

DYNAMICS OF SEASONAL CROP YIELD PREDICTION UNDER WEATHER AND CLIMATE
EXTREMES

By

Abhijeet Abhishek

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Civil Engineering – Doctor of Philosophy

2023

ABSTRACT

The challenges of predicting seasonal crop yields amidst fluctuating weather and climate extremes is a pressing concern, given the increasing unpredictability brought on by global climate shifts and their socio-economic implications. This research delves into the multifaceted dynamics surrounding this challenge, placing focus on the profound implications of droughts, especially in vulnerable regions like Cambodia. Droughts, characterized by extended periods of scant precipitation, have historically been disruptors of agricultural systems. Their ripple effects often cascade through economies, underpinning socio-economic upheavals that stretch beyond agricultural boundaries. In the study's initial phase, Cambodia's agricultural patterns from 2000 to 2016 were closely examined. The results were notable: despite the recurrent and debilitating droughts, rice yields consistently rose. This anomaly was traced back to shifts in agricultural practices, particularly the heightened application rates of chemical fertilizers post-2008. This finding is both encouraging and cautionary, hinting at adaptive resilience in the face of adversity but also flagging the potential environmental implications of intensified chemical usage. The research then transitioned to an evaluation of the efficacy of seasonal climate forecasts in predicting interannual crop yields. The inherent uncertainties of such forecasts, their potential pitfalls, and their pivotal importance were all evaluated. By integrating hydrologic models with probabilistic forecasts, a novel methodology was formulated, aiming to bridge the gulf between historical data sets and real-time climatic shifts. This approach was tested across Cambodia, offering a comparative perspective that enriched the research's findings. In regions heavily reliant on rainfed agricultural systems, the value of precise, timely forecasting cannot be overstated. The unpredictability of rainfall patterns, exacerbated by climate change, places immense strain on these

systems, making accurate forecasting a backbone for effective agricultural planning. However, forecasting, no matter how advanced, is still beset with challenges. Additionally, the study revealed that among the myriad variables affecting crop yields, minimum air temperature and dry spells wielded the most significant impact, underscoring their critical role in agricultural yield dynamics. The subsequent phase of the research, therefore, ventured into data assimilation, exploring its potential in refining crop yield predictions. In essence, this comprehensive study not only sheds light on the intricate interplay between climate extremes and crop yield predictions but also charts a potential way forward. By blending traditional research methodologies with advanced technologies, it underscores the need for proactive, informed, and adaptable agricultural strategies. The findings from this research are expected to aid efforts aimed at achieving agricultural sustainability and bolstering food security in an era characterized by climatic uncertainties and evolving challenges.

ACKNOWLEDGEMENTS

I extend my sincere appreciation to my advisor, Dr. Mantha S. Phanikumar, for his invaluable mentorship and unwavering belief in my abilities. His consistent enthusiasm and dedication to academic research have been a guiding light throughout my Ph.D. journey.

I am deeply indebted to Dr. Narendra N. Das, my co-advisor, whose steadfast support, and expert guidance have been instrumental in steering my research in the right direction. His wisdom and patience have been a constant source of inspiration.

My sincere gratitude extends to my thesis committee members, Dr. Jeffrey Andresen, Dr. Yadu Pokhrel, and Dr. Alicia Sendrowski, for their invaluable comments, constructive suggestions, and unwavering support in reviewing and enhancing this thesis. In particular, Dr. Sendrowski played a pivotal role by providing invaluable assistance during the development of my second paper.

Furthermore, I wish to express my heartfelt appreciation to my former academic co-advisor, Dr. Amor V.M. Ines, whose early guidance, and insightful contributions played a pivotal role in shaping the foundation of my Ph.D. research.

I am deeply grateful to my lab mates, past and present, whose camaraderie, intellectual discussions, and collaborative spirit have enriched my research experience. Your insights and shared experiences have been an integral part of my academic journey.

To my family, including my parents, my aunt, and Ipsi, I extend my profound appreciation for their continuous support, unwavering encouragement, and boundless love. Your belief in me has been my greatest source of strength.

TABLE OF CONTENTS

Chapter 1.....	1
INTRODUCTION	1
1.1. Background	1
1.2. Research Objectives	7
1.3. Significance of the Research	8
1.4. Organization of Thesis	9
Chapter 2.....	10
EVALUATING THE IMPACT OF DROUGHT ON AGRICULTURAL CROP YIELDS	10
2.1. Introduction	10
2.2. Materials and Methods	17
2.2.1. Study Area	17
2.2.2. Model Description	19
2.2.3. Data	22
2.2.4. Methodology	24
2.3. Results and Discussions	29
2.3.1. Technical Validation	29
2.3.2. Modeling Results	33
2.3.3. Correlation with Crop Yield	41
2.4. Conclusion.....	46
Chapter 3.....	48
EFFICACY OF SEASONAL CLIMATE FORECASTS IN PREDICTING INTERANNUAL CROP YIELDS	48
3.1. Introduction	48
3.2. Materials and Methods	57
3.2.1. Study Area	57
3.2.2. Modeling Framework.....	58
3.2.3. Data	62
3.2.4. Methodology	65
3.2.5. Mutual Information.....	72
3.3. Results and Discussion.....	75
3.3.1. Efficacy of seasonal climate forecast on crop yield.....	75
3.3.2. Influence of geophysical and drought variables on crop yield	80
3.3.3. Process Network Connectivity.....	84
3.4. Conclusion.....	93
Chapter 4.....	97
EVALUATING THE CROP YIELD PREDICTABILITY THROUGH SEQUENTIAL DATA ASSIMILATION	97
4.1. Introduction	97
4.2. Data and Methods.....	102
4.2.1. Study Area	102
4.2.2. Model Description	104

4.2.3.	Data	107
4.2.4.	Assimilation Data.....	108
4.2.5.	Methodology	111
4.3.	Results and Discussion.....	121
4.3.1.	Soil Moisture Analyses	121
4.3.2.	Yield Trends Across Years	129
4.3.3.	Statistical Analysis.....	135
4.3.4.	Comparative Crop Yield Estimation Analysis.....	142
4.4.	Conclusion.....	152
Chapter 5.....		155
CONCLUSION.....		155
BIBLIOGRAPHY.....		158
APPENDIX A. IMPLICATIONS OF DROUGHT ON INTERANNUAL CROP YIELDS		174
APPENDIX B. EQUATIONS FOR VARIOUS STATISTICAL MEASURES		177

Chapter 1.

INTRODUCTION

The dynamics of seasonal crop yield prediction amid weather and climate extremes are multifaceted, shaped by a confluence of environmental, technological, and socio-economic factors. This chapter aims to provide an overview of the existential and future problems happening globally, having severe repercussions on the environment and society. Drawing from both global trends and localized observations, the chapter encapsulates the key challenges and the existing knowledge base, setting the stage for the subsequent in-depth explorations.

1.1. Background

The intricate relationship between land and climate is central to understanding the pressing environmental challenges faced globally. This connection between different aspects of the biogeophysical processes is vital for maintain equilibrium in the climate system. Thus, the changes on the earth's surface plays a critical role in regional and global weather patterns. However, the interplay between climate and land has become an area of escalating concerns in recent decades. As anthropogenic activities have continuously reshaped the landscape, they have imposed profound pressure on our natural ecosystems, often linked to habitat destruction, biodiversity loss, human health, and a myriad of environmental complications (Mahmoud and Gan, 2018; Burrell et al., 2020). Furthermore, emerging megatrends such as population growth, rapid urbanization, and land-use alterations have accelerated global warming, with temperatures on an upward trajectory, extensive land and soil degradation, and the frequency and intensity of extreme weather events on the rise (Lehmann et al., 2018). Increased anthropogenic activities, such as increased greenhouse emissions, land alterations from agriculture and deforestation, etc., have considerably altered the weather patterns and natural landscape over the years (Arora, 2019). Such multi-scale changes,

both directly resulting from human intervention and indirectly through climate change implications, have profound repercussions. Warmer temperatures, intense precipitation events, and increased exposure to weather extremities such as drought, flood, hurricanes, heatwaves, wildfires are some of the observable effects of the rapidly changing and unpredictable nature of climate that have resulted in huge socioeconomic losses (Estrada et al., 2023; Robinson et al., 2021). Such transformations not only impact local climates but also significantly affect water availability, biodiversity, ecosystem health, and other critical sectors.

Developing regions, particularly in Africa and Asia, are bearing the brunt of these climatic adversities. The burgeoning demand for food, coupled with the dependency on dwindling natural resources, has introduced formidable socioeconomic challenges (Bandara et al., 2014). Changes in land conditions, including deforestation, afforestation, urbanization, and agricultural expansion, has disturbed the natural climatic patterns, hydrological regimes, and agricultural outputs. For instance, conversion of forests into agricultural lands in Southeast Asia have led to deterioration of the soil quality (e.g., reduced carbon dioxide storage), leading to reduction in evapotranspiration (ET) and localized warming, altering precipitation patterns and thus further influencing land conditions in a cascading effect (Tolimir et al., 2020; Vadrevu et al., 2019). The challenges posed by these interactions are further exacerbated by extreme weather events. The increasing frequency and intensity of phenomena such as droughts, floods, and heatwaves underscore the vulnerability of land resources (Qi et al., 2022). Although drought and floods are naturally occurring hazards, human-induced/anthropogenic factors (e.g., land use changes, overutilization of resources, etc.) have further aggravated the risks associated with these climate extremities globally. Particularly, the increased frequency and intensity (severity) of drought and floods have threatened food security and sustainability globally, affecting critical sectors such as agriculture, water resources,

terrestrial biodiversity, forestry, and fisheries, among others (Tabari et al., 2023). These findings have been corroborated by the Intergovernmental Panel on Climate Change (IPCC) (Shukla et al., 2019), highlighting the urgency for adaptive and mitigation strategies. Specifically, the report outlined the importance of land-atmosphere interactions in influencing regional weather patterns, and its direct impact on agricultural productivity and food-security in climate vulnerable regions. For example, climatic fluctuations combined with human-induced global warming has led to the reduction of agricultural productivity in mid- and low- latitude areas over the past 50 years (IPCC AR-6; Yadav et al., 2021). Similarly, the construction of hydropower (dams) projects in the Amazon and Mekong River basins, along with the rapid land use changes, and the El Nino Southern Oscillation (ENSO) events have resulted in severe repercussions for local climates, water availability, and biodiversity, leaving millions of people exposed to acute food insecurity and water scarcity (Milhorange and Bursztyn, 2019; Mishra et al., 2021). These threats have enormous impact on certain sections of the society, specifically the poor rural settings, low-income and other marginalized communities have been vulnerable to the risks associated with seasonality. The direct impacts from these climatic shocks have resulted in massive socio-economic losses in food productivity, crop yields and food prices.

The complexities of climatic perturbations and their subsequent impact on agricultural systems have been a subject of intense research and deliberation. Among the myriad climatic anomalies, drought stands out as the most complex weather-related hazard. Due to its unique characteristics—slow onset, pronounced spatial and temporal variability, it is extremely complex to understand and mitigate drought. Recently, many studies have outlined the concerning increase in drought frequency and intensity not merely as a natural fluctuation but intrinsically linked to anthropogenic activities (Dai, 2013). Longer intense droughts in the Horn of Africa (Blamey et al.,

2018) and Southeast Asia are some of the glimpses of the worst climate-related disasters over the past few decades. One of the threatening outcomes of these intensified droughts is their detrimental effect on food production systems worldwide. The situation is evident in the increasing instances of crop failures, reduced agricultural yields, and rising food prices, especially in regions that are heavily reliant on rain-fed agriculture. Hence, there is an urgency to understand the relationship between climate and food security through the holistic cognizance of land-climate interactions.

The intricate relationship between climatic variables and agricultural productivity has long been a focal point of research, but the exacerbating effects of global warming have brought this issue to the forefront of global policy discourse. The agricultural sector is inherently sensitive to climatic variations. Recent climatic anomalies, characterized by increasing temperatures, erratic rainfall patterns, and the frequency of extreme weather events, have disrupted traditional agricultural practices, particularly in climate-sensitive regions (Wiebe et al., 2015). Such climate extremes can have profound implications, leading to harvest failures and threatening the food security of communities at both local and global scales. Studies have consistently shown that a rise in temperature beyond the optimum level can significantly reduce crop yields. For instance, for every 1°C increase in temperature, wheat yields can decline by 6%, a stark figure considering the global reliance on this staple crop (Asseng et al., 2013). Irregular precipitation patterns, coupled with the increasing frequency of droughts, have heightened water scarcity concerns. This has direct implications for irrigated agriculture, which accounts for a significant portion of global food production. Similarly, elevated temperatures and altered humidity levels have shifted the dynamics of pests and diseases, leading to increased incidences of crop losses, while coastal agricultural lands are facing the brunt of sea-level rise, with increased soil salinity affecting crop health and productivity. Moreover, the effects of climate change on food security are not limited to crop

yields. The entire food supply chain, from production, post-harvest processing, to distribution, is susceptible. Furthermore, the socio-political implications of these biophysical changes are significant. Food price volatility, driven in part by fluctuating agricultural yields due to climatic anomalies, can lead to socio-economic unrest. The global food price crisis of 2007-2008, for instance, triggered widespread protests in several countries, underscoring the intricate links between food security, climate, and societal stability (Tadasse et al., 2016).

To navigate this intricate web of challenges, technological advancements have emerged as a beacon of hope. The integration of remote sensing products, offering continuous measurements of geophysical parameters, has revolutionized the monitoring of water resources and crop growth (Khanal et al., 2020; Whitecraft et al., 2019). Remote sensing has paved a new era of precision agriculture, offering an array of continuous measurements of geophysical parameters pivotal for agricultural practices. High-resolution satellite imagery, such as the Synthetic Aperture Radar (SAR), provides invaluable insights into pivotal geophysical parameters like soil moisture, groundwater, and precipitation. Unlike traditional optical sensors, which are hampered by cloud cover and require daylight, SAR operates effectively across diverse weather conditions and times of day. This radar system captures high-resolution images by emitting radio waves and measuring their reflections off the Earth's surface. Such capabilities make SAR invaluable for monitoring geophysical parameters like soil moisture, groundwater levels, and precipitation. These parameters, traditionally measured sporadically and often inaccurately due to the vast spatial scales, can now be monitored continuously, offering granular insights into the subtlest of changes. Although, remote sensing has its technical limitations, i.e., they have intermitted observations of data, they are coupled with land surface models (LSM) to simulate a continuous interaction between the atmosphere and the Earth's surface. By integrating remote sensing raw data into

computational models, there is a more nuanced understanding of the synergies between hydrological processes and crop growth (Huang et al., 2019; Simons et al., 2017). The synergy between remote sensing and LSM is particularly evident in agriculture. For instance, understanding soil moisture is paramount to efficient irrigation. While remote sensing provides data on surface soil moisture, LSM delves deeper, offering insights into root zone soil moisture, a critical determinant of plant health and water stress. This amalgamation of data and analytics equips farmers with real-time information, enabling them to make informed decisions on irrigation, potentially conserving water resources and optimizing crop yields. Concurrently, advancements in computational capabilities and artificial intelligence have significantly enhanced the analytical prowess of LSM, potentially predicting crop yields and water requirements with exceptional accuracy (Liu, 2020; Jha et al., 2019). The development of sophisticated land surface schemes in hydrologic and crop growth models has enabled to tackle the intricate relationships between the different parameters of the soil-water-atmosphere continuum.

Despite the technological advancements, there exists a diverse set of challenges which has been outlined in subsequent chapters. One of the major bottlenecks in such studies is the persistence of uncertainties (or errors), either in the raw meteorological data, model itself, or a combination of both. Commonly, meteorological datasets from ground-based observatories, or remote sensing platforms inherently possess uncertainties due to the sensor inaccuracies, calibration errors, or spatial and temporal coverage gaps. Likewise, the model parameters that are highly complex to be modeled are sometimes represented in a simplified manner, leading to potential deviations from actual observations. In the context of agriculture, an erroneous prediction about soil moisture or an impending heatwave, rooted in model uncertainties, can lead farmers to make ill-advised decisions, affecting both their livelihood and the larger food supply chain. Hence,

this compounded uncertainty can be particularly problematic when making long-term predictions or when planning for extreme events.

While it becomes imperative to discern the nuances of climate extremes, particularly drought, on agriculture, it's equally pertinent to bridge the gap between hydrologic and agricultural dynamics, ensuring a cohesive understanding of these interlinked geophysical systems. This study showcases the ability to facilitate a multi-pronged approach to understanding drought and its repercussions on agricultural crop yields, leveraging multiple techniques and tools to understand the dynamics of both systems. By employing a multi-modeling coupled framework, remote sensing, data analytics and statistics, data assimilation, the study endeavors to provide actionable insights, ensuring enhanced agricultural sustainability and resilience against the looming challenge of climate extremes.

1.2. Research Objectives

The overarching aim of this dissertation is to undertake a detailed exploration into the dynamics of agricultural crop yield prediction across climatically sensitive regions, with a particular focus on understanding the intricate interplay between hydrological dynamics and agricultural productivity. To achieve this, the study is structured around the following objectives:

- i. Assessment of historic evaluation of drought across the Lower Mekong River basin (LMB) countries in Southeast Asia, and the impact of such events on interannual crop yields over the years.
- ii. Quantifying the uncertainties associated with crop yield forecasts and analyzing the intricate relationship between hydrologic variables, especially drought indicators, and crop yields to understand their combined influence on agricultural outcomes.

- iii. Investigating the potential implications of advanced data assimilation techniques in refining agricultural yield predictions over different crops in East Africa and Southeast Asia.

In essence, this research endeavor seeks to bridge the existing knowledge gaps, offering a holistic view of drought's influence on agriculture, and paving the way for more resilient and sustainable agricultural practices in the face of ever-evolving climatic challenges.

1.3. Significance of the Research

Climate extremes, particularly drought, have emerged as critical challenges in the context of agriculture and hydrology. The impacts of these extremes can ripple through socio-economic systems, influencing food security, livelihoods, and overall economic stability. In regions with a significant agricultural footprint, understanding the intricacies of drought and its cascading effects on crop yields is not only vital for immediate resilience but also for long-term sustainable development. The significance of this research lies in its multidimensional approach to a problem of global concern. Drought, often characterized by its subtle and gradual onset, poses a complex challenge to regions heavily reliant on agriculture. As global demands for food surge, the constraints posed by climate-related changes are hampering growth rates of essential crops. Our study closely focuses on climate sensitive regions that have been the most vulnerable to the negative impacts of climate extremes. Leveraging advanced methodologies, such as earth observation systems, in-situ data collection, and state-of-the-art statistical modeling techniques, this research aims to bridge existing knowledge gaps. The research not only focuses on the present outlook but aspires to offer actionable and robust data-driven insights for improved decision-making during critical agricultural phases.

1.4. Organization of Thesis

This thesis is meticulously structured into six chapters. The "Introduction" establishes the foundation for the entire research. It offers a broader view of the multifaceted challenges of climate extremes, particularly drought, and its repercussions on agricultural yields. Additionally, it outlines the objectives of the study and significance of the research. Chapter 2 "Evaluating the impact of drought on agricultural yields" closely looks into the empirical examination of drought's influence on agricultural output. The chapter is self-contained, encompassing its own introduction, materials and methods, results, discussions, and conclusions, providing a detailed account of the study's findings and their implications. The next set of chapters would be exploring two unique approaches for improving the crop yield prediction over time. Chapter 3 "Efficacy of seasonal climate forecast on interannual crop yields" explores the potential of seasonal climate forecasts in predicting interannual crop yields by quantifying the uncertainties from historic climate forecasts. Additionally, this chapter explores the intricate dynamics between climatic forecasts and crop outputs. Chapter 4 "Improving the crop yield prediction through data assimilation" employs a scientific approach to predict agricultural crop yields using data assimilation. It provides vital insights to the enhanced accuracy of crop yield prediction, accounting for a myriad of factors. The final chapter "Conclusion" offers a consolidated view of the research's findings. It not only synthesizes the empirical findings from Chapters 2, 3, and 4 but also places them in the broader context of the research theme, highlighting their implications and suggesting avenues for future exploration.

Chapter 2.

EVALUATING THE IMPACT OF DROUGHT ON AGRICULTURAL CROP YIELDS

This chapter was published in Journal of Hydrology, Vol 599, Abhijeet Abhishek, Narendra N Das, Amor VM Ines, Konstantinos M Andreadis, Susantha Jayasinghe, Stephanie Granger, Walter L Ellenburg, Rishiraj Dutta, Nguyen Hanh Quyen, Amanda M Markert, Vikalp Mishra, Mantha S Phanikumar, Evaluating the impacts of drought on rice productivity over Cambodia in the Lower Mekong Basin, 126291, Copyright Elsevier (2021).

2.1. Introduction

Drought, often regarded as one of the complex weather-related extremes, stands out as a potential disruptor of agricultural systems. Often characterized by a prolonged deficit of precipitation and long periods of abnormally dry conditions, there isn't a universally accepted definition of drought (Van Loon et al., 2016a). Drought manifest from a range of hydrometeorological processes that reflect the long-term imbalance between water supply and water demand. Due to their high variability, both spatially and temporally, the effects of drought can be localized, thus making its characterization complex and uncertain. Historically, recurring droughts have had far-reaching impacts, crippling the socioeconomic conditions of countries and affecting agriculture, water resources, human livelihoods, food security, and natural ecosystems (Wilhite, 2000). Numerous studies across diverse geographies—including the United States (Andreadis et al., 2005; Zhang et al., 2017), India (Aadhar and Mishra, 2017), the Amazon (Duffy et al., 2015; Wongchuig Correa et al., 2017), Europe (Grillakis, 2019), Africa (Gebremeskel Haile et al., 2019; Sheffield et al., 2014), and China (Zhang and Jia, 2013; Zhang et al., 2016)—have documented the variability, severity, duration, and timing of droughts.

Arguably, the frequency and intensity of droughts have been on an upward trend, attributed mainly to external factors such as population growth, rapid urbanization, land-use changes, and altered precipitation patterns (Dai, 2013; Van Loon et al., 2016b). These factors, exacerbated by the impacts of climate change, have intensified pressures on countries, raising grave concerns related to food and water security (Godfray et al., 2010). These physical changes, driven in part by anthropogenic activities such as global warming (McCabe et al., 2004; Dai, 2011; Trenberth et al., 2014), are anticipated to contribute to extreme temperatures and the intensification of drought hotspots around the globe.

The increased consumptive use of water resources attributed to agriculture has heavily weighed on the water budget, with extensive irrigation, increased groundwater extractions, obsolete management of water resources further aggravating the complexity of the situation (Mishra and Singh, 2010). This has resulted in seasonal water shortages, markedly seen in areas like the southern parts of the Mekong Basin in Southeast Asia (Thilakarathne and Sridhar, 2017). The Mekong River, traversing through four riparian countries—Lao PDR, Thailand, Cambodia, and Vietnam—accounts for approximately 77% of the flow and holds paramount environmental and economic significance to the region (MRC, 2014). Yet, the Mekong region grapples with challenges: dwindling natural resources, reducing croplands, and heightened exposure to extreme events like droughts and floods. These challenges are exacerbated by land-use changes and rapid urbanization (MRC, 2009). Along with natural alterations, the recent infrastructure (e.g., hydropower) and irrigation developments in the upstream Mekong have changed the magnitude and seasonality of flow (Hoang et al., 2019; Pokhrel et al., 2018), especially in the lower stretches of the basin (near Tonle Sap Lake, Cambodia). As a result, the Tonle Sap Lake has experienced a sizable reduction in volume, leading the surrounding wetlands to dry up (Wang et al., 2020). With

ongoing and anticipated anthropogenic activities, there is potential for even more pronounced seasonal variations in the basin's storage, especially concerning lake volume and water level fluctuations. Given the ongoing and proposed anthropogenic developments, the basin's active storage could be further subjected to significant seasonal changes, particularly lake volume and water level fluctuations. Impacts of extreme drought events in the past decades, notably in 1997-98, 2003-05 (IPCC, 2007) and late 2015-2016 (Guo et al., 2017; Son et al., 2012) were predominantly visible within the agricultural sector in the Lower Mekong countries. Despite the increased vulnerability of agriculture to drought, the latter's impacts have been somewhat overlooked, and arguably less understood than flood (Kim et al., 2019).

About 70 percent of the population are actively engaged in agricultural activities in the LMB region of Southeast Asia. A majority of these agricultural systems are dependent on rainfall, rendering them susceptible to the uncertainties and risks linked with seasonal variability (Johnston et al., 2012; MRC, 2003). Given the absence of alternative income sources beyond agriculture, the repercussions of seasonality-induced crop losses are particularly harsh, severely affecting the livelihoods of lower-income and marginalized communities. Rice, the region's staple crop, is acutely sensitive to climate fluctuations and has witnessed detrimental impacts from climatic changes. Areas near Tonle Sap Lake and the deltaic regions of Vietnam, being low-lying, are especially prone to these environmental challenges. The escalating demand for rice—with projections reaching 3720 kg/ha in Cambodia and 6530 kg/ha in Vietnam by 2025 (MRC, 2014)—underscores the imperative of alleviating risks posed by seasonal extremes while concurrently boosting agricultural output. Hence, addressing these challenges calls for a multi-faceted approach. A holistic perspective, encompassing both regional and national drought risk management strategies, is gaining traction as an effective way to diminish community vulnerabilities (ADB,

2009). Additionally, there's a pressing need to continuously monitor hydrological conditions, track drought status, and estimate crop yields, ensuring that this vital data is routinely available to stakeholders. Providing decision-makers with timely insights into inter-seasonal and intra-annual hydrological patterns, drought metrics, and rice yield forecasts will empower them to make informed decisions, bolstering risk management and strategic planning.

In the context of this study, particularly concerning rice cultivation in regions like the LMB, it's important to clarify the term 'drought.' While it's true that traditional paddy farming involves saturated soil conditions, these are not always maintained through active irrigation. Instead, many regions rely on natural rainfall and associated hydrological processes to provide the necessary water. Therefore, in this research, 'drought' refers to prolonged periods of below-average rainfall that lead to insufficient natural water availability in the rice paddies. This condition is not solely about the absence of rain; it's about the resultant lack of standing water critical for rice cultivation, which in this case, is predominantly rainfed. During drought periods, there may be insufficient precipitation to sustain typical surface water or groundwater levels, impacting the natural flooding of paddies. This understanding is especially relevant in the context of Cambodia, where the rice farming systems are predominantly dependent on seasonal rains, and the capacity for supplemental irrigation is limited or non-existent. Here, a meteorological drought directly translates to a deficit in the water necessary for rice cultivation, as the standing water in paddies is primarily sourced from rainfall and natural hydrological contributions, rather than controlled irrigation systems. By this definition, the study recognizes that 'drought' encompasses not just a reduction in rainfall, but also the broader implications this has for natural water availability critical to maintaining suitable conditions for rice growth in paddy fields.

Despite the advent of numerous drought monitoring tools, such as the United States Drought Monitor (Svoboda et al., 2002), and the Global Drought Early Warning Monitoring Framework (GDEWF: Pozzi et al., 2013), the accurate assessment and prediction of drought characteristics still remains a formidable challenge. Historically, inconsistent records and subpar of spatial coverage data from ground-based observational networks have been major obstacles to extensive hydro-agricultural studies. However, recent technological advancements in dynamic modeling and the use of satellite-based observations have created the potential for operational drought monitoring and forecasting (Klisch and Atzberger, 2016). Spaceborne sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) at visible and infrared frequencies, Tropical Rainfall Measuring Mission (TRMM) at higher microwave frequencies, Soil Moisture Active Passive (SMAP) at lower microwave frequency, Gravity Recovery and Climate Experiment (GRACE) gravity anomaly measurements, collect data across various wavelengths of the electromagnetic spectrum at varying spatial resolutions, offering a broader and more consistent view of hydrological processes. This has resulted in generating reliable data products via different methods (e.g., data assimilation) for effectively tracking the changes in the physical processes of the land-atmosphere continuum. With the integration of targeted drought indicators into these models, monitoring techniques have significantly improved. These indicators have not only improved the existing monitoring techniques (and capabilities) but facilitated insightful information on early warning, preventive measures, and mitigation strategies. These indices not only quantify the meteorological/agricultural/hydrological drought characteristics, but also provide a holistic comprehensive assessment of the location, severity, and duration of drought events in a specific area at multiple timescales. Several indices such as the Standardized Precipitation Index (SPI: McKee et al., 1993), Soil Moisture Deficit Index (SMDI: Narasimhan

and Srinivasan, 2005), Palmer Drought Severity Index (PDSI: Palmer, 1965), Normalized Difference Vegetation Index (NDVI: Rouse et al., 1973), Vegetation Condition Index (VCI: Kogan, 1995), Standardized Runoff Index (SRI), among others, have been proposed to address drought prediction. Subsequently, a number of advanced indicators were developed by combining multiple indices. A combination of multivariate indices (Scaled Drought Condition Index (SDCI) and Synthesized Drought Index (SDI)) for characterizing agricultural drought, carried out by Du et al. (2013) and Rhee et al. (2010), yielded even more promising results, outperforming many existing metrics. Hence, using a combination of multivariate remote sensing-based indices (or data), the numerical representation of drought characteristics from a range of extensive hydroclimatic conditions provides useful information on the overall qualitative state of drought (Hao et al., 2015). These studies have partly yielded better results because of the improvements from the use of remote sensing data (and the use of multi-sensor products).

A recurring challenge in the field of scientific applications is the disconnect between the technology and its end-users, resulting in the partial implementation of effective adaptation strategies (Andreadis et al., 2017). This fundamental disconnect is distinct within the agriculture sector, so bridging the gap is essential for better decision making. Though many software frameworks have explored the integration of modeling systems, the internal representation of datasets has always proved to be challenging, often resulting in compatibility issues within the constituent models (Beran and Piasecki, 2009). Addressing this challenge, we utilized a newly developed integrated information framework- the Regional Hydrologic Extremes Assessment System (RHEAS: Andreadis et al., 2017), that couples a physically based hydrologic and crop model, and uses data assimilation (primarily using remote sensing observations) and projections to provide nowcasts and seasonal forecasts (3-6 months in advance) of yield estimates and

associated drought indicators. RHEAS was developed at the Jet Propulsion Laboratory (NASA-JPL) and has been implemented at the different NASA-SERVIR hubs, including the Regional Centre for Mapping of Resources for Development (RCMRD) in Nairobi and the Asian Disaster Preparedness Center (ADPC) in Bangkok. Covering regions over East Africa and the Lower Mekong, RHEAS operates at a 25 km spatial resolution, extending to 5 km in specific areas for diverse water resources applications. A notable feature of RHEAS is its model-agnostic data storage, simplifying data transfer within the framework. This design facilitates effortless downloading, extraction, and visualization of datasets, ensuring a user-friendly experience. In this study, the capabilities of RHEAS are harnessed by integrating readily available observations of meteorological (e.g., precipitation) and hydrologic parameters (e.g., soil moisture) via a coupled hydro-agricultural model to identify changes in drought conditions and evaluate their effects on rice production in the Lower Mekong region.

The RHEAS integrated system offers tailored insights by delivering key hydrologic variables, drought indicators, and crop productivity values directly associated with rice yield estimates. These insights are invaluable to decision-makers and regional agencies for proactive resource allocation and mitigation strategies against drought impacts. Recognizing its potential, regional entities such as the Mekong River Commission (MRC), and the Vietnam Academy of Water Resources (VAWR) have incorporated RHEAS outputs to issue guidelines and advisories to the federal agencies across the LMB countries. While RHEAS covers the entirety of the LMB, our study narrows its focus to Cambodia, given the nation's heavy reliance on agriculture and the profound impacts of drought on its livelihoods and food security. This research pursues two main objectives: i) first, to assess the changes in drought (and hydrological processes) and rice yields over time, and ii) second, to gauge the potential implications of such changes on Cambodia's rice

yields. Leveraging a diverse set of hydrologic datasets, the derived model simulations (e.g., soil moisture, yield, etc.) were analyzed with remote sensing observations to evaluate the model efficacy. Building upon the successful preliminary study conducted by Abhishek in 2018, the approach and results outlined in this paper are anticipated to significantly enhance our understanding of the current environmental conditions affecting hydrology and crop yield. This enhancement is achieved by providing a quantitative historical perspective, which is instrumental in supporting and refining decision-making processes.

2.2. Materials and Methods

2.2.1. Study Area

The Mekong River, located in Southeast Asia, stands as one of the world's most significant rivers. Stretching approximately 4,350 km and boasting a mean annual discharge of around 475 km³, it ranks as the tenth largest river globally in terms of flow (Liu et al., 2007). Originating from hilly mountain ranges, it meanders through deep gorges, expansive floodplains spanning 2,600 km, eventually flowing into the South China Sea. Its journey encompasses a diverse landscape, from the towering terrains of Lao (~4500 m elevation) to Vietnam's deltaic plains, draining an area of 810,000 km² (MRC, 2014; Lauri et al., 2014). Typically, the Mekong River Basin (MRB) is typically portioned into the "upper basin" (Upper Mekong Basin or UMB), and the "lower basin" (LMB). The LMB holds particular environmental and economic significance, covering about 77% (606,000 km²) of the basin and spanning four countries: Cambodia, Lao PDR, Thailand, and Vietnam, covering an approximate area of 618,783 km² (Figure 1.). This region, marked by a typical monsoon climate, experiences two distinct seasons- i) wet Southwest monsoon (May-Oct), and ii) a dry Northeast monsoon (Nov-Apr), with annual precipitation between 1200 and 2500 mm/year respectively. Most of the year's precipitation (~85%) occurs during the wet season, with

temperatures and evaporation rates fluctuating based on elevation (Kite, 2001). Temperature and evaporation rates throughout the basin vary with elevation, typically ranging between 22-28°C and 1,000-2,000 mm/year respectively. The river flows through extensive wetland habitats supporting productive ecosystems, particularly the Tonle Sap Lake in Cambodia. The lake and its surrounding floodplains in the heart of Cambodia play an important role in inducing a seasonal change in the direction of the flow in the river. This seasonality greatly influences the floodplains and the agricultural productivity downstream (lower provinces of Cambodia) and the Mekong delta. The Tonle Sap floodplains, combined with the Mekong delta, support a population of 35 million and are responsible for over 90% of the region's paddy plantation (MRC, 2014). However, despite strides in the agricultural sector, Cambodia still faces challenges. Disparities in mechanization and irrigation capabilities have curtailed its production potential, lagging behind its neighboring nations.

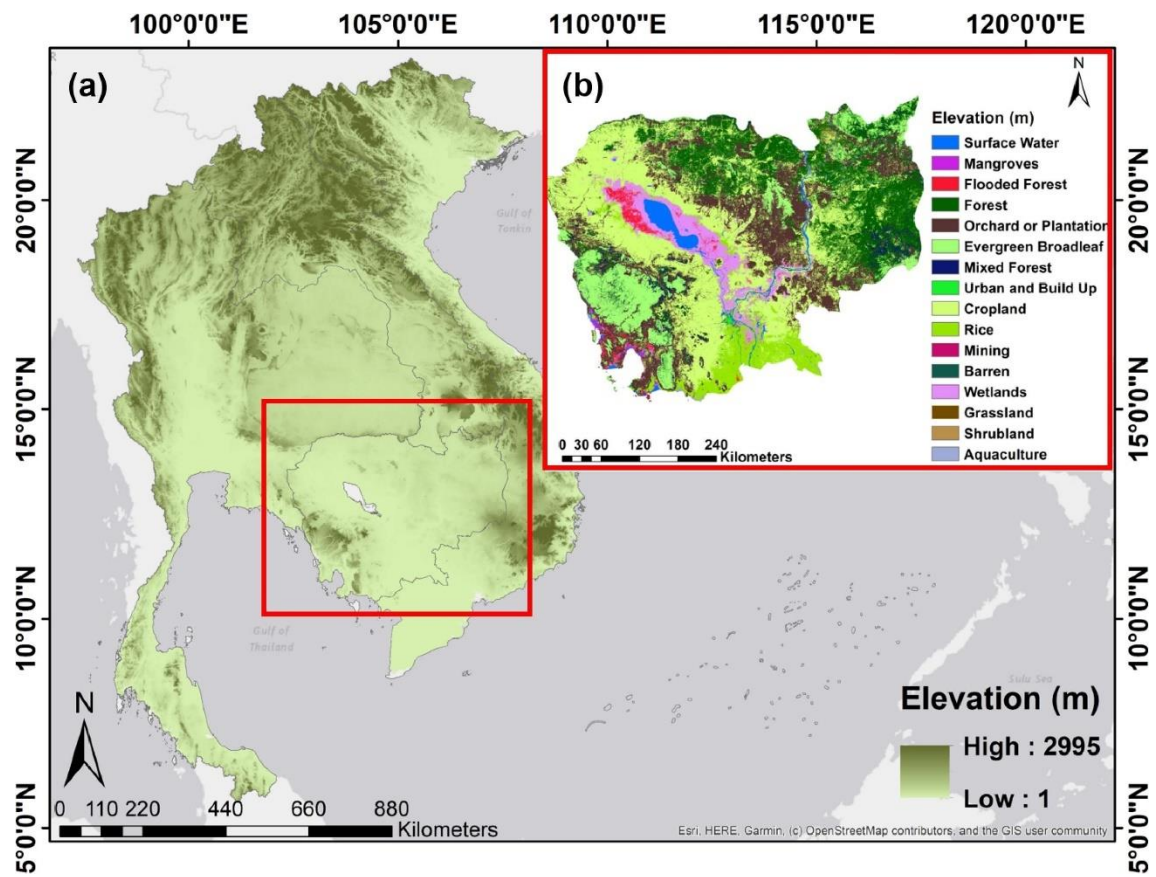


Figure 1. (a) Map showing the Lower Mekong Basin (LMB) countries with Cambodia highlighted; (b, inset) Regional land cover/land use map of the study area Cambodia.

2.2.2. Model Description

RHEAS integrates multiple remote sensing products across different components of the terrestrial water cycle to effectively monitor and estimate the drought status, water stress and interannual crop yields. The main component of the RHEAS architecture hosts a spatially enabled relational (PostGIS) database that ingests a diverse suite of earth science products, from model datasets to satellite observations. This PostGIS extension, built on the PostgreSQL database, harnesses the SQL language (combined with other features) to manage and query spatial geometries, setting RHEAS apart from other information systems. Additionally, RHEAS follows a hybrid design, enabling the seamless coupling of the component models, thus making its design

distinct and unique from other information systems. Such design of combining modular and object-oriented programming has several advantages- i) efficient transferability of data across models, and ii) the system's modularity which means each model interfaces directly with the PostGIS database, bypassing the need to adapt to other models' internal formats. This arrangement also allows for straightforward implementation and customization with minimal input requirements from the end-users, thus extending the system's applicability. For a visual representation, Figure 2. showcases a simplified flowchart of RHEAS's architecture, detailing its constituent models and meteorological inputs. This integration strategy captures a comprehensive view of hydrological processes and the dynamics of the soil-plant-atmosphere continuum. Detailed information about the model design, architecture, installation, and operation is readily available at <https://github.com/nasa/RHEAS>.

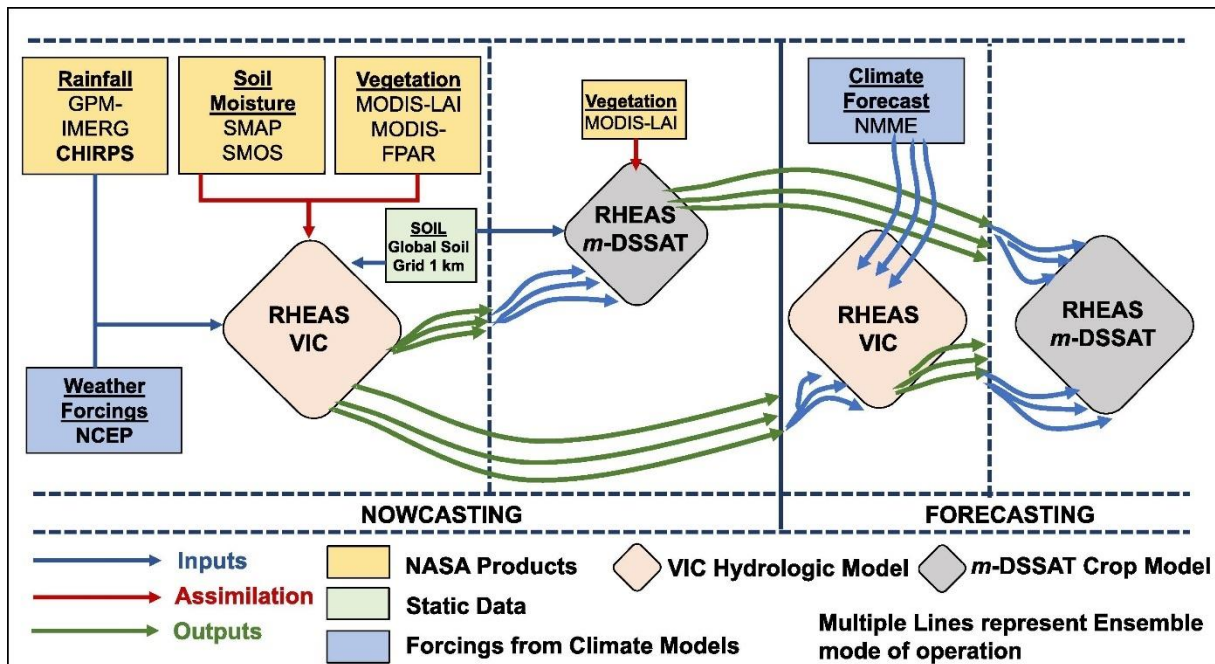


Figure 2. Simplified flow chart of the RHEAS algorithm: RHEAS assimilates multiple remote sensing observations to provide different hydroclimatic states and drought indicators. VIC: Variable Infiltration Capacity hydrologic model; m-DSSAT: modified Decision Support System

Figure 2. (cont'd)

for Agrotechnology Transfer crop model; CHIRPS: Climate Hazards Group Infrared Precipitation with Station data; NCEP: National Centre for Environmental Prediction; SMAP: Soil Moisture Active Passive; NMME: North American Multi-Model Ensemble. Nowcasting is the present agri-hydrologic states based on past, historic records, while the forecast simulations require a model ensemble to predict the drought characteristics and crop yields at lead times of one to three months.

The macroscale Variable Infiltration Capacity (VIC: Liang et al., 1994) hydrologic model is the primary component in the RHEAS information system. The VIC hydrological model simulates the land-atmosphere fluxes and computes the energy (and water) balance at the land surface at daily time-steps. In addition, VIC generates a multitude of hydrologic variables and drought indicators that are used to quantify the uncertainties across different components of the hydrologic cycle. These variables and indicators are tightly constrained by in situ and satellite observations of soil moisture, precipitation, runoff, evapotranspiration (ET), groundwater, and snow (if needed).

The process-based Decision Support System for Agrotechnology Transfer (DSSAT: Jones et al., 2003) crop model is integrated within RHEAS. It simulates the crop growth, development, and yield, considering diverse different management practices and soil conditions. A notable enhancement within RHEAS is the modified DSSAT (*m*-DSSAT) version. The *m*-DSSAT is a modified version of the baseline DSSAT that has the functionality to run 40-50 ensemble members and can stop and restart every day. This feature of the *m*-DSSAT makes it stand out compared to the baseline DSSAT or other crop growth models that generally runs continuously from sowing until maturity/harvest (Ines et al., 2013). This modification was necessary to facilitate data

assimilation of leaf area index (LAI) and soil moisture during different phases of crop growth. In addition, this refinement also helps to better capture the history of crop growth towards the harvest season by producing a realistic crop yield forecast than what could otherwise be obtained by just using model-based forcing from a seasonal climate forecast.

