Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Scienze e Gestione dei Cambiamenti Climatici
Ciclo 29, Anno di discussione 2017

Impacts Of Climate Change And Variability On Crop Yields Using Emulators And Empirical Models

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: SECS-P/06, SECS-S/01, AGR/02. Tesi di Dottorato di Malcolm N. Mistry, matricola 956110

Coordinatore del Dottorato

Prof. Carlo Barbante

Tutore del Dottorando

Prof. Francesco Bosello

Co-tutori del Dottorando
Dr. Enrica De Cian
Prof. Silvio Gualdi
Prof. Ian Sue Wing

Impacts Of Climate Change And Variability On Crop Yields Using Emulators And Empirical Models

Malcolm N. Mistry

Tutor: Prof. Francesco Bosello

co-tutors: Dr. Enrica De Cian, Prof. Silvio Gualdi, Prof. Ian Sue Wing

PhD Program Coordinator: Prof. Carlo Barbante

A thesis submitted to the Department of Economics and Graduate School (Ca' Foscari University, Venice)
in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in Science and Management of Climate Change (29th Cycle)

December, 2016

Venice, Italy

About the author

Prior enrolling in the PhD program in 2013, Malcolm Mistry completed his Bachelor degree in Information Technology from India, and a Master of Science (MSc) degree in Weather Climate and Modelling from the University of Reading, U.K. His MSc dissertation focused on the Intraseasonal Variability of Indian Summer Monsoon (ISM) in observed and General Circulation Model (GCM). Malcolm is also a qualified navigating officer of merchant and offshore vessels, having sailed earlier onboard research and seismic survey ships engaged in data acquisition of sea bed profiles for offshore oil and gas industry.

Having a background in meteorology and climate modeling, his core research interests include assessing the impacts of climate change and variability on agriculture and health. He has experience in handling large datasets of Earth System Models (ESMs), numerous scientific data management/analyses tools and bridging the link between climate and impact modelers. But perhaps his biggest passion and strength lie in communicating the complex science of climate change and its' implications to the wider public. As part of his initiative, Malcolm often creates an awareness of human attribution to climate change as an invited speaker to public schools and other educational organizations.

Among his other experiences, Malcolm has previously worked as a Research Associate at the Department of Earth Science, Barcelona Supercomputing Centre (Barcelona, Spain), where he provided technical support for the development of an Earth System Modelling Framework (ESMF). He is presently engaged as a Junior Researcher at the Fondazione Eni Enrico Mattei (FEEM), based in Venice, Italy. His PhD thesis focuses on applying econometric methods to build a multi-crop model emulator for predicting impacts of climate change on agriculture.

In memory of some loved ones who have left a void in my life. My late father, who would have been the first person (back home) to read this thesis. Next, my dear friend Dr. Chris Bell (Chris), whose ability to explain complex problems with simplified models played a huge role in developing my passion for modeling. And finally my childhood nanny Jivi (who passed away while writing this thesis), who would always ask me if I ever planned to stop studying.

Without their blessings, I wouldn't be able to complete this thesis.

"The woods are lovely, dark and deep.

But I have promises to keep,

And miles to go before I sleep,

And miles to go before I sleep."

……Robert Frost (1874-1963)

Abstract (English)

The thesis assesses impacts of climate change and variability on regional and global crop yields using econometric approaches to analyze global gridded data. Using a large dimension panel data of six Global Gridded Crop Models (GGCMs) for four rainfed crops (maize, rice, soybeans and wheat) an emulator suitable/amenable of being integrated into Integrated Assessment Models (IAMs) is built. The performance of the emulator is evaluated against observational-based, empirical models at regional scale by building a statistical model calibrated on historical observed crop yields data for United States (U.S.) counties. Chapter 1 provides the background of existing research methodologies in agronomic literature. The gaps in existing research and scope for research are laid down as motivation and objectives of the research that follows in the subsequent chapters. Chapter 2 discusses the data, methodology and framework used in the construction of a simple statistical emulator of the response of crops to weather shocks simulated by crop models. To facilitate the integration of the emulator into IAMs, the simplest model using a base specification of linear fixed effect with time trend interactions is developed. Chapter 3 investigates modifications to the base specification with a series of robustness checks exploring the suitability of an additional predictor variable, the stratification of coefficients geographically by groups of Agro-Ecological Zones (AEZs); and most importantly, the role of spatial dependence in variables by applying a spatial model. Chapter 4 compares the performance of the statistical emulator calibrated on crop model results, with an empirical models of crop responses based on historical data. The comparison focuses on U.S. counties. The base specification from Chapter 2 together with historical observed data from the U.S. Department of Agriculture (USDA), are utilized in an inter-comparison exercise for divergence in results and subsequent implications. Collectively, the three chapters (2-4) address several important questions: (1) what do reduced-form statistical response surfaces trained on crop model outputs from various simulation specifications look like; (2) do model-based crop response functions vary systematically over space (e.g., crop suitability zones) and across crop models?, (3) how do model-based crop response functions compare to crop responses estimated using historical observations? and (4) what are the implications for the characterization of future climate risks? Chapter 5 concludes the thesis providing a summary of key contributions and suggestions for future work.

Abstract (Italian)

La tesi valuta gli impatti dei cambiamenti climatici e della variabilità climatica sulla produttività agricola a scala regionale e globale analizzando dati ad alta risoluzione spaziale con metodi econometrici. La tesi utilizza dati provenienti da sei modelli globali delle rese agricole per quattro coltivazioni non irrigate (mais, riso, soia, e grano) per costruire un emulatore da integrare in modelli di valutazione integrata (IAMs). La prestazione dell'emulatore statistico è valutata su scala regionale utilizzando modelli empirici basati su osservazioni storiche per gli Stati Uniti.

Il Capitolo 1 fornisce il contesto della ricerca esistente e descrive le metodologie disponibili nell'ambito dell'agronomia. Introduce la motivazione e gli obiettivi della ricerca sviluppata nei capitoli successivi. Il Capitolo 2 discute i dati, la metodologia usata per sviluppare un semplice emulatore statistico della funzione di risposta delle rese agricole a shock meteorologici simulati da modelli di processo. Per facilitare l'integrazione dell'emulatore in modelli IAMs, questo capitolo testa un modello semplice ad effetti fissi con l'interazione con trend temporali. Il Capitolo 3 esplora delle varianti del modello base che esplorano 1) altre variabili esplicative 2) variazioni geografiche in base a diverse aree agronomiche (Agro-Ecological Zones, AEZs), 3) il ruolo della dipendenza spaziale nei dati. Il Capitolo 4 confronta la performance dell'emulatore statistico calibrato sui dati dei modelli di processo con dei modelli empirici basati su dati storici. Il confronto analizza i dati per gli Stati Uniti. Si basa sul modello base sviluppato nel Capitolo 2 e dati storici per gli Stati Uniti dal Dipartimento dell'Agricoltura (USDA).

Nel loro insieme i tre capitoli 2-4 affrontano diverse importanti domande: 1) come si caratterizzano le funzioni di risposta in forma ridotta stimate a partire da dati generati da modelli di processo 2) come queste variano geograficamente e in base al modello che genera i dati 3) come queste differiscono rispetto a funzioni di risposta stimate a partire dai dati osservati storicamente e 4) quali sono le implicazioni per l'analisi del rischio climatico. Il Capitolo 5 conclude la tesi con un riassunto dei contributi chiave e suggerimenti per lavori futuri.

Acknowledgements

"The whole art of teaching is only the art of awakening the natural curiosity of young minds for the purpose of satisfying it afterwards" ~ **Anatole France (1844 – 1924)**

The journey to complete this PhD thesis has been long and cumbersome (as for most students). Yet, this journey has been both stimulating and exciting. Without certain key people involved all along, I cannot even imagine how I could have accomplished the research undertaken in this thesis.

My sincere gratitude to all my supervisors, who have shown remarkable patience in dealing with me all along. The stimulating discussions and innovative ideas with all of them helped me more than anything to enjoy my research. Coming from a non-statistical background, Dr. Enrica De Cian's initial introduction to econometrics and impact models intrigued me. Her passion and energy in spite of a million other engagements is simply unmatched. Special mention must be made of my external supervisor (Prof. Ian Sue Wing), whom ironically till the time of writing the thesis I haven't met. In spite of the trans-Atlantic distance, Ian's persistence in encouraging me to improve my work is probably the biggest factor (next to my kids) in losing my sleep. Yet, the extra push to cross the line has been worth all the efforts. My sincere thanks to Prof. Bosello for arranging my research work and additional funding at FEEM. Your immense background on adaptation and economic modeling helped me to structure the research better from an application tool. Prof Gualdi, without your crystal clear explanation of climate models and the inherent limitations in modeling processes, I wouldn't be able to quantify the unknowns better in my work.

Would also like to thank my two external reviewers (Profs. Karen Fisher-Vanden and Alessandro Olper) who provided valuable suggestions in structuring the final thesis better. Christoph Muller and Joshua Elliott (AgMIP project), for providing most of the finer details of the data used in my study. And all my department professors; especially our PhD coordinator (Prof. Carlo Barbante) for constantly taking our feedback in improving the layout of the PhD program, as well as paying special attention to the challenges faced by international students.

My entire family; but in particular my wife, my mother and my mother-in-law who have made a number of sacrifices in ensuring I wasn't distracted by my children during the PhD. Zubina (my wife), without your moral support all along, I could have dropped out of the PhD program in the very first year. My two beautiful children (Jaden and Faye) who have made it most challenging for me to focus on work at home; the ability to write codes with remote control cars trespassing my papers is owed to you.

This research work wouldn't be possible without the funding from Ca' Foscari University of Venice (through the Italian Ministry of Public Education), and FEEM. The IT staff at Boston University also need to be thanked for facilitating my work on their supercomputer.

Finally, I would like to thank admin staff (especially Federica our PhD secretary, and Monica at FEEM), colleagues in FEEM, and in particular my classmates (Fabio, Michele, Francesca and Arthur). Without your support and feedback in internal seminars, as well as in dealing with the bureaucracy in a foreign land; completing this thesis would remain a distant dream.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background and motivation

There is widespread concern that trends and variability in weather induced by climate change will detrimentally affect global agricultural productivity and food supplies (Parry *et al* 1999, 2004). Reliable quantification of the risks of negative impacts at regional and global scales is a critical research need, which has so far been met by forcing state-of-the-art global gridded crop models with outputs of Earth System Model (ESM) simulations in exercises such as the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP)-Fastrack (Warszawski *et al* 2013). Together with Integrated Assessment Models (IAMs), the predictions for future regional and global food production form the backbone in our understanding of the pressures exerted by projected climate change and variability. Over the last few decades taking advantage of the exponential rise in computing capacity and data availability, the modelling community has made rapid strides in the understanding of discrete scale processes involved in crop development (e.g., Elliott *et al* 2014). Yet, both gaps and scope for research remain which lays the framework of my research thesis.

1.1.1 Key tools used in modelling crop yield responses to climate

Responses of crop yields to climate are broadly examined using (i) process-based, also referred to as 'mechanistic' models (e.g., Rosenzweig *et al* 2014, Elliott *et al* 2014) or (ii) empirical approaches (e.g., Lobell and Burke 2010, Schlenker and Roberts 2009, Urban *et al* 2015). Mechanistic models rely on numerical simulation of the key processes associated with crop phenology and require extensive input data on management, soil conditions, cultivar amongst others (Rosenzweig *et al* 2014, Lobell and Burke 2010). On the contrary, statistical techniques utilize historical data of crop yield and associated parameters in conjunction with weather variables, to estimate the underlying relationship between the crop growth and climate (Lobell and Burke 2010).

The strengths and limitations of the two approaches are equally well recognized in literature. The strengths of a statistical approach often emphasized are its limited reliance on field calibration data, with lower computational requirements compared to process-based approach (Lobell and Burke 2010) . In contrast, two of its shortcomings often subject of criticism and skepticism are the inability to encapsulate adaptation and assumptions of stationarity, both are partly true depending on the choice of regression

methodology (Lobell *et al* 2011). Moreover, due to near perfect collinearity of gently increasing carbon dioxide (CO_2) concentrations with other time trends (such as technology advancements) (Schlenker and Roberts 2009, Sue Wing *et al* 2015), empirical estimates often find it difficult to explicitly account for CO_2 fertilization effects (CFE).

1.1.2 Framework of recent global crop models inter-comparison exercises

The potential impacts of climate change and variability on agriculture and the consequent spillover effects on global food production and economy have been an active area of research in recent decades (e.g., Parry *et al* 2004, Lobell *et al* 2011, Challinor and Wheeler 2008, Porter *et al* 2014, Nelson *et al* 2014b). A number of crop modelling tools such as process based crop models, agro-ecosystem models and statistical models have been utilized for furthering our understanding of key processes involved in crop yield responses to agro-meteorological element (such as temperature, precipitation, soil conditions).

In recent years, cumbersome exercises to quantify the impacts of climate change on global crop yields were initiated within the frameworks of the Agricultural Model Intercomparsion and Improvement Project (AgMIP[1]—Rosenzweig *et al* 2013, 2014, Elliott *et al* 2014) and ISI-MIP (Warszawski *et al* 2013). Unlike earlier studies that focused on the potential impacts of climate change on agriculture at a coarser resolution and often at regional scales (such as Parry *et al* 2004, Challinor and Wheeler 2008, Bassu *et al* 2014), ISIMIP-Fastrack (ISIMIP-FT) is the first to attempt the same at a global-gridded scale, encompassing different Global Gridded Crop Models (GGCMs), crops and scenarios, with a systematic harmonization of simulation protocol (Rosenzweig *et al* 2014). Bringing together a combination of different ESMs and GGCMs, ISI-MIP makes a concrete effort in quantifying the uncertainties of climate change impacts on global crop yields.

Notwithstanding the vast array of ESMs-GGCMs combinations used in the ISIMIP-FT, the lack of a pure statistical crop model[2] participating in the simulation exercise is the primary motivating factor for this research. As highlighted in Lobell and Burke 2010, Oyebamiji *et al* 2015 and Blanc and Sultan 2015, a tool capable of replicating yields for a variety of important crops, from a wide ensemble of heterogeneous crop models, under different climate change scenarios and in particular, at a global-gridded scale, is

[1] AgMIP is the umbrella project that coordinated and provided simulation data of crop models to the agriculture sector of ISI-MIP. A quick overview can be found here https://www.agmip.org/ag-grid/ggcmi/ and https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/about

[2] Although there is a hybrid GGCM (PEGASUS) participating in ISIMIP-FT; only one other recent study 'Blanc and Sultan (2015)' implements a statistical emulator using ISIMIP-FT data. However, they focus only on crop maize and as shown in my research, the methodology and comprehensive set of robustness checks employed here enables the emulator to be applied across broad range of climate impacts assessment.

highly essential. The multi-crop emulator developed as part of this study would contribute to this need identified by the research community, as well as to a set of broader objectives outlined below.

1.2 Objectives and research questions

I begin by reviewing the existing state of knowledge and empirical methodology used in assessments of climate change impacts on agriculture. I focus on econometric methods commonly employed to assess response of crop yields to climate change, the uncertainties associated with estimation methods and gaps in existing empirical literature. As vital contribution to empirical agronomic work, I develop an emulator capable of capturing the impacts of weather shocks on crop yields via statistically-estimated, reduced-form response surfaces of GGCMs. Though simple, the IAM community needs such a parsimonious emulator that could be easily combined with multiple realizations of future climate to develop an analysis of future climate risk on crops. The development of the emulator broadly sets the outline for not only developing new research methodologies, but also for expanding existing impact assessment approaches. The doctoral research is organized around the following research questions:

1. What does the most basic emulator of the heat and moisture effects of climate change look like, controlling completely for the confounding effects of CO_2 by basing it on no-CFE GGCMs' runs?
2. Can the emulator replicate GGCMs' yield responses to future climate for multiple crops, considering the heterogeneity across the GGCMs, crops and simulation setups?
3. At what spatial and temporal scales does the emulator show strength and weakness and whether crops' response to climate vary geographically?
4. How do the coefficient estimates from an empirical model trained on historical observed crop yields compare with an identical specification of the emulator?
5. What are the underlying meta-parameters that contributed most to the divergence in GGCMs' yield responses?

To address these questions, I utilize data from two different sources. Since agriculture is largely influenced by regional climate, soil conditions and management practices; the study uses global gridded data at fine scale resolution from ISIMIP-FT. In addition, historical observed data covering U.S. counties from the U.S. Department of Agriculture (USDA) are also used for my research.

1.3 Outline of Doctoral thesis

This thesis is organized as three core chapters (excluding this introductory chapter), and an additional final chapter summarizing the key contributions of the research and proposing scope for future work. Each chapter examines a different methodology (econometric specification) as part of stability and robustness tests.

My first core chapter (Chapter 2) discusses the framework and methodology involved in the building of a robust statistical emulator for crop yield responses to climate change. Utilizing crop yields for major rainfed crops (maize, rice, soybeans and wheat) and climate data from ISIMIP-FT, I develop a simple and flexible emulator that can be rapidly integrated with IAMs. The reduced-form response functions to temperature (T) and precipitation (P) are characterized and assessed for their stationarity across time and different models—predictive power, and potential diagnostic utility. Here the emulator is built independently as six GGCMs and in addition, as a multi-GCCM calibrated on a large merged panel of six GGCMs. This offers flexibility in application as both ensemble of individual emulators and/or a multi-GGCM emulator The parsimony is at the core of design implementation. Yet, its holds considerable potential as a diagnostic methodology to elucidate uncertainties in the processes simulated by GGCMs, and to support the development of climate impact inter-comparison exercises within the IAM community. This chapter is currently being finalized for submission to Environment Research Letters (ERL) and is co-authored by Enrica De Cian and Ian Sue Wing.

Chapter 3 addresses the broad range of robustness tests on the emulator built in chapter 2. Apart from testing the suitability of an additional predictor variable -Vapor Pressure Deficit (VPD)-, two other regression specifications are examined for their performance and practical application in comparison to the simple base specification used in chapter 2. To examine heterogeneity of crop yield response across geographic regions, I group the eighteen AEZs (Lee *et al* 2005) which are a combination of a climate region and growing period length, into six broader zones. The covariates are then allowed to interact with the six AEZs giving coefficient estimates stratified by each AEZ. The second regression specification takes advantage of recent developments in spatial panel econometrics (SPE). I utilize a Spatial Lag of X (SLX) model that accounts for dependence in the predictor variables. This chapter is under preparation for submission to Journal of Agriculture and Forest Meteorology, and is co-authored by Enrica De Cian and Ian Sue Wing.

Chapter 4 shifts from global to regional perspective, focusing on U.S. counties. For this, I extract U.S. county data from my GGCMs' panel data to match the spatial scales of the historical observed data from USDA. I then evaluate how well GGCMs' internal representations of crop growth compare with the

responses of yields to heat and moisture to weather variation under the current climate estimated by econometric model trained on observations. I provide first glimpse into the origins and implications of the divergence in yield impacts, both among GGCMs, and between GGCMs and historical observations. Furthermore, I assess the implications of the differences between GGCMs' aggregated responses under future climate change scenario. This chapter is submitted to the special issue of ERL ('An Inter-method Comparison of Climate Change Impacts on Agriculture'), and is co-authored by self, Enrica De Cian and Ian Sue Wing.

To conclude, Chapter 5 provides a summary of the contributions made by the thesis, discusses the caveats, and offers suggestions for future work.

Chapter 2: Robust Statistical Emulation of Process Model Crop Yield Responses to Climate Change

Preface

Attention given to the impacts of climate change on agriculture at both regional and global scales has gathered momentum in recent years. With the projected rise in world population, mitigating the effects of climate change on agriculture and identifying the pressures on adaptation to future crop productivity require wide-ranging modelling tools. Historically, majority of the work assessing impacts of climate change on agriculture have focused on coarser resolutions (districts, counties and usually countries). However, recent advancements in computing capacities have led interest shifting to fine scale resolutions, typically at global gridded resolution.

This chapter discusses the design, calibration and implementation of a statistical emulator, at a global gridded fine scale resolution. An Emulator is a Surrogate Model (Fast Statistical Approximation or a Cheap Computational Model) that can be rapidly coupled with the output of ESMs.

The proceedings of this chapter are being prepared for submission to Environment Research Letters (ERL), and are co-authored by Enrica De Cian and Ian Sue Wing. Baring few minor superficial changes to the figures and text, the manuscript is largely unchanged from the version of the paper under preparation. I designed and performed research, analyzed the data and wrote the paper. Enrica De Cian and Ian Sue Wing provided scientific input. All co-authors are involved in the revision of the final text for submission to ERL.

Main Text

Abstract

A rapidly growing literature employs historical observations or pseudo-data generated by Global Gridded Crop Models (GGCMs) to empirically estimate reduced-form crop yield responses to meteorology. The resulting fitted response surfaces, when forced by Earth System Model (ESM) simulations of future climate, function as computationally tractable statistical emulators of climatic shocks to crop productivity that can be coupled with Integrated Assessment Models (IAMs) to evaluate the broader energy and economic implications of the agricultural climate change impacts. We document the development of a statistical emulator of the yields of four major cereal crops (maize, rice, wheat, and soybeans), over 1972-2099 under two climate change scenarios (Representative Concentration Pathways 4.5 and 8.5). We access the suitability of panel fixed effects techniques, using the data from Inter-Sectoral Impact Model Intercomparison Project (ISIMIP)-Fastrack: a combination of six GGCMs and one

ESM, under rainfed cultivation regimes not accounting for carbon dioxide (CO_2) fertilization effects (CFE). We characterize the reduced-form response functions to temperature and precipitation, and assess their stationarity across time and different models—predictive power, and potential diagnostic utility. Using a simple trend interaction specification, we demonstrate how adaptation plays a contrasting role across GGCMs in reducing the potential negative percentage yield shocks on crops in future. A key contribution our work is the ability to capture and control for the potential confounding impacts of adaptation that inadvertently lead to double counting of shocks in IAMs. Our results show that the statistical emulator has considerable agreement in estimating relative changes in crop yields in future vis-à-vis the underlying GGCMs on which it is calibrated. The degree of agreement measured by way of suitable statistical indices, varies across GGCMs and mean future periods. The potential reasons of the disparity and possible techniques to account for them are highlighted. Built as six independent emulators that can be applied as an ensemble, as well as a multi-GGCM emulator calibrated on a large merged panel of six GGCMs; we demonstrate that our simple and flexible statistical emulator holds considerable potential as a diagnostic methodology to elucidate uncertainties in the processes simulated by GGCMs, and to support the development of climate impact inter-comparison exercises within the integrated assessment modelling community.

2.1 Introduction

Concern abounds that shifting weather patterns driven by climate change will detrimentally affect global agricultural productivity and food supplies (Portmann *et al* 2010). Understanding the potential threat necessitates reliable quantification of the risks of negative impacts on crop yields at broad geographic scales. Thus far, such assessments have involved forcing Global Gridded Crop Models (GGCMs) with outputs of Earth System Model (ESM) simulations in the context of the Agricultural Model Intercomparsion and Improvement Project (AgMIP—Rosenzweig *et al* 2013, 2014, Elliott *et al* 2014) and ISI-MIP (Warszawski *et al* 2013, Frieler *et al* 2015). GGCMs' key advantage is their ability to be calibrated on relatively few observations and then simulated over large geographic areas, generating realizations of crop yields with comprehensive spatial and temporal coverage at fine resolution. Even so, it is challenging to make direct use of these results to assess risks to agriculture because of their dimensionality—projections vary across discrete combinations of specific warming scenarios, ESMs, and crop models. Flexibly incorporating agricultural impacts into IAMs requires an encompassing envelope of model-and/or scenario-averaged responses to meteorology that is capable of faithfully reproducing GGCMs' simulated shocks to crop yields over a wide range of future climatic conditions.

The latter (scenario-averaged) more general responses are the key output of empirical climate economics studies, which capture the impacts of weather shocks on crop yields via statistically-estimated, reduced-form response surfaces made up of the marginal effects of time exposure to a vector of intervals of different meteorological variables (e.g. see Schlenker and Roberts 2009, Schlenker and Lobell 2010).

Response surfaces play the role of an *emulator*. Letting i and t index locations and time periods, they relate observed crop yields, Y, to a vector of meteorological covariates, $\boldsymbol{X}$, via a multi-dimensional envelope, $F(\cdot)$, which incorporates response parameters, $\boldsymbol{\beta}$: $Y_{i,t} = F(\boldsymbol{X}_{i,t}, \boldsymbol{\beta})$. In the empirical economic literature, F is commonly specified as a variant of the linear cross-section time-series econometric model

$$y_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + \alpha_i + \tau_t + \varepsilon_{i,t} \tag{2.1}$$

where y is the logarithm of historically observed yields, the vector $\boldsymbol{X} = \{T_1, \dots, T_J, P_1, \dots, P_K\}$ records the length of time over the growing season that each location historical spent in J intervals of temperature (T) and K intervals of precipitation (P), α is a fixed effect capturing the influence of time-invariant unobserved idiosyncratic factors at each location, τ is a time effect capturing the influence of common time-varying shocks, and ε is a random disturbance term. The elements of $\boldsymbol{\beta}$ are semi-elasticities that represent the temporally and spatially averaged yield impacts of each category of exposure. The limitation of equation (2.1) is its limited geographic scope—typically the locations i are sub-national administrative units within a single country.

Here we document the construction of a crop yield emulator that satisfies the competing mandates of GGCMs' global coverage and econometric models' parsimonious representation of the weather responsiveness of yields. The first question we ask is, what is $\boldsymbol{X}$, the set of meteorological variables that adequately and parsimoniously captures the envelope of GGCMs' response. Second, we ask what is $F(\cdot)$, in terms of the shape of the reduced-form response surface that best captures the relationship between $\boldsymbol{X}$ and y, and its stationarity across GGCMs with different characteristics, and over time. Finally, we investigate the ability of the resulting best-fit model to reproduce the percentage changes in yields generated by GGCMs under future warming scenarios relative to a reference base period.

We are not the first to ask these questions. Our approach seeks to both encompass and extend recent efforts by Oyebamiji *et al* (2015), who emulate multiple crops' responses utilizing data for a single GGCM (LPJmL), and Blanc and Sultan (2015), who emulate only the responses for maize[3] using multiple GGCMs, albeit independently. Typically, the outputs of model inter-comparison exercises record only a subset of the endogenously-varying internal processes of their constituent models. Whereas Oyebamiji *et*

[3] Blanc (2016) expands the earlier work of Blanc and Sultan (2015) to multiple crops as done in this chapter. However, at the time of preparing this paper for submission to ERL, Blanc (2016) is unpublished in peer-reviewed journal. Nevertheless, the methodology and objectives addressed in this chapter have differences with both Blanc and Sultan (2015) and Blanc (2016).

al 2015 account for potential changes in management practices[4], Blanc and Sultan 2015 do not control for the potential confounding impacts of adaptation that inadvertently lead to double counting of shocks in IAMs.

We focus on predicting the impacts on future rainfed crop yields as a response to climate change gross and net of any potential form of adaptation for four major crops (maize, rice, soybeans and wheat). Using an analytical framework that is simple, flexible and robust; we focus on two contrasting Representative Concentration Pathway scenarios (RCP), RCP 4.5. and RCP 8.5 (Moss *et al* 2010). In order to focus on the key underlying mechanisms that play the most important role in crop growth process, namely heat stress and moisture interactions; our study focuses on GGCMs' crop yields not accounting for CO_2 Fertilization Effect (CFE).

We find that our simple specification using daily intervals of T and P defined within the crop growing season, along with their individual interactions with time trend; is able to emulate the climate responses of the GGCMs. We show the robustness of our specification by way of both in-sample and out-of-sample validations. Further, to illustrate the contribution of adaptation to moderating yield shocks over the two epochs (2030~2064 and 2065~2099, RCP 8.5), we demonstrate the different responses to changes in future crop yields, using a trend-interaction specification. The latter findings are crucial to agronomic studies since the inability to correctly account for adaptation in future yield responses are often considered short-comings of statistical methods (Lobell *et al* 2011).

The rest of the chapter is organized as follows. Section 2.2 describes our econometric methodology for modelling climate-yield relationships. Section 2.3 presents the resulting relative changes in crop yields in future, and a comparison with the GGCMs. We summarize our findings with a discussion of the caveats and scope for work in future in Section 2.4.

2.2 Methods

The starting point for our analysis is the "perfect model approach (PMA)" developed by Lobell and Burke (2010) and used by Holzkämper *et al* (2012) and Blanc and Sultan (2015). Yields under different climate forcings are simulated by GGCMs, statistical models of the underlying relationship between climate and yields are developed and estimated, and the ability of the resulting emulators to reproduce GGCM yields are tested (figure 2.1). Importantly, at the broad geographic scale of GGCM grid cells, true yield

[4] To achieve this goal, the study utilizes a series of control runs of GGCM LPJmL (setup under different initial conditions) and builds an emulator based on a complex calibration process requiring a large number/forms of predictor variables.

responses across much of the world are not known. Thus testing the ability of our emulator to recreate the behaviour of the model is the only practical way to assess our ability to emulate the underlying yield response in nature.

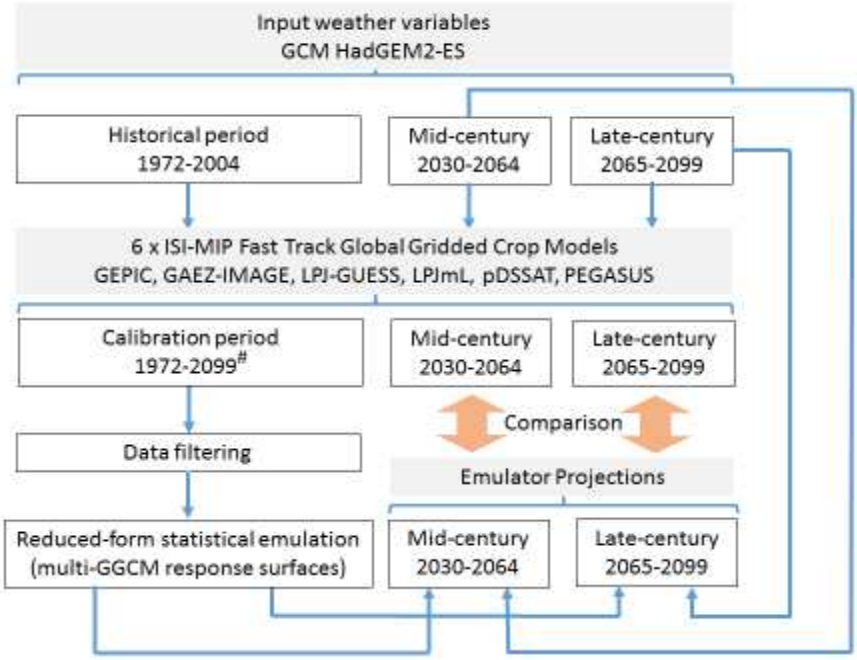


Figure 2.1. Methodological approach used in this study. The 6 x ISIMIP-FT GGCMs are used independently as well as a combined multi-GGCM panel. For an out-of-sample validation, data for 1972~2089 was used for calibration and then emulator projections in 2090~2099 were compared with GGCMs' outputs.

2.2.1 Data

Our data taken from the ISIMIP-FT exercise, uses six GGCMs[5] to simulate yields of maize, rice, soybeans and wheat on a 0.5° grid[6] for historical and future years (1972-2004 and 2005-2099, respectively), assuming cultivation of crops in all grid cells under rainfed conditions, not accounting for CFE[7]. All GGCM runs are forced with bias-corrected climate inputs (Hempel *et al* 2013) from HadGEM2-ES (Jones *et al* 2011), but the individual GGCMs differ substantially in their parameterizations, calibration, input variables, management practices and representations of farmer

[5] The six GGCMs utilized in this study are GEPIC (Liu *et al* 2007), GAEZ-IMAGE (van Vuuren *et al* 2006), LPJ-GUESS (Sitch *et al* 2003), LPJmL (BONDEAU *et al* 2007, Sitch *et al* 2003), pDSSAT (Elliott *et al* 2013, Jones *et al* 2003) and PEGASUS (Deryng *et al* 2011). Details of the modelling groups involved are provided in Section 1 of Appendix A.

[6] Approx 55 km^2 at the equator.

[7] This is a sensitivity run, where CO_2 concentrations for the future period (2005-2100) were held constant on present day levels.

adaptation (see table 2A in Appendix A, and Rosenzweig *et al* 2014 SI for futher details). All GGCMs simulated yields in historical simulation period (1972-2004) keeping management practices constant to year 2000.

To estimate statistical yield response, we use gridded annual yields (tons/hectare, t/ha) over the entire historical and future period (1972-2099)[8]. We mask grid cells using global rainfed cultivated areas for each crop from the monthly irrigated and rainfed crop areas around the year 2000 (MIRCA2000) dataset (Portmann *et al* 2010), before identifying and dropping cells with anomalously high or low crop yield values (Section 2 of Appendix A). The result is a balanced 128-year panel for each GGCM x crop combination[9], each of which has its own spatial coverage and distinct number of observations.

With regard to weather forcings in the ISIMIP-FT exercise, growing seasons vary across crops and GGCMs, as well as over the historical and future periods of their simulations. For example, the crop planting dates in GGCMs LPJ-GUESS and LPJmL vary according to annual weather conditions, thus leading to potential endogeneity of growing season exposures. To keep the analyses tractable, we subsume this heterogeneity and adopt a common, fixed, four-month growing season differentiated by latitude: for the northern hemisphere, the months May-August ($MJJA$) of the year of each observation of yields, and for the southern hemisphere, November-December of the year preceding each observation as well as January-February of the observation's year ($NDJF$).

We matched HadGEM2-ES climate forcings to GGCM generated realizations of yield for each year of the future period two RCPs (4.5 and 8.5), using the methods described above. For consistency, we used the identical crop-specific spatial filters and growing season truncations across the different crops and models. Statistical models that we train on climate and yields over the calibration period can then be linked with climate data for the prediction period[10] to generate synthetic yield projections capable of being compared with GGCM outputs.

[8] As separate robustness checks, we also estimate the statistical yield response models on the historical period (1972-2004). However, across most GGCMs we find that the marginal response of log yield to extreme heat is larger over the historical period than in the future; there by suggesting an acceleration of endogenous adaptation post-historical period. This makes it difficult to accurately estimate the trends from the historical period and then apply them to make unbiased projections for future periods.

[9] Some exceptions (e.g., PEGASUS does not simulate rice, GAEZ-IMAGE has different number of years etc.) are discussed in Section 2 of Appendix A.

[10] The two periods (2030-2064 and 2065-2099 in RCPs 4.5 and 8.5) used for validating our future predictions with GGCMs' are within the sample of our calibration space, thus amounting to in-sample validation. For an out-of-sample validation, we re-calibrate the model on 1972-2089 and then compare our predictions in 2090-2099.

2.2.2 Variable selection and empirical specification

Drawing on the empirical climate economics literature (such as Lobell and Burke 2010, Schlenker and Roberts 2009, Deschênes and Greenstone 2012), our empirical analysis framework relies on panel-data fixed effect models. For our covariates we employ intervals ("bins") of T and P, as well as the interaction of the bins with a linear time trend. The bins $\{T_1, \dots, T_J, P_1, \dots, P_K\}$ are counts of number of days over the growing season at each grid-cell spent in J intervals[11] of $T(Degree\ Celcius, °C)$ and K intervals of $P(millimiter\ per\ day, mm/d)$, where:

$J = \{< 5,\ 5\sim7.5, 7.5\sim10, 10\sim12.5, 12.5\sim15, 15\sim17.5, 17.5\sim20, 20\sim22.5, 22.5\sim25, 25\sim27.5, 27.5\sim30, > 30\}$ and

$K = \{< 3,\ 3\sim4,\ 4\sim5, 5\sim10, 10\sim15, 15\sim20,\ > 20\}$

The bins $J = 15\sim17.5°C$ and $K = 5\sim10mm/d$ are omitted in regressions as reference category. Thus with reference to equation (2.1), each coefficient of T (P) indicates the impact on $log\ yield$ of an additional day in the Jth (Kth) interval, relative to a day in the dropped T (P) bin. The rationale behind the binning approach is discussed in Section 5 of Appendix A.

