Terrestrial fluxes in major Indian agroecosystems: A modeling study using the Community Land Model (CLM5)

Kangari Narender Reddy

Centre for Atmospheric Sciences
Indian Institute of Technology Delhi

August 2025

Terrestrial fluxes in major Indian agroecosystems: A modeling study using the Community Land Model (CLM5)

by

Kangari Narender Reddy

Centre for Atmospheric Sciences

submitted

in fulfillment of the requirements for the degree of Doctor of Philosophy

to

Indian Institute of Technology Delhi

August 2025

Certificate

This is to certify that the thesis entitled "**Terrestrial fluxes in major Indian agroecosystems: A modeling study using the Community Land Model (CLM5)**" submitted by **Mr. Kangari Narender Reddy** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy** is a record of original bonafied research carried out by him. Mr. Narender Reddy has worked under my guidance and supervision and has fulfilled the requirements for the submission of this thesis. The results contained in this thesis have not been submitted in part or full to any other University or Institute for the award of any degree or diploma.

Dr. Somnath Baidya Roy
Professor
Center for Atmospheric Sciences
Indian Institute of Technology Delhi
Hauz Khas, New Delhi- 110016
India

Acknowledgments

I would like to express my sincere gratitude to my parents and family for their constant support, encouragement, and sacrifices throughout my doctoral journey. Their unwavering belief in me has been a source of strength and motivation.

I am deeply indebted to my supervisor, **Dr. Somnath Baidya Roy**, for his invaluable guidance, insightful suggestions, and constant encouragement. His mentorship has been instrumental in shaping my research skills and academic development.

I gratefully acknowledge the contributions of my colleagues, **Vangala Gayatri, Venkateshwara Varma, and Riya Yadav**, whose efforts were crucial in assembling the dataset used in this work.

I also extend my gratitude to my lab mates for their support and cooperation, which made the research environment stimulating and collaborative. Finally, I am thankful to my friends at IIT Delhi for their companionship, which made my time here both memorable and fulfilling.

Kangari Narender Reddy
IIT Delhi
August 2025

Abstract

Agroecosystems cover nearly 50% of the land in India, with wheat and rice accounting for approximately 80% of the farmlands. Several site-scale investigations explored terrestrial carbon, water, and energy fluxes (terrestrial fluxes – from now in the thesis) within Indian agroecosystems (Patel et al., 2011 and 2021; Bhattacharya et al., 2013). However, these do not accurately reflect the extensive and climatically varied farming environments present in India. Consequently, the primary aim of this study is to quantitatively understand the long-term trends and regional variations in yield, growth parameters, and terrestrial fluxes of Indian agroecosystems. Land surface models (LSMs) are valuable tools for such studies because they can be used to conduct numerical experiments to examine the effects of natural and human drivers on agroecosystems across regional and global scales. The Community Land Model version 5.0 (CLM5) is a state-of-the-art LSM that has been widely used to study agroecosystems. However, the depiction of Indian agroecosystems in CLM5 is not realistic, leading to biases in the simulation of yield and irrigation compared to observed data (Lombardozzi et al., 2020; Mathur and AchuthaRao, 2020). The primary challenge in improving the representation of Indian crops is the scarcity of publicly accessible crop data necessary for model calibration and evaluation (Gahlot et al., 2020).

This thesis aims to investigate the spatio-temporal variability in crop phenology and terrestrial fluxes across Indian agroecosystems and the role of natural and anthropogenic drivers in the observed variability. The specific objectives are outlined as follows: (1) to understand the crop phenology of Indian agroecosystems from site-scale observational data, (2) to improve the CLM5 model using site-scale observations, (3) to evaluate the improved CLM5 model at a regional scale, (4) to estimate the regional and inter-annual variability of crop growth parameters and terrestrial fluxes within Indian agroecosystems and the role of various drivers using CLM5 model, and finally (5) to evaluate the improved CLM5 model against the state-of-the-art LSM representing Indian agroecosystems.

First, Chapter 2 addresses the objective 1 that attempts to solve the main bottleneck in accurately simulating Indian agroecosystems, specifically the absence of publicly accessible crop datasets that hinders our ability to quantitatively understand and model agroecosystems. This thesis compiled a novel crop dataset focusing on two significant crops in India: rice and wheat. This was done by digitizing the archived masters and doctoral thesis from agricultural universities across India. The crop dataset spans from 1970 to 2020, encompassing site-scale

observations across 47 sites and 147 growing seasons. This dataset served as the foundation for calibrating the CLM5 model. The LAI data from the site-scale observations shows that the major growing season for rice is June to October and for wheat it is November to April. Major observations from the data analysis are that rice and wheat have a statistically significant increasing trend in yield (38 kg/ha/yr– 97.4% increase, for rice and 34 kg/ha/yr– 83% increase, for wheat) over the 1970-2020 period. However, there is no considerable difference or trend observed in the growing season length of the crops during this period. Interestingly, a statistically significant increasing trend in harvest dates of wheat crop is observed from the dataset; harvest dates extended by ~14 days over the 50-year period ($p<0.1$).