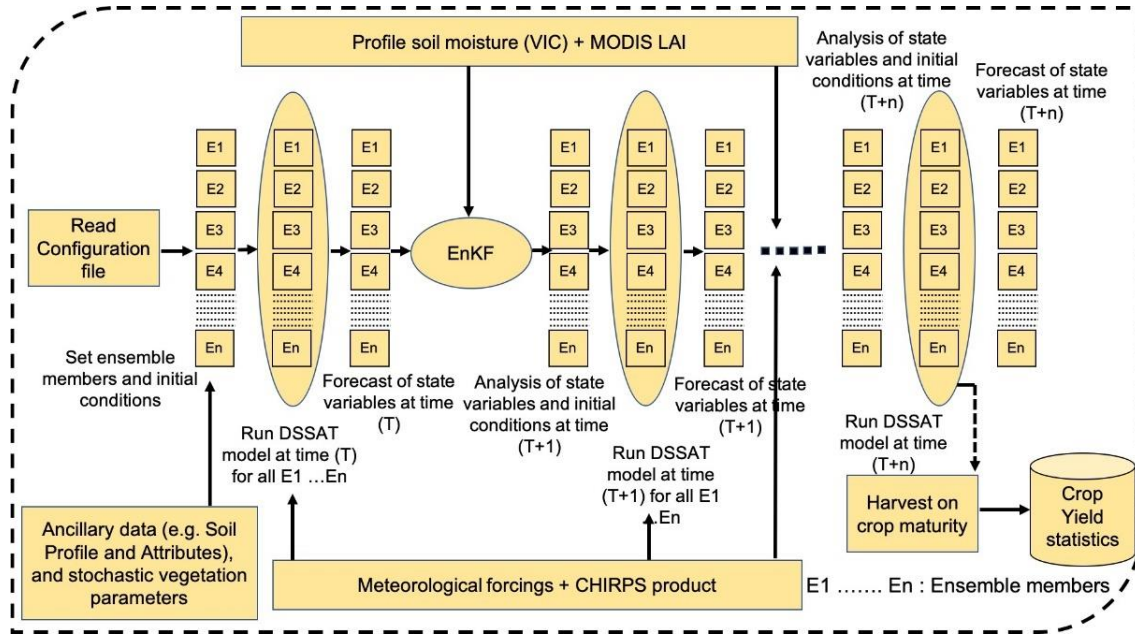


Figure 3. Schematic of the *m*-DSSAT crop model with multiple ensemble members.

2.2.3. Data

At a minimum, all the land surface models require high-quality meteorological forcings (e.g., precipitation, air temperature, and wind speed) and land cover/type information (e.g., soil properties, land cover, elevation) (Mizukami et al., 2014). The high-resolution (0.05°), near-real-time daily, and pentad Climate Hazards Group Infrared Precipitation with Station data (CHIRPS: Funk et al., 2015) gridded precipitation product was used to force the hydrologic model from 1981 through the present. Due to its long historical records, low latency, and temporally consistent datasets, the CHIRPS precipitation product was used to simulate the near-real-time initial hydrologic conditions. Many precursory studies have successfully used the CHIRPS precipitation

product for flood and drought monitoring (Katsanos et al., 2016; Toté et al., 2015). Other variables, such as air temperature and wind speed, were obtained using the National Centre for Environmental Prediction (NCEP: Kalnay et al., 1996) gridded reanalysis product. The land coverage information was obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS-500m) global product at yearly intervals (Friedl et al., 2010). Information on soil characteristics is an essential component of crop model simulations. Absence of any well-established soil database and gridded product over the Mekong region has made modeling soil processes difficult (Arrouays et al., 2014). Because of the gridded nature of RHEAS, a high-resolution (10-km) (Han et al., 2019), gridded soil database, based on the SoilGrids1km, was used for regional crop modeling in this study. Other ancillary information, such as fertilizer application rates, cultivar varieties, were obtained from various external sources (e.g., local bodies, government agencies, etc.) and previous studies. Table 1. briefly summarizes a list of datasets that are available within the RHEAS database.

Table 1. RHEAS input datasets along with their spatial and temporal resolution.

Variable	Product	Spatial Resolution	Temporal Resolution	Period Availability
Precipitation	GPM IMERG	10 km	Daily	2014 - present
	TMPA	25 km	Daily	1998 - present
	CMORPH	8 km	Daily	1998 - present
	CHIRPS	5 km	Daily	1981 - present
	PERSIANN-CDR	25 km	Daily	1983 - present
Temperature	NCEP	1.875 deg	Daily	1948 - present
Wind Speed	NCEP	1.875 deg	Daily	1948 - present
Soil Moisture	SMAP	~9 km	2-3 days	2015 - present
	SMOS	40 km	2-3 days	2010 - present
SAR Backscatter	Sentinel 1A/B	<1km	12 days	2016 - present
Evapotranspiration	MOD16	1 km	8 days	2000 - present
Leaf Area Index	MOD15	1 km	8 days	2002 - present
Seasonal Climate Forecast	NMME	2.5	Monthly	2000 - present

Note: Green: forcings, Blue: data assimilation, Orange: ancillary data for planting information

2.2.4. Methodology

A comprehensive regional study spanning 1981-2019 was conducted over the Lower Mekong Basin (LMB) region, with Cambodia as the primary focus. Using a text-based configuration file, simulations were executed to provide information on the mode of operation (either nowcast or forecast) and the chosen simulation type (crop via m-DSSAT or hydrologic via VIC). After the ingestion of the required forcing (and ancillary information), the VIC hydrologic model explicitly computes the hydrologic and drought characteristics for each grid cell (0.25°) over the study region. VIC has a unique representation of sub-grid variability and better characterization of vertical soil moisture distribution, thus making it an ideal choice for hydrologic simulations.

In this study, we have solely focused on agricultural and meteorological droughts. Agricultural drought is especially intricate and difficult to distinguish from other types due to the high disparity of water requirement for different crops. This study primarily centered on agricultural and meteorological droughts, with agricultural drought's intricacy stemming from varying water requirements across different crops. RHEAS uses the Soil Moisture Deficit Index (SMDI) and Drought Severity indicator for characterizing agricultural drought. The SMDI computes weekly soil water deficit to represent the overall soil water availability in the root zone. Using the long-term median of weekly averaged soil water, and maximum and minimum weekly soil water recording, SMDI provides the soil moisture deficit (in percent) at various depths of the soil profile. This information is especially used to discern the optimal crop water requirements of different crops at various stages of growth. Likewise, the drought severity index allows for the categorization of various levels of moisture availability (dryness/wetness), wherein the moisture content is expressed as a percentile of the cell's model climatology for each grid cell. Precipitation

is the most critical variable that is subjected to major fluctuations compared to other hydrologic states such as runoff, soil moisture, groundwater, etc. Thus, quantifying the impact of precipitation deficit on other hydrologic states (e.g., streamflow, groundwater, etc.) will help characterize meteorological drought. Hence, the standardized Precipitation Index (SPI) was used to reflect the precipitation deficit at multiple timescales (notably 1-,3-,6- and 12-month). Based on the historic long-term precipitation records, the SPI provides a standardized measure of conditions, indicating dryness or wetness. Tables, 2 and 3, detail the standardized scales for SPI and SMDI, wherein positive values represent wet conditions, and a negative value represents dry conditions. Moreover, RHEAS produces other indicators like the Standardized Runoff Index (SRI) and Dryspells, vital for understanding hydrologic drought conditions. Dryspells, in particular, supplement the system with information regarding the frequency of drought (i.e., number of days between two drought events) and plays a vital role in decision making activities during the growing season.

Table 2. Meteorological conditions based on classification of standardized precipitation index (SPI) (source: McKee et al., 1993).

SPI Value	Meteorological Condition
≥ 2.0	Extremely Wet
1.5 to 1.99	Very Wet
1.0 to 1.49	Moderately Wet
-0.99 to 0.99	Near Normal
-1.0 to -1.49	Moderately Dry
-1.5 to -1.99	Severely Dry
≤ -2.0	Extremely Dry

Table 3. Meteorological conditions based on classification of Soil Moisture Deficit Index (SMDI) (source: Narasimhan and Srinivasan, 2005).

SMDI Value	Meteorological State
3 to 4	Extremely Wet
2 to 3	Very Wet
1 to 2	Moderately Wet
1 to 0.5	Near normal
0.5 to -0.5	Normal
-0.5 to -1	Near normal
-1 to -2	Moderately Dry
-2 to -3	Severely Dry
-3 to -4	Extreme Dry

Following a similar technique as mentioned above, the crop model simulations were carried out by providing the essential information through a configuration file. Using the same information and model forcings (same as VIC), the modified DSSAT (*m*-DSSAT) model is run on a yearly basis to generate the inter-annual yields. The *m*-DSSAT model simulates the daily crop growth and developmental stages, considering the genetic attributes of the crop and prevailing environmental conditions. It delves into the intricate soil-plant-atmosphere continuum, factoring in soil moisture levels, nutrient dynamics, plant water consumption, and both potential and actual evapotranspiration rates. Furthermore, *m*-DSSAT requires detailed input on agricultural management practices, including but not limited to planting schedules, fertilizer application protocols, and irrigation strategies, all of which profoundly impact the eventual yield predictions. Hence, it becomes extremely important to supplement the crop model with specific information on local practices, such as cultivar types in the region, planting dates from crop calendar, etc., for ensuring optimal agricultural productivity. Typical management practice information was obtained

from local bodies, governments, and previous studies. Although, no specific information on cultivar varieties and fertilizer application rates were available, we used the information from previous studies and World Bank/FAO (Food and Agriculture Organization). While specific data on cultivar varieties and fertilizer application rates were elusive, we relied on insights from prior research and the aforementioned organizations. The genotype coefficients of rice varieties, as presented in Table 4., were adopted from a similar study by Wang et al. (2017).

In addition, the model is adept at gauging the crop's response to various stressors such as water scarcity, temperature extremes, and nutrient deficiencies, recalibrating growth, and yield accumulation in response to these factors. Ultimately, the *m*-DSSAT calculates the final yield estimates by assessing the simulated biomass production and harvest index, thereby reflecting both the crop's inherent genetic potential and the array of environmental and management influences it has encountered. This multifaceted simulation approach affords a nuanced exploration of diverse agricultural scenarios and their consequent effects on crop productivity, providing pivotal insights for strategic agricultural planning and management.

Table 4. Genotype coefficients of the two rice varieties computed through the *m*-DSSAT crop model.

Rice Variety	Phenological coefficients				Growth Coefficients			
	P1	P2R	P5	P2O	G1	G2	G3	G4
Sen Pidao	554.400	87.7	251.100	13.00	68.67	0.021	1.00	1.15
Phka Rumduol	435.100	295.100	388.00	11.28	58.96	0.026	1.00	1.20

Phenology genetic coefficients

P1: Time period in °C (above a base temperature of 9 °C) from seedling emergence to the end of juvenile phase. Expressed as growing degree days (GDD)

P2R: Extent to which phase development leading to panicle initiation is delayed for each hour increase in photoperiod above P2O. Expressed as growing degree days (GDD)

P5: Time period from beginning of grain filling to physiological maturity with a base temperature of 9 °C. Unit: GDD

P2O: Critical photoperiod or the longest day length at which development occurs at a maximum rate. Expressed in hours

Growth genetic coefficients

G1: Potential spikelet number coefficients as estimated from the number of spikelets per g of main culm dry weight (less leaf blades and sheaths plus spikes) at anthesis. Unit: Spikelets per g of main culm

G2: Single grain weight under ideal growing conditions, i.e., non-limiting light, water, nutrients, and absence of pests and diseases. Expressed in grams

G3: Tillering coefficient (scalar value) relative to IR64 cultivar under nonlimiting conditions

G4: Temperature tolerance coefficient

2.3. Results and Discussions

2.3.1. Technical Validation

2.3.1.1. Hydrological Modeling Validation

While many hydrological parameters are spatially heterogeneous, it becomes challenging to capture these large-scale changes, thus resulting in a disparate abstraction of the real system. However, with considerable adjustment and sufficient validation, the optimization of differences between model estimates and observations can present a near-accurate representation of the actual states. Model validation addresses the mismatch and associated uncertainties between simulated and recorded observations by adjusting key parameters. As the VIC model had been previously developed for the region, further calibration was not carried out in this study (Chang et al., 2019). Instead, both the constituent models (VIC and m-DSSAT) within RHEAS were validated against available observations, following a two-stage process. Firstly, the VIC model was validated against soil moisture data. The composite estimates of daily land surface conditions derived from the Soil Moisture Active Passive (SMAP) Level-3 radiometer soil moisture product (SPL3SMP) at a 36 km resolution was analyzed for Cambodia. The validation of annual mean soil moisture presented good accordance with observations ($r = 0.8$) for the top 5 cm soil layer, as shown in Figure 4a. The correlation between model outputs and observations consistently indicates a high correlation over most places, except some coastal areas and the Tonle Sap Lake, where the SMAP measurements over land are of inferior quality due to the presence of water bodies (Figure 5b.). To circumvent the limitation and reliability of SMAP soil moisture data near water bodies, the direct measurement of soil moisture and precipitation, and the exploitation of alternative remote sensing tools that are less susceptible to water interference can be employed for improved validation. However, in our case, those regions account a mere proportion of the total area, thus making the comparative

analysis of SMAP surface soil moisture with model simulated soil moisture an effective and reliable tool for ensuring consistency in model predictions. As the SMAP soil moisture is an independent product, it was deemed appropriate for validation against the VIC-generated profile soil moisture outputs.

In this study, we opted for the Root Mean Square Error (RMSE) and unbiased Root Mean Square Error (uRMSE) for model evaluation, owing to their straightforward interpretability, sensitivity to extreme soil moisture values critical for drought assessment, and capacity for absolute error quantification, offering an unambiguous measure of model accuracy. The uRMSE, in particular, demonstrates robustness as it remains unaffected by mean values, ensuring consistent performance evaluation across diverse datasets. Moreover, our focus was on the precise estimation of soil moisture magnitudes — essential for agricultural applications — rather than the temporal alignment of wet and dry periods, a focus of metrics like the Nash-Sutcliffe Efficiency (NSE) and Kling-Gupta Efficiency (KGE). RMSE/uRMSE also facilitate universal benchmarking, with clear performance criteria that enhance the communicability of results to both technical and non-technical stakeholders. While these metrics primarily gauge the magnitude of errors, they complement other statistics that assess different aspects of model performance, contributing to a comprehensive evaluation framework.

In addition, validation also examines the model performance with observations. This is essential to ensure the accurate characterization of the drought indices from the validated parameters. As a result, the moisture deficit in the soil can be used to define agricultural drought and provide critical information for early warning. However, it's essential to note the limitations of our validation process. Due to a lack of access to other vital hydrologic observations, such as streamflow, our validation was exclusively confined to soil moisture.

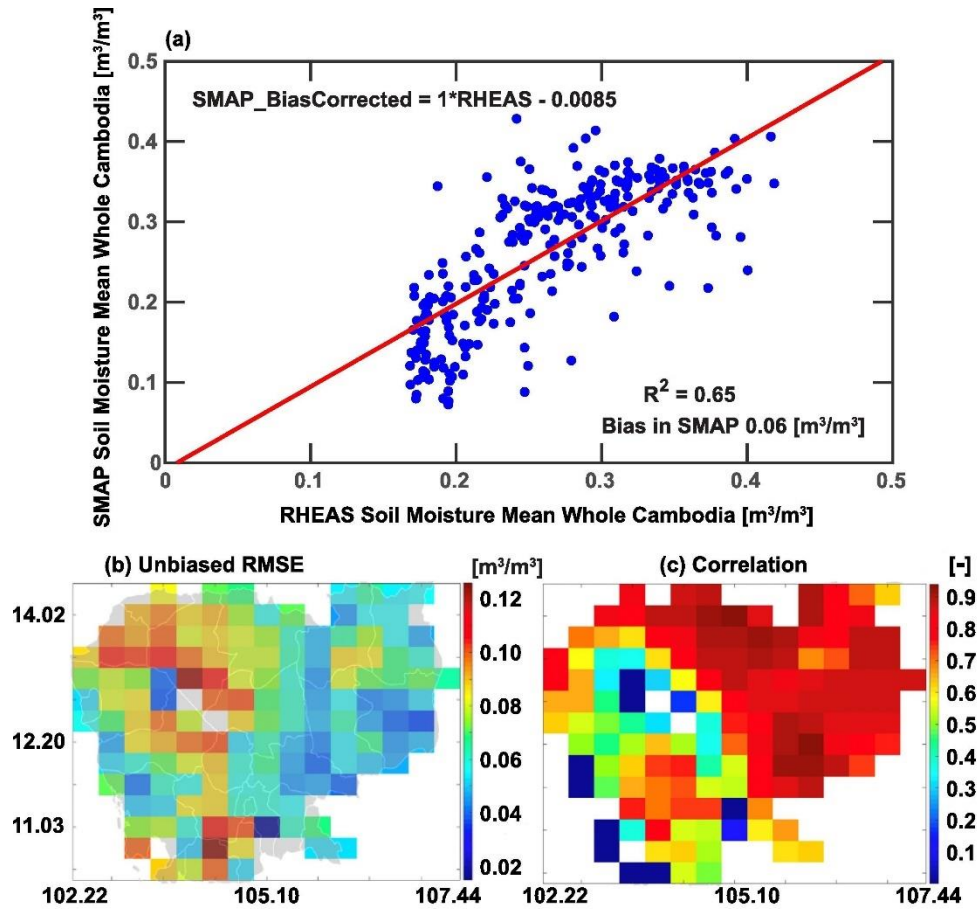


Figure 4. (a, top panel) RHEAS-generated surface soil moisture compared with SMAP soil moisture in 2017 over Cambodia (b, bottom panel) Unbiased RMSE and (c, bottom panel) Correlation for RHEAS-based surface soil moisture for 2016 and 2017 data when compared with the SMAP surface soil moisture data (L3_SM_P) at 36 km EASE2 grid resolution.

2.3.1.2. Crop Modeling Validation

For the period between 2000 and 2016, we carried out a regional-scale validation using end-of-season rice yields, aggregated at the country level for Cambodia. The interannual rice yields from *m*-DSSAT over Cambodia were validated against an independent source of yield records obtained from the FAO. Figure 5. presents a comparison of observed and model-generated annual end-of-year yields. Notably, the soil and phenological data were collected for the period 2010–2013, used in Wang et al. (2017). Using the same genetic coefficients in our simulations, the rice

yields estimated. The results, as shown in Figure 5., exhibited a good correlation ($R^2 \sim 0.84$) with very low bias between observed and RHEAS simulated yield estimates. However, extending the simulations to an earlier timeframe (2000–2008) (Figure 1., Appendix A) showed a substantial bias compared to the latter part of the simulation period, i.e., 2008–2016. The large bias (~750 kg/ha) in the rice yields observed for the initial 5 years, i.e., between 2000–2004 can be attributed to the following: i) lack of farm management data, such as fertilizer application rates, and ii) subpar rice cultivar genetic coefficients. While we utilized calibrated cultivar genetic coefficients from Wang et al. (2017) throughout our simulation duration, the lack of specific data about the rice cultivars used in the early 2000s introduces a layer of uncertainty. A possible shift in rice cultivars during these years could explain the observed yield discrepancies. Incorporating a calibrated cultivar variety for years 2000–2004 could potentially improve the rice yields. However, such calibration would require data currently unavailable to us.

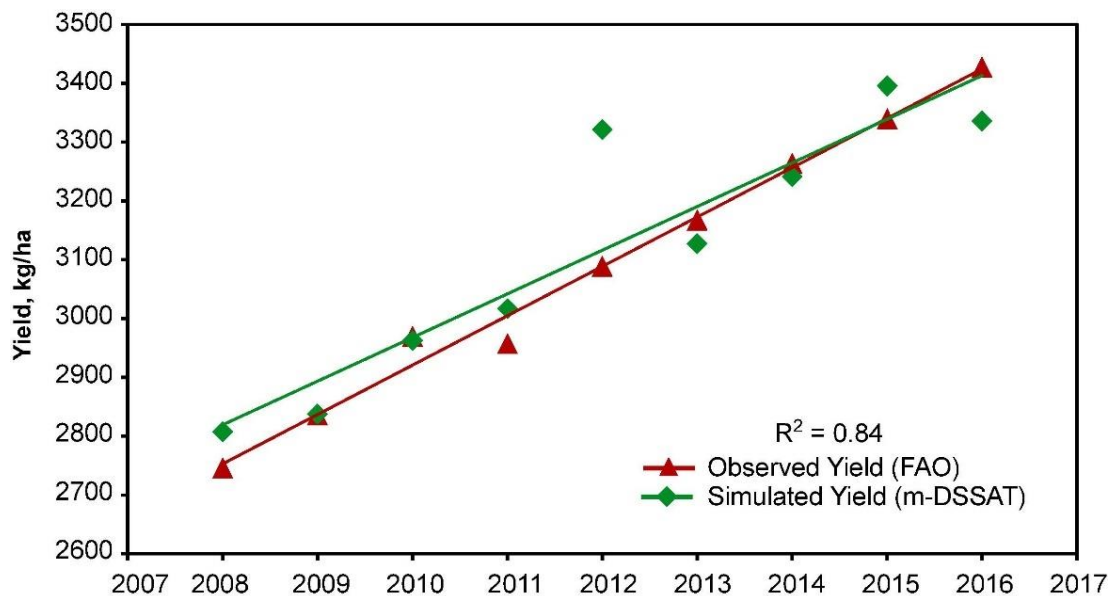


Figure 5. Comparison of RHEAS (*m*-DSSAT)-simulated rice yields with actual observations (FAO) over Cambodia between 2008 and 2016.

2.3.2. Modeling Results

2.3.2.1. Hydrological Modeling

The VIC hydrologic model served as the foundational tool to derive vital drought monitoring indices for the LMB region. Figure 6. illustrates the sample outputs based on RHEAS drought indicators for 2014. This depiction provides an in-depth view of fundamental drought indicators, including the 3-month standardized precipitation index, agricultural drought severity, soil moisture deficit, dry spells, and standardized runoff index. A closer look at these indicators reveals that the Mekong plain, encompassing the low-lying areas of Cambodia, Lao PDR, and the Mekong delta, experienced significant stress throughout the year. The 3-month SPI, for instance, offers insights into short-term moisture conditions by aggregating seasonal precipitation probabilities. The drought severity, on the other hand, presents a standardized measure (i.e., fluctuations/anomalies) of moisture deficits (dry/wet) at the root zone. Notably, spatial averaging might cause the 3-month SPI to exhibit slightly subdued values; in certain grids, these values could have fallen as low as -3. The moisture deficit (SMDI) index appraises the potential crop water extraction amount from different soil depths during various stages of crop growth.

Interestingly, spatial patterns in the drought indicators underscore the vast variability in wet and dry conditions across the region. For instance, the northern parts of the LMB display an SMDI value of +3, indicating higher root zone moisture. Similarly, the bottom rows in Figure 6. shows the overall dryspells, annual precipitation totals, and the standardized runoff index (SRI). Furthermore, despite the diverse methodologies behind each index, a consistent pattern emerges across all of them. As most of the drought indices are based on the core hydrologic variables (such as precipitation and soil moisture), the indices follow a kindred behavior. This behavior is quite evident, wherein the areas under stress remain congruent for all the indices irrespective of their

nature of quantification (e.g., ‘precipitation’ for ‘SPI’, and ‘soil moisture’ for ‘SMDI/Severity’, etc.). For the snapshots shown in Figure 6., RHEAS showcased reasonable performance in capturing the drought stresses. As shown, all the indicators exhibited uniform stress patterns in drought-affected areas. Whether looking at the 3-month SPI, precipitation, or SRI3 indicators, the consistent spatial stress over regions such as Cambodia, southern Lao, the lower parts of Thailand, and the Vietnam delta is undeniable. Similarly, SMDI and drought severity depicted a similar behavior over the same regions.

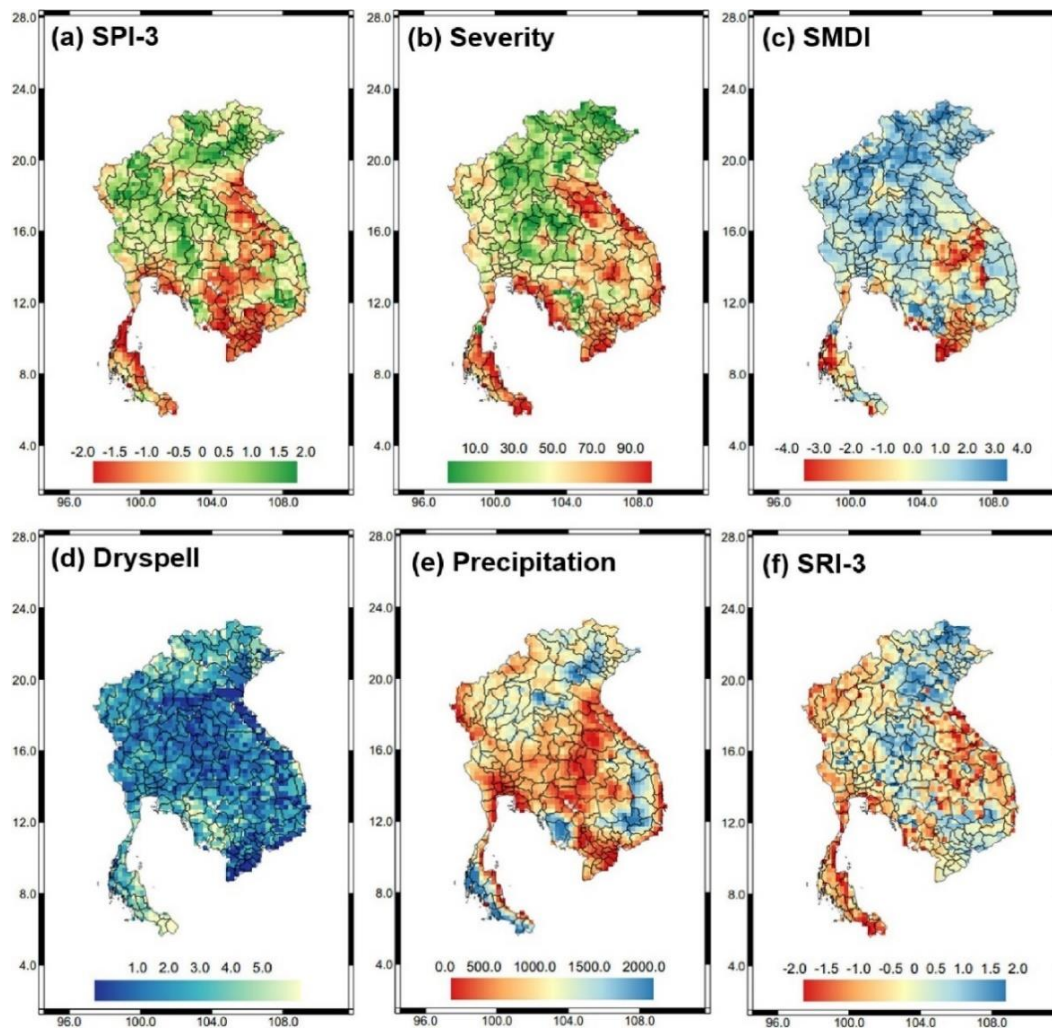


Figure 6. Spatial distribution of RHEAS drought products over LMB: (a) 3-month Standardized Precipitation Index (SPI3) for March-April-May (MAM) (b) Agricultural drought severity for

Figure 6. (cont'd)

MAM (c) Soil Moisture Deficit Index values (SMDI) during the last week of May (d) Dry spells for the MAM period (e) Precipitation totals (in mm) for MAM (f) 3-month Standardized Runoff Index (SRI3) for MAM.

To further dissect the persistence and behavior of drought in the region, an extensive study focused on Cambodia was undertaken. Figure 7. illustrates the differences in the behavior of SPI and SRI indices over Cambodia at various timescales for 1990–2019. As SPI and SRI indices are based on the standard deviations of precipitation and runoff respectively from the long-term median values, they were pivotal for characterizing drought on various timescales and understanding the region's exposure to extreme conditions over three decades. In other words, the analysis indicated the exposure of the region to extreme dry/wet conditions over a period of 30 years. Particularly, the SPI3 exhibited a frequent change in dry/wet conditions than the longer duration equivalent (i.e., SPI6 and SPI12). The SPI3 (and SPI1) were more effective in capturing the precipitation trends and short-to medium-term moisture availability, especially during the rice-growing season. The SPI3 and SRI3 better captured the seasonal anomalies (proportion of dry or wet months) within a year, notably during the 1988–99 and 2015–16 drought events (Guo et al., 2017). However, these shorter timescale indices (SPI1/SRI1 or SPI3/SRI3) can be misleading in regions with normal to mild wetness (or normal dryness) for a particular 3-month period. This interpretation emerges due to potentially inaccurate accumulation of impacts on shorter time scales, wherein the wetness (or dryness) for that 3-month period depicts a temporary wet (dry) period. Hence, the larger timescale meteorological indices (SPI6/SRI6 or SPI12/SRI12) are much more effective in quantifying long-term wet or dry season trends. The SPI and SRI exhibit a similar pattern at larger timescales (12-month). The 3-month SPI and SRI behave distinctly wherein the

SRI tends to be less variable than SPI due to the storage of incoming precipitation as soil moisture that limits the amount of surface runoff. This behavior can be clearly seen during the drought events of 1998–99 and 2015–16, where the 3-month SPI and SRI follow a slightly different pattern with a lag as compared to the respective 12-month timeframe. As the 12-month indices are based on the cumulative result of shorter periods, SPI and SRI show a lower frequency of positive and negative standardized values when compared to the 3-month values. Figure 2. Appendix A shows the mean annual precipitation and runoff over Cambodia for the years 1990–2019. There was a clear indication of seasonal influence in the region, with prevailing wet periods at the start of the growing season (June onwards), and subsequent dry conditions from Nov-Mar.

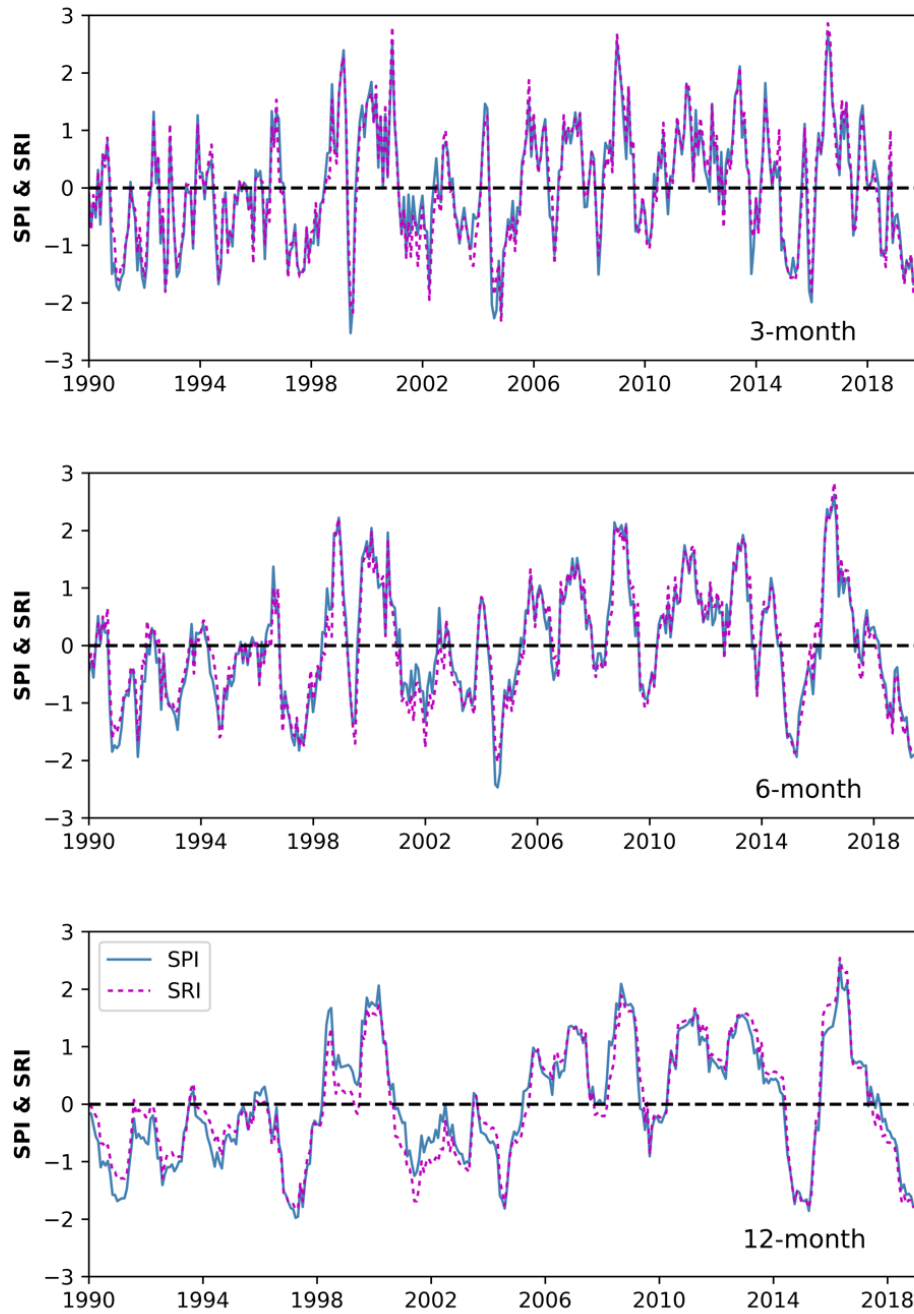


Figure 7. Historical temporal variations of SPI and SRI values at 3-, 6- and 12-month timescales over Cambodia.

Furthermore, a detailed regional study was carried out for the 2015 drought year, illustrated in Figure 8. The spatial variability in drought severity (intensity) for Mar-Apr-May (MAM) and

Jun-Jul-Aug (JJA) time periods was effectively well captured by RHEAS, with higher stress observed in the initial months compared to the latter (Figure 8a.). Similarly, the spatial distribution of the SMDI showed good agreement with drought severity (Figure 8c.). A comparison of the moisture deficit between the 22nd week (i.e., the 4th week of May) and 33rd week (i.e., the 2nd week of Aug) clearly showed higher moisture deficit during the initial period of the year. Though the overall annual precipitation totals increased over the years (except for the drought years), the interannual spatial variability shows lower levels of rainfall during MAM time period as compared to JAS time period (Figure 9d.). The 3-month SPI during the initial months (FMA) exhibited a relative moderate to severe dry conditions over the region while the latter period (JJA) presented a near normal to moderate wet conditions due to the onset of the southwest monsoon (Figure 8b.). The SPI3 index provides a better estimation of precipitation deficit and short-to-medium term moisture conditions and is considered critical in capturing the precipitation trends during reproductive and early grain-filling stages. Similarly, dryspells are used to detect significant changes in drought frequency due to the fluctuations in water/energy fluxes. Overall, there had been a consistent prevalence of high-stress conditions during the initial time period (April onwards) as compared to the mid-months, with extreme stress conditions in the southeast provinces of Prey Veng and Takeo, and normal to moderate dry conditions over the western provinces of Pailin and Battambang.

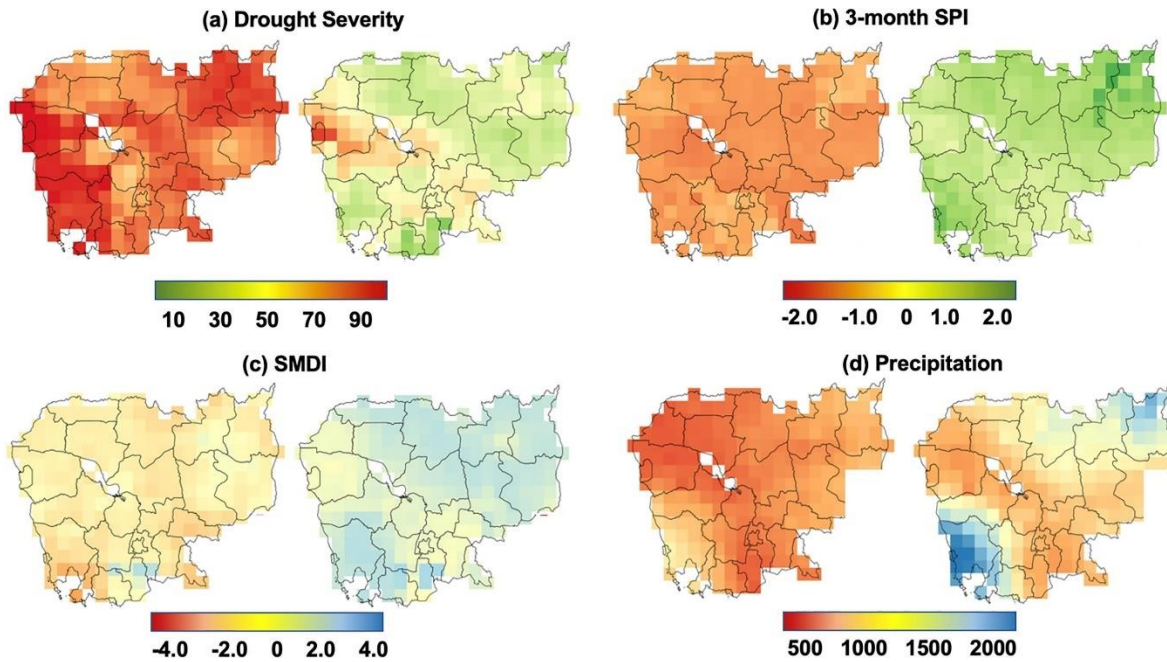


Figure 8. Drought data products: Maps of a) Agricultural Drought Severity for MAM and JJA b) 3-month Standardized Precipitation Index (SPI3) during FMA and JJA c) Soil Moisture Deficit Index during 22nd and 33rd week d) Precipitation totals during MAM and JAS, showing the interannual variability for the 2015 drought year over Cambodia.

2.3.2.2. Crop Modeling

The secondary objective of this investigation entailed the quantification of the interannual variability in rice yields across Cambodia. Such quantifications serve a dual purpose: offering agricultural stakeholders' insights into yield trends and aiding policymakers in devising strategies to curtail agricultural deficiencies resulting from drought adversities. Figure 9. presents the simulated annual yields over Cambodia for two pivotal years, 2005 and 2015. There is a clear indication of an increase in yields (2900 kg/ha in 2005 to 3550 kg/ha in 2015) over the years; however, the drought-striven provinces manifested consistent low yield trajectories across the decade. This stagnation is primarily attributed to recurrent and extended dry spells characterizing these regions. Conversely, regions encompassing the Tonle Sap basin and nearby areas manifested

a positive yield trend over the decade. This enhancement can be partially ascribed to advancements in agricultural technology and optimized agronomic practices. Figure 3., further detailed in the Supplementary Material, offers a spatial breakdown of average yields across Cambodian provinces for the 17-year study window (2000–2016), clearly indicating low yields in the southeast regions. This observation harmonizes with drought metrics procured from the VIC hydrological model. Regions with pronounced precipitation deficits and heightened drought severity invariably reported compromised agricultural outputs. However, the western provinces, exemplified by Battambang, registered superior yields even amidst considerable drought exposures. Such anomalies can be rationalized by regional soil heterogeneities and the inherent adaptability of paddy cultivars to localized conditions. Ongoing endeavors are channeled towards fine-tuning the RHEAS framework on a regional granularity to enhance its efficacy and precision. The forthcoming phase of research aims to incorporate and assimilate diverse state variables, such as the Normalized Difference Vegetation Index (NDVI), to further refine the model's predictive capabilities.

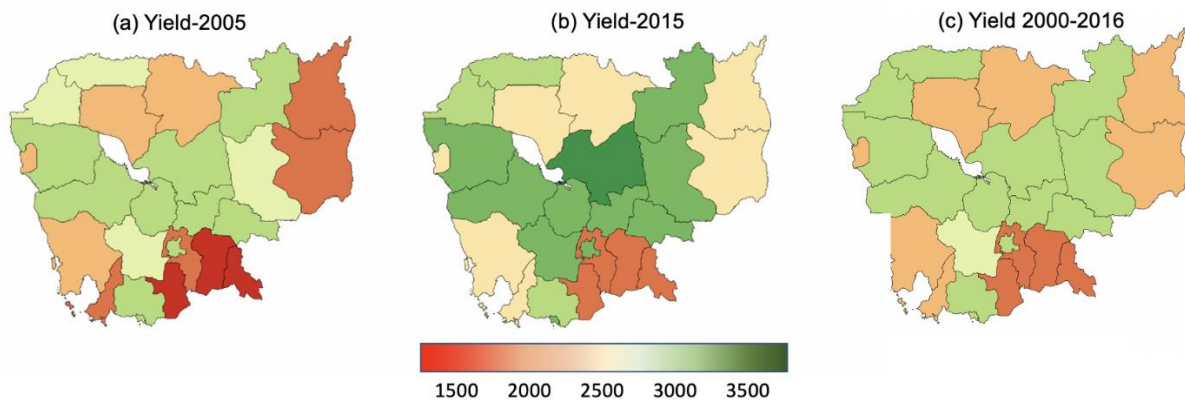


Figure 9. Provincial rice yields (from *m*-DSSAT) over Cambodia in 2005, 2015, and 2000-2016.

2.3.3. Correlation with Crop Yield

Figure 10. presents an intricate comparison of the impacts of rice-growing season (JJASO) yields in tandem with the standardized drought indices and hydrologic variables, encapsulating the period from 2000 to 2016 over Cambodia. For this evaluation, parameters such as drought severity, cumulative precipitation, and the frequency of days registering temperatures surpassing 30 °C during the growing season were meticulously assessed against both observed and model-simulated yields. While simulations based on crop models anticipate water supply substantially driving crop yields, capturing the relationship between weather and crop yields shows no clear association. As shown in Figure 10., the climatic patterns inferred from drought indices remained relatively consistent, exerting a negligible influence on the annual rice yields. A paradigmatic instance of this observation is the 2015 drought year. Contrary to expectations, annual yields amplified despite the pervasive drought conditions blanketing the nation. However, this situation can be better interpreted by the decrease in yield in 2016, mainly due to the persistence of drought spell from 2015 until mid-2016.

Overall, we see a consistent upward trajectory in annual yields irrespective of the stress conditions, and no conclusive pattern in the behavior of drought parameters and crop yields can be deduced on a regional basis. Hence, more pronounced, and discernible trends might emerge at finer spatial resolutions, such as provincial scales, as highlighted in Figure 3., Appendix A. While simulated yield data was available for specific provinces like Pailin and Preah Vihear, the absence of empirical yield data at these scales stymied our capacity to draw definitive conclusions. Notably, there was good correlation between severity index and accumulated precipitation, with the prevalence of elevated stress conditions (i.e., high severity and lower precipitation) during the onset of growing season in June. These stress conditions gradually reduced in the ensuing months,

culminating in diminished severity and increased precipitation by September and October. Furthermore, an exploration into the number of days registering average temperatures exceeding 30 °C (as illustrated in Figure 10e.) during each growing season was undertaken to gauge potential thermally induced stresses on crops. It is well-established that crop energetics, from sowing to harvest, are quantified using the Growing-Degree-Day (GDD), which is computed leveraging maximum, minimum, and mean ambient temperatures. However, the plant productivity (photosynthetic capacity) almost becomes negligible or diminishes drastically above an average temperature of 30 °C, impacting productivity. This threshold was thus utilized to ascertain thermal stresses on crops. The analysis showed a record prevalence of days surpassing this threshold during stress periods, most notably during the 2015–2016 drought phase.

