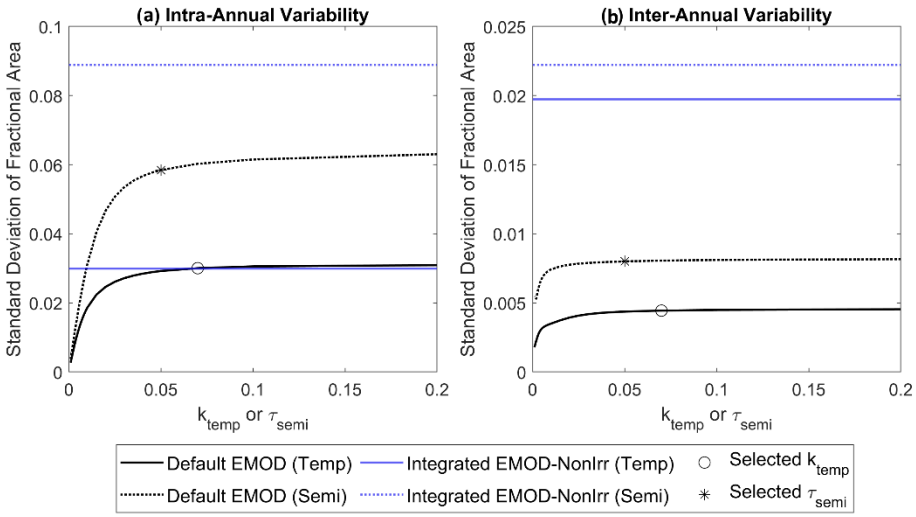


Figure C.5: Adjustment of decay factors for temporary and semi-permanent habitats in default EMOD function. The objective is to match the (a) intra-annual variability and (b) inter-annual variability of the habitats from the hydrologic model as closely as possible. Decay factor of semi-permanent habitat also must be smaller than that of temporary habitat.

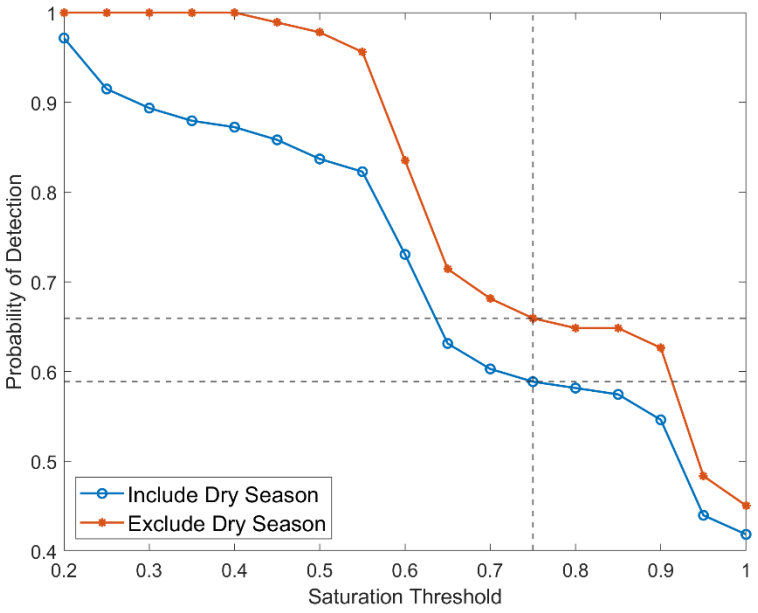


Figure C.6: Sensitivity analysis of saturation threshold to POD. The POD determines if the model can predict an aquatic habitat successfully and was calculated based on the ratio of the number of successful predictions to the total number of observations.