We analyze each of our six GGCMs multi-crop dataset of weather and yields using a traditional panel data econometric approach (equation 2.2, referring to it as our 'base specification'):

$$y_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + (\boldsymbol{X}_{i,t} * \tau_t)\boldsymbol{\theta} + \alpha_i + \varepsilon_{i,t} \tag{2.2}$$

where τ is now a linear time trend, i are the spatial units (grid-cells) and t the time dimension (year). The elements of $\boldsymbol{\theta}$ capture the additional effect of changing factors[12] (within the GGCMs) on the marginal response to T and P. The remaining terms are the same as in equation (2.1). However, as discussed later (Section 2.3.1), our stability tests indicate that for three of the GGCMs (GAEZ-IMAGE, pDSSAT and PEGASUS), the interaction term in equation (2.2) gives us implausible results. We therefore omit the interaction term for this set of GGCMs, and rewrite our equation (2.2) as equation (2.3), referring to it as 'base specification without interaction'

$$y_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + \alpha_i + \varepsilon_{i,t} \tag{2.3}$$

[11] For each T and P bin (except the extreme lower and upper bins), the lower range is included in the count (e.g. in temperature bin $7.5\sim10$, $T \geq 7.5$ is included in the count). The extreme T and P bins are open-ended.

[12] These factors can be considered as the moderating effects on the response to extreme heat that need to be correctly accounted for.

Further, since no GGCM is considered 'perfect' or ranked in their performance (Rosenzweig *et al* 2014), we take our PMA a step further by calibrating our model on a combined multi-GGCM panel of all six GGCMs. The equation (2.2) would now include an additional dimension representing model-specific fixed effects, λ_m (m^{th} GGCM) and can be now written as:

$$y_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + \boldsymbol{\delta} * (\boldsymbol{X}_{i,t} * \tau_t)\boldsymbol{\theta} + \alpha_i * \lambda_m + \varepsilon_{i,t} \quad (2.4)$$

where $\boldsymbol{\delta}$ is a dummy variable; $\boldsymbol{\delta} = \mathbf{0}$ for GAEZ-IMAGE, pDSSAT and PEGASUS, and $\boldsymbol{\delta} = \mathbf{1}$ otherwise

By incorporating a multi-GGCM panel as in equation (2.4) our study is the first to attempt construction of a comprehensive emulator that encompass the heterogeneous behaviour across GGCMs[13]. It should be mentioned here that by applying both approaches (equations 2.2 or its variant equation 2.3; and 2.4), we offer added flexibility in the application of emulators, either as an ensemble of six GGCMs or as a single multi-GGCM.

We run our regression specifications in R package $Linear\ Fixed\ Effects\ (LFE)$ (Gaure 2013), which can handle arbitrary number of factors[14] and is tailored for fixed effect estimation on large panel data. To account for heteroscedasticity and autocorrelation in the error term ($\varepsilon_{i,t}$), we use robust standard errors (S.E.)[15] clustered by grid-cells.

2.2.3 Climate scenarios used in future projections

The RCPs are atmospheric greenhouse gas (GHG) concentration trajectories representing different level of radiative forcing (Moss *et al* 2010). While the RCP 8.5 scenario encompasses the highest level of global warming compared to historical conditions and projects the highest level of GHG concentration by year 2100, RCP 4.5 is representative of an intermediate scenario with relatively lower GHG concentration and changes in temperature by the end of the twenty-first century. The differing climate variability in the combination of two future periods and RCP scenarios (figure 5A of Appendix A) enables us to undertake a more robust validation. It must be noted though that considering the crop yield data used in our study maintains the CO_2 concentrations for the future period (2005-2100) constant to present day levels, the

[13] Although Blanc and Sultan (2015) and Blanc (2016) include climate variables from multiple ESMs and utilize five GGCMs, the studies do not attempt a combined multi-GGCM.

[14] The factors in equations (2.2) and (2.3) correspond to α_i and λ_m, (i.e. grid-cell ID and GGCM), with i and m as the levels. In econometric jargon, factors are invariably referred to as dummy variables.

[15] The S.E.s are adjusted for the reduced degrees of freedom (DOF) coming from the dummies which are implicitly present. They are also small-sample corrected

relative difference in the GHG concentration for the two RCPs would be less relevant compared to the changes in the variability and mean state of the climate.

2.2.4 Emulator projections and diagnostic comparisons with GGCM outputs

To assess our emulator projections with GGCMs, we use relative bias (RB) as shown in equation (2.5) computed at each i^{th} grid-cell. Positive (negative) values of RB indicate that the emulated changes in yields at that grid-cell exceeds (understates) the corresponding GGCM projections.

$$RB_i = [\%\, Change\ y_i^{Emu} / \%\, Change\ y_i^{GGCM}] - 1 \quad (2.5)$$

Our performance comparison of the emulator proceeds in two stages. First, we undertake an in-sample test, estimating equations (2.2-2.4) over 1972-2099. We apply the resulting estimated coefficients to transformed weather variables to emulate % changes in yields over two periods (2030-2064 and 2065-2099) and compare our results with the corresponding GGCM's % changes in yields. For an out-of-sample validation, we use a subset of the full panel (i.e. 1972-2089) for re-estimating the equations (2.2-2.4), and then compare our % changes in yields for the left out period (2090-2099) with the corresponding GGCM's % changes in yields. The performance comparison is undertaken for both RCP scenarios.

It is worth re-iterating the six GGCMs differ in model types, processes, fertility inputs, and calibration procedures (see table 2A in Appendix A). The use of relative yield changes rather than absolute yield values is thus an important measure of the suitability of our emulator estimates.

2.3 Results

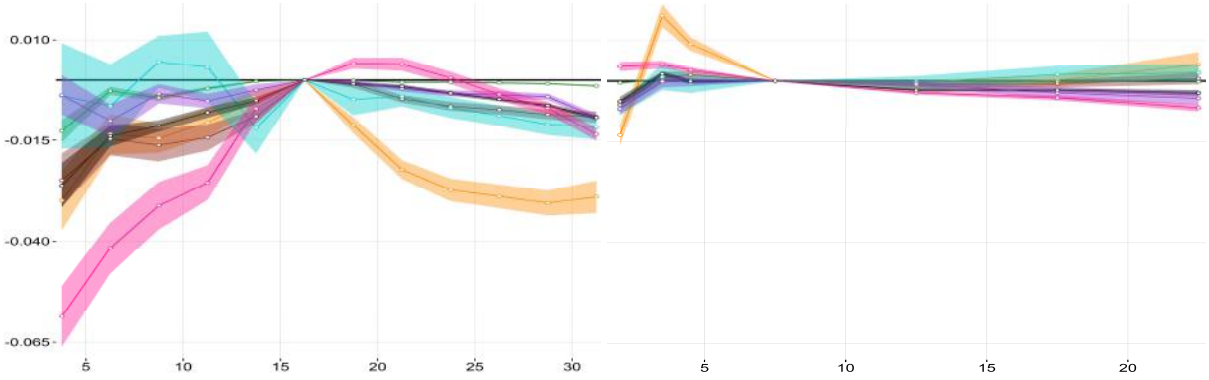
2.3.1 Regression analyses (1972-2099 RCP 8.5)

We begin by assessing the ability of the emulators to mimic the nonlinear response of crop yields to heat uncovered by empirical studies (Schlenker and Roberts 2009, Lobell *et al* 2011). For the sake of clarity, it is worth reminding that the set of GGCMs (GAEZ-IMAGE, pDSSAT and PEGASUS), the regression specification is based on equation (2.3). The other GGCMs (GEPIC, LPJ-GUESS and LPJmL) incorporated interaction term and the specification took the form of equation (2.2).

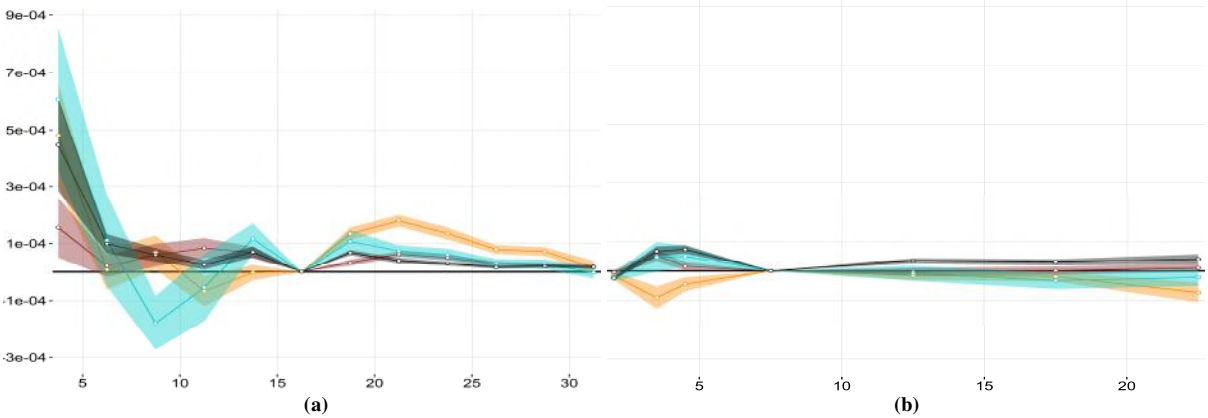
Figure 2.2A shows the estimated coefficients of T and P bins for maize with robust S.E.s (see figure 1A of Appendix A for corresponding plots of other crops). Since the dependent variable is $log\ yield$, the vertical difference (y-axis) implies a percentage change in yields (t/ha) for each additional day in a particular bin, relative to the bin dropped in regression and holding all other bins constant. For example,

consider two points on T coefficients (panel 'A') for GEPIC (orange). Moving from a day with daily average T between $15{\sim}17.5°C$ to a day at $> 30°C$ would result in a predicted marginal yield decline of 3%, holding all other bins at the same level.

A. Nonlinear relationship between (a) T and log $yield$ (b) P and log $yield$



B. Nonlinear relationship between (a) 'T' interaction terms and log $yield$ (b) 'P' interaction terms and log $yield$



GEPIC GAEZ-IMAGE LPJ-GUESS LPJmL pDSSAT PEGASUS Multi-GGCM

Figure 2.2. Response of Maize $\mathbf{log}\,\boldsymbol{yield}$ to (a) $\boldsymbol{T}$ and (b) $\boldsymbol{P}$ bins. Coefficient estimates are for (A) main $\boldsymbol{T}$ and $\boldsymbol{P}$ (B) interaction of $\boldsymbol{T}$ and $\boldsymbol{P}$ bins with time trend (GEPIC, LPJ-GUESS, LPJmL and Multi-GGCM). S.E.s are robust to heteroscedasticity and autocorrelation. Graphs display changes in yield (%) for exposure to an additional day within a particular $\boldsymbol{T(P)}$ bin interval, relative to bins $\boldsymbol{T = 15{\sim}17.5\,°\,C}$ ($\boldsymbol{P = 5{\sim}10\ mm/d}$). The **95%** confidence band (CI) is adjusted for spatial correlation. The horizontal black line corresponds to $\boldsymbol{x - axis = 0}$ reference. CI intersecting the horizontal **0** reference line would imply that the corresponding coefficient is insignificant $\boldsymbol{(i.e.\ p > 0.05)}$

As seen in figure 2.2A, across all GGCMs, the response of yields to P over a wide range are remarkably muted in comparison to the corresponding responses to T. Moreover, the responses of different GGCMs exhibit considerable heterogeneity[16] for T, with an additional day over $30°C$ causing yields to decline between $0.4 - 3\%$ from their peak at the optimum T. Most GGCMs exhibit similar characteristic shape, with the mid-range temperatures ideal for maize growth process and lower/higher thresholds having detrimental impact on yields, in line with earlier works (see Schlenker and Roberts 2009, Lobell *et al* 2012, Blanc and Sultan 2015). There are a few exceptions though with LPJmL (cyan) having largely insignificant coefficients ($p > 0.05$) for lower T bins, and GAEZ-IMAGE (dark green) having an overall muted response. However, the multi-GGCM response of yields to T (in black) embraces the broad pattern of all six GGCMs remarkably well.

As noted by Schlenker and Roberts 2009, we also observe that the response of crop yields is robust to modification in specifications. A similar pattern of nonlinear effects of T remains even if the homogeneous time variant factors (e.g. technology change) are controlled for by using year-fixed effects rather than interaction of T and P with time trends (results available upon request).

Focusing attention on figure 2.2B which shows the response of maize yields to the coefficient estimates of interaction terms ($\boldsymbol{\theta}$) for the three GGCMs (GEPIC, LPJ-GUESS, LPJmL), the base response to T is altered to varying degrees by trends in the marginal response of extreme heat reflecting the impact of adaptation. For these GGCMs, the trends offset yield declines, suggesting that their internal processes facilitate adaptation, reducing losses by $1{\sim}3$ % for $> 30°C$ days[17].

For GGCMs pDSSAT and PEGASUS that do not incorporate the interaction term in the regression specification, our stability tests indicate that the introduction of the same generates implausible positive base responses to extreme heat (figure 2A in Appendix A). GAEZ-IMAGE with its overall muted response to T and P has negligible difference in implementation of either specification (equations 2.2 or 2.3). These findings therefore vindicate omitting the interaction term from the regression specifications of GAEZ-IMAGE, pDSSAT and PEGASUS.

[16] The varying responses of GGCM yields to T can be largely attributed to the characteristics of GGCMs' sensitivity to temperature changes and acute heat stress (Rosenzweig *et al* 2014 SI). The divergence in GGCMs' yield responses are investigated further in Chapter 4.

[17] Comparison of T bin coefficients in figure 2.2A with the corresponding T bin coefficients in figure 2.2B. The difference in vertical scale between the two values would imply a shift in response of % yield.

2.3.2 Assessing emulator goodness of fit

Bearing in mind the low base yields across a sizeable number of grid-cells in majority of the GGCMs, our RB results are likely to be influenced by outliers[18]. We begin by assessing the bias in our emulators as distribution of RB weighted by grid-cell mean annual production[19] in tons (t), for all crops, in periods 2030~2064 and 2065~2099 under both RCP 4.5 and 8.5 scenarios (figure 2.3).

[18] Since we are making a relative comparison of two 'percentage changes in future yields', the low base yields (t/ha) in either or both GGCMs and emulators historical period can result in large RB across such grid-cells. Therefore, to enable a systematic assessment of the RB, we need to weigh the RB by the corresponding grid-cell mean annual production (t).

[19] Since the harvested area ($hectares, ha$) varies by grid-cells in the GGCM simulations, the RB is weighted by production (t) and not by yields (t/ha)

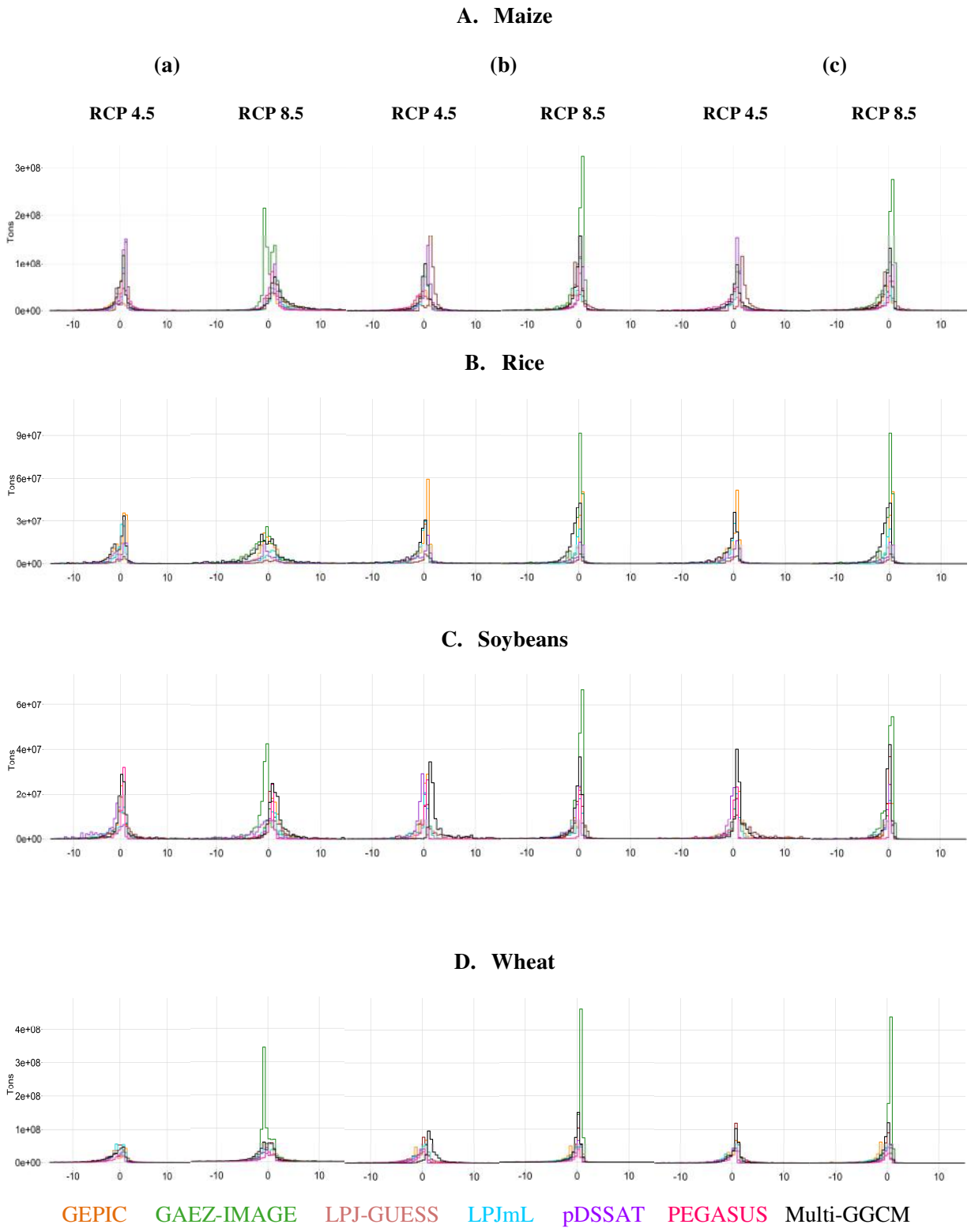


Figure 2.3. Distribution of RB (%) weighted by grid-cell mean annual production (***t***) for all four crops (a) 2030-2064 (In-sample validation) (b) 2065-2099 (In-sample validation) (c) 2090-2099 (Out-of-sample validation)

As evident from the above distributions, the RB for each crop-emulator (including the multi-GGCM) is low (within +/-5) across grid-cells with major share of present day global crop production. To facilitate easier interpretation on the share of bias, table 2.1 shows the % of maize global production across grid-cells under different intervals of RB, for all three periods under the two RCP scenarios. For instance, emulator LPJmL over-predicts ($RB > 5$) yields for approx. one-tenth of total global maize production over the period 2030~2064 RCP 8.5, but the RB is low ($< -5 \sim > 5$) for the major share of global production.

Table 2.1. Share of maize production (%) with the corresponding RB intervals, for periods 2030~2064, 2065~2099 shown in parentheses, and 2090~2099[a] shown in square braces, in (A) RCP 4.5 and (B) RCP 8.5 scenarios.

(A) RCP 4.5

RB	GEPIC (%)	GAEZ-IMAGE[b] (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	1.33 (0.75) [3.84]	-	1.49 (1.96) [3.53]	1.54 (2.37) [3.92]	0.03 (0.00) [0.16]	3.48 (1.91) [1.31]	0.69 (1.38) [2.44]
5 ~ 10	1.33 (0.74) [4.15]	-	1.30 (2.17) [4.64]	1.85 (2.87) [3.75]	0.24 (0.06) [0.11]	3.65 (1.57) [1.59]	1.01 (1.53) [3.17]
0 ~ 5	24.85 (27.66) [47.66]	-	77.44 (84.51) [83.27]	63.64 (50.87) [63.77]	76.78 (69.18) [75.37]	57.60 (30.27) [34.34]	54.12 (40.70) [58.08]
0 ~ -5	62.00 (60.91) [36.34]	-	19.56 (11.32) [8.58]	29.51 (39.47) [24.72]	20.96 (28.79) [22.56]	27.60 (53.89) [43.32]	39.15 (51.86) [31.47]
-5 ~ -10	6.00 (5.46) [4.19]	-	0.10 (0.04) [0.00]	1.76 (2.42) [1.95]	1.02 (1.13) [1.08]	3.58 (6.32) [8.68]	2.53 (2.44) [2.62]
< -10	4.50 (4.48) [3.83]	-	0.12 (0.00) [0.00]	1.70 (2.01) [1.89]	0.98 (0.84) [0.74]	4.09 (6.04) [10.65]	2.50 (2.10) [2.22]

(B) RCP 8.5

RB	GEPIC (%)	GAEZ-IMAGE (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	9.16 (0.14) [0.13]	2.50 (0.04) [0.02]	11.47 (2.06) [1.51]	5.85 (0.04) [0.04]	2.02 (0.01) [0.01]	0.84 (0.07) [0.07]	6.58 (0.16) [0.13]
5 ~ 10	7.35 (0.10) [0.05]	2.90 (0.04) [0.02]	11.48 (1.74) [1.41]	4.87 (0.04) [0.05]	2.71 (0.01) [0.01]	1.12 (0.17) [0.04]	7.02 (0.17) [0.16]
0 ~ 5	59.93 (19.36) [14.98]	46.89 (52.10) [51.26]	62.12 (27.68) [22.52]	73.89 (39.40) [39.29]	65.97 (55.57) [55.93]	64.61 (62.41) [58.64]	71.63 (33.68) [32.87]
0 ~ -5	21.58 (70.47) [73.08]	46.00 (41.89) [42.70]	14.62 (62.91) [64.44]	15.89 (57.25) [58.84]	26.77 (37.28) [37.06]	28.46 (32.89) [35.39]	13.39 (60.77) [60.55]
-5 ~ -10	0.67 (5.00) [5.78]	0.81 (2.51) [3.34]	0.11 (2.68) [6.11]	0.14 (1.96) [0.89]	1.34 (3.35) [3.44]	2.38 (2.22) [3.04]	0.77 (2.52) [2.93]
< -10	1.32 (4.95) [5.98]	0.89 (3.43) [2.26]	0.20 (2.93) [6.00]	0.18 (1.32) [0.90]	1.19 (3.78) [3.56]	2.60 (2.24) [2.81]	0.61 (2.70) [3.37]

[a] Regressions for out-of-sample validation were run on a panel spanning 1972~2089, in contrast to the in-sample which are on 1972~2099

[b] Data not available

In order to assess the spatial pattern of the predicted changes in future emulator yields with those of the GGCMs', figure 2.4 illustrates the RB of the emulators geographically, as calculated by equation (2.5) at each grid-cell.

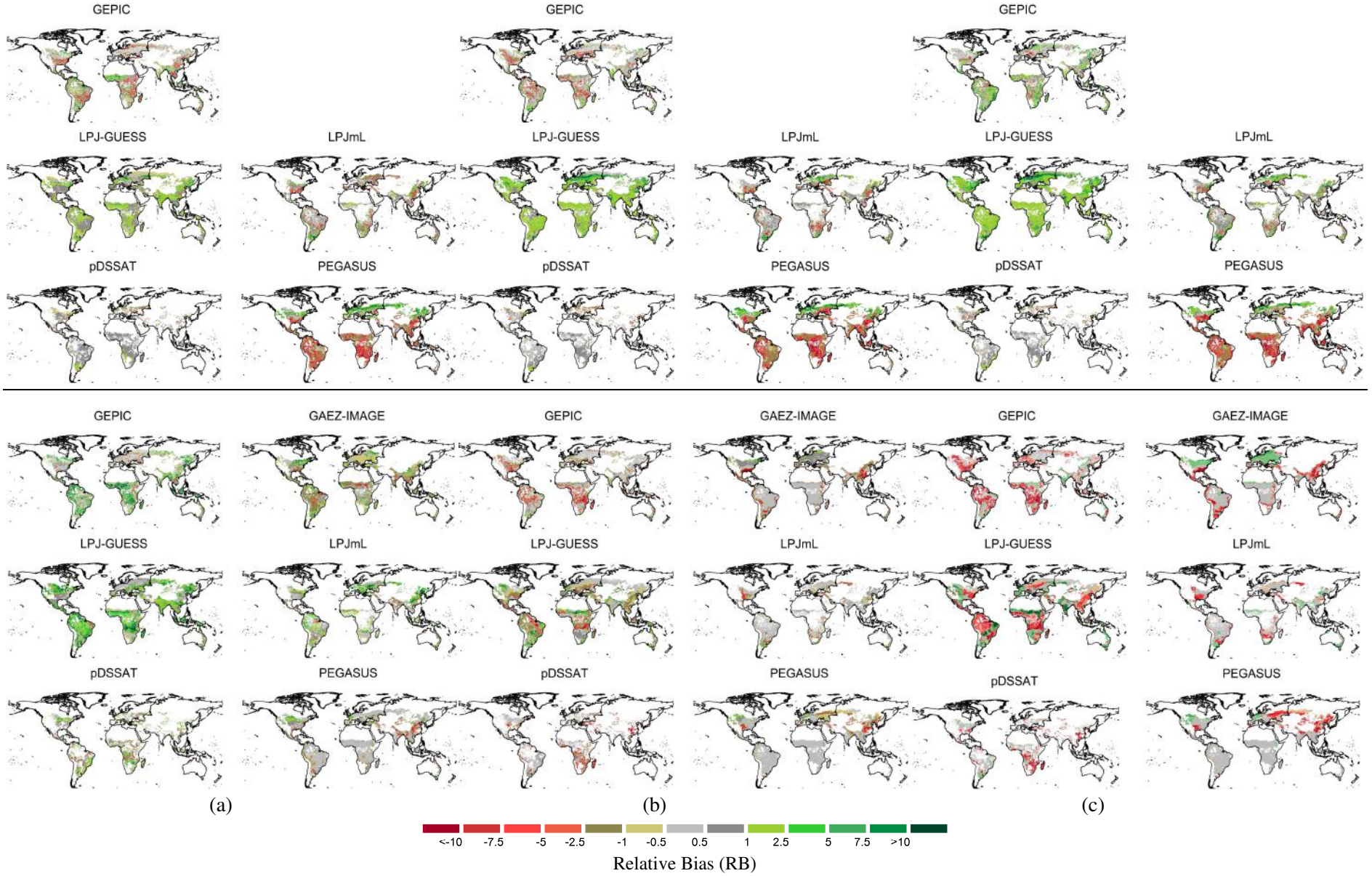


Figure 2.4. Performance of six maize emulators (RB) in predicting % changes in yields for RCP 4.5 (top panel) and RCP 8.5 (bottom panel) in (a) 2030~2064 (b) 2065~2099 and (c) 2090~2099, relative to 1972~2004 baseline. White regions denote regions where crop is either not grown (as per MIRCA 2000) or is filtered out in the data-cleaning steps of each GGCM. The spatial coverage is thus different across the maps of six GGCMs. Data was unavailable for GAEZ-IMAGE in RCP 4.5. The out-of-sample predictions (2090~2099) were made applying the coefficient estimates from regression run on 1972~2089 panel. Positive (negative) RB over a grid-cell implies that the % changes in yields predicted by emulator are higher (lower) than the % changes in yields of the corresponding GGCM.

Across both RCPs and all three future periods, emulators built on GAEZ-IMAGE, LPJmL and pDSSAT have fairly good agreement in predicting % changes in the yields relative to the baseline historical period. This is evident from the low RB (within the range of $-5 \sim 5$) over most regions, with a few exceptions where the low-base values result in higher disparity in the relative impacts of emulator and GGCM. The results are in line with figure 2.4 and the corresponding summary (table 2.1), both of which show the RB to be low where the production share of maize is high.

For PEGASUS, the RB shows contrasting range in the same future period of the two RCPs, with far better agreement in the emulation of relative changes in yields for RCP 8.5 scenario. Although the reason for this disparity are beyond the scope of this paper requiring detailed knowledge of the simulation setup used by the GGCM in the two climate scenarios; a possible explanation could be the growing seasons (harvesting dates etc.) that are dynamically assigned by the GGCM in its simulation run[20] fall outside the definition of our growing seasons months.

For GGCM GEPIC, the growing seasons are again dynamic as in PEGASUS. Moreover, the simulations in GEPIC are run independently every decade to take into account soil nutrient depletion (Rosenzweig *et al* 2014 SI). The ability of our emulator to replicate GEPIC's non-linear responses to future climate would thus be challenging with a generic linear time trend interaction.

The last of the six individual emulators (LPJ-GUESS) generally over-estimates the % changes in yields under all period-scenario combinations, showing a marginal positive RB ($1 \sim 2.5$). It must be noted that GGCM LPJ-GUESS predicts potential yields unlimited by nutrient or management constraints[21]. Any analysis of its yields are suitable only if the potential maximum yield in future are being investigated.

Although the broader stable performance of a multi-GGCM emulator constructed using equation (2.4) is evident from figure 2.4 and table 2.1, the maps in figure 2.5 illustrate how the overall responses of the six GGCMs' yields can be harnessed as a single blend of multi-model impacts assessment.

---

[20] See table 2A in Appendix A for GGCMs' meta-parameters. As highlighted in Rosenzweig *et al* 2014 SI, some GGCMs (PEGASUS, GEPIC) allow for automatic adjustments of planting and harvesting dates as per the annual weather conditions. Although not examined in detail, it is likely that for RCP 4.5 (where the RB in PEGASUS is higher compared to RCP 8.5), the GGCM dynamic growing season months are not in agreement with our fixed growing season window for all crops~GGCM~period~scenarios combinations. The contrasts in the RB for emulator GEPIC across the scenario~periods could also be partly attributed to this reason. Contrasts in inter-model divergence of responses are investigated in Chapter 4.

[21] These caveats are discussed in https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/data-archive/fast-track-data-archive/data-caveats.
Although GAEZ-IMAGE also predicts potential yields, unlike LPJ-GUESS its yields are driven by climate and soil moisture condition (soil moisture, soil water holding capacity).

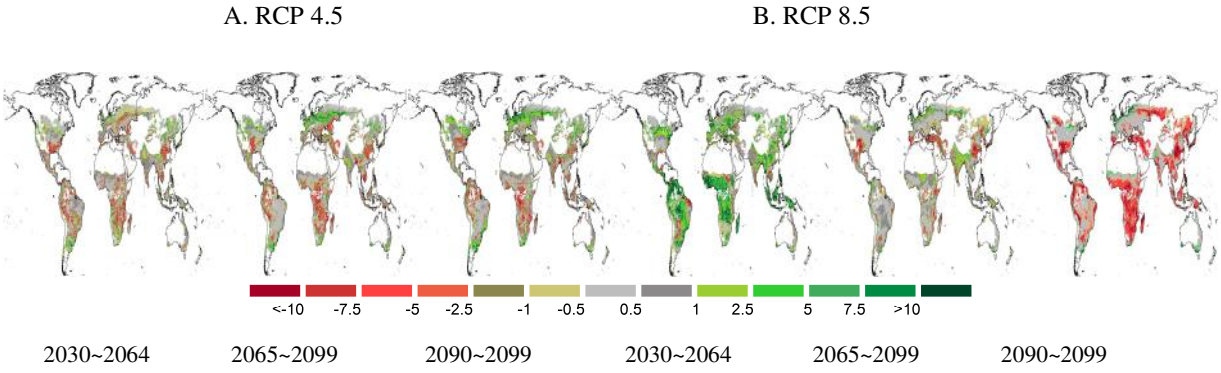


Figure 2.5. Performance of multi-GGCM maize emulator (RB) in predicting % changes in yields, relative to 1972~2004 baseline for (A) RCP 4.5 and (B) RCP 8.5. The white regions are grid-cells where crop is either not grown (as per MIRCA2000) or is filtered out in the data-cleaning steps of each GGCM. The spatial coverage will encompass the earlier map of each of the six GGCMs. Data was unavailable for GAEZ-IMAGE in RCP 4.5 and the multi-GGCM is therefore built on five GGCMs. The out-of-sample predictions were made applying the coefficient estimates from regression run on 1972~2089 multi-GGCM panel. Positive (negative) RB over a grid-cell implies that the % changes in yields predicted by multi-GGCM emulator are higher (lower) than the % changes in mean yields of the six GGCMs.

For both scenarios and across all GGCMs, the emulator RB in the out-of-sample validation are similar to in-sample validation (table 2.1 and figures 2.3, 2.4 and 2.5), thus showing the robustness to estimation methodology. Considering the heterogeneity across the six GGCMs as well as in their simulation setups for the four crops (Rosenzweig *et al* 2014 SI), it is not surprising that the results do vary by emulator-crop-period-scenario combinations (figures 2.3-2.4, table 2.1 here and table 4A in Appendix A). More importantly, there is no single emulator that consistently outperforms the rest across all four crops. Suffice to say that by incorporating multiple crops, GGCMs in different future RCP~epoch combinations; our estimation methodology and results are neither overtly dependent on the choice of the underlying processes involved in a particular GGCM, nor on the choice of a particular crop whose yield response is inadvertently replicated.

2.3.3 How much is adaptation beneficial?

We now investigate the ability of our statistical specification to capture the intrinsic adaptation associated with the GGCMs' simulations in historical (1972~2004) and future (2005~2099). For the three emulators that incorporated the interaction term (GEPIC, LPJ-GUESS and LPJmL), we make a comparison of how much larger the % shocks to yields would be if we stripped out the effects of the types of adaptation the GGCMs appear to be assuming.

We use the coefficient estimates of T and P bins i.e. $\boldsymbol{\beta}$ term in equation (2.2), estimated from the regressions run on 1972~2099 RCP 8.5 for maize. We then predict the changes in yields in the two epochs (2030~2064 and 2065~2099) relative to the historical period (1972~2004) as before. The changes in maize yields estimated here would imply impacts of climate on yields not taking the potential benefits of adaptation into account in the future, relative to the baseline historical period. We call these 'without-adaptation' predictions.

For comparison, we retain our earlier predictions of the relative changes in maize yields in future, that in addition to $\boldsymbol{\beta}$, also accounted for the coefficients of the interaction terms i.e. $\boldsymbol{\theta}$ term of equation (2.2). These are our base specification estimates accounting for potential benefits of adaptation in the future, again relative to the same baseline historical period. We call these 'with-adaptation' predictions.

The difference in the percentage changes in yields between without-adaptation and with-adaptation -measured as percentage points (pp)-, would give us the estimated potential benefits of using adaptation in reducing the impacts of climate change[22]. To illustrate these graphically, plots in figure 2.6 (shown as avoided % loss of production across maize grid-cells) depict the contributions of the adaptation (implicit to the three GGCMs) in moderating yield declines or in symmetrically amplifying yield increases.

As evident from figure 2.6, we find noteworthy ($> 5\ pp$) potential positive benefits of adaptation across major maize growing regions for the three emulators. The added benefit is highest for LPJ-GUESS which predicts potential yields in ISIMIP-FT and shows a consistent pattern across all regions. GEPIC and LPJmL on the other hand show spatial heterogeneity in potential benefits with some regions showing contrasting impacts of adaptation (especially across India and central Africa).

[22] This is based on the assumption that the present crop growing regions used in our study (based on MIRCA2000) would remain constant in the future. However, due to lack of available data for land-use from ISIMIP-FT and to remain consistent within the assumptions of GGCMs' simulation setup, we base our analysis on this notion.