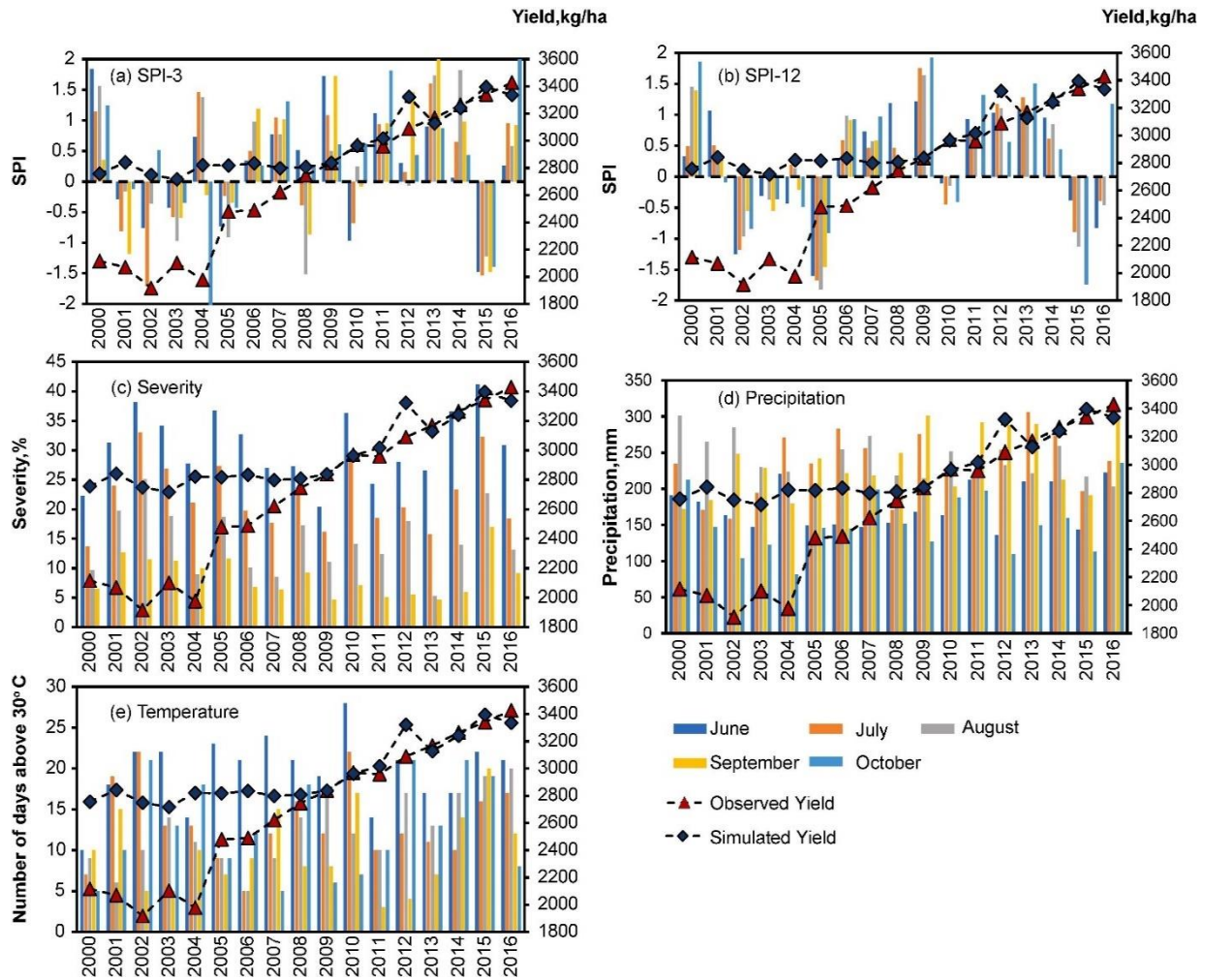


Figure 10. Comparison of yield responses to: (a) growing season 3-month SPI (b) 12-month SPI (c) agricultural drought severity (d) precipitation totals (e) number of days above 30 °C for the growing season (Jun-Oct) from 2000 to 2016.

Likewise, the standardized drought indices, encompassing 3 and 12-month SPI, along with drought severity and accumulated precipitation, were meticulously evaluated. These indices provide a holistic understanding of the evolving drought dynamics, offering invaluable insights for agronomists and policymakers to devise informed strategies. Figure 10a. and 10b. provide a visual representation of the 3- and 12-month Standardized Precipitation Index (SPI) juxtaposed with rice yields during the growing season, which spans from June to October. The 3-month index

was used to capture the short- and medium-term moisture conditions and estimate the precipitation trends/dynamics during the important crop growth stages. In contrast, the 12-month SPI provides an aggregate perspective, shedding light on enduring precipitation patterns, effectively broadening the temporal lens. Additionally, RHEAS was able to capture the decrease in yield corresponding to drought events manifesting in 2002, 2004 and 2008. However, it exhibited a contrasting behavior in the context of subsequent droughts, specifically those occurring in 2010 and 2015. It should be noted that the agreement between the drought indices and yield only accounts for the physical changes within the growing season. Nonetheless, post-2005, drought variability during the growing season consistently hovered in the spectrum from normal to mildly wet (as evidenced by SPI values surpassing 0). This trend aligns with the upward trajectory observed in rice yields throughout these years.

As shown in Figure 10., the growing season (Jun-Oct) exhibited negligible impact from water stress conditions. However, the study does not categorically indicate the prevalence of stress-free conditions throughout the study period. Instead, as shown in Figure 8., stress conditions mostly occurred during the initial months (Feb-May) which subsequently alleviated to mild/near-normal conditions at the onset of the growing season. As the drought conditions were notably significant in capturing the interannual yield variability, we evaluated other management factors (such as fertilizer application rates, cultivar varieties) that could possibly drive the associated changes. Since the meteorological variables (e.g., precipitation, ambient temperature, etc.) and cultivar varieties (from Wang et al., 2017) remained same throughout the study period, the only variable factor in the analysis was the rate of fertilizer application (adopted from FAO/World Bank).

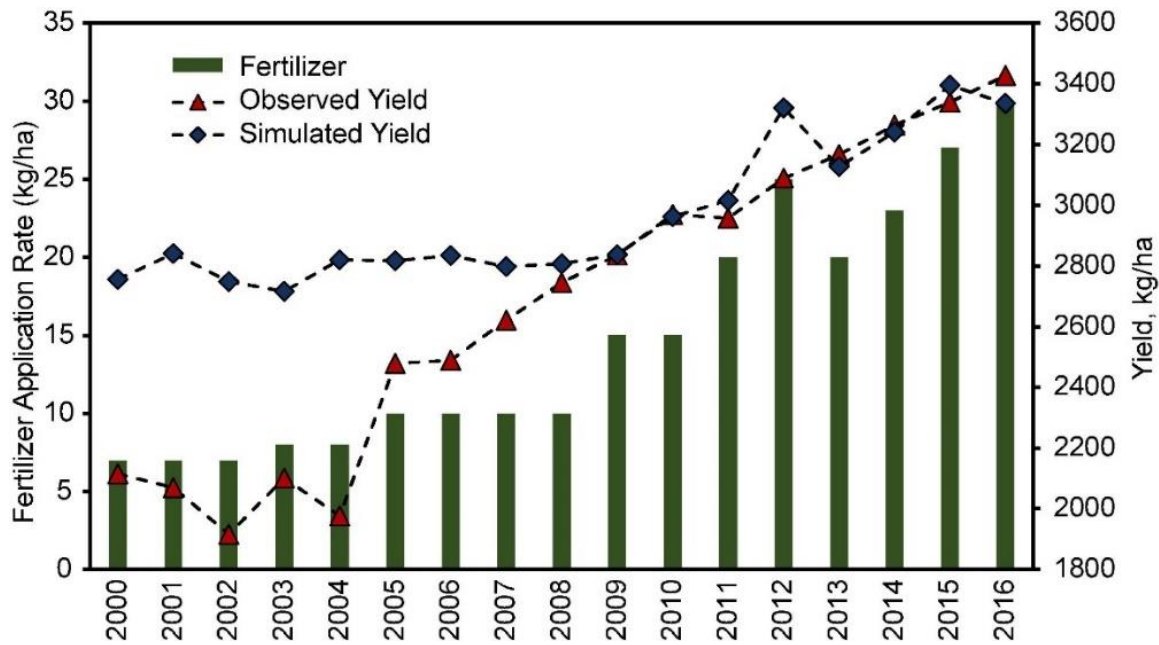


Figure 11. Fertilizer application rates for the study period (2000–2016) compared with annual rice yields over Cambodia.

From 2008, there has been a steady increase in fertilizer application rates, echoing a strong resemblance with observed and simulated yields, as visualized in Figure 11. A strong correlation is noted between observed yields and fertilizer application rates during the 2008–2016 period ($R^2 \sim 0.84$) with a similarly high value between simulated yields and fertilizer rates ($R^2 \sim 0.92$). Although less information was available on the local application rates, a significant pattern can be noted from the analysis. With increased consumption of fertilizer rates since 2008, the interannual yields have steadily increased each year. Prior to the aforementioned period, the fertilizer application rates remained the same, hence the initial timeframe, i.e., 2000–2005, arrayed lower yields. As the management practices (e.g., fertilizer application rates) in a country depend on the socioeconomic conditions of the farmers and the support of local governments, the fertilizer application rates in Cambodia have still not reached optimum levels (compared to the neighboring Mekong countries such as Vietnam and Thailand) and exhibit abundant scope for improvement in

the near future. Hence, considering the results obtained from RHEAS, the consistent increase in yield can be significantly attributed to the application of chemical-based fertilizers.

2.4. Conclusion

The RHEAS framework is uniquely designed, offering holistic modeling capabilities that enable a detailed comprehension of the interplay between evolving meteorological conditions, hydrological phenomena, climate variations, and agronomic practices, and their collective impact on crop growth and yield outcomes. Based on the current application, RHEAS showed reasonable performance in capturing the intricacies of both agricultural and meteorological drought variabilities, as inferred from the analyzed drought indices.

The results from the study depicted a common persistence of mild to medium dry conditions over Cambodia during the initial months. However, this preliminary stress tends to attenuate with the commencement of the monsoon season and the subsequent rice-growing phase, which initiates in June. In essence, the February to May window is characterized by pronounced hydrological stress, but the ensuing period, starting June, typically evolves towards near-normal or even moderately wet conditions. While the mean annual precipitation metrics have showcased a positive trend over the years, interrupted sporadically by certain drought events, there's a discernible decline in dry season rainfall, attributable to erratic weather patterns. The consecutive and severe drought episodes, especially the 2015-16 event, have intensified apprehensions regarding the recurrence of such adverse conditions in regions inherently susceptible to drought.

Considering the results obtained from RHEAS, there is a clear indication of a significant increase of drought prevalence across vast swathes of the LMB, that may or may not have serious implications on the agriculture sector. Remarkably, the rice yields in Cambodia have charted a consistent upward trajectory over the past decade. This positive shift can be primarily attributed to

escalating fertilizer application rates, complemented by the progressive modernization of agricultural methodologies, the advent of mechanization, and technological advancements. Corroborating these findings, the Mekong River Commission, in their annual report (MRC, 2014), highlighted a similar uptrend in the annual output of major crops, inclusive of rice. RHEAS was able to adequately capture the interannual variability of rice yields with observations ($R^2 \sim 0.84$) with low bias from 2005 onwards. The growing season hydrologic stress conditions when compared with annual yields also did not reflect any significant effect on the crop status barring a few drought years. Nonetheless, the drought indices exhibited lower agricultural yields in the water-stress provinces, predominantly located in Southeast Cambodia, reflecting the impact of drought on crop status at finer, provincial scales. While the aforementioned situation envisages lower future yields due to elevating stress conditions, substantial arguments can be made on comparing drought conditions with annual yields. Drought solely does not affect agricultural yields as Cambodia inherently gets affected by flash flooding/waterlogging, possibly resulting in decreased yields and crop losses. As the meteorological variables and cultivar type were kept the same throughout the study period, the only variable factor for attaining higher yields can be attributed to the use of chemical fertilizers. Although the fertilizer application rates in Cambodia are well behind the neighboring Mekong countries (Vietnam and Thailand), there is much scope for improvement in the future.

Chapter 3.

EFFICACY OF SEASONAL CLIMATE FORECASTS IN PREDICTING INTERANNUAL CROP YIELDS

This chapter was published in Agricultural and Forest Meteorology, Vol 341, Abhijeet Abhishek, Mantha S Phanikumar, Alicia Sendrowski, Konstantinos M Andreadis, Mahya GZ Hashemi, Susantha Jayasinghe, PV Vara Prasad, Roberts J Brent, Narendra N Das, Dryspells and Minimum Air Temperatures Influence Rice Yields and their Forecast Uncertainties in Rainfed Systems, 109683, Copyright Elsevier (2023).

As discussed in the first chapter, which delved into the historical trajectory of drought and its consequent ramifications on agricultural crop outputs, the ensuing chapter pivots towards a more prospective lens: the forecasting of agricultural yields. The primary objective of this chapter is to meticulously examine the forecast proficiency related to crop yields, and quantify the inherent uncertainties encapsulated within such forecasting endeavors.

3.1. Introduction

The Earth is witnessing a series of unprecedented shifts in climate patterns, escalating the frequency and severity of extreme weather events globally. These alterations, predominantly driven by anthropogenic influences, have led to multifaceted consequences, ranging from ecological disturbances to socio-economic ramifications. A paramount concern emanating from these trends are notably characterized by elevated temperatures, erratic precipitation patterns, altered seasonality, and heightened susceptibility to extreme meteorological phenomena such as droughts and floods (Alizadeh et al., 2020; Clarke et al., 2022). These climatic disturbances threaten agricultural resilience and jeopardize food security, especially in vulnerable regions like Asia, Africa, and Latin and South America (Challinor, 2011; Jägermeyr et al., 2021). Here, the

confluence of increasing natural and anthropogenic factors have amplified the crisis, particularly in the realm of food production (Ahmadalipour et al., 2019; Kang et al., 2021).

Climate and weather conditions significantly influence crop growth and development, underscoring the importance of accurate and timely weather forecasting for effective agricultural planning. Seasonal climate forecasts provide crucial insights for rainfed agricultural systems, aiding farmers in optimizing practices based on expected climate patterns (Lal and Stewart, 2018). Timely access to critical weather information, such as rainfall and temperature variability, enables informed decisions, optimizing practices from planting to post-harvest activities. Particularly, farmers can adjust strategies based on expected patterns selecting drought-tolerant varieties during anticipated dry periods or aligning harvest schedules with impending extreme events. Furthermore, a reliable forecast and an understanding of climate-crop links enhance productivity and risk minimization. For instance, forecasted yield estimates allow farmers to assess production and income risks, prompting adaptive strategies like crop insurance and diversification. Likewise, anticipating market conditions enables informed decisions on timing, volume, and pricing, fostering stability. Numerous earlier studies have demonstrated that improved yield forecasts enhance farmer's decision-making abilities (Australia: Brown et al., 2018; India: Kushwaha et al., 2022; United States: Lacasa et al., 2023). In contrast, unreliable forecasts can result in suboptimal practices such as inadequate water management, inaccurate fertilization, incorrect planting dates. Although, uncertainties in seasonal climate forecasts, stemming from meteorological influences and/or model intricacies, are inevitable, identifying the sources of uncertainties in seasonal climate forecasts, and implementing mitigation strategies are vital for effective decision-making and ensuring sustainable crop productivity.

Substantial research underscores the intricate relationship between climate variations and crop yields, particularly concerning staple crops like maize, rice, wheat, and soybeans (Innes et al., 2015; Mavromatis, 2015). As outlined in many of these studies, there is compelling evidence of the adverse effects of changing climatic conditions on interannual crop yields, thus emphasizing the urgent need to address this burgeoning crisis of food security (Minoli et al., 2022; Vogel et al., 2019; Rifai et al., 2019; Shukla et al., 2019). Factors such as shortened growing seasons, altered precipitation patterns, warmer growing season temperatures (i.e., the Growing Degree Days, GDD), and heightened extremities of flood and drought occurrences threaten water availability and agricultural productivity in developing agrarian nations (Challinor et al., 2016; Ray et al., 2015; Zhao et al., 2017). For instance, Ortiz-Bobea et al. (2021) observed a temperature-related linear decline in agricultural productivity over the growing season, varying across countries. Hence, the existing literature underscores the urgency of addressing the impacts of climate variability on agricultural crop yields through sustainable and climate-resilient approaches.

Skillful forecasts of crop yields at seasonal and sub-seasonal timescales are essential for facilitating improved crop management and decision-making given the myriad uncertainties in agricultural production (Basso and Liu, 2019; Togliatti et al., 2017). Weekly to monthly yield projections not only ensure food security but also provide an overview of the weather conditions in the adjacent growing season. Such critical pieces of information can allow governments, research institutions, and policymakers, to act accordingly, and implement effective adaptation measures to build resilient agricultural systems (e.g., early warning systems for extreme events, water management techniques, etc.). Numerous previous studies have forecasted crop yields at varying spatial and temporal scales using process-based models and satellite remote sensing techniques (Karthikeyan et al., 2020; Konduri et al., 2020). However, relying on a combination of

approaches—use of computational modeling, data analytics, automation, advisory systems, remote sensing, and appropriate agronomic practices—can yield better results compared to a single approach. Incorporating satellite remote sensing data into process-based models allows for a comprehensive assessment of crop conditions and facilitates targeted interventions. Generally, satellite imagery captures the vegetation dynamics, and crop attributes including leaf area index (LAI), land surface temperature at varied temporal and spatial coverages monitor the real-time crop growth and yield potential. Additionally, the weather forcings (e.g., precipitation, air temperature, etc.) that are provided to the models, assimilation of remotely sensed state variables (e.g., soil moisture), calibration and validation of the models, among others are some of the key components used to capture crop characteristics (e.g., phenology, yield, etc.). Similarly, combining satellite datasets with process-based models leads to improved simulation accuracy while enhancing mechanistic understanding of the physiological and biochemical processes of crop growth. Specifically, they track the crop development (phenology) over the growing season based on climate, soil properties, agronomic practices, etc. Likewise, the use of data (statistical) analytics (e.g., historic trends/patterns of cropping), and knowledge about local agronomic practices (e.g., planting/sowing date, fertilizer application, cultivars, etc.) enables accurate predictions, optimized resource allocation, whilst identifying yield gaps and inefficiencies.

Harnessing an integrated approach amalgamates the advantages of diverse techniques, bolstering decision-making efficacy and operational efficiency. In other words, the integration of advanced modeling techniques with real-time climate data offers farmers the potential to enhance productivity, reduce risks, and improve resource management. Particularly, by combining hydrologic and crop models, we overcome the limitations of representing the complex hydrological processes solely within the crop models and get a more accurate representation of

water availability and drought conditions, which are essential factors influencing crop growth and productivity. The hydrologic model provides valuable insights on the soil moisture dynamics and water availability (compared to the simplified soil water balance module employed in crop models), enabling us to assess the water requirement for crop growth and development during periods of stress. Similarly, a combination of data analytics and automation enables the optimization of agricultural operations by analyzing large sets of data and leveraging technology to streamline processes, maximize productivity, and reduce wastage. Recently, several studies have incorporated new technologies such as machine learning, and deep learning specifically using neural networks, that sample the input variable space for a particular vegetation type (Ermida et al., 2017). These new systems boast sophisticated algorithms that provide the flexibility to analyze a large and diverse set of datasets, enabling a holistic understanding of agricultural systems. Such studies have exhibited improved prediction of interannual crop yields, when remote sensing information, and accurate agronomic management information, were supplemented with deep learning and machine learning-based models (Feng et al., 2020; Paudel et al., 2023). Although integrated approaches demonstrate improved accuracy, their implementation is often hindered due to technical complexities, data availability and quality (e.g., inconsistent data), expensive computational (and maintenance) requirements, and contextual factors (e.g., differences in local agronomic practices).

While existing forecasting methods have advanced our understanding of crop yield dynamics, there are several limitations that hinder the accurate and timely estimation of yield projections, particularly in regions vulnerable to climate extremes. The LMB, specifically Cambodia, serves as a compelling case study, as it has experienced significant impacts from weather extremities and anthropogenic factors, leading to adverse effects on rice production and

the livelihoods of farmers (Guo et al., 2017; Hoang et al., 2019; Son et al., 2012; Thilakarathne and Sridhar, 2017). Hence, addressing these limitations requires advancements in data availability, innovative modeling approaches, and robust stakeholder engagement.

Implementing the abovementioned approaches would greatly benefit in mitigating the impacts of such natural climatic shocks, whilst significantly reducing crop losses. Generally, the impact of climate-related variables that are subjected to a higher degree of uncertainty can be marginally modulated/controlled during the growing season through a multi-approach framework (Hansen et al., 2006). Hence, changes in crop phenology and development, especially from grain-filling until mature/harvest, are highly dependent on weather, so having a reliable long-term seasonal forecast of weather variables (e.g., precipitation and temperature) is key to achieving optimal yield forecasts. Generally, the meteorological data (input) forced into the model often contribute a significant proportion of uncertainty (Müller et al., 2021). This can be easily interpreted by referring to Figure 12., which shows a considerable portion of uncertainty (highlighted in blue color) at the start of the growing season that gradually decreases over a period of time until it reaches crop maturity/harvest. While uncertainties are an inherent challenge, historical weather data was employed in this study to delineate climate variability-associated uncertainties (illustrated by the white dashed boundary around the blue in Figure 12.). This study focused in on quantifying the uncertainties associated with climate inputs.

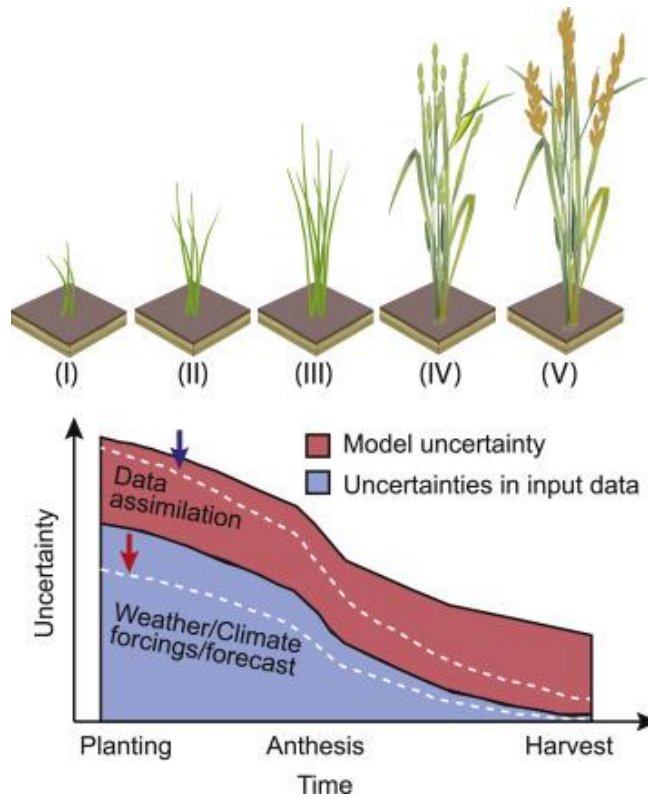


Figure 12. Anatomy of crop yield prediction: The stacked plot (bottom left panel) exhibits the contribution of uncertainties, either from model parameters or input forcings, based on various stages of crop (rice) growth: (I) Seedling, (II) Vegetative stages (Tillering, Stem elongation, Panicle Initiation), (III) Reproductive stage (Booting, Heading, Flowering), (IV) Grain filling, and (V) Maturity (top left panel). The white dashed line represents the reduction in model uncertainty (red) through data assimilation while the white dashed line in input climate forcings (blue color) represents the significant reduction in uncertainty using multi-model ensembles of seasonal climate forecast data, over the growing season, i.e., from planting through harvest.

Forecast data holds potential to minimize the detrimental effects of adverse weather, such as drought stress, on crop phenology, translating to decreased losses and higher yields. Hence, recognizing the importance of such geophysical changes on specific sectors (e.g., agriculture, water resources) becomes imperative for understanding the complex web of relationships between

these environmental processes and crop yields in the forecast mode. While these large-scale geophysical systems are intricate and often under-analyzed, the full potential of representing physical processes and interactions in computational models remains untapped. This missing information can therefore have a large impact on understanding and quantifying the links between different geophysical parameters (Goodwell et al., 2020). Information Theory (IT) provides a comprehensive method to characterize the cause-and-effect relationships of large-scale geophysical processes through the extraction of information from single/multiple signal(s) (i.e., time series of a state variable) (Shannon, 1948). To date, a substantial amount of work has been done in hydrology using IT (Sendrowski and Passalacqua, 2017; Thiesen et al., 2019), however, no studies have implemented such an approach on an agro-hydrologic coupled system. As coupled systems have non-linear relationships between them, IT provides a systematic quantification of the missing links between various physical processes.

The primary goal of our study was to evaluate the probabilistic seasonal rice yield forecasts (for the growing season) over Cambodia using the historic North American Multi-Model Ensemble (NMME) forecast datasets (Jha et al., 2019; Slater et al., 2019). We formulated the following research questions: i) How do uncertainties in seasonal climate forecasts of crop yields evolve as the season progresses? and ii) What are the dominant hydrologic variables or derived indicators that influence crop yield forecasts through the growing season? Based on our questions, we hypothesized that seasonal climate forecast (independent variable) improves crop yield (dependent variable) predictions over time as more weather inputs (i.e., observations) are available (from planting through harvest). Thus, we implemented this using the RHEAS integrated framework that provides an end-to-end composite hydrologic (and drought) and crop yield probabilistic nowcast (current) and seasonal forecast over Cambodia. Although RHEAS provides holistic real-time

monitoring of drought and yield nowcasts covering a wide range of areas, this study is a significant extension to the last study (Chapter 2), wherein we examine the efficacy/skill of seasonal forecasts based on a multi-model ensemble forecast system. Specifically, we implemented a novel approach to estimating yields based on past weather observations (NMME) up to a certain forecast date, and weather data from other years are sampled for the rest of the season. In other words, the model integrates the observed weather data (from NMME) up to the current date with the rest of the input data sampled from climatology for the remaining season. In the end, we get an approximation of the climatic component of uncertainty (error) during each time window that represents the discrepancy between simulated and observed yields. In short, by using up-to-date climate information, we account for the temporal variability of climate variables during the growing season, capturing the effects of changing weather patterns on yield outcomes. Furthermore, we investigated the performance (impact) of associated drought/hydrologic (forecast) variables with crop yields by measuring the shared information between crop yield and hydrologic (and drought) variables. The combination of such statistical technique with hydrologic-crop modeling, and seasonal climate forecast sets this research apart, offering a more comprehensive and dynamic understanding of the factors influencing crop yield predictions. While it is widely known that uncertainties decrease over time, the novelty behind the approach lies in its comprehensive assessment and identification of key variables that primarily influence the uncertainties in seasonal climate forecasts of crop yields over the growing season. In other words, the study provides a detailed analysis of how uncertainties evolve (over time), and identify the key geophysical variables (i.e., hydrologic variables and derived drought indicators) that influence crop yield forecasts throughout the growing season.

3.2. Materials and Methods

3.2.1. Study Area

The study region Cambodia, as shown in Figure 13., has a diverse and unique ecological landscape, often characterized by a low-lying central plain that is encompassed by uplands and low mountains. The central plains host the vast freshwater lake Tonle Sap and the upper stretches of the Mekong River. The presence of the Tonle Sap and the Mekong River greatly influences the region's hydrology, agricultural practices, etc. The Mekong River and its tributaries (and floodplains) are crucial for irrigation, transportation, and fishing activities, serving as the lifeline for the local communities. During the wet season, the Mekong River swells, causing its tributaries and the Tonle Sap to reverse their flow. This inundation enriches the surrounding floodplains with nutrient-rich sediment, creating fertile conditions for agriculture. Likewise, the Tonle Sap plays a crucial role in supporting diverse ecosystems. The lake acts as a natural reservoir, expanding and contracting with the annual monsoon rains and the flow of the Mekong River (MRC, 2014). In addition, the central plains serve as the agriculturally fertile areas, with regions around Tonle Sap and the deltaic regions of the Mekong River hosting large areas of rice, sugarcane, cassava, and corn production. Rice is the primary grown crop in the region, with a large proportion of the population dependent on it for their livelihood (Mainuddin et al., 2013). Typically, rice is grown twice a year—a) wet season (influenced by the southwest monsoon), which lasts from May to October, and b) dry season (influenced by the northeast monsoon), which lasts from November to April. The annual precipitation ranges between 1200 and 2500 mm/year, with ~85% of the rainfall during the wet season (Lauri et al., 2014). Likewise, the average annual temperatures and evaporation rates range between 21 to 35°C and 1,000–2,000 mm/year respectively depending on

the location (Liu et al., 2007). Overall, these prominent features greatly influence agriculture, hydrology, and the natural ecosystem in the region.

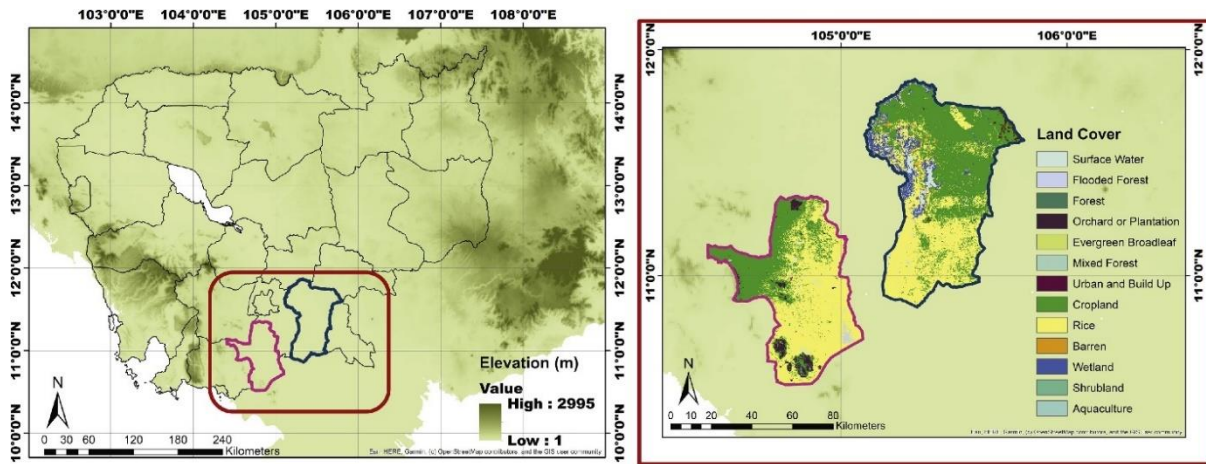


Figure 13. Map showing the topography of Cambodia, with the two study areas—Takeo and Prey Veng provinces—highlighted in red and blue respectively. The inset shows the general landcover pattern over both provinces.

3.2.2. Modeling Framework

The study implemented the state-of-the-art integrated RHEAS framework specifically designed to amalgamate information from the intricate components of the terrestrial water cycle, predominantly focusing on the soil-plant-atmosphere continuum. This synergy effectively facilitates the real-time monitoring of crop yields and corresponding hydrologic stresses. With its modularity, RHEAS stands out for its adeptness to seamlessly incorporate a myriad of datasets through a range of modeling systems. As discussed in the second chapter, the unique architecture of RHEAS is founded upon the loose coupling of the VIC hydrologic model with the *m*-DSSAT crop model. This coupling offers a holistic representation of both hydrologic and agronomic processes, with Figure 14. detailing the structural intricacies of the RHEAS framework.

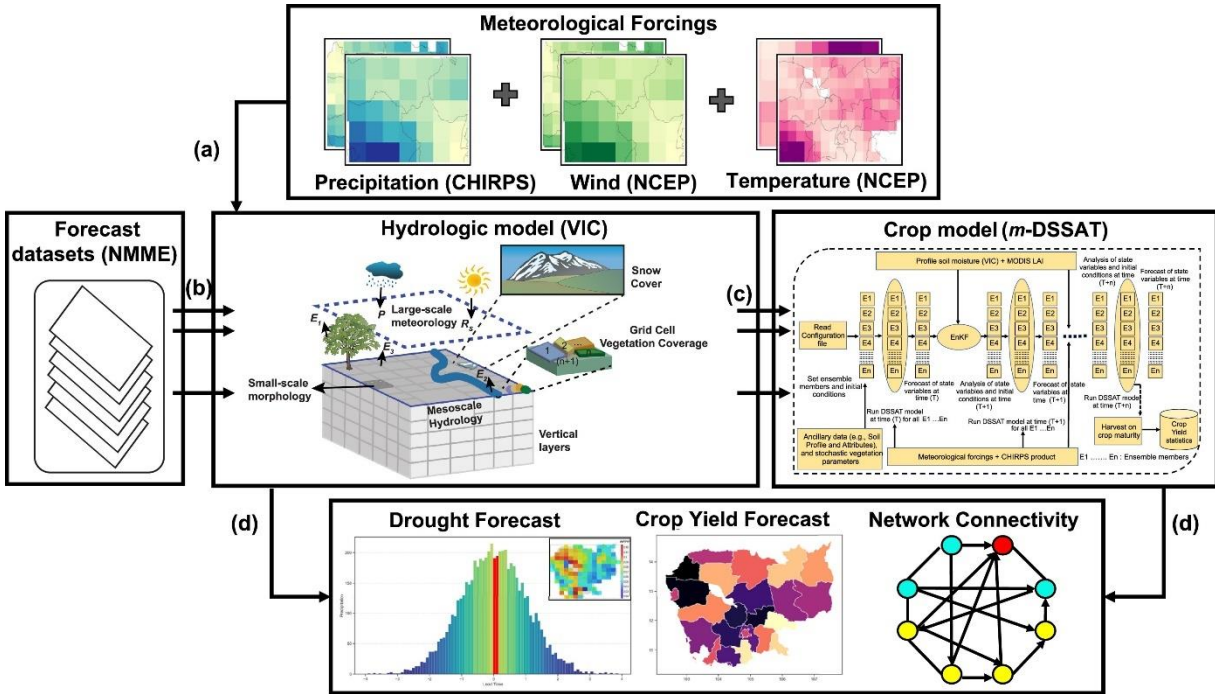


Figure 14. Simplified illustration of the RHEAS framework. RHEAS follows a modular hybrid approach of loosely coupling a hydrologic model with a crop model. a, Using a set of meteorological forcings, the VIC hydrologic model simulates the core hydrologic variables and drought states in a nowcast (nowcast refers to the present conditions) mode; b, An ensemble of NMME datasets is next ingested into the model for seasonal forecast operations; c, Using the VIC-simulated outputs, the m-DSSAT crop model simulates an ensemble of crop yields, both in nowcast and forecast mode; d, Outputs of the model, specifically hydrologic variables (e.g., soil moisture) and drought indices (e.g., severity) are then used to construct the network connectivity/links between different components of the model outputs.

At the core of RHEAS, the VIC hydrological model captures the complex hydrological processes that govern the movement of water by incorporating spatially distributed information on various hydrological variables such as precipitation (PREC), minimum and maximum surface air temperature (TMIN and TMAX), evapotranspiration (ET), runoff, etc. Specifically, it simulates the hydrological variables and drought states (indices) over a gridded domain at a spatial resolution

of 0.25° at a daily time-step in nowcast (i.e., estimation of current conditions) and forecast (i.e., estimation of future conditions) modes of operation. It incorporates the spatially distributed information on various hydrological variables to construct the key drought indices (or indicators), including the Standardized Precipitation Index (SPI), drought Severity (SEV), the Soil Moisture Deficit Index (SMDI), Dryspells, etc. The VIC hydrological model exhibits several advantages when integrated with other environmental models (e.g., a crop model in our case). Firstly, it solves the energy and water balance equations to provide a realistic estimation of water availability for crop growth and development. Secondly, it captures the dynamics of water movement through the plant-soil-atmosphere continuum, which is critical for understanding the impact of different hydrological processes on interannual crop yields. Furthermore, the VIC has been extensively tested and validated over different regions, ensuring its stability and reliability.

Although the VIC hydrologic model has traditionally been used for standalone real-time applications (Hamman et al., 2018), the modular framework of RHEAS makes it coherent for the VIC to interact with other crop growth models (e.g., DSSAT) via the PostgreSQL relational database system (Holzworth et al., 2015). Firstly, such a design allows the seamless flow of information and smooth handling of internal files/datasets between the constituent models (i.e., VIC and *m*-DSSAT), whilst yielding a composite hydrologic-agriculture data product. Secondly, the integration within RHEAS adds another layer of robustness to our approach, enabling us to make accurate predictions and assessments of crop yields under different climate and management scenarios.

Using the VIC-generated hydrologic variables (e.g., rainfall, solar radiation, surface air temperature, etc.), and other agronomic management information (cultivar genotype coefficients, fertilizers) as input, the *m*-DSSAT crop model simulates the growth, crop stage development, and

end-of-season yield over the growing season. Due to the extensive testing and calibration using empirical data from various regions and cropping systems, DSSAT offers several advantages over other crop models (Kadiyala et al., 2015; Subash and Ram Mohan, 2012). Its inherent ability to incorporate information on agronomic practices, such as fertilizer application rates, cultivar genotype coefficients, irrigation rates, etc., enables the assessment of crop behavior under different management scenarios. Additionally, DSSAT has been extensively validated over different regions, including our study area (Abhishek et al., 2021). However, unlike the traditional DSSAT, the m-DSSAT within RHEAS is a customized version that can run multiple ensembles member (~40 ensembles) for different soil properties and agronomic management practices (Ines et al., 2013).

Although DSSAT is widely used and recognized for its capability to simulate crop growth and yield, there are few notable limitations, such as the simplified soil water balance model employed within DSSAT, that may not accurately represent the complex hydrological processes. As a result, it can lead to potential discrepancies in simulating water availability and its impact on crop growth and yield. Therefore, to address this limitation, our study incorporates a hydrologic model within the RHEAS framework. By integrating the hydrologic model with the m-DSSAT crop model, we enhance the accuracy and robustness of our predictions. The hydrologic model captures the complex hydrological processes, including infiltration, evapotranspiration, and runoff, which are crucial factors influencing soil water availability and crop growth. Additionally, the VIC hydrological model simulates the land-atmosphere fluxes to capture the spatial and temporal dynamics of soil water availability over the growing season. Consequently, this improves the accuracy of crop yield predictions, as the hydrologic model accounts for the intricate

interactions between water availability, plant physiological responses, and agronomic management practices.

3.2.3. Data

The high-resolution CHIRPS precipitation product and the NCEP air temperature and wind speed gridded reanalysis product was used as meteorological/initial forcing into the hydrologic model. The CHIRPS product is a merged product that incorporates satellite-derived rainfall estimates with ground-based station data to produce a gridded precipitation dataset. Due to the availability of consistent (1981-present), high-resolution (0.05 degrees or ~5 km) blended products with extensive validation, it has been widely used for various applications, including agricultural monitoring, climate analysis, drought monitoring, and water resource management, especially in data-scarce regions (Katsanos et al., 2016; Toté et al., 2015). Likewise, the NCEP reanalysis product provides a comprehensive dataset of global atmospheric and oceanic observations. It utilizes a data assimilation approach to merge historical observations with model simulations, resulting in a consistent representation of the Earth's climate system over long periods. Thus, they have been widely used for climate research and weather analysis, contributing to our understanding of past climate conditions, and aiding in the development of climate-related studies and applications (Eini et al., 2019; Hagemann and Dümenil Gates, 2001; Maurer et al., 2010). Additionally, the land cover information in the region was derived from the Moderate Resolution Imaging Spectroradiometer (MODIS-500 m) global product (Friedl et al., 2010), whilst the gridded soil information (10 km) was based on the SoilGrids1km by Han et al. (2019). A list of all the datasets used within the RHEAS framework are summarized in Table 5.

Table 5. List of various data products incorporated into RHEAS with their spatial and temporal resolution.

Variable	Product	Spatial Resolution	Temporal Resolution	Period Availability
Precipitation	CHIRPS	5 km	Daily	1981 - present
Temperature	NCEP	1.875 deg	Daily	1948 - present
Wind Speed	NCEP	1.875 deg	Daily	1948 - present
Seasonal Climate Forecast	NMME (CCSM4 and CFSv2)	1 deg	Daily/ 6-hourly	2011 - 2020
Topography/ Global DEM	GTOPO30	~1 km	static	1996
Land cover	MODIS	500 m	1-2 days	2000 - present
Soil Profile	SoilGrid-250m 2.0	250 m	static	2017

Note: CHIRPS: Climate Hazards Group Infrared Precipitation with Station data, NCEP: National Centre for Environmental Prediction, NMME: North American Multi-Model Ensemble, CCSM4: Community Climate System Model 4.0, CFSV2: Climate Forecast System version 2, DEM: Digital Elevation Model, MODIS: Moderate resolution Imaging Spectroradiometer

For the seasonal climate forecasts, we used the NMME seasonal prediction system that combines multiple climate models in an ensemble-based approach to provide skillful predictions of climate variables on seasonal to interannual timescales. The NMME dataset offers a comprehensive set of predictions, enabling the assessment of uncertainty and the exploration of future climate scenarios at a global scale. Examples of commonly predicted climate variables include surface air temperature, the net solar flux at the surface, specific humidity, precipitation, etc. All these variables are extensively evaluated and validated with observations to assess their skill and reliability. Thus, they have illustrated the potential in predicting the key climate phenomena such as El Niño-Southern Oscillation (ENSO), climate risk assessment on water resources and agriculture, etc. (Thober et al., 2015; Wang et al., 2022). Although the NMME data

incorporates several state-of-the-art climate models, we used the climate forecasts from the Climate Forecast System version 2 (CFSv2) and the Community Climate System Model 4.0 (CCSM4) coupled modeling systems. Each model consists of multiple ensembles of climate prediction, with each ensemble based on different initial conditions, model physics, and data assimilation techniques. In practice, the multi-model ensemble approach has demonstrated exceptional skill/efficacy in quantifying prediction uncertainty and consistently outperforming individual model ensembles, thus making it an ideal choice for this study (Becker and van den Dool, 2016; Tippett et al., 2019). Specifically, we used the daily records of climate forecast data, particularly precipitation, and temperature, from both models in this study (discussed further in methods).

For validating the simulated crop yields, the actual yields (i.e., observations) were collected by the Department of Agricultural Land Resources Management, General Directorate of Agriculture, Phnom Penh, Cambodia. They provide food and agriculture information (including crop yield, fertilizer rates, etc.) over selected provinces in Cambodia and nearby areas by establishing mechanisms to collect data through field visits, surveys, or reports submitted by farmers. These data are aggregated at the provincial or national level, often involving quality control measures, including data validation, cross-checking with other sources, and addressing any inconsistencies or errors in the collected data. Other relevant information, such as the local agronomic practices (e.g., cultivar varieties, fertilizer rates, etc.), were obtained from published reports, local stakeholders, or agricultural databases (e.g., the Food and Agriculture Organization, FAO).