Chapter 3 addresses objective 2, where the CLM5 model is improved and evaluated against site-scale observations. The default version of the CLM5 model failed to replicate the crop phenology and growth seen in the site-scale observations. To address this drawback, the crop module in CLM5 is calibrated and validated at a point scale using the novel Indian crop dataset compiled in this thesis. The representation of crops in CLM5 is improved through a two-step approach: (1) calibrating the growing season and crop parameters specific to crop growth; (2) integrating regional variations in crop growth parameters, which is crucial for a country like India with its diverse climate conditions. The calibration of the growing season and crop parameters involved 30 sensitivity simulations. I evaluated the base temperature used in calculating the growing degree days for the crop at each site to account for the regional variability in crop growth. One hundred fifty sensitivity experiments are carried out to establish a relationship between latitude and base temperature for a crop. The calibration of the model significantly improved the simulation of the growing season, crop physiological parameters, crop phenology, and yield compared to site-scale observations. RMSE for Leaf Area Index (LAI) in wheat and rice reduced by ~50%. The RMSE in yield estimations for wheat and rice reduced by ~40% and ~60%, respectively. The RMSE of the growing season length for the wheat crop reduced by ~75%.

Chapter 4 addresses objective 3, where the improved CLM5 is evaluated at regional scale by conducting numerical simulations for the 2000–2014 period. For this purpose, the values of yield, irrigation amounts and patterns, and terrestrial fluxes simulated by the default and improved models are compared against multiple data sources. The results show that the CLM5 improvements reduce the RMSE in rice yield by 25% and the RMSE in GPP by about 50%. Major improvements are also observed in water and energy fluxes at the regional scale after improvements to the crop module. The irrigation pattern and amount irrigated saw a considerable improvement and are now closer to observations. The findings from the site-scale and regional-scale assessment of the improved model underscore the importance of employing region-specific crop parameters to accurately simulate agroecosystems and regional land surface processes.

Chapter 5 addresses objective 4. The chapter examines the regional and inter-annual variability in crop yield and terrestrial fluxes within Indian agroecosystems, utilizing multi-decadal simulations that cover the period from 1970 to 2014. Furthermore, numerical experiments are conducted to investigate the impact of two natural and two agricultural management factors on the observed trends in crop physiology and terrestrial fluxes throughout the simulation period. The LAI, yield, total dry matter of wheat and rice, and the carbon fluxes associated with the crops in their respective growing regions demonstrated statistically significant increasing trends ($p<.05$) from 1970 to 2014. The numerical experiments revealed that management practices predominantly drive the upward trend, while the rise in CO_2 is also a significant driver. The application of nitrogen fertilization had a significant impact on rice yield, which is cultivated in warm and wet seasons. Irrigation primarily influenced wheat yield, which is grown in dry and cold seasons. Even with the detrimental effects of climate on crop growth and terrestrial fluxes, higher CO_2 levels and increased human activities, like fertilization and irrigation, alleviated these challenges, leading to a rise in crop yields over the study period.

Chapter 6 of the thesis performs a comparative analysis of the CLM5 model and the Integrated Science Assessment Model (ISAM), which is the only other land surface model calibrated and assessed (Gahlot et al., 2020) to simulate the Indian crops to the best of my knowledge. This thesis has two advantages over the Gahlot et al. (2020). First, the CLM model is calibrated using data from 14 sites covering 33 growing seasons. Second, this study investigates both rice and wheat as opposed to Gahlot et al. (2020), who look at only spring wheat. Numerical experiments conducted in Chapter 5 with CLM5 were repeated using the ISAM model. The findings demonstrate that CLM5 simulates the observed variation in growing season length more precisely with an RMSE of 14 days, in contrast to the ISAM model, which has an RMSE of 28 days. The wheat yield in CLM5 shows a strong sensitivity to nitrogen fertilization, while in ISAM, the yield demonstrates a significant sensitivity to irrigation. The AR/GPP ratio exhibits variability in CLM5, while it remains constant in ISAM. The ratio in ISAM is comparatively low when evaluated against the values reported in multiple studies. Examining climate and human management impacts on carbon fluxes reveals that although net primary productivity (NPP) trends in CLM5 and ISAM are similar, the effects of human management practices on carbon fluxes vary considerably. Nitrogen fertilization has a significant impact in CLM5, while irrigation shows a more substantial effect in ISAM. Comparing the current state-of-the-art representation of Indian agroecosystems in LSMs, i.e., CLM5 and ISAM, major conclusions are drawn. Both models have a lot of scope for improvement. CLM5 can incorporate the precipitation variability to simulate regional variability in sowing dates. ISAM can incorporate an improved carbon-nitrogen cycle so that the simulated carbon fluxes over agroecosystems are closer to observations.

This is the first study to address the long-term spatial and inter-annual variability of agroecosystem physiological parameters and their terrestrial fluxes in India using a robustly calibrated LSM. One of the primary contributions of this work is making the Indian crop data accessible to all model developers to enhance Indian crop simulation in models, which is a long-lasting bottleneck in the accurate representation of Indian crops.