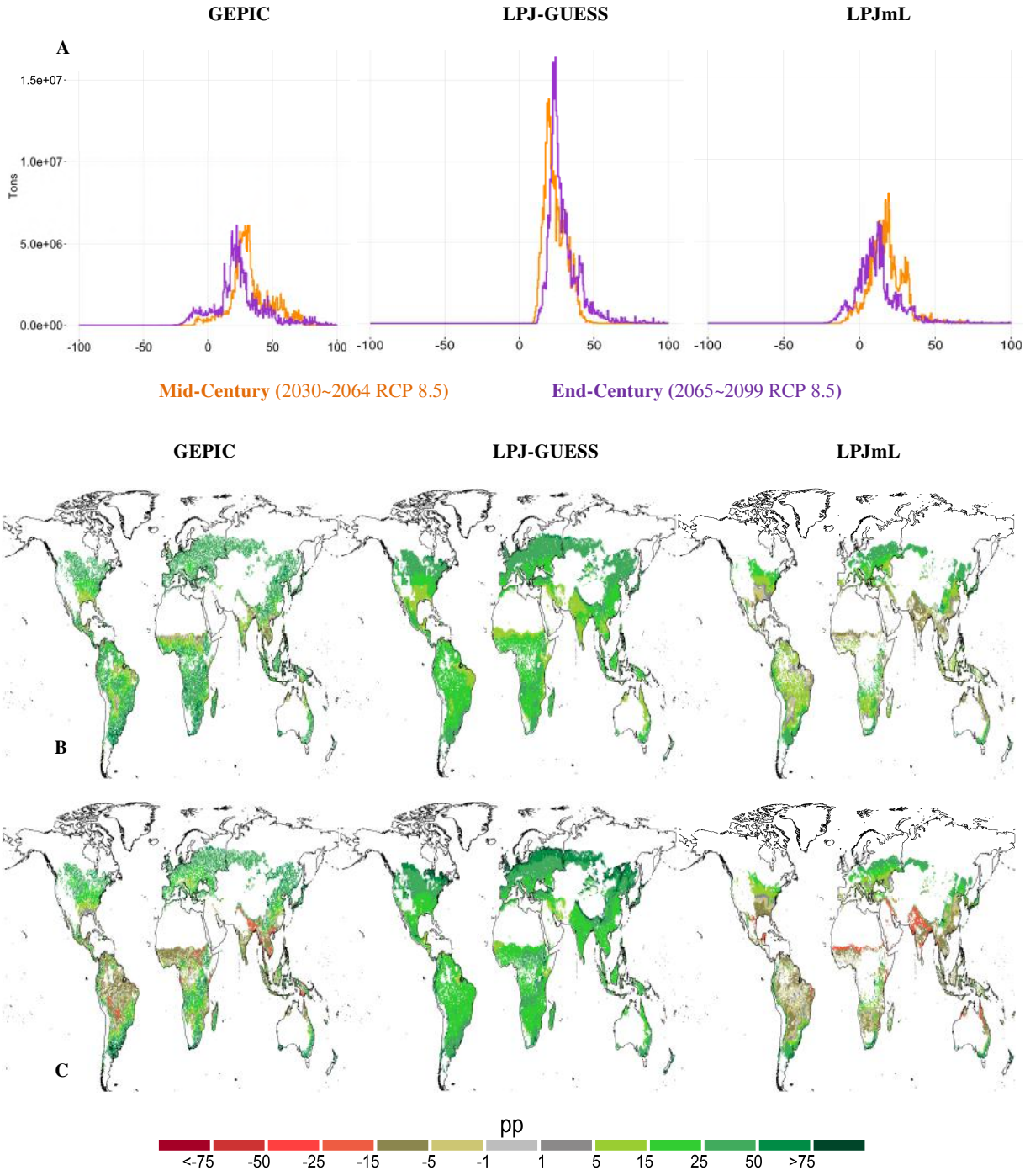


Figure 2.6. (A) Change in % maize yield shock (***pp***) in RCP 8.5 scenario, weighted by grid-cell mean annual production (***t***), in 2030~2064 (orange) and 2065~2099 (purple); both relative to 1972~2004. Maps of change in % maize yield shock (***pp***) for RCP 8.5 scenario, in (B) 2030~2064 and (C) 2065~2099, relative to 1972~2004. The difference is calculated as [% changes in emulator yields accounting for adaptation] – [% changes in emulator yields not accounting for adaptation].

It must be borne in mind that the type of adaptation varies across the three GGCMs (Rosenzweig *et al* 2014 SI) and in the absence of finer details of how adaptation is modelled at regional scales, the implausible benefits of adaptation across these pockets of growing regions would be difficult to decipher.

Conclusions

We develop an ensemble of statistical emulators for four different rainfed crops, at a global, fine scale, gridded resolution. Constructed independently on six different GGCMs as well as on a combined multi-model panel of six GGCMs; our simple, flexible and robust emulator can have wide-ranging applications in studies assessing impacts of climate change on crop yields. But perhaps the biggest attraction of our reduced form emulator is the rapid implementation as a surrogate model, where-in only two weather variables (T and P; as bins) are required to be constructed from the output of a ESM. Further, by incorporating interactions of T and P bins with a simple linear time trend, the emulator is capable of capturing the underlying adaptation and management practices implicit in the GGCMs' simulations. Though simple, the emulator can be easily combined with multiple realizations of future climate for analysis of climate impacts[23] on crops and ultimately be linked to IAMs.

Yet, we recognize the possible limitations of this study and scope for further studies. For instance, our choice of fixed growing season months ($NDJF$ for Southern Hemisphere. and $MJJA$ for Northern Hemisphere.) across all six GGCMs that do have heterogeneous growing seasons could potentially forego heat-moisture interaction falling outside the growing season bins. It is envisaged that future simulation protocols of ISI-MIP2[24] would ensure homogeneity for GGCM growing season, addressing the caveat in this study.

We have not attempted to fit a statistical model using the data from the irrigated regime of GGCMs simulations. This is primarily because the GGCMs use highly varying degree of parameterization schemes in their simulation setup and the same is difficult to specify as a single regression fitting across all GGCMs. Again, ISI-MIP2 is anticipated to include systematic harmonization of simulation runs, and thus be more suitable for such an exercise.

The motive to use crop yield data not accounting for CFE was deliberate. By basing our study on constant- CO_2 runs, our aim has been to examine what a basic emulator of heat and moisture effects of climate change would look like, controlling completely for the confounding effects of CO_2. Nonetheless,

[23] The wider scope of such an emulator therefore has a potential application for both with and without adaptive strategies in agriculture.

[24] See ISI-MIP2 protocol, www.isi-mip.org

future yields accounting for CFE can be predicted using our present methodology, by incorporating a post-estimation correction (as done recently by Sue Wing *et al* 2015). Left for work to be done in future, such a technique would be analogous to 'ON/OFF' switch, thereby negating the need to re-calibrate the emulator independently on crop yield data with CFE.

Closing Remarks

The emulators I develop in this paper are deliberately simple, as the objective was to develop emulators that can be rapidly combined with multiple realizations of future climate for climate risk analysis and ultimately be linked to IAMs. Because the GGCMs' simulations in ISIMIP-FT are not harmonized with a common set of input parameters (such as crop growing seasons, adaptation and management practices etc.), it is difficult to identify the intrinsic parameters within the six GGCMs that play a prominent role in the divergence of results. The lack of calibration (or contrasting calibration techniques) further make it difficult to rank the GGCMs in their overall performance of replicating historical crop yields. It is therefore recommended to apply the GGCMs as a multi-model ensemble in impact estimation studies.

The work in the succeeding chapter focuses on testing the robustness to different meteorological variables and empirical specifications, thus refining the statistical emulators built in this chapter. Moreover, in Chapter 4 where the focus shifts to U.S. counties, I attempt to investigate the underlying drivers of divergence in the GGCMs' yield responses. Undertaking such an exhaustive task is the need of the hour in inter-model and inter-method comparison exercises.

Chapter 3: Robustness Tests of Statistical Emulator

Preface

The emulator developed in Chapter 2 is evaluated by a series of robustness checks. This chapter explores possible refinements of the statistical emulator with respect to: i) the choice of explanatory variables and ii) the statistical specification, by considering also the potential role of further stratifying the response of crop yields by Agro-Ecological Zones (AEZs). Specifically, I investigate the role of Vapor Pressure Deficit (VPD) as an added predictor variable and compare the performance of emulators with the specifications of Chapter 2 that were built on T and P. Further, to ensure that the results summarized in Chapter 2 did not depend on an overly specific choice of regression specification, and in line with common practice in statistical modelling (e.g. Urban *et al* 2015, Schlenker and Roberts 2009, Baylis *et al* 2011), I consider two further suites of regression models (a) base specification stratified by group of six AEZs and (b) Spatial Lag of Covariates (SLX). I then discuss the degree of improvement these further refinements can achieve over the base specification considered in Chapter 2. For brevity of space, all regression specifications in this chapter are re-run on the panel data for crop maize (Chapter 2), for 1972~2089 RCP 8.5. The predictions for the out-of-sample (2090~2099 RCP 8.5) would thus enable a comparison with the corresponding out-of-sample estimates made in Chapter 2. Unless explicitly stated otherwise, the definitions/interpretations of all parameters (such as growing season months, methodology to calculate relative bias, interpretation of coefficient estimates etc.) remain the same as in Chapter 2.

The proceedings of this chapter are in preparation for submission to Journal of Agriculture and Forest Meteorology, and are co-authored by Enrica De Cian and Ian Sue Wing. Baring few minor superficial changes to the figures and minor changes to the text, the manuscript is largely unchanged from the version of the paper under preparation. I designed and performed research, analyzed the data and wrote the paper. Enrica De Cian and Ian Sue Wing provided scientific input. All co-authors are involved in the revision of the final text for submission to the journal.

Main Text

3.1 Vapor Pressure Deficit (VPD)

VPD is a meteorological variable that measures the dryness of the atmosphere thereby providing an indication of the current evaporation potential of the air. Expressed in standard pressure units such as millibars (mb) or hectoPascals (hPa)[25], VPD is commonly used as a guidance parameter in agriculture to determine condensation threat, as well as irrigation and environmental control management decisions (Wang *et al* 2004). As emphasized by Anderson (1936) -"the strain under which an organism is placed in maintaining a water balance during temperature changes is much more clearly shown by noting the VPD than by recording the relative humidity (RH)"-, the advantage of using VPD over RH is that the former is an absolute measure of atmospheric moisture independent of temperature (Seager *et al* 2015). VPD has also been discussed in recent agronomic studies (notably Lobell *et al* 2013, Cai *et al* 2012) and in the absence of soil moisture data[26], can be considered as a useful proxy[27] for determining the plant water stress. The optimum range for VPD varies with both crop type as well as its growth stage (see table 3.1 for a general guidance on range of optimum VPD values for crops).

Table 3.1. Typical optimum ranges of VPD for most crops (Source: http://www.just4growers.com/). Orange: ideal; White: Acceptable, Light blue: too humid; Dark blue: too dry.

TEMP		RELATIVE HUMIDITY													
°C	°F	100%	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%
15	59	0.0	0.8	1.7	2.5	3.4	4.2	5.1	5.9	6.8	7.6	8.5	9.4	10.2	11.1
16	61	0.0	0.9	1.8	2.8	3.7	4.6	5.5	6.4	7.3	8.2	9.1	10.0	10.9	11.8
17	63	0.0	1.0	2.0	2.9	3.9	4.9	5.8	6.8	7.8	8.8	9.7	10.6	11.6	12.6
18	64	0.0	1.0	2.0	3.1	4.1	5.1	6.2	7.2	8.2	9.3	10.3	11.3	12.4	13.4
19	66	0.0	1.1	2.2	3.3	4.4	5.5	6.6	7.7	8.8	9.9	11.0	12.1	13.2	14.3
20	68	0.0	1.2	2.4	3.5	4.7	5.9	7.0	8.2	9.4	10.6	11.7	12.8	14.0	15.2
21	70	0.0	1.2	2.4	3.7	4.9	6.2	7.4	8.6	9.9	11.1	12.4	13.7	14.9	16.1
22	72	0.0	1.3	2.6	3.9	5.3	6.6	7.9	9.2	10.5	11.9	13.2	14.5	15.8	17.2
23	73	0.0	1.4	2.8	4.2	5.6	7.0	8.5	9.9	11.3	12.7	14.1	15.4	16.8	18.2
24	75	0.0	1.5	3.0	4.5	5.9	7.4	8.9	10.4	11.9	13.4	14.9	16.4	17.9	19.4
25	77	0.0	1.6	3.2	4.8	6.4	8.0	9.5	11.1	12.7	14.3	15.9	17.4	19.0	20.6
26	79	0.0	1.7	3.4	5.1	6.7	8.4	10.1	11.8	13.4	15.1	16.8	18.4	20.1	21.8
27	81	0.0	1.8	3.5	5.3	7.1	8.9	10.7	12.4	14.2	16.0	17.8	19.6	21.3	23.1
28	82	0.0	1.9	3.8	5.7	7.6	9.5	11.4	13.3	15.1	17.0	18.9	20.7	22.6	24.5
29	84	0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0	22.1	24.1	26.1
30	86	0.0	2.1	4.2	6.4	8.5	10.6	12.7	14.8	17.0	19.1	21.2	23.3	25.4	27.5
31	88	0.0	2.2	4.5	6.7	9.0	11.2	13.4	15.7	17.9	20.2	22.4	24.6	26.9	29.1
32	90	0.0	2.4	4.7	7.1	9.5	11.9	14.2	16.6	19.0	21.3	23.7	26.1	28.4	30.8
33	91	0.0	2.5	5.0	7.5	10.0	12.5	15.0	17.6	20.1	22.6	25.1	27.6	30.1	32.6
34	93	0.0	2.7	5.3	8.0	10.6	13.3	15.9	18.6	21.2	23.9	26.5	29.2	31.8	34.5

[25] $1mb = 1hPa$

[26] Unlike data for soil moisture, both mean daily RH and mean daily T are available in ISIMIP-FT.

[27] By itself though, VPD is not a direct or actual measurement of water loss/needs for crops.

Mathematically, VPD is the difference between the saturation vapor pressure (e_s) of the air and its water vapor content or actual vapor pressure(e_a), calculated as shown below in equations 3.1-3.3 (Anderson 1936). Given mean daily relative humidity (RH) in percent and mean daily temperature (T) in degree Celsius ($°C$),

$$VPD(hPa) = e_s - e_a \tag{3.1}$$

where

$$e_s(hPa) = 6.11e^{[\frac{17.27T}{T+237.3}]} \tag{3.2}$$

$$e_a(hPa) = e_s \times \frac{(100-RH)}{100} \tag{3.3}$$

3.2 Results with VPD as an added predictor variable

3.2.1 Regression analyses (1972~2089 RCP 8.5)

I revisit the base specifications defined in Chapter 2 for the two sets of GGCMs (i.e. equation 2.2 for GGCMs GEPIC, LPJ-GUESS and LPJmL; and equation 2.3 for GGCMs GAEZ-IMAGE, pDSSAT and PEGASUS). However, in addition to T and P, the vector of meteorological covariates ($\boldsymbol{X}$) would now also include VPD. I refer to the new specification as the base_VPD specification.

The bins $\{VPD_1, \ldots, VPD_l\}$ are counts of number of days over the growing season at each grid-cell spent in L intervals of $VPD(hPa)$ where: $L = \{<3,\ 3\sim5,\ 5\sim7,\ 7\sim9,\ 9\sim11,\ >11\}$[28]

3.2.1 Assessing emulator goodness of fit (GOF), base_VPD specification

I repeat the out-of-sample validation as done earlier in Chapter 2 to evaluate the relative bias (RB) of each of the six GGCM emulators. Figure 3.1A shows the distribution of RB weighted by grid-cell mean annual production in tons (t), for maize, in 2090~2099 under RCP 8.5 scenario. In order to assess the spatial pattern of the predicted changes in future emulator yields with those of the GGCMs', figure 3.1B illustrates the RB of the emulators geographically.

[28] The bin $L = 7\sim9\ hPa$ is omitted in regressions as reference category. As with T and P; for each VPD bin except the extreme lower and upper bins, the lower range is included in the count. The extreme bins are open-ended.

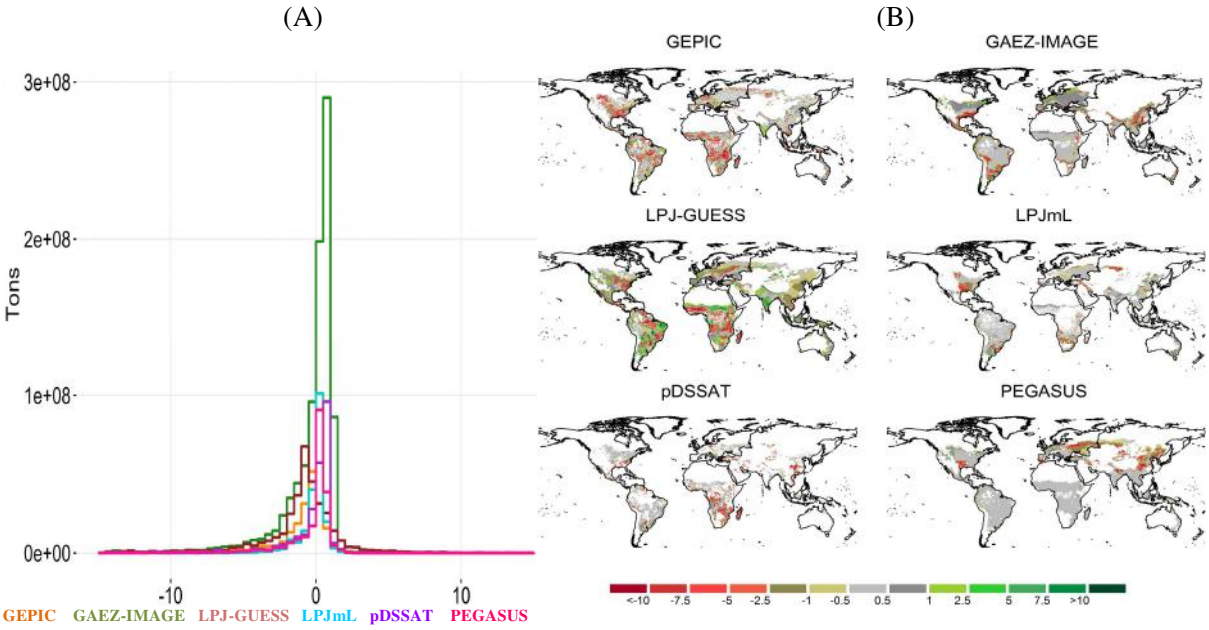


Figure 3.1 (A) Distribution of RB (%) weighted by grid-cell mean annual production $(tons, t)$ and **(B)** Performance of six maize emulator (RB) in predicting % changes; for maize in 2090~2099 RCP 8.5, relative to 1972~2089 baseline.

I find that including VPD as a predictor variable by and large does not improve the emulators' GOF vis-à-vis the base specification. Although the estimated coefficients of VPD are statistically significant ($p < 0.05$) across all GGCMs (table 4B in Appendix B), the predictions of the individual emulators (gauged by the RB) are only marginally better in comparison to the base specification. This is evident when comparing the figures 3.1A and 3.1B with the corresponding 2090~2099 RCP8.5 scenario figures 2.3A(c) and 2.4(c) of Chapter 2. Moreover, as summarized in table 3.2, for each emulator the share of maize production (%) by RB intervals are near identical for the base and base_VPD specifications (Table 2.1 of Chapter 2).

Table 3.2. Share of maize production (%) with the corresponding RB intervals, for 2090~2099 RCP8.5 using base_VPD and base (in square braces) specifications.

RB	GEPIC (%)	GAEZ-IMAGE (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)
> 10	0.17 [0.13]	0.02 [0.02]	1.35 [1.51]	0.05 [0.04]	0.00 [0.01]	0.06 [0.07]
5 ~ 10	0.13 [0.05]	0.04 [0.02]	1.33 [1.41]	0.08 [0.05]	0.01 [0.01]	0.09 [0.04]
0 ~ 5	17.78 [14.98]	50.86 [51.26]	21.51 [22.52]	40.43 [39.29]	55.98 [55.93]	54.51 [58.64]
0 ~ -5	70.71 [73.08]	43.53 [42.70]	63.37 [64.44]	57.72 [58.84]	37.53 [37.06]	40.08 [35.39]
-5 ~ -10	5.52 [5.78]	3.15 [3.34]	6.11 [6.11]	0.87 [0.89]	3.27 [3.44]	2.90 [3.04]
< -10	5.70 [5.98]	2.41 [2.26]	6.32 [6.00]	0.86 [0.90]	3.22 [3.56]	2.36 [2.81]

3.3 Robustness checks with additional regression specifications

To ensure my results did not depend on an overly specific choice of regression specification, and in line with common practice in statistical modelling (e.g. Urban *et al* 2015, Schlenker and Roberts 2009, Baylis *et al* 2011), I reanalyze the multi-GGCM dataset of weather and yields using two additional regression models[29].

3.3.1 Base specification stratified by Agro-Ecological Zones (AEZ)

A common criticism of empirical approaches in agronomic studies draws from the notion that they have limited (agronomic) meaning, even if statistically correct (Lobell *et al* 2011). For instance, the dynamics of plant water uptake (or soil-plant-atmosphere continuum) are conveniently ignored, either due to lack of data or for simplicity. Moreover, the base specification (in Chapter 2) was built on the assumption that the response of crop yields to weather, are uniform across all grid cells. The coefficient estimates when not differentiated according to the responses of crop growth process by regions (e.g. having different soil types), can become questionable (Cai *et al* 2012).

Driven by these reasons, I begin with a specification that tests the spatial stability of the base parameters by geographically stratifying $\boldsymbol{\beta}$ by six broad crop suitability regions (indexed by z) derived from Lee et

[29] The additional specifications are run with T and P as covariates, in line with the base specification of Chapter 2.

al's (2005) agro-ecological zones (AEZs). Each AEZ is formed as a combination of the duration of crop growing period and a climate region. The eighteen AEZs originally defined are further consolidated into six broader AEZs[30] (details in Section 1 of Appendix B).

$$y_{i,t} = \Sigma_z \boldsymbol{X}_{i(z),t}\boldsymbol{\beta}_z + \alpha_i + \varepsilon_{i,t} \quad (3.4)$$

As evident from equation (3.4), stratifying $\boldsymbol{\beta}$ would amount to interacting the individual T and P bins with a dummy variable (AEZ-Group), thus involving a further number of T and P interaction terms in the regression. I therefore restrict my analyses to the group of GGCMs (namely GAEZ-IMAGE, pDSSAT and PEGASUS) that did not include the time trend interaction with covariates (equation 2.3 in Chapter 2).

3.3.2 Spatial Panel Model (SPM)

As shown by Auffhammer *et al* (2013), climate variables exhibit inherent correlation across space and time. However, most empirical work in agronomic literature (e.g. Schlenker and Roberts 2009, Lobell *et al* 2012) implement robust standard errors (SEs) as proposed by Conley (1999) or (Hsiang 2010). This makes inferences 'robust', conditional on the spatial dependence (of unknown form) being confined to the error term. Nevertheless, by using robust SEs (as also implemented in Chapter 2), spatial dependence in the data generating process (here climate variables) do not get addressed. Not accounting for spatial dependence in dependent and/or independent variables could lead to biased and/or inconsistent estimates of the coefficients (Elhorst 2010, Auffhammer *et al* 2013). In contrast, the cost of ignoring spatial dependence in the disturbances if any, results only in a loss of efficiency (Elhorst 2010). Both these potential important implications could not be more relevant than in a gridded panel data, like the one in this study constructed from the ESM data, at a fine scale resolution. Subsequently, a key takeaway of this study reiterates that the coefficient estimates from studies not accounting for spatial dependence of any form need to be taken with caution (Baylis *et al* 2011).

My second specification for robustness check includes spatially lagged regressors constructed with a weight matrix (W) as additional explanatory variables. The Spatial Lag of Covariates model (SLX) described by equation (3.5) is an extension of the base specification (equation 2.3):

$$y_{i,t} = \boldsymbol{W}\boldsymbol{X}_{i,t}\boldsymbol{\lambda} + \boldsymbol{X}_{i,t}\boldsymbol{\beta} + \alpha_i + \varepsilon_{i,t} \quad (3.5)$$

[30] To facilitate comparison of results, the six broad zones used in the AEZ specification are grouped using the same definition as in Blanc and Sultan (2015).

where $\boldsymbol{W}$ is a row-standardized matrix of spatial weights constructed using the "k-nearest neighbour" method ($k = 4$) [31], $\boldsymbol{\beta}$ identifies the direct effects, and the spatial lag parameter, $\boldsymbol{\lambda}$, captures the spillover effects of meteorological predictors in surrounding cells. Following the reasons cited for restricting the number of interaction terms, I examine the suitability of SLX on the same set of GGCMs that were used in the AEZ specification. The regression results of the three GGCMs are summarized in table 6B of Appendix B.

3.4 Assessing emulator GOF, using AEZ and SLX specifications

3.4.1 AEZ Specification

I see evidence that response of crop yields to weather variables varies geographically, thus implying differential responses of crop yields under heterogeneous cropping zones. The substantial heterogeneity in the coefficients across the six AEZs is noticeable from the regression summary of the three GGCMs (table 5B in Appendix B). The improvement in results compared to the base specification are marginal and in line with earlier findings of Blanc and Sultan (2015).

For a comparison of results with the base specification, I repeat the out-of-sample validation as done earlier for base_VPD specification. To facilitate easier comparison between the base and AEZ specifications (as well as between base and SLX specifications), I overlay their respective GGCM distributions of RB weighted by grid-cell mean annual production (figure 3.2).

[31] Constructing W with alternate forms (such as distance cut off), as well as changing the k parameter, did not alter the results significantly. This is in line with findings in spatial econometrics literature -e.g. LeSage and Pace (2014)-. Due to brevity of space, although other SPMs (such as Spatial Durbin Error Model, SDEM) were examined to capture the true data generating process (DGP), the discussion is kept limited to the SLX model. Nonetheless, for a more comprehensive discussion of different spatial models, readers are guided to LeSage 2008 and Elhorst 2010.

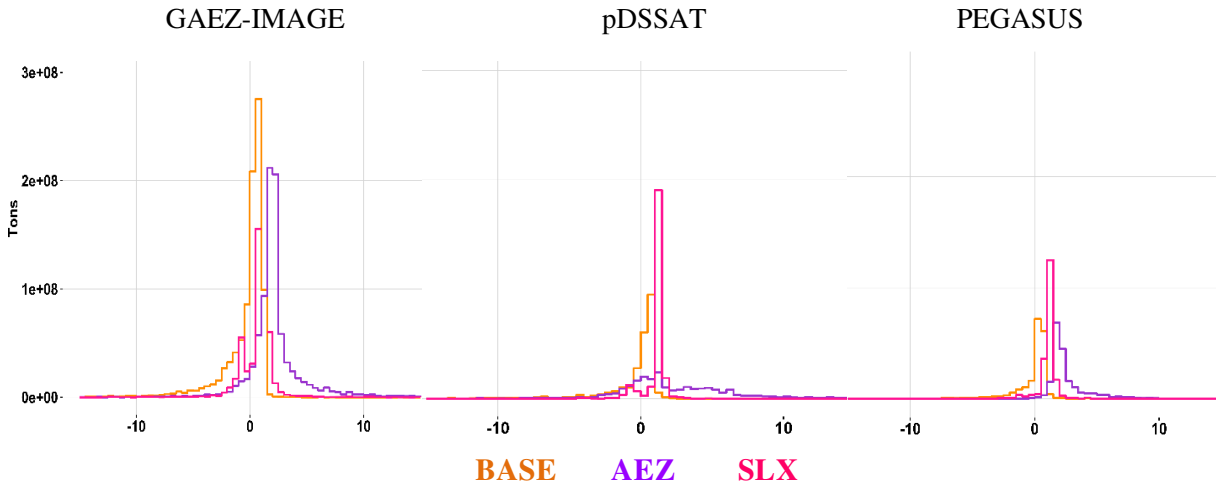


Figure 3.2 Distribution of RB weighted by grid-cell mean annual production (t) for maize in 2090~2099 RCP 8.5, relative to 1972~2089 baseline. The three specifications (a) Base (orange) (b) AEZ (purple) and (c) SLX (pink) are overlaid for easier interpretation

In each of the three GGCM emulators, the distribution for AEZ specification is leptokurtic (fatter tails or positive excessive kurtosis) compared to the platykurtic distribution (thinner tails or negative excessive kurtosis) of the base specification. This is most prominent for pDSSAT, implying higher RB across more grid-cells with lower share of mean annual production. The distributions for the AEZ specifications are also negatively skewed, compared to the slight positive skew of the corresponding base specifications for GAEZ-IMAGE, and near normal for pDSSAT and PEGASUS. This implies a general tendency of the AEZ specification to estimate lesser negative impacts (or higher positive benefits) vis-à-vis the corresponding GGCMs' actual estimates. Both these characteristics (kurtosis and skewness) imply generally insignificant improvements (if not poorer) in results for the AEZ in comparison to the base specification.

To further facilitate the comparison in results of between the base and AEZ specifications, figure 3.3 shows empirical cumulative distribution function (ECDF) for the same crop out-of-sample validation. The ECDF of AEZ specification across all three GGCM emulators has a similar positive shift in RB in comparison to the corresponding base specification. This implies a general tendency of the AEZ specification to systematically over-predict vis-à-vis the corresponding GGCMs' actual estimates.

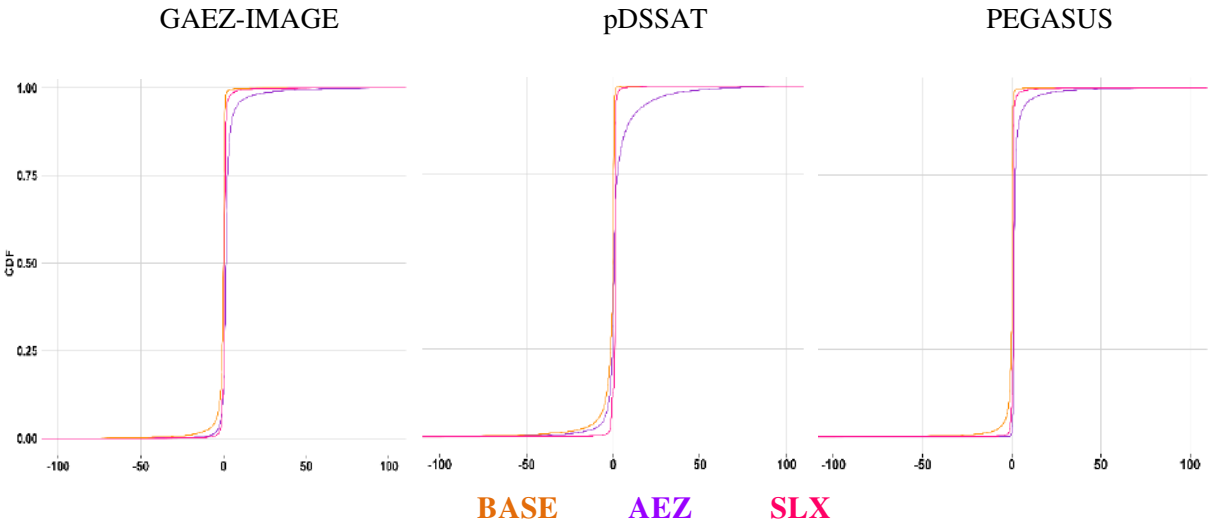


Figure 3.3 ECDF of RB weighted by grid-cell mean annual production (t) for maize in 2090~2099 RCP 8.5, relative to 1972~2089 baseline. The three specifications (a) Base (orange) (b) AEZ (purple) and (c) SLX (pink) are overlaid for easier interpretation

It must be noted that following the data cleaning steps in the GGCMs' panels, and the subsequent grouping of the grid-cells into six AEZs; the sample size (number of grid-cells) in most AEZs reduce remarkably (table 3B in Appendix B). Consequently, the AEZ specification gives statistically insignificant $(p > 0.05)$ estimates for a large number of coefficients across each AEZ (table 5B in Appendix B). It is therefore plausible that in a larger and reliable sample, using coefficients for spatially heterogeneous zones would improve the results even further.

3.4.2 SLX Specification

Focusing back on figure 3.2, the distribution of the SLX specification (pink) shows a clear improvement in using a specification accounting for spatial dependence in the explanatory variables. The same two characteristics (kurtosis and skewness) reveal patterns that not only imply lower RB across more grid-cells with lower share of mean annual production, but also a near neutral mean bias suggesting the SLX specification agrees more with the corresponding GGCMs' actual estimates compared to the base specification. The same is even more prominent when comparing the ECDFs of the SLX (pink) and base (orange) specifications in figure 3.3. The bias in the variance is much lower for the SLX compared to the base specification.

Conclusions

The simple specification used in construction of emulators in Chapter 2 were refined here and put through a series of comparison and robustness checks. Including an additional predictor variable (VPD) to the base specification of Chapter 2 did not reveal significant differences in results, thus retaining the model specification using only T and P as the preferred choice of predictor variables.

To investigate the potential geographic heterogeneity of the response of crop yields to weather, the next exercise in my sensitivity and robustness checks made use of AEZs. Modifying the base specification of Chapter 2, regressions were repeated on three GGCMs allowing for the slope parameter to interact with six groups of AEZs carved out of the original eighteen AEZs. The results show that although differentiating the coefficient estimates of the covariates has a basis for stratifying the responses of crop yields geographically, the smaller sample size in the AEZ specification does not improve the overall bias of the three emulators' vis-à-vis their base specification.

Finally, utilizing the potential strengths of SPMs from a relatively new subject of spatial econometrics, I investigated the suitability of a simple SPM (i.e. SLX) in the construction of the emulator. The results of the SLX model reveal the importance of explicitly accounting for spatial dependence in a fine scale gridded data, in order to obtain reliable and consistent estimates. A detailed examination of various SPMs could potentially lead to a further improvised version of an emulator and is recommended for future work. However, the further improvement in results could come at the cost of added complexity, thus deviating from the core theme of building an emulator, that of simplicity and flexibility.

Closing remarks

Taking a step further in chapter 4, the focus shifts from global to regional spatial domains by restricting the GGCM data to United States (U.S.) counties. Moreover, for a head-to-head comparison with the coefficient estimates derived from the GGCM data, historical observed U.S. county crop yields from U.S. Department of Agriculture (USDA) are utilized for calibrating an empirical model. The comparison will throw light on divergence of estimates made using the two sets of data, and implications of the differences between the GGCMs' aggregated responses under future climate change scenarios.

Chapter 4: Simulated vs. Empirical Weather Responsiveness of Crop Yields: U.S. Evidence and Implications for the Agricultural Impacts of Climate Change

Preface

This chapter focuses on an Inter-method comparison between the coefficient estimates of a statistical emulator (calibrated on data from GGCMs as done in Chapter 2), with empirical models built on historical observed data. Shifting focus from global to regional scales, the data used here (for both GGCMs and historical observed) spans the United States (U.S.) counties. For calibrating empirical models, I utilize the historical observed crop yields from the U.S. Department of Agriculture (USDA).

The core objective of this study is to compare:

(i) GGCM simulated historical mean yields (1972-2004) with the actual historical observed yields for the U.S. counties (1972-2004)

(ii) The estimated coefficients of the temperature (T) and precipitation (P) bins across the six emulators (from regressions run on ISIMIP-FT data for the U.S. counties)

(iii) The emulators' estimated T and P coefficients with those of the empirical model (calibrated on USDA panel data); and the subsequent implications on predicted yield changes under future climate warming scenario.

The proceedings of this chapter co-authored by Enrica De Cian and Ian Sue Wing, are submitted to the special issue of Environmental Research Letters (ERL): Focus on An Inter-method Comparison of Climate Change Impacts on Agriculture. Baring few minor superficial changes to the figures and minor changes to the text, the manuscript is largely unchanged from the version of the paper submitted. I designed and performed research with key scientific inputs from Enrica De Cian and Ian Sue Wing. Further, I analyzed the data and wrote the paper. All co-authors were involved in the revision of the final paper submitted to ERL, with the same title.

Main Text

Abstract

Global gridded crop models (GGCMs) are the workhorse of assessments of the agricultural impacts of climate change. Yet the changes in crop yields projected by different models in response to the same meteorological forcing can differ substantially. Through an inter-method comparison, we provide a first glimpse into the origins and implications of this divergence—both among GGCMs and between GGCMs and historical observations. We examine yields of rainfed maize, wheat, and soybeans simulated by six GGCMs as part of the Inter-Sectoral Impact Model Intercomparison Project-Fast Track (ISIMIP-FT) exercise, comparing 1972-2004 hindcast yields over the coterminous United States (U.S.) against U.S. Dept. of Agriculture (USDA) time series for >1,000 counties. Leveraging the empirical climate change impacts literature, we estimate reduced-form econometric models of crop yield responses to temperature and precipitation exposures for both GGCMs and observations. We find that 15-50% of the variance in both simulated and observed yields is attributable to weather variation. GGCMs have difficulty reproducing the observed distribution of percentage yield anomalies, and exhibit aggregate responses that show yields to be more weather-sensitive than in the observational record—not only in response to adverse exposures to extreme high temperature or low precipitation, but over the entire range of heat and moisture conditions. This disparity is largely attributable to heterogeneity in GGCMs' responses, as opposed to uncertainty in historical weather forcings, and is responsible for widely divergent impacts of climate on future crop yields.