3.2.4. Methodology

Although RHEAS has been successfully implemented over large areas (such as the LMB) for assessing the spatial-temporal behavior of end-of-season rice yields and associated drought conditions, in this study, the modeling domain was configured over Cambodia (at a spatial resolution of 25 km). As is well-known, Cambodia has been at the forefront of frequent extreme events, such as droughts, that have adversely affected the agriculture sector, and the rural economy. Hence, in this study, our intent was to be region/country-specific, and we scaled down our analysis to provincial levels (often called administrative units or provinces). There are multiple reasons for doing so: i) a number of previous studies have been carried out at the country level while there are hardly any district-specific studies, ii) country-level crop yields are very subjective and do not reflect the actual production/yield for major cropping districts, iii) evaluation of hydrologic/drought conditions at a district-level can reveal the outcome/performance of the final yields. Herein, two major rainfed rice-growing provinces were chosen- Takeo and Prey Veng, based on the information from published reports and government agencies.

As RHEAS has been designed to provide tightly constrained drought and crop yield information in data-scarce regions, the key objective of the study focused on maximizing the use of available scarce observations. Hence, the modeling setup in our study was based on the global calibration of the VIC model by Sheffield and Wood, (2007) and Zhang et al., (2018). Using this calibrated version of VIC, we computed the different hydrologic variables. In addition, the VIC hydrologic model was sufficiently validated against observations. As shown by Abhishek et al. (2021), the VIC-generated surface soil moisture was validated against the SMAP surface soil moisture (L3_SM_P) data, with the simulated surface soil moisture exhibiting a good agreement (correlation) with SMAP observations (Colliander et al., 2017). As soil moisture from SMAP is

regarded as an independent product, validating the model against the observations provides confidence in model stability and performance. Regarding the crop model calibration, the *m*-DSSAT was calibrated by adjusting the crop cultivar genotype coefficients (refer Table 6.). Specifically, the phenological and growth coefficients were optimized based on an initial set of cultivar genotype coefficients, and yield records for the last 10-12 years. The Shuffled Complex Evolution (SCE) algorithm was implemented to obtain an optimal version of cultivar coefficients by running the crop model multiple times (~350 iterations) (Rahnamay Naeini et al., 2018). As a result, the calibrated crop parameters capture the dynamics of crop growth and development as influenced by the hydrologic parameters and other variables within the system.

Table 6. Estimates of genotype coefficients of the rice varieties used in Takeo and Prey Veng provinces of Cambodia. The *m*-DSSAT crop model was calibrated using the Shuffled Complex Evolution (SCE) algorithm for getting an optimized set of phenological and growth parameters.

Phenological coefficients					Growth Coefficients			
Province	P1	P2R	P5	P2O	G1	G2	G3	G4
Takeo (1)	532.501	122.161	440.195	13.823	68.7939	0.027	1.00	1.20
Prey Veng (2)	554.400	87.7	251.100	13.0	68.67	0.021	1.00	1.15
Phenology genetic coefficients								
<p>P1: Time period in °C (above a base temperature of 9 °C) from seedling emergence to the end of juvenile phase. Expressed as growing degree days (GDD)</p> <p>P2R: Extent to which phase development leading to panicle initiation is delayed for each hour increase in photoperiod above P2O. Expressed as growing degree days (GDD)</p> <p>P5: Time period from beginning of grain filling to physiological maturity with a base temperature of 9 °C. Unit: GDD</p> <p>P2O: Critical photoperiod or the longest day length at which development occurs at a maximum rate. Expressed in hours</p>								
Growth genetic coefficients								
<p>G1: Potential spikelet number coefficients as estimated from the number of spikelets per g of main culm dry weight (less leaf blades and sheaths plus spikes) at anthesis. Unit: Spikelets per g of main culm</p> <p>G2: Single grain weight under ideal growing conditions, i.e., non-limiting light, water, nutrients, and absence of pests and diseases. Expressed in grams</p> <p>G3: Tillering coefficient (scalar value) relative to IR64 cultivar under nonlimiting conditions</p> <p>G4: Temperature tolerance coefficient</p>								

At first, the VIC hydrologic model executes nowcast simulations from 1981 through 2011 at a daily time step, computing a large range of hydrologic variables and associated drought indices. These drought indicators are constructed based on the historic climatology (1981-2010) of different hydrologic data streams (e.g., precipitation, soil moisture, etc.). The nowcast simulations are a representation of the current hydrologic states in the region and serve as a base for future prediction. However, the objective was to examine the efficacy of seasonal yield and drought forecasts for the major growing season (generally the wet season, Jun-Nov), a comprehensive method of forecasting the yield and hydrologic indicators from 2012 through 2020 was devised. For convenience in demonstrating the results, a three-year scenario was selected, each representing a wet (2013), moderate (mod) (2019), and dry (2015) year. As the study focuses on individual administrative units, a comparison of the yield and associated drought conditions for the abovementioned years can be critical in understanding the logic behind the approach.

To test our hypothesis that seasonal climate forecasts improve crop yield predictions over time, we used the multi-model (CFSv2: Climate Forecast System version 2; CCSM4: Community Climate System Model 4.0) ensemble climate forecasts from the NMME seasonal forecasting system as the historic forecast datasets. Each monthly NMME dataset (e.g., Jun) comprises 6-7 month's forecast of precipitation and temperature variables. These seasonal forecasts are expressed as probabilities of occurrence below, near, and above normal relative to a long-term climatology of total precipitation and mean temperature. Furthermore, these multi-model ensemble datasets were converted to plausible daily weather sequences (e.g., precipitation, etc.) and ingested as input to the VIC hydrologic model. Using these datasets, the VIC runs forecast simulations for 10 ensemble members to predict the future (Jul-Nov) drought and hydrologic conditions. Following the hydrologic simulation, the m-DSSAT crop model subsequently runs a series of (40-50)

ensembles to predict the end-of-year (Oct/Nov) crop yield. However, prior to the run, a series of ancillary information, such as cultivar genotype coefficients and agronomic management information (e.g., fertilizer rates), are provided to the crop model. Especially, the type of cultivars, and the amount of fertilization are two key agronomic parameters, that are critical for enhancing (either increasing or decreasing) the crop model performance. Generally, crop models have shown better performance when optimal cultivars and balanced nutrient (fertilizer) application rate information were available. In Cambodia, most of the agricultural setups are rainfed (i.e., mostly paddy rice), thus no irrigation was considered in the study. Although some forms of irrigation activities have recently been introduced around the Tonle Sap Lake, herein, we focused on the provinces that have no irrigation.

After the forecast run of drought and crop yield at the end of June, the hydrologic model simulates the present condition (nowcast) for the subsequent month, i.e., July. A similar approach is then followed in the next run, wherein the NMME forecast data of July (i.e., from August to February) is ingested into the RHEAS, and the drought and yield forecast is generated for Aug-Dec. It should be noted that the hydrologic model provides forecast beyond the harvest month (i.e., Nov), however, the crop model is simulated specifically for the long (wet) growing season (i.e., Jun-Nov) as rice is generally harvested in Oct/Nov. Likewise, the process is replicated for each monthly (Jun-Nov) NMME forecast datasets until we reach the end of harvest, i.e., Nov. It is noteworthy that during end of each incremental monthly forecasting of drought and crop yields, the number of months (days) of real forcing data increases. Ultimately, the drought and crop yields simulated for the month of November is done entirely through the real forcing data. A detailed step-by-step flow of the methodology employed is provided in Figure 15. The general idea behind such an approach is to see the dependence on reducing climate forecasts inputs (i.e., forcings) as

we move forward in time during the growing season. For instance, the yield from the Jun NMME forecasts is completely dependent on the historic projection datasets. However, in the Jul NMME forecast run, we see the crop model avails the Jun nowcast and the rest of the NMME forecast to provide the end-of-year yield. Similarly, when we reach Nov, the model is the least dependent on the historic forecast datasets and more on the actual conditions (nowcast).

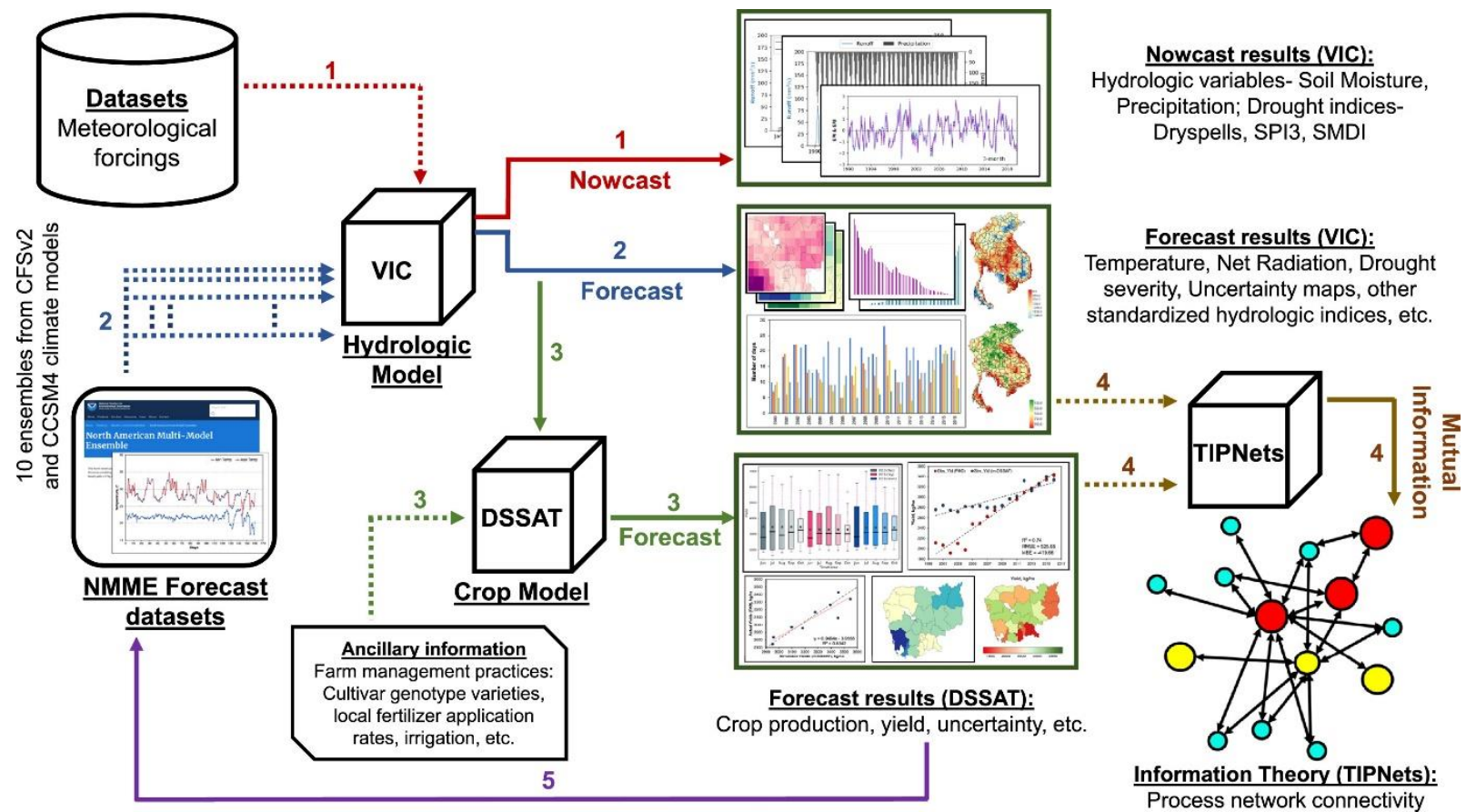


Figure 15. Detailed representation of the methodology employed within RHEAS for the study. The color (and number) of the arrows represents the steps incorporated in the study chronologically. At the core, the Variable Infiltration Capacity (VIC) hydrologic model loosely couples with the Decision Support System for Agrotechnology Transfer (DSSAT) crop model to provide information on interannual crop yields and associated drought states at current (nowcast) and future (forecast) conditions. For prediction of future

Figure 15. (cont'd)

states, the forcing datasets consist of multiple coupled models (CFSv2: Climate Forecast System version 2; CCSM4: Community Climate System Model 4) from the North American Multi-Model Ensemble (NMME) forecasting system. Steps 1-4 is carried out for each timeframe (Jun-Oct) over the growing season. Finally, the Temporal Information Partitioning Network (TIPNets) reflects the significant information (or link) shared between the associated variables.

3.2.5. Mutual Information

Our aim was to examine the relationships between the parameters of the two coupled dynamic systems, i.e., the hydrologic and crop models, through the information that was exchanged between both systems. As many of these relationships were likely nonlinear, the statistics of IT provide an approach to capture this nonlinearity without requiring any prior assumptions about the variables. Having a general quantification of the connections between different components of the hydrologic (e.g., soil moisture) and crop (e.g., phenology) modeling systems can help identify the inherent parameters/processes affecting interannual crop yields. Specifically, Mutual Information (MI) measures the amount of shared information or synchronization between two variables and is well-equipped to analyze the dynamics of the state variables of a nonlinear coupled system. Figure 16. illustrates a simplified representation of the various aspects of MI. Here, we treated the hydrologic and drought parameters (e.g., precipitation, dryspells, etc.) as source variables, while the target variable was end-of-year yield. Before MI is computed, we quantify the uncertainties contained in a variable (say ' X ') using the Shannon entropy (Shannon, 1948):

To measure MI for two variables, X and Y , the uncertainty in a variable ($H(X)$) were computed by:

$$H(X) = - \sum_{i=1}^N p(x_i) \log_2(p(x_i)) \quad (1)$$

where $H(X)$ (measured in bits) value measures the amount of uncertainty contained in X and is based on its probability density function $p(x_i)$ (PDF). Now, using the idea of Shannon entropy, we compute the MI between two variables X (target variable) and Y (source variable). For simple understanding, we can understand this easily by considering X as ‘crop yield’ and Y as ‘hydrologic variables or drought parameters’. Thus, the information shared between X and Y can be measured using Eq. 2.

$$MI(X; Y) = H(X) + H(Y) - H(X, Y) = H(X) - H(X|Y) \quad (2)$$

where, MI essentially measures the reduction in uncertainty of X (the target variable) by having this new information about Y (the source variable) and is based on the joint probability of the two variables, and $H(X, Y)$ is the joint entropy. MI can be measured for two variables over a time lag or can be measured without a time lag (often referred as zero-lag MI), provided by:

$$MI(X; Y) = \sum p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (3)$$

where, $p(x, y)$ is the joint probability of lagged X and Y variable. When, we assess the contribution of uncertainty from multiple (source) variables (e.g., precipitation, temperature, etc.) to a single target variable (e.g., crop yield), generally Normalized mutual information (NMI) is preferred, given by:

$$NMI(X, Y) = \frac{MI(X; Y)}{H(X, Y)} \quad (4)$$

where, $NMI(X, Y)$ is the NMI between X and Y . By dividing the mutual information between the source and target variables by the entropy of the target variable, NMI provides a normalized measure that ranges between 0 and 1, wherein a value of 0 indicates no shared information or

explanatory power, while a value of 1 indicates that the source variable completely explains the uncertainty in the target variable.

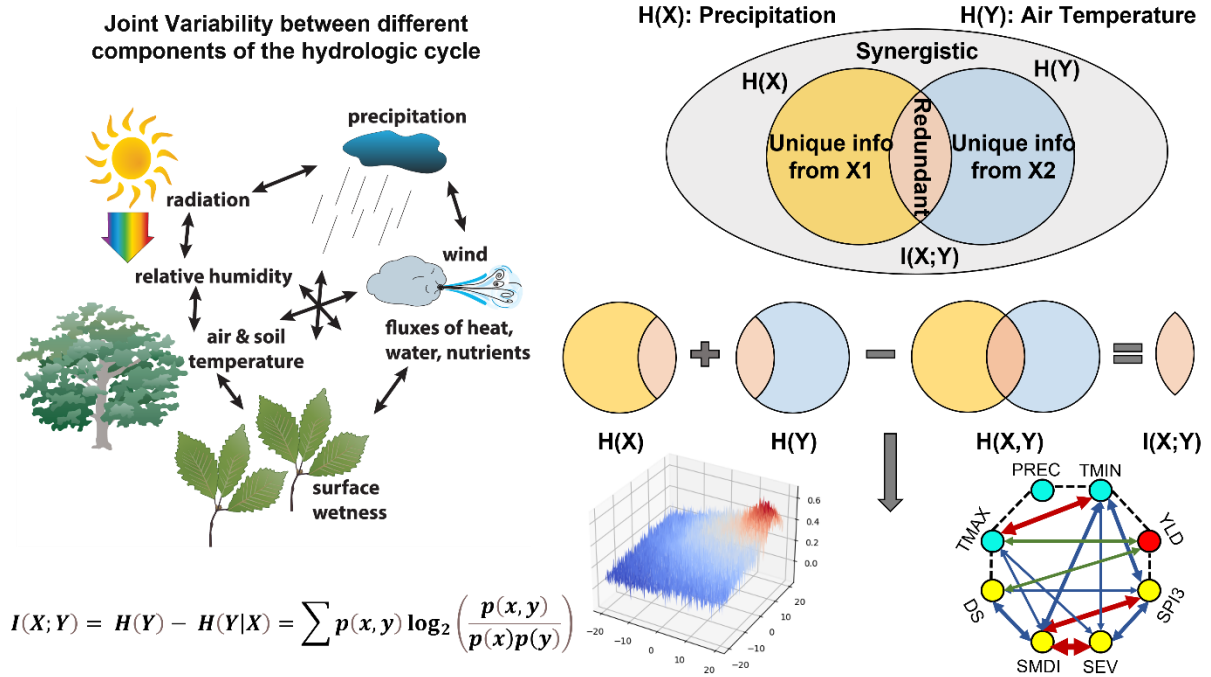


Figure 16. Simplified illustration of the concept of Mutual Information. It shows an integrative depiction of the hydrologic cycle's intricate connections with agricultural systems, emphasizing the mutual information shared between the hydrologic and agricultural parameters in the soil-vegetation-atmosphere continuum.

Here, we used a Matlab-based interface- the Temporal Information Partitioning Network (TIPNets: Goodwell and Kumar, 2017) system to quantify the MI between all hydrologic variables and yield. The software first takes the hydrologic time series and yield ensembles as inputs and creates the PDFs of the data using either the fixed interval or kernel density estimation (KDE) binning scheme. KDE is generally employed when there are few data points, but for this work, we had sufficient data to use the fixed interval binning scheme. The joint PDFs of the principal hydrologic parameters, drought indicators, and yield were determined using the fixed interval binning scheme with 10 bins used for computing the PDF (Ruddell and Kumar, 2009a, 2009b).

The TIPNets compute the Shannon entropy ($H(X)$) for each variable and MI for all possible variable pairs. Here, we only focused on the zero-lag MI for the relationships with yield, as it was not a time series variable. Statistical significance among the time series variables was found using a shuffled surrogate method where all time series were randomly shuffled, and the MI statistics were recomputed using the random series. If the original MI values surpass the random MI values following a t-test at a 95% confidence threshold, then the relationships were considered statistically significant. We performed this analysis for the two study areas over the three years, and the five time periods in each year. The MI values were then normalized by the Shannon entropy of the yield variable to see the portion of yield uncertainty explained by each variable, termed as NMI.

3.3. Results and Discussion

3.3.1. Efficacy of seasonal climate forecast on crop yield

The primary aim of the study was to understand the evolution of uncertainties of climate forecasts on seasonal crop yields. Based on our working hypothesis, we conducted a monthly forecast of interannual yields based on a 1-month moving window. Herein, we implemented the approach on two major rainfed rice-grown provinces in Cambodia- i) Takeo, and ii) Prey Veng for three distinct years- 2013 (wet), 2015 (dry) and 2019 (mod). These administrative units host large areas of rice production (part of the Mekong Delta) (MRC, 2014) and have been historically impacted by drought/climate extremities, which makes them ideal case studies. To characterize the uncertainties in climate forecast, we used a 40-ensemble yield product from the *m*-DSSAT crop model for each forecast cycle. Each cycle (represented as I, II, III, IV, V) refers to the nature of input data (Figure 17.), wherein the first cycle (i.e., Jun) of a year comprises 1-month of real meteorologic conditions ('nowcast' that is simulated by the model), while the rest of the input data

are based on the ingested NMME seasonal forecast. Similarly, with each subsequent cycle, we have a higher degree of present conditions (nowcast) and lesser reliance on forecasts. Figure 17a. and 17b. present the end-of-season yield over Takeo and Prey Veng during the study period. By dividing the analysis into five-time intervals, we aimed to capture the temporal dynamics of yield prediction and evaluating their accuracy throughout the growing season. Here, the boxplot with a distinct color for each year showcases the extent of simulated yield values for each cycle over a cropping season (i.e., 5 months). For instance, in Takeo, most yield values for Jun 2013 ranged between ~2100 kilogram per hectare (kg/ha) and ~4800 kg/ha (barring the outliers), whereas as we progress in time, the yield values substantially decrease, ranging between ~2500 kg/ha and ~4000 kg/ha (a reduction of ~1200 kg/ha) respectively (Figure 17a.). In line with the observed trend, similar patterns were also observed in other years for both provinces. Specifically, a reduction in yield was recorded, amounting to approximately 400 kg/ha (Prey Veng, 2019) (Figure 17b.) and 1300 kg/ha (Takeo, 2019) (Figure 17a.). These yield reductions indicate a notable decline in crop productivity during those respective years. Overall, we see a clear indication of reducing uncertainties in yield over the cropping season from planting (i.e., ‘I’ timeframe or Jun) through harvest (i.e., ‘V’ timeframe or Oct). In addition, the average yields from the crop model were compared against an independent source of yield records (obtained from Department of Agricultural Land Resources Management, General Directorate of Agriculture, Phnom Penh, Cambodia). In Figure 17a. and 17b., we see the median (simulated) yields (represented by the black line) in accordance with observations (represented by a star mark) over both provinces. When comparing ensembles of simulated yield against observations, the median offers a favorable alternative to the mean due to its resilience to extreme values or outliers. The mean can be significantly distorted by a few outlying values within the ensemble of simulated yields. In

contrast, the median, which represents the middle value when the data is arranged in ascending order, diminishes the impact of extreme values, thus yielding a more robust measure of central tendency. Additionally, since observations typically consist of single values, while an ensemble represents multiple values, focusing on the median aligns the comparison by emphasizing the central tendency of the ensemble rather than its mean, thus ensuring a more appropriate and accurate assessment of the model's performance against observations. At the beginning (i.e., 'I' timeframe in 2013 or Jun, 2013), when we have less information on the meteorological conditions, the end-of-season yield (~2930 kg/ha) in Takeo does not reproduce the observations (~3500 kg/ha) (Figure 17a.). However, with subsequent inclusion of actual meteorologic conditions into the model at later timeframes (e.g., IV (Sep) and V (Oct) timeframes), there is an improvement in the yield prediction. Notably, in Takeo, we see the model underestimating yield when compared against observations during I-III timeframe (i.e., Jun-Jul-Aug for all three years, but inches closer towards the actual yields during harvest (V timeframe or Oct). Additionally, the model tends to capture the uncertainty and observed values during the dry ($R^2=0.8$) and moderate ($R^2=0.9$) years (as compared to the wet year) (Figure 17c.). Similarly, the model performed considerably better over Prey Veng (than Takeo) in capturing the yield uncertainties whilst accurately predicting/matching the observed yields. While there were noteworthy instances of simulated yield aligning with observations across various timeframes (such as June, July, and October 2015), the model demonstrated a progressive enhancement in its ability to match yield observations in Prey Veng during 2019, achieving an R^2 value of 0.9. Figure 17c. and 17d. shows the correlation between simulated and observed yields over different timeframes for both provinces. It should be noted that the average simulated yield of Takeo exhibits a substantial bias (~300 kg/ha in 2013 and 2015 to ~430 kg/ha in 2019,) compared to the observation, especially during the initial timeframes

of the year. This large bias can be attributed to the mismatch/absence of – i) rice cultivar genetic coefficients, and ii) appropriate data on local management practices (e.g., fertilizer rates, planting date, etc.). Although sufficient calibration (of the crop model) was carried out for each province the lack of information on explicit rice cultivars and management practices during the entire period (3 years) could be a major source of uncertainty (see Table 6. for calibrated values of cultivar genotype coefficients). Additionally, Figure 4., Appendix A illustrates the relative yield anomaly during different timeframes within a season over both provinces. Hence, having appropriate knowledge about such key information (e.g., agronomic management information, daily weather information, etc.) can be deemed imperative for quantifying yield forecasts. Not only do these results support our hypothesis, but they also provide a substantial ground for using such a unique approach for improving crop yield forecasting.

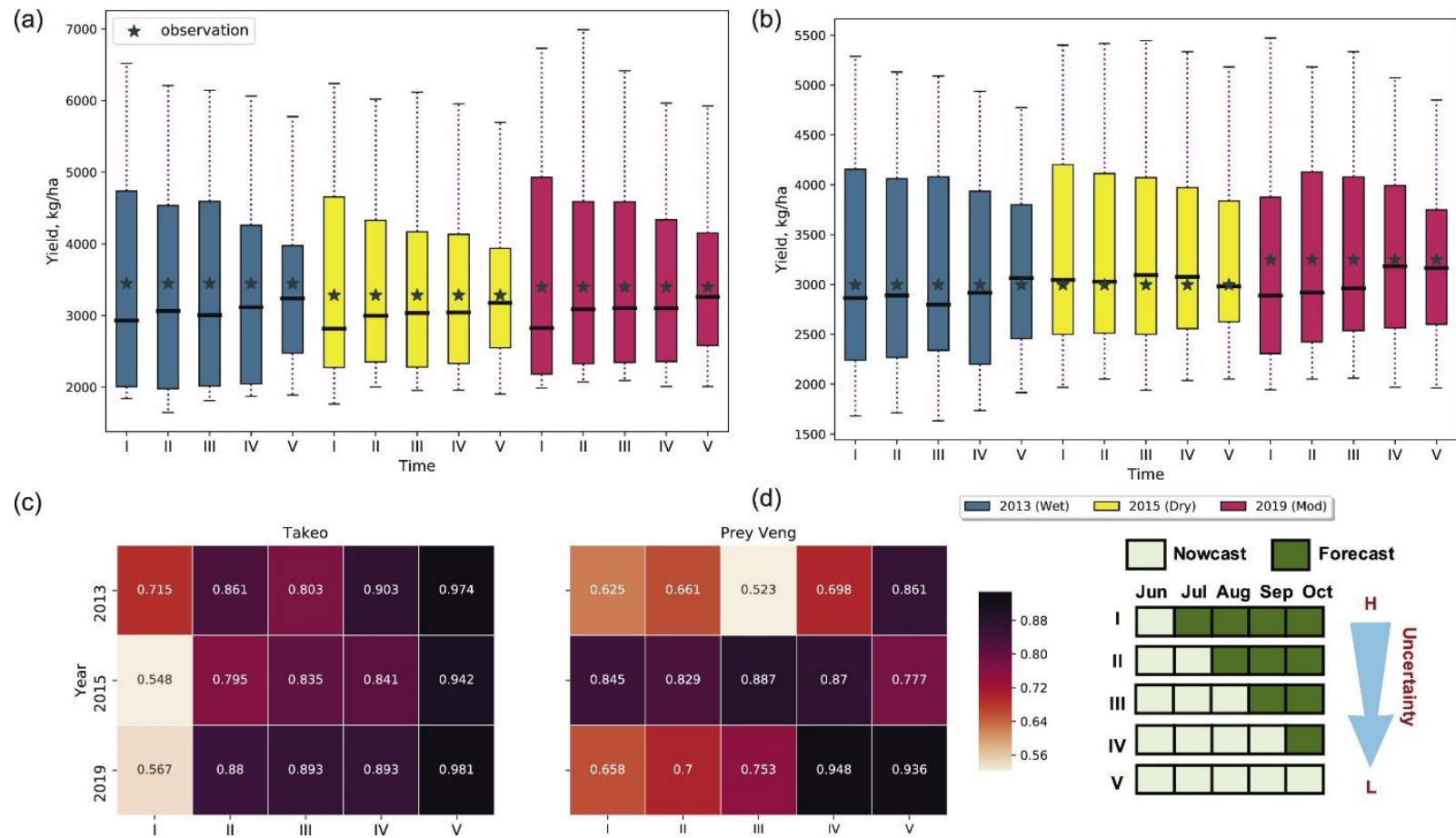


Figure 17. Uncertainties in yield over the growing season from *m*-DSSAT crop model. a) Boxplot of simulated rice yields from 40 ensembles of the crop model over Takeo during three (wet, dry, mod) distinct years. Here, a season consists of 5 timeframes with a 1-month moving window. Each timeframe represents the end-of-year yield scenarios based on the weightage of nowcast/forecast input forcings. The I, II, III, IV, V timeframes essentially represent the yield values wherein the input forcings comprised of 1-month

Figure 17. (cont'd)

nowcast and 4-month of forecasts. With each subsequent timeframe, there is an increase in nowcast forcings and decreasing reliance on forecasts. The black line on the boxplots represents the median yield over that particular timeframe, while the grey stars represent the observations.

b) Boxplot of simulated yields over Prey Veng during 2013, 2015 and 2019. c-d) Skill of crop yield predictions over the growing season: correlation coefficient between simulated and observed yields over different timeframes in Takeo and Prey Veng.

3.3.2. Influence of geophysical and drought variables on crop yield

To address the uncertainty related to weather, the impact of various hydrologic (and drought) variables were analyzed to examine the contribution of each variable toward the end-of-season yields. We computed the NMI among all the hydrologic and drought variables with simulated rice yields to quantify the synchronous dependence of one variable on another. Normalization of MI is commonly performed to scale its values between 0 and 1, making it more interpretable and comparable across different analyses. Various methods can be employed for this normalization, such as dividing the mutual information by the square root of the product of the entropies of the variables involved. This normalization guarantees that MI falls within the range of [0, 1], irrespective of the variables' specific ranges or units. By normalizing MI, it becomes possible to compare the relationship strengths between variables across diverse time frames or datasets. It is important to note that the range of a variable itself does not directly impact the calculation of normalized mutual information, as the focus lies on the probabilistic distributions of the variables rather than their ranges. However, if a variable has a limited or constrained range, it may affect its overall variability, potentially influencing its contribution to the mutual information calculation. Here, we focused on crop yield as the target variable to identify the active/predominant

parameters impacting yield through the growing season as the forecasts are updated. Figure 18. shows the NMI between the hydrologic variables and drought indices on crop yields over different timeframes in a year (i.e., how much information is mutually shared between the hydrologic (and drought) variables and crop yield), where the stars highlight the variable with the greatest synchronization with yield. In Takeo, we observed a significant flow of information between the independent variables and crop yield during different timeframes in a season (Figure 18a.). Each timeframe exhibited a varied degree of information flow through a given season, with TMIN and DS exhibiting a consistently higher synchronization with yield over the entire study period, i.e., TMIN had a strong connection with crop yield. During the wet year (2013), TMIN (~ 0.24 bits/bits, bi/bi), DS (~ 0.24 bi/bi), and the SMDI (SMDI: (Narasimhan and Srinivasan, 2005) (~ 0.12 bi/bi) showcased higher synchronization, whereas the net Short Wave radiation (SW) (~ 0.08 bi/bi), and the 3- and 6-month Standardized Precipitation Index (i.e., the SPI3, and SPI6: (McKee et al., 1993) (~ 0.08 bi/bi) arrayed lower levels of MI. SMDI is a standardized drought index that is based on the surface soil moisture and root zone soil moisture, often used to characterize agricultural drought. It considers the soil's water-holding capacity and calculates the moisture deficit by comparing the current soil moisture level with the long-term average. A positive value (e.g., +1.0, +1.5, +2.0) represents wetter conditions (higher than normal long-term average) whereas negative values (e.g., -1.0, -1.5, -2.0) represents drier soil moisture (lower than normal long-term average). Likewise, the 3- and 6-month SPI provides information on the precipitation deficit or surplus over a three-month period respectively. These indices are constructed by accounting the long-term average precipitation for the respective timeframes, enabling the identification of abnormal conditions. The SPI values are an indication of the dryness/wetness in a region, with positive SPI values (i.e., +0.5, +1.0, +1.5, etc.) indicating wetter conditions (i.e., above-average precipitation),

and negative SPI values (i.e., -0.5, -1.0, -1.5, etc.) indicating drier conditions (i.e., below-average precipitation). Having knowledge about these parameters offers valuable insights into drought conditions, helping stakeholders respond effectively. Additionally, we saw an increase in the information flow from the start of the cropping season (Jun) until harvest (Oct) during the wet year. This essentially means that the strength between the nodes, each representing a geophysical parameter (e.g., DS, SW, etc.), increased over time. In 2013, there was an increase of ~47% (DS) and ~128% (TMIN), while a few parameters saw a decrease of ~18% (PREC) and 24% (SW) during the same period. A similar pattern was observed during the dry year (2015), wherein the information flow was higher during the later timeframes (i.e., Sep, and Oct). These percentages represent the relative increase or decrease in the values of DS and SW from the beginning (June) to the end (October) of the season. The fluctuations observed throughout the year, with variables increasing and then decreasing, are captured by these percentage changes, which highlights the seasonal dynamics and variations in DS and SW. Although the mod year (2019) depicted irregular behavior, it showcased a higher synchronization than the dry year. Overall, we observed yield exhibiting a higher synchronization with the hydrologic variables (10 out of 15 months) compared to the drought indicators with TMIN as the strongest link. Also, the drought variables displayed strong links during the earlier months (Jun, Jul) of the wet and dry years, while the latter months (Sep, Oct) array a strong synchronization with the hydrologic variables. During the initial months of both wet and dry years, we found strong connections between drought indices (e.g., Standardized Precipitation Index and Standardized Soil Moisture Deficit Index) and crop yield. This indicates that drought conditions significantly influenced crop performance in the early stages of the growing season. These findings were consistent with previous research (Hussain et al., 2016) emphasizing the importance of early-season drought on subsequent crop outcomes. However, as

we approached the harvest period, we noticed a shift in the significance of variables. Hydrologic factors like precipitation and soil moisture exhibited stronger associations with crop yields compared to the drought parameters. This suggests that hydrologic variables played a more influential role in determining final crop yields as the growing season progressed. This shift can be attributed to the direct impact of hydrologic variables on water availability and nutrient supply during critical growth and development stages. The increasing synchronization between crop yield and hydrologic variables in the later months of the growing season can be explained by crop physiology. As crops mature and approach the harvest stage, their water requirements and sensitivity to soil moisture conditions become more pronounced. Consequently, hydrologic variables that directly influence soil moisture and water availability play a critical role in determining crop yields during this crucial period. Likewise, we saw a similar pattern of information flow in Prey Veng wherein DS, TMIN, TMAX, and PREC had a substantial impact on yield (Figure 18b.). Compared to Takeo, the amount of information mutually exchanged between the models was relatively stable throughout the months, which could potentially be due to the better performance of the crop model over Prey Veng. In contrast to Takeo which arrayed strong predominant links with hydrologic variables (10 out of 15 months), the NMI for Prey Veng showcased strong linkages with 8 hydrologic and 7 drought variables over the study period, with DS dominating the information sharing with yield compared to other variables (6 out of 15 months). It is also noteworthy that a dominant link between two nodes at an initial timeframe may not necessarily have a significant synchronization at a later timeframe. For instance, PREC exhibited the strongest synchronization with crop yields at the start of the year in 2019, but the strength between the links gradually decreased (or became insignificant) during the later timeframe (i.e., IV and V timeframe in 2019). Instead, DS maintained a consistent link throughout the season

and had the highest information flow during the end-of-season timeframe, thus making it the most suitable variable that considerably influences crop yield.

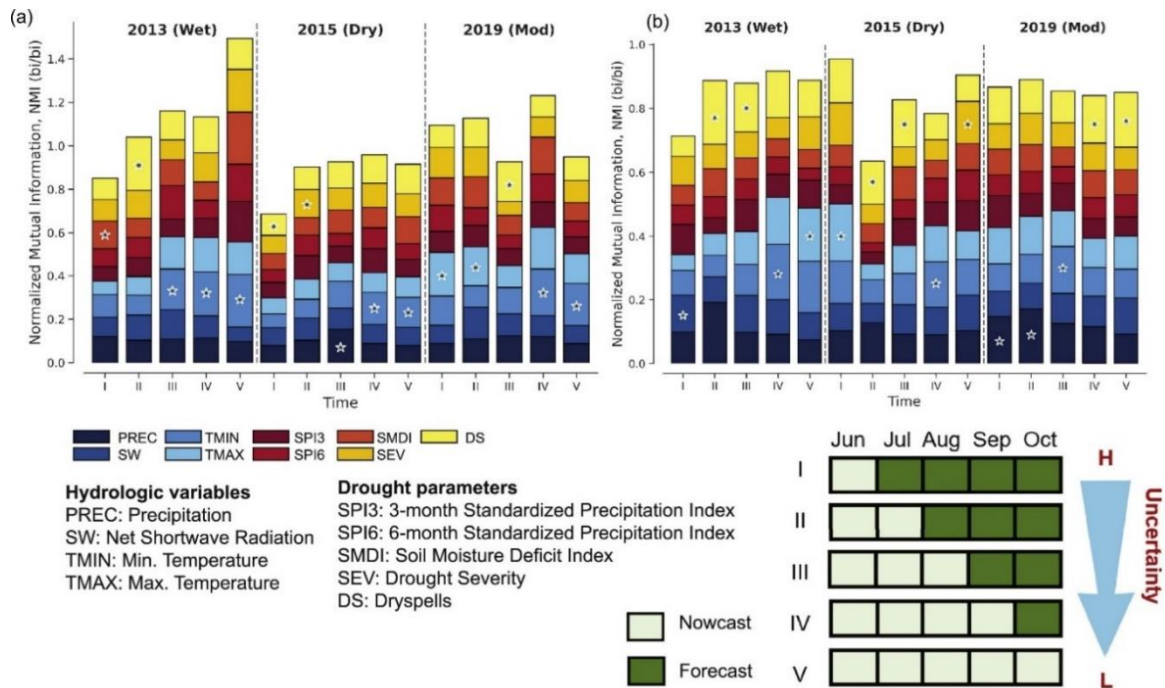


Figure 18. Stacked bar plot showing the NMI between different hydrologic variables (PREC, TMIN, TMAX, SW) and crop (rice) yield, and NMI between drought indicators (SPI3, SMDI, SPI6, SEV, DS) and crop yield: a) Mutual exchange of information over Takeo; b) Mutual exchange of information over Prey Veng. The stacked bars represent the contribution of the independent variables and crop yield at different timeframes. Shades of blue color (navy to sky blue) represents different hydrologic variables while the shades of red color (red to yellow) represent the different drought indicators. The star (in each stacked bar) represents the variable/indicator having the maximum influence on crop yield during that time window.

3.3.3. Process Network Connectivity

In our study, we examined the various network connections between the hydrologic variables, drought indicators, and end-of-season yield in Takeo and Prey Veng region. To achieve this, we constructed a network plot that provides a comprehensive visualization of the relationships

between hydrologic variables and crop yield during different time frames (see Figure 19. and Figure 20.). We visualized these connections using process networks by analyzing the strength, and direction of the relationships among these variables throughout the growing season. Typically, it involves creating a network plot, where each variable is represented as a node (or point), with each node encompassing a particular variable under consideration (e.g., PREC, TMAX, TMIN, DS, etc.) (Figure 19. and Figure 20.). These nodes typically represent the timeseries (based on the PDF) of a particular variable. These nodes are connected via arrows that illustrates the connection between different variables. These connections between different variables are established based on the NMI values calculated by the TIPNets model. In simple terms, the arrows show the statistical dependence between two variables, considering their joint distribution and individual distributions. Additionally, the intensity and thickness of the arrows indicates the strength of the connectivity between variables and visualizing the dependencies within the system. For e.g., a thinner arrow indicates a weaker connection, signifying a low NMI value. Conversely, a thicker arrow represents a stronger association, indicating a high NMI value. The varying thickness of the arrows are a visual depiction of the various relationships between different variables/parameters, aiding in the identification of dominant interactions. By examining the network plots as a whole, we identified the influential nodes (or variables) that act as key drivers within the system during different time period.

To provide more clarity on the strength of the connections, we categorized the thickness and color of the arrows based on the NMI values. Specifically, the red and blue color arrows were employed to identify distinct relationships between the hydrologic variables and drought parameters. On the other hand, the green color arrows were used to depict the connections between hydrologic variables and crop yield, as well as drought parameters and crop yield. Similarly, the

thickness of the arrows represented the strength and intensity of the connections between two variables based on the NMI values. A thicker red arrow (for e.g., SPI3 and SEV in Jun, 2013) represents a strong connection, indicating NMI values greater than 0.55 bi/bi. Similarly, a relevantly thin red arrow depicts strong connection with NMI values ranging between 0.4-0.54 bi/bi. Likewise, the blue arrows represent the moderate connections that have NMI values ranging between 0.3-0.4 bi/bi, while thin blue arrows show the weak connections with NMI values ranging between 0.1 to 0.3 bi/bi. Classifying the connections based on the NMI values allows for a better understanding of the varying strength of the connections between the variables. Regarding the black dashed line in the network plots, they were merely used to ensure the visual completeness and aesthetics of the network plot, and do not carry any specific meaning or significance.

Due to limited space within the network plot, we prioritized variables based on their importance, relevance, and existing research findings. Hence, we illustrated the process connectivity between selected variables—SPI3, SMDI, SEV, DS, TMAX, TMIN, PREC and crop yield (represented as YLD), thus leading to the exclusion of other variables such as SPI6 and SW from the network plot. During the initial months of the study years (specifically, Jun and Jul), we observed significant network connections between variables, particularly in the wet year in Takeo (see Figure 19.). In Jun, there was a strong association between SEV and SMDI (with NMI of approximately 0.55 bi/bi), while in Jul, a strong connection was observed between SPI3 and SPI6 (with NMI of approximately 0.6-0.7 bi/bi). Conversely, the links between DS and TMAX (NMI of ~0.08 bi/bi), and PREC and SEV (NMI of ~0.06 bi/bi) were weak or insignificant. However, as we progressed further in time (e.g., Sep, Oct), the connections between the variables became more explicit and prominent, i.e., the variables that exhibited higher NMI values in the previous months (i.e., Jun and Jul) continued to show strong connections, while the other links with lower NMI

values during that period (i.e., Jun and Jul) became weak or insignificant. Overall, in 2013, DS and SPI3, SMDI and SPI3, DS and SEV, TMIN and SPI3 represented a consistent association with each other during the growing season. This outcome was seen as a result of the changing nature of input forcings onto the model. To elaborate, by incorporating real forcing data (i.e., less observed weather data from NMME and more input data sampled from climatology for the remaining season), the bulk of the noisy links weakened over time and certain strong links persisted throughout the growing season. For e.g., in 2013, we consistently observed strong connections between SPI3 and SMDI, SPI6 and SPI3, SEV and SPI3, and SMDI and DS, while other links became insignificant over the course of the growing season. This information not only indicates the robustness and stability of the dominant connections but enables us to identify key parameters that aids in understanding the influence of different variables on water availability. Additionally, we observed significant connections during the dry and mod years. Specifically, during the dry year (2015), we found a consistent relationship between SMDI and SEV, SPI3 and SEV, and DS and SMDI, while in 2019, SEV and TMIN, SMDI, and TMIN, SMDI and SEV, TMAX and TMIN, and TMIN and SPI3 exhibited a strong consistent relation. Thus, having information on the relationships between the key hydrologic variables and drought parameters during different time periods enables us to determine alternative agricultural outcomes.