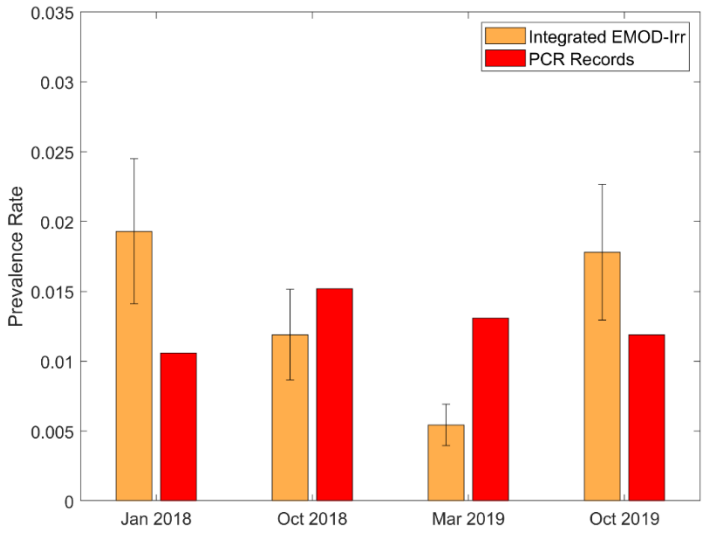


Figure C.7: Comparison of simulated monthly average prevalence rate with one standard error and sampled PCR. and was calculated based on the ratio of the number of successful predictions to the total number of observations.

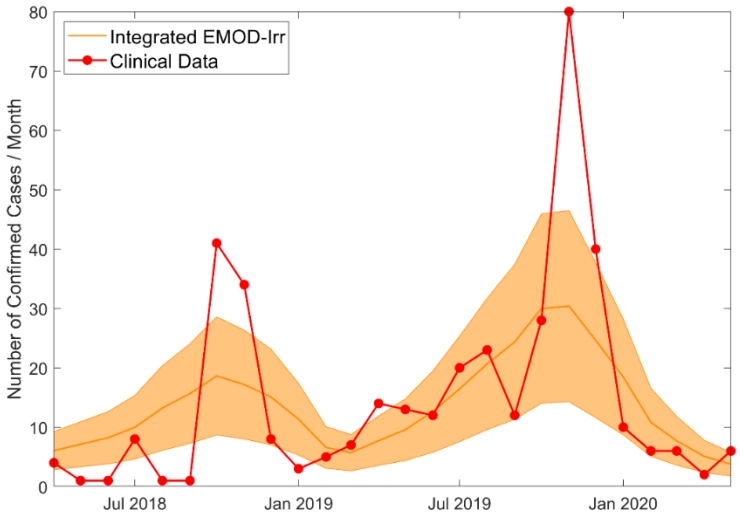


Figure C.8: Comparison of simulated monthly confirmed cases and clinical data. The 95% confidence interval is indicated by the orange band.