4.1 Introduction

Climate change will adversely impact a wide range of human systems (IPCC 2014). The agriculture sector, particularly field crops, is especially vulnerable because production is both strongly weather dependent and exposed to meteorological shifts. Exposure of maize, wheat, soybeans and other food staples to high temperature and low precipitation extremes portend declining yields. This has been forcefully demonstrated by the empirical climate change economics literature, through statistical estimation of reduced-form responses of yields to weather shocks using historically observed production, harvested area, temperature and precipitation in many locations across multiple years (e.g. Lobell *et al* 2011, Porter *et al* 2014, Schlenker and Lobell 2010, Tack *et al* 2015). Additional evidence comes from process-based crop models, which simulate the detailed influences on plant growth of a wide array of weather variables, plant genotypes, environmental factors such as the carbon dioxide (CO_2) fertilization effect (CFE), soil quality or pests, and agronomic adaptations such as irrigation, fertilizer application, and the timing of planting and harvesting (Elliott *et al* 2014, Bassu *et al* 2014, Rosenzweig *et al* 2014).

Whereas the geographic domain of empirical studies is often limited to individual countries or regions with a sufficient number of historical observations[32], a growing number of process-based global gridded crop models (GGCMs) generate results on crop growth under different climatic conditions driven by earth system model (ESM) projections at the grid cell level across the globe (see Deryng *et al* 2011, Rosenzweig *et al* 2014, and Elliott *et al* 2014 for further discussion). This capability enables GGCMs to generate a consistent picture of climate change impacts on crop yield at broad spatial scales.

However, confidence in the resulting projections of agricultural impacts turn on the fidelity with which GGCMs capture the effects of changing meteorology on yields. GGCMs represent the dynamics of plant growth through a high number of parameters that require calibration, whose values are uncertain and may vary geographically. Customary techniques for validating parameterized models involve statistical evaluation of their ability to reproduce point estimates of yields at different locations, for example at field trial sites or over spatially aggregated production regions under year-to-year variation in weather conditions (for excellent recent examples, see Morell *et al* 2016, Müller *et al* 2016 *under review*).

However, comparatively little attention has been paid to how the aggregate responses of yields to heat and moisture simulated by GGCMs stack up against corresponding empirically-derived responses for real-world agricultural systems[33]. Recent applications of econometric modeling techniques to cross-section/time-series datasets of crop yields generated by GGCM inter-comparison exercises have focused on constructing reduced-form statistical emulators of single (Oyebamiji *et al* 2015) or multiple-GGCM (Blanc and Sultan 2015) simulations of one or more crops. However, we are not aware of published head-to-head comparison between process simulations and econometric models trained on observations. It is this gap in the literature that we seek to address.

Our strategy is to elucidate and compare the aggregate responses of observed and GGCM-simulated yields to observed and ESM-simulated temperature and precipitation under current climatic conditions. We pose six key questions:

Q.I How well do the outputs of GGCM hindcast simulations match historically observed yields?

Q.II Are GGCMs able to reproduce the correlations between observed yields and adverse (i.e., high temperature and low precipitation) weather extremes seen in the observational record?

[32] For examples, see Iglesias *et al* 2000 for Spain, Lobell and Burke 2010 for U.S. counties, Lobell *et al* 2012 for India, Schlenker and Lobell 2010 for Sub-Saharan Africa.

[33] For instance, see (Lobell and Burke 2010 and Watson *et al* 2015), although both discussions are limited to a single crop model.

Q.III How similar are GGCM-simulated and observed yield responses, under not only adverse extremes, but the full range of weather conditions over crops' growing seasons.

Q.IV Do differences between GGCMs and observations in the weather-responsive component of yields arise primarily because of divergent meteorological forcings (i.e., the difference in exposures between reanalysis data and ESM historical simulations) or divergence in GGCMs' simulated responses and actual agricultural system responses to these forcings?

Q.V What do the estimated response functions imply for the impacts of climate change-driven shifts in temperature and precipitation on future United States (U.S.) crop yields?

Q.VI Which GGCMSs' attributes are correlated with the divergence of crop yield responses from the empirical estimates based on historical observations?

To obtain answers we use statistical methods to extract and compare the responses of yield to weather shocks for two sets of data that span the same temporal and spatial domain: rainfed maize, wheat and soybeans in the coterminous U.S. over the period 1972-2004. For crop models we use the outputs of runs of a suite of six GGCMs fielded by the Inter-Sectoral Impact Model Intercomparison Project Fast-Track (ISIMIP-FT) exercise (Warszawski *et al* 2013, Rosenzweig *et al* 2014, Frieler *et al* 2015), along with their meteorological forcings (Hempel *et al* 2013). For historical observations we use U.S. Dept. of Agriculture (USDA) multi-decadal time series of production and harvested area at the fine spatial scale of counties—whose areal extents are comparable in size to GGCMs' grid cells across U.S. farm states, matched to high-frequency weather exposures from a climate reanalysis dataset.

The rest of the chapter is organized as follows. Section 4.2 discusses our data and elaborates the methods we use to answer questions I-V. A discussion of the results is provided in section 4.3. We summarize our findings with the associated caveats and recommendations for future research in section 4.4.

4.2 Methods

Our data consist of a set of unbalanced panels of maize, wheat and soybean yields (Y) that are either observed or modeled at i areal units over t years, and matching daily observed or simulated growing season temperature (T) and precipitation (P) for the same locations and periods. Historical crop yields were computed from U.S. county records of production and harvested area tabulated by the USDA

National Agricultural Statistics Service's Quickstats 2.0 database.[34] Simulated hindcast yields are drawn from (Rosenzweig *et al* 2014) for six GGCMs: GEPIC (Liu *et al* 2007), GAEZ-IMAGE (van Vuuren *et al* 2006), LPJ-GUESS (Sitch *et al* 2003), LPJmL (BONDEAU *et al* 2007, Sitch *et al* 2003), pDSSAT (Elliott *et al* 2013, Jones *et al* 2003) and PEGASUS (Deryng *et al* 2011). GGCMs produce yields on a 0.5° grid, and are forced by historical bias-corrected meteorology simulated by the HadGEM2-ES climate model (Jones *et al* 2011) at the same resolution.[35] Our source of historical weather was the Global Land Data Assimilation System (GLDAS2) forcing files of 3-hourly meteorological fields on a 1° grid (Rodell *et al* 2004), spatially interpolated to U.S. counties. Further details of the data and models are given in sections 1-3 of Appendix C.

Assessing GGCMs' skill (Q.I) is not as simple as it might seem, since the GAEZ-IMAGE and LPJ-GUESS models simulate potential yields, while the remainder simulate actual yields, making apples-to-apples comparison difficult. As well, different models are calibrated using historical yields from different sources, whereas others are not calibrated (see Rosenzweig *et al* 2014 SI for further details) Mindful of these caveats, our approach is to characterize the distribution of the differences between the cross-section/time-series yield anomalies of GGCMs and observations, ${}^{*}Y_{i,t}^{GGCM} - {}^{*}Y_{i,t}^{USDA}$. To facilitate comparison, we use a normalization that expresses the anomalies as fractional deviations from each location's long-run mean, ${}^{*}Y_{i,t} = Y_{i,t}/\bar{Y}_i - 1$. If ${}^{*}Y_{i,t}^{GGCM}$ and $Y_{i,t}^{USDA}$ are similar, then we would expect the probability density function (PDF) of the anomaly difference (defined above) to be sharply peaked with zero mean.

To address question Q.II we elucidate the covariation between yield anomalies and adverse weather—which we define as high-temperature and low-precipitation extremes in the respective forcing datasets. We apply a fixed annual growing season[36] to our GGCM input and climate reanalysis data, within which we calculate the cumulative days of each county's exposure to j intervals of temperature, ξ_j^T, and k intervals of precipitation, ξ_k^P. For each county we then compute the temporal correlations between ${}^{*}Y_i$ and the extreme bins of these variables ($j: T > 30$°C, $k: P \leq 5$mm) in our ESM-simulated and observational datasets.[37]

[34] http://quickstats.nass.usda.gov/

[35] GGCM inputs and outputs were downloaded from the ISIMIP-FT archive: https://esg.pik-potsdam.de/search/isimip-ft/

[36] For both datasets, we define the growing season as May-August (MJJA). See section 1.3 in Appendix C

[37] See section 3 in Appendix C for details of binning intervals used in regressions.

Our answer to Q.III extends the foregoing analysis to the entire range of growing season temperature and precipitation exposures, and constitutes the meat of the paper. We quantify the potentially nonlinear influence of climate on yields using a semi-parametric cross-section/time-series regression model of the kind developed in the empirical climate-change impacts literature (Schlenker and Roberts 2006, 2009, Deschênes and Greenstone 2007, 2012, Lobell *et al* 2011, Ortiz-Bobea 2013, Burke and Emerick 2015). The dependent variable is the natural logarithm of annual yield (y), predictors are a vector of location-specific effects (μ, which capture the influence of unobserved time-invariant local characteristics such as topography and soils), a vector of time effects (τ, which capture the influence of unobserved common time-varying shocks) and the vectors of climatic covariates ξ_j^T and ξ_k^P described above, while ε is a random disturbance term:

$$y_{i,t} = \mu_i + \tau_t + \Sigma_j \beta_j^T \xi_{j,i,t}^T + \Sigma_k \beta_k^P \xi_{k,i,t}^P + \varepsilon_{i,t} \quad (4.1)$$

Eq. (4.1) is estimated via ordinary least squares on our observational dataset, the datasets of simulated weather inputs and yield outputs corresponding to each of our GGCMs, and multi-model panel consisting of the combined inputs and outputs of the six GGCMs. The latter merged regression model includes a GGCM dimension along which there is likely to be idiosyncratic variation. We control statistically for this by introducing an additional model-specific factor into eq. (4.1).

Of interest in eq. (4.1) are the estimated parameters $\boldsymbol{\beta}^T$ and $\boldsymbol{\beta}^P$, vectors of semi-elasticities that indicate the percentage shift in yields relative to their conditional mean levels in response to an additional day in a given interval of heat or moisture. Each of their constituent elements captures the distinct marginal effect of exposure within the corresponding interval (e.g., the average impact of an additional day with 25-27°C versus >30°C average temperature). Collectively, the elements flexibly trace out the aggregate response of yields to heat and moisture as piecewise linear splines. The latter are statistically identified from the contemporaneous covariation between observed yields and meteorology within each interval, as well as the distribution of weather exposures across intervals in our transformed datasets.

Empirical and ISIMIP-FT studies of agricultural impacts of climate change employ different meteorological inputs, with the former using weather station observations or reanalysis datasets and the latter using the outputs of ESM simulations. The lack of standardization between the two approaches complicates comparison of USDA and GGCM responses of yield to weather and motivates Q.IV, which our empirical modeling strategy provides a way to address.

The weather-responsive component of log yield at each location is determined by the fitted temperature and precipitation semi-elasticities ($\widehat{\boldsymbol{\beta}}^T$ and $\widehat{\boldsymbol{\beta}}^P$):

$$\psi(\boldsymbol{T}_i, \boldsymbol{P}_i) = \Sigma_j \hat{\beta}_j^T \xi_{j,i}^T + \Sigma_k \hat{\beta}_k^P \xi_{k,i}^P \tag{4.2}$$

The difference between the weather-responsive components of each GGCM's historical run and the observations is

$$\Delta\psi = \psi(\boldsymbol{T}_i, \boldsymbol{P}_i)^{GGCM} - \psi(\boldsymbol{T}_i, \boldsymbol{P}_i)^{USDA} = \Sigma_j \hat{\beta}_j^{T,GGCM} \xi_{j,i}^{T,ESM} + \Sigma_k \hat{\beta}_k^{P,GGCM} \xi_{k,i}^{P,ESM} - \left(\Sigma_j \hat{\beta}_j^{T,USDA} \xi_{j,i}^{T,Reanal} + \Sigma_k \hat{\beta}_k^{P,USDA} \xi_{k,i}^{P,Reanal}\right) \tag{4.3}$$

Adding and subtracting cross-terms on the right-hand side of eq. (4.3) and evaluating the exposure covariates in the resulting expression at their 1972-2004 climatic means allows us to decompose the difference above into two terms, one of which captures the effect of divergence due to differences in climate forcing and the other capturing the effect of divergence in the response to climate of GGCMs relative to observations:

$$\Delta\psi = \underbrace{\Sigma_j \hat{\beta}_j^{T,GGCM}\left(\bar{\xi}_{j,i}^{T,ESM} - \bar{\xi}_{j,i}^{T,Reanal}\right) + \Sigma_k \hat{\beta}_k^{P,GGCM}\left(\bar{\xi}_{k,i}^{P,ESM} - \bar{\xi}_{k,i}^{P,Reanal}\right)}_{\text{Climate component } (\Delta\psi^{Climate})} + \underbrace{\Sigma_j\left(\hat{\beta}_j^{T,GGCM} - \hat{\beta}_j^{T,USDA}\right)\bar{\xi}_{j,i}^{T,Reanal} + \Sigma_k\left(\hat{\beta}_k^{P,GGCM} - \hat{\beta}_k^{P,USDA}\right)\bar{\xi}_{k,i}^{P,Reanal}}_{\text{Response component } (\Delta\psi^{Response})} \tag{4.4}$$

The relative importance of the two components can then be assessed by comparing their distributions across locations.

We address question Q.V by quantifying the changes in yields that result from combining our fitted responses with future meteorology under climate warming. We force our log yield response functions ψ with meteorological exposures from HadGEM2-ES simulations for our hindcast period, as well as mid-21st century (2033-2065) and late century (2067-2099) climates under the RCP 8.5 (Moss *et al* 2010) high-warming scenario. In each epoch the simulated daily temperature and precipitation ($\widetilde{\boldsymbol{T}}_i$ and $\widetilde{\boldsymbol{P}}_i$) fields are binned into the j and k intervals, respectively, to construct analogues of the weather covariates, $\tilde{\xi}^T$ and $\tilde{\xi}^P$, for current and future years.[38] These serve as inputs to eq. (4.2), enabling the resulting weather-responsive log yields to be used to compute a normalized multi-decadal index of climate impact, given by

[38] We assume the same growing season for future climates as for the historical period.

the ratio of each location's average yield under a future climate to its average yield under the present climate. Using the expectation operator $\mathbb{E}$ to denote average over the years within each epoch, the index is:

$$\Psi_i = \mathbb{E}\left[\exp\left\{\psi\left(\widetilde{\boldsymbol{T}}_i^{Climate^{Future}}, \widetilde{\boldsymbol{P}}_i^{Climate^{Future}}\right) - \psi\left(\widetilde{\boldsymbol{T}}_i^{Climate^{Current}}, \widetilde{\boldsymbol{P}}_i^{Climate^{Current}}\right)\right\}\right] \tag{4.5}$$

Assuming that the current geographic distribution of harvested areas persists into the future, the projected change in the production of each crop due to heat and moisture changes is simply the product of our index and the long-run mean crop yield under the current climatic conditions, $\Psi_i \times \overline{Y}_i^{Climate^{Current}}$. We stress that Ψ_i will almost surely diverge from the factional change in yields between current and future decades simulated by GGCMs. Major reasons are the CFE and climate adaptations assumptions implicitly incorporated into GGCMs models, particularly endogenous or unrecorded prescribed future changes in fertilizer application rates, crop calendars, or crop genotypes.[39] We follow Schlenker and Roberts (2009) and restrict attention to counties east of the 100th meridian (excluding Florida) where rainfed cultivation is likely to remain concentrated.

Motivated by the important shortcoming of inter-model comparison exercises in identifying which GGCMs' parameters influence their responses to weather the most; our final exercise addresses Q. VI. We undertake a thorough analysis of the GGCMs' meta-response parameters, identifying the effect of each dimension in amplifying or attenuating the differences in GGCMs' responses. Exploiting the key similarities and differences across the six GGCMs (as documented in Rosenzweig *et al* 2014, Elliott *et al* 2014), we first categorize the GGCMs by parameter dimensions[40] that likely influence the divergence in the inter-model responses. Together with the GGCMs' and USDA's estimated coefficient responses, we then formulate six different specifications (eqs. 4.6 - 4.11) to carry out robust statistical meta-analyses.

[39] For instance, see Rosenzweig *et al* 2014 SI for details on adaptations accounted for by the GGCMs, and Elliott *et al* 2014 for revised protocols in the next phase of GGCMs' simulations to introduce harmonization in GGCMs' simulation runs.

[40] The key GGCMs' characteristics used as dimension dummies in our analyses are: (i) Type of yield simulated (Actual or Potential) (ii) Cultivar adaptation (iii) Heat stress (iv) Dynamic planting window adaptation and (v) Type of calibration (See table 2A in Appendix for further details of the GGCMs' broader characteristics). To account for reduced number of observations and loss of degrees of freedom, the parameter dummies (iv) and (v) are omitted in eqs. 4.8 and 4.11 (see table 4.1 in Section 3).

The dependent variable in eqs. 4.6 and 4.9 is a vector of differences for each of the GGCMs' and USDA's $\widehat{\boldsymbol{\beta}}^{T}$ and $\widehat{\boldsymbol{\beta}}^{P}$(from eq. 4.1), thus amounting to 72 observations[41]. For explanatory variables, in addition to the parameter dimensions that are used as a vector of dummies in eq. (4.6), the specification in eq. (4.9) also incorporates the interaction of the parameter dummies with the extreme high temperature and low precipitation bins, defined by dummies *ht* and *lp,* as $\{25\sim27.5, 27.5\sim30, > 30\}℃$ and $\{< 5, 5\sim10\}\ mm/day$ respectively[42].

In contrast to the combined set of $\widehat{\boldsymbol{\beta}}^{T}$and $\widehat{\boldsymbol{\beta}}^{P}$ used as a dependent variable in eqs 4.6 and 4.9; eqs. 4.7, 4.8, 4.10 and 4.11 now restrict the analyses to the individual set of temperature and precipitation coefficients. The dependent variable therefore is a vector of differences for each of the GGCMs' and USDA's $\widehat{\boldsymbol{\beta}}^{T}$(eqs. 4.7, 4.10) and $\widehat{\boldsymbol{\beta}}^{P}$(eqs. 4.8, 4.11), thus corresponding to 60 and 12 observations (for the six GGCMs).

$$\Delta\zeta = \zeta\left(\widehat{\boldsymbol{\beta}}^{T}, \widehat{\boldsymbol{\beta}}^{P}\right)^{GGCM} - \zeta\left(\widehat{\boldsymbol{\beta}}^{T}, \widehat{\boldsymbol{\beta}}^{P}\right)^{USDA} = \widehat{\omega}_1\delta_1 + \widehat{\omega}_2\delta_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots \widehat{\omega}_m\delta_m + \varepsilon \tag{4.6}$$

$$\Delta\zeta = \zeta\left(\widehat{\boldsymbol{\beta}}^{T}\right)^{GGCM} - \zeta\left(\widehat{\boldsymbol{\beta}}^{T}\right)^{USDA} = \widehat{\omega}_1\delta_1 + \widehat{\omega}_2\delta_2 + \ldots\ldots\ldots\ldots\ldots\ldots \widehat{\omega}_m\delta_m + \varepsilon \tag{4.7}$$

$$\Delta\zeta = \zeta\left(\widehat{\boldsymbol{\beta}}^{P}\right)^{GGCM} - \zeta\left(\widehat{\boldsymbol{\beta}}^{P}\right)^{USDA} = \widehat{\omega}_1\delta_1 + \widehat{\omega}_2\delta_2 + \ldots\ldots\ldots\ldots\ldots\ldots \widehat{\omega}_m\delta_m + \varepsilon \tag{4.8}$$

$$\begin{aligned}\Delta\zeta = \zeta\left(\widehat{\boldsymbol{\beta}}^{T}, \widehat{\boldsymbol{\beta}}^{P}\right)^{GGCM} - \zeta\left(\widehat{\boldsymbol{\beta}}^{T}, \widehat{\boldsymbol{\beta}}^{P}\right)^{USDA} &= \widehat{\omega}_1\delta_1 + \widehat{\omega}_2\delta_2 + \ldots\ldots \widehat{\omega}_m\delta_m + \widehat{\omega}_4(\delta_1 * \delta_{ht}) + \\ &\widehat{\omega}_5(\delta_2 * \delta_{ht}) + \ldots\ldots \widehat{\omega}_n(\delta_m * \delta_{ht}) + \widehat{\omega}_7\left(\delta_1 * \delta_{lp}\right) + \\ &\widehat{\omega}_8\left(\delta_2 * \delta_{lp}\right) + \ldots \widehat{\omega}_k(\delta_m * \delta_{lp}) + \varepsilon\end{aligned} \tag{4.9}$$

$$\begin{aligned}\Delta\zeta = \zeta\left(\widehat{\boldsymbol{\beta}}^{T}\right)^{GGCM} - \zeta\left(\widehat{\boldsymbol{\beta}}^{T}\right)^{USDA} &= \widehat{\omega}_1\delta_1 + \widehat{\omega}_2\delta_2 + \ldots\ldots \widehat{\omega}_m\delta_m + \widehat{\omega}_1(\delta_1 * \delta_{ht}) + \\ &\widehat{\omega}_2(\delta_2 * \delta_{ht}) + \ldots\ldots \widehat{\omega}_n(\delta_m * \delta_{ht}) + \varepsilon\end{aligned} \tag{4.10}$$

$$\begin{aligned}\Delta\zeta = \zeta\left(\widehat{\boldsymbol{\beta}}^{P}\right)^{GGCM} - \zeta\left(\widehat{\boldsymbol{\beta}}^{P}\right)^{USDA} &= \widehat{\omega}_1\delta_1 + \widehat{\omega}_2\delta_2 + \ldots\ldots \widehat{\omega}_m\delta_m + \widehat{\omega}_1\left(\delta_1 * \delta_{lp}\right) + \\ &\widehat{\omega}_2\left(\delta_2 * \delta_{lp}\right) + \ldots\ldots \widehat{\omega}_k(\delta_m * \delta_{lp}) + \varepsilon\end{aligned} \tag{4.11}$$

[41] 13 bins (in eq. 1) x 6 GGCMs = 72 observations.

[42] $\boldsymbol{\delta} = \mathbf{1}$ for the temperature and precipitation bins defined by *ht* and *lp*, and $\boldsymbol{\delta} = \mathbf{0}$ otherwise.

Of interest in eqs. 4.6 – 4.11 are parameters $\widehat{\omega}$, coefficients of the broader characteristics of the GGCMs that indicate the average impact of the incorporated dimensions (denoted by dummy variable δ)[43], on the entire pattern of GGCM's crop yield response.

The first three sets of specifications (eqs. 4.6 – 4.8) enable us to examine the influence on the parameter dimensions averaged across all temperature and precipitation bins. In contrast, eqs. (4.9 – 4.11) enable us to attribute the key parameters that influence the divergence in GGCMs' responses to the extreme bins (as defined above), where one can anticipate the largest variation across the GGCMs.

4.3 Results

4.3.1 Differences between GGCM simulated and historically observed yield anomalies

Figure 4.1 tabulates the distributions of the differences in percentage yield anomalies between GGCMs and USDA records for our three crops over the 1972-2004 period. The wide support of the distribution suggests that the ISIMIP-FT GGCMs struggle to reproduce the PDF of yield anomalies generated by a real-world agricultural system. For the half of our county sample lying within the interquartile range the model-observation divergence is on order of -/+30%, while in the majority of remaining locations simulated yields can dramatically overstate or understate the observations.

While this pattern persists across crops, GGCMs' performance—as judged by the variance of the distributions—tends to be generally better for wheat and especially maize compared to soybeans, which exhibits much larger dispersion. The modes of the individual annual cross-county PDFs (shown in light colors) also shift substantially from one harvest to another, but these positive and negative fluctuations do not follow a predictable temporal sequence that might suggest systematic bias. The marked differences across models and crops in the annual and aggregate PDFs also suggest that no single GGCM has a clear advantage in modeling all crops. Rather, an individual GGCM may exhibit skill in modeling yields of one crop versus another (e.g., wheat relative to soybeans simulated by LPJmL), while for any given crop some GGCMs outperform others (e.g., maize simulated by GAEZ-IMAGE relative to GEPIC).

[43] $\boldsymbol{\delta = 1}$ when the parameters (in table S#) are implemented in the GGCMs, and $\boldsymbol{\delta = 0}$ otherwise.

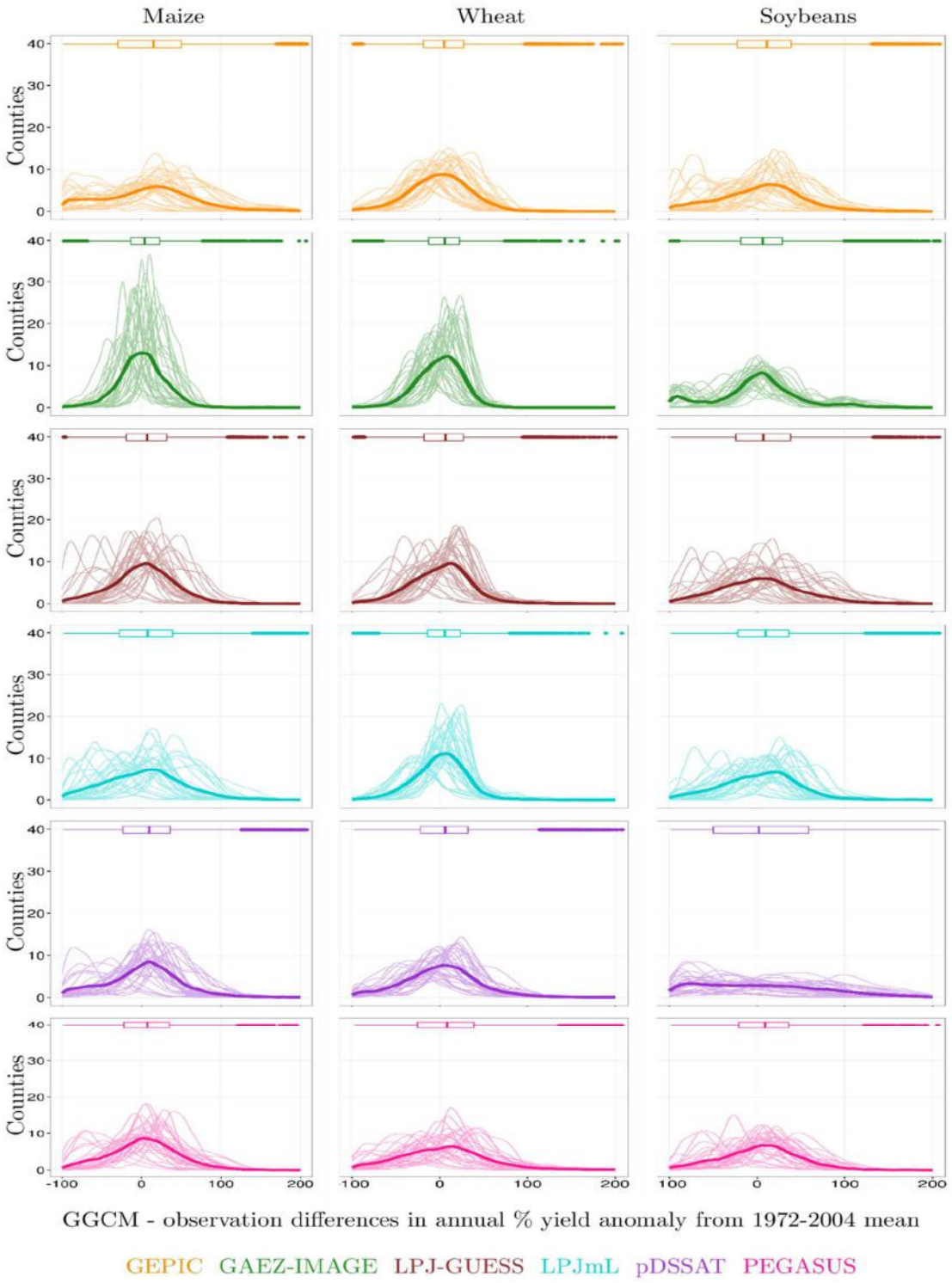


Figure 4.1. Cross-county distribution of the GGCM - USDA difference in percentage yield anomalies. Anomalies are calculated as the % deviation of every county's yield from its own 1972-2004 mean (eq. 4.1). Light lines show the annual distribution of county differences between each model and observations. Heavy lines show the distribution across counties and years.

4.3.2 Correlations between yield anomalies and extreme temperature and precipitation

While GGCMs' ability to reproduce observed yields has been comprehensively analyzed, with a focus on establishing a global-scale benchmark for valid comparisons (Müller *et al* 2016 *under review*), we argue that where such benchmarks are available (such as at sub-national scales within the U.S.), the critical focus of evaluation should be models' skill in reproducing the *response* of yields to climatic forcings observed in real-world agricultural systems. In this regard, our preliminary indicator is the historical correlations between annual yield anomalies and extreme high temperatures and low precipitation, respectively, for both GGCMs and observations.

Figure 4.2 visualizes the map of the correlations between yields and annual growing season exposures to the extreme high temperature and extreme low precipitation bins (respectively) as a bivariate density, providing a first glimpse into the origins of the divergence between GGCM-simulated and observed yields. Not surprisingly, both correlations are negative in 50-75% of counties, however the magnitudes of the correlations differ both across models and among crops. Yields are more strongly correlated with adverse weather exposures for maize and soybeans than for wheat, which exhibits a pattern of equivocal response in both the observations and the simulations (except for LPJ-GUESS and pDSSAT). However, with the exception of GAEZ-IMAGE, simulated maize and soybean responses exhibit excess weather sensitivity compared to observations, with GEPIC, LPJ-GUESS, LPJmL and pDSSAT showing tight clustering of negative impacts across counties. This may be due to differences between modeled and observed management practices, length and number of the growing seasons, and adaptation strategies.

The strength of the association between heat and moisture impacts is indicated by the best-fit line, which is generally more steeply sloped for GGCMs than for the observations, point to models' comparatively higher sensitivity to low precipitation exposures. Wheat appears to be less sensitive to both heat and low moisture, which could be attributed to multiple cropping seasons (see Müller *et al* 2016 *under review* for similar discussions).

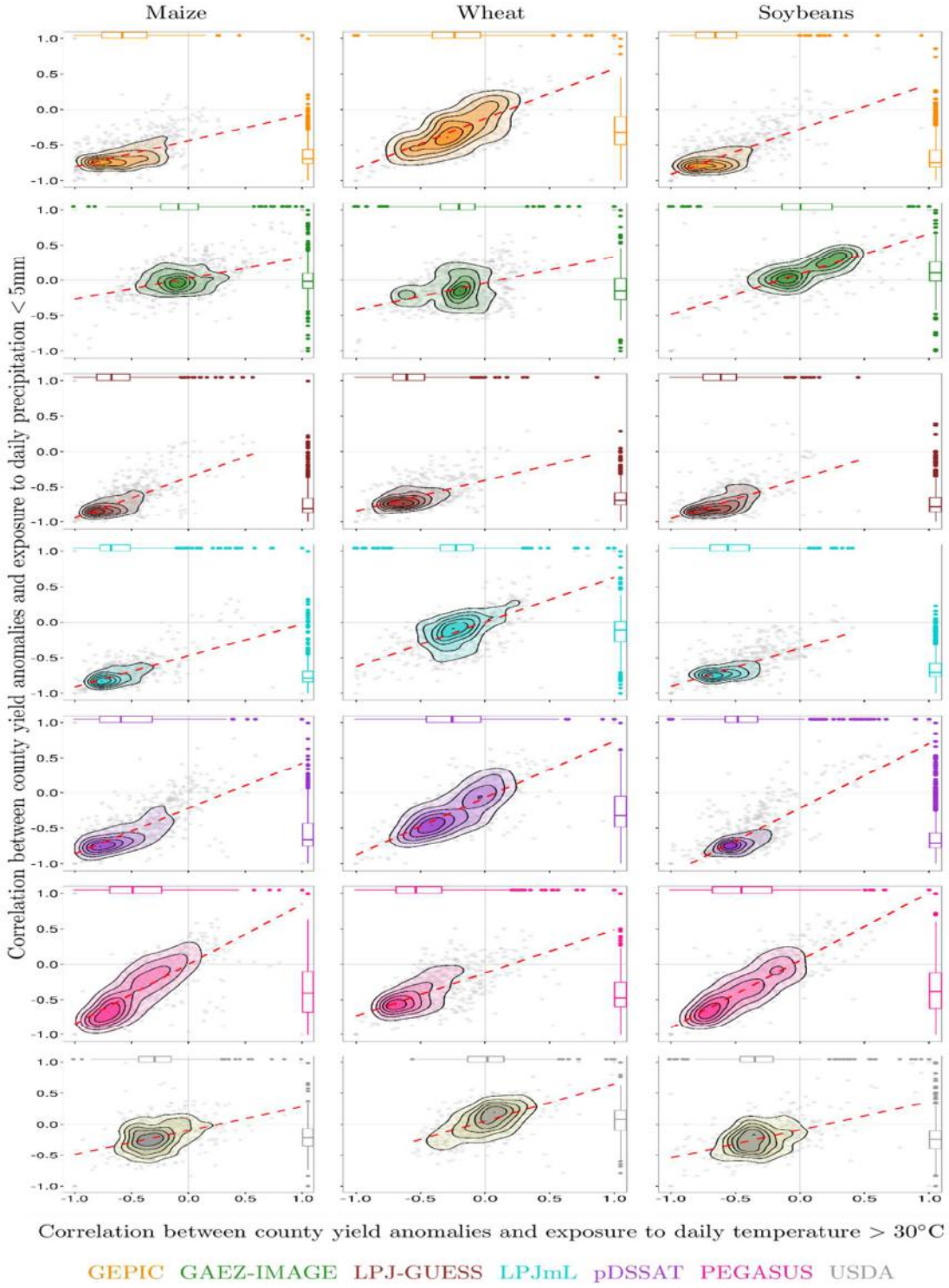


Figure 4.2. Correlations between % yield anomalies and extreme high temperature exposures ($Corr(\ ^{*}Y, \xi^{T}_{>30°\mathrm{C}})$, horizontal axis) and extreme low precipitation exposures ($Corr(\ ^{*}Y, \xi^{P}_{<5\ \mathrm{mm}})$, vertical axis) for six GGCMs and observations. Dashed red lines are the linear fit indicating the cross-county pattern of association between temperature and precipitation exposure correlations.

4.3.3 Empirical modeling of simulated and observed yield responses to weather

The foregoing comparisons leave uncontrolled a wide variety of factors that might reasonably be anticipated to affect yield. One is management practices, whose variation across sub-national locations and years is not recorded (or made readily available), either by USDA or as part of the ISIMIP-FT exercise. Another is non-extreme weather: negative yield impacts of more frequent extreme low moisture and/or high heat days could conceivably be offset by near-optimal growing conditions throughout the remainder of the growing season. Conversely, yields might well be lower in counties and years that experience fewer extreme adverse days, but more frequent non-extreme but nonetheless sub-optimal weather.

The advantage of the econometric model in eq. (4.1) is its ability to account for both sets of factors, transparently partitioning the variance in simulated and observed yields between idiosyncratic influences potentially associated with unobserved shifts in management, and the mean deterministic effects of the full range of heat and moisture conditions experienced by crops. The latter are shown in figure 4.3 as piecewise linear splines that trace out the responses of log yield to ranges of temperature and precipitation. The empirical models are precisely estimated, with the covariates explaining 75% of the cross-section/time-series yield variation on average (table 3C in Appendix C), but weather responses accounting for between 0-51% (table 4C). Estimates derived from both the GGCM and observational datasets are qualitatively consistent with empirical evidence on the critical negative effects of extreme heat (cf. Schlenker and Roberts 2009, Tack *et al* 2015). Aside from GEPIC maize and pDSSAT soybean simulations, these effects are understated by the corresponding responses to low moisture, also a feature of empirical findings. But there is considerable heterogeneity in crop models' responses to both extreme and non-extreme weather, and, compared to the observational benchmark (panels D-F and J-L) GGCMs overstate crops' sensitivity (cf. Figure 4.3).