सारांश

भारत की लगभग 50% भूमि agroecosystems (Agroecosystems) से आच्छादित है, जिनमें wheat और rice लगभग 80% कृषि भूमि पर उगाए जाते हैं। कई स्थल-स्तरीय अध्ययनों ने भारतीय agroecosystems में स्थलीय कार्बन, जल और ऊर्जा प्रवाहों (आगे इस शोध में इन्हें terrestrial fluxes कहा गया है) की जांच की है (Patel et al., 2011 and 2021; Bhattacharya et al., 2013)। तथापि, ये अध्ययन भारत की व्यापक और जलवायु की दृष्टि से विविध कृषि-परिस्थितियों को सही रूप में नहीं दर्शाते। फलस्वरूप, इस शोध का मुख्य उद्देश्य भारतीय agroecosystems की उपज, वृद्धि-परिमाणों तथा terrestrial fluxes में दीर्घकालिक प्रवृत्तियों और क्षेत्रीय विविधताओं को मात्रात्मक रूप से समझना है। Land Surface Models (LSMs) इस प्रकार के अध्ययनों के लिए उपयोगी साधन हैं क्योंकि इनके माध्यम से प्राकृतिक और मानव-प्रेरित कारकों के प्रभावों की जांच क्षेत्रीय और वैश्विक स्तर पर की जा सकती है। Community Land Model version 5.0 (CLM5) एक अत्याधुनिक LSM है जिसका agroecosystems के अध्ययन में व्यापक उपयोग हुआ है। तथापि, CLM5 में भारतीय agroecosystems का चित्रण यथार्थपरक नहीं है, जिसके कारण उपज और सिंचाई के अनुकरण में प्रेक्षित आँकड़ों की तुलना में पक्षपात उत्पन्न होता है (Lombardozzi et al., 2020; Mathur and AchuthaRao, 2020)। भारतीय फसलों के बेहतर प्रतिनिधित्व की प्रमुख चुनौती मॉडल के अंशशोधन और मूल्यांकन हेतु सार्वजनिक रूप से उपलब्ध फसल-आँकड़ों की कमी है (Gahlot et al., 2020)।

यह शोध भारतीय agroecosystems में फसल की फेनोलॉजी और terrestrial fluxes की स्थान-कालिक परिवर्तनशीलता तथा प्रेक्षित विविधताओं में प्राकृतिक और मानव-जनित कारकों की भूमिका का परीक्षण करने का उद्देश्य रखता है। इसके विशिष्ट उद्देश्य निम्नलिखित हैं: (1) स्थल-स्तरीय प्रेक्षणों से भारतीय agroecosystems की फसल फेनोलॉजी को समझना, (2) इन प्रेक्षणों का उपयोग कर CLM5 मॉडल को सुधारना, (3) सुधारित CLM5 मॉडल का क्षेत्रीय स्तर पर मूल्यांकन करना, (4) CLM5 मॉडल की सहायता से भारतीय agroecosystems की फसल-वृद्धि परिमाणों एवं terrestrial fluxes की क्षेत्रीय और वार्षिक परिवर्तनशीलता तथा विभिन्न प्रेरक कारकों की भूमिका का अनुमान लगाना, तथा अंततः (5) सुधारित CLM5 मॉडल का भारतीय agroecosystems के प्रतिनिधित्व करने वाले अन्य अत्याधुनिक LSM से तुलनात्मक मूल्यांकन करना।

प्रथम, अध्याय 2 उद्देश्य (1) को संबोधित करता है जो भारतीय agroecosystems के सटीक अनुकरण में प्रमुख बाधा को हल करने का प्रयास करता है, अर्थात सार्वजनिक रूप से उपलब्ध फसल-आँकड़ों का अभाव, जो agroecosystems को मात्रात्मक रूप से समझने और मॉडल करने की हमारी क्षमता को बाधित करता है। इस शोध में भारत की दो प्रमुख फसलों—धान और गेहूँ—पर केंद्रित एक नवीन फसल-आँकड़ा संकलित किया गया। यह कार्य देशभर के कृषि विश्वविद्यालयों में संग्रहित मास्टर्स एवं डॉक्टोरल शोध प्रबंधों का डिजिटलीकरण कर किया गया। यह आँकड़ा 1970 से 2020 की अवधि को कवर करता है तथा 47 स्थलों और 147 फसल ऋतुओं के स्थल-स्तरीय प्रेक्षणों को सम्मिलित करता है। यह आँकड़ा CLM5 मॉडल के अंशशोधन का आधार बना। स्थल-स्तरीय प्रेक्षणों से प्राप्त पत्ती क्षेत्र सूचकांक (LAI) आँकड़े दर्शाते हैं कि धान की प्रमुख फसल ऋतु जून से अक्तूबर और गेहूँ की नवंबर से अप्रैल है। आँकड़ा विश्लेषण से प्रमुख अवलोकन यह हैं कि 1970–2020 की अवधि में धान और गेहूँ की उपज में सांख्यिकीय रूप से महत्त्वपूर्ण बढ़ती प्रवृत्ति देखी गई (धान के लिए 38 किग्रा/हेक्टेयर/वर्ष – 97.4% वृद्धि; गेहूँ के लिए