Similarly, we analyzed the NMI to quantify the influence of the independent variables (i.e., both hydrologic variables and drought parameters) on crop yields. The variables that had the greatest influence (i.e., variables exhibiting higher NMI values) on crop yields were shown in the network plot by green arrows. During the wet year, we found TMIN, SMDI, SEV, and SPI3 had the maximum influence on end-of-year yields, indicating a significant statistical relationship with crop yield. Similarly, in the dry year (2015), we observed a significant statistical relationship

between crop yield and DS, TMIN, while TMAX, DS, and TMIN exhibited a significant relationship during the mod year (2019). Likewise, we conducted a similar process network connectivity analysis over Prey Veng (see Figure 20.), and the findings aligned with those of Takeo. The analysis revealed a similar conclusion—i) Connection density is higher when there is more reliance on seasonal forecasts (i.e., during the initial months, Jun-Jul), ii) Over time, the information flow networks become distinct and stronger as real (nowcast) data is incorporated into the model, and the consistent links persist, and iii) TMIN, TMAX and DS showed the maximum influence on end-of-season yield, reinforcing the significance of these variables in determining crop outcomes. Thus, the interpretation of such connections can substantially help us understand the variability and predictability of crop yields during different time in a season.

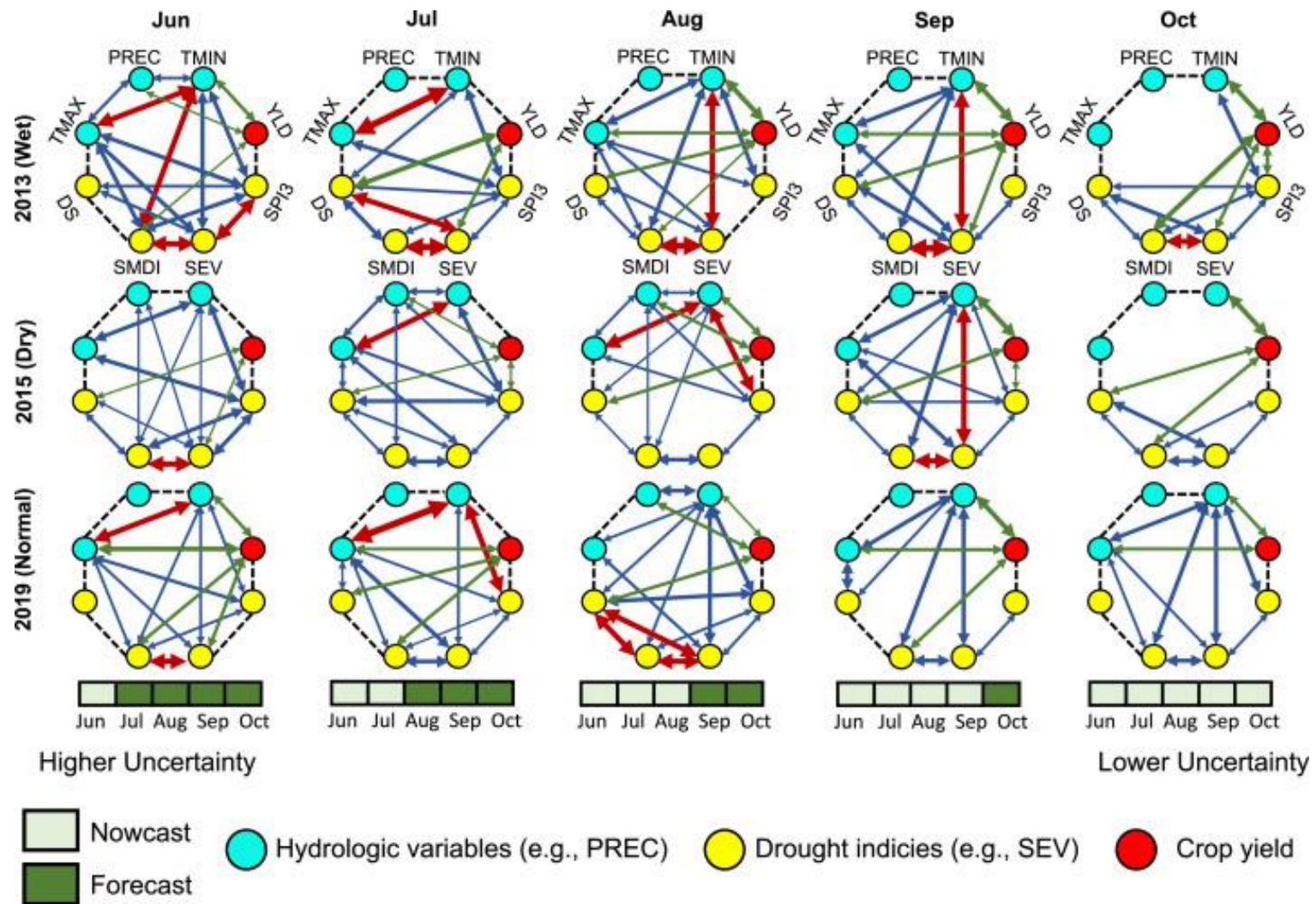


Figure 19. Process network connectivity. Establishing the potential links (connectivity) between the core hydrologic variables (represented by aqua colored circles), associated drought indicators (represented by yellow circles) and end-of-season crop (rice)

Figure 19. (cont'd)

yields (represented by red circles) over the growing season (Jun-Oct) in Takeo during the wet year (2013) (top panel), dry year (2015) (middle panel), and mod year (2019) (bottom panel) respectively. Each variable, also called a node, represents the timeseries of the related variable. The thickness of the links represents the strength of the relationship. The color variations in the plot reflect the varying degrees of significance, with darker colors indicating stronger mutual information. Red arrows represent the most dominant network connections, while the blue arrows represent moderate-low strength, and green arrows represent the strength of the independent variables that have significant influence on seasonal yields. The dashed black line merely provides completeness and visual aesthetics to the network plot. TMIN: Min. Temperature; TMAX: Max. Temperature; PREC: Precipitation; SPI3: 3-month Standardized Precipitation Index; SEV: Drought Severity; SMDI: Soil Moisture Deficit Index; DS: Dryspells; YLD: Crop yield.

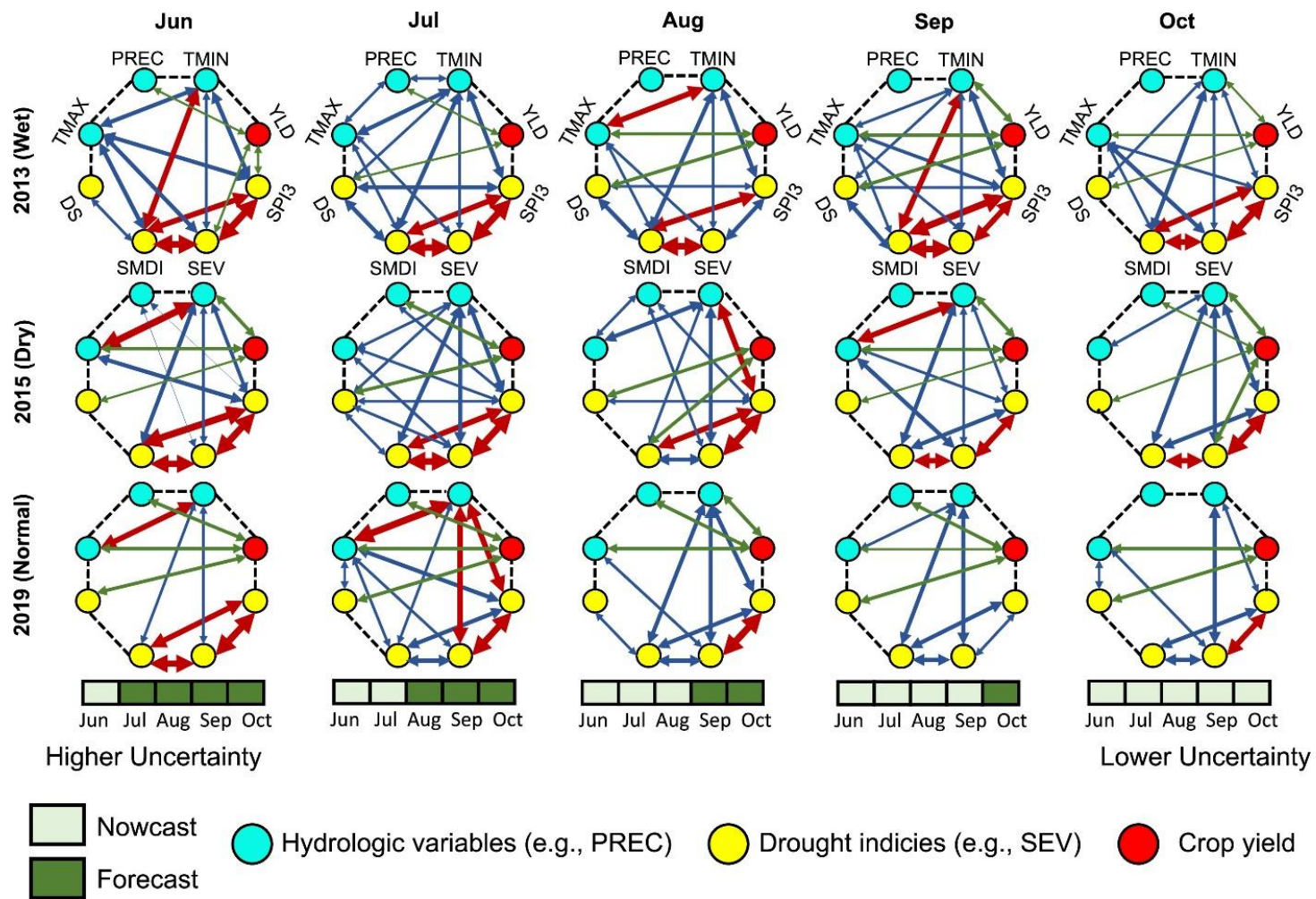


Figure 20. Network connectivity over Prey Veng. Process networks showing the connectivity between the hydrologic parameters (aqua circles), drought indicators (yellow circles) and yield (red circles) over Prey Veng during a wet (2013) (top panel), dry (2015) (middle panel) and normal (2019) (bottom panel) year respectively. TMIN: Min. Temperature; TMAX: Max. Temperature; PREC:

Figure 20. (cont'd)

Precipitation; SPI3: 3-month Standardized Precipitation Index; SEV: Drought Severity; SMDI: Soil Moisture Deficit Index; DS: Dryspells; YLD: Crop yield.

Our findings indicated DS and TMIN having the maximum influence on crop yields. This can be explained through an understanding of the physiological responses of these crops to environmental conditions. Rice yields are sensitive to minimum temperatures, as cooler nights can slow metabolic activity and induce flower sterility, particularly during the grain-filling and reproductive stages. Dry spells are also critical, as rice, typically cultivated in water-rich environments, suffers from water stress that hampers photosynthesis, nutrient uptake, and overall growth when water is scarce. Conversely, maize is more sensitive to maximum temperatures than rice. High heat during maize's flowering period can dry out the female flowers and disrupt pollination, leading to poor kernel formation. Though rice can withstand higher temperatures due to its aquatic growth environment, its yield is more susceptible to the effects of low temperatures and dry periods. Thus, understanding these crop-specific environmental sensitivities is key for effective agricultural management and forecasting.

3.4. Conclusion

We employed a novel technique for evaluating the efficacy of historic seasonal climate forecasts based on a multi-model ensemble. Based on the current application, RHEAS performed reasonably well in quantifying the yield uncertainties associated with seasonal climate forecasts. Particularly, the *m*-DSSAT crop model was able to capture the uncertainties in end-of-season yield, whilst significantly complementing observations over the growing season. To characterize the uncertainties, the *m*-DSSAT model provides a 40-ensemble yield scenario based on a 1-month timeframe moving window. These model ensembles simulate yield based on the nature of—i) meteorological forcings (from the VIC hydrologic model), ii) cultivar varieties, iii) fertilizer application rates, whilst capturing the model uncertainties, either arising from model parameters or agronomic practices. The results from the study showed the extent of yield values becoming

better constrained (i.e., reduction of uncertainty) with each subsequent timeframe over both provinces. Overall, the model provided a near-accurate yield forecast with reasonable bias (that improves over every forecast/month) at longer lead times (see Figure 17.). In addition, the model performed reasonably well in matching the observations over both provinces. However, due to the absence of real input forcings during the initial months (Jun-Jul-Aug) of the season, there were substantial discrepancies (gaps or bias) between the simulated and observed yields. However, with the availability of more nowcasts in subsequent months (Sep-Oct), the average simulated yield mimics the observations, thus significantly improving the initial bias. As no parameters were changed other than the meteorologic forcings, this inherently connotes the quality of initial forcings as one of the critical drivers for yield prediction. Not only do these results support our hypothesis, but also provide a substantial ground for using such an approach to improve crop yield forecasting for agriculture sustainability.

Even though the impacts of weather-related uncertainties can be modeled satisfactorily, understanding the relationships between the key hydrologic and agriculture variables is essential for identifying the dominant pathways and variables affecting interannual crop yields. Herein, we employed the concept of MI to discern the synchronous flow of information between the key hydrologic and agriculture variables. Particularly, we looked at the information mutually exchanged between the hydrologic (and drought) parameters (independent variable) and crop yield (dependent variable). Our results showed a significant increase in NMI during the initial timeframes (Jun-Jul) as compared to the later timeframes (Sep-Oct), most notably during the wet (Takeo, Prey Veng) and dry year (Takeo) (see Figure 18.). In other words, such behavior essentially implies the significant transfer of information at a timeframe wherein we have the maximum weightage of the present (nowcast) conditions, i.e., when our yield predictions are

completely based on real simulated data (and not on historic forecasts). Whilst higher yield uncertainties arrayed less NMI during the Jun-Jul timeframe, there were few instances of relatively stable (2019, Prey Veng) or irregular (2019, Takeo and 2015, Prey Veng) behavior of the independent variables towards yield. Over the study duration, DS and TMIN maintained a consistent level of statistical significance with yield, i.e., TMIN and DS were the dominant parameters having the maximum influence on interannual yields. In addition, we observed that hydrologic variables, specifically temperature, demonstrated a stronger correlation with yield in 10 out of 15 months in the Takeo region (i.e., temperature have a greater impact on crop yield in Takeo). Conversely, in Prey Veng, we found a higher correlation between yield and 8 hydrologic variables and 7 drought variables throughout the study period (i.e., a broader range of hydrologic and drought-related factors influence crop yield in Prey Veng). It is noteworthy, while it suggests a higher significance of particular variables during a certain timeframe, it is important to consider that the significance may vary depending on specific years, regions, and prevailing conditions.

Additionally, we looked at the process network connectivity between the key hydrologic variables, drought indicators, and yield. Our results indicated the prevalence of multiple network connections between the independent variables during the start of the season (Jun) (see Figure 19. and Figure 20.). However, the bulk of these links diminishes over time and the strong prominent links persist at the end of the season. Such a pattern can be attributed to the forcing information provided to the model, i.e., the associated uncertainties in the forecast. During harvest, the model simulates the hydrologic parameters based on the real conditions (nowcast) which reduce the noisy links and highlight the explicit networks. Likewise, we presented the dominant links between the independent variables and yield, with DS and TMIN being the most influential parameters over both provinces. Thus, to obtain a sound and reliable forecast of crop (rice) yield with manageable

uncertainties for a growing season, it is desirable to have a seasonal climate forecast that encapsulates the information on DS and TMIN optimally. Currently, all the available seasonal forecasts do not have such information. Hence, including skillful information on DS and TMIN will effectively enhance the forecast capability with uncertainty estimates essential for agricultural sustainability.

Chapter 4.

EVALUATING THE CROP YIELD PREDICTABILITY THROUGH SEQUENTIAL DATA ASSIMILATION

This chapter delves deeper into the realm of crop yield predictability, specifically exploring the potential enhancements through sequential data assimilation. While the preceding chapter focused on using historic climate forecasts to refine yield predictions, our current focus shifts to the integration and assimilation of pivotal remote sensing data, aiming to further elevate the precision of crop yield estimations.

4.1. Introduction

Physically based models in hydrology serve as pivotal tools for macro-scale land surface modeling. These models enable the estimation of a myriad earth science variables like temperature and solar radiation, and effectively encapsulate the nuances and dynamics of the water cycle's various components. At their core, these systems are designed to simulate the historical, present-day, and potential future hydrologic responses arising from intricate earth-atmosphere interactions. However, regardless of the developments in hydrologic modeling, the uncertainties associated with the models are inevitable. Significant scale differences and sparse, unreliable ground-based (gauge network) estimates, together preclude the reliable applications of land surface models in scientific studies. In addition, the input and structural components of physical models include substantial uncertainties (or errors), often leading to unreliable estimates of predictive outputs. Abhishek et al. (2023) discussed the uncertainties associated with crop yield forecasts by evaluating the efficacy of historic seasonal climate forecasts. One of the major hurdles lies in the mathematical representation within these models, which inherently possesses a degree of uncertainty. A prime example of this uncertainty can be seen in the modeling of 'soil moisture' (SM). This parameter is

crucial for delineating surface and sub-surface water fluxes. Yet, the uncertainties stemming from satellite retrievals, meteorological inputs, ground-based (and in-situ) measurements, and model parameterization lead to distinctive error characteristics.

This uncertainty is not limited to hydrological models alone. Crop models too are subjected to observational errors, often as a result of sampling and measurement inaccuracies. Identifying and tracking these uncertainties is vital to guide future data collection and model refinement strategies. In the realm of data acquisition, while the ongoing trend of miniaturization and the increasing use of commercial drones have facilitated high-quality, field-scale data collection, satellite observations remain the gold standard for large-scale vegetation monitoring. Development of vegetation indices (such as the Normalized Difference Vegetation Index, NDVI) and other growth attributes such as LAI derived from satellite products have significantly improved the characterization of vegetation parameters.

SM is a key variable that links the water and energy balances at the land-atmosphere interface, thus forming a critical component of the hydrologic cycle. This proportion of water content in the unsaturated soil zone not only controls the numerous processes relevant to the water and energy cycles, but also impacts the exchange of trace gases, such as carbon dioxide, over land. Although, SM is a storage component for water (from precipitation), it is instrumental in driving feedback mechanisms (e.g., SM-climate feedback such as land surface precipitation and evaporation, etc.) within the local and regional climate systems, exemplified by its influence on processes like land surface precipitation and evaporation. The return of water to the atmosphere, through mechanisms like plant transpiration and bare soil evaporation, is intrinsically tied to the availability of water within the unsaturated soil. Such feedback loops are strikingly evident in regions like the Amazon and the Mississippi River basin, where a fraction of the soil moisture

contributes to atmospheric water content, subsequently influencing rainfall patterns. This underscores the significance of soil moisture as a primary regulator for various hydrological responses, from rainfall-runoff dynamics to drought predictions. It plays an important role in forewarning impending drought conditions (Narasimhan and Srinivasan, 2005), managing water resources (Dobriyal et al., 2012), tracking crop phenology (D'Odorico et al., 2007), and development of weather patterns. Accurate estimation of soil moisture can be deemed crucial for the identification of agricultural drought, specifically the areas under heat and moisture stress can be given special attention prior in time. As a result, the losses/damage associated with crops/plants can be minimized to an extent, thus ensuring optimal productivity.

Many studies have highlighted the impact of weather extremes, hydrologic stresses over rainfed agriculture systems (Abhishek et al., 2023, Butler and Huybers, 2015, Zhao et al., 2017, Lobell et al., 2013). However, no substantial evidence has been established to explain the nonlinearity in the relationship between soil moisture and crop yields (Schauberger et al., 2017, Urban, Sheffield and Lobell, 2015). Although, remote sensing observations provides a comprehensive measurement of soil moisture at large spatial scales, they lack information on a temporal basis (i.e., they represent a snapshot in time). Due to the large revisit interval of the satellites (usually in days), there is a gap between successive observations, which may not reflect the changes amidst the two consecutive overpasses. For instance, soil moisture is critical for plant growth (especially during germination, reproductive stages) and drought monitoring, but the poor temporal coverage of satellites limits the optimal productivity of crops (Mladenova et al., 2017), and inadequate monitoring of drought severity, water stress (Martinez-Fernandez et al., 2016). Although soil moisture does not change much on a day-to-day basis apart from external events

such as precipitation or irrigation, these physical changes still have a substantial impact on the overall moisture content over the topsoil and rootzone layers.

Similarly, LAI is an important biophysical parameter that provides information on multiple facets of vegetation dynamics, ranging from light interception and photosynthesis to canopy structure. Its relevance extends further, playing a pivotal role in modeling the intricate exchanges of water and carbon between the biosphere and atmosphere (Asner et al., 2003). Oftentimes, LAI is commonly assimilated into process-based crop growth models to monitor the vegetation dynamics and estimate the biomass and grain yield (Quaife et al., 2008). Unlike SM, the relationship between ‘yield’ and ‘LAI’ is not straightforward, as LAI varies significantly during different stages of crop growth (especially during silking and grain filling), thus having a significant influence on end-of-season yields. For instance, due to the extreme sensitivity of crops to warmer temperatures, the crops exhibit an accelerated leaf senescence (Chen et al., 2010). While in-situ measurements and model-based evaluations offer one avenue for LAI estimation, the MODIS LAI product has emerged as a popular alternative, providing broad spatial (~500 m) and temporal (~8 day) coverage. This product has been extensively used to monitor the impact of drought (and other environmental stressors) on vegetation, crop productivity, and predicting interannual yields, and optimization of management practices. However, the structural complexity of canopy, heterogeneity in land coverage, spatial distinction of finer-scale features (due to sensor spatial resolution), presence of clouds and aerosols limit the ability of the product, hence resulting in underestimations relative to reality (i.e., observations).

Even if, perfect model states and good quality observation retrievals are provided, the uncertainties cannot be expected to disappear completely. However, with considerable adjustment and sufficient validation (from various sources), the use of ‘Data Assimilation’ in scientific

hydrologic (and other) studies present the near-accurate representation of the actual states. Data assimilation allows the optimal merging of the model estimates (simulations) and observations (station/remote sensing) to produce the best possible state of present conditions. Observations are often incorporated into modeling frameworks to summarize the current state of the system. As remote sensing observations only provide information at a single time step, dynamic models are used to predict the spatial and temporal variations in a system. Analysis using different data assimilation techniques not only provides the best estimates of the state of a physical system, but melds a powerful methodology for accurate, reliable prediction (Abaza et al., 2014, Vrugt et al., 2006). In addition, DA is used for model initialization, updating model parameters (Hendricks and Kinzelbach, 2008) and filling observational gaps in data-scarce regions (Lahoz et al, 2010). Numerous precursory studies have focused on the use of assimilation techniques for data handling and addressing model complexities (Andreadis et al., 2006; Bosilovich et al., 2007; Margulis et al., 2002).

Among all the techniques, the Monte Carlo based Ensemble Kalman Filter (EnKF: Evensen, 2003) has been extensively used to integrate real-time data in hydrologic models. EnKF allows the optimal merging of model simulations with observations in a Bayesian sense by applying an ensemble of model states to represent the error statistics in both. These uncertainties from model parameters and observations are then optimally weighted and adjusted towards the actual measurements. Because of its sophisticated structure, easy implementation, optimum performance, and capability to handle nonlinearities in hydrologic models, the EnKF is often used instead the ordinary Kalman filter (Pasetto et al., 2012). The implementation of this methodology has been successfully applied in hydrologic studies, such as surface-groundwater interaction (Kurtz et al., 2014), rainfall-runoff modeling (Vrugt et al., 2006), flood forecasting (Seo et al.,

2009), land surface modeling (Crow and Wood, 2003), and integrated frameworks (combined information system) (Andreadis et al., 2017), among others (Das, Mohanty, Cosh and Jackson, 2008; Dunne and Entekhabi, 2005; Reichle, Walker, Koster and Houser, 2002; Zhou et al., 2006; Das and Mohanty, 2006).

The primary objective of the study was to quantify the enhancements in crop yield prediction for two key cereal crops, maize and rice, by assimilating remotely sensed soil moisture and LAI observations. Specifically, we sought to understand the differential impacts, if any, before and after the assimilation of these remote sensing observations across distinct crop types. We hypothesized that—the assimilation of satellite-derived soil moisture and LAI would lead to notable improvements in the seasonal yield prediction for maize, while minor enhancements might be observed in rice-dominated areas. Our focus regions for this study were Kenya and Cambodia, both of which have extensive rainfed agricultural areas dedicated to maize and rice cultivation, respectively. Given the global significance of maize and rice as staple food sources, achieving an optimal forecast of their interannual yields is paramount. To accomplish our objectives, we employed the comprehensive RHEAS framework, specifically tailored for assimilating remote sensing data directly into the crop modeling system. The primary intent behind leveraging RHEAS was to offer a more refined and constrained crop yield prediction, particularly valuable in regions with limited data availability.

4.2. Data and Methods

4.2.1. Study Area

Our study regions—Kenya and Cambodia—together hosts large areas of maize and rice production respectively. Like most developing countries, Kenya, and Cambodia (Figure 21.) have a predominant agrarian economy, with majority of the population actively engaged in agricultural

activities. Located in the Horn of Africa, the agricultural regions of Kenya (primarily in the West) mostly inherit their climate from Lake Victoria. Typically classified as a rainforest, the precipitation pattern is bimodal, wherein the long rain season spans between Apr-Jul and short rain season spans between Oct-Dec (Midega et al. 2015). With a mean annual precipitation of 900-1700 mm and annual temperatures ranging between 18-19 °C, Kenya has intensive cropping with no irrigation. Common crops grown in the region include wheat, barley, maize, coffee, and sunflower. The western provinces, specifically Trans Nzoia, Uasin Gishu, Bungoma, Kakamega, Nakuru are considered as some of the agricultural productive regions of Africa. On the other end, Cambodia derives most of its hydroclimatic features from the monsoon regimes and the flow of the Mekong River. During the major southwest monsoon (May-Oct/Nov), the Mekong River flows into the Tonle Sap Lake, whereas the short season (Nov/Dec-Apr) exhibits a flow reversal from the Tonle Sap Lake to the Mekong River (MRC, 2014). Such seasonality has significant influence on the wetland ecosystems and agricultural production. Although the Tonle Sap floodplains and the southeast provinces (part of the Mekong delta) host ~90 percent of paddy plantation, large discrepancies in mechanization, technology, irrigation capacity, etc. have limited their prospects of growth when compared to the neighboring countries (Mainuddin et al., 2013). In addition, climate change has threatened the economic development, food security and environmental sustainability of both countries, thus adding complexity to the existing challenges with food and water resources.

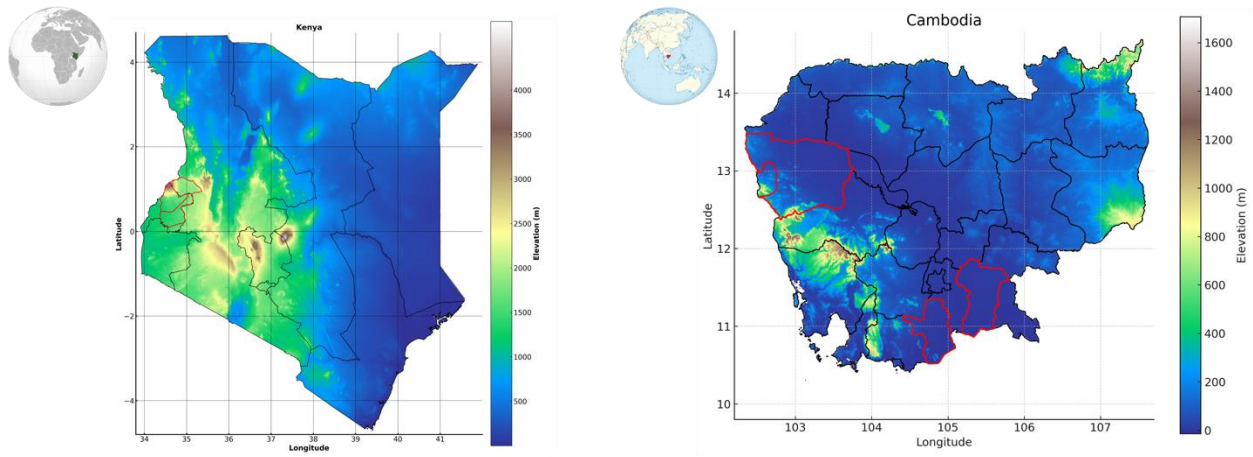


Figure 21. Elevation maps of Kenya (left) and Cambodia (right), with respective study areas highlighted in red.

4.2.2. Model Description

In this study, we implemented the RHEAS framework over both regions and used data assimilation to provide seasonal forecasts of yield estimates and associated drought status. RHEAS enables the (loose) coupling of the VIC hydrological model with the DSSAT crop model. Due to system modularity and flexibility of RHEAS to couple with other environmental models (in this case, a crop model), the hybrid architecture of RHEAS allows the easy accessibility of model datasets (input and outputs), forcing files, remote sensing retrievals, etc. from a centrally located Postgres enabled relational GIS database (Holzworth et al., 2015), whilst performing end-to-end validation/calibration, nowcast/forecast operations, data assimilation, among others.

Using a suite of meteorological forcings, the VIC hydrologic model computes the energy and water balance over a gridded domain with a 0.05° spatial resolution (~ 5 km). While the VIC has found wide applicability across domains (Hamman et al., 2018), RHEAS's modular design allows VIC to harmoniously interface with other models, such as *m*-DSSAT in our case. The *m*-DSSAT, as incorporated into RHEAS, is a modified version of the standard model, emphasizing a multi-ensemble approach. This design caters specifically to data assimilation, permitting periodic

perturbations of the crop model throughout the growing season based on incoming remote sensing observations. In essence, *m*-DSSAT can halt and resume its simulation in alignment with the assimilation of observations, primarily SM and LAI.

Within the RHEAS framework, the assimilation of data is achieved through an EnKF scheme. By merging the predicted state of the system from the model with incoming observations, the EnKF refines the state estimates iteratively. This dynamic updating reduces discrepancies between the model's predictions and real-world observations, enhancing the model's accuracy. The assimilation process in RHEAS is designed to handle non-Gaussian errors and incorporates a sequential updating scheme that utilizes observational data as they become available. This ensures that the most recent and relevant data are consistently integrated into the model, optimizing predictive capabilities. Figure 22. shows the structure of the assimilation setup used in the study.

Figure 22. (cont'd)

Moisture (SM) data and Leaf Area Index (LAI) data into the *m*-DSSAT system. At the center, the *m*-DSSAT employs the EnKF algorithm ensuring the seamless fusion of these diverse data sources, ultimately enhancing the precision and reliability of the model's output.

4.2.3. Data

For the VIC model, a daily timeseries of meteorological forcings is paramount. Precipitation data for this study were sourced from the high-resolution CHIRPS product available at a spatial and temporal resolution of 0.05° and 1981-present respectively. Meanwhile, the requisite air temperature and wind speed data were derived from the NCEP gridded reanalysis product. Complementing these, soil-texture data (capturing aspects like permanent wilting point and hydraulic conductivity) and land-use classifications (differentiating vegetation types, bare soil, etc.) were incorporated to compute various hydrological fluxes and states, such as evaporation and soil moisture. Likewise, the *m*-DSSAT crop model in RHEAS harnesses VIC's outputs to simulate crop yield predictions. The crop model requires daily weather records, comprehensive soil profile data, and a detailed record of agricultural management practices. Crucially, inputs pertaining to cultivar types, irrigation methods, and fertilizer application rates play a pivotal role in determining plant growth trajectories, developmental stages, and eventual yields. To bolster the robustness of our model, we integrated additional ancillary data, collated from previous research, published reports, and well-known databases like the Food and Agriculture Organization (FAO) and The World Bank. Table 7 lays out a list of used datasets in the study.

Table 7. List of datasets used in the study with their spatial and temporal resolution.

Variable	Product	Spatial Resolution	Temporal Resolution	Period Availability
Precipitation	CHIRPS	5 km	Daily	1981 - present
Temperature	NCEP	1.875 deg	Daily	1948 - present
Wind Speed	NCEP	1.875 deg	Daily	1948 - present
Soil Moisture	SMAP	~9 km	2-3 days	2015 - present
Leaf Area Index	MCD15	500 m	8 days	2002 - present
Topography/ Global DEM	GTOPO30	~1 km	static	1996
Land cover	MODIS	500 m	1-2 days	2000 - present
Soil Profile	SoilGrid-250m 2.0	250 m	static	2017

Note: CHIRPS: Climate Hazards Group Infrared Precipitation with Station data, NCEP: National Centre for Environmental Prediction, SMAP: Soil Moisture Active Passive, DEM: Digital Elevation Model, MODIS: Moderate resolution Imaging Spectroradiometer

4.2.4. Assimilation Data

The SMAP mission collects consistent, systematic data across the microwave portions of the electromagnetic spectrum at varying spatial resolutions, providing valuable information on the surface SM (and other ancillary information, e.g., brightness temperature) content over the topsoil layer (~5 cm), every 2-3 days. The SMAP mission was intended to improve the estimates of water and energy transfer between land and atmosphere (Entekhabi et al., 2010a). Although, the original footprint of the SMAP product is approximately 36 km, the enhanced Level-3 (L3) SM product provides a composite of daily estimates of land surface conditions at a 9 km grid resolution. This enhanced product is derived from the Level-2 (L2) SM product (using the Backus-Gilbert optimal interpolation techniques), posted to a 9 km Equal-Area Scalable Earth Grid, Version 2.0 (EASE-Grid 2.0) cylindrical projection. Using the downscaled brightness temperature (T_b) and radar backscatter observations of the gridded 36 km SMAP product, the enhanced SMAP 9 km product

is created. Detailed information about the SPL3SMP_E product can be found in O'Neill et al. (2020) and Xing et al. (2023). Das et al. (2018) laid out the advantages of using the enhanced 9 km product wherein finer details about the moisture content over some areas were captured. Hence, using this product can have a significant influence on agricultural parameters (e.g., crop growth, phenology, and yield). In addition, the accurate characterization of moisture content at 9 km spatial resolution can provide substantial information on the prevailing hydrologic conditions and drought stress over agricultural areas, so proper attention can be given to the areas under stress prior in time. Herein, we assimilated the SMAP enhanced 9 km product into the RHEAS framework to see the improvement in drought and crop yield predictability over different crop types.

Similarly, Leaf Area Index (LAI) is an essential bio-geophysical parameter that plays a pivotal role in monitoring and understanding vegetation health, vigor, and growth dynamics. LAI, which quantifies the amount of leaf material in canopies, provides critical insights into the potential photosynthesis and the overall energy balance of vegetative systems, directly influencing the interaction between the earth and its atmosphere. For our study, we utilized the MCD15 MODIS product, which offers robust and reliable LAI estimations. The MODIS (Moderate Resolution Imaging Spectroradiometer) instruments onboard the Terra and Aqua satellites have been instrumental in delivering consistent and widespread observations of the earth's surface. Specifically, the MCD15 product offers LAI data at a 500m resolution on an 8-day composite period, ensuring a fine balance between spatial granularity and temporal frequency. MCD15 is generated using a radiative transfer algorithm, which models the interaction between sunlight and the earth's surface. The algorithm ingests reflectance data from various MODIS bands and then outputs LAI after accounting for different vegetation types and associated biome-specific parameters. This methodology ensures that the derived LAI values are not just mere statistical

extrapolations but are rooted in the actual physical and biological characteristics of the vegetation. The MODIS LAI product, given its technical robustness and extensive validation, has been employed in a myriad application, from tracking forest health, monitoring agricultural fields, to understanding the dynamics of climate change (Myneni et al., 2002; Yang et al., 2006). Its ability to capture rapid changes in vegetation structure, especially during critical growth phases, makes it invaluable in studies like ours, where understanding vegetation dynamics in real-time is essential. Moreover, previous studies have shown that the MCD15 MODIS product provides a reliable and accurate representation of LAI, even in challenging conditions like dense canopies or varied terrains (Zheng & Moskal, 2009; Fensholt & Sandholt, 2003).

The assimilation of LAI data into the crop models serves multiple pivotal functions. Firstly, it enhances the model's capability to monitor real-time changes in vegetation dynamics, especially during critical growth phases. This is instrumental in crop models since LAI dynamically changes through the crop's growth stages, directly influencing light interception, photosynthesis, and water use efficiency, all of which are crucial components in determining the final yield. Moreover, LAI is a sensitive indicator of plant stress, such as those caused by drought, pests, or nutrient deficiencies. By assimilating LAI data, the models can promptly detect any deviation from normal growth conditions, allowing for a more accurate representation of the crop's actual health and performance in the field. This is particularly important for forecasting yields, as it enables the model to adjust yield predictions based on the crop's observed vitality and stress conditions throughout the growing season.

4.2.5. Methodology

4.2.5.1. Open-loop Simulations

Firstly, we perform an open-loop simulation, wherein the hydrologic model simulates the core hydrologic variables over Cambodia and Kenya. Daily timeseries of precipitation, air temperature, and wind speed are forced into the VIC model to generate the current hydrologic states (e.g., soil moisture, transpiration, runoff generation, etc.). Specifically, we looked into the hydrologic conditions over the respective (major) growing seasons, i.e., Apr-Jul (over Kenya) and Jun-Oct (over Cambodia). Prior to the model simulations, the VIC hydrologic model within RHEAS had been calibrated at a global (and regional) extent based on the study by Zhang et al., 2018, and Sheffield and Wood, 2007. Previous application of RHEAS at coarse spatial resolutions (0.25°) were based on the aforementioned calibration methodology. However, herein we attempted to calibrate the VIC (at a finer resolution of 0.05°) using the SM and streamflow data over East Africa and the Mekong region. As the region of interest boasts large spatial variability, an accurate representation of hydrologic states is essential for simulating the hydrologic response. Using the Shuffle Complex Evolution (SCE) algorithm (Duan et al., 1992), the initial set of parameters (e.g., soil hydraulic conductivity, soil bulk density, etc.) values were adjusted to minimize the difference between model outputs and observations.

Following the hydrologic simulation, the crop model simulates the crop phenological development, growth from emergence until maturity. However, before running the simulation, other ancillary information, such as cultivar genotypes, fertilizer rates, etc., are complemented to the crop model. Based on the cultivar coefficient parameters, provided in Table 8., for both crops (rice and maize), the selected regions are sufficiently calibrated using the best available data (e.g., yield observations). Based on previous approaches (by Wang et al., 2010), we implemented the

SC algorithm for calibrating the rice and maize genotype coefficient parameters for the respective regions. Following a complex approach of multiple periodic shuffling, sampling, and competitive evolution (Naeini et al., 2018), the SC algorithm optimizes the selected genotype parameters to provide a set of calibrated parameters. The SC divides the parameter space into a set of subspaces and randomly selects values from each subspace to generate a new set of parameter values (in this case, the various growth and phenological parameters such as P1, P2O, P5, G1, G2, etc. (laid out in detail in Table 8. and Table 9.)). The objective function evaluates the parameter values, and the best set of values are retained for the next iteration. The above steps are repeated until the objective function converges to a minimum, indicating that the optimal parameters have been found. Using these calibrated set of cultivar parameters, and other ancillary information (e.g., fertilizer application rates, etc.), the *m*-DSSAT generates an ensemble (~40 members) of yield values over each season/year. The yield estimates within a season are computed based on the underlying hydrologic forcings that are fed by the hydrologic model, thus exhibiting a coupling approach between both modeling systems. Due to the scarce availability of agronomic data (and resources), proper attention has not been given to these developing agricultural productive regions. However, we selected provinces that host large percentage of agricultural productivity (specifically, maize for Kenya and rice for Cambodia) based on previous studies, published literatures/reports, etc. Generally, having the correct information about cultivar varieties and fertilization application rates can significantly enhance the crop model performance. Detailed functioning of the modeling systems has been explained in Andreadis et al. (2017) and Abhishek et al. (2021).

Table 8. Estimates of genotype coefficients of the maize varieties used in Kenya. The *m*-DSSAT crop model was calibrated using the SCE algorithm for getting an optimized set of phenological and growth parameters.

Province	Phenological coefficients			Growth coefficients		
	P1	P2	P5	G2	G3	Phint
Bungoma	121.757	0.730186	570.086	491.85	9.25675	36.78
Transnzoia	114.658	0.742115	694.132	546.364	9.21766	36.78
Kakamega	148.866	0.791032	671.426	417.251	9.32528	36.78

Phenology genetic coefficients

P1: Thermal time from seedling emergence to the end of juvenile phase (°C day)

P2: Delay in development for each hour that day-length is above 12.5 hours (days)

P5: Thermal time from silking to time of physiological maturity (°C day)

Growth genetic coefficients

G2: Maximum kernel number per plant

G3: Kernel growth rate during linear grain filling stage under optimum conditions (mg day⁻¹)

Phint: Thermal time between successive leaf tip appearance (°C day tip⁻¹)

Table 9. Estimates of genotype coefficients of the rice varieties used in Cambodia. The *m*-DSSAT crop model was calibrated using SCE algorithm for getting an optimized set of phenological and growth parameters.

Province	Phenological coefficients				Growth Coefficients			
	P1	P2R	P5	P2O	G1	G2	G3	G4
Takeo	455.707	303.639	398.178	13.2159	49.5898	0.0311578	1.00	1.20
Prey Veng	524.362	356.726	365.103	13.6226	51.0014	0.0232092	1.00	1.20
Battambang	381.722	258.035	338.95	13.2988	53.9522	0.0313879	1.00	1.20

Phenology genetic coefficients

P1: Time period in °C (above a base temperature of 9 °C) from seedling emergence to the end of juvenile phase. Expressed as growing degree days (GDD)

P2R: Extent to which phase development leading to panicle initiation is delayed for each hour increase in photoperiod above P2O. Expressed as growing degree days (GDD)

P5: Time period from beginning of grain filling to physiological maturity with a base temperature of 9 °C. Unit: GDD

P2O: Critical photoperiod or the longest day length at which development occurs at a maximum rate. Expressed in hours

Growth genetic coefficients

G1: Potential spikelet number coefficients as estimated from the number of spikelets per g of main culm dry weight (less leaf blades and sheaths plus spikes) at anthesis. Unit: Spikelets per g of main culm

G2: Single grain weight under ideal growing conditions, i.e., non-limiting light, water, nutrients, and absence of pests and diseases. Expressed in grams.

G3: Tillering coefficient (scalar value) relative to IR64 cultivar under nonlimiting conditions; **G4:** Temperature tolerance coefficient

4.2.5.2. EnKF Assimilation Framework

The EnKF is an advanced assimilation technique tailored for non-linear models. It provides a sequential estimation technique to update model predictions with observational data, doing so through the use of an ensemble of model states to capture the error statistics. Its fundamental philosophy rests on the idea of using an ensemble of model states to represent the inherent uncertainties of a system, bridging the gap between model predictions and real-world observations. These ensembles are a collection of potential model states representing different possible realizations of the system. Rather than relying on a singular deterministic forecast, the EnKF propagates this ensemble to capture the non-linear dynamics and uncertainties inherent in complex systems. These ensemble members are typically birthed by introducing minor perturbations to a base state, ensuring diversity within the ensemble. Generally, EnKF follows a two-step approach at every time step: i) forecast step, and ii) update step.

- i) Forecast step: The first stage in the EnKF cycle is the forecast (or prediction) step. The ensemble members are propagated forward in time using the model dynamics:

$$x_i^f = M(x_i^a) \quad (5)$$

where, x_i^f is the forecast of the i^{th} ensemble member, M represents the model, and x_i^a is the analyzed state from the previous assimilation step.