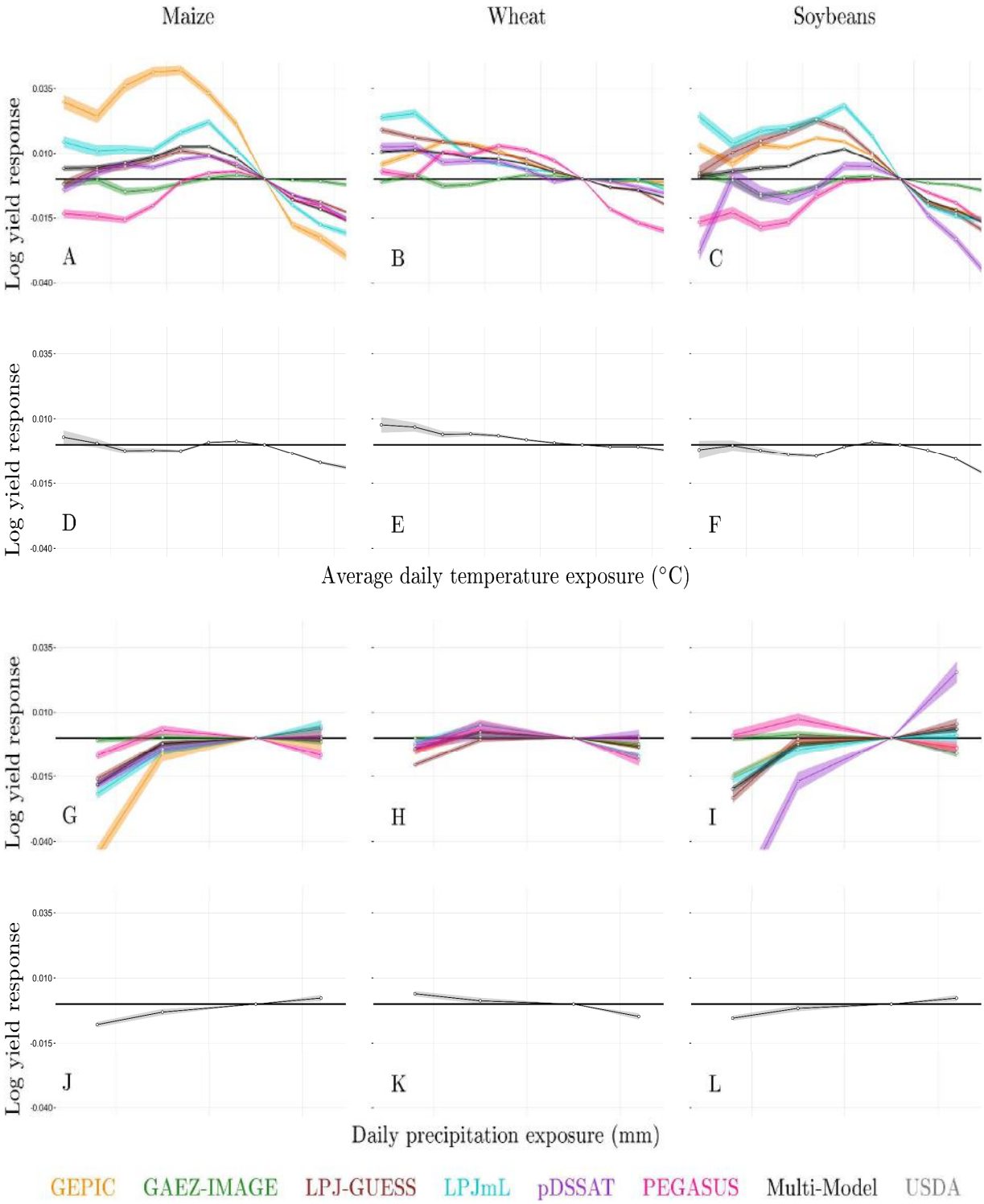


Figure 4.3. Log yield impacts of temperature and precipitation exposures for maize, wheat and soybeans, mean responses (solid lines) and confidence intervals (shaded areas). Responses are normalized relative to the number of days with temperatures 22.5-25°C and precipitation 10-15 mm, represented by the heavy horizontal axis. Standard errors are robust to heteroscedasticity, and temporal and spatial autocorrelation.

In the observational dataset, exposure to an additional day >30°C reduces annual maize and soybean yields by 1% but generates wheat yield losses one-tenth as large. The corresponding GGCM responses are more elastic, between 0.2-3% for maize, 0.5-3.6% for soybeans, and 0.1-2% for wheat. Exposure to an additional day with precipitation <5 mm reduces maize and soybean yields by about 0.25% in the observational dataset. Here too, GGCMs exhibit larger losses across all crops, between 0 and 4.5% (1.9% at the multi-model mean). Even so, no GGCM exhibits consistent positive or negative biases relative to the observational response.

4.3.4 Decomposing the divergence between GGCM- and observationally-calibrated yield responses

The foregoing divergence may result from several influences. One potential culprit is omitted variable bias, particularly the contaminating effects on $\widehat{\boldsymbol{\beta}}^{T}$ and $\widehat{\boldsymbol{\beta}}^{P}$ of management practices that are correlated with weather and unrecorded in the observational dataset, but omitted from GGCM simulations. A second is simply differences between the aggregate responses to weather shocks implied by process models' internal representation of crop growth and the responses of real-world agricultural systems. A third is differences in the exposures implied by GLDAS for the observations as opposed to HadGEM2-ES for the GGCMs. Although omitted variable bias is not something we can address, we can establish the importance of the first and second influences relative to the third by decomposing the difference between GGCM and USDA yield responses into climatic uncertainty and response uncertainty, using eq. (4.4). Figure 4.4 show the results of this calculation.

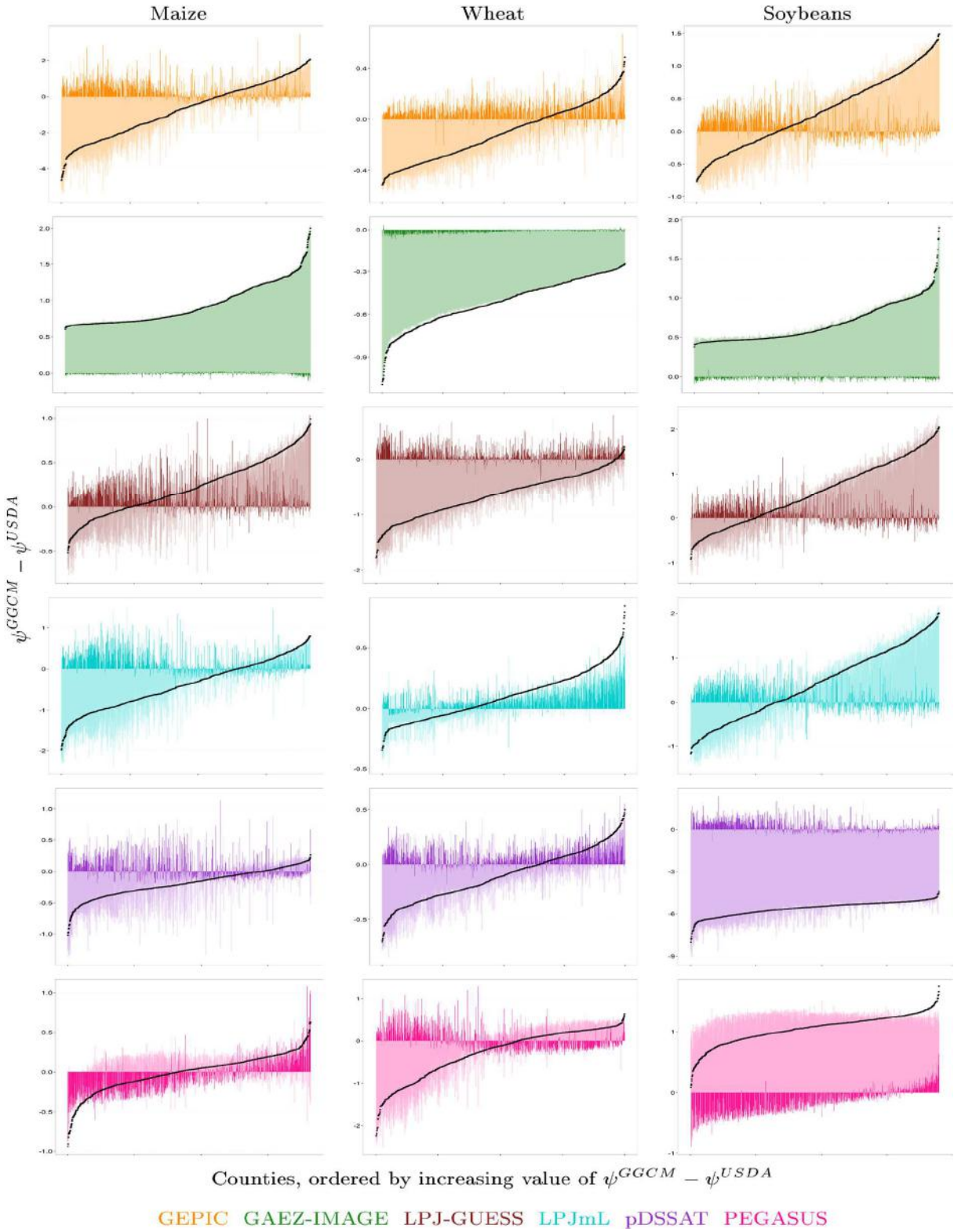


Figure 4.4. Decomposition of predicted weather component of GGCM yield - predicted weather component yield of observed yield for 950 counties showing the total difference (black dots), climate component (dark bars), and response component (light bars).

The horizontal axis rank-orders counties from the largest negative to the largest positive values of the difference between the weather-responsive portion of each GGCM's historical run and the observations, $\Delta\psi$. The magnitude of this divergence is measured on the vertical axis and indicated by black dots. Corresponding to these, for each county a light-colored bar indicates the response component ($\Delta\psi^{Response}$), while a dark-colored bar identifies the climate component ($\Delta\psi^{Climate}$). The majority of GGCM-crop combinations show patterns of total divergence that mirror the cross-county trend in $\Delta\psi^{Response}$, with $\Delta\psi^{Climate}$mostly adding noise. This suggests that the differences in the splines in Figure 4.3 are mostly attributable to GGCMs' internal responses, not differences in meteorological inputs.[44]

4.3.5 Future U.S. crop yields under climate change

Using eq. (4.5) to quantify the implications of our estimated response functions for the impacts of climate change on yields generates patterns of changes summarized in figure 4.5. Differences in GGCMs' responses in figure 4.3 translate into starkly contrasting projection of yield change. The response functions for models such as GEPIC that exhibit strong negative correlations between yields and high temperature or low precipitation predict losses of more than 75%, while those for GAEZ-IMAGE counterintuitively predicts yield gains.

We emphasize that our projection methodology is based solely on weather impacts, and does not attempt to account for the potentially beneficial offsetting effects of either the CFE, exogenous future adaptations or the endogenous adjustments (e.g. changes in cultivars and growing seasons) simulated by models like LPJ-GUESS and LPJmL. By comparison, the adverse impacts projected by our less elastic observationally calibrated model are substantially attenuated, especially for wheat that suffers negligible damages for both future epochs.

[44] The reverse is true for PEGASUS maize and soybeans simulations. In the LPJmL wheat simulation the climate component dominates for $\Delta\psi > 0$ while the response component dominates for $\Delta\psi < 0$. Weather appears to play a negligible role in GAEZ-IMAGE simulations, with $\Delta\psi$ almost entirely driven by the estimated crop response.

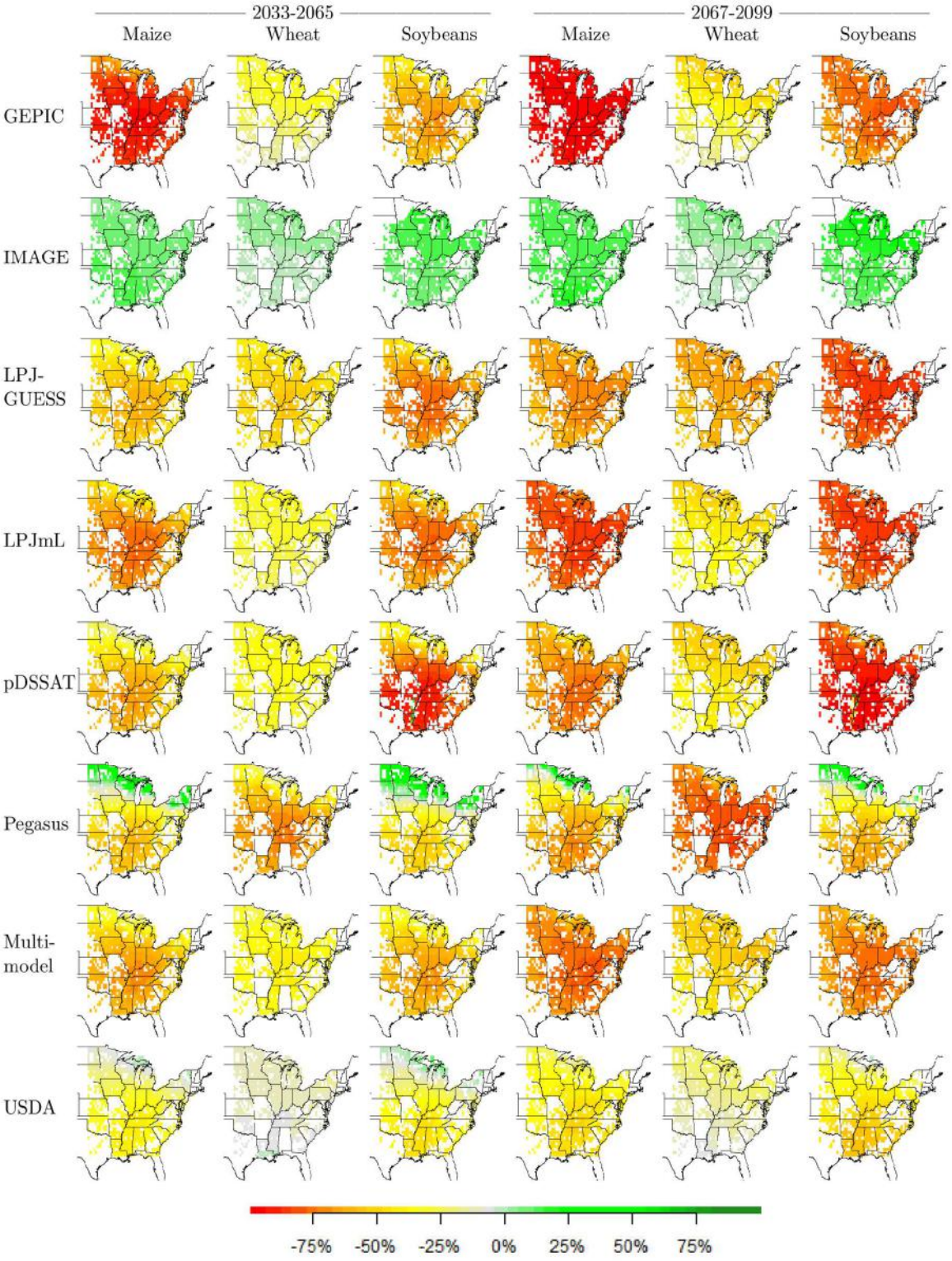


Figure 4.5. Projected % changes in crop yields for two future periods under an RCP 8.5 warming scenario simulated by HadGEM2-ES (see figure 2C in Appendix C for changes in exposure of future temperature and precipitation simulated by HadGEM2-ES under RCP 8.5, relative to the historical period).

4.3.6. What drives the divergences in GGCMs' crop yield responses to weather?

We finally examine what drives the divergence in results across GGCMs. In particular, the divergence of the estimated crop yield responses to temperature and precipitation (figures 4.2 and 4.3), not only across the six GGCMs, but also between the GGCMs and the empirical specification of USDA. As emphasized in literature (e.g. Rosenzweig *et al* 2014, Elliott *et al* 2015, Müller *et al* 2016), understanding these drivers of model responses are of paramount importance, and yet seldom systematically initiated (see Nelson *et al* 2014 and Bassu *et al* 2014).

Table 4.1 summarizes the results of the statistical meta-analysis (eqs. 4.6 – 4.11) for crop maize (See results in tables 5C for wheat and soybeans).

Table 4.1: Regression summary of meta-analysis for crop maize. The six model specifications denoted by (4.6 – 4.11) in the table correspond to the specifications represented by eqs. (4.6 – 4.11) respectively. Robust standard errors (S.E.) are reported in parenthesis.

Dependent variable: Difference in Estimated Coefficients (GGCMs – USDA) as defined in eqs.(6-11)

	(4.6)	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)
Potential Yield	-0.007	-0.012	0.007	-0.025***	-0.031***	-0.001
	(0.007)	(0.009)	(0.005)	(0.005)	(0.004)	(0.002)
Change in Cultivar	0.011	0.017*	-0.006	0.032***	0.039***	0.002
	(0.008)	(0.009)	(0.005)	(0.006)	(0.005)	(0.001)
Dyn. planting Window	-0.002	-0.003		-0.008***	-0.008***	
	(0.002)	(0.002)		(0.002)	(0.002)	
Heat stress	-0.018**	-0.025**	0.006	-0.037***	-0.044***	-0.009***
	(0.008)	(0.010)	(0.006)	(0.007)	(0.007)	(0.000)
GGCM Calibration (Site)	0.001	0.001		0.003***	0.001	
	(0.001)	(0.002)		(0.001)	(0.002)	
I(hi_t * Pot_Yield)				0.051***	0.063***	
				(0.007)	(0.006)	
I(hi_t * Cultivar)				-0.061***	-0.074***	
				(0.007)	(0.006)	
I(hi_t * Plant_window_Dyn)				0.019***	0.019***	
				(0.003)	(0.003)	
I(hi_t * H_stress)				0.049***	0.062***	
				(0.008)	(0.007)	
I(hi_t * Calib_Site)				-0.009***		
				(0.002)		
I(lo_p * Pot_Yield)				0.037***		0.013*
				(0.008)		(0.007)
I(lo_p * Cultivar)				-0.045***		-0.012*
				(0.008)		(0.006)
I(lo_p * Plant_window_Dyn)				0.013***		
				(0.003)		
I(lo_p * H_stress)				0.048***		0.023***
				(0.009)		(0.006)
I(lo_p * Calib_Site)				-0.006***		
				(0.002)		
Adjusted F Statistic and	4.289***	4.885***	1.553	42.577***	24.052***	2.325*
Degress of Freedom	(df = 4;77)	(df = 4;59)	(df = 2;17)	(df = 14;77)	(df = 8;59)	(df = 5;17)
Observations	78	60	18	78	60	18
Adjusted R^2	0.107	0.186	0.031	0.604	0.679	0.238
Residual Std. Error	0.013 (df = 73)	0.014 (df = 55)	0.006 (df = 15)	0.009 (df = 63)	0.009 (df = 51)	0.005 (df = 12)

Note: *p<0.1; **p<0.05; ***p<0.01

A few features become immediately apparent by focusing on the significant ($p < 0.05$) coefficient estimates of the broad GGCM characteristics (table 4.1). The specifications (4.8) and (4.11) are largely insignificant, given the small sample size of precipitation bins used in our regression specification (eq. 4.1). In specifications (4.7 and 4.10), the treatment of heat stress[45] parameter induces a negative and significant bias in the temperature coefficients, relative to the USDA estimated coefficients. However, this effect is completely offset in the high-temperature and low-precipitation bins (interaction terms in models 4.9 – 4.10, table 4.1).

The second important significant and prominent characteristic is cultivar adaptation, which (expectedly) has the opposite effect to heat stress. Switching over to high yield cultivars (climate adaptation) should potentially offset the negative yield losses under warming climate; but once again, this effect is offset in the extreme temperature and precipitation bins. Other broader characteristics such as dynamical changes in planting window and type of GGCM calibration, have limited significant effect on the divergence in responses of GGCMs' yields to weather variables relative to USDA empirical responses.

The results in table 4.1 are largely consistent for soybeans, but to a lesser extent for wheat (see table 5C in Appendix C). While understanding the underlying differences in treatment effects induced by the same parameters across the three crops is beyond the scope of this paper, a likely reason for the disparity could emanate from the marginal heterogeneity in the GGCMs' inter-crop simulation setups. To summarize, our novel approach in identifying the potential drivers of divergence in responses could prove even more beneficial when a broader suite of GGCMs are incorporated in inter-comparison exercises (such as in ISI-MIP2[46] and Global Gridded Crop Model Intercomparison (Elliott *et al* 2014)).

4.4 Discussion and Conclusions

This paper presents a head-to head comparison of process-based and econometric models that are commonly utilized to examine climate change impacts on crop yields. We compare GGCMs' internal representations of crop growth with the responses of yields to heat and moisture variation under the current climate as estimated by econometric models trained on observations. Our focus is on rainfed yields of maize, wheat and soybeans in the U.S. counties.

Although when comparing with observations, we can expect a given GGCM not to be able to exactly reproduce the level of yield in a particular year and county, even across all counties that produce a given

[45] Among the six GGCMs used in our study, only PEGASUS explicitly accounts for heat-stress in ISIMIP-FT simulations (see table 2A in Appendix A for broader GGCMs' characteristics)
[46] See ISI-MIP2 protocol, www.isi-mip.org

crop, under a range of climatic conditions, the GGCMs examined in this paper have difficulties at approximating the PDF of percentage yield anomalies, defined as the deviation of GGCM's annual county yields from its own 1972-2004 mean. Our analysis shows that the main driver of this difference is the diversity in estimated crop response function, whereas uncertainty in the different sources of climate inputs plays a secondary role.

The empirical estimates of crop yield responses to temperature and precipitation indicate that GGCMs generally overstate crops' sensitivity to both meteorological variables. This result could reflect unmeasured farmer adaptations that influence the response functions estimated using the econometric model but that are absent from the crop models. However, we take a step further by asking an important question as to what drives the divergence in responses across the GGCMs, and between the GGCMs and the empirical estimates. Our statistical meta-analysis is able to attribute the average impact on the entire pattern of GGCMs' response to a few key model simulation parameters. For instance, we establish that the treatment of heat stress in PEGASUS induces a negative and significant average bias in the temperature coefficients relative to the USDA coefficient estimates. This largely explains the earlier documented hypothesis of PEGASUS's generally pessimistic predictions in future crop yields (see Rosenzweig *et al* 2014, Müller *et al* 2015). Our contribution should therefore further facilitate benchmarking of results in inter-model comparison exercises (such as being addressed by Müller *et al* 2016); but more importantly, by taking a critical step further in elucidating the underlying GGCMs' characteristics responsible for the divergence in their crop yield responses.

It must be acknowledged though that the GGCMs' simulations from ISIMIP-FT exercise are at global domain and are not optimized for U.S. counties. Care should therefore be exercised in interpreting the results outside the context of this study. Moreover, it is envisaged that the next phase of model inter-comparison exercises (e.g. ISI-MIP2[47], Global Gridded Crop Model Intercomparison (Elliott *et al* 2014)) would include systematic harmonization of GGCMs in their simulation setup. A similar broad scale inter-model exercise as done by us could throw more insight on the underlying reasons for divergence in GGCMs' performance, and potentially set precedence of benchmarks in modeling exercises.

Closing remarks

The following chapter systematically summarizes the results covered in the thesis. The caveats associated with the overall study, and scope for future research are also highlighted.

[47] See ISI-MIP2 protocol, www.isi-mip.org

Chapter 5: Conclusion

Preface

In spite of rapid strides in technology in agriculture sector (e.g. Green revolution[48]), crop productivity in both developed and developing countries is largely influenced by weather variables, notably temperature and precipitation (Lobell *et al* 2011, Porter *et al* 2014). A primary reason for the persistent pressure faced by agriculture due to weather elements stem from the fact that crop yields are more vulnerable to the variability in regional climate patterns, rather than the changes in the mean regional or global climate (Hatfield *et al* 2011). To understand the implications of future changes in the mean and variability of climate on food production and economy, at both regional and global scales; the research community and policy makers employ a large array of modelling tools, such as process based or mechanistic crop models, IAMs and computable general equilibrium models (CGEs).

Empirical approaches have been increasingly applied in impact studies across various domains, such as Energy (e.g. De Cian *et al* 2013), Health (e.g. Egbendewe-Mondzozo *et al* 2011), and most notably Agriculture (e.g. Lobell and Field 2007, Schlenker and Roberts 2009 and Greenstone 2012). Its reliance on computationally less demanding resources makes it an attractive tool in inter-sectorial impact studies (Lobell and Burke 2010). In addition, its simple rapid integration with other models (such as ESM, IAM) makes it a very attractive tool to supplement complex and computationally demanding mathematical models. It is only natural that utilizing the strengths of an appropriate econometric approach (within the boundaries of its limitations) in agronomic studies, could have far reaching applications in supplementing process based crop models' simulations of future crop yields.

Recent work by Oyebamiji *et al* (2015) and Blanc and Sultan (2015) have built a generic framework for the construction of crop-model emulators, albeit with certain limitations such as the former employing only one GGCM (LPJmL), and the latter on only one crop (maize)[49]. Moreover, their work not only utilizes a large number of climate and non-climatic predictor variables, but also comes with certain constraints in the practical application that are inherently bound to the methodology used in their respective calibrations.

---

[48] Largely attributed to the eminent American scientist Norman Borlaug, "Green Revolution" refers to the rapid advancements in agricultural technology and practices that led to a rapid rise in global food production beginning early 1960s. Technological advances include but are not limited to, improved crop varieties, better fertilizer applications and management practices, genetically modified organisms, irrigation infrastructures etc.

[49] At the time of writing this thesis, Blanc (2016) expands the work of Blanc and Sultan (2015) to include wheat, soybeans and rice in building a statistical emulator. However, notable key differences between these works and the work undertaken in this thesis remain.

Putting simplicity, flexibility and above all robustness in the forefront, this thesis has investigated the suitability and application of a statistical emulator for estimating impacts of climate change on agriculture, at varying spatial and temporal scales, under different future climate scenarios (RCPs 4.5 and 8.5). Such a simple, robust and flexible reduced form statistical response function to emulate crop yields could provide useful information to policy makers on both the impacts (social and economic costs), as well as the pressures on adaptation in the coming years. Moreover, it is envisaged that such an approach would help to perform sensitivity analyses in order to further understand the compounding driving factors responsible for changes in global crop yields.

This chapter summarizes the key results of the thesis and how they address the overall objectives laid down in the introductory chapter (Chapter 1). I discuss some caveats associated with the data and methodology, providing suggestions for interpreting the results with caution. Finally, I provide suggestions for future research, with a broader goal of adding further value-added tools for policy and decision making.

5.1 Overview of results

5.1.1 Chapter 2

Together Chapters 2 and 3 broadly address the research questions 1-3 outlined in the objectives (Section 1.2) of Introduction. Chapter 2 began by introducing the framework of recent global inter-model, inter-comparison exercises focusing on the impacts of climate change on crop yields (namely AgMIP and ISI-MIP). The cumbersome model realizations implemented by such exercises under varying simulation setups provide an understanding of the model uncertainties in the future projections of changes in crop yields, as well as much needed data for calibration in empirical studies. Baring a few exceptions[50], historical observed global crop yield data are neither easily nor reliably available, specifically at a fine scale (gridded or district/county) spatial resolution over a significant temporal dimension fit for statistical analyses. Under such limitations, employing data from the output of process based GGCMs simulations is one logical alternative (Lobell and Burke 2010).

The PMA framework, on which the emulators are constructed, was discussed in detail highlighting the potential advantages as well as the pitfalls of the methodology, and the need to interpret the results with an element of caution. For instance, as Lobell and Burke (2010) emphasize, the biggest advantage in using a PMA is that the "true" response to climate change can be calculated, synonymous to laboratory

[50] A few notable exceptions include the historical observed crop yield data over India at district level , and for U.S. counties, and the recent global scale contributions of (Ray *et al* 2012, Iizumi *et al* 2014)

(modeling world) control experiments (runs). In contrast, when using actual historical observed data for calibrating a statistical model, the future responses are unknown. Naturally, using data for calibration/validation from only one source (such as a single crop model and/or crop, as done by the above mentioned earlier studies focusing on emulators), could make the results conditional to the choice of the crop/model. A key take away message of this chapter has thus been to apply the PMA to a broad array of crops and models that represent many processes influencing yields. Moreover, apart from the option to use the projections of individual emulators as an ensemble of six emulators, the development of a multi-GGCM emulator (not attempted in earlier studies) provides a tool for rapid amalgamation of heterogeneous GGCMs, thus having a potential to better quantify the uncertainty of climate impacts on crop yields.

Focusing on the key results, barring a few exceptions, the responses (coefficient estimates) of GGCMs' yields to T and P bins, exhibit similar characteristic shape (such as the mid-range T ideal for most crops' growth process and lower/higher thresholds having detrimental impact on yields), in line with earlier works of Schlenker and Roberts (2009), Lobell *et al* (2012) and Blanc and Sultan (2015). The results of the in-sample and out-of-sample validations under both RCPs 4.5 and 8.5 show the emulators have low relative bias (RB) across grid-cells with a higher share of global crop production. Given the heterogeneity across the GGCMs, their simulation setups, and most importantly dynamic adaptation in some of them (e.g. LPJ-GUESS and PEGASUS), the contrasting results across the six GGCM emulators are expected. Yet the multi-GGCM emulator shows encouraging results by replicating the broad behavior of the underlying six GGCMs on which it was trained.

Further, by incorporating a trend interaction regression specification on three of the GGCMs, I illustrate the contribution of adaptation (implicit to GGCMs), in moderating yield shocks over the two future epochs (2030~2064 and 2065~2099) under RCP 8.5. The findings are crucial to agronomic studies since the inability to correctly account for adaptation in future yield responses are often considered short-comings of statistical methods (Lobell *et al* 2011). Moreover, the outputs of model inter-comparison exercises typically record only a subset of the endogenously-varying internal processes of their constituent models. Not controlling for the potential confounding impacts of adaptation can therefore inadvertently lead to double counting of shocks in IAMs/CGEs.

The chapter also examines sensitivity of regression parameters and specifications, a key necessity in any empirical study. Additionally, by focusing on different validation periods under contrasting future climate scenarios, I provide a preliminary robustness check of the estimation methodology. The findings (that of overall better performance of emulators in the mid-century validation period, compared to the end-

century) intuitively supports the known hypothesis that statistical models tend to perform better when the variability in the mean climate is similar in the training and forecast periods (Lobell and Burke 2010), as evident from the climate projections made for the two epochs by HadGEM2-ES.

Moving on to potential caveats, notwithstanding the encouraging results from the most basic emulator of the heat and moisture effects of climate change, it is worth reiterating that further detailed analyses of model specifications could lead to refinements in the fitted model. For instance, the lack of soil-moisture data (in ISIMIP-FT) often considered a key parameter in determining water balance and water requirements during crop growth processes, would have proven beneficial in replicating the responses of GGCMs' crop yields better. Albeit another variable (Vapor Pressure Deficit, VPD) was used as a proxy (in Chapter 3), the inclusion of soil-moisture as a predictor variable could have improved the overall model fit. Moreover, omitting key control variables in empirical estimates could incorrectly attribute responses to predictor variables included in regression specification.

Another potential caveat in this study has been the omission of a different set of climate variables, as addressed by Oyebamiji *et al* (2015) and Blanc and Sultan (2015). Utilizing data from GGCMs driven by only one global climate model (GCM) (HadGEM2-ES in this study), along with the weather variables from the same GCM used for calibration and validation, could potentially restrict the span of the likely climate input space. Again, constrained by data availability, the ISIMIP-FT data used in this study has been restricted to HadGEM2-ES as the input climate model[51]. It may also be noted though that realizations of climate variables from GCMs are not necessarily a true climate input space, but more of climate model space (Oyebamiji *et al* 2015). Moreover, as later demonstrated in Chapter 4, the contribution of input climate variables from different sources have marginal effect in driving the divergence in results across the GGCMs.

The lack of detailed model documentation concerning the assumptions of adaptation and management practices used by the GGCMs, has been a major impediment[52] in capturing the adaptive capacity of GGCMs in their future period simulations. It is envisaged that ISI-MIP2 would not only have better harmonization across GGCM simulations, but also detailed documentation of the same, thus improving

[51] It must be noted that although Blanc and Sultan (2015) also utilized ISIMIP-FT data in their study, their focus on only crop maize enabled them to employ data from other GCMs. For the rest of the crops (wheat, rice and soybeans) not used in their study and which are used in this thesis, data from only GCM HadGEM2-ES is available in the simulation runs. Oyebamiji *et al* (2015) on the other hand had the luxury to recalibrate their process based model (LPJmL) using different input climate forcings. Needless to say, their empirical specifications allowed them to use the same multiple GCMs.

[52] Although Chapter 4 investigates the cofounding divergence in GGCMs' yield responses, a further detailed documentation of GGCMs' simulation parameters could facilitate a more thorough analysis.

the bias and precision of the statistical estimates. Lastly, although data from irrigation regime was available, the thesis did not attempt to focus on building an emulator suitable for the same. This partly stems from the core research questions outlined in the introductory chapter, emphasizing the need to build a tool capable of highlighting the pressures on adaptation on future crop yields. Nevertheless, utilizing data from irrigated regime is highly recommended as work for future.

5.1.2 Chapter 3

The chapter is an extension to the work done in the previous chapter (Chapter 2). In line with practices in empirical studies, a series of robustness analyses of the methodology used in fitting the regression models, as well as explicitly accounting for spatial dependence in model data were investigated. I begin by revisiting the base (regression) specification of Chapter 2, using an additional predictor variable (VPD). Although there were no noticeable improvements in the overall results over the base specification involving only T and P; VPD is an important variable for understanding soil-crop-atmosphere continuum. In fact, to broadly encompass the feedback mechanisms of T, P and VPD, as well as to eliminate any potential multicollinearity on including additional predictor variables, exploring a composite variable - such as Standardized Precipitation-Evapotranspiration Index ($SPEI$)[53] (Vicente-Serrano *et al* 2010)- as a standalone predictor variable is highly recommended for future work. Such an approach could not only reap beneficial results (better model fit and agreement with the GGCMs' future crop yields predictions), but also obviate the need to depend on variables that are not reported by the outputs of GCMs/GGCMs (such as soil moisture, evapotranspiration etc. in ISIMIP-FT).

The AEZ specification that explored the potential geographic heterogeneity in the slopes (coefficients), did not improve the results over the base specification, in line with Blanc and Sultan (2015). Although a further detailed analyses to understand the potential reasons were not attempted, part of this outcome could be attributed to the data used in this study. Recalling that the data for calibration and validation comes from GGCMs of ISIMIP-FT wherein the modelling groups were permitted to use their best/existing setups for simulation runs; it is conceivable that the crop modelers themselves did not have the responses of crop yields to weather and soil parameters stratified by geographic sub-regions in their control runs. Another potential reason for the poor results emanating from this specification stems from the reduced sample size (observations) of the six AEZ groups. This could have hampered both the precision and the bias of the estimated coefficients.

[53] $SPEI$ is a relatively new drought index, based on the earlier Standardized Precipitation Index (SPI). The index accounts for the effect of T on drought development through a simplified water balance computation.

The third and final regression model (Spatial lag of X, SLX) emanates from the family of spatial panel models (SPM). Apart from a few earlier studies (e.g. Cai *et al* 2012) that have applied spatial econometric methods in agriculture on a much smaller regional scale, the application of a SPM on a fine scale global gridded agricultural study is the first of its kind in this thesis. The potential pitfalls of ignoring spatial dependences in dependent and/or independent variables were highlighted in this chapter, and the results of such non-spatial techniques even though accounting for spatial correlation in the error terms were also recommended to be used with caution. Although the SLX specification (and in general, all of the SPMs) requires the construction of a spatial weight matrix (W), the improvement in overall results compared to the non-spatial specification clearly suggests the added exercise is worth the effort. It is also important to highlight potential scope of improvements when applying SPMs. The SPMs[54] used in this study are not only static, but also not geographically weighted (i.e. the coefficient estimates do not vary by grid-cells; unlike for e.g. Cai *et al* 2012). Using dynamic and/or geographically weighted spatial models becomes challenging to interpret and apply at a global gridded scale; and is left as scope for future work.

5.1.3 Chapter 4

Shifting focus from global to regional domain, as well as initiating an inter-comparison of results between statistical models trained on GGCMs' and historical observed data, Chapter 4 laid an important foundation on contributing work to model inter-comparison exercises. The work in this Chapter addresses the remaining questions (4 and 5) raised in the objectives of the introductory chapter.