34 किग्रा/हेक्टेयर/वर्ष – 83% वृद्धि)। तथापि, इस अवधि में फसल ऋतु की लंबाई में कोई महत्त्वपूर्ण अंतर या प्रवृत्ति नहीं देखी गई। दिलचस्प रूप से गेहूँ की कटाई तिथियों में सांख्यिकीय रूप से महत्त्वपूर्ण बढ़ती प्रवृत्ति देखी गई; कटाई तिथि लगभग 14 दिन बढ़ गई ($p<0.1$)।

अध्याय 3 उद्देश्य (2) को संबोधित करता है, जहाँ CLM5 मॉडल को सुधारा गया और स्थल-स्तरीय प्रेक्षणों के विरुद्ध मूल्यांकन किया गया। CLM5 का डिफ़ॉल्ट संस्करण फसल की फेनोलॉजी और वृद्धि का अनुकरण करने में असफल रहा। इस कमी को दूर करने हेतु, इस शोध में संकलित भारतीय फसल आँकड़े का उपयोग कर □□□5 के फसल मॉड्यूल को बिंदु स्तर पर अंशशोधित एवं सत्यापित किया गया। फसल-प्रतिनिधित्व को दो-चरणीय दृष्टिकोण द्वारा सुधारा गया: (1) फसल वृद्धि-विशिष्ट फसल ऋतु और परिमाणों का अंशशोधन, (2) क्षेत्रीय विविधताओं का समावेशन, जो भारत जैसे विविध जलवायु वाले देश के लिए महत्त्वपूर्ण है। फसल ऋतु और परिमाणों के अंशशोधन में 30 संवेदी अनुकरण किए गए। प्रत्येक स्थल पर फसल के लिए बढ़ती डिग्री दिवसों की गणना हेतु प्रयुक्त आधार तापमान का मूल्यांकन किया गया ताकि क्षेत्रीय विविधताओं को सम्मिलित किया जा सके। फसल के लिए अक्षांश और आधार तापमान के बीच संबंध स्थापित करने हेतु 150 संवेदी प्रयोग किए गए। मॉडल का अंशशोधन स्थल-स्तरीय प्रेक्षणों की तुलना में फसल ऋतु, फसल शारीरिक परिमाण, फेनोलॉजी और उपज के अनुकरण को महत्त्वपूर्ण रूप से सुधारता है। गेहूँ और धान के LAI के लिए RMSE लगभग 50% कम हुआ। उपज अनुमान में गेहूँ और धान के लिए RMSE क्रमशः 40% और 60% कम हुआ। गेहूँ की फसल ऋतु लंबाई के लिए RMSE लगभग 75% कम हुआ।

अध्याय 4 उद्देश्य (3) को संबोधित करता है, जहाँ सुधारित CLM5 का क्षेत्रीय स्तर पर 2000–2014 की अवधि हेतु संख्यात्मक अनुकरणों के माध्यम से मूल्यांकन किया गया। इसके लिए उपज, सिंचाई की मात्रा और पैटर्न, तथा terrestrial fluxes को डिफ़ॉल्ट और सुधारित मॉडल द्वारा अनुकरणित कर बहुविध आँकड़ा स्रोतों से तुलना की गई। परिणाम दर्शाते हैं कि सुधारित CLM5 में धान की उपज हेतु RMSE 25% तथा GPP हेतु RMSE लगभग 50% कम हुआ। जल और ऊर्जा प्रवाहों में भी महत्त्वपूर्ण सुधार देखे गए। सिंचाई पैटर्न और सिंचित मात्रा अब प्रेक्षणों के अधिक निकट हैं। स्थल-स्तरीय और क्षेत्रीय मूल्यांकन से स्पष्ट है कि क्षेत्र-विशिष्ट फसल परिमाणों का उपयोग agroecosystems और क्षेत्रीय भूमि सतह प्रक्रियाओं के सटीक अनुकरण हेतु आवश्यक है।

अध्याय 5 उद्देश्य (4) को संबोधित करता है। इसमें 1970–2014 की अवधि को कवर करने वाले बहुदशकीय अनुकरणों का उपयोग कर भारतीय agroecosystems की उपज और terrestrial fluxes की क्षेत्रीय और वार्षिक परिवर्तनशीलता का परीक्षण किया गया। साथ ही चार संख्यात्मक प्रयोग किए गए ताकि दो प्राकृतिक और दो कृषि-प्रबंधन कारकों के प्रभावों का विश्लेषण हो सके। परिणाम दर्शाते हैं कि LAI, उपज, गेहूँ और धान की कुल शुष्क पदार्थ मात्रा, तथा उनके कार्बन प्रवाह 1970–2014 की अवधि में सांख्यिकीय रूप से महत्त्वपूर्ण बढ़ती प्रवृत्ति ($p<0.05$) दर्शाते हैं। प्रयोगों से स्पष्ट हुआ कि यह प्रवृत्ति मुख्यतः प्रबंधन प्रथाओं द्वारा संचालित है, जबकि CO_2 वृद्धि भी एक महत्त्वपूर्ण कारक है। नाइट्रोजन उर्वरीकरण का धान की उपज पर महत्त्वपूर्ण प्रभाव देखा गया, जो गर्म और आर्द्र ऋतु में उगाई जाती है। गेहूँ की उपज मुख्यतः सिंचाई से प्रभावित होती है, जो शुष्क और ठंडी ऋतु में उगाई जाती है। जलवायु के प्रतिकूल प्रभावों के बावजूद CO_2 वृद्धि और प्रबंधन प्रथाएँ (जैसे उर्वरीकरण और सिंचाई) इन चुनौतियों को कम कर फसल उपज में वृद्धि का कारण बनीं।