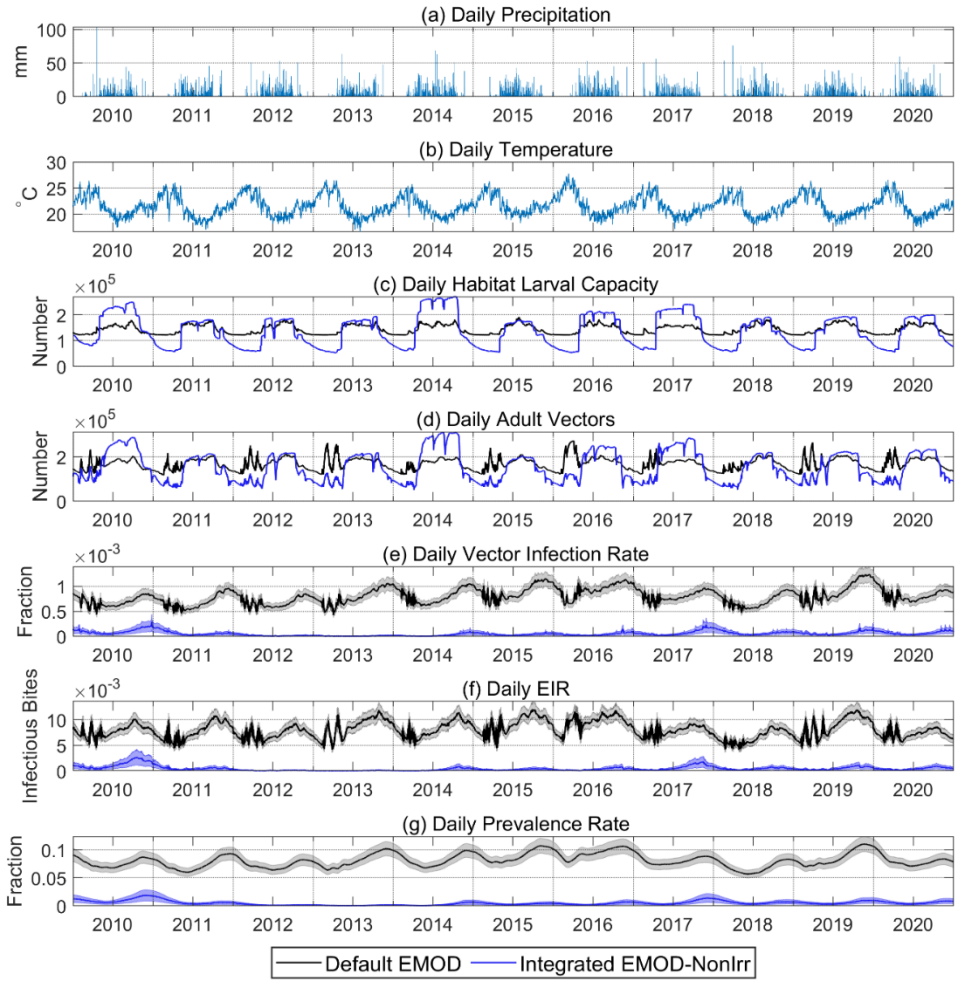


Figure C.9: Time series of daily climate data and comparison of simulated daily malaria transmission results between "Default EMOD" scenario and "Integrated EMOD-NonIrr" scenario. Climate data include (a) precipitation and (b) temperature. Malaria transmission results include (c) habitat larval capacity (d) adult vector abundance (e) adult vector infection rate (f) entomological inoculation rate and (g) parasite prevalence rate. The simulation was performed for a 20-year period from 2000 to 2020 but here we only show the results from 2010 to 2020 for simplicity.

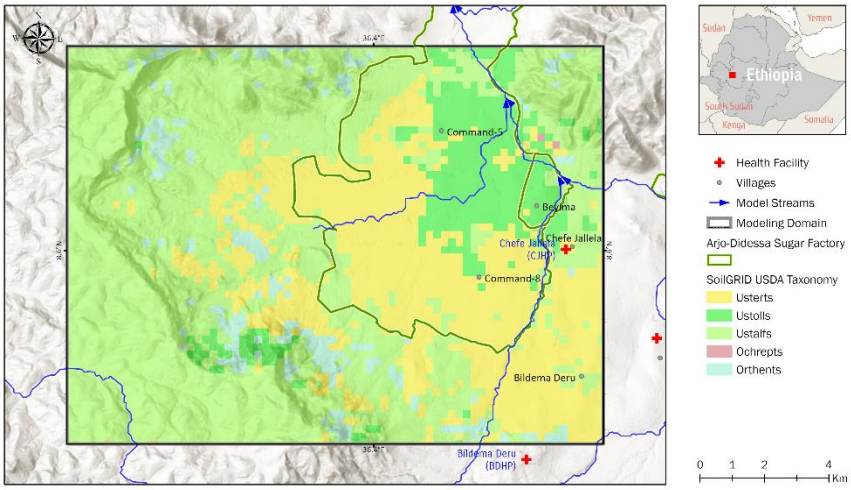


Figure C.10: The distribution of the top two-meter soil types in USDA soil taxonomy from SoilGrids250m TAXOUSDA dataset. Most soil types in this area are characterized as clay or clay loam with low permeability ranging from 0.0015 to 0.015 m/h.