For this, I focus on U.S. counties, utilizing observed (USDA) and modeled (ISIMIP-FT GGCMs) data. I undertake a head-to head comparison of process-based and econometric models that are commonly utilized to examine climate change impacts on crop yields. My results reveal that across all U.S. counties that produce a given crop under a range of climatic conditions, the GGCMs examined failed to replicate the PDF of % yield anomalies, defined as the deviation of GGCMs' annual county yields from its own 1972-2004 mean. Whereas uncertainty in the different sources of climate inputs does play a marginal role, the key driver of the heterogeneous results across GGCMs is the diversity in estimated crop response function. Perhaps the biggest contribution of the chapter to inter-model comparison exercises is the outcome of the meta-parameter analyses. The results show two key drivers of divergence in GGCMs' yield responses, namely heat-stress and cultivar adaptation. As with every empirical study, the interpretation of results outside the focus of study area and climate scenario need not be applicable. The GGCMs from ISIMIP-FT are certainly not optimized for U.S. crop growing regions. The subsequent

[54] Apart from SLX, other SPMs investigated include the Spatial Error Model (SEM) and the Spatial Durbin Error Model (SDEM). Results were not reported in Chapter 3 due to close agreement with those of SLX

implications of the projected changes in GGCMs' yields under future climate scenarios should therefore be taken with an element of caution.

5.2 Concluding discussion

In comparison to the current literature on agro-econometric and climate impacts, the main innovations of the thesis lie in the following five key elements:

(i) The use of very large and heterogeneous multi-GGCM panel data as a calibration dataset for empirical model.

(ii) The use of a parsimonious and flexible econometric specification.

(iii) The tenacity to account for additional covariates (such as VPD) not previously applied in empirical estimates, and more importantly, to explicitly account for spatial dependence and correlation in model specifications.

(iv) The use of both observed and modeled data for U.S. counties, to do a head-to-head comparison of processed based vs. econometric models.

(v) Finally, by attributing the average impact on the entire pattern of GGCMs' response to a few key model simulation parameters, the thesis provides a first insight as to why the GGCMs vary in projections of future yield changes.

The construction of a multi-crop, multi-GGCM emulator required detailed understanding of crop growth processes, key parameters and variables; and a detailed investigation of available sources of agriculture and climatological datasets. The exercise at a fine scale global gridded resolution comes with its own confounding dynamic responses of crop yields to climate. Constructed independently on six different GGCMs as well as on a combined multi-model panel of six GGCMs; my simple, flexible and robust emulator can have wide-ranging applications in studies assessing impacts of climate change on yields of four important crops. It is envisaged that the work done in this thesis would lay a further empirical roadmap, towards building a tool capable of replicating yield responses for wider variety of crops and heterogeneous crop models, predominantly with regional impetus.

Appendix A: Supplementary material for Chapter 2

1. Global Gridded Crop Models (GGCMs) used in this study along with the contact details of the modelling groups

(i) Geographic Information System (GIS)-based Environmental Policy Integrated Climate (GEPIC) (Liu *et al* 2007)

(ii) Global Agro-Ecological Zone model in the Integrated Model to Assess the Global Environment (GAEZ-IMAGE) (Van Vuuren *et al* 2006)

(iii) Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) (Bondeau *et al* 2007)

(iv) Lund Potsdam-Jena managed Land (LPJmL) (Bondeau *et al* 2007, Sitch *et al* 2003)

(v) parallel Decision Support System for Agro-technology Transfer (pDSSAT) (Elliott *et al* 2013, Jones *et al* 2003)

(vi) Predicting Ecosystem Goods And Services Using Scenarios (PEGASUS) (Deryng *et al* 2011)

Table 1A. GGCMs used in this study, with the home institution and contact details.

Model	Institution	Contact Person/Web address
GEPIC	EAWAG (Switzerland)	Christian Folberth/Hong Yang christian.folberth@eawag.ch hong.yang@eawag.ch
GAEZ-IMAGE	Netherland Environmental Assessment Agency, PBL (Netherland)	Elke Stehfest/Kathleen Neumann elke.stehfest@pbl.nl kathleen.neumann@pbl.nl
LPJ-GUESS	Lund University (Sweden), IMK-IFU, Karlsruhe Institute of Technology (Germany)	Stefan Olin/Thomas Pugh stefan.olin@nateko.lu.se thomas.pugh@imk.fzk.de
LPJmL	PIK (Germany)	Christoph Muller christoph.mueller@pik-potsdam.de
pDSSAT	University of Chicago (USA)	Joshua Elliott, jelliott@ci.uchicago.edu
PEGASUS	Tyndall Centre, University of East Anglia (UK)	Delphine Deryng d.deryng@uea.ac.uk

Table 2A: Similarities and differences in ISIMIP-FT GGCMs used in this study, adapted from (Rosenzweig *et al* 2014, Elliott *et al* 2014, Nelson *et al* 2014b, Müller *et al* 2015)

Parameters	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS
Model Type	Site-based	AEZ-model	Agro-ecosystem	Agro-ecosystem	Site-based	Agro-ecosystem
Crop Yield	Actual	Potential	Potential	Actual	Actual	Actual
Crop Cultivars (Adaptation 1)	Yes	Yes	Yes	No	No[55]	Yes
Planting window (Adaptation 2)	Dynamic[56] (climate adaptation)	Implicit Planting dates (climate adaptation)	Fixed[57]	Fixed planting date	Fixed (by taking the historical avg, for all years in future)	Dynamic (climate adaptation)
Nitrogen (N) fertilization	Yes	No	No	No	Yes	Yes
Type of stresses	Water, Temperature(T), Heat, Oxygen(O_2), N[58], Phosphorous, Bulk Density, Aluminum	Water, T, Bulk Density	Water, T	Water, T	Water, T, Heat, O_2, N	Water, T, Heat, N, Phosphorous, Potassium
Light Utilization (Photosynthesis)[59]	RUE	RUE	Leaf	Leaf	RUE (Leaf for Soybeans)	RUE

[55] For pDSSAT, cultivar choice, fertilizer application etc. are fixed by the historical average of all future years
[56] Dynamic: Automatic adjustments of planting and harvesting dates due to annual weather conditions; an internal model process.
[57] Fixed: planting windows are determined using historical values based on literature. LPJ-GUESS allows planting dates adaptation within +/-15 days of calculated optimum values, but planting window is fixed.
[58] GGCMs without N stress (GAEZ-IMAGE, LPJ-GUESS, LPJmL) tend to be more optimistic in yield response, *ceteris paribus*.
[59] RUE: Radiation Used Efficiency. Leaf: Leaf-level photosynethesis. In general, Leaf-level represents better water efficiency and photosynthesis processes. Such GGCMs would show lower yield losses (especially when accounting for CFE), *ceteris paribus*.

Model Calibration and Type_*SpatialResolution*	Yes Site-specific_*National*	No	No	Yes Global_*National*	Yes Site-specific_*Field-scale*	Yes Global_*Gridcell*
Method used in Evapotranspiration (ET) calculation	Penman-Monteith	Priestly-Taylor	Priestly-Taylor	Priestly-Taylor	Priestly-Taylor	Priestly-Taylor
Other Remarks	GEPIC accounts for soil fertility erosion, which requires the simulations to be run independently for each decade, in order to equilibrate soil processes	Crop yields are calculated for an interval of 5 years, starting from 1970. For the years in between, yields are linearly interpolated.	In addition to allowing change in cultivars, yields are unlimited by nutrient or management constraints. This is another form of adaptation i.e. through increased fertilizer supply	Like LPJ-GUESS, LPJmL also assumes adaptation to changed climatic conditions (e.g. T, CO_2 etc), through increased ferlilizer supply. Moreover, LPJmL decides internally whether to grow winter or spring wheat (Müller *et al* 2015)	Can be considered as the only GGCM that truly does not account for changes in adaptation and management practices	PEGASUS explicitly accounts for heat stress and thus typically projects more significant reductions in agricultural productivity than the rest of the GGCMs (despite allowing for adaptation in sowing dates and varieties. It is typically the most pessimistic of all GGCMs) (Müller *et al* 2015)

2. Steps involved in data cleaning and number of observations by GGCMs

One notable difference in the ISIMIP-FT data vis-à-vis data from other agronomic studies is the spatial coverage of crops. In addition to present day croplands, crop modellers were allowed to simulate yields across present day non-growing regions[60] as well. In order to use a reliable set of times series for our cross-section analyses, we follow a three stage procedure to identify and filter out noisy data.

Step 1: For each rainfed crop used in our study, we mask the grid-cells of our original full global panel to retain only the global rainfed cultivated areas using the MIRCA2000 (Portmann *et al* 2010) global gridded database on cropland for different crops.

Step 2: Next, to identify cells with anomalously high or low crop yield values or yield values showing questionable reliability, we filter grid-cells where annual yield remained the same for three or more consecutive years in the period 1972-2099.

Step 3: Finally, we retain grid-cells where $yield\ (t/ha) > 0$ for all 128 years (1972-2099)[61]. By doing so, we ensure the panel is balanced (i.e. we have same number of observations across all grid-cells).

Although there is no strict requirement for a balanced panel in our regression specification, it not only facilitates easier computation of means of all variables at each grid-cell, but also enables consistent evaluation of mean responses of yield to changes in climate variables.

The abovementioned data cleaning steps have been applied to all crops as well as to the additional panel used in out-of-sample calibration (i.e. 1972-2089). The total number of years in this case would be 118.

Table 2 summarizes the sample panel used in the regression analysis resulting from the three-stage data cleaning procedure.

[60] Crop is grown across all grid-cells (land area) where soil and weather conditions permit it.

[61] There are a few exceptions to this. GAEZ-IMAGE does not have data in RCP 4.5 for any of the crops. The multi-GGCM panel in RCP 4.5 would therefore consist of five GGCMs. For RCP 8.5, GAEZ-IMAGE does not report data for years 2001-2004. Thus the panel for 1972-2099 would correspond to 124 years. To maintain consistency, the multi-GGCM panel also drops data for 2001-2004 from the remaining five GGCMs. Thus the multi-GGCM in RCP 8.5 would also span 124 years.

Table 3A. Summary information of panel used in calibration, by Crop~GGCM~RCP. Observations used in regression with number of grid-cells in square braces.

(i) RCP 4.5 (1972~2099)

Crop	GEPIC	GAEZ-IMAGE[a]	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS[b]	Multi-GGCM
Maize	2,715,136 [21,212]	-	3,346,688 [26,146]	2,258,816 [17,647]	1,713,792 [13,389]	2,825,728 [22,076]	12,860,160 [30,214]
Rice	768,896 [6,007]	-	603,136 [4,712]	749,312 [5,854]	292,992 [2,289]	-	2,414,336 [18,862]
Wheat	1,519,872 [11,874]	-	1,746,688 [13,646]	1,735,296 [13,557]	875,648 [6,841]	1,473,024 [11,508]	7,350,528 [57,426]
Soybeans	879,360 [6,870]	-	1,058,560 [8,270]	909,696 [7,107]	510,976 [3,992]	672,640 [5,255]	4,031,232 [31,494]

(ii) RCP 8.5 (1972~2099)

Crop	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS	Multi-GGCM
Maize	2,680,448 [20,941]	2,912,512 [23,488]	3,341,696 [26,107]	2,165,632 [16,919]	1,570,432 [12,269]	2,840,320 [22,190]	15,117,336 [30,619]
Rice	765,184 [5,978]	481,368 [3,882]	566,784 [4,428]	724,480 [5,660]	294,400 [2,300]	-	2,758,752 [22,248]
Wheat	1,510,400 [11,800]	1,382,228 [11,147]	1,749,120 [13,665]	1,714,304 [13,393]	866,816 [6,772]	1,363,328 [10,651]	8,361,072 [67,428]
Soybeans	879,616 [6,872]	7,97,072 [6,428]	1,041,152 [8,134]	893,568 [6,981]	435,328 [3,401]	643,840 [5,030]	4,568,904 [36,846]

[a] No data for RCP 4.5. No data for years 2001-2004 in RCP 8.5.
[b] Rice is not simulated by PEGASUS

3. Splines of T and P coefficients for rice, soybeans and wheat, with 95% CI bands

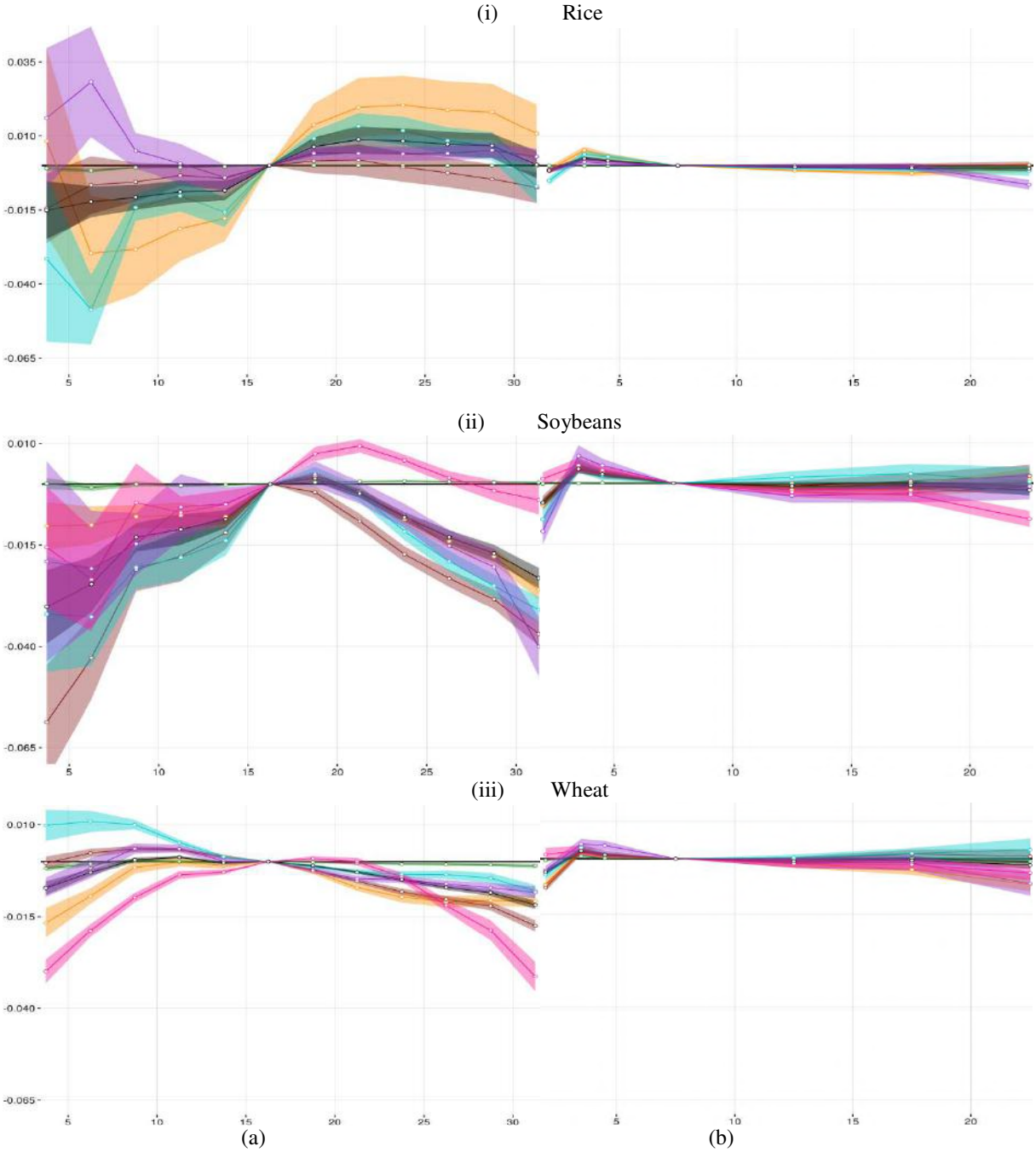
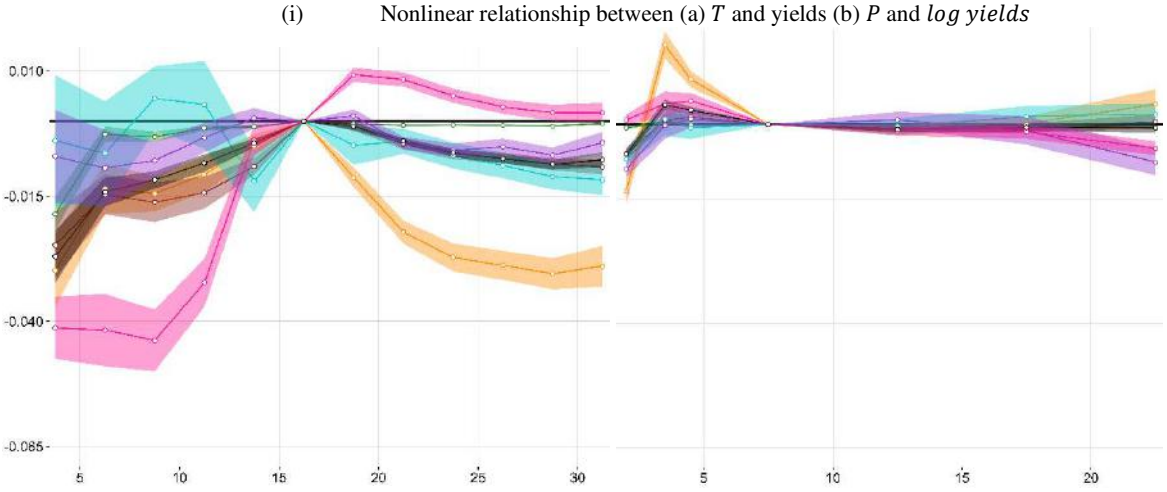


Figure 1A: Response of (i) Rice (ii) Soybeans (iii) Wheat $log\ yields$ to (a) T and (b) P bins, with standard errors (SEs) robust to heteroscedasticity and autocorrelation. Graphs display changes in yield (%) for exposure of one day to a particular $T(P)$ bin interval, relative to bins $T = 15{\sim}17.5\,°C$ $(P = 5{\sim}10\,mm/d)$. The 95% confidence band (CI) is adjusted for spatial correlation. The horizontal black line corresponds to x-axis=0 reference. CI intersecting the horizontal 0 reference line would imply that the corresponding coefficient is insignificant (i.e. $p > 0.05$). Rice is not simulated by PEGASUS, hence the multi-GGCM was run as a merged panel of five GGCMs.

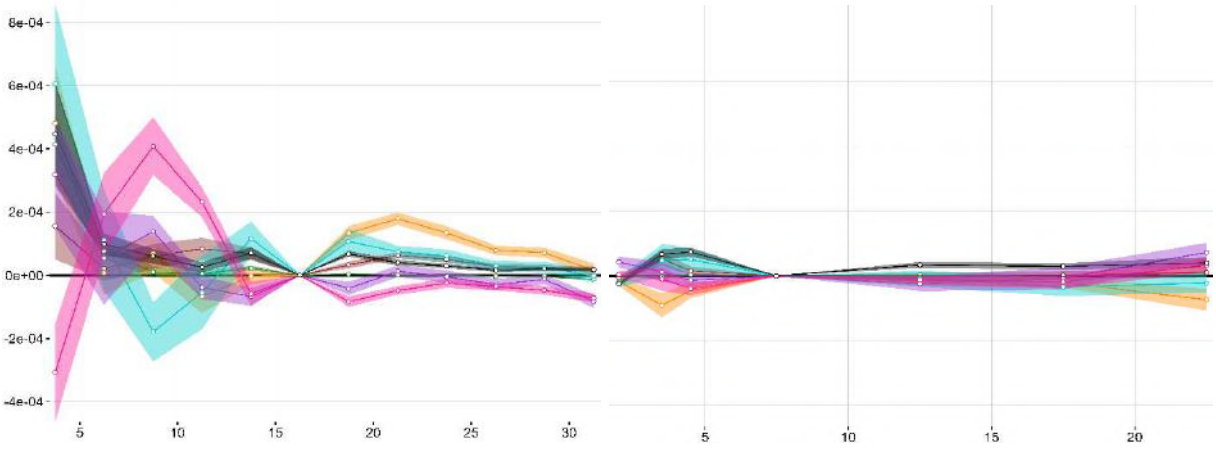
4. The perplexing nature of adaptation in GAEZ-IMAGE, pDSSAT and PEGASUS

Returning to the discussion in Section 2.3.1 of Chapter 2 for crop maize, we repeat the regressions on GAEZ-IMAGE, pDSSAT and PEGASUS with the inclusion of the interaction terms. The regression thus takes the base specification form of equation 2.2 (main text). In comparison to the responses shown in figure 2A of main text, we observe contrasts in the behaviour of responses to T for pDSSAT and PEGASUS (figure 2A here).

(i) Nonlinear relationship between (a) T and yields (b) P and $log\ yields$



(ii) Nonlinear relationship between (a) 'T' interaction term and $log\ yields$ (b) 'P' interaction term and $log\ yields$



(a) (b)

GEPIC **GAEZ-IMAGE** **LPJ-GUESS** **LPJmL** **pDSSAT** **PEGASUS** **Multi-GGCM**

Figure 2A: Response of Maize $log\ yields$ to (a) T and (b) P bins. Coefficient estimates are for (i) T and P (ii) interaction of T and P bins with time trend using the base specification (equation 2.2 in main text) for all six GGCMs. SEs are robust to heteroscedasticity and autocorrelation. Graphs display changes in yield (%) for exposure of one day to a particular $T(P)$ bin interval, relative to bins $T = 15{\sim}17.5\,°\,C$ $(P = 5{\sim}10\,mm/d)$. The 95% confidence band (CI) is adjusted for spatial correlation. The horizontal black line corresponds to $x - axis = 0$ reference. CI intersecting the horizontal 0 reference line would imply that the corresponding coefficient is insignificant (i.e. $p > 0.05$)

As evident from figure 2A, GAEZ-IMAGE (dark green) has an overall muted response of T and P, thus not revealing noticeable difference in responses vis-à-vis the specification used in main text (equation 2.3). However, the response of T for pDSSAT (purple) suggest that an additional day in the $[> 30℃]$ bin, is not more detrimental to yields than what is observed in figure 2A of main text. On the contrary, for PEGASUS (pink), an implausible positive response to extreme heat is observed suggesting that a higher than threshold temperature (29 ℃ for maize) is marginally beneficial for crop growth process. It is also interesting to note the negative response of interaction terms to extreme heat (figure 2B) suggesting that any adaptation in both pDSSAT and PEGASUS would have further marginal negative impacts on maize yields. Although lack of thorough documentation of the GGCM simulation runs make it challenging to provide concrete reasons behind this perplexing nature of adaptation (especially in PEGASUS), we discuss a few possible reasons.

pDSSAT does not account for adaptation in the future simulation runs of ISIMIP-FT. Suffice to say, including interaction term results in a model misspecification. PEGASUS on the other hand is the only GGCM used in our study that which explicitly accounts for heat stress and typically projects more significant reductions in agricultural productivity than the rest of the model ensemble (Rosenzweig *et al* 2014). Therefore, the higher mean growing season T (or increase in the frequency of future growing season days falling in $> 30 \circ C$ bin, figure 5) in RCP 8.5 would support our earlier findings (figure 2B) and explain why the interaction term inadvertently captures a negative trend and attributes it (at least partially) to detrimental impacts of adaptation.

The omission of interaction terms in GAEZ-IMAGE, pDSSAT and PEGASUS in further motivated by our sensitivity checks. Figure 3A shows the empirical cumulative distribution function (ECDF) for the emulators, ‘with-adaptation’ and ‘without-adaptation’ as defined in the main text.

(i) 2030-2064 (ii) 2065-2099

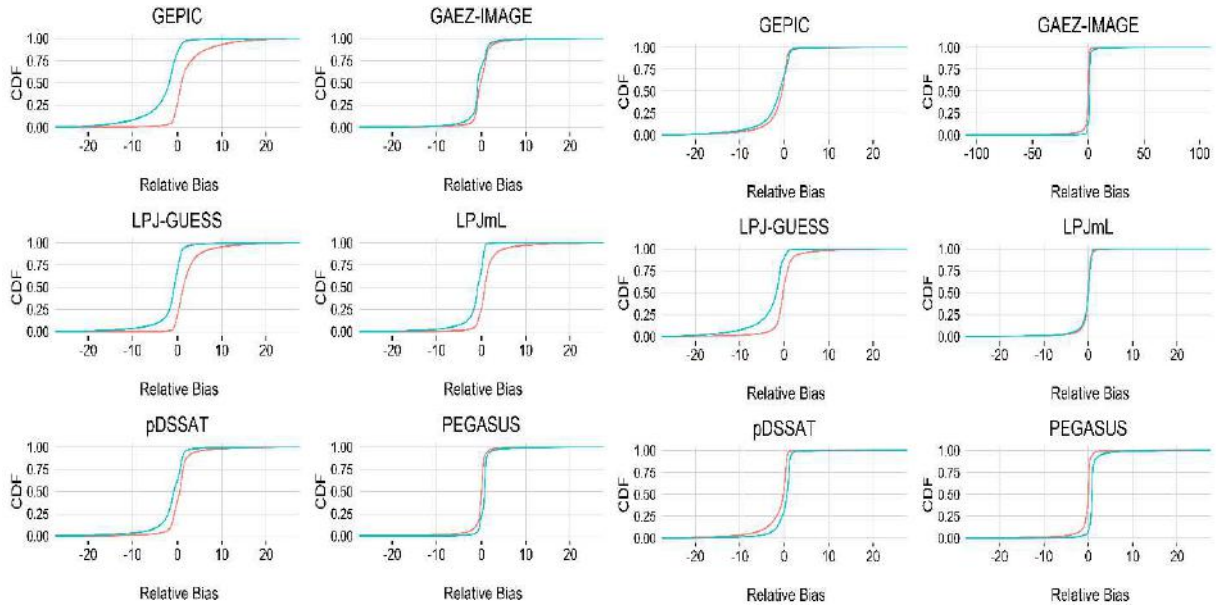


Figure 3A: ECDF of Relative Bias (%) in with-adaption (red) and without-adaptation (cyan) specifications of six maize emulators in RCP 8.5 scenario

The ECDFs for GAEZ-IMAGE, pDSSAT and PEGASUS show that the with-adaptation model doesn't appear to make a huge difference to the RB. We therefore refrain from adding any complexity to its's simple specification.

However, for the remaining three emulators (GEPIC, LPJ-GUESS, LPJmL), the reduction in relative bias (RB) is notable. For these emulators, the RB of 'without-adaptation' is overwhelmingly negative, indicating that by omitting the interaction terms from our base specification, we systematically under-predict the relative changes in yields. Instead, by using the 'with-adaptation' model we trade off reduction in negative bias for the introduction of slight positive bias, especially at mid-century.

5. Choice of explanatory variables and non-parametric binning approach used in regression

In line with agronomic literature (such as Schlenker and Roberts 2009, Lobell and Burke 2010), T and P play the most crucial role in various stages of crop growth. With our primary goal to construct a parsimonious emulator that can be rapidly linked to future climate ensembles for climate risk analysis, T and P as non-parametric bins not only captures the envelope of underlying response, but also reduces the reliance on other complex predictors and/or their forms as used for example in some recent studies (Oyebamiji *et al* 2015, Blanc and Sultan 2015).

Although a specification involving T and P have been used in earlier empirical agronomic studies in various linear and non-linear transformations (see Schlenker and Roberts 2009, Lobell and Burke 2010, Lobell *et al* 2011, Wolfram and Lobell 2010), at best we are not aware of any work incorporating the growing season counts of daily T and P measures as bins. We now summarize the key advantages and trade-offs of the binning approach implemented in our methodology.

The biggest advantage of the non-parametric binning approach comes from laying less restrictions in the functional form of the predictor variables (here T and P) and thus able to capture the piece-wise linear responses of crop yields to T and P better. In contrast, a conventional approach that of using mean growing season/months T or its' quadratic form (T_2); as well as growing degree days (GDD), put a higher functional form restricts. Moreover, the GDD thresholds developed and used in earlier studies (notably Schlenker and Roberts 2009, and Schlenker and Lobell 2010) were developed within the framework of agriculture in the United States (U.S.). Responses of crop yields to T can vary geographically e.g. maize cultivated in a tropical weather could have different thresholds to extreme temperatures compared to those cultivated in Europe. And since we examine four crops whose responses (especially to threshold temperatures) would be different, a GDD approach would involve defining varying thresholds.

Having said that, the flexibility of our methodology do come with certain trade-offs. Among these, the most notable is the assumption of additive separability where in irrespective of the contrast in the mean growing season temperatures of any two years, the marginal effect of a day in a particular bin will be identical. The other restriction imposed in our framework is that the effects would remain constant within each bin of T and P. So to say, the marginal impact of an additional day in the lower T bin (relative to the omitted bin) would be the same as the impact of an additional day in the upper T bin

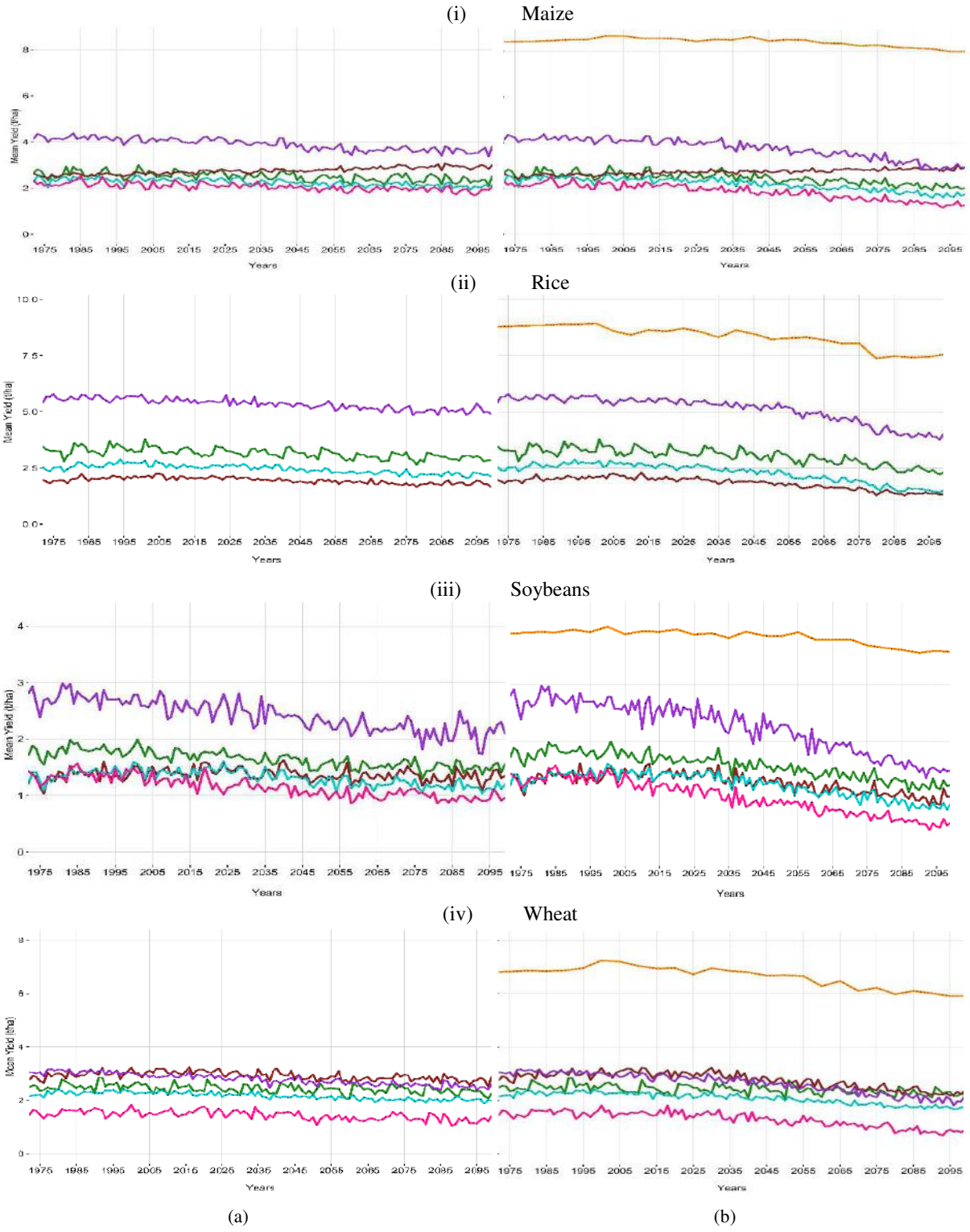


Figure 4A: GGCMs' yields (t/ha) 1972-2099 averaged over grid-cells used in analyses in (a) RCP 4.5. (b) RCP 8.5 for rainfed (i) Maize (ii) Rice (iii) Soybeans and (iv) Wheat. Rice is not simulated by PEGASUS. There is no data for GAEZ-IMAGE in RCP 4.5

6. Contrasting variability in mean climate for the two future periods in RCP 8.5 relative to baseline historical period (1972~2004) is illustrated in figure 5A. The changes in the frequency of extreme T and P bins in 2065~2099 shows a larger variation compared to the same in 2030~2064.

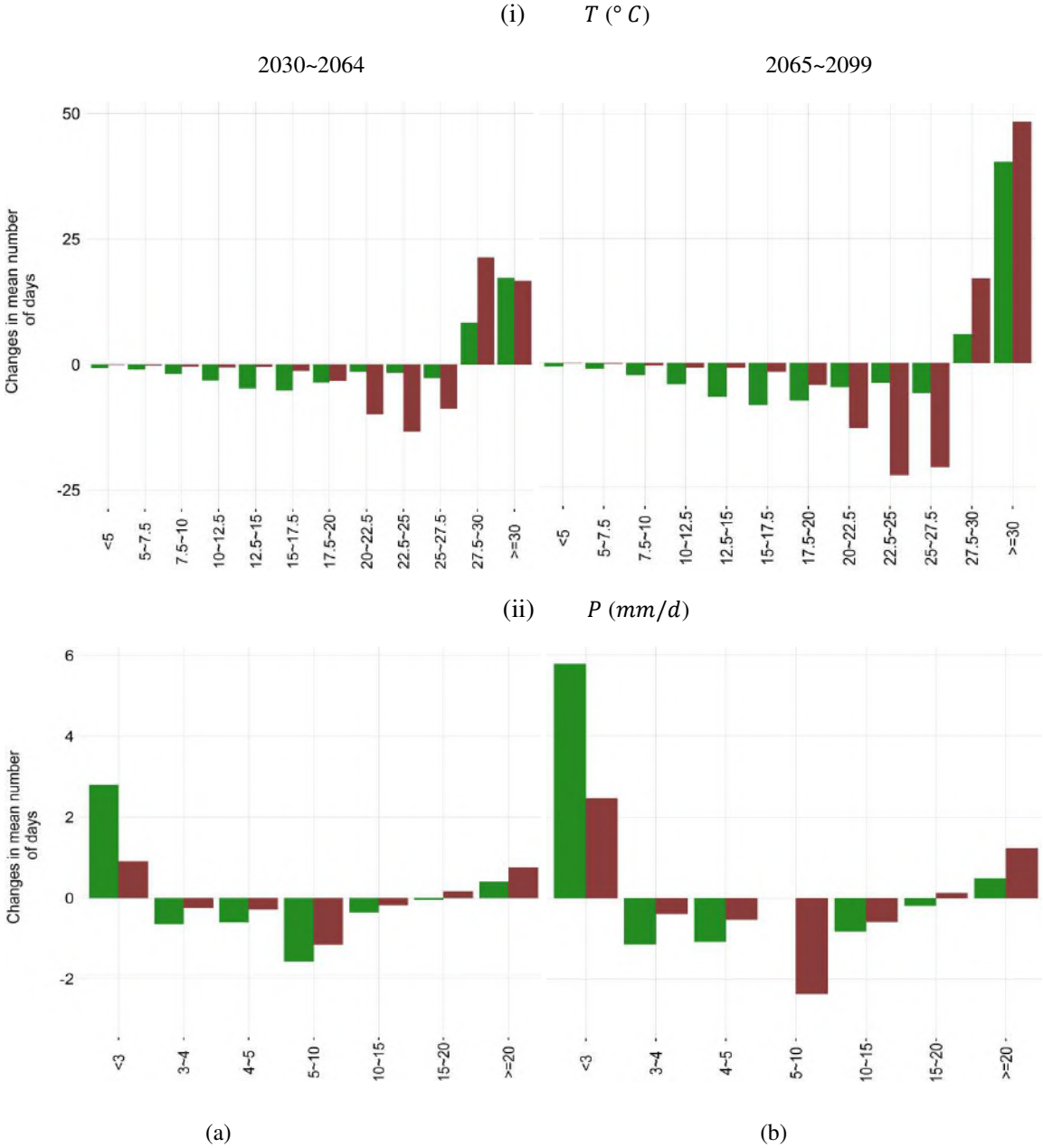


Figure 5A: Change in distribution of (T ° C) and (P mm/d) bins in mean periods (a) 2030~2064 and (b) 2065~2099 in RCP 8.5, relative to the baseline historical period (1972~2004). Each bin is indicated by the upper limit, e.g. $4-5$ mm in P corresponds to mean P (4,5]. The lowest and the highest bins in both variables do not have bounds. The bins are averaged over summer months in each hemisphere (**Northern**, **Southern**), across crop maize growing grid-cells. Thus the changes in mean number of days imply as changes in the summer months of each hemisphere.