अध्याय 6 में CLM5 और इंटीग्रेटेड साइंस असेसमेंट मॉडल (ISAM) का तुलनात्मक विश्लेषण किया गया, जो भारतीय फसलों को अनुकरण हेतु अंशशोधित एवं मूल्यांकित किया गया एकमात्र अन्य LSM है (Gahlot et al., 2020)। इस शोध के दो प्रमुख लाभ हैं: (1) CLM मॉडल का अंशशोधन 14 स्थलों और 33 फसल ऋतुओं के आँकड़ों से किया गया; (2) इस अध्ययन में धान और गेहूँ दोनों का विश्लेषण किया गया जबकि Gahlot et al. (2020) ने केवल वसंत ऋतु के गेहूँ पर ध्यान केंद्रित किया। अध्याय 5 में किए गए संख्यात्मक प्रयोग ISAM मॉडल से भी दोहराए गए। परिणाम दर्शाते हैं कि CLM5 फसल ऋतु की लंबाई का अनुकरण अधिक सटीक रूप से करता है (RMSE = 14 दिन) जबकि ISAM में RMSE 28 दिन है। CLM5 में गेहूँ की उपज नाइट्रोजन उर्वरीकरण के प्रति अधिक संवेदनशील है, जबकि ISAM में यह सिंचाई के प्रति संवेदनशील है। CLM5 में AR/GPP अनुपात परिवर्तनशील है, जबकि ISAM में यह स्थिर रहता है और प्रेक्षणों की तुलना में कम है। कार्बन प्रवाहों पर जलवायु और प्रबंधन प्रभावों की जाँच से पता चलता है कि यद्यपि CLM5 और ISAM में NPP प्रवृत्तियाँ समान हैं, परंतु मानव प्रबंधन प्रभावों में काफ़ी भिन्नता है। CLM5 में नाइट्रोजन उर्वरीकरण प्रमुख प्रभाव डालता है, जबकि ISAM में सिंचाई का प्रभाव अधिक है। LSMs में भारतीय agroecosystems के प्रतिनिधित्व की वर्तमान स्थिति (CLM5 और ISAM) की तुलना से स्पष्ट है कि दोनों मॉडलों में सुधार की व्यापक संभावनाएँ हैं। CLM5 में वर्षा परिवर्तनशीलता को सम्मिलित कर बुवाई तिथियों के क्षेत्रीय अंतर का अनुकरण किया जा सकता है। ISAM में कार्बन-नाइट्रोजन चक्र को सुधारकर कार्बन प्रवाहों को प्रेक्षणों के अधिक निकट लाया जा सकता है।

यह प्रथम अध्ययन है जिसने भारतीय agroecosystems के शारीरिक परिमाणों और उनके terrestrial fluxes की दीर्घकालिक स्थानिक और वार्षिक परिवर्तनशीलता को एक सुदृढ़ अंशशोधित LSM द्वारा संबोधित किया है। इस कार्य का एक प्रमुख योगदान भारतीय फसल आँकड़ों को सार्वजनिक रूप से उपलब्ध कराना है ताकि मॉडल विकासकर्ता भारतीय फसलों के अनुकरण को बेहतर बना सकें, जो लंबे समय से एक प्रमुख बाधा रहा है।

Contents

Certificate i

Acknowledgments ii

Abstract iii

सारांश vii

List of Figures xvii

List of Tables xviii

Abbreviations xix

1 Introduction 1

1.1 Motivation 2

1.2 Objectives 6

2 Understanding Indian agroecosystems phenology using site-scale data 9

2.1 Introduction 10

2.1.1 The rich history of agricultural research in India 10

2.2 Methodology 11

2.2.1 The crop data for wheat and rice 11

2.3 Results 11

2.3.1 The details of crop data 11

2.3.2 The experiments and variables from data 13

2.3.3 The data used for model calibration and validation 14

2.4 Summary and Conclusions 15

3 CLM5 crop module calibration and evaluation at site-scale 18

3.1 Introduction 19

3.1.1 Community Land Model (CLM5) 20

3.2 Methodology 24

3.2.1 Site-scale simulations 24

3.2.2 Parameters impacting the crop phenology in CLM5 25

3.2.3 Improvements in CLM5_Mod1 . . . 25
3.2.4 Mod2 case parameters: varying base temperature by latitude . . . 29
3.2.5 Evaluation Metrics . . . 33
3.2.6 Model Evaluation at the site-scale . . . 34
3.3 Results . . . 34
3.3.1 Wheat . . . 34
3.3.2 Rice . . . 40
3.4 Summary and Conclusions . . . 44