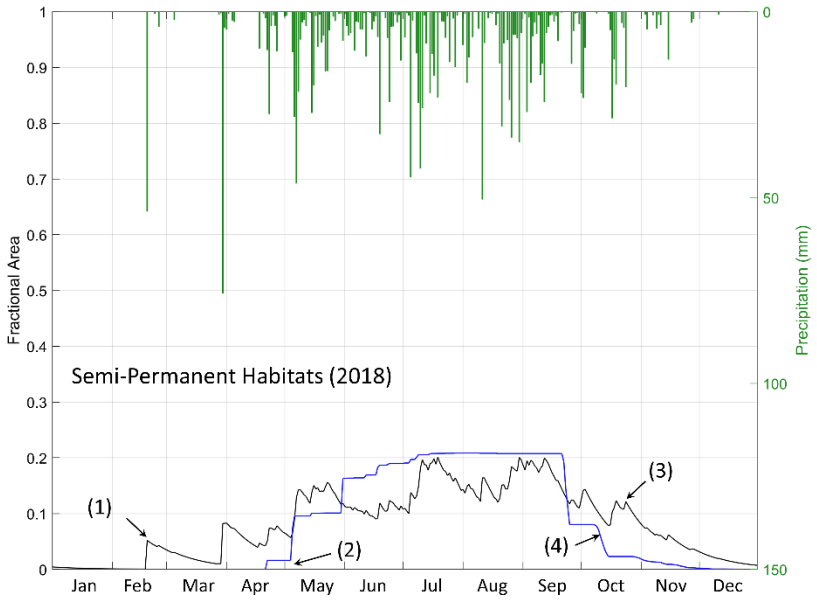


Figure C.11: Comparison of simulated semi-permanent habitats between "Default EMOD" scenario (black line) and "Integrated EMOD-NonIrr" scenario (blue line) in 2018. Earlier

rising limb in "Default EMOD" scenario: (1) no infiltration, new ponds created instantaneously by rainfall; (2) ponds formed some time after rainfall when soil saturation exceeded threshold. Delayed falling limb in "Default EMOD" scenario: (3) habitat area continued to increase with rainfall; (4) pond drained/dried up and soil became unsaturated after a period without rainfall, new rainfall insufficient to create ponding as soil saturation remained below threshold.