- ii) Update step: With the availability of new observations, the ensemble members are updated. The analysis ensemble mean and members are computed using:

$$\bar{x}^a = \bar{x}^f + K(y^0 - H\bar{x}^f) \quad (6)$$

$$x_i^a = x_i^f + K(y^0 + \epsilon - Hx_i^f) \quad (7)$$

where, y^0 is the observations, H is the observation operator and ϵ is the observational error.

iii) Computation of Kalman Gain: Then, the Kalman Gain ‘K’ is computed:

$$K = P^f H^T (H P^f H^T + R)^{-1} \quad (8)$$

where, P^f is the forecast error covariance matrix estimated from the ensemble, R is the observational error covariance matrix.

In our study, the EnKF integrates the remotely sensed SM and LAI observations into the *m*-DSSAT crop model. This assimilation refines the crop model predictions, enhancing the accuracy of crop yield forecasts. Each ensemble member captures a possible growth trajectory of the crop. The assimilation ensures that these trajectories align with observed growth patterns.

4.2.5.3. Assimilation of SM and LAI

Post the open-loop simulation, we used DA to refine the prediction of crop yields. This process serves as the bridge between remote sensing observations and model predictions, allowing us to anchor our model's forecasts in the reality captured by satellite data. The first step involved downloading the SMAP enhanced Level-3 (SPL3SMP_E) product from 2015 to 2022 over Kenya and Cambodia. These datasets are often provided as HDF5 files, so the first step involves understanding the structure of the dataset, encompassing the various variables, their dimensions, and metadata. Following that, the datasets are transformed to a geographic coordinate system. For each grid point (i.e., latitude and longitude), a time series of SM values is extracted by iterating over each grid point and collecting values across time. The SMAP data also includes quality flags, which help in filtering out potentially unreliable (due to cloud cover, interference, etc.) data points. Then, we filter out the data based on selected agriculturally productive provinces in both study regions. The selected data within these provinces are aggregated by taking the mean SM for each day across all the years, creating a climatology of daily mean SM values. These updated values

serve as the input for assimilation into the *m*-DSSAT crop model. The same process is repeated for LAI observations from the MCD15 MODIS product.

To capture and represent the inherent uncertainties within the model, an ensemble of model simulations is generated. This is achieved by introducing slight perturbations to the model's input parameters, essentially creating a multitude of possible scenarios. These ensemble members serve as the foundational state for the EnKF. As the season progresses and as more observations are integrated, this ensemble is systematically refined to better represent the true state of the system. When SM and LAI observations are available on a particular day, they are incorporated into the crop model through the EnKF mechanism. It follows a systematic endeavor- i) An 'innovation' or 'observation-minus-forecast' term is computed, representing the discrepancy between the model's prediction and the actual observation, ii) The Kalman Gain is then computed. This matrix quantifies the weight given to the innovation in updating the model state. It is determined by considering uncertainties in both the model prediction and the observation, iii) This innovation, scaled by the Kalman Gain, is used to adjust the model's state, thereby updating the ensemble members to better represent reality, and iv) These updated ensemble members serve as the starting point for the next cycle of simulations.

Throughout the growing season, the model is in a state of dynamic evolution. With every new observation, the model's state is updated, reflecting the latest data. This iterative mechanism of model simulation, data assimilation, and state updating is a continuous process that spans the entire growing season, ensuring that the model remains attuned to the most recent observations. At the culmination of the growing season, the assimilated model, which has been periodically informed by observations, is leveraged to predict crop yields. These forecasted yields are compared with the actual observed yields to gauge the model's accuracy. Figure 23. shows a simplified

representation of the methodology employed in the study. To dissect the impact and the potential synergies of assimilating different datasets, the study followed three distinct prediction scenarios:

- i) Assimilating SM alone: In this scenario, only the remotely sensed soil moisture data is assimilated into the crop model. Soil moisture plays a pivotal role in determining the water availability to crops, influencing their growth and, ultimately, their yield. By integrating real-time soil moisture data, the model aims to achieve a more accurate representation of the current water conditions in the soil profile, thus refining the yield predictions.
- ii) Assimilating LAI alone: LAI, a key indicator of plant health and vitality, provides insights into the photosynthetic capacity of the crop. In this scenario, the model assimilates only the LAI data, allowing it to better track the growth and health of the crop canopy. By understanding the crop's photosynthetic potential at different stages of growth, the model can offer more precise yield forecasts.
- iii) Joint assimilation of SM and LAI: Recognizing that both soil moisture and plant health are instrumental in determining crop yield, this scenario seeks to harness the combined power of both datasets. By assimilating both SM and LAI data, the model aims to create a more holistic representation of the crop's environment and its growth conditions. This dual assimilation might capture the intricate interplay between soil conditions and plant health, potentially leading to the most accurate yield predictions.

Post-assimilation, the study evaluates the accuracy of the yield predictions for each scenario against actual observed yields. By comparing the predictions from the three scenarios, the study aspires to ascertain which dataset – SM, LAI, or their combination – provides the most

accurate and reliable yield forecasts. The comparative analysis offers insights into the relative importance of soil conditions versus plant health in influencing yield, and whether there's a synergistic effect when both datasets are assimilated together.

Additionally, the crops depend primarily on precipitation for water in rainfed agriculture. The depth of crop roots, which influences the zone of soil from which plants extract water, varies among crop types. For maize, roots can reach depths of more than one meter, often around 1.5 meters, allowing access to moisture available in deeper soil layers. In contrast, rice, especially when cultivated in traditional paddy systems, has shallower root systems, generally around 15-30 cm, as it is grown in fields deliberately flooded with water. In reality, the majority of soil water utilized by plants is typically found in the upper layers of the soil profile, approximately in the top 6-7 cm. This is particularly true for rice, which is often grown in saturated or near-saturated soil conditions. Maize, on the other hand, can extract water from deeper layers, especially during dry periods. These differences in root depth and water usage between maize and rice have significant implications for our data assimilation process. For maize, the model needs to effectively represent soil moisture conditions at greater depths, up to 1.5 meters. This is crucial during the dry season when maize relies on deeper soil moisture reserves. However, the current model configuration and the SMAP satellite observations primarily focus on surface soil moisture (up to 5 cm), which might not fully capture the soil moisture status in the entire root zone of maize, potentially leading to discrepancies in model performance. For rice, the presence of standing water in paddies, a typical scenario especially in rice cultivation, presents additional complexities. The current model might not accurately represent the saturated soil conditions and the dynamics of standing water, leading to underestimations of soil moisture and subsequent impacts on yield predictions. Hence, to enhance the model's performance, particularly for maize, incorporating soil moisture observations

from deeper layers, if available, or adjusting the model's root zone soil moisture estimations based on known relationships between surface and deep soil moisture can significantly improve the assimilation efficiency. This could involve calibrating the model to account for the typical root depth of maize and the corresponding soil moisture profile. On the other hand, for rice, refinements should include better representation of paddy field hydrology, such as the dynamics of flooding and draining, and their impact on soil moisture and plant growth. This might necessitate adjustments in model parameters or the incorporation of ancillary data reflecting paddy field management practices. By tailoring the data assimilation process to the specific water usage characteristics and root profiles of maize and rice, it can lead to improved soil moisture estimations and, consequently, improved crop yield predictions.

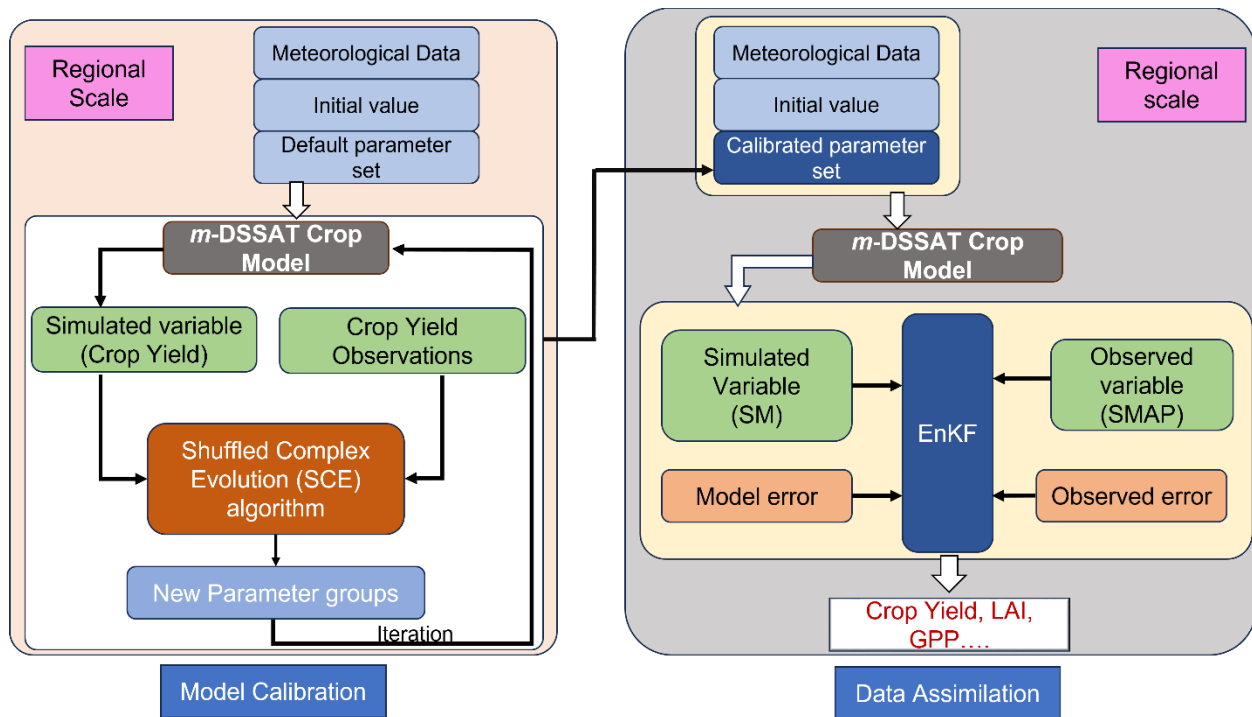


Figure 23. Detailed schematic of the methodology employed in the study. The EnKF is responsible for assimilating remote sensing observations into the *m*-DSSAT modeling system.

4.3. Results and Discussion

4.3.1. Soil Moisture Analyses

In the effort to refine the crop yield predictions through data assimilation, the accurate representation and validation of modeled soil moisture becomes imperative. This section presents a systematic comparison of our hydrologic model outputs against the well-established SMAP satellite observations. Such a comparative analysis is crucial not just for model validation, but also to discern the model's strengths and areas requiring calibration or refinements. The choice of SMAP as our benchmark for validation is underpinned by several compelling reasons: i) SMAP soil moisture observations are derived independently, with no association or reliance on our hydrologic model parameters or underlying assumptions. This ensures a robust and unbiased evaluation of our model's performance, and ii) SMAP provides frequent and consistent observations (i.e., high temporal resolution), capturing dynamic soil moisture variations. In this section, we delve deeper into the model's performance across specific provinces in Kenya and Cambodia, highlighting the key trends, deviations, and potential reasons behind them.

Each province, with its unique climatic and agricultural backdrop, presents distinct challenges and insights. By dissecting these individual case studies, as shown in Figure 24., we aim to paint a comprehensive picture of our model's performance across diverse environments. In Bungoma, the model-predicted soil moisture largely follows a similar trend as that of the SMAP satellite observations, suggesting that the model performed well in capturing the soil moisture dynamics for the entire period. However, there were few instances of underestimation by the model during specific periods. Notably, between early January to March 2021 and 2022, there was more pronounced overestimation by the model. Similarly, the predicted soil moisture from the model closely follows the SMAP observations for most of the years in Kakamega except for a noticeable

divergence in 2021 where the model significantly overestimates the soil moisture levels. In Transnzoia, there seems to be a consistent underprediction of soil moisture by the model throughout the years, especially noticeable during the middle months of each year. This divergence could be attributed to local agricultural practices or micro-climatic conditions not adequately captured by the model. Across all the maize-grown provinces discussed above, there seems to be a trend where the model occasionally overpredicts soil moisture, especially in early parts of the year. This could possibly be due to the specific agricultural practices associated with maize cultivation. For example, the onset of the maize growing season, which typically starts with land preparation and planting, might influence soil moisture dynamics.

While analyzing the data over Cambodia, the model slightly over predicts SM in Takeo, most notably during the middle of years. Other provinces in Cambodia such as Prey Veng and Battambang exhibited good correlation between the model predictions and the SMAP observations. However, there are sporadic instances where the model either over-predicts or under-predicts, notably during the mid-parts of 2021 and 2022. To sum up, there's a recurring theme of under-prediction for the rice-grown provinces, especially in Takeo. This is an indication that the model might not be capturing the specific hydrological conditions associated with rice paddies, which are flooded for extended periods. Unlike other agricultural fields, rice paddies maintain a layer of standing water on their surface for extended periods. This surface ponding, if not properly parameterized in the model, can lead to an underestimation of soil moisture, as the model might not account for this additional water content. As a result, the microwave emissions captured by SMAP are strongly influenced by the water's presence, leading to a misinterpretation of actual soil moisture levels. The other reason for the inferior performance of the model over rice in Cambodia (as compared to maize in Kenya) can be related to the inaccurate representation of standing water

dynamics in rice fields. Most hydrological models are designed to represent soil moisture in unsaturated conditions. However, the presence of standing water in rice paddies for extended periods can push the models beyond their designed operational limits. Additionally, the models don't account the associated agricultural practices (like the timing of flooding, use of bunds to retain water, or periodic draining), which can impact moisture levels, leading to discrepancies in predictions.

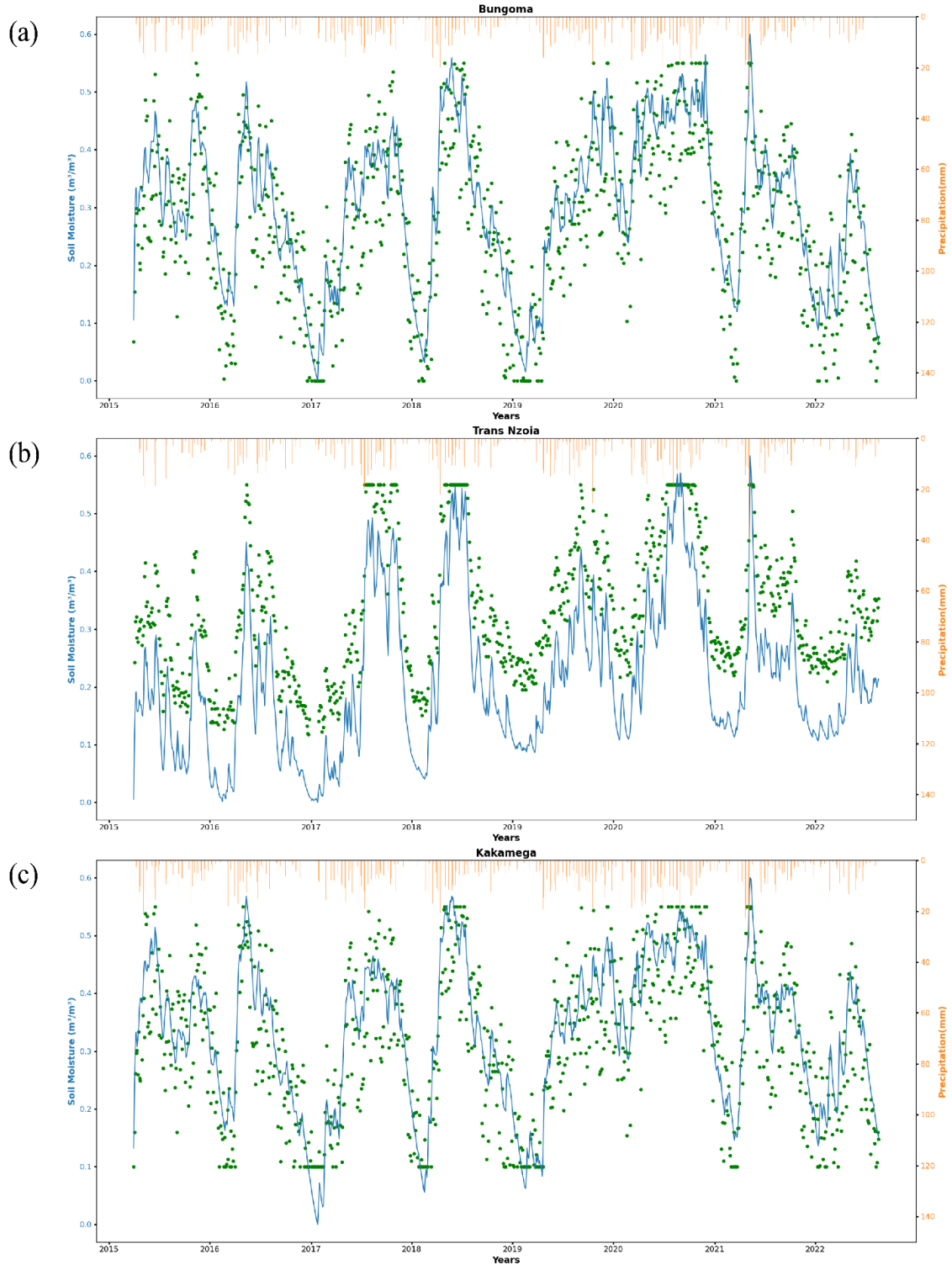
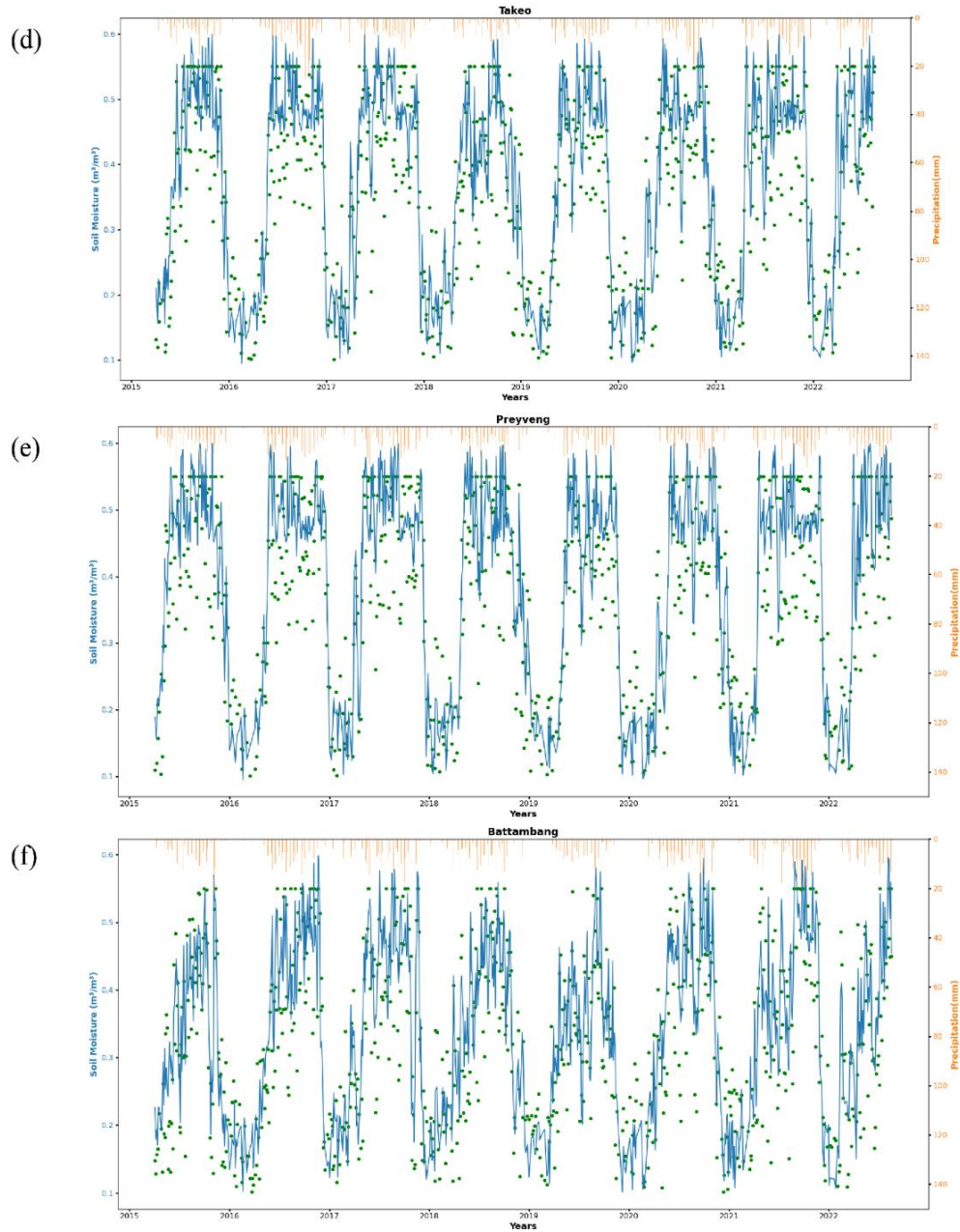


Figure 24. Temporal variations of Soil Moisture (SM) across six distinct locations: a) Bungoma, b) Trans-Nzoia, c) Kakamega, d) Takeo, e) Prey Veng, and f) Battambang. The blue line

Figure 24. (cont'd)

represents the model-generated SM, green dots depict observations from the SMAP satellite, and the orange line illustrates precipitation data from the CHIRPS satellite. The plots highlight the interplay between modeled soil moisture, satellite observations, and precipitation, capturing the overall trend and seasonality while revealing occasional discrepancies.



In summary, the model performed satisfactorily across most years and locations except for a few sporadic periods. Specifically, the model seems to have challenges in predicting extreme soil moisture conditions, either over-predicting or under-predicting. This might be due to the model's sensitivity to certain parameters or external forcings that need to be finely tuned for specific local conditions. The discrepancies between the model predictions and SMAP observations could be attributed to multiple factors. Local agricultural practices, micro-climatic conditions, irrigation practices, and even land use changes can influence soil moisture levels in a region. Additionally, the inherent uncertainties in remote sensing observations and the model's representation of complex land surface processes can also result in these discrepancies.

Additionally, we analyzed the SM performance over different locations using the unbiased RMSE spatial maps from 2015 to 2022. These spatial maps (Figure 25.) not only offer valuable insights into the performance of soil moisture estimation methods but also provide a visually accessible means to gauge their reliability against observed data. For Kenya (Figure 25a.), the RMSE values predominantly range from light yellow to deep orange, suggesting a moderate level of estimation accuracy across much of the region.

In Kenya, a large proportion of the region exhibits dark red colors, indicating a higher unbiased RMSE values, mostly in the eastern half of the country. Areas with darker shades of blue and yellow, primarily seen in the central and western part of Kenya, denoted lower unbiased RMSE values, indicating diminishing accuracy in those regions. These deeper shades, particularly towards the western border of Kenya, indicate higher RMSE values, reflecting less accurate soil moisture estimations. Several factors can explain these spatial variations. The central region of Kenya is characterized by its diverse agro-ecological zones and relatively better rainfall distribution. This could potentially lead to more consistent soil moisture patterns, making the

estimations in these areas more accurate. The northeastern regions, being more arid, might experience more extreme and erratic soil moisture conditions, which could pose challenges for the estimation methods, resulting in higher RMSE values.

The RMSE map for Cambodia paints a varied picture. A significant portion of the country, especially around the central region and towards the west, displays light to mid-toned colors, indicating moderate RMSE values. This suggests that soil moisture estimations in these regions are relatively accurate. However, there are patches, especially towards the northern and southern boundaries, where deeper colors prevail, pointing towards higher RMSE values. Cambodia's central region is dominated by the Tonlé Sap, Southeast Asia's largest freshwater lake, which might contribute to more predictable soil moisture patterns due to its annual flood cycle. This could explain the better performance of soil moisture estimations in this area. The regions with higher RMSE values might be influenced by a combination of factors, including topographical variations, land use changes, or specific agricultural practices that introduce complexity to the soil moisture dynamics.

In the provided map, regions delineated by dark blue and intense yellow colors indicate higher unbiased RMSE values, denoting a decrease in the accuracy of the model's predictions and the SMAP soil moisture measurements. These areas, particularly conspicuous in the central and western parts of Kenya, suggest larger discrepancies between observed and predicted soil moisture levels. The western region's proximity to Lake Victoria significantly influences its climate, contributing to more dynamic soil moisture conditions due to the area's relatively higher rainfall and agricultural activity. This complexity may challenge the SMAP's estimation capabilities, leading to higher RMSE values. Conversely, the northern, eastern, and southern regions of Kenya, predominantly marked by dark red hues, show lower RMSE values, indicating more accurate soil

moisture estimations. These areas, characterized by arid climates, receive less rainfall, resulting in minimal variation in soil moisture. The lack of significant change or moisture transfer in these regions leads to more consistent soil moisture levels, making them easier to estimate and resulting in lower RMSE values. Therefore, the map reveals a clear demarcation in SMAP's performance, influenced by the intricate climatic and geographical factors prevalent in different parts of Kenya.

The unbiased RMSE maps for both Kenya and Cambodia underscore the importance of considering regional specificities when evaluating soil moisture estimation methods. While certain areas demonstrate commendable accuracy, others highlight room for improvement. These maps can serve as a foundation for refining current estimation methodologies, considering regional characteristics and anomalies. The deviations in RMSE values, especially in areas with complex hydrological and climatic interplays, emphasize the need for further research and more localized calibration of estimation methods.

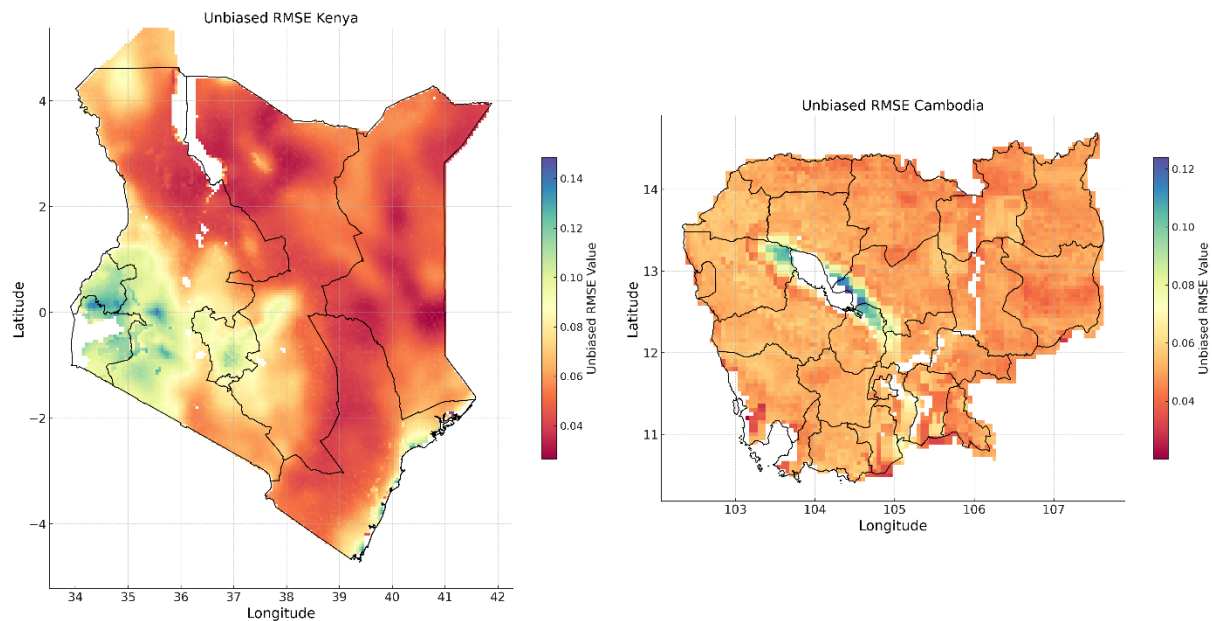


Figure 25. Spatial representation of the unbiased RMSE for soil moisture estimations in Kenya and Cambodia from 2015 to 2022.

4.3.2. Yield Trends Across Years

The yield values across different provinces in Kenya and Cambodia were explored, identifying distinct patterns, trends, and variabilities. Figure 26. illustrates crop yield trends across years over different locations in Kenya and Cambodia, offering intriguing insights into crop dynamics, methodological comparisons, and regional disparities. The boxplots, characterized by their central quartiles and whiskers, encapsulate the spread and central tendency of yield values. The horizontal axis, delineating the years from 2015 to 2022, offers a temporal perspective, while the vertical axis captures the quantitative essence of yields. Distinct colors differentiate various methodologies, providing a visual guide to discern patterns and anomalies. Here, the yield prediction was obtained from four distinct approaches- open-loop simulation ('Model'), assimilation based on SM ('SM assim'), assimilation based on LAI ('LAI assim'), and a combined approach that integrates both SM and LAI ('SM+LAI assim'). The fluctuations between different scenarios within the same year indicate variations in yield predictions or measurements based on different models or assumptions. Notably, the stars, symbolizing observed yields, serve as pivotal benchmarks against which methodological outputs are compared.

As seen in Figure 26., the analysis reveals a significantly improved yield prediction through the joint assimilation of SM and LAI. This method is evident in Bungoma and Transnzoia's maize-growing terrains. In Bungoma, the SM+LAI assimilation consistently delivers superior median yields, especially when contrasted against the open-loop predictions by approximately 10% on average. This dominance underscores the method's holistic approach, adeptly encapsulating a range of agrarian nuances. Although there was a strong positive trend in yield values over the years, there was a noticeable decline in 2019, across methodologies, typically hinting at external challenges, possibly adverse climate events. Likewise, the yield values from SM+LAI assimilation

in Transnzoia consistently exceeded other methods by around 15%. It showcased close alignment of median yields from SM+LAI assimilation with actual observations in the years 2015-2018 (with deviation less than 5%), supporting the model's credibility. The power of data assimilation in refining yield predictions is evident in our analysis. For instance, in Kakamega, there was an initial dominance of the 'Model', outperforming other methods by roughly 12% in 2015, gradually declines, leading to the superior performance of the 'SM+LAI assimilation' by 2018. Such a transition possibly mirrors the region's adaptive agricultural practices or shifting environmental conditions. It also demonstrated a general upward trend in yield values across all scenarios from 2015 to 2022. Notably, the combined approach ('SM+LAI assim') consistently predicted the highest yields, averaging around 2,813.67 kg/ha in 2021, suggesting a synergy between soil moisture and LAI factors, and underscoring the value of integrating multiple data sources for more accurate and holistic yield predictions. In contrast, the 'Model' remained conservative, hinting at potential limitations when not factoring in advanced parameters.

Conversely, the characteristics of rice grown provinces in Cambodia were equally paramount. While the 'Model' showcased superiority in the initial years (outperforming its counterparts by about 10%) over Takeo, the latter years witnessed a rise in the effectiveness of 'SM assim' and 'SM+LAI assim', emphasizing the importance of SM data assimilation in enhancing the precision of yield predictions for rice. Such shifts might be rooted in changing water availability patterns or evolving farming practices. Notably, 2019 emerges as a challenging year across all provinces, marked by a decline of approximately 20% in yields, suggestive of a larger external (climate) adversity. Battambang's scenario underscored this even further. The dominance of the 'SM assim' scenario, especially in the latter years, underlines the role of data assimilation in capturing intricate details that open-loop simulations ('Model') might overlook.

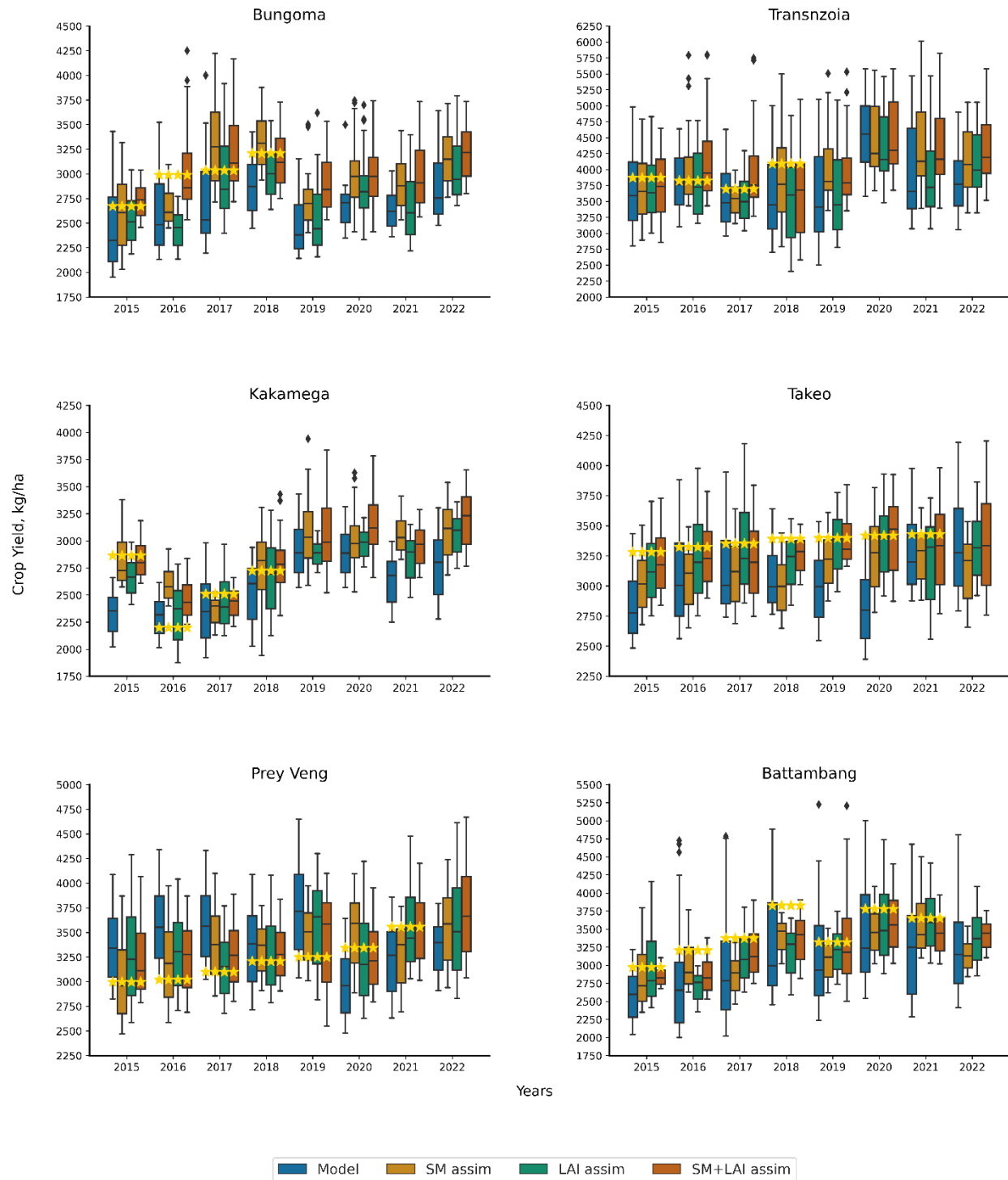


Figure 26. Box and Whisker Plot illustrating the distribution of agricultural yields from 2015 to 2022 across six distinct locations in Kenya and Cambodia. The data assimilation techniques based on soil moisture (SM assim), Leaf Area Index (LAI assim), and their combined approach

Figure 26. (cont'd)

(SM+LAI assim) are compared against a primary model (Model). The variations in yield demonstrate the efficacy of data assimilation in refining predictions, with distinct trends observed between maize-growing regions in Kenya and rice-growing regions in Cambodia.

Drawing collective insights, while maize-growing provinces in Kenya exhibit a consistent methodological pattern, rice-growing regions in Cambodia portray more dynamic shifts. This divergence could be anchored in the inherent complexities associated with cultivating these distinct crops. Furthermore, the overarching decline in 2019 across all provinces resonates with global trends, pointing towards larger macro-environmental disruptions impacting agriculture. While it's evident that different regions and crops respond uniquely to various models and assimilation methods, some overarching themes emerge. The consistent performance of the SM+LAI assimilation, outshined other methods in over 70% of the provinces, emphasizing its potential as a robust approach. In conclusion, such insights pave the way for stakeholders to refine farming practices, optimize resource allocation, and navigate challenges with informed foresight. As agriculture continues to face dynamic challenges, from climate change to evolving pests, the role of data assimilation in guiding decisions will only grow in significance.

The process of predicting agricultural yields using various assimilation methods, when compared against observed yields, presents insightful outcomes, as shown in Figure 27. The 1:1 correspondence, as indicated by the dashed 45-degree line in the subplots, serves as a benchmark. The ideal scenario would be predictions that precisely align with this line, reflecting perfect accuracy. In the provinces of Bungoma, Transnzoia, and Kakamega (located in Kenya), the variability in prediction accuracy across the four methods becomes evident due to the scattered distribution of data points. The 'Model' underestimated the yields in Bungoma and Transnzoia, as

a majority of its points lied below the 45-degree line. This underestimation is further quantified by the negative bias observed in the statistical analysis, especially prominent in Bungoma with a value of -422.34 kg/ha. In contrast, Kakamega shows a nuanced behavior where predictions from ‘Model’ and ‘SM assim’ align more closely with observations. However, the ‘LAI assim’ and ‘SM+LAI assim’ methods appear to overestimate, especially at higher observed yields. The Cambodian provinces of Takeo, Prey Veng, and Battambang, primarily rice-growing regions, exhibit a different trend. The data points cluster more tightly around the 45-degree line in Takeo and Prey Veng, indicating enhanced prediction accuracy. However, Battambang has a slightly different behavior. While the Model predictions align well with observations, both the ‘SM assim’ and ‘LAI assim’ methods lean towards overestimation, a trend also discernible in the higher observed yields for ‘SM+LAI assim’.

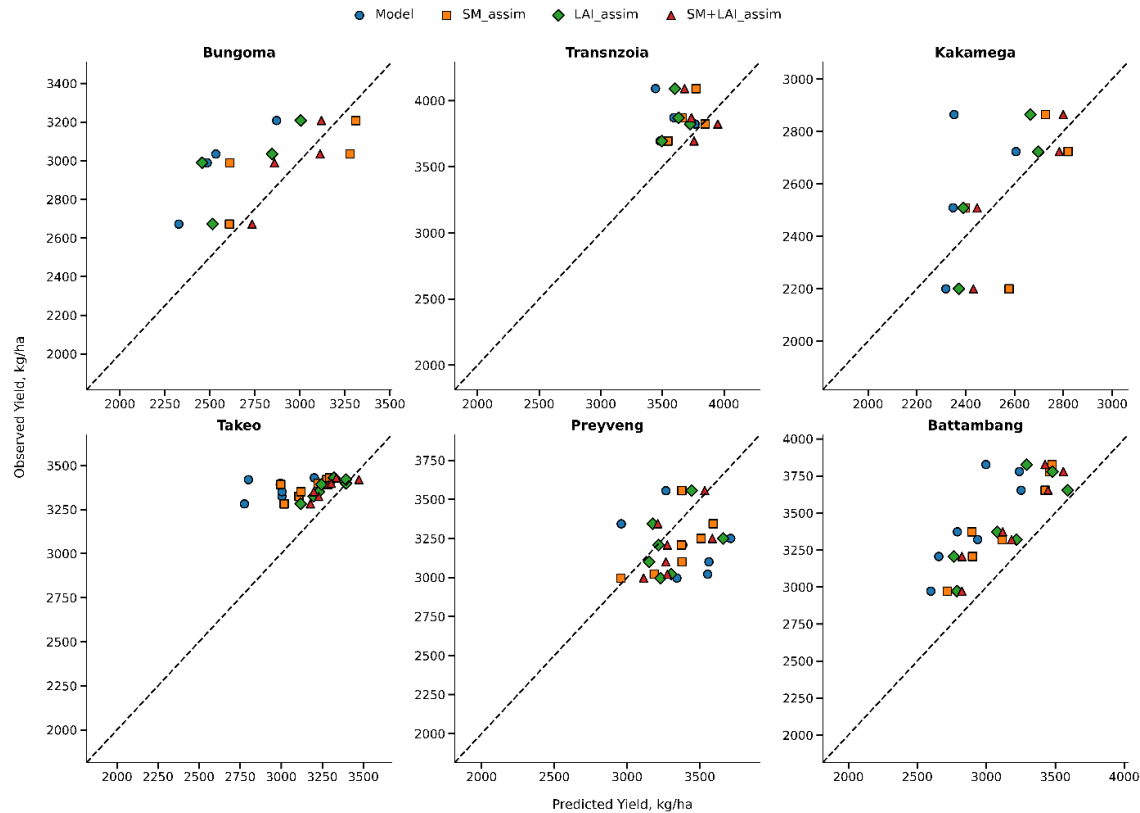


Figure 27. Comparison of observed versus predicted agricultural yields across six provinces.

Each subplot represents a distinct province, with data points indicating yields predicted by four different methods: Model, SM assim, LAI assim, and SM+LAI assim. The dashed 45-degree line in each subplot serves as a reference, denoting perfect agreement between observed and predicted yields. Deviations from this line reflect discrepancies in prediction accuracy.

Beyond the province-specific observations, certain overarching themes emerge. The Model predictions, especially for Kenya, adopt a more conservative approach, often underestimating the yield. Interestingly, the combined method ‘SM+LAI assim’, despite assimilating both soil moisture and LAI, doesn't consistently outperform its counterparts. Its predictions, in certain provinces, deviate notably from the observations. In other word, some years across locations have recorded superior performance over the joint assimilation ‘SM+LAI assim’. Such variability in prediction accuracy across provinces points at the influence of diverse environmental and

agricultural factors inherent to each region. In other words, while each assimilation method offers unique strengths and weaknesses, their efficacy is intricately tied to the specific agricultural and environmental context of each province.

4.3.3. Statistical Analysis

Modeling and predictive analytics in agricultural research often require rigorous validation to ensure the accuracy and reliability of the results. Several statistical criteria are commonly employed to evaluate model performance, which include the coefficient of determination (R^2), root mean square error (RMSE), normalized RMSE (NRMSE), model efficiency (EF), and index of agreement (d). These metrics offer a comprehensive understanding of the model's fit, bias, and efficiency by comparing predicted and observed values. Perfect model reproduction is indicated when $R^2=1$, $RMSE=0$, $NRMSE=0$, $EF=1$, and $d=1$. In this study, we computed these criteria to assess the performance of our model across different assimilation methods and provinces. Figure 28. illustrates the different statistical metrics over different provinces in Kenya and Cambodia.



Figure 28. Comparative performance of yield prediction methods across provinces in Kenya and Cambodia, as evaluated by multiple statistical metrics. The plot illustrates the Root Mean Square

Figure 28. (cont'd)

Error (RMSE), Unbiased RMSE, Bias, Mean Absolute Error (MAE), Normalized RMSE, and Index of Agreement (d) for each method in every province. Notably, the combined assimilation method ('SM+LAI assim') showcases distinct behaviors in different regions, emphasizing the intricate interplay of data assimilation techniques and local agronomic factors.