4 Evaluation of improved CLM5 at regional scale 45
4.1 Introduction . . . 46
4.2 Methodology . . . 46
4.2.1 Regional scale simulations . . . 46
4.2.2 Model evaluation at regional scale . . . 46
4.3 Results . . . 48
4.3.1 Yield . . . 48
4.3.2 Irrigation . . . 50
4.3.3 GPP and LAI . . . 51
4.3.4 Latent heat flux . . . 55
4.3.5 Sensible heat flux . . . 57
4.4 Summary and Conclusions . . . 60

5 Spatio-temporal variability of terrestrial fluxes in Indian agroecosystems and the role of drivers 62
5.1 Introduction . . . 63
5.2 Methodology . . . 63
5.2.1 Details of experiment . . . 63
5.2.2 Evaluation of crop physiological parameters . . . 64
5.2.3 Evaluation of terrestrial fluxes . . . 65
5.2.4 Impact of drivers on crops and their terrestrial fluxes . . . 66
5.3 Results . . . 67
5.3.1 Evaluation of crop physiological parameters . . . 67
5.3.2 Evaluation of carbon fluxes . . . 67
5.3.3 Evaluation of water and energy fluxes . . . 71

5.3.4 Trends in carbon fluxes . . . 73
5.3.5 Trends in crop physiological parameters . . . 79
5.3.6 Trends in water and energy fluxes . . . 83
5.3.7 Annual Yield . . . 85
5.3.8 Seasonal carbon fluxes . . . 85
5.3.9 Water and energy fluxes from global studies . . . 87
5.4 Summary and conclusions . . . 87

6 Evaluation of the improved CLM5 with the state-of-the-art LSM 90
6.1 Introduction . . . 91
6.1.1 Process representation in LSMs . . . 91
6.1.2 Differences in crop modules in CLM5 and ISAM . . . 92
6.1.3 Impact of external drivers on crops . . . 94
6.2 Methodology . . . 95
6.2.1 Experiments conducted . . . 95
6.2.2 Evaluation of crop physiology and carbon fluxes . . . 95
6.3 Results . . . 96
6.3.1 Evaluation against site-scale observations . . . 96
6.3.2 Annual wheat crop yield . . . 97
6.3.3 Comparison of carbon fluxes . . . 100
6.3.4 Impact of drivers on carbon fluxes . . . 103
6.4 Summary and conclusion . . . 104

7 Conclusions and discussions 107
7.1 Summary . . . 108
7.2 Conclusions and future-scope . . . 109

References 112

A Appendix 120
A.1 Carbon fluxes in CLM5 for C3 crops . . . 120
A.2 Code changes made to CLM5 source code in this study . . . 120
A.2.1 Code added in *CNPhenologyMod.F90* module . . . 120
A.2.2 Code changes in *CropType.F90* module . . . 121

Curriculum Vitae 122

List of Figures

1.1 Mechanistic representation of crop growth in a model and the natural and human management factors impacting crop growth (Chenu et al., 2017). 3

1.2 Schematic showing the evolution of LSMs since their inception (Fisher and Koven, 2020). 4

2.1 The number of growing seasons data at each site for (a) rice crop, (b) wheat crop, and (c) the location of the site and number of growing seasons at each site available in the data (Varma et al., 2024). 12

2.2 The number of growing seasons data in each decade for (a) rice crop and (b) wheat crop. 13

2.3 Analyzing the observational data for (a) LAI, (b) yield, and (c) growing season length in rice and wheat crops. 14

2.4 Analyzing the observational data for (a) sowing date and (b) harvest date in rice and wheat crops. 14

2.5 Location of sites used in the current study for calibrating and validating the major Indian crops (1) wheat and (2) rice. 15

3.1 Comparing the CLM5 default simulation of crop growth against the site-scale LAI observations and irrigation data (BCM/day—Billion Cubic Meters/day) from Biemans et al. (2016). 26

3.2 The overall bias in the site scale simulations during the sensitivity study of base temperature (x-axis) for (a) wheat and (b) rice. The y-axis shows the overall bias (mean of absolute bias in LAI, yield, and growing season length). 31

3.3 Latitudinal variation in base temperature calibration (a) wheat and (b) rice. . 32

3.4 Site scale LAI simulated by three versions of CLM5 against observations for wheat. 35

3.5 Site-scale CLM5 performance against observations (1) wheat; (2) rice. Crop variables compared are (a) max. LAI during the growing season, (b) yield, and (c) growing season length. 37

3.6 Comparing site scale yield of wheat simulated by CLM5 against the observations. 39

3.7 Site scale LAI simulated by three versions of CLM5 against observations for rice. 42

3.8 Comparing site scale yield of rice simulated by CLM5 against the observations. . 43

4.1 Domain used in simulations of CLM5 for regional experiments. 47

4.2 Yield estimates of (a) wheat and (b) rice by (1) EarthStat 2005, and (2-4) difference in yield between CLM5 (mean 2003-2007) versions and EarthStat data. 49