Table 4A. Share of crop production (%) with the corresponding RB intervals, for periods 2030~2064, 2065~2099 shown in parentheses, and 2090~2099[a] shown in square braces, in (i) RCP 4.5 and (ii) RCP 8.5

Rice

(i) RCP 4.5

RB	GEPIC (%)	GAEZ-IMAGE[b] (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS[b] (%)	Multi-GGCM (%)
> 10	0.07 (0.12) [0.09]	-	0.38 (0.71) [1.30]	0.42 (0.09) [0.14]	0.14 (0.06) [0.21]	-	0.31 (0.66) [0.40]
5 ~ 10	0.23 (0.08) [0.07]	-	0.36 (0.78) [1.66]	0.29 (0.07) [0.12]	0.13 (0.08) [0.09]	-	0.52 (0.51) [0.33]
0 ~ 5	53.58 (61.90) [56.66]	-	49.03 (49.56) [51.02]	48.78 (32.20) [48.84]	45.95 (45.44) [46.04]	-	41.51 (25.30) [34.77]
0 ~ -5	39.59 (33.88) [37.23]	-	45.20 (45.54) [42.82]	43.33 (67.26) [50.08]	46.43 (41.98) [42.25]	-	47.77 (62.20) [54.80]
-5 ~ -10	3.26 (1.82) [2.55]	-	2.61 (1.82) [1.88]	4.17 (0.30) [0.57]	2.77 (4.86) [6.52]	-	4.69 (5.80) [4.17]
< -10	3.27 (2.20) [3.40]	-	2.43 (1.58) [1.31]	3.01 (0.09) [0.28]	4.57 (7.57) [4.89]	-	5.19 (5.53) [5.54]

(ii) RCP 8.5

RB	GEPIC (%)	GAEZ-IMAGE (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	4.20 (0.03) [0.02]	1.51 (0.15) [0.11]	3.03 (0.05) [0.29]	3.31 (0.11) [0.08]	2.05 (0.04) [0.03]	-	3.48 (0.04) [0.03]
5 ~ 10	2.82 (0.01) [0.03]	1.58 (0.14) [0.03]	2.57 (0.18) [0.13]	5.60 (0.10) [0.22]	2.27 (0.06) [0.06]	-	3.86 (0.06) [0.06]
0 ~ 5	37.19 (63.38) [62.67]	22.81 (49.47) [58.58]	47.91 (41.73) [25.26]	45.59 (32.45) [31.23]	23.89 (14.52) [15.87]	-	30.13 (14.52) [15.87]
0 ~ -5	52.00 (32.42) [33.74]	57.52 (43.78) [36.35]	42.03 (53.95) [66.61]	29.90 (67.23) [68.02]	60.71 (80.51) [79.57]	-	49.81 (80.51) [79.57]
-5 ~ -10	2.24 (1.75) [1.85]	8.40 (3.15) [1.78]	2.50 (1.94) [4.21]	6.90 (0.08) [0.36]	6.61 (2.70) [2.01]	-	5.71 (2.70) [2.01]
< -10	1.56 (2.40) [1.70]	8.17 (3.31) [3.16]	1.95 (3.06) [3.56]	8.69 (0.03) [0.09]	4.46 (2.16) [2.25]	-	7.00 (2.16) [2.46]

[a] Regressions for out-of-sample validation were run on a panel spanning 1972~2089, in contrast to the in-sample which are on 1972~2099

[b] Data not available

Wheat

(i) RCP 4.5

RB	GEPIC (%)	GAEZ-IMAGE[b] (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	0.54 (0.14) [1.04]	-	0.67 (0.66) [3.36]	0.03 (0.00) [0.53]	0.00 (0.00) [0.00]	2.35 (0.20) [0.21]	0.24 (2.98) [0.81]
5 ~ 10	0.34 (0.15) [1.06]	-	0.55 (1.76) [2.55]	0.04 (0.00) [1.11]	0.04 (0.00) [0.00]	2.33 (0.14) [0.18]	0.42 (4.50) [0.83]
0 ~ 5	35.27 (38.57) [58.71]	-	30.25 (39.26) [66.59]	35.45 (42.26) [48.65]	23.03 (21.04) [31.40]	35.98 (29.55) [42.89]	28.23 (87.05) [68.52]
0 ~ -5	50.39 (53.59) [34.94]	-	53.85 (46.89) [24.73]	56.28 (49.62) [43.71]	61.80 (66.43) [52.85]	47.86 (57.92) [46.70]	54.31 (5.15) [26.87]
-5 ~ -10	6.96 (3.57) [2.29]	-	8.56 (5.52) [1.53]	4.30 (4.29) [3.14]	7.83 (6.25) [6.81]	5.23 (5.56) [5.46]	7.89 (0.18) [1.27]
< -10	6.50 (3.98) [1.96]	-	6.13 (5.90) [1.24]	3.91 (3.82) [2.86]	7.30 (6.26) [8.95]	6.25 (6.64) [4.57]	8.90 (0.14) [1.69]

(ii) RCP 8.5

RB	GEPIC (%)	GAEZ-IMAGE (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	1.99 (0.01) [0.03]	1.27 (0.02) [0.02]	1.99 (0.02) [0.04]	0.97 (0.00) [0.00]	0.10 (0.00) [0.00]	1.02 (0.03) [0.02]	0.84 (0.01) [0.00]
5 ~ 10	2.02 (0.05) [0.09]	1.66 (0.01) [0.01]	1.95 (0.12) [0.04]	0.62 (0.00) [0.00]	0.19 (0.00) [0.00]	1.09 (0.02) [0.06]	0.95 (0.01) [0.02]
0 ~ 5	37.75 (33.88) [21.45]	25.10 (76.56) [74.31]	42.79 (35.21) [31.41]	40.02 (31.75) [25.63]	24.18 (30.27) [23.00]	41.85 (38.12) [29.68]	32.41 (34.84) [32.92]
0 ~ -5	53.58 (56.75) [65.07]	69.59 (21.54) [23.60]	46.76 (54.53) [62.30]	54.74 (56.06) [61.48]	64.63 (56.25) [63.32]	49.75 (55.82) [64.19]	56.14 (60.91) [63.87]
-5 ~ -10	2.43 (4.80) [7.07]	1.19 (1.04) [1.04]	3.41 (4.12) [3.82]	1.92 (5.56) [6.64]	5.37 (5.67) [6.89]	3.10 (3.49) [3.07]	4.92 (2.20) [1.91]
< -10	2.22 (4.52) [6.29]	1.19 (0.83) [1.03]	3.09 (6.10) [2.39]	1.73 (6.63) [6.24]	5.53 (7.81) [6.78]	3.20 (2.52) [2.97]	4.73 (2.05) [1.27]

[a] Regressions for out-of-sample validation were run on a panel spanning 1972~2089, in contrast to the in-sample which are on 1972~2099

[b] Data not available

Soybeans

(i) RCP 4.5

RB	GEPIC (%)	GAEZ-IMAGE[b] (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	0.38 (0.86) [1.32]	-	3.70 (7.38) [11.85]	1.20 (0.23) [0.28]	0.00 (0.00) [0.00]	0.01 (0.00) [0.03]	1.17 (8.78) [1.27]
5 ~ 10	0.92 (0.10) [1.48]	-	3.22 (4.70) [6.62]	0.77 (0.43) [0.22]	0.00 (0.00) [0.00]	0.00 (0.00) [0.00]	1.71 (8.30) [2.03]
0 ~ 5	52.16 (61.62) [65.75]	-	41.33 (38.02) [53.28]	42.36 (46.50) [62.34]	11.00 (9.27) [34.11]	71.01 (64.54) [62.70]	50.82 (82.13) [87.02]
0 ~ -5	41.38 (32.44) [24.88]	-	44.55 (42.70) [25.08]	42.63 (44.87) [30.89]	61.61 (86.84) [61.18]	24.96 (33.80) [30.24]	42.62 (0.71) [9.04]
-5 ~ -10	2.71 (2.13) [3.32]	-	3.46 (4.32) [1.58]	7.37 (2.79) [2.71]	16.29 (1.88) [2.50]	1.30 (0.85) [3.93]	1.61 (0.07) [0.41]
< -10	2.45 (2.84) [3.24]	-	3.73 (2.88) [1.59]	5.68 (5.18) [3.55]	11.09 (2.01) [2.21]	2.72 (0.81) [3.11]	2.06 (0.00) [0.24]

(ii) RCP 8.5

RB	GEPIC (%)	GAEZ-IMAGE (%)	LPJ-GUESS (%)	LPJmL (%)	pDSSAT (%)	PEGASUS (%)	Multi-GGCM (%)
> 10	4.51 (0.02) [0.01]	0.84 (0.03) [0.02]	3.70 (0.10) [0.09]	3.45 (0.01) [0.03]	2.23 (0.00) [0.00]	0.23 (0.00) [0.00]	4.33 (0.02) [0.03]
5 ~ 10	4.90 (0.01) [0.02]	0.63 (0.03) [0.02]	5.64 (0.10) [0.08]	2.81 (0.08) [0.02]	1.10 (0.00) [0.00]	0.22 (0.00) [0.00]	4.10 (0.02) [0.06]
0 ~ 5	71.28 (54.69) [48.38]	18.25 (63.89) [51.64]	67.01 (48.46) [32.54]	53.00 (36.09) [31.60]	26.99 (6.65) [25.23]	63.87 (81.25) [80.97]	76.47 (39.87) [23.51]
0 ~ -5	18.68 (40.82) [46.19]	71.72 (34.07) [45.51]	22.89 (48.89) [64.33]	33.61 (58.37) [65.81]	52.41 (90.21) [65.23]	28.24 (17.95) [18.37]	13.36 (56.64) [72.14]
-5 ~ -10	0.14 (2.01) [2.24]	4.52 (1.04) [1.28]	0.18 (1.16) [1.35]	2.85 (4.76) [1.39]	7.79 (2.49) [2.95]	4.15 (0.40) [0.36]	0.86 (1.91) [2.09]
< -10	0.49 (2.44) [3.16]	4.04 (0.95) [1.53]	0.57 (1.28) [1.64]	4.27 (0.68) [1.14]	9.48 (0.65) [6.59]	3.30 (0.40) [0.30]	0.88 (1.54) [2.18]

[a] Regressions for out-of-sample validation were run on a panel spanning 1972~2089, in contrast to the in-sample which are on 1972~2099

[b] Data not available

Appendix B: Supplementary material for Chapter 3

1. **Agro-Ecological Zones (AEZ)**

A detailed discussion of AEZ is provided in (Lee *et al* 2005) . The map in figure 1B shows the 18 AEZs formed by overlaying the combination of 6 categories of duration of growing period (DGP) with the 3 climatic zones (table 1B).

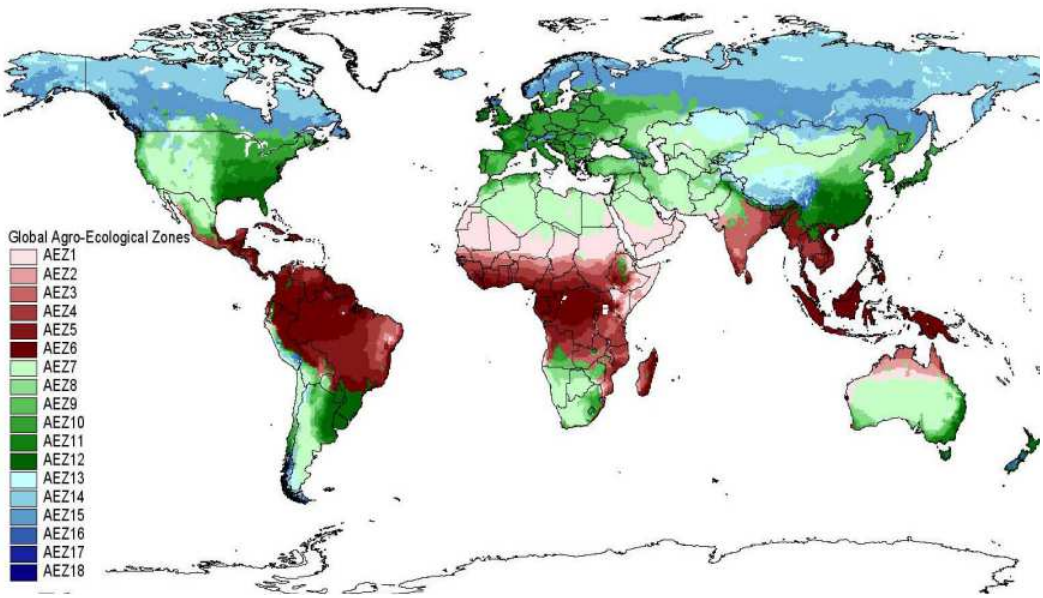


Figure 1B Map of 18 global AEZs at 0.5 degree grid cell resolution (Source Lee *et al* 2005)

Table 1B Definition of 18 global AEZs as defined by DGP and climate zone (Source Lee *et al* 2005)

DGP in days[a]	Climate Zones		
	Tropical	**Temperate**	**Boreal**
0-59	AEZ1	AEZ7	AEZ13
60-119	AEZ2	AEZ8	AEZ14
120-179	AEZ3	AEZ9	AEZ15
180-239	AEZ4	AEZ10	AEZ16
240-299	AEZ5	AEZ11	AEZ17
>300	AEZ6	AEZ12	AEZ18

[a] DGP $<$ 120 days is classed as Short Growing Period (SGP) and DGP $>$ 120 days as Long Growing Period (LGP)

Table 2B Grouping of 6AEZs (from the parent 18 AEZs) used in this study (patterned after Blanc and Sultan 2015)

AEZ groups	Growing Period	Climate Zones	AEZ
AEZ_G1	SGP	Tropical	1, 2
AEZ_G2	LGP	Tropical	3, 4, 5, 6
AEZ_G3	SGP	Temperate	7, 8
AEZ_G4	LGP	Temperate	9, 10, 11, 12
AEZ_G5	SGP	Boreal	13, 14
AEZ_G6	LGP	Boreal	15, 16, 17, 18

Table 3B Number of observations used in regressionsa for 3 GGCMs by 6 AEZ groups.

AEZ groups	GAEZ-IMAGE	pDSSAT	PEGASUS
AEZ_G1	77,862	72,452	96,760
AEZ_G2	1,137,948	633,600	1,143,302
AEZ_G3	236,778	169,566	381,376
AEZ_G4	961,134	500,792	876,622
AEZ_G5	19,608	10,266	41,182
AEZ_G6	66,348	4,130	92,040

[a] Regression was run over a 118-years balanced panel spanning 1972-2089

Table 4B Regression summary with Clustered Robust Standard Errors (S.E.s) in parentheses: base_VPD specification[62]

Dependent variable: log. Yield (Maize)						
	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS
tas_5lo	-0.0177336***	-0.0118239***	-0.0184172***	-0.0012387	-0.0042704*	-0.0498848***
	(0.0037375)	(0.0013739)	(0.0031860)	(0.0063179)	(0.0023194)	(0.0033731)
tas_5_7p5	-0.0051788*	-0.0028661***	-0.0102919***	-0.0030695	-0.0105207***	-0.0366603***
	(0.0027153)	(0.0004868)	(0.0019782)	(0.0051481)	(0.0020467)	(0.0028614)
tas_7p5_10	-0.0064849***	-0.0042128***	-0.0120987***	0.0046670	-0.0038158***	-0.0279764***
	(0.0017033)	(0.0003216)	(0.0018656)	(0.0028712)	(0.0011401)	(0.0027144)
tas_10_12p5	-0.0062979***	-0.0018308***	-0.0118638***	0.0023509	-0.0059679***	-0.0226616***
	(0.0018142)	(0.0002009)	(0.0015682)	(0.0035341)	(0.0008813)	(0.0018919)
tas_12p5_15	-0.0014921	-0.0002855***	-0.0073496***	-0.0095538***	-0.0024764***	-0.0058462***
	(0.0011153)	(0.0001017)	(0.0008666)	(0.0026029)	(0.0007470)	(0.0014180)
tas_17p5_20	-0.0085895***	-0.0001049*	-0.0006156	-0.0028565	-0.0004941	0.0037487***
	(0.0009173)	(0.0000615)	(0.0005218)	(0.0019033)	(0.0003095)	(0.0006311)
tas_20_22p5	-0.0186937***	-0.0004481***	-0.0040602***	-0.0018595	-0.0020648***	0.0029576***
	(0.0010020)	(0.0000876)	(0.0004863)	(0.0013047)	(0.0003790)	(0.0008083)
tas_22p5_25	-0.0242674***	-0.0003823***	-0.0063491***	-0.0049439***	-0.0030794***	-0.0012274
	(0.0011594)	(0.0001060)	(0.0005168)	(0.0012811)	(0.0004131)	(0.0007862)
tas_25_27p5	-0.0264449***	-0.0006084***	-0.0076854***	-0.0074324***	-0.0031882***	-0.0062677***
	(0.0012334)	(0.0001228)	(0.0005319)	(0.0012530)	(0.0005006)	(0.0008033)
tas_27p5_30	-0.0264329***	-0.0008097***	-0.0081796***	-0.0085444***	-0.0037655***	-0.0111963***
	(0.0013229)	(0.0001380)	(0.0005754)	(0.0012950)	(0.0005059)	(0.0008416)
tas_g30	-0.0232541***	-0.0012768***	-0.0080615***	-0.0084584***	-0.0083611***	-0.0184335***
	(0.0016513)	(0.0001548)	(0.0006450)	(0.0014628)	(0.0006022)	(0.0009513)
p_3lo	-0.0135086***	-0.0003510***	-0.0073808***	-0.0049246***	-0.0054347***	-0.0005318
	(0.0014306)	(0.0000959)	(0.0006058)	(0.0006405)	(0.0006168)	(0.0005150)
p_3_4	0.0104888***	0.0003215***	-0.0015452**	-0.0001066	0.0003909	0.0021593***
	(0.0012410)	(0.0000448)	(0.0006354)	(0.0013515)	(0.0005614)	(0.0004079)
p_4_5	0.0060332***	0.0001272***	-0.0004421	-0.0011114	0.0000141	0.0012588***
	(0.0007366)	(0.0000334)	(0.0004293)	(0.0009217)	(0.0003991)	(0.0002811)
p_10_15	0.0027802***	-0.0000711**	0.0011158**	0.0013492*	-0.0009468***	-0.0017750***
	(0.0006500)	(0.0000292)	(0.0004456)	(0.0007229)	(0.0002922)	(0.0002249)
p_15_20	0.0023372***	-0.0000949***	0.0010680**	0.0027287***	-0.0024371***	-0.0028451***
	(0.0008117)	(0.0000364)	(0.0004263)	(0.0010147)	(0.0006366)	(0.0002868)
p_20up	0.0042405***	-0.0001573**	0.0021316***	0.0029687***	-0.0034095***	-0.0044260***
	(0.0014670)	(0.0000705)	(0.0006640)	(0.0010361)	(0.0007378)	(0.0004158)
vpd_3lo	-0.0111219***	-0.0003247***	-0.0078033***	-0.0088972***	-0.0010695*	-0.0048970***
	(0.0007712)	(0.0000964)	(0.0004960)	(0.0011717)	(0.0006019)	(0.0006626)
vpd_3_5	-0.0057143***	0.0001085**	-0.0026612***	-0.0020935***	0.0019487***	-0.0037303***
	(0.0005477)	(0.0000470)	(0.0002672)	(0.0005008)	(0.0004875)	(0.0003684)
vpd_5_7	-0.0007941	0.0001024***	0.0002537	-0.0008199**	0.0001878	-0.0013454***
	(0.0004968)	(0.0000257)	(0.0003691)	(0.0004126)	(0.0003215)	(0.0002584)
vpd_9_11	0.0056991***	0.0001385**	0.0028883***	0.0010107*	0.0034148***	0.0019872***
	(0.0008826)	(0.0000541)	(0.0004212)	(0.0005975)	(0.0003440)	(0.0003639)
vpd_11up	-0.0095447***	0.0000398	-0.0028746***	-0.0058014***	-0.0005386*	0.0037319***
	(0.0007443)	(0.0000609)	(0.0002904)	(0.0004569)	(0.0003227)	(0.0003640)
tas_5lo:Time_trend	0.0004842***		0.0001605***	0.0005807***		
	(0.0000953)		(0.0000549)	(0.0001295)		

[62] The bins $T = 15{\sim}17.5\,°C$, $P = 5{\sim}10\,mm/d$ and $VPD = 7{\sim}9\,hPa$ were omitted in regressions as reference category. GGCMs GAEZ-IMAGE, pDSSAT and PEGASUS do not include interaction terms in regression.

Table 4B…*continued*

	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS
tas_5_7p5:Time_trend	0.0000143		0.0000112	0.0000729		
	(0.0000388)		(0.0000199)	(0.0000807)		
tas_7p5_10:Time_trend	0.0000402		0.0000451**	-0.0001566***		
	(0.0000303)		(0.0000192)	(0.0000471)		
tas_10_12p5:Time_trend	-0.0000748***		0.0000818***	-0.0000581		
	(0.0000286)		(0.0000183)	(0.0000506)		
tas_12p5_15:Time_trend	-0.0000297**		0.0000618***	0.0000766***		
	(0.0000143)		(0.0000106)	(0.0000233)		
tas_17p5_20:Time_trend	0.0000767***		0.0000145**	0.0000672***		
	(0.0000109)		(0.0000065)	(0.0000194)		
tas_20_22p5:Time_trend	0.0001258***		0.0000394***	0.0000427***		
	(0.0000101)		(0.0000054)	(0.0000122)		
tas_22p5_25:Time_trend	0.0000950***		0.0000297***	0.0000344***		
	(0.0000107)		(0.0000054)	(0.0000121)		
tas_25_27p5:Time_trend	0.0000444***		0.0000103**	0.0000077		
	(0.0000080)		(0.0000049)	(0.0000099)		
tas_27p5_30:Time_trend	0.0000111		-0.0000056	-0.0000068		
	(0.0000087)		(0.0000051)	(0.0000104)		
tas_g30:Time_trend	-0.0000439***		-0.0000147***	-0.0000348***		
	(0.0000123)		(0.0000057)	(0.0000119)		
p_3lo:Time_trend	0.0000064		0.0000094**	-0.0000217***		
	(0.0000076)		(0.0000041)	(0.0000077)		
p_3_4:Time_trend	-0.0000511***		0.0000670***	0.0000585**		
	(0.0000189)		(0.0000107)	(0.0000235)		
p_4_5:Time_trend	-0.0000328***		0.0000276***	0.0000513***		
	(0.0000112)		(0.0000067)	(0.0000157)		
p_10_15:Time_trend	-0.0000479***		-0.0000135**	-0.0000279**		
	(0.0000096)		(0.0000068)	(0.0000117)		
p_15_20:Time_trend	-0.0000244***		-0.0000003	-0.0000422***		
	(0.0000095)		(0.0000064)	(0.0000146)		
p_20up:Time_trend	-0.0000453***		0.0000180***	-0.0000313**		
	(0.0000157)		(0.0000061)	(0.0000143)		
vpd_3lo:Time_trend	0.0000182***		0.0000057	0.0000923***		
	(0.0000066)		(0.0000072)	(0.0000206)		
vpd_3_5:Time_trend	-0.0000110*		-0.0000088**	0.0000030		
	(0.0000057)		(0.0000043)	(0.0000077)		
vpd_5_7:Time_trend	-0.0000140*		-0.0000102**	0.0000032		
	(0.0000075)		(0.0000050)	(0.0000070)		
vpd_9_11:Time_trend	0.0000644***		0.0000197***	0.0000389***		
	(0.0000107)		(0.0000048)	(0.0000090)		
vpd_11up:Time_trend	0.0000479***		0.0000221***	0.0000418***		
	(0.0000074)		(0.0000032)	(0.0000060)		
Observations	2,519,654	2,684,928	3,091,600	2,039,512	1,500,606	2,642,138
R2	0.5376031	0.6882976	0.8544042	0.8208722	0.7312246	0.8593720
Adjusted R2	0.5336430	0.6855367	0.8531577	0.8193373	0.7289236	0.8581690
Residual S.E.	0.6938001(df=2498257)	0.1128816(df=2661354)	0.2712849(df=3065356)	0.3683743(df=2022184)	0.4007698(df=1487867)	0.4729892(df=2619725)
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 5B Regression summary with Clustered Robust S.E.s in parentheses: AEZ specification[63]

	Dependent variable: log. Yield (Maize)		
	GAEZ-IMAGE	pDSSAT	PEGASUS
tas_12p5lo	-0.0175499	0.6082852***	-0.0286904
	(0.0432454)	(0.1983485)	(0.0963355)
tas_12p5_15	0.0263049	-0.0299457	0.0589668
	(0.0328767)	(0.0877869)	(0.0435692)
tas_17p5_20	-0.0019484	0.0027642	0.0061861
	(0.0030453)	(0.0052020)	(0.0125026)
tas_20_22p5	-0.0014653	0.0027368	0.0061610
	(0.0031670)	(0.0051457)	(0.0124762)
tas_22p5_25	-0.0014710	0.0009220	-0.0000347
	(0.0031542)	(0.0057823)	(0.0124104)
tas_25_27p5	-0.0017343	0.0039287	0.0002837
	(0.0031696)	(0.0062820)	(0.0125418)
tas_g27p5	-0.0022958	-0.0013277	-0.0129658
	(0.0031351)	(0.0064768)	(0.0126447)
p_3lo	-0.0039287***	-0.0093082***	0.0067191***
	(0.0009376)	(0.0014002)	(0.0019596)
p_3_4	0.0014507***	0.0004925	0.0032451***
	(0.0004297)	(0.0020038)	(0.0011202)
p_4_5	0.0001957	-0.0006230	0.0026380***
	(0.0004185)	(0.0013057)	(0.0005947)
p_10_15	0.0010980**	-0.0017090	-0.0067325***
	(0.0005111)	(0.0015430)	(0.0009499)
p_15_20	0.0031227***	-0.0040748	-0.0114888***
	(0.0007553)	(0.0025684)	(0.0017919)
p_20up	0.0058055***	-0.0086988***	-0.0151243***
	(0.0014258)	(0.0032690)	(0.0026495)
tas_12p5lo:AEZ_G2	0.0168003	-0.6071749***	0.0357766
	(0.0432470)	(0.1983567)	(0.0963590)
tas_12p5_15:AEZ_G2	-0.0266874	0.0309361	-0.0547973
	(0.0328758)	(0.0877865)	(0.0435384)
tas_17p5_20:AEZ_G2	0.0018640	-0.0037239	-0.0114066
	(0.0030429)	(0.0052089)	(0.0124797)
tas_20_22p5:AEZ_G2	0.0012130	-0.0048883	-0.0164041
	(0.0031668)	(0.0051475)	(0.0125043)
tas_22p5_25:AEZ_G2	0.0012274	-0.0047881	-0.0151293
	(0.0031532)	(0.0057676)	(0.0124496)
tas_25_27p5:AEZ_G2	0.0013668	-0.0067630	-0.0181971
	(0.0031699)	(0.0062541)	(0.0125893)
tas_g27p5:AEZ_G2	0.0014055	-0.0059235	-0.0125035
	(0.0031357)	(0.0064413)	(0.0126957)
p_3lo:AEZ_G2	0.0036221***	0.0041239***	-0.0039663**
	(0.0009212)	(0.0015953)	(0.0019684)
p_3_4:AEZ_G2	-0.0013396***	-0.0028215	0.0006563
	(0.0004351)	(0.0022910)	(0.0011570)
p_4_5:AEZ_G2	-0.0001180	-0.0008333	0.0002160
	(0.0004217)	(0.0014731)	(0.0006422)
p_10_15:AEZ_G2	-0.0011785**	0.0005461	0.0035842***
	(0.0005095)	(0.0015855)	(0.0009917)
p_15_20:AEZ_G2	-0.0033065***	0.0007171	0.0071287***
	(0.0007556)	(0.0027543)	(0.0018157)
p_20up:AEZ_G2	-0.0061944***	0.0034092	0.0071017***
	(0.0014242)	(0.0034329)	(0.0026451)
tas_12p5lo:AEZ_G3	0.0156022	-0.6110540***	0.0203225
	(0.0432144)	(0.1983033)	(0.0964250)
tas_12p5_15:AEZ_G3	-0.0267040	0.0298569	-0.0591969
	(0.0328889)	(0.0878120)	(0.0433849)
tas_17p5_20:AEZ_G3	0.0024903	-0.0026099	-0.0020008
	(0.0030616)	(0.0052179)	(0.0125785)
tas_20_22p5:AEZ_G3	0.0021101	-0.0041914	0.0022694
	(0.0031907)	(0.0052423)	(0.0126057)

[63] The bins $T = 15{\sim}17.5\ ^{\circ}C$ and $P = 5{\sim}10\ mm/d\ hPa$, along with their interactions with AEZ groups were omitted in regressions as reference category.

Table 5B…*continued*

	GAEZ-IMAGE	pDSSAT	PEGASUS
tas_22p5_25:AEZ_G3	0.0023109	-0.0047541	0.0052403
	(0.0031546)	(0.0058588)	(0.0124963)
tas_25_27p5:AEZ_G3	0.0038144	-0.0066287	0.0033217
	(0.0031721)	(0.0063039)	(0.0125872)
tas_g27p5:AEZ_G3	0.0023485	-0.0014976	0.0082639
	(0.0031514)	(0.0064892)	(0.0127296)
p_3lo:AEZ_G3	0.0008879	-0.0091029***	-0.0145943***
	(0.0010354)	(0.0024081)	(0.0027622)
p_3_4:AEZ_G3	-0.0014108**	-0.0037512	0.0048413***
	(0.0005536)	(0.0024040)	(0.0015877)
p_4_5:AEZ_G3	-0.0006080	-0.0016682	0.0021868*
	(0.0005180)	(0.0016339)	(0.0011268)
p_10_15:AEZ_G3	-0.0005960	0.0108563***	0.0046658***
	(0.0006548)	(0.0023252)	(0.0016220)
p_15_20:AEZ_G3	-0.0015856	0.0191247***	0.0075970***
	(0.0010071)	(0.0037649)	(0.0027004)
p_20up:AEZ_G3	-0.0055651***	0.0284752***	0.0032197
	(0.0015688)	(0.0070636)	(0.0036175)
tas_12p5lo:AEZ_G4	0.0159382	-0.6148303***	-0.0050338
	(0.0432406)	(0.1983412)	(0.0963158)
tas_12p5_15:AEZ_G4	-0.0270213	0.0265906	-0.0659889
	(0.0328770)	(0.0877692)	(0.0435489)
tas_17p5_20:AEZ_G4	0.0020431	-0.0029484	-0.0018705
	(0.0030449)	(0.0051972)	(0.0123051)
tas_20_22p5:AEZ_G4	0.0011826	-0.0041325	-0.0018988
	(0.0031675)	(0.0051750)	(0.0122773)
tas_22p5_25:AEZ_G4	0.0011486	-0.0041279	0.0000644
	(0.0031538)	(0.0057424)	(0.0122857)
tas_25_27p5:AEZ_G4	0.0012374	-0.0071803	-0.0025078
	(0.0031663)	(0.0062554)	(0.0124362)
tas_g27p5:AEZ_G4	0.0013617	-0.0054574	0.0023431
	(0.0031297)	(0.0064643)	(0.0125383)
p_3lo:AEZ_G4	0.0038778***	0.0004712	-0.0057653***
	(0.0009381)	(0.0015863)	(0.0020985)
p_3_4:AEZ_G4	-0.0012409***	0.0008060	-0.0021349
	(0.0004417)	(0.0020342)	(0.0013547)
p_4_5:AEZ_G4	-0.0001007	0.0019279	-0.0016823**
	(0.0004248)	(0.0013436)	(0.0008085)
p_10_15:AEZ_G4	-0.0012675**	0.0024359	0.0040539***
	(0.0005157)	(0.0016111)	(0.0011260)
p_15_20:AEZ_G4	-0.0033486***	0.0053961**	0.0073943***
	(0.0007515)	(0.0027010)	(0.0019398)
p_20up:AEZ_G4	-0.0060063***	0.0100730***	0.0093106***
	(0.0014306)	(0.0037032)	(0.0028415)
tas_12p5lo:AEZ_G5	0.0140662	-0.6169890***	0.0096670
	(0.0432403)	(0.1983570)	(0.0963748)
tas_12p5_15:AEZ_G5	-0.0253089	0.0247339	-0.0657158
	(0.0328887)	(0.0878145)	(0.0435797)
tas_17p5_20:AEZ_G5	0.0067638**	-0.0011473	0.0041585
	(0.0033593)	(0.0053854)	(0.0125922)
tas_20_22p5:AEZ_G5	0.0042321	-0.0049094	0.0072913
	(0.0033614)	(0.0051924)	(0.0125540)
tas_22p5_25:AEZ_G5	0.0023931	-0.0049593	0.0131044
	(0.0033867)	(0.0058143)	(0.0125070)
tas_25_27p5:AEZ_G5	0.0048066	-0.0086230	0.0157380
	(0.0035692)	(0.0063737)	(0.0126329)
tas_g27p5:AEZ_G5	-0.0028015	-0.0042500	0.0042584
	(0.0033098)	(0.0065836)	(0.0127530)
p_3lo:AEZ_G5	-0.0010889	-0.0119311***	0.0006122
	(0.0015680)	(0.0017555)	(0.0028969)
p_3_4:AEZ_G5	-0.0024456*	-0.0129316***	-0.0000085
	(0.0013810)	(0.0024720)	(0.0022930)
p_4_5:AEZ_G5	-0.0010316	-0.0089168***	-0.0011482
	(0.0010144)	(0.0017763)	(0.0018550)
p_10_15:AEZ_G5	0.0018873	0.0133250***	0.0034796
	(0.0018996)	(0.0028346)	(0.0025361)

Table 5B…*continued*

	GAEZ-IMAGE	pDSSAT	PEGASUS
p_15_20:AEZ_G5	0.0024267	0.0347399***	0.0130403***
	(0.0034244)	(0.0080764)	(0.0045838)
p_20up:AEZ_G5	-0.0017109	0.0624820***	0.0181468***
	(0.0052586)	(0.0147142)	(0.0053252)
tas_12p5lo:AEZ_G6	0.0123906	-0.6244849***	0.0021541
	(0.0432429)	(0.1983523)	(0.0963382)
tas_12p5_15:AEZ_G6	-0.0259563	0.0227506	-0.0628861
	(0.0328784)	(0.0878291)	(0.0436082)
tas_17p5_20:AEZ_G6	0.0024082	-0.0023164	0.0021435
	(0.0030532)	(0.0055523)	(0.0125461)
tas_20_22p5:AEZ_G6	0.0010396	-0.0020706	0.0059469
	(0.0031798)	(0.0053659)	(0.0125001)
tas_22p5_25:AEZ_G6	-0.0002792	-0.0058395	0.0155059
	(0.0031655)	(0.0059670)	(0.0124283)
tas_25_27p5:AEZ_G6	0.0002128	-0.0098301	0.0122426
	(0.0031858)	(0.0064858)	(0.0125626)
tas_g27p5:AEZ_G6	0.0000228	-0.0125823*	0.0137153
	(0.0031738)	(0.0066752)	(0.0126853)
p_3lo:AEZ_G6	0.0039661***	0.0034641	0.0075283***
	(0.0009637)	(0.0028234)	(0.0020805)
p_3_4:AEZ_G6	-0.0009404*	0.0003256	0.0010862
	(0.0005670)	(0.0053222)	(0.0013721)
p_4_5:AEZ_G6	0.0000282	0.0021495	-0.0009648
	(0.0005339)	(0.0037311)	(0.0009847)
p_10_15:AEZ_G6	-0.0013295*	0.0129819***	0.0024959*
	(0.0006842)	(0.0046049)	(0.0013293)
p_15_20:AEZ_G6	-0.0030981***	0.0059540	0.0043432*
	(0.0009343)	(0.0080411)	(0.0022483)
p_20up:AEZ_G6	-0.0034198**	0.0116821	0.0026233
	(0.0015682)	(0.0099285)	(0.0030226)
Observations	2,499,678	1,390,866	2,631,282
R2	0.6732797	0.7242579	0.8559615
Adjusted R2	0.6703781	0.7218856	0.8547261
Residual S.E.	0.1146734 (df = 2477673)	0.4096623 (df = 1379001)	0.4787755 (df = 2608905)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6B Regression summary with Clustered Robust S.E.s in parentheses: SLX specification [64]

	Dependent variable: log. Yield (Maize)		
	GAEZ-IMAGE	pDSSAT	PEGASUS
tas_5lo	-0.0230029***	-0.0269091***	-0.0638193***
	(0.0021500)	(0.0045765)	(0.0058359)
tas_5_7p5	-0.0079817***	-0.0193372***	-0.0459680***
	(0.0008647)	(0.0027104)	(0.0039757)
tas_7p5_10	-0.0053334***	-0.0090300***	-0.0321966***
	(0.0004089)	(0.0016311)	(0.0026436)
tas_10_12p5	-0.0017591***	-0.0049676***	-0.0173139***
	(0.0002414)	(0.0008198)	(0.0015128)
tas_12p5_15	-0.0004220***	-0.0015532***	-0.0028145**
	(0.0001029)	(0.0003411)	(0.0013020)
tas_17p5_20	0.0001053	0.0000145	0.0017176***
	(0.0000855)	(0.0002531)	(0.0004515)
tas_20_22p5	-0.0000471	-0.0003695	0.0009369
	(0.0001570)	(0.0004670)	(0.0006430)
tas_22p5_25	0.0001152	-0.0002868	-0.0007593
	(0.0002152)	(0.0006555)	(0.0008028)
tas_25_27p5	0.0000389	0.0000459	-0.0018625*
	(0.0002609)	(0.0008241)	(0.0010231)
tas_27p5_30	-0.0000372	0.0003257	-0.0025739**
	(0.0002974)	(0.0009896)	(0.0012235)
tas_g30	-0.0001746	-0.0003705	-0.0020536
	(0.0003324)	(0.0011687)	(0.0014116)
p_3lo	0.0001473**	-0.0022185***	-0.0013968***
	(0.0000702)	(0.0005018)	(0.0003799)
p_3_4	0.0002448***	0.0001290	-0.0002076
	(0.0000571)	(0.0003064)	(0.0002697)
p_4_5	0.0000435	-0.0004630**	-0.0001603
	(0.0000317)	(0.0002146)	(0.0001857)
p_10_15	-0.0000713**	-0.0000586	0.0001496
	(0.0000329)	(0.0002305)	(0.0002082)
p_15_20	-0.0001403**	-0.0007140*	0.0009527***
	(0.0000555)	(0.0004088)	(0.0003418)
p_20up	-0.0003027***	-0.0016558**	0.0017679***
	(0.0001026)	(0.0006474)	(0.0004848)
tas_5lo_slag	0.0145492***	0.0240006***	0.0112411*
	(0.0021011)	(0.0051637)	(0.0067975)
tas_5_7p5_slag	0.0050995***	0.0079912***	0.0086478*
	(0.0009276)	(0.0030695)	(0.0050011)
tas_7p5_10_slag	0.0011268**	0.0063787***	0.0055387
	(0.0004996)	(0.0017170)	(0.0037834)
tas_10_12p5_slag	-0.0006770**	-0.0010081	-0.0067201***
	(0.0003299)	(0.0011925)	(0.0025029)
tas_12p5_15_slag	-0.0000551	-0.0022408**	-0.0038046***
	(0.0001738)	(0.0010097)	(0.0012225)
tas_17p5_20_slag	-0.0002621**	-0.0000638	0.0060855***
	(0.0001293)	(0.0005550)	(0.0006977)
tas_20_22p5_slag	-0.0004846***	-0.0019612***	0.0060967***
	(0.0001869)	(0.0006757)	(0.0008104)
tas_22p5_25_slag	-0.0004864**	-0.0032427***	0.0038807***
	(0.0002364)	(0.0008046)	(0.0009836)
tas_25_27p5_slag	-0.0006530**	-0.0035751***	0.0006559
	(0.0002742)	(0.0009700)	(0.0012372)
tas_27p5_30_slag	-0.0007619**	-0.0044215***	-0.0022494
	(0.0003078)	(0.0010748)	(0.0014485)
tas_g30_slag	-0.0011223***	-0.0090862***	-0.0091685***
	(0.0003337)	(0.0012385)	(0.0016962)
p_3lo_slag	-0.0005219***	-0.0043274***	0.0062364***
	(0.0001337)	(0.0007592)	(0.0006468)

[64] The bins $T = 15{\sim}17.5\ ℃$ and $P = 5{\sim}10\ mm/d$ along with their spatial lagged terms were omitted in regressions as reference category.