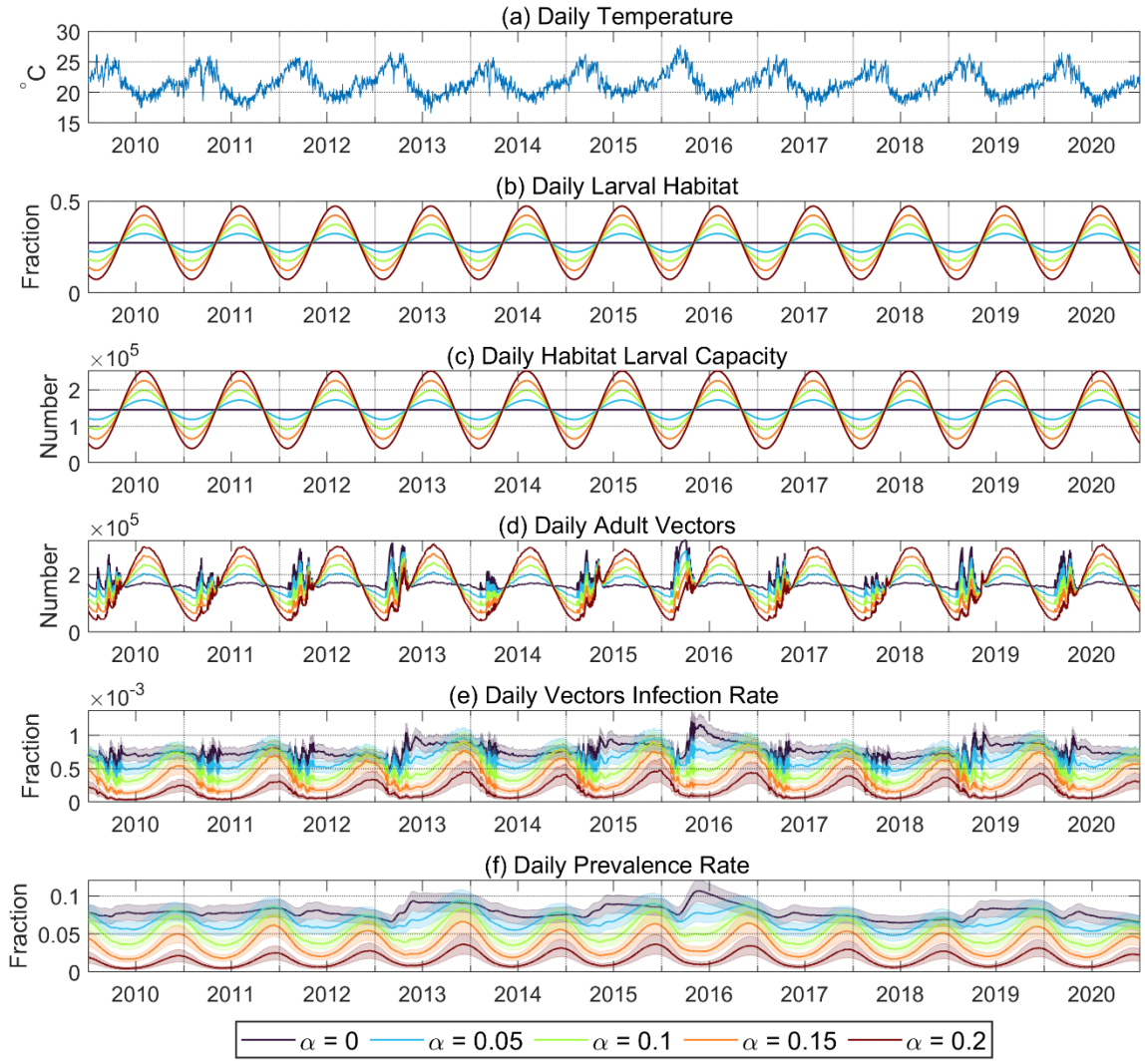


Figure C.12: Simulation results from sensitivity analysis of malaria transmission to different amplitudes of larval habitat seasonality, α. Time series include (a) daily temperature (b) synthetic sinusoidal larval habitat (c) habitat larval capacity (d) adult vector abundance (e) adult vector infection rate and (f) parasite prevalence rate.

C.3 Supplementary tables

Table C.1: Input data for ParFlow-CLM and EMOD.

Variable	Resolution	Input Type	Latency	Source
Topography	5-meter	Distributed	-	ALOS WORLD 3D Topographic Data (Takaku et al., 2016; Takaku & Tadono, 2017)
Precipitation	0.04°×0.04°, 3-hourly	Distributed	~1 hour	PERSIANN-CCS-CDR (Sadeghi et al., 2021)
Surface Solar Radiation Downwards	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5 (Hersbach et al., 2018, 2020)
Surface Thermal Radiation Downwards	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
Air Temperature (2m above ground surface)	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
Skin Temperature	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
Surface Pressure	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
Water-vapor specific humidity	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
North-to-South Component of Wind Speed (10m above ground surface)	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
East-to-West Component of Wind Speed (10m above ground surface)	0.25°×0.25°, 1-hourly	Distributed	5 days	ERA5
Land use (2000)	30-meter	Distributed	-	GlobeLand30 (Chen et al., 2015)
Land use (2010)	30-meter	Distributed	-	GlobeLand30

Land use (2020)	30-meter	Distributed	-	GlobeLand30
Soil type	250-meter	Distributed	-	SoilGrids250m, TAXOUSDA (Hengl et al., 2017)
Depth to Bedrock	250-meter	Distributed	-	SoilGrids250m, BDRICM (Hengl et al., 2017)
Near Surface Permeability (< 100 m)	Regional Scale	Distributed	-	GLHYMPS, 2.0 (Gleeson et al., 2014)

Table C.2: Calibrated parameters in EMOD.

Parameter	**Value**
Acquire_Modifier	0.25
Base_Gametocyte_Production_Rate	0.0615
Gametocyte_Stage_Survival_Rate	0.565
Base_Gametocyte_Mosquito_Survival_Rate	0.002
Antigen_Switch_Rate	1.2×10^{-10}
Aquatic_Arrhenius_1	85,884,000,000
Aquatic_Arrhenius_2	7,495
Infected_Arrhenius_1	121,726,800,000
Infected_Arrhenius_2	6,751
Scaling factor for larval capacity	1.15×10^{-7}

Table C.3: Spatial average of adult vector and parasite prevalence rate from the dry season (November 2016 to April 2017) and the rain season (May 2017 to October 2017).

Parameter	**Dry Season**		**Rainy Season**	
	Adult Vectors (# /km^2)	**Prevalence** (Fraction)	**Adult Vectors** (# /km^2)	**Prevalence** (Fraction)
Default EMOD	1.56×10^3	0.0895	1.78×10^3	0.0785
Integrated EMOD-NonIrr	1.01×10^3	0.0486	2.14×10^3	0.0505
Integrated EMOD-Irr	1.98×10^3	0.0718	2.53×10^3	0.1010