4.3 Comparing the total yield of (a) wheat and (b) rice against the FAO data from 2000 to 2014. 50
4.4 Comparison of water added through irrigation simulated by CLM5 and water demand data from Biemans et al. (2016). 51
4.5 Crop areas of wheat and rice used in CLM5. 51
4.6 Spatial variation of GPP simulated by CLM5 against MODIS data. The data shows the monthly GPP averaged over 2000-2014. 53
4.7 Spatial variation of LAI simulated by CLM5 against MODIS data. The data shows the monthly LAI averaged over 2000-2014. 54
4.8 Comparison of MODIS (a) GPP and (b) LAI against CLM5 simulations over the period 2001 to 2014. 55
4.9 Comparing CLM5 simulated (a) LAI, (b) GPP, (c) LH, and (d) SH against observations. The data used here is the monthly mean from 2000 to 2014. 56
4.10 Comparison of FLUXCOM latent heat product against the three setups of CLM5. Data shown here is the monthly average taken over the period 2001 to 2014. 58
4.11 Comparison of FLUXCOM sensible heat product against the three setups of CLM5. Data shown here is the monthly average taken from 2001 to 2014. 59
4.12 Comparison of FLUXCOM(a) latent heat and (b) sensible heat with CLM5 three setups over the period 2001 to 2014. 60
5.1 Comparing CLM5 simulated rice (a.1, a.2, and a.3) and wheat (b.1, b.2, and b.3) crop phenology against the observations from 1970-2014 68
5.2 Comparison of yield simulated by CLM5 for major Indian crops (a) rice, and (b) wheat against the FAO observations 68
5.3 Comparing CLM5 carbon fluxes against observations of wheat at IARI, New Delhi, from Reddy et al. (2023) (a, b, and c), (d) wheat cropland at Meerut, from Patel et al. (2011), (e) rice cropland at Cuttack from Bhattacharya et al. (2013), and (f) scatter plot of all monthly mean carbon flux observations against the CLM5 simulations 69
5.4 Spatial plots of a. GPP, b. TER, and c. NEP from 1. FLUXCOM, and 2. CLM5 for the rice-growing regions. The difference between the CLM5 and FLUXCOM data is shown in column 3. Period: 1970-2014 71

5.5 Spatial plots of a. GPP, b. TER, and c. NEP from 1. FLUXCOM, and 2. CLM5 for the wheat-growing regions. The difference between the CLM5 and FLUXCOM data is shown in column 3. Period: 1970-2014 72
5.6 Comparing FLUXCOM energy fluxes against the CLM5 simulated data over the rice-growing region during 1970-79 and 2005-14. The second and fourth columns show the differences between CLM5 and FLUXCOM data, r-value, and MAB computed over ten years 73
5.7 Comparison of monthly water and energy fluxes simulated by CLM5 CTRL run against the FLUXCOM data during the rice-growing season (July to October) . 74
5.8 Comparing FLUXCOM energy fluxes against the CLM5 simulated data over the wheat-growing region during 1970-79 and 2005-14. The second and fourth columns show the differences between CLM5 and FLUXCOM data, r-value, and MAB computed over ten years 75
5.9 Comparison of monthly water and energy fluxes simulated by CLM5 CTRL run against the FLUXCOM data during the wheat-growing season (December to March) 76
5.10 The mean of annual carbon fluxes in rice-growing regions (GPP, AR, and NPP) in the CTRL simulation (a.1, b.1, and c.1) from 1970 to 2014 is shown by a dark grey line. Columns 2 to 5 show the impact of each driver. 77
5.11 Spatial trend observed in carbon fluxes (GPP, AR, and NPP) in rice-growing regions in the CTRL simulation (a.1, b.1, and c.1, respectively). Columns 2 to 5 show the impact of each driver. Regions stippled implies the trend is statistically significant at $p<0.05$. 77
5.12 The mean of annual carbon fluxes in wheat-growing regions (GPP, AR, and NPP) in the CTRL simulation (a.1, b.1, and c.1) from 1970 to 2014 is shown by a dark grey line. Columns 2 to 5 show the impact of each driver. 78
5.13 Spatial trend observed in carbon fluxes (GPP, AR, and NPP) in wheat-growing regions in the CTRL simulation (a.1, b.1, and c.1, respectively). Columns 2 to 5 show the impact of each driver. Regions stippled implies the trend is statistically significant at $p<0.05$. 79
5.14 The crop parameters of rice (max.LAI, yield, and dry matter) in the CTRL simulation (a.1, b.1, and c.1) from 1970 to 2014 is shown by a dark grey line. Columns 2 to 5 show the impact of each driver. 80