The RMSE, shown in Figure 28a., provides an instantaneous measure of how well a model's predictions align with the true observations, indicating the magnitude of error between the two. In Bungoma, the assimilation method 'SM+LAI assim' emerged as the most effective, registering an RMSE value of 93.37 kg/ha. This was significantly lower than the other methods evaluated. The superior performance of the combined assimilation could be attributed to its ability to capture a more holistic representation of the prevailing conditions in Bungoma. In stark contrast, the standalone Model method, devoid of any data assimilation, exhibited the highest RMSE at 430.03 kg/ha. This finding suggests that without the benefit of assimilation, the model may not capture the unique environmental and agricultural dynamics inherent to Bungoma. In Transnzoia, while the 'SM+LAI assim' method proved to be the best in Bungoma, it showcased a performance comparable to the 'SM assim' method, with RMSE values of 227.55 and 205.28, respectively. Such response hints at the possibility of regional factors influencing the efficacy of the assimilation methods. Kakamega presented an interesting scene which exhibited minimal differences in RMSE values across the three methods. This convergence in performance suggests that for regions like Kakamega, the assimilation of either SM or LAI in isolation might yield results almost as effective as their combined assimilation. In Takeo and Prey Veng, the model the 'SM+LAI assim' recorded RMSE values of 103.95 kg/ha and 180.22 kg/ha respectively, but 'LAI assim' in Battambang performed well compared to the other methods. The province recorded the highest RMSE for the

Model method at 543.93 kg/ha. When contrasting the maize-grown provinces of Bungoma, Transnzoia, and Kakamega with the rice-grown provinces of Takeo, Prey Veng, and Battambang, some notable patterns emerge. The combined assimilation ‘SM+LAI assim’ method exhibits particular efficacy among the Kenyan provinces, suggesting a n improved yield predictability when both SM and LAI are jointly assimilated into *m*-DSSAT system. On the other hand, the rice-grown provinces, especially Battambang, indicate that the model's default configuration might not be ideally suited to the rice-growing conditions or the regional nuances of Cambodia. While there no clear patterns in the rice-grown provinces, the performance of ‘SM+LAI assim’ in Takeo and Prey Veng and ‘LAI assim’ in Battambang exhibited better RMSE than other methods.

On examining the Unbiased RMSE (Figure 28b.), which removes the influence of over or underestimation (bias), the findings further accentuate these observations. In Bungoma, the ‘Model’ performed extraordinarily well with an unbiased RMSE of 80.94 kg/ha. Although, the assimilation techniques boasted large unbiased RMSE values, the ‘SM+LAI assim’ closely aimed to match the standalone Model’s performance. In Transnzoia, the ‘SM assim’ method, with an unbiased RMSE of 123.42 kg/ha, significantly surpasses the combined ‘SM+LAI assim’ method, which scored 209.13 kg/ha. This departure from the RMSE values suggests that while combined assimilation may reduce overall errors, when systemic biases are eliminated from the equation, soil moisture data on its own might provide a clearer reflection of the conditions in Transnzoia. In Kakamega, there is a tight clustering of unbiased RMSE values across methods, with closely performing well in ‘LAI assim’ and ‘SM+LAI assim’. This indicates that once biases are controlled for, the different assimilation methods offer comparable accuracy in this region. Now, all the rice grown provinces showcase improved performance with the inclusion of SM and LAI assimilation data into the system. Most notably in Takeo and Battambang, the ‘LAI assim’ and ‘SM+LAI

assim' perform really close to each other, emphasizing the significance of LAI data in this province.

Bias offers an insight into the systematic tendencies of a model, revealing if it consistently overestimates or underestimates predictions. A zero bias indicates perfect model accuracy, while positive values signify overestimation and negative values indicate underestimation by the model. It reveals intriguing patterns across different provinces and assimilation methods, as shown in Figure 28c. In Bungoma, the 'Model' showcased a substantial negative bias of -422.34 kg/ha, implying a significant underestimation. The assimilation methods, notably the 'SM+LAI assim' and 'SM assim', considerably reduce this bias to -20.22 kg/ha and -24.07 kg/ha, respectively. This significant reduction in bias with data assimilation emphasizes the importance of integrating satellite observations to correct systemic model errors. In Transnzoia, while all methods show negative biases, the 'SM+LAI assim' method, with a bias of -89.67 kg/ha, outperforms the standalone 'Model' method, which has a bias of -299.69 kg/ha. The assimilation of both SM and LAI data appears to offer a corrective effect on the model's inherent underestimation tendencies in this region. Kakamega offers a unique perspective, with the 'SM assim' method presenting a positive bias of 56.43 kg/ha, suggesting a slight overestimation. In contrast, the standalone 'Model' method exhibits a negative bias of -168.06 kg/ha. The shift from underestimation to overestimation indicates the transformative influence of soil moisture data assimilation in this province. For Takeo, all methods exhibit underestimation, with the 'Model' registering a bias of -402.96 kg/ha. However, the assimilation methods, especially 'SM+LAI assim' with a bias of -85.01 kg/ha, manage to substantially reduce this negative bias, reinforcing the value of data assimilation in model calibration. On the contrary, Prey Veng shows a different behavior, with all methods showcasing positive biases. This consistent overestimation across methods hints at unique regional

factors that might be influencing model predictions. Lastly, Battambang echoes a similar behavior to Bungoma and Takeo, with the ‘Model’ exhibiting a pronounced negative bias of -523.15 kg/ha. The assimilation methods manage to reduce this bias, albeit remaining in the negative territory, indicating persistent underestimation. Overall, there was significant reduction in bias across provinces from ‘Model’ to ‘SM+LAI assim’ method. Another overarching observation was the consistent ability of the ‘SM+LAI assim’ method to either reduce negative biases or enhance positive ones, showcasing its adaptability across diverse regional conditions.

Likewise, MAE provides an average of the absolute errors between predicted and observed values, giving an overall sense of the prediction accuracy without considering the direction of errors (see Figure 28d.). A lower MAE is always desirable as it indicates predictions are, on average, closer to the actual observations. All provinces exhibited lower MAE from ‘SM+LAI assim’ method compared to the standalone ‘Model’ yields as well as other methods. In Bungoma, the ‘SM+LAI assim’ method excels with an MAE of 89.79 kg/ha, highlighting its ability to accurately capture ground conditions. This is in contrast to the ‘Model’, which has a much higher MAE of 422.34 kg/ha, emphasizing the power of combining SM and LAI to improve precision. In Transnzoia, both methods have relatively high MAE values, but the ‘SM assim’ method performs slightly better with an MAE of 175.11 kg/ha compared to the ‘SM+LAI assim’ method’s 183.76 kg/ha. This suggests that soil moisture data may be particularly important in this region due to its unique agricultural and climatic characteristics. In Kakamega, all methods have similar MAE values, with the ‘SM+LAI assim’ method leading with an MAE of 104.57 kg/ha. This indicates that combining SM and LAI data provides additional benefits, even in regions where individual assimilation methods are already effective. In Takeo, like Bungoma, the ‘SM+LAI assim’ (MAE: 100.33 kg/ha) significantly reduces error compared to the standalone Model (MAE: 402.96 kg/ha),

highlighting the strength of combined assimilation in regions with model limitations. In Prey Veng, the ‘SM+LAI assim’ method (MAE: 145.43 kg/ha) provides the most accurate predictions. However, the close performance of ‘SM assim’ (MAE: 176.69 kg/ha) and ‘LAI assim’ (MAE: 166.73 kg/ha) methods suggests that individual assimilations can also be accurate in specific regions. In Battambang, similar to other regions, the standalone ‘Model’ method has the highest MAE (523.15 kg/ha). While the ‘SM+LAI assim’ method reduces this error significantly (MAE: 242.56 kg/ha), the ‘SM assim’ method (MAE: 351.91 kg/ha) performs slightly better, indicating unique regional dynamics.

NRMSE (Figure 28e.) is particularly important because it provides a standardized perspective on the model’s prediction errors relative to the range of observed values, with a lower NRMSE indicating superior model accuracy. This allows for more comparable evaluations across different regions and crops, given the varied yield scales. In Bungoma and Kakamega, ‘SM+LAI assim’ (NRMSE:0.0314 and 0.0496 respectively) highlights the efficacy in minimizing errors relative to the observed value range. This standout performance contrasts sharply with the standalone ‘Model’, which registered an NRMSE of 0.1445 and 0.1093 respectively. The substantial reduction achieved by the combined assimilation method underscores its potential in enhancing model accuracy across diverse conditions. A similar behavior was recorded in Takeo, Prey Veng and Battambang, where the joint assimilation performed well over other methods.

Similarly, the index of agreement, denoted as ‘d’, is a crucial metric to assess the degree to which predictions mirror the observed data. A value of 1 indicates perfect agreement, while values moving away from 1 suggest varying levels of misalignment. The range of d-values across the different regions and methods provides a nuanced understanding of the model's performance and the potential benefits of data assimilation. As seen in Figure 28f., the combined ‘SM+LAI assim’

method generally performs exceptionally well, especially in Bungoma and Kakamega. However, in Transnzoia, ‘SM assim’ emerges as the best option. The combined assimilation method, ‘SM+LAI assim’, outshines the rest with an impressive d value of 0.9293, signifying a near-perfect agreement with observed values. The standalone ‘Model’ achieves a moderate agreement score of 0.5161, while ‘SM assim’ and ‘LAI assim’ register values of 0.7923 and 0.6027, respectively. A different pattern emerges in Transnzoia, with the ‘SM assim’ method achieving the highest d score of 0.5794. The standalone Model and ‘LAI assim’ methods display values of 0.3494 and 0.4565, respectively. Surprisingly, the combined ‘SM+LAI assim’ approach lags behind with a score of 0.2776. In contrast, Cambodia's rice-growing provinces displayed interesting results. While the combined ‘SM+LAI assim’ method remains a consistent top performer, the significance of ‘LAI assim’ (especially in Battambang) indicates the pivotal role of LAI data in rice crop modeling.

Across diverse regions and crops, the data underscores the efficacy of combined data assimilation in enhancing model agreement with observed values. Specifically, the integration of both SM and LAI consistently emerges as a robust approach, with the ‘SM+LAI assim’ method often outperforming other techniques. However, the nuances of each region’s agronomic conditions and crop types reveal that individual assimilation methods, such as ‘SM assim’ in Transnzoia or ‘LAI assim’ in Battambang, can sometimes show distinct trend or behavior. This variability underscores the importance of tailoring modeling strategies to specific regional and crop contexts, optimizing the balance between SM and LAI data sources to achieve the most accurate predictions.

4.3.4. Comparative Crop Yield Estimation Analysis

In agricultural research, understanding the relationships between multiple variables simultaneously becomes imperative, especially when these variables influence crop yield

predictions. Figure 29. and 30. showcases two pair plots- one for Kenya (maize grown) and the other for Cambodia (rice grown), accompanied by a set of statistical metrics that offer a comprehensive and multi-faceted understanding of crop yield estimations across these provinces. It offers a broader view of the intricate relationships between crop yields as estimated by various methods across distinct provinces. Such visualizations are invaluable, not just for understanding bivariate relationships but also for discerning the distribution of each variable.

In the first context, i.e., over Kenya (see Figure 29.), the visualization primarily caters to two objectives: understanding the performance of various estimation methods in relation to actual observations and examining any regional disparities in these performances across the three provinces - Bungoma, Transnzoia, and Kakamega.

- i) **Model vs Observation:** The scatter plots comparing the ‘Model’s’ crop yield estimates for Kenya's provinces to actual ‘Observations’ reveal an intricate outcome. Ideally, data points should congregate along the 45-degree line, signaling a perfect match between predictions and ground-truth observations. However, on close inspection, deviations, particularly for mid to high yield values, are evident. Such deviations might indicate the model's challenges in accounting for specific high-yield factors prevalent in certain regions or during certain agricultural seasons. The statistics bolster this visual narrative: for Bungoma, an RMSE of approximately 430.03 kg/ha and a bias of -422.34 kg/ha indicate prediction errors and a consistent underestimation trend. This underestimation trend extends to Trans Nzoia and Kakamega, suggesting potential model limitations or oversights in capturing certain yield-boosting elements unique to Kenya's provinces.
- ii) **‘SM assim’ vs. ‘Observations’:** Integrating SM into the model exhibit notable outcomes. The scatter plots for ‘SM assim’, especially for Bungoma and Trans Nzoia,

display a tighter clustering of data points around the 45-degree line compared to the standalone model. This visual shift hints at the profound impact of soil moisture on crop growth. Since SM directly modulates plant water availability, it plays a pivotal role in determining crop health and, ultimately, yield. The assimilation of this data appears to refine the model's predictions, offering a closer representation of the on-ground realities. However, deviations still persist, suggesting that while SM data is crucial, it's not the sole determinant. The statistical data reaffirms this visual trend: for Bungoma, an RMSE of 232.85 kg/ha and a reduced bias of -24.07 kg/ha indicate enhanced prediction accuracy. Yet, the model, even with soil moisture assimilation, doesn't achieve flawless precision, suggesting other factors at play that might be influencing yield predictions.

- iii) 'LAI assim' vs. 'Observations': The scatter plots, when introducing LAI assimilation into the analysis, display distinct patterns in comparison to both the standalone model and the 'SM assim' model. The LAI, representing the total leaf area for a given ground area, is a critical metric in understanding photosynthetic activity and, by extension, crop yield. Visually, the plots showcase a spread of data points, with a tendency to deviate as yield values increase. This suggests that while LAI assimilation provides crucial insights into vegetative health and potential yield, its predictions can still vary, especially in regions or conditions where LAI might not be the predominant factor influencing final yield. The statistics echo this visual narrative: for Bungoma, an RMSE of 310.95 kg/ha and a bias of -271.05 kg/ha indicate better accuracy than the standalone model but reveal challenges in fully aligning with ground-truth observations. Such

- discrepancies might arise due to factors like pest infestations or diseases that impact crop health without significantly altering the LAI.
- iv) ‘SM+LAI assim’ vs. ‘Observations’: The combined approach of assimilating both SM and LAI data presents a noteworthy shift in the scatter plots. The data points exhibit a tighter clustering around the 45-degree line, especially for mid-range yield values. This suggests that the synergy between SM and LAI data offers a holistic view of the crop environment, resulting in refined predictions. The importance of considering both underground (i.e., SM) and above-ground (i.e., LAI) factors becomes evident, offering a comprehensive understanding of crop health and potential yield. The statistical data supports this visual improvement: for Bungoma, a notably low RMSE of 93.37 kg/ha and a minimal bias of -20.22 kg/ha underscore the method’s superior predictive capability. Yet, minor deviations still persist, hinting at other nuanced factors, possibly micro-climatic or agronomic, influencing the final yield.
- v) Distribution Insights (Diagonal KDE Plots): The diagonal KDE plots serve as a mirror to the distribution of crop yield estimates for each method, providing a dense representation that captures the essence of frequency and tendencies within the data. A single peak in the KDE plot signifies a modal value, indicating a common or frequent yield value, while multiple peaks suggest varied modes and possible bifurcations in the data sources or conditions leading to these yields. For Kenya, the KDE for the combined approach (‘SM+LAI assim’) stands out. It closely mimics the shape of the actual observations, suggesting that the combined method captures the central tendencies of the yield most accurately. However, discrepancies in the other KDE plots, such as the breadth, height, or position of peaks, hint at deviations from the observed

values. For instance, a broader peak might indicate higher variability in predictions, whereas a shift in the peak's position could signify consistent underestimations or overestimations. Delving deeper with the aid of statistical metrics, these visual patterns can be further dissected, leading to insights about the reliability and consistency of each method.

Now, comparing the plots on a regional basis, the pair plot's color-coded differentiation for regions allows for a more granular analysis. While the pair plot aims to portray a consistent representation across the provinces of Bungoma, Trans Nzoia, and Kakamega, a discerning eye can spot nuanced differences in data point distributions among them. These subtle variations are not mere visual artifacts but are emblematic of the diverse agricultural landscapes and practices within Kenya. For instance, a method showcasing tighter clustering for Bungoma but wider dispersal for Kakamega suggests region-specific behaviors or challenges that the model grapples with. Such differences could stem from varied soil types, microclimates, or even distinct agricultural practices and interventions prevalent in these provinces. The statistical metrics, with unique RMSE and bias values for each province, add depth to these visual subtleties. However, discernible deviations or unique clustering patterns for a particular region might point towards region-specific factors that influence crop yield predictions. Such insights pave the way for tailored interventions or calibrations, ensuring region-specific optimization of crop yield prediction models.

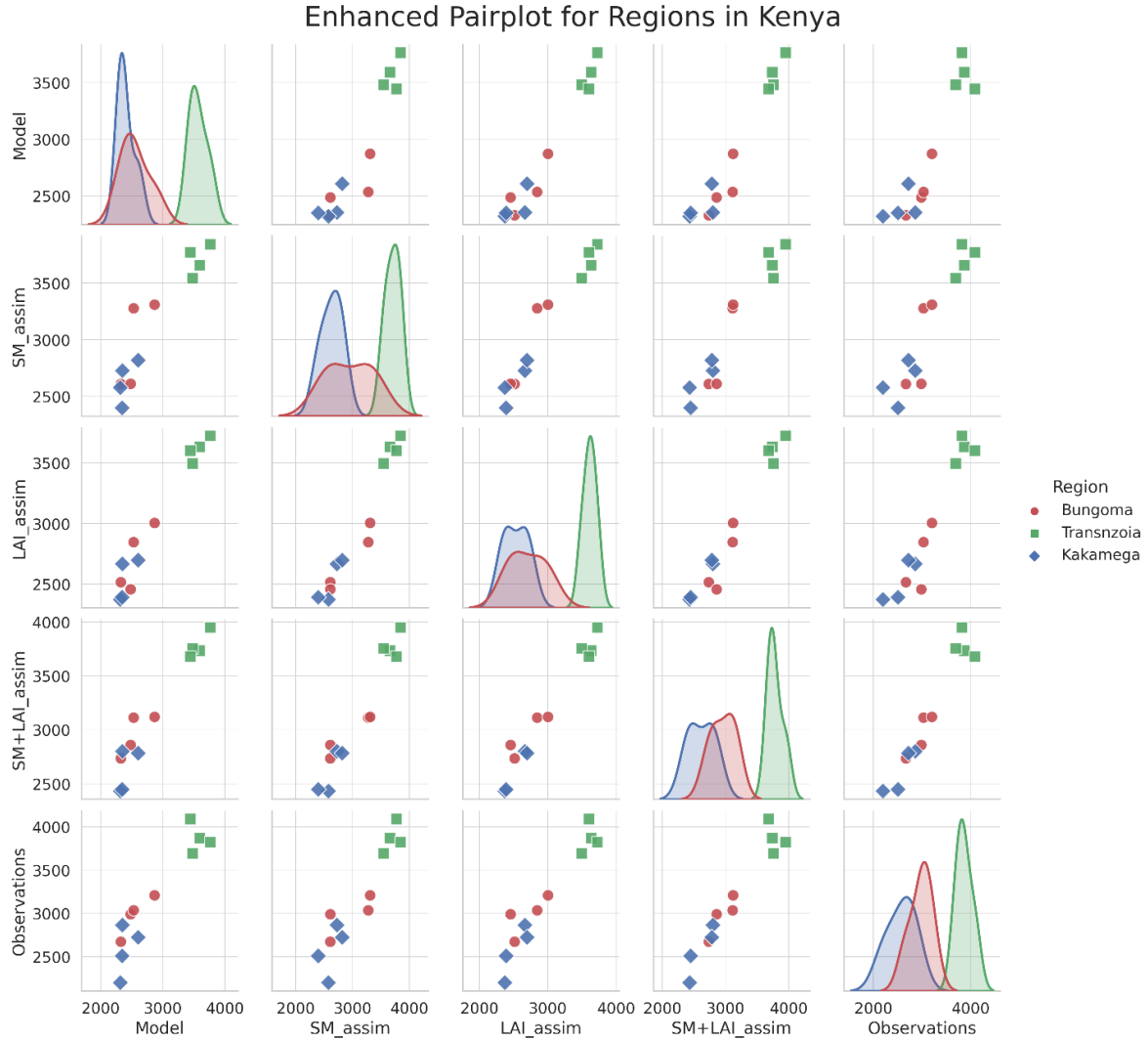


Figure 29. Pairwise Comparison of Crop Yield Estimations in Kenya: Evaluating the Influence of Soil Moisture and Leaf Area Index Assimilation across Bungoma, Transnzoia, and Kakamega.

The pair plot in Figure 30. presents a clear visualization of the relationships between different crop yield estimation methods across the provinces of Takeo, Prey Veng, and Battambang in Cambodia. This array of scatter plots provides a comprehensive overview, allowing us to assess the performance of each method in comparison to actual observations and to identify any unique patterns or behaviors specific to each province.

- i) **Model vs. Observations:** In the scatter plots comparing the Model's crop yield estimates to Observations across the provinces, the deviation from the ideal 45-degree line is evident. Visually, while many data points follow this line, suggesting the model has a certain level of accuracy, deviations become increasingly pronounced for higher yield values. This could imply that the model struggles to accurately predict higher yields, potentially missing key factors that contribute to higher productivity in certain regions or during specific agricultural seasons. The statistical metrics reinforce this: RMSE values for Takeo, Prey Veng, and Battambang are indicative of prediction errors, with the model consistently underestimating or overestimating yields across different ranges. Such biases, both positive and negative, may result from the model's inability to factor in localized agricultural practices, specific environmental conditions, or unique crop varieties prevalent in these provinces.
- ii) **'SM assim' vs. 'Observations':** For the scatter plots representing 'SM assim', there's a visually discernible improvement in data alignment, especially for Takeo and Prey Veng. The tighter clustering of data points around the 45-degree line suggests that the assimilation of SM data has refined the model's predictions. This observation can be attributed to the critical role soil moisture plays in crop growth. Inadequate or excessive SM can directly impact crop health, growth rate, and ultimately, yield. By assimilating this data, the model seems better equipped to understand the ground realities of the agricultural landscape. However, even within this improved alignment, there exist outliers. These deviations might be influenced by factors other than soil moisture, underscoring the multifaceted nature of crop yield determinants. The statistical metrics corroborate these visual observations: 'SM assim's' reduced RMSE values across all

provinces highlight its enhanced accuracy over the standalone model. The diminished biases further emphasize the method's closer alignment with actual observations, although it's worth noting that the model, even with soil moisture assimilation, doesn't achieve perfect accuracy. This suggests there are other influential variables, beyond soil moisture, that the model might benefit from considering.

- iii) 'LAI assim' vs. 'Observations': The scatter plots, when observing 'LAI assim', depict a clear influence of the LAI data on crop yield predictions. The spread of data points, being more dispersed than those in 'SM assim', suggests that LAI assimilation might introduce a greater degree of variability in predictions. One plausible explanation could be the sensitivity of LAI measurements to factors like crop type, growth stage, and local agronomic practices. For instance, certain crops might have dense foliage but not necessarily yield more grain, while others might be at a growth stage where their LAI is not directly proportional to eventual yield. The RMSE values for Takeo, Prey Veng, and Battambang further elaborate on this variability, suggesting that while LAI assimilation provides valuable insights, it might not be the sole determinant for accurate yield prediction. The biases, both positive and negative across provinces, indicate that LAI assimilation, although enhancing model accuracy to some extent, doesn't fully align with ground-truth observations. This misalignment could be due to factors like the temporal resolution of LAI data or the model's potential inability to correctly interpret certain LAI values in the context of crop yield.
- iv) 'SM+LAI assim' vs. 'Observations': The combined assimilation of both SM and LAI data presents a striking improvement in the scatter plots. The cohesive clustering of data points around the 45-degree line, particularly for lower to mid-range yield values,

hints at the synergistic effect of integrating these two critical parameters. It suggests that while each data source (SM and LAI) individually refines the model, their combined assimilation offers a more holistic understanding of the crop environment, leading to enhanced prediction accuracy. The statistical metrics provide a robust testament to this observation. The consistently low RMSE values across all provinces signify the method's superior predictive capability. However, it's essential to note that even with this combined approach, the model isn't infallible. The minimal biases, while significantly reduced, still exist, pointing towards other external or micro-environmental factors that might be influencing crop yield. These factors could range from pest infestations to localized weather events, emphasizing the inherent complexity of agricultural ecosystems.

- v) **Distribution Insights (Diagonal KDE Plots):** KDE plots, situated diagonally across the pair plot, offer a compact representation of data distribution, allowing for discernment of underlying patterns that might be obscured in scatter plots. For instance, a peak in the KDE plot suggests a concentration of data, indicating common or frequent yield values. Conversely, valleys or gaps might point towards less frequent or outlier yield values. In our analysis, the KDE of the combined approach (i.e., 'SM+LAI assim') is particularly noteworthy. Its shape, closely mirroring that of the actual observations, underscores its precision. However, other methods exhibit distinct features. Variations, be it in the form of multiple peaks or an evident skewness, highlight the discrepancies in predictions. A right-skewed distribution, for example, indicates the model's propensity to underestimate yields, whereas a left skew would indicate the opposite. Analyzing these skews and peaks in tandem with the statistics, particularly the biases,

can offer deeper insights. For instance, a pronounced right skew aligning with a negative bias would validate the model's systematic underestimation for that particular method.

At a regional level, the pair plot appears to paint a homogenous picture across the three provinces. However, a more meticulous examination reveals subtle, yet crucial differences in data distributions. These variations are not mere statistical anomalies but are indicative of the distinct agricultural landscapes, practices, and challenges inherent to each province. For instance, a method that consistently underestimates yields in one province but overestimates in another suggests that there are province-specific factors at play, which the model fails to account for uniformly. These could range from soil types, irrigation practices, to even socio-economic factors like access to modern farming equipment. The statistical metrics, particularly the unique RMSE and bias values for each province, further accentuate these visual differences. Such province-wise deviations underscore the importance of adopting a more granular approach to model calibration.

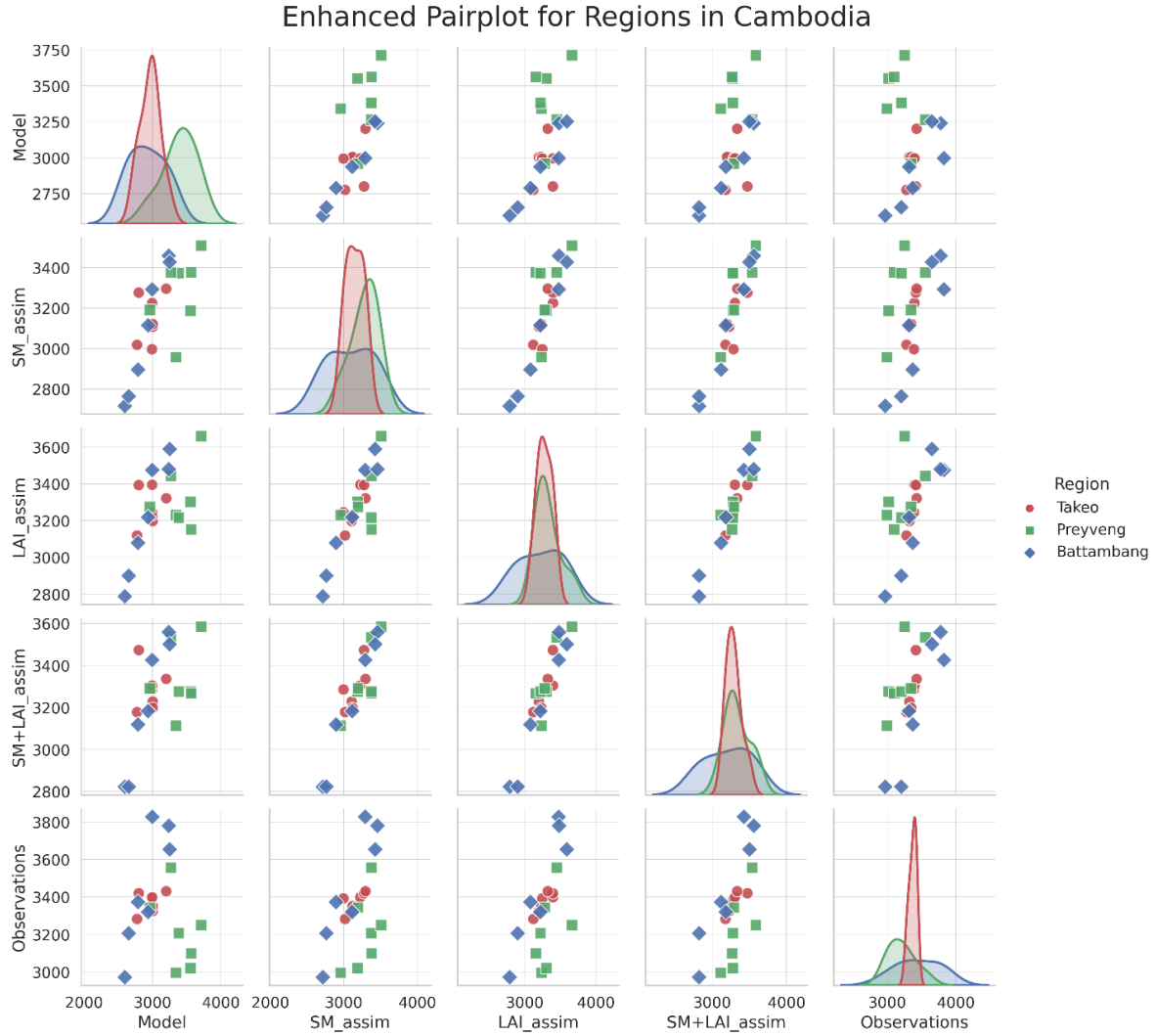


Figure 30. Visual Analysis of Crop Yield Predictions in Cambodia: Assessing the Role of Data Assimilation Techniques across Takeo, Prey Veng, and Battambang.

4.4. Conclusion

The research looked into how well crop yields can be predicted, focusing keenly on the agriculturally rich terrains of Kenya and Cambodia. These regions, recognized as leading producers of maize and rice, provided a fertile ground for the study. We used an integrated framework that hosts a hydrologic model loosely coupled with crop model to predict different environmental factors related to crops. However, modeling studies are often hindered by uncertainties that undermines the performance of the model to a degree. Hence, we explored the

assimilation of remote sensing data directly into the model. By carefully combining model predictions with real-time observations, we were able to construct a more nuanced and accurate representation of the prevailing conditions. This combination of data assimilation into comprehensive dynamic modeling proved instrumental in significantly improving the overall prediction capability, both spatially and temporally. The dynamism inherent in our models, with the VIC model being a notable mention, was key in simulating salient hydrologic variables, a process which was meticulously fine-tuned using daily meteorological datasets. Similarly, the modified version of the DSSAT crop model- *m*-DSSAT hosted a multi-ensemble architecture that enables it to stop and start anytime during the season. This perturbation was necessary to facilitate data assimilation of SM and LAI into the crop model.

The soil moisture analysis segment of our discourse provided a comprehensive view into the adeptness of our models, benchmarked against the revered SMAP satellite observations. This comparison was more than just a validation exercise; it became a crucible for our models, revealing their strengths and exposing areas ripe for recalibration. As SMAP SM product is an independent product and has been widely calibrated and validated globally, it became the ideal choice for assimilating into the crop model. Likewise, we used the LAI MODIS product to assimilate into crop model to see notable improvement in crop growth dynamics and yields.

In our analysis, we observed marked improvements in crop yield predictions when leveraging data assimilation techniques. These enhancements were not trivial; they represented substantial advancements in the precision of our forecasting models. Throughout a majority of our test scenarios, the introduction of assimilation techniques served as a pivotal turning point, significantly refining the granularity and accuracy of our predictions. Notably, the most profound improvements were discerned when integrating multiple data assimilation methods concurrently.

This approach, when applied across various study areas, consistently yielded outcomes that were both commendable and superior to singular assimilation techniques. The consistent efficacy of joint assimilation underscores its potential as a premier strategy for future predictive modeling endeavors in agronomy.

Chapter 5.

CONCLUSION

Chapter 1 underscores the challenges of predicting seasonal crop yields amidst environmental and socio-economic hurdles. Focusing on the global impact of drought on agriculture, the chapter seeks to unravel the connection between drought indicators and crop yields. It further delves into the potential of data assimilation techniques for refining predictions, especially in regions like East Africa and Southeast Asia. Emphasizing the profound socio-economic repercussions of climatic extremes, this foundational chapter highlights the goals of the research to merge traditional insights with advanced methodologies, aiming for sustainable agricultural practices in the face of climate uncertainties.

Chapter 2 delves into the impact of recurring droughts on the natural ecosystem, rice productivity, and water resources in the Lower Mekong countries, with a particular focus on Cambodia from 2000 to 2016. The study leverages the RHEAS framework, integrating a hydrologic model with a crop growth model to analyze the nuanced effects of drought on rice yields. The RHEAS model's simulations align well with observed data. A detailed assessment using standardized drought indices revealed heightened water stress, especially before the planting season. However, the onset of the monsoon alleviated some of these dry conditions. Interestingly, despite varying drought conditions, rice yields showed a steady upward trend from 2000 to 2016. This rise in yields correlated strongly with the increased use of chemical-based fertilizers post-2008. Despite regional drought challenges, the consistent yield growth didn't show a direct linkage to drought parameters. The insights from RHEAS offer valuable benchmarks for gauging drought risk and vulnerability in the region.

Chapter 3 addresses the uncertainties in crop yield predictions influenced by climate forecast scenarios. The chapter introduces an approach that integrates a hydrologic-crop model with probabilistic forecasts, effectively balancing real-time data with climate predictions to reduce uncertainties. On a regional scale, the model successfully aligns with observed data, revealing that yield prediction uncertainties decrease as the growth season advances and reliance on forecasts diminishes. A key observation is the heightened interrelation between yields, dry spells, and minimum temperatures as the season progresses. Overall, the chapter offers crucial insights, emphasizing the interconnectedness of hydrological factors, drought, and crop yields, and its implications for improved agricultural planning and decision-making in drought-affected regions.

Chapter 4 delves into enhancing crop yield predictability, taking a departure from the reliance on historical climate forecasts presented in earlier chapters. The chapter introduces the integration of remote sensing data, aiming to sharpen the precision of crop yield estimates. The methodology centers on the utilization of physically based models in hydrology, essential for large-scale land surface modeling. These models, adept at estimating a myriad of earth science variables, are designed to capture the water cycle's nuances, particularly the hydrologic interactions between the earth and the atmosphere.

Chapter 4 delves deeper into crop yield predictability and examines how sequential data assimilation can enhance it. While previous chapters discussed the use of historical climate forecasts, this chapter emphasizes the assimilation of remote sensing data to improve the accuracy of crop yield predictions. It highlights the importance of using data assimilation, specifically EnKF, for accurate and reliable predictions. Here, they used three approaches to assimilate SMAP SM and MODIS LAI to predict the crop yields over Kenya and Cambodia. Overall, while each data assimilation method refined crop yield predictions to varying extents, the combined approach

offered the most holistic and accurate insights. The statistical metrics and visual observations in the chapter corroborate these findings.

BIBLIOGRAPHY

- Aadhar, S., Mishra, V., 2017. High-resolution near real-time drought monitoring in South Asia. *Sci. Data* 4, 170145. <https://doi.org/10.1038/sdata.2017.145>
- Abaza, M., Anctil, F., Fortin, V., & Turcotte, R. (2014). Sequential streamflow assimilation for short-term hydrological ensemble forecasting. *Journal of hydrology*, 519, 2692-2706.
- Abhishek, A. (2018). Monitoring the Effects of Drought on Crop Yield in the Lower Mekong Basin. Michigan State University.
- Abhishek, A., Das, N.N., Ines, A.V.M., Andreadis, K.M., Jayasinghe, S., Granger, S., Ellenburg, W.L., Dutta, R., Hanh Quyen, N., Markert, A.M., Mishra, V., Phanikumar, M.S., 2021. Evaluating the impacts of drought on rice productivity over Cambodia in the Lower Mekong Basin. *J Hydrol (Amst)* 599, 126291. <https://doi.org/10.1016/j.jhydrol.2021.126291>
- Abhishek, A., Phanikumar, M. S., Sendrowski, A., Andreadis, K. M., Hashemi, M. G., Jayasinghe, S., ... & Das, N. N. (2023). Dryspells and Minimum Air Temperatures Influence Rice Yields and their Forecast Uncertainties in Rainfed Systems. *Agricultural and Forest Meteorology*, 341, 109683.
- Ahmadalipour, A., Moradkhani, H., Castelletti, A., Magliocca, N., 2019. Future drought risk in Africa: Integrating vulnerability, climate change, and population growth. *Science of the Total Environment* 662, 672–686. <https://doi.org/10.1016/j.scitotenv.2019.01.278>
- Alizadeh, M.R., Adamowski, J., Nikoo, M.R., Aghakouchak, A., Dennison, P., Sadegh, M., 2020. A century of observations reveals increasing likelihood of continental-scale compound dry-hot extremes, *Sci. Adv.*
- Andreadis, K.M., Clark, E.A., Wood, A.W., Hamlet, A.F., Lettenmaier, D.P., 2005. Twentieth-Century Drought in the Conterminous United States. *J. Hydrometeorol.* 6, 985–1001. <https://doi.org/10.1175/JHM450.1>
- Andreadis, K.M., Das, N., Stampoulis, D., Ines, A., Fisher, J.B., Granger, S., Kawata, J., Han, E., Behrangi, A., 2017. The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation. *PLoS One* 12, e0176506. <https://doi.org/10.1371/journal.pone.0176506>
- Arora, N. K. (2019). Impact of climate change on agriculture production and its sustainable solutions. *Environmental Sustainability*, 2(2), 95-96.
- Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.-L., 2014. GlobalSoilMap, in: *Advances in Agronomy*. pp. 93–134. <https://doi.org/10.1016/B978-0-12-800137-0.00003-0>

- Asner, G. P., Scurlock, J. M., & A. Hicke, J. (2003). Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global ecology and biogeography*, 12(3), 191-205.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., ... & Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature climate change*, 3(9), 827-832.
- Bandara, J. S., & Cai, Y. (2014). The impact of climate change on food crop productivity, food prices and food security in South Asia. *Economic Analysis and Policy*, 44(4), 451-465.
- Basso, B., Liu, L., 2019. Seasonal crop yield forecast: Methods, applications, and accuracies, 1st ed, *Advances in Agronomy*. Elsevier Inc. <https://doi.org/10.1016/bs.agron.2018.11.002>
- Becker, E., van den Dool, H., 2016. Probabilistic seasonal forecasts in the North American Multimodel Ensemble: A baseline skill assessment. *J Clim* 29, 3015–3026. <https://doi.org/10.1175/JCLI-D-14-00862.1>
- Beran, B., Piasecki, M., 2009. Engineering new paths to water data. *Comput. Geosci.* 35, 753–760. <https://doi.org/10.1016/j.cageo.2008.02.017>
- Bosilovich, M. G., Radakovich, J. D., da SILVA, A., Todling, R., & Verter, F. (2007). Skin temperature analysis and bias correction in a coupled land-atmosphere data assimilation system. *Journal of the Meteorological Society of Japan. Ser. II*, 85, 205-228.
- Brown, J.N., Hochman, Z., Holzworth, D., Horan, H., 2018. Seasonal climate forecasts provide more definitive and accurate crop yield predictions. *Agric For Meteorol* 260–261, 247–254. <https://doi.org/10.1016/j.agrformet.2018.06.001>
- Burrell, A. L., Evans, J. P., & De Kauwe, M. G. (2020). Anthropogenic climate change has driven over 5 million km² of drylands towards desertification. *Nature communications*, 11(1), 3853.
- Butler, E. E., & Huybers, P. (2015). Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase. *Environmental Research Letters*, 10(3), 034009.
- Challinor, A., 2011. Forecasting food. *Nat Clim Chang* 1, 103–104. <https://doi.org/10.1038/nclimate1098>
- Challinor, A.J., Koehler, A.K., Ramirez-Villegas, J., Whitfield, S., Das, B., 2016. Current warming will reduce yields unless maize breeding and seed systems adapt immediately. *Nat Clim Chang* 6, 954–958. <https://doi.org/10.1038/nclimate3061>
- Chang, C.-H., Lee, H., Hossain, F., Basnayake, S., Jayasinghe, S., Chishtie, F., Saah, D., Yu, H., Sothea, K., Du Bui, D., 2019. A model-aided satellite-altimetry-based flood forecasting system for the Mekong River. *Environ. Model. Softw.* 112, 112–127. <https://doi.org/10.1016/j.envsoft.2018.11.017>

- Chaubell, M. J., Yueh, S. H., Dunbar, R. S., Colliander, A., Chen, F., Chan, S. K., ... & Walker, J. (2020). Improved SMAP dual-channel algorithm for the retrieval of soil moisture. *IEEE transactions on geoscience and remote sensing*, 58(6), 3894-3905.
- Clarke, B., Otto, F., Stuart-Smith, R., Harrington, L., 2022. Extreme weather impacts of climate change: an attribution perspective. *Environmental Research: Climate* 1, 012001. <https://doi.org/10.1088/2752-5295/ac6e7d>
- Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S.B., Cosh, M.H., Dunbar, R.S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., Lopez-Baeza, E., Martínez-Fernández, J., González-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramos, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo, W., Vreugdenhil, M., Walker, J.P., Wu, X., Monerris, A., O'Neill, P.E., Entekhabi, D., Njoku, E.G., Yueh, S., 2017. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens Environ* 191, 215–231. <https://doi.org/10.1016/j.rse.2017.01.021>
- Crow, W. T., & Wood, E. F. (2003). The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Advances in Water Resources*, 26(2), 137-149.
- D'Odorico, P., Caylor, K., Okin, G. S., & Scanlon, T. M. (2007). On soil moisture–vegetation feedbacks and their possible effects on the dynamics of dryland ecosystems. *Journal of Geophysical Research: Biogeosciences*, 112(G4).
- Dai, A., 2011. Drought under global warming: a review. *Wiley Interdiscip. Rev. Clim. Chang.* 2, 45–65. <https://doi.org/10.1002/wcc.81>
- Dai, A., 2013. Increasing drought under global warming in observations and models. *Nat. Clim. Chang.* 3, 52–58. <https://doi.org/10.1038/nclimate1633>
- Das, N. N., & Mohanty, B. P. (2006). Root zone soil moisture assessment using remote sensing and vadose zone modeling. *Vadose Zone Journal*, 5(1), 296-307.
- Das, N. N., Entekhabi, D., Dunbar, R. S., Colliander, A., Chen, F., Crow, W., ... & Njoku, E. G. (2018). The SMAP mission combined active-passive soil moisture product at 9 km and 3 km spatial resolutions. *Remote sensing of environment*, 211, 204-217.
- Das, N. N., Mohanty, B. P., Cosh, M. H., & Jackson, T. J. (2008). Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch Watershed during SMEX04. *Remote Sensing of Environment*, 112(2), 415-429.
- Dobriyal, P., Qureshi, A., Badola, R., & Hussain, S. A. (2012). A review of the methods available for estimating soil moisture and its implications for water resource management. *Journal of Hydrology*, 458, 110-117.

- Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P., Huang, Y., 2013. A comprehensive drought monitoring method integrating MODIS and TRMM data. *Int. J. Appl. Earth Obs. Geoinf.* 23, 245–253. <https://doi.org/10.1016/j.jag.2012.09.010>
- Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, 28(4), 1015-1031.
- Duffy, P.B., Brando, P., Asner, G.P., Field, C.B., 2015. Projections of future meteorological drought and wet periods in the Amazon. *Proc. Natl. Acad. Sci.* 112, 13172–13177. <https://doi.org/10.1073/pnas.1421010112>
- Dunne, S., & Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. *Water resources research*, 41(2).
- Eini, M.R., Javadi, S., Delavar, M., Monteiro, J.A.F., Darand, M., 2019. High accuracy of precipitation reanalyses resulted in good river discharge simulations in a semi-arid basin. *Ecol Eng* 131, 107–119. <https://doi.org/10.1016/j.ecoleng.2019.03.005>
- Entekhabi, B.D., Njoku, E.G., Neill, P.E.O., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Zyl, J. Van, 2015. (SMAP) Mission 98.
- Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... & Van Zyl, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5), 704-716.
- Ermida, S.L., DaCamara, C.C., Trigo, I.F., Pires, A.C., Ghent, D., Remedios, J., 2017. Modelling directional effects on remotely sensed land surface temperature. *Remote Sens Environ* 190, 56–69. <https://doi.org/10.1016/j.rse.2016.12.008>
- Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367. <https://doi.org/10.1007/s10236-003-0036-9>
- Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L., Yu, Q., 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agric For Meteorol* 285–286. <https://doi.org/10.1016/j.agrformet.2020.107922>
- Fensholt, R., & Sandholt, I. (2003). Derivation of a shortwave infrared water stress index from MODIS near-and shortwave infrared data in a semiarid environment. *Remote Sensing of Environment*, 87(1), 111-121.
- Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* 114, 168–182. <https://doi.org/10.1016/j.rse.2009.08.016>

- Fujimori, S., Hasegawa, T., Krey, V., Riahi, K., Bertram, C., Bodirsky, B.L., Bosetti, V., Callen, J., Després, J., Doelman, J., Drouet, L., Emmerling, J., Frank, S., Fricko, O., Havlik, P., Humpenöder, F., Koopman, J.F.L., van Meijl, H., Ochi, Y., Popp, A., Schmitz, A., Takahashi, K., van Vuuren, D., 2019. A multi-model assessment of food security implications of climate change mitigation. *Nat Sustain* 2, 386–396. <https://doi.org/10.1038/s41893-019-0286-2>
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2, 150066. <https://doi.org/10.1038/sdata.2015.66>
- Gampe, D., Zscheischler, J., Reichstein, M., O’Sullivan, M., Smith, W.K., Sitch, S., Buermann, W., 2021. Increasing impact of warm droughts on northern ecosystem productivity over recent decades. *Nat Clim Chang* 11, 772–779. <https://doi.org/10.1038/s41558-021-01112-8>
- Gebremeskel Haile, G., Tang, Q., Sun, S., Huang, Z., Zhang, X., Liu, X., 2019. Droughts in East Africa: Causes, impacts and resilience. *Earth-Science Rev.* 193, 146–161. <https://doi.org/10.1016/j.earscirev.2019.04.015>
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. *Science* (80-.). 327, 812–818. <https://doi.org/10.1126/science.1185383>
- Goodwell, A.E., Jiang, P., Ruddell, B.L., Kumar, P., 2020. Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback. *Water Resour Res* 56, 1–12. <https://doi.org/10.1029/2019WR024940>
- Goodwell, A.E., Kumar, P., 2017. Temporal Information Partitioning Networks (TIPNets): A process network approach to infer ecohydrologic shifts. *Water Resource Research* 53, 5899–5919. <https://doi.org/10.1002/2016WR020216>
- Grillakis, M.G., 2019. Increase in severe and extreme soil moisture droughts for Europe under climate change. *Sci. Total Environ.* 660, 1245–1255. <https://doi.org/10.1016/j.scitotenv.2019.01.001>
- Guo, H., Bao, A., Liu, T., Ndayisaba, F., He, D., Kurban, A., De Maeyer, P., 2017. Meteorological Drought Analysis in the Lower Mekong Basin Using Satellite-Based Long-Term CHIRPS Product. *Sustainability* 9, 901. <https://doi.org/10.3390/su9060901>
- Hagemann, S., Dümenil Gates, L., 2001. Validation of the hydrological cycle ECMWF and NCEP reanalyses using the MPI hydrological discharge model. *Journal of Geophysical Research Atmospheres* 106, 1503–1510. <https://doi.org/10.1029/2000jd900568>
- Hamman, J.J., Nijssen, B., Bohn, T.J., Gergel, D.R., Mao, Y., 2018. The variable infiltration capacity model version 5 (VIC-5): Infrastructure improvements for new applications and

- reproducibility. *Geosci Model Dev* 11, 3481–3496. <https://doi.org/10.5194/gmd-11-3481-2018>
- Han, E., Ines, A.V.M., Koo, J., 2019. Development of a 10-km resolution global soil profile dataset for crop modeling applications. *Environ. Model. Softw.* 119, 70–83. <https://doi.org/10.1016/j.envsoft.2019.05.012>
- Hansen, J, Challinor, A, Ines, A, Wheeler, T, Moron, V., Hansen, Jw, Challinor, A, Ines, A, Wheeler, T, Moron, V., Hansen, J.W., Challinor, Andrew, Ines, Amor, Wheeler, Tim, Moron, V., 2006. Translating climate forecasts into agricultural terms : advances and challenges To cite this version : HAL Id : hal-02894588 Translating climate forecasts into agricultural terms : advances and challenges.
- Hao, C., Zhang, J., Yao, F., 2015. Combination of multi-sensor remote sensing data for drought monitoring over Southwest China. *Int. J. Appl. Earth Obs. Geoinf.* 35, 270–283. <https://doi.org/10.1016/j.jag.2014.09.011>
- Hendricks Franssen, H. J., & Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44(9).
- Hoang, L.P., van Vliet, M.T.H., Kumm, M., Lauri, H., Koponen, J., Supit, I., Leemans, R., Kabat, P., Ludwig, F., 2019. The Mekong's future flows under multiple drivers: How climate change, hydropower developments and irrigation expansions drive hydrological changes. *Science of The Total Environment* 649, 601–609. <https://doi.org/10.1016/j.scitotenv.2018.08.160>
- Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: Current status and future prospects. *Environmental Modelling and Software* 72, 276–286. <https://doi.org/10.1016/j.envsoft.2014.12.013>
- Hussain, M., Waqas-ul-Haq, M., Farooq, S., Jabran, K., & Farroq, M. (2016). The impact of seed priming and row spacing on the productivity of different cultivars of irrigated wheat under early season drought. *Experimental Agriculture*, 52(3), 477-490.
- Ines, A.V.M., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* 138, 149–164. <https://doi.org/10.1016/j.rse.2013.07.018>
- Innes, P.J., Tan, D.K.Y., Van Ogtrop, F., Amthor, J.S., 2015. Effects of high-temperature episodes on wheat yields in New South Wales, Australia. *Agric For Meteorol* 208, 95–107. <https://doi.org/10.1016/j.agrformet.2015.03.018>
- IPCC, 2007. Climate change 2001, Weather. <https://doi.org/10.1256/wea.58.04>
- Jägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J.A., Fuchs, K., Guarin, J.R., Heinke, J., Hoogenboom, G., Iizumi,

- T., Jain, A.K., Kelly, D., Khabarov, N., Lange, S., Lin, T.S., Liu, W., Mialyk, O., Minoli, S., Moyer, E.J., Okada, M., Phillips, M., Porter, C., Rabin, S.S., Scheer, C., Schneider, J.M., Schyns, J.F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F., Rosenzweig, C., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat Food* 2, 873–885. <https://doi.org/10.1038/s43016-021-00400-y>
- Jha, B., Kumar, A., Hu, Z.Z., 2019. An update on the estimate of predictability of seasonal mean atmospheric variability using North American Multi-Model Ensemble. *Clim Dyn* 53, 7397–7409. <https://doi.org/10.1007/s00382-016-3217-1>
- Johnston, R., Hoanh, C.T., Lacombe, G., Lefroy, R., Pavelic, P., Fry, C., 2012. Managing water in rainfed agriculture in the Greater Mekong Subregion. <https://doi.org/10.5337/2012.201>
- Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., Ritchie, J., 2003. The DSSAT cropping system model. *Eur. J. Agron.* 18, 235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)
- Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. *European Journal of Agronomy* 18, 235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)
- Kadiyala, M.D.M., Jones, J.W., Mylavarapu, R.S., Li, Y.C., Reddy, M.D., 2015. Identifying irrigation and nitrogen best management practices for aerobic rice-maize cropping system for semi-arid tropics using CERES-rice and maize models. *Agric Water Manag* 149, 23–32. <https://doi.org/10.1016/j.agwat.2014.10.019>
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bull Am Meteorol Soc* 77, 437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kang, H., Sridhar, V., Mainuddin, M., Trung, L.D., 2021. Future rice farming threatened by drought in the Lower Mekong Basin. *Sci Rep* 11, 1–15. <https://doi.org/10.1038/s41598-021-88405-2>
- Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *J Hydrol (Amst)* 586, 124905. <https://doi.org/10.1016/j.jhydrol.2020.124905>
- Katsanos, D., Retalis, A., Michaelides, S., 2016. Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period. *Atmos. Res.* 169, 459–464. <https://doi.org/10.1016/j.atmosres.2015.05.015>
- Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, 12(22), 3783.

- Simons, G., Poortinga, A., Bastiaanssen, W. G., Saah, D., Troy, D., Hunink, J., ... & Clinton, N. (2017). On Spatially Distributed Hydrological Ecosystem Services: Bridging the Quantitative Information Gap Using Remote Sensing and Hydrological Models.
- Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., ... & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and forest meteorology*, 276, 107609.
- Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*, 2, 1-12.
- Liu, S. Y. (2020). Artificial intelligence (AI) in agriculture. *IT Professional*, 22(3), 14-15.
- Whitcraft, A. K., Becker-Reshef, I., Justice, C. O., Gifford, L., Kavvada, A., & Jarvis, I. (2019). No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework. *Remote Sensing of Environment*, 235, 111470.
- Kim, S., Shao, W., Kam, J., 2019. Spatiotemporal patterns of US drought awareness. *Palgrave Commun.* 5, 107. <https://doi.org/10.1057/s41599-019-0317-7>
- Kite, G., 2001. Modelling the Mekong: hydrological simulation for environmental impact studies. *J. Hydrol.* 253, 1–13. [https://doi.org/10.1016/S0022-1694\(01\)00396-1](https://doi.org/10.1016/S0022-1694(01)00396-1)
- Klisch, A., Atzberger, C., 2016. Operational Drought Monitoring in Kenya Using MODIS NDVI Time Series. *Remote Sens.* 8, 267. <https://doi.org/10.3390/rs8040267>
- Kogan, F.N., 1995. Droughts of the Late 1980s in the United States as Derived from NOAA Polar-Orbiting Satellite Data. *Bull. Am. Meteorol. Soc.* 76, 655–668. [https://doi.org/10.1175/1520-0477\(1995\)076<0655:DOTLIT>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0655:DOTLIT>2.0.CO;2)
- Konapala, G., Mishra, A.K., Wada, Y., Mann, M.E., 2020. Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nat Commun* 11, 1–10. <https://doi.org/10.1038/s41467-020-16757-w>
- Konduri, V.S., Kumar, J., Hargrove, W.W., Hoffman, F.M., Ganguly, A.R., 2020. Mapping crops within the growing season across the United States. *Remote Sens Environ* 251, 112048. <https://doi.org/10.1016/j.rse.2020.112048>
- Kurtz, W., Franssen, H. J. H., Kaiser, H. P., & Vereecken, H. (2014). Joint assimilation of piezometric heads and groundwater temperatures for improved modeling of river-aquifer interactions. *Water Resources Research*, 50(2), 1665-1688.
- Kushwaha, N.L., Rajput, J., Shirsath, P.B., Sena, D.R., Mani, I., 2022. Seasonal climate forecasts (SCFs) based risk management strategies: A case study of rainfed rice cultivation in India. *Journal of Agrometeorology* 24, 10–17. <https://doi.org/10.54386/jam.v24i1.775>

- Lacasa, J., Messina, C.D., Ciampitti, I., 2023. A probabilistic framework for forecasting maize crop yield response to agricultural inputs with sub-seasonal climate predictions. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/acd8d1>
- Lahoz, B. K. W., & Menard, R. (2010). *Data assimilation*. Springer-Verlag Berlin Heidelberg.
- Lal, R., & Stewart, B.A. (Eds.). (2018). *Soil and Climate* (1st ed.). CRC Press. <https://doi.org/10.1201/b21225>
- Lauri, H., Räsänen, T.A., Kumm, M., 2014. Using Reanalysis and Remotely Sensed Temperature and Precipitation Data for Hydrological Modeling in Monsoon Climate: Mekong River Case Study. *J Hydrometeorol* 15, 1532–1545. <https://doi.org/10.1175/jhm-d-13-084.1>
- Lehmann, J., Mempel, F., & Coumou, D. (2018). Increased occurrence of record-wet and record-dry months reflect changes in mean rainfall. *Geophysical Research Letters*, 45(24), 13-468.
- Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* 99, 14415. <https://doi.org/10.1029/94JD00483>
- Liu, S., Lu, P., Liu, D., Jin, P., 2007. Pinpointing source of Mekong and measuring its length through analysis of satellite imagery and field investigations. *Geo-spatial Inf. Sci.* 10, 51–56. <https://doi.org/10.1007/s11806-007-0011-6>
- Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature climate change*, 3(5), 497-501.
- Mahmoud, S. H., & Gan, T. Y. (2018). Impact of anthropogenic climate change and human activities on environment and ecosystem services in arid regions. *Science of the Total Environment*, 633, 1329-1344.
- Mainuddin, M., Kirby, M., Hoanh, C.T., 2013. Impact of climate change on rainfed rice and options for adaptation in the lower Mekong Basin. *Natural Hazards* 66, 905–938. <https://doi.org/10.1007/s11069-012-0526-5>
- Margulis, S. A., McLaughlin, D., Entekhabi, D., & Dunne, S. (2002). Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water resources research*, 38(12), 35-1.
- Martínez-Fernández, J., González-Zamora, A., Sánchez, N., Gumuzzio, A., & Herrero-Jiménez, C. M. (2016). Satellite soil moisture for agricultural drought monitoring: Assessment of the SMOS derived Soil Water Deficit Index. *Remote Sensing of Environment*, 177, 277-286.
- Maurer, E.P., Hidalgo, H.G., Das, T., Dettinger, M.D., Cayan, D.R., 2010. The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow

- in California. *Hydrol Earth Syst Sci* 14, 1125–1138. <https://doi.org/10.5194/hess-14-1125-2010>
- Mavromatis, T., 2015. Crop–climate relationships of cereals in Greece and the impacts of recent climate trends. *Theor Appl Climatol* 120, 417–432. <https://doi.org/10.1007/s00704-014-1179-y>
- McCabe, G.J., Palecki, M.A., Betancourt, J.L., 2004. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proc. Natl. Acad. Sci.* 101, 4136–4141. <https://doi.org/10.1073/pnas.0306738101>
- McKee et al., 1993 T.B. McKee, N.J. Doesken, J. Kleist The relationship of drought frequency and duration to time scales *Proceedings of the Eighth Conference on Applied Climatology, American Meteorological Society, Anaheim, CA (1993), pp. 179-183*
- Mekong River Commission, 2009. Initiative on sustainable hydropower work plan. Me-kong River Commission. (Available at: <http://www.mrcmekong.org/programmes/hydropower/hydropower-pub.htm>. Accessed on: 17/12/2010).
- Midega, C. A., Bruce, T. J., Pickett, J. A., Pittchar, J. O., Murage, A., & Khan, Z. R. (2015). Climate-adapted companion cropping increases agricultural productivity in East Africa. *Field Crops Research*, 180, 118-125.
- Milhorance, C., & Bursztyn, M. (2019). Climate adaptation and policy conflicts in the Brazilian Amazon: prospects for a Nexus+ approach. *Climatic Change*, 155(2), 215-236.
- Minoli, S., Jägermeyr, J., Asseng, S., Urfels, A., Müller, C., 2022. Global crop yields can be lifted by timely adaptation of growing periods to climate change. *Nat Commun* 13. <https://doi.org/10.1038/s41467-022-34411-5>
- Mishra, A., Ketelaar, J. W., Uphoff, N., & Whitten, M. (2021). Food security and climate-smart agriculture in the lower Mekong basin of Southeast Asia: Evaluating impacts of system of rice intensification with special reference to rainfed agriculture. *International Journal of Agricultural Sustainability*, 19(2), 152-174.
- Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. *J. Hydrol.* 391, 202–216. <https://doi.org/10.1016/j.jhydrol.2010.07.012>
- Mizukami, N., P. Clark, M., G. Slater, A., D. Brekke, L., M. Elsner, M., R. Arnold, J., Gangopadhyay, S., 2014. Hydrologic Implications of Different Large-Scale Meteorological Model Forcing Datasets in Mountainous Regions. *J. Hydrometeorol.* 15, 474–488. <https://doi.org/10.1175/JHM-D-13-036.1>
- Mladenova, I. E., Bolten, J. D., Crow, W. T., Anderson, M. C., Hain, C. R., Johnson, D. M., & Mueller, R. (2017). Intercomparison of soil moisture, evaporative stress, and vegetation indices for estimating corn and soybean yields over the US. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4), 1328-1343.

- MRC, 2003. State of the Basin Report State of the Basin Report Executive Summary.
- MRC, 2014. Crop production for food security and rural poverty Baseline and pilot modelling.
- MRC, 2014. Crop production for food security and rural poverty Baseline and pilot modelling.
- Müller, C., Franke, J., Jägermeyr, J., Ruane, A.C., Elliott, J., Moyer, E., Heinke, J., Falloon, P.D., Folberth, C., Francois, L., Hank, T., Izaurrealde, R.C., Jacquemin, I., Liu, W., Olin, S., Pugh, T.A.M., Williams, K., Zabel, F., 2021. Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios. *Environmental Research Letters* 16. <https://doi.org/10.1088/1748-9326/abd8fc>
- Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., ... & Running, S. W. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote sensing of environment*, 83(1-2), 214-231.
- Naeini, M. R., Yang, T., Sadegh, M., AghaKouchak, A., Hsu, K. L., Sorooshian, S., ... & Lei, X. (2018). Shuffled complex-self adaptive hybrid evolution (SC-SAHEL) optimization framework. *Environmental Modelling & Software*, 104, 215-235.
- Narasimhan, B., Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. *Agric. For. Meteorol.* 133, 69–88. <https://doi.org/10.1016/j.agrformet.2005.07.012>
- Ortiz-Bobea, A., Ault, T.R., Carrillo, C.M., Chambers, R.G., Lobell, D.B., 2021. Anthropogenic climate change has slowed global agricultural productivity growth. *Nat Clim Chang* 11, 306–312. <https://doi.org/10.1038/s41558-021-01000-1>
- Padrón, R.S., Gudmundsson, L., Decharme, B., Ducharne, A., Lawrence, D.M., Mao, J., Peano, D., Krinner, G., Kim, H., Seneviratne, S.I., 2020. Observed changes in dry-season water availability attributed to human-induced climate change. *Nat Geosci* 13, 477–481. <https://doi.org/10.1038/s41561-020-0594-1>
- Palmer, W.C., 1965. Palmer_1965.pdf. Res. Pap.
- Pandey, S., Bhandari, H., Ding, S., Prapertchob, P., Sharan, R., Naik, D., Taunk, S.K., Sastri, A., 2007. Coping with drought in rice farming in Asia: insights from a cross-country comparative study. *Agric. Econ.* 37, 213–224. <https://doi.org/10.1111/j.1574-0862.2007.00246.x>
- Pasetto, D., Camporese, M., & Putti, M. (2012). Ensemble Kalman filter versus particle filter for a physically-based coupled surface–subsurface model. *Advances in water resources*, 47, 1-13.
- Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., Athanasiadis, I.N., 2023. Interpretability of deep learning models for crop yield forecasting. *Comput Electron Agric* 206. <https://doi.org/10.1016/j.compag.2023.107663>

- Pokhrel, Y., Shin, S., Lin, Z., Yamazaki, D., Qi, J., 2018. Potential Disruption of Flood Dynamics in the Lower Mekong River Basin Due to Upstream Flow Regulation. *Sci. Rep.* 8, 17767. <https://doi.org/10.1038/s41598-018-35823-4>
- Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., Heim, R.R., Brewer, M.J., Svoboda, M., Westerhoff, R., van Dijk, A.I.J.M., Lloyd-Hughes, B., Pappenberger, F., Werner, M., Dutra, E., Wetterhall, F., Wagner, W., Schubert, S., Mo, K., Nicholson, M., Bettio, L., Nunez, L., van Beek, R., Bierkens, M., de Goncalves, L.G.G., de Mattos, J.G.Z., Lawford, R., 2013. Toward Global Drought Early Warning Capability: Expanding International Cooperation for the Development of a Framework for Monitoring and Forecasting. *Bull. Am. Meteorol. Soc.* 94, 776–785. <https://doi.org/10.1175/BAMS-D-11-00176.1>
- Qi, Wei, Lian Feng, Hong Yang, and Junguo Liu. "Increasing concurrent drought probability in global main crop production countries." *Geophysical Research Letters* 49, no. 6 (2022): e2021GL097060.
- Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B. E., Disney, M., & Bowyer, P. (2008). Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, 112(4), 1347-1364.
- Rahnamay Naeini, M., Yang, T., Sadegh, M., AghaKouchak, A., Hsu, K. lin, Sorooshian, S., Duan, Q., Lei, X., 2018. Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL) optimization framework. *Environmental Modelling and Software* 104, 215–235. <https://doi.org/10.1016/j.envsoft.2018.03.019>
- Ray, D.K., Gerber, J.S., Macdonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. *Nat Commun* 6. <https://doi.org/10.1038/ncomms6989>
- Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002). Extended versus ensemble Kalman filtering for land data assimilation. *Journal of hydrometeorology*, 3(6), 728-740.
- Rhee, J., Im, J., Carbone, G.J., 2010. Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. *Remote Sens. Environ.* 114, 2875–2887. <https://doi.org/10.1016/j.rse.2010.07.005>
- Rifai, S.W., Li, S., Malhi, Y., 2019. Coupling of El Niño events and long-term warming leads to pervasive climate extremes in the terrestrial tropics. *Environmental Research Letters* 14. <https://doi.org/10.1088/1748-9326/ab402f>
- Robinson, A., Lehmann, J., Barriopedro, D., Rahmstorf, S., & Coumou, D. (2021). Increasing heat and rainfall extremes now far outside the historical climate. *npj Climate and Atmospheric Science*, 4(1), 45.
- Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resour. Technol. Satell. Symp.* 1, 309–317. <https://doi.org/citeulike-article-id:12009708>

- Ruddell, B.L., Kumar, P., 2009a. Ecohydrologic process networks: 1. Identification. *Water Resour Res* 45, 1–23. <https://doi.org/10.1029/2008WR007279>
- Ruddell, B.L., Kumar, P., 2009b. Ecohydrologic process networks: 2. Analysis and characterization. *Water Resour Res* 45, 1–14. <https://doi.org/10.1029/2008WR007280>
- Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., ... & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature communications*, 8(1), 13931.
- Sendrowski, A., Passalacqua, P., 2017. *Water Resources Research*. *Water Resour Res* 53, 1841–1863. <https://doi.org/10.1002/2016WR019768>
- Seo, D. J., Cajina, L., Corby, R., & Howieson, T. (2009). Automatic state updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, 367(3-4), 255-275.
- Shannon, C.E., 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Sheffield, J., Wood, E.F., 2007. Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *J Geophys Res* 112, D17115. <https://doi.org/10.1029/2006JD008288>
- Sheffield, J., Wood, E.F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S., Ogallo, L., 2014. A Drought Monitoring and Forecasting System for Sub-Saharan African Water Resources and Food Security. *Bull. Am. Meteorol. Soc.* 95, 861–882. <https://doi.org/10.1175/BAMS-D-12-00124.1>
- Slater, L.J., Villarini, G., Bradley, A.A., 2019. Evaluation of the skill of North-American Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA. *Clim Dyn* 53, 7381–7396. <https://doi.org/10.1007/s00382-016-3286-1>
- Son, N.T., Chen, C.F., Chen, C.R., Chang, L.Y., Minh, V.Q., 2012. Monitoring agricultural drought in the Lower Mekong Basin using MODIS NDVI and land surface temperature data. *International Journal of Applied Earth Observation and Geoinformation* 18, 417–427. <https://doi.org/10.1016/j.jag.2012.03.014>
- Subash, N., Ram Mohan, H.S., 2012. Evaluation of the impact of climatic trends and variability in rice-wheat system productivity using Cropping System Model DSSAT over the Indo-Gangetic Plains of India. *Agric For Meteorol* 164, 71–81. <https://doi.org/10.1016/j.agrformet.2012.05.008>
- Summary., I.T., Shukla, P.R., Skea, J., Slade, R., Diemen, R. van, Haughey, E., Malley, J., Pathak, M., Pereira, J.P., 2019. Foreword Technical and Preface. *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems* 35–74.

- Svoboda, M., LeCompte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., et al. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, 83, 1181–1190. <https://doi.org/10.1175/1520-0477-83.8.1181>
- Tabari, H., & Willems, P. (2023). Sustainable development substantially reduces the risk of future drought impacts. *Communications Earth & Environment*, 4(1), 180.
- Tadasse, G., Algieri, B., Kalkuhl, M., & Von Braun, J. (2016). Drivers and triggers of international food price spikes and volatility. *Food price volatility and its implications for food security and policy*, 59-82.
- Thiesen, S., Darscheid, P., Ehret, U., 2019. Identifying rainfall-runoff events in discharge time series: A data-driven method based on information theory. *Hydrol Earth Syst Sci* 23, 1015–1034. <https://doi.org/10.5194/hess-23-1015-2019>
- Thilakaratne, M., Sridhar, V., 2017. Characterization of future drought conditions in the Lower Mekong River Basin. *Weather Clim. Extrem.* 17, 47–58. <https://doi.org/10.1016/j.wace.2017.07.004>
- Thober, S., Kumar, R., Sheffield, J., Mai, J., Schäfer, D., Samaniego, L., 2015. Seasonal soil moisture drought prediction over Europe using the North American Multi-Model Ensemble (NMME). *J Hydrometeorol* 16, 2329–2344. <https://doi.org/10.1175/JHM-D-15-0053.1>
- Tippett, M.K., Ranganathan, M., L’Heureux, M., Barnston, A.G., DelSole, T., 2019. Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Clim Dyn* 53, 7497–7518. <https://doi.org/10.1007/s00382-017-3721-y>
- Togliatti, K., Archontoulis, S. V., Dietzel, R., Puntel, L., VanLoocke, A., 2017. How does inclusion of weather forecasting impact in-season crop model predictions? *Field Crops Res* 214, 261–272. <https://doi.org/10.1016/j.fcr.2017.09.008>
- Tolimir, M., Kresović, B., Životić, L., Dragović, S., Dragović, R., Sredojević, Z., & Gajić, B. (2020). The conversion of forestland into agricultural land without appropriate measures to conserve SOM leads to the degradation of physical and rheological soil properties. *Scientific Reports*, 10(1), 13668.
- Toté, C., Patricio, D., Boogaard, H., van der Wijngaart, R., Tarnavsky, E., Funk, C., 2015. Evaluation of Satellite Rainfall Estimates for Drought and Flood Monitoring in Mozambique. *Remote Sens.* 7, 1758–1776. <https://doi.org/10.3390/rs70201758>
- Trenberth, K.E., Dai, A., van der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. *Nat. Clim. Chang.* 4, 17–22. <https://doi.org/10.1038/nclimate2067>
- Urban, D. W., Sheffield, J., & Lobell, D. B. (2015). The impacts of future climate and carbon dioxide changes on the average and variability of US maize yields under two emission scenarios. *Environmental Research Letters*, 10(4), 045003.

- Vadrevu, K., Heinimann, A., Gutman, G., & Justice, C. (2019). Remote sensing of land use/cover changes in South and Southeast Asian Countries. *International Journal of Digital Earth*, 12(10), 1099-1102.
- Van Loon, A.F., Gleeson, T., Clark, J., Van Dijk, A.I.J.M., Stahl, K., Hannaford, J., Di Baldassarre, G., Teuling, A.J., Tallaksen, L.M., Uijlenhoet, R., Hannah, D.M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., Rangecroft, S., Wanders, N., Van Lanen, H.A.J., 2016b. Drought in the Anthropocene. *Nat. Geosci.* 9, 89–91. <https://doi.org/10.1038/ngeo2646>
- Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* 23, 1696–1718. <https://doi.org/10.1175/2009JCLI2909.1>
- Vogel, E., Donat, M.G., Alexander, L. V., Meinshausen, M., Ray, D.K., Karoly, D., Meinshausen, N., Frieler, K., 2019. The effects of climate extremes on global agricultural yields. *Environmental Research Letters* 14. <https://doi.org/10.1088/1748-9326/ab154b>
- Vrugt, J. A., Gupta, H. V., Nualláin, B., & Bouten, W. (2006). Real-time data assimilation for operational ensemble streamflow forecasting. *Journal of Hydrometeorology*, 7(3), 548-565.
- Wang, L., Ren, H.L., Xu, X., Huang, B., Wu, J., Liu, J., 2022. Seasonal-Interannual Predictions of Summer Precipitation Over the Tibetan Plateau in North American Multimodel Ensemble. *Geophys Res Lett* 49. <https://doi.org/10.1029/2022GL100294>
- Wang, Q., Chun, J.A., Lee, W.-S., Sanai, L., Seng, V., 2017. Shifting Planting Dates and Fertilizer Application Rates as Climate Change Adaptation Strategies for Two Rice Cultivars in Cambodia. *J. Clim. Chang. Res.* 8, 187–199. <https://doi.org/10.15531/kscrr.2017.8.3.187>
- Wang, Y., Feng, L., Liu, J., Hou, X., Chen, D., 2020. Changes of inundation area and water turbidity of Tonle Sap Lake: responses to climate changes or upstream dam construction? *Environ. Res. Lett.* 15, 0940a1. <https://doi.org/10.1088/1748-9326/abac79>
- Wiebe, K., Lotze-Campen, H., Sands, R., Tabeau, A., van der Mensbrugghe, D., Biewald, A., ... & Willenbockel, D. (2015). Climate change impacts on agriculture in 2050 under a range of plausible socioeconomic and emissions scenarios. *Environmental Research Letters*, 10(8), 085010.
- Wilhite, D.A., 2000. Chapter1 Drought as a Natural Hazard, in: *Drought: A Global Assessment*. pp. 3–18.
- Wongchuig Correa, S., Paiva, R.C.D. de, Espinoza, J.C., Collischonn, W., 2017. Multi-decadal Hydrological Retrospective: Case study of Amazon floods and droughts. *J. Hydrol.* 549, 667–684. <https://doi.org/10.1016/j.jhydrol.2017.04.019>

- Xing, Z., Li, X., Fan, L., Colliander, A., Frappart, F., de Rosnay, P., ... & Wigneron, J. P. (2023). Assessment of 9 km SMAP soil moisture: Evidence of narrowing the gap between satellite retrievals and model-based reanalysis. *Remote Sensing of Environment*, 296, 113721.
- Yadav, P., Jaiswal, D. K., & Sinha, R. K. (2021). Climate change: Impact on agricultural production and sustainable mitigation. In *Global climate change* (pp. 151-174). Elsevier.
- Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., ... & Myneni, R. B. (2006). MODIS leaf area index products: From validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), 1885-1898.
- Zhang, A., Jia, G., 2013. Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data. *Remote Sens. Environ.* 134, 12–23. <https://doi.org/10.1016/j.rse.2013.02.023>
- Zhang, J., Mu, Q., Huang, J., 2016. Assessing the remotely sensed Drought Severity Index for agricultural drought monitoring and impact analysis in North China. *Ecol. Indic.* 63, 296–309. <https://doi.org/10.1016/j.ecolind.2015.11.062>
- Zhang, L., Jiao, W., Zhang, H., Huang, C., Tong, Q., 2017. Studying drought phenomena in the Continental United States in 2011 and 2012 using various drought indices. *Remote Sens. Environ.* 190, 96–106. <https://doi.org/10.1016/j.rse.2016.12.010>
- Zhang, Y., Pan, M., Sheffield, J., Siemann, A.L., Fisher, C.K., Liang, M., Beck, H.E., Wanders, N., MacCracken, R.F., Houser, P.R., Zhou, T., Lettenmaier, D.P., Pinker, R.T., Bytheway, J., Kummerow, C.D., Wood, E.F., 2018. A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. *Hydrol Earth Syst Sci* 22, 241–263. <https://doi.org/10.5194/hess-22-241-2018>
- Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.L., Elliott, J., Ewert, F., Janssens, I.A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peñuelas, J., Ruane, A.C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., Asseng, S., 2017. Temperature increase reduces global yields of major crops in four independent estimates. *Proc Natl Acad Sci U S A* 114, 9326–9331. <https://doi.org/10.1073/pnas.1701762114>
- Zheng, G., & Moskal, L. M. (2009). Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. *Sensors*, 9(4), 2719-2745.
- Zhou, Y., McLaughlin, D., & Entekhabi, D. (2006). Assessing the performance of the ensemble Kalman filter for land surface data assimilation. *Monthly Weather Review*, 134(8), 2128-2142.

APPENDIX A. IMPLICATIONS OF DROUGHT ON INTERANNUAL CROP YIELDS

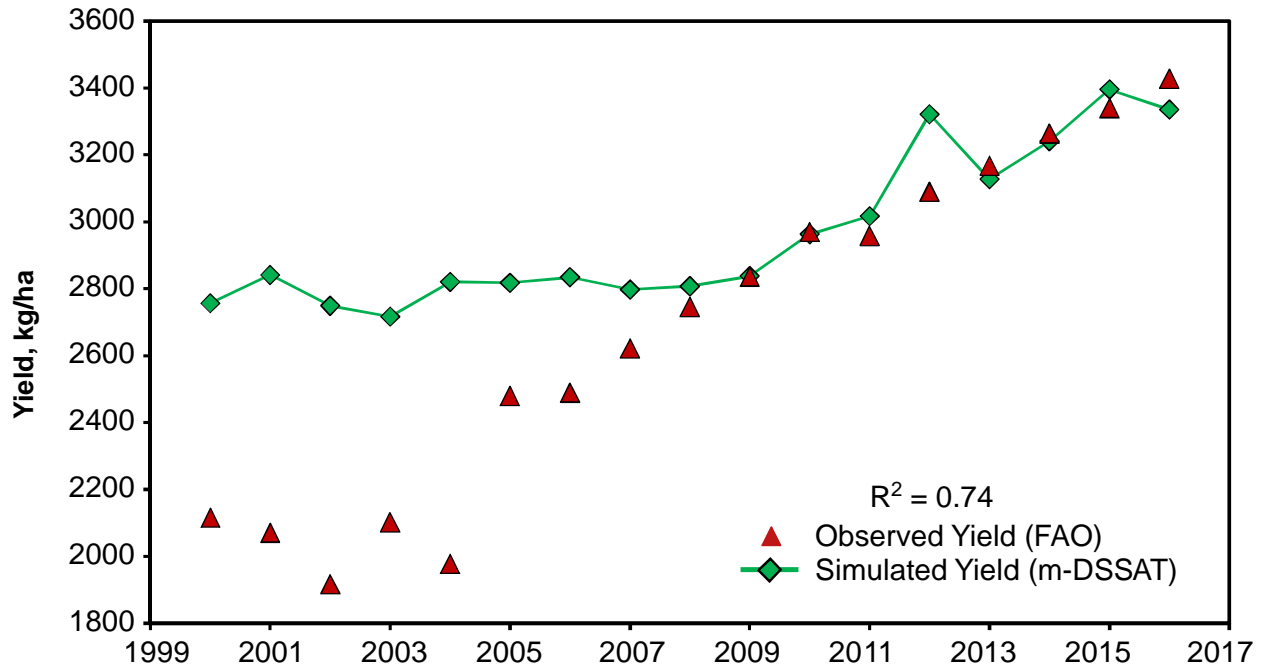


Figure 31. Historic temporal variation of observed (FAO) and simulated (*m*-DSSAT) rice yields over Cambodia between 2000 and 2016.

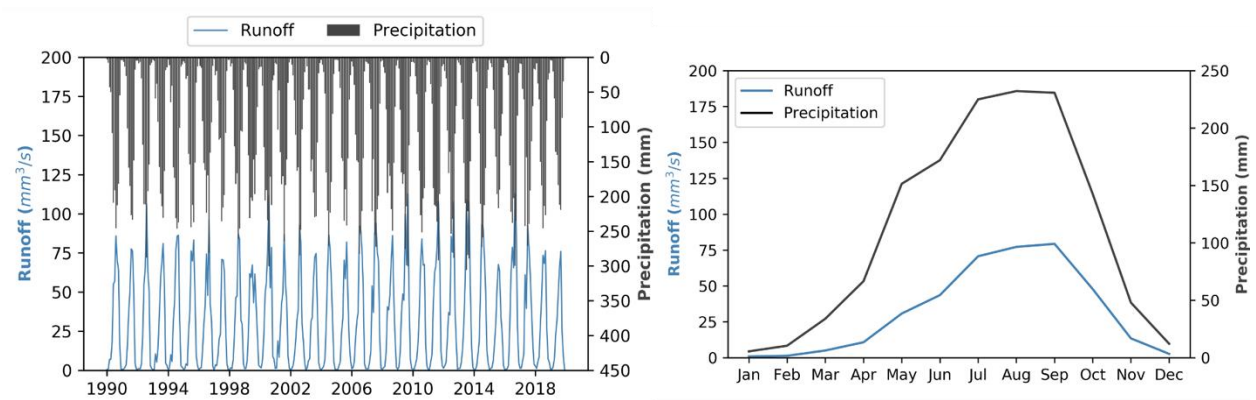


Figure 32. Historical temporal variations of annual runoff and precipitation totals over Cambodia between 1990-2019 (left); Mean monthly variation of average precipitation and runoff between 1990-2019 (right).

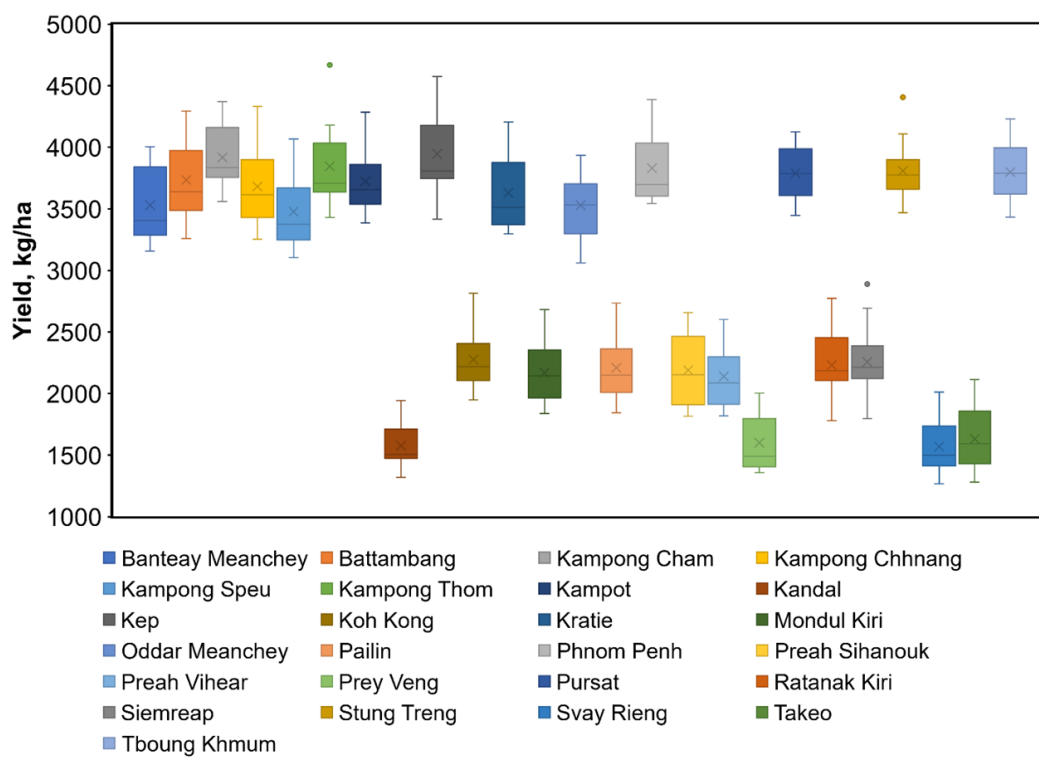


Figure 33. Provincial rice yields (from *m*-DSSAT) over Cambodia from 2000-2016.

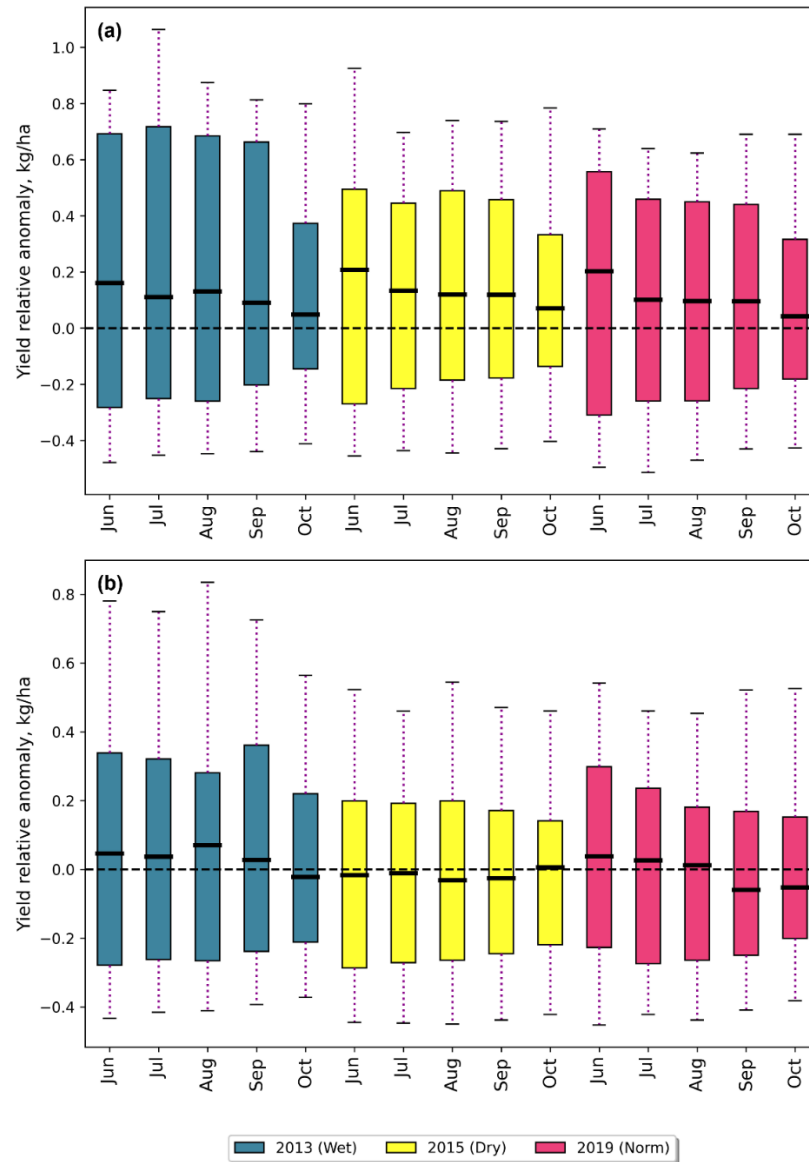


Figure 34. Relative yield anomaly. a) Boxplot of the relative yield anomalies over Takeo and b) Prey Veng provinces during different timeframes of the study period. Each year is represented by a distinct color and consists of different timeframes within that particular season. Yield anomalies are based on the median value of the observation yields and the ensembles of simulated yields to show the relative losses.

APPENDIX B. EQUATIONS FOR VARIOUS STATISTICAL MEASURES

Detailed formulations of key statistical measures utilized in this study are presented below, including equations for Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Bias, among others. These equations are fundamental to the analytical methods employed for data analysis and interpretation.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (B.1)$$

$$RNMRSE = RMSE / \bar{O} \quad (B.2)$$

$$EF = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (B.3)$$

$$d = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (B.4)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (B.5)$$