5.15 Spatial trend observed in crop phenology of rice-growing regions in the CTRL simulation (a.1, b.1, and c.1). Columns 2 to end show the impact of each driver. Regions stippled implies the trend is statistically significant at $p<0.05$. 81
5.16 The crop parameters of wheat (max.LAI, yield, and dry matter) in the CTRL simulation (a.1, b.1, and c.1) from 1970 to 2014 is shown by a dark grey line. Columns 2 to 5 show the impact of each driver. 82
5.17 Spatial trend observed in crop phenology of wheat-growing regions in the CTRL simulation (a.1, b.1, and c.1). Columns 2 to end show the impact of each driver. Regions stippled implies the trend is statistically significant at $p<0.05$. 82
5.18 The mean of annual water and energy fluxes in rice-growing regions (GPP, AR, and NPP) in the CTRL simulation (a.1, and b.1) from 1970 to 2014 is shown by a dark grey line. Columns 2 to 5 show the impact of each driver. 84
5.19 The mean of annual water and energy fluxes in wheat-growing regions (GPP, AR, and NPP) in the CTRL simulation (a.1, and b.1) from 1970 to 2014 is shown by a dark grey line. Columns 2 to 5 show the impact of each driver. . . . 84
5.20 Impact of management drivers on crop yield of (a) wheat and (b) rice. 85
6.1 Comparing the climate drivers and management drivers in CLM5 and ISAM models (CTRL - S_{driver} in (a and d); and absolute values used in model in (b and c)). 95
6.2 Evaluation of site scale wheat simulation of LAI by CLM5 and ISAM. The numbers on the right top corner are the growing season length simulated by CLM5 (blue), ISAM (orange), and from observations (black). 98
6.3 Yield estimates of wheat by (a) EarthStat 2005, and (b and c) difference in yield simulated by CLM5 and ISAM (mean 2003-2007) against EarthStat data. 99
6.4 Evaluation of country scale yield simulation of wheat by CLM5 and ISAM against FAO annual yield estimates and the impact of management drivers on yield simulated by CLM5 and ISAM, rolling mean with a period of 5. Table in the figure shows the mean yield over 2010-14 period simulated by CLM5 and ISAM in CTRL, S_{Irrig} and S_{NFert} simulations. 100

6.5 Site scale monthly mean carbon fluxes (GPP, TER, and NEP) at (a-c) IARI, New Delhi, and (NEP) at (d) Meerut against the monthly mean fluxes in the grid cell having the site simulated by CLM5 (blue) and ISAM (brown). Panel (e) shows the scatter plot of carbon fluxes simulated at sites against their monthly model estimations. Table shows the Pearson's r, MAB and RMSE of CLM5 and ISAM. 101
6.6 Regional scale mean growing season carbon fluxes (a) GPP, (b) AR, and (c) NPP in wheat growing regions simulated by CLM5 and ISAM. 103
6.7 The ratio of AR/GPP in CLM5 and ISAM models. 103
6.8 Comparison of the impact of climate and human management drivers in carbon fluxes, (a) GPP, (b) AR, and (c) NPP, simulated by (1) CLM5 and (2) ISAM over the past 40 years. 104

List of Tables

2.1 Site-scale data from Varma et al. (2024) used for calibrating and validating the CLM5 model . . . 16
3.1 Sensitivity experiment values for CLM5_Mod1 case. . . . 27
3.2 Parameter values for wheat and rice in the CLM5 crop module . . . 28
3.3 Base temperature at lowest bias observed during the sensitivity studies for wheat and rice . . . 32
3.4 Latitudinal variation parameters for wheat and rice . . . 33
3.5 Evaluation of wheat and rice across three CLM5 setups at site scale . . . 41
4.1 Evaluation of CLM5 simulations at the regional scale against MODIS (LAI and GPP) and FLUXCOM (LH and SH) data. . . . 56
5.1 Description of CLM5 simulations conducted in this study and the constraints used to extract the impact of individual drivers. . . . 64
5.2 Comparing GPP simulated by CLM5 for rice and wheat against global studies. . . 86
6.1 Description of ISAM simulations conducted in this study and the constraints used to extract the impact of individual drivers. . . . 97
6.2 Comparison of seasonal carbon fluxes simulated by CLM5 and ISAM against various studies. Note: CLM5_Rf is rainfed CLM5 wheat and CLM5_Ir is irrigated CLM5 wheat. . . . 102
6.3 Comparison of carbon fluxes against various studies . . . 102

Abbreviations

AD mode	Accelerated Decomposition mode
Agro-IBIS	Agroecosystem Integrated Biosphere Simulator
AR	Autotrophic Respiration
BGC-Crop	Biogeochemistry-Crop module in CLM
CFT	Crop Functional Type
CLM5	Community Land Model version 5.0
CTRL	CLM5 control simulation
DM	Total Dry Matter
GDD	Growing Degree Days
GPP	Gross Primary Productivity
GSWP3	Global Soil Wetness Project Phase 3
GY	Grain Yield
HR	Heterotrophic Respiration
IGP	Indo-Gangetic Plains
ISAM	Integrated Science Assessment Model
KGE	Kling Gupta Efficiency
LAI	Leaf Area Index
LH	Latent Heat Flux
LSM	Land Surface Model
MAB	Mean Absolute Bias
NEP	Net Ecosystem Respiration
NPP	Net Primary Productivity
PFT	Plant Functional Type
r	Pearson's correlation coefficient
RMSE	Root Mean Square Error
SH	Sensible Heat Flux
S_{Clim}	CLM5/ISAM simulation forced with 1901-1920 atmospheric data
S_{CO_2}	CLM5/ISAM simulation forced with pre-industrial [CO_2]
S_{Irrig}	CLM5/ISAM simulation with no irrigation to crops
S_{NFert}	CLM5/ISAM simulation with no nitrogen fertilization to crops
TER	Total Ecosystem Respiration