Table 6B…*continued*

	GAEZ-IMAGE	pDSSAT	PEGASUS
p_3_4_slag	0.0005469***	0.0034490**	0.0065028***
	(0.0000890)	(0.0013739)	(0.0010031)
p_4_5_slag	0.0000192	-0.0004045	0.0035296***
	(0.0000596)	(0.0008515)	(0.0004865)
p_10_15_slag	0.0000112	-0.0005484	-0.0032466***
	(0.0000537)	(0.0005301)	(0.0003826)
p_15_20_slag	0.0000184	-0.0027286**	-0.0041564***
	(0.0000782)	(0.0011673)	(0.0005789)
p_20up_slag	0.0001189	-0.0018568	-0.0089135***
	(0.0001436)	(0.0012359)	(0.0008010)
Observations	2,684,928	1,500,606	2,642,138
R2	0.6891828	0.7313940	0.8584584
Adjusted R2	0.6864283	0.7290922	0.8572468
Residual S.E.	0.1127215 (df = 2661342)	0.4006451 (df = 1487855)	0.4745243 (df = 2619713)
Note:			*p<0.1; **p<0.05; ***p<0.01

Appendix C: Supplementary material for Chapter 4

1. Data

Since the study primarily compares crop models' simulated data with historical observed data; and the resulting coefficient estimates derived from regressions run on their respective panel, I discuss the two data independently.

1.1 USDA historical observed data

For comparison of the GGCMs' yields with the factual yields, I employ historical observed annual county level production $(bushel, bu)$ and harvested areas $(acre, a)$, made available by the USDA Quickstats database[65]. The data utilized in this study covers ~1500 - 1800 counties in the U.S. over the period 1972–2004 (33 years). Crop yields (bu/a) for each county are calculated as the ratio of production to harvested area. The conversion from bu/a to t/ha for each crop (for consistency with GGCMs' yield units) is described as below (table 1C).

Table 1C. Conversion from bu/a to t/ha for each crop

Crop	bu/a	t/ha
Maize	1	0.0628
Wheat	1	0.0673
Soybeans	1	0.0673

Some limitations to the USDA data include the inability to differentiate the irrigated and rainfed production by county. In line with my earlier analyses, I focus on rainfed regime of the ISIMIP-FT GGCMs' simulations. Therefore, to maintain consistency with model data, I need to formulate a methodology to differentiate the irrigated and rainfed crop production by county. Following Sue Wing *et al* (2015), I utilize data on irrigation infrastructure collected from the U.S. Army Corps of Engineers' National Inventory of Dams (USACE-NID, see Section 2 for further details), and assign the crops as rainfed of those counties where the cumulative quantity of water that is stored and potentially available for application to crops is zero. The final counties retained in regression analyses for each crop are shown in figure 1C (USDA panel). The result is an unbalanced panel (spanning years 1972-2004; see details of observations and number of counties in table 2C).

[65] http://quickstats.nass.usda.gov/

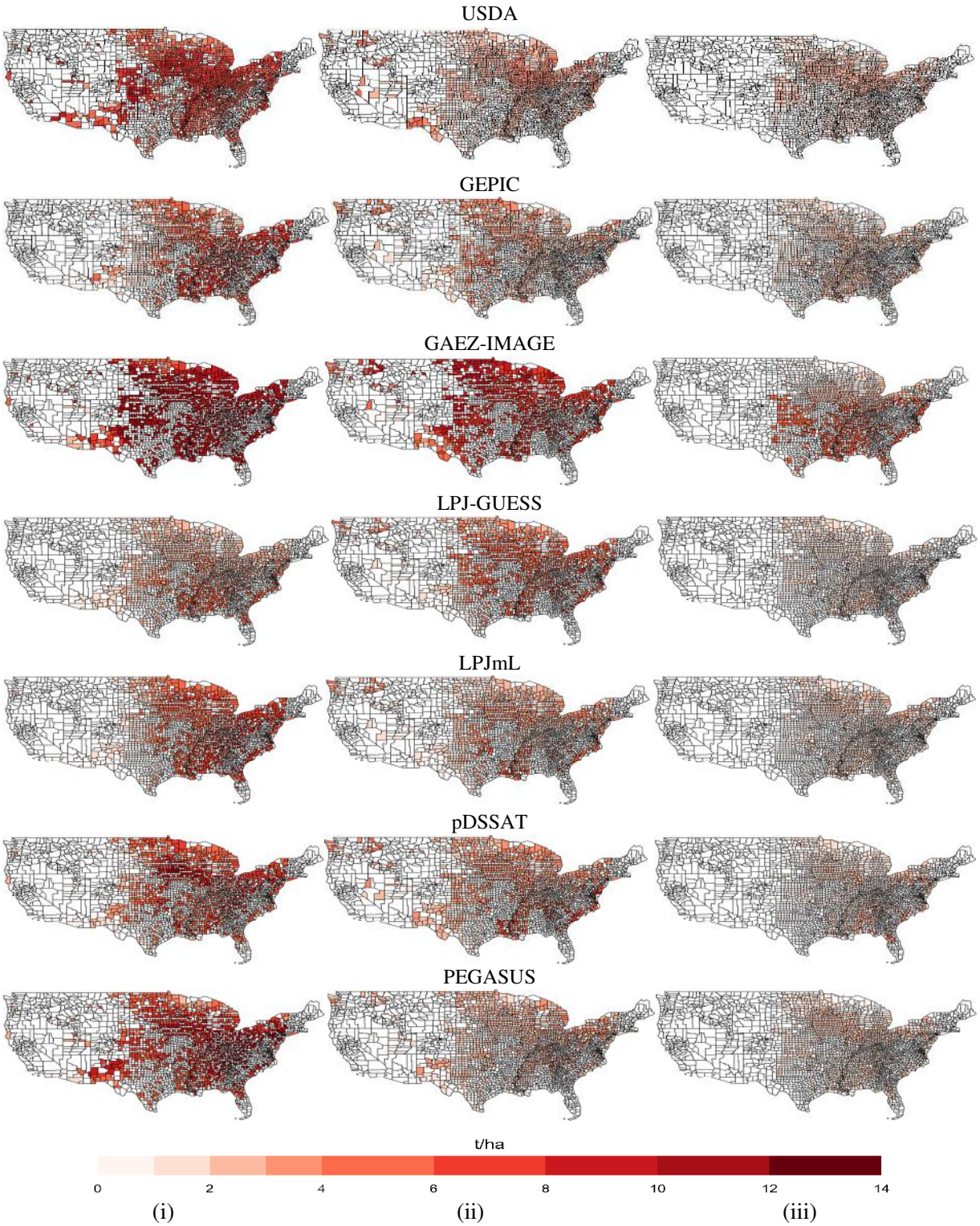


Figure 1C. Maps of USDA and GGCMs' historical (1972~2004) mean county yields (t/ha) for (i) Maize (ii) Wheat and (iii) Soybeans.

1.2. GGCM Data

I utilize the annual gridded rainfed crop yields from the six GGCMs[66] of ISIMIP-FT (Hempel *et al* 2013, Rosenzweig *et al* 2014), as in Chapters 2 and 3. The data used in this study however accounts for carbon dioxide (CO_2) fertilization effects (CFE) and spans over the GGCMs' historical simulation period 1972-2004 (33 years). The global gridded data is subset to retain only the grid-cells covering mainland U.S., using the U.S. county polygon shapes files[67] in R.

An important difference to highlight in Chapter 4 is the spatial scales at which the comparison of GGCMs' and USDA yields are undertaken. Since the ISIMIP-FT data is at 0.5° gridded resolution covering the entire U.S. (recalling that crop modelers were allowed to simulate yields over all land), for consistency with USDA county data I use U.S. county borders to mask the GGCMs' data. The masked GGCM's data would therefore retain all grid-cells that are encompassed within the overall county borders. Although this would result in a marginal higher number of observations for the GGCMs (compared to the USDA county data), my sensitivity checks reveal similar results if the GGCM gridded data are aggregated to the coordinates of the centroids of each county polygon shape. For consistency with the number of counties utilized in analyses of USDA crop yields, I retain the same counties in the GGCMs' panels (table 2C). There are a few exceptions though with the smaller counties ($< 0.5°$ gridded resolution) not encompassing any grid-cells, and thus dropping out in the extraction process. The number of grid-cells, counties and observations used in each crop~GGCM combination regression are summarized in table 2C.

1.3 Crop growing seasons

In line with the reasons highlighted in Section 2.2.1 (Chapter 2), for both USDA and GGCMs, I adopt a common, fixed, four-month crops growing season (CGS) as May-August ($MJJA$)[68]. The CGS varies not only for each crop-GGCM combination, but also in the historical and future periods of GGCMs' simulated data (e.g. LPJ-GUESS and LPJmL mimic planting dates according to climatic conditions). However, the definition of CGS as done in this study by and large encapsulates the broader CGS across the GGCMs and crops. By doing so, I avoid the potential endogeneity problem with crop modellers'

[66] For details of modeling groups, see Section 1 in Appendix A.

[67] Shape files were downloaded from here https://www.census.gov/geo/maps-data/data/cbf/cbf_counties.htmlhttps://www.census.gov/geo/maps-data/data/cbf/cbf_counties.html and processed using R packages raster, sp, and rgdal.

[68] Since all counties/grid-cells would fall in the northern hemisphere, the definition of crop growing season is consistent with the months for the northern hemisphere (as in Chapters 2 and 3)

definition of when planting and harvesting begin in each year. Nevertheless, it is important to highlight that the definition of CGS is also not likely to be consistent with the actual CGS for the USDA data (e.g. wheat is partly grown outside these months in the U.S.) and so the overall results of wheat could be marginally influenced by this assumption (for e.g., see Lobell and Field 2007 where in the results are fairly insensitive to the choice of CGS months for six crops examined in their study)

1.4 Historical weather exposure for empirical analysis of USDA crop yields

Using the Global Land Data Assimilation System (GLDAS) 3-hourly assimilated T and P observations at 1° grid (Rodell *et al* 2004), I construct daily T and P $bins$ as weather inputs for my empirical analysis. The T and P variables were bilinearly interpolated to U.S. county boundaries and counts of 3-hour exposure were accumulated over each annual growing season ($MJJA$).

1.5. Historical weather exposure for empirical analysis of GGCMs' crop yields

All GGCM historical (1972-2004) crop yield simulation runs are forced with bias-corrected climate inputs (Hempel *et al* 2013) from HadGEM2-ES (Jones *et al* 2011). In line with analyses in Chapters 2 and 3, I matched the bias corrected HadGEM2-ES daily T and P as bins, to GGCM generated realizations of yield for each year of the historical period using the earlier methods described in Section 2.2.1 (Chapter 2).

2. U.S. Army Corps of Engineers' National Inventory of Dams (USACE-NID) data

To ensure that the counties retained in the USDA panel regressions represent rainfed crop yields, a proxy indicator that can identify counties with potential irrigation facilities is required. As discussed in detail by Sue Wing *et al* (2015) Supplementary Information (SI), I employ data on irrigation infrastructure collected from the USACE-NID, which records facilities' location, primary purpose, storage capacity (of dams) and in-service dates. For simplicity and to avoid any potential discrepancy, I restrict my analyses to those counties where the maximum storage for counties (in the USACE-NID dataset) is assigned '0' for all the years (1972~2004).

3. Binning structure of temperature and precipitation in regressions

The meteorological covariates are defined as the cumulative exposure to intervals ("bins") of T and P during the CGS of each year in both USDA and GGCMs' regression specifications. The bins $\{T_1, \ldots, T_J, P_1, \ldots, P_K\}$ are counts of number of days over the growing season at each GGCM grid-cell

(county for USDA regression) spent in j intervals[69] of $T(Degree\ Celcius, °C)$ and k intervals of $P(millimiter\ per\ day, mm/d)$, where:

$j = \{< 7.5, 7.5\sim10, 10\sim12.5, 12.5\sim15, 15\sim17.5, 17.5\sim20, 20\sim22.5, 22.5\sim25, 25\sim27.5, 27.5\sim30, > 30\}$

and

$k = \{< 5, 5\sim10, 10\sim15, > 15\}$

The bins $j = 22.5\sim25\ °C$ and $k = 10\sim15\ mm/d$ were omitted in regressions as reference category. Thus with reference to equation (4.1) in Chapter 4, each coefficient of T (P) indicates the impact (on $log\ yield$) of an additional day in the j^{th} (k^{th}) interval, relative to a day in the dropped T (P) bin. All our regression specifications were run in R package Linear Fixed Effects (LFE) (Gaure 2013), which can handle arbitrary number of factors (fixed effects) and is tailored for fixed effect estimation on large panel data. To account for heteroscedasticity and autocorrelation in the error term $\varepsilon_{i,t}$ (equation 4.1 of Chapter 4), I used robust standard errors (SE)[70] clustered by grid-cells.

[69] For each T and P bin except the extreme lower and upper values, the lower range is included in the count. The extreme bins are open-ended.

[70] The S.E.s are adjusted for the reduced degrees of freedom (DOF) coming from the dummies which are implicitly present. They are also small-sample corrected

Table 2C. Number of observations, counties (in parentheses) and total grid-cells (in square parentheses) used in GGCMs and USDA regressions

GGCM	Maize	Wheat	Soybeans
GEPIC	43,426 (1,092) [1,352]	41,056 (999) [1,266]	34,908 (950) [1,092]
GAEZ-IMAGE	44,889 (1,107) [1,393]	42,463 (1014) [1,301]	28,279 (900) [1,005]
LPJ-GUESS	44,864 (1,083) [1,360]	42,305 (999) [1,283]	36,332 (943) [1,101]
LPJmL	45,853 (1,107) [1,413]	43,643 (1014) [1,325]	37,019 (964) [1,145]
pDSSAT	41,273 (1,096) [1,381]	41,787 (1014) [1,303]	34,367 (958) [1,121]
PEGASUS	43,901 (1,086) [1,344]	41,281 (1258) [1,621]	33,678 (942) [1,082]
Multi-GGCM[a]	264,206 (1,107) [1,413]	252,535 (1014) [1,325]	204,583 (964) [1,145]
USDA, full panel[b]	72,967 (2,691)	66,249 (2,479)	56,412 (2,193)
USDA, rainfed panel	51,089 (1,828)	42,752 (1,653)	41,510 (1,605)

[a] Multi-GGCM regressions were run on a combined panel of six GGCMs, with an additional factor (GGCM) in the regression specification (equation 4.1 in Chapter 4)

[b] As part of sensitivity checks, regression on USDA panel was also run without differentiating between irrigated and rainfed yields. The estimated were found to be similar to those from the USDA rainfed panel

Table 3C. Regression summary with Clustered Robust S.E. (in parentheses) for rainfed (i) Maize (ii) Soybeans and (iii) Wheat

i. Dependent variable: log. Yield (Maize)

	USDA	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS	Multi-GGCM
tas_7p5lo	0.0029414**	0.0503259***	0.0014086	0.0062907***	0.0267604***	0.0015136	-0.0202699***	0.0109714***
	(0.0013286)	(0.0026835)	(0.0008723)	(0.0008973)	(0.0016423)	(0.0011419)	(0.0012738)	(0.0008560)
tas_7p5_10	0.0006017	0.0219230***	-0.0031222***	0.0055467***	0.0084488***	0.0017254*	-0.0307178***	0.0006218
	(0.0008493)	(0.0020344)	(0.0008188)	(0.0008556)	(0.0014140)	(0.0010307)	(0.0013004)	(0.0006711)
tas_10_12p5	-0.0023824***	0.0371449***	-0.0008933	0.0095065***	0.0132254***	0.0020942***	-0.0254167***	0.0060057***
	(0.0005364)	(0.0021383)	(0.0009162)	(0.0006822)	(0.0012250)	(0.0007898)	(0.0010213)	(0.0007005)
tas_12p5_15	-0.0021785***	0.0285498***	-0.0028130***	0.0088264***	0.0085917***	0.0015205**	-0.0155689***	0.0048474***
	(0.0004239)	(0.0014379)	(0.0007020)	(0.0006246)	(0.0008280)	(0.0006259)	(0.0006721)	(0.0004767)
tas_15_17p5	-0.0024319***	0.0300566***	-0.0009853**	0.0107690***	0.0133147***	0.0043416***	-0.0035697***	0.0089734***
	(0.0002623)	(0.0013913)	(0.0004857)	(0.0006117)	(0.0006988)	(0.0006221)	(0.0005482)	(0.0004859)
tas_17p5_20	0.0009156***	0.0318692***	0.0004603	0.0083624***	0.0138893***	0.0071929***	-0.0005061	0.0101204***
	(0.0001750)	(0.0012546)	(0.0005578)	(0.0003769)	(0.0006858)	(0.0005435)	(0.0004528)	(0.0004641)
tas_20_22p5	0.0013427***	0.0168075***	-0.0000663	0.0041365***	0.0059725***	0.0044163***	0.0017186***	0.0054607***
	(0.0001733)	(0.0010087)	(0.0003261)	(0.0002941)	(0.0005090)	(0.0004666)	(0.0003815)	(0.0003561)
tas_25_27p5	-0.0034302***	-0.0209518***	0.0001067	-0.0067520***	-0.0096412***	-0.0061303***	-0.0084942***	-0.0085681***
	(0.0001786)	(0.0009893)	(0.0002312)	(0.0003380)	(0.0005347)	(0.0004972)	(0.0004656)	(0.0003811)
tas_27p5_30	-0.0069894***	-0.0291915***	0.0002903	-0.0093059***	-0.0186805***	-0.0113753***	-0.0109311***	-0.0132110***
	(0.0002855)	(0.0014062)	(0.0003721)	(0.0004252)	(0.0006409)	(0.0004970)	(0.0006126)	(0.0004266)
tas_g30	-0.0091841***	-0.0404368***	0.0020532***	-0.0140636***	-0.0210601***	-0.0177628***	-0.0191294***	-0.0183610***
	(0.0004095)	(0.0012407)	(0.0003705)	(0.0003789)	(0.0008649)	(0.0005561)	(0.0005968)	(0.0003730)
p_5lo	-0.0078686***	-0.0243232***	-0.0009258**	-0.0098828***	-0.0159419***	-0.0117320***	-0.0031789***	-0.0109355***
	(0.0005292)	(0.0017548)	(0.0004168)	(0.0005973)	(0.0009124)	(0.0008551)	(0.0009504)	(0.0006264)
p_5_10	-0.0030719***	-0.0057338***	-0.0005488	-0.0017629***	-0.0052483***	-0.0035353***	0.0010602	-0.0026195***
	(0.0006067)	(0.0017815)	(0.0003842)	(0.0005230)	(0.0009161)	(0.0008469)	(0.0010986)	(0.0005843)
p_15up	0.0023167***	0.0046236**	0.0009314*	0.0059325***	0.0099634***	0.0059294***	-0.0045659***	0.0038275***
	(0.0006275)	(0.0022096)	(0.0005152)	(0.0006384)	(0.0010869)	(0.0010365)	(0.0012310)	(0.0007284)
Observations	51,089	43,426	44,889	44,864	45,853	41,273	43,901	264,206
R2	0.7364114	0.7868651	0.6892234	0.8764847	0.8492243	0.7262140	0.6110207	0.7406698
Adjusted R2	0.7263854	0.7797858	0.6789455	0.8724944	0.8442760	0.7164230	0.5983207	0.7322723
Residual S.E.	0.216176(df=49216)	0.7957024(df=42029)	0.257544(df=43451)	0.23659(df=43459)	0.461562(df=44395)	0.371543(df=39847)	0.462408(df=42512)	0.566539(df=255918)

Note: *p<0.1; **p<0.05; ***p<0.01

ii. Dependent variable: log. Yield (Soybeans)

	USDA	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS	Multi-GGCM
tas_7p5lo	-0.0020043	0.0215894***	0.0024596*	0.0069627***	0.0419370***	0.0165029***	-0.0394296***	0.0102104***
	(0.0018295)	(0.0007812)	(0.0014743)	(0.0015333)	(0.0018472)	(0.0024836)	(0.0024415)	(0.0008192)
tas_7p5_10	-0.0003167	0.0094857***	-0.0069068***	0.0106153***	0.0054723***	0.0056684*	-0.0251333***	0.0009813
	(0.0010250)	(0.0007340)	(0.0015411)	(0.0012686)	(0.0017655)	(0.0029096)	(0.0017582)	(0.0008585)
tas_10_12p5	-0.0022235***	0.0146659***	-0.0068461***	0.0194705***	0.0111955***	-0.0092969***	-0.0265141***	0.0030275***
	(0.0006599)	(0.0007808)	(0.0012079)	(0.0008921)	(0.0015397)	(0.0021949)	(0.0015251)	(0.0008491)
tas_12p5_15	-0.0038558***	0.0083743***	-0.0050278***	0.0180865***	0.0137083***	-0.0143583***	-0.0201772***	0.0020978***
	(0.0004538)	(0.0004157)	(0.0007662)	(0.0006835)	(0.0009947)	(0.0013297)	(0.0009570)	(0.0004434)
tas_15_17p5	-0.0044738***	0.0084544***	-0.0028182***	0.0177465***	0.0198297***	-0.0060135***	-0.0065699***	0.0071700***
	(0.0002892)	(0.0003687)	(0.0006305)	(0.0004995)	(0.0008519)	(0.0013938)	(0.0007196)	(0.0003976)
tas_17p5_20	-0.0008074***	0.0103046***	-0.0007707**	0.0149984***	0.0179208***	0.0017002	0.0015366**	0.0091276***
	(0.0001783)	(0.0003055)	(0.0003701)	(0.0004451)	(0.0007614)	(0.0010537)	(0.0006518)	(0.0003502)
tas_20_22p5	0.0010007***	0.0069777***	-0.0010801***	0.0089160***	0.0117341***	0.0029999***	0.0012337**	0.0057479***
	(0.0001792)	(0.0003059)	(0.0002659)	(0.0003708)	(0.0005912)	(0.0008024)	(0.0005658)	(0.0002931)
tas_25_27p5	-0.0021464***	-0.0101982***	0.0000124	-0.0095951***	-0.0105547***	-0.0186935***	-0.0058658***	-0.0091249***
	(0.0001498)	(0.0003279)	(0.0001497)	(0.0004097)	(0.0006123)	(0.0008785)	(0.0005620)	(0.0003160)
tas_27p5_30	-0.0055394***	-0.0140244***	0.0004097**	-0.0135460***	-0.0172370***	-0.0277973***	-0.0076833***	-0.0131326***
	(0.0002096)	(0.0004302)	(0.0001847)	(0.0005978)	(0.0007450)	(0.0010783)	(0.0006546)	(0.0003290)
tas_g30	-0.0112451***	-0.0179849***	0.0010055***	-0.0217638***	-0.0205560***	-0.0408052***	-0.0173798***	-0.0176961***
	(0.0003987)	(0.0003963)	(0.0002760)	(0.0005838)	(0.0007441)	(0.0010999)	(0.0005459)	(0.0003451)
p_5lo	-0.0053936***	-0.0054759***	-0.0000242	-0.0063017***	-0.0073103***	-0.0545069***	0.0085216***	-0.0103624***
	(0.0004981)	(0.0005561)	(0.0006078)	(0.0007053)	(0.0012004)	(0.0018881)	(0.0013132)	(0.0006158)
p_5_10	-0.0016208***	-0.0016163***	0.0001564	-0.0017874**	-0.0062830***	-0.0144566***	0.0071557***	-0.0026706***
	(0.0005983)	(0.0005957)	(0.0006964)	(0.0007600)	(0.0013410)	(0.0019881)	(0.0015483)	(0.0006256)
p_15up	0.0022689***	0.0013255*	0.0009891	0.0128435***	0.0052602***	0.0170603***	0.0003457	0.0064598***
	(0.0006124)	(0.0006874)	(0.0007347)	(0.0008430)	(0.0014521)	(0.0023110)	(0.0016643)	(0.0007254)
Observations	41,510	34,908	28,279	36,332	37,019	34,367	33,677	204,582
R2	0.7018623	0.8304705	0.3707549	0.8264802	0.6715554	0.6934535	0.4087950	0.6141671
Adjusted R2	0.6895284	0.8247679	0.3465132	0.8208337	0.6606558	0.6826970	0.3883435	0.6013249
Residual S.E.	0.180927(df=39860)	0.230332(df=33771)	0.238308(df=27229)	0.292614(df=35186)	0.552259(df=35829)	0.782370(df=33201)	0.493503(df=32550)	0.590675(df=197991)

Note: *p<0.1; **p<0.05; ***p<0.01

iii. Dependent variable: log. Yield (Wheat)

	USDA	GEPIC	GAEZ-IMAGE	LPJ-GUESS	LPJmL	pDSSAT	PEGASUS	Multi-GGCM
tas_7p5lo	0.0076782***	0.0108508***	-0.0017360***	0.0258254***	0.0238067***	0.0169180***	-0.0022206*	0.0124392***
	(0.0015201)	(0.0007761)	(0.0006249)	(0.0006505)	(0.0014434)	(0.0015442)	(0.0012516)	(0.0005245)
tas_7p5_10	0.0068445***	0.0085632***	-0.0001466	0.0153312***	0.0180041***	0.0131653***	-0.0070583***	0.0080738***
	(0.0008948)	(0.0008460)	(0.0006709)	(0.0005764)	(0.0010328)	(0.0012163)	(0.0011550)	(0.0004702)
tas_10_12p5	0.0040388***	0.0137081***	-0.0020298**	0.0154553***	0.0150130***	0.0097757***	0.0058075***	0.0096794***
	(0.0006349)	(0.0006415)	(0.0008767)	(0.0004852)	(0.0005973)	(0.0011735)	(0.0008803)	(0.0003383)
tas_12p5_15	0.0041869***	0.0131250***	-0.0029618***	0.0140379***	0.0079762***	0.0135785***	0.0055747***	0.0085229***
	(0.0004188)	(0.0005313)	(0.0006803)	(0.0004136)	(0.0004535)	(0.0009281)	(0.0008104)	(0.0003005)
tas_15_17p5	0.0035122***	0.0128343***	-0.0003093	0.0115348***	0.0060941***	0.0107958***	0.0078107***	0.0080490***
	(0.0003223)	(0.0004798)	(0.0004888)	(0.0003725)	(0.0004097)	(0.0007988)	(0.0007311)	(0.0002868)
tas_17p5_20	0.0019380***	0.0085079***	0.0003924	0.0081175***	0.0045173***	0.0057965***	0.0089977***	0.0059748***
	(0.0002147)	(0.0003900)	(0.0004285)	(0.0003685)	(0.0003027)	(0.0008266)	(0.0006991)	(0.0002703)
tas_20_22p5	0.0007430***	0.0024996***	0.0001386	0.0028780***	0.0018667***	0.0015289*	0.0041844***	0.0021834***
	(0.0001712)	(0.0003194)	(0.0003042)	(0.0003115)	(0.0002796)	(0.0007946)	(0.0006052)	(0.0002372)
tas_25_27p5	-0.0007894***	-0.0003115	-0.0004048	-0.0027728***	-0.0012307***	0.0004751	-0.0128928***	-0.0029281***
	(0.0001538)	(0.0003003)	(0.0002726)	(0.0003744)	(0.0002685)	(0.0006626)	(0.0006403)	(0.0002486)
tas_27p5_30	-0.0007809***	-0.0005185	0.0001230	-0.0054077***	-0.0001344	-0.0031565***	-0.0192035***	-0.0047206***
	(0.0002100)	(0.0003180)	(0.0003158)	(0.0005311)	(0.0003294)	(0.0006731)	(0.0008093)	(0.0002697)
tas_g30	-0.0020952***	-0.0016770***	-0.0004210	-0.0114572***	-0.0056180***	-0.0055151***	-0.0251358***	-0.0083518***
	(0.0002941)	(0.0004226)	(0.0003090)	(0.0005358)	(0.0003842)	(0.0007326)	(0.0008163)	(0.0003035)
p_5lo	0.0039684***	-0.0014137**	0.0000899	-0.0056861***	0.0024935***	-0.0003045	0.0043478***	-0.0001153
	(0.0005209)	(0.0006488)	(0.0003645)	(0.0005644)	(0.0006043)	(0.0013119)	(0.0011592)	(0.0004151)
p_5_10	0.0012606**	0.0008180	0.0000058	-0.0011378*	0.0027146***	0.0010022	0.0036213***	0.0011557**
	(0.0006058)	(0.0007201)	(0.0003469)	(0.0006014)	(0.0007624)	(0.0015896)	(0.0013033)	(0.0004561)
p_15up	-0.0047483***	-0.0008077	0.0004688	0.0037713***	-0.0025329***	0.0012860	-0.0061317***	-0.0007149
	(0.0006256)	(0.0007622)	(0.0004195)	(0.0006691)	(0.0007555)	(0.0017229)	(0.0014108)	(0.0004917)
Observations	42,752	41,056	42,463	42,305	43,643	41,787	41,281	252,535
R2	0.6803230	0.6347984	0.7157260	0.8943015	0.7605653	0.3618249	0.5995471	0.7003024
Adjusted R2	0.6671089	0.6227613	0.7064270	0.8908785	0.7528113	0.3405676	0.5865052	0.6907759
Residual S.E.	0.199791(df=41054)	0.293276(df=39745)	0.239063(df=41117)	0.246312(df=40977)	0.317761(df=42273)	0.569432(df=40439)	0.553398(df=39978)	0.429219(df=244754)

Note: *p<0.1; **p<0.05; ***p<0.01

4. Variation in historical observed (USDA) and simulated (GGCMs) yields empirically attributed to weather (T and P bins)

The adjusted R^2 (in table 3C) derived from the regression specification (equation 4.1 in Chapter 4) track how much of the cross-section/time-series variation in yields is explained by not only the predictor variables (T and P bins), but also by the grid-cell fixed effects (μ_i) and the time effects (τ_t). In order to gauge how much, on average, the weather variables (T and P bins) explain the cross-section/time-series variation in yields, table 4C summarizes the adjusted R^2 by stripping out the influencing effects of the idiosyncratic unobserved shocks (μ_i and τ_t). These are obtained directly from R LFE package 'Projected Model $adj - R^2$'

Table 4C. Percentage of variation explained by the covariates (T and P)

GGCM	Maize	Wheat	Soybeans
GEPIC	37%	9%	50%
GAEZ-IMAGE	0%	0%	0%
LPJ-GUESS	44%	33%	51%
LPJmL	33%	5%	23%
pDSSAT	26%	1%	33%
PEGASUS	16%	16%	11%
Multi-GGCM	15%	3%	12%
USDA	8%	0%	11%

Across regions (here counties), GGCMs by and large have homogenous setup in non-weather parameters such as management practices, soil types, sowing dates, crop varieties etc. The proportion of variance in yields explained by weather is therefore by and large higher for GGCMs than in the empirical (USDA) model (table 4C). For the latter, a higher proportion of variance is explained by the heterogeneous non-weather parameters.

Focusing on GGCMs (table 4C), GAEZ-IMAGE is not well captured by the regression specification (equation 4.1 of Chapter 4). Furthermore, the R^2 is generally low in wheat for all GGCMs, except for LPJ-GUESS and PEGASUS. However, by and large the regression specification is able to explain between one-fifth to half of the variance in each of the six GGCMs, and slightly lower in the Multi-Model regression specification.

For the empirical model (USDA), the yield variance left unexplained by the regression specification is relatively larger to that of the GGCMs. The negligible variance explained by T and P for USDA wheat could be partly explained by the fact that across U.S. counties, wheat is grown in multiple seasons outside the months $MJJA$ (CGS defined in this study), and is largely irrigated. For maize and soybeans, the R^2 is marginally higher than wheat, but yet lower to the corresponding GGCMs' crops specifications. The unexplained variance in USDA maize and soybeans likely reflects those non-weather variables that are omitted from the analysis, and that are subsequently absorbed by the fixed and/or year effects.

5. Average change in exposure across HadGEM2-ES temperature and precipitation bins in RCP 8.5 scenario, relative to historical (1972-2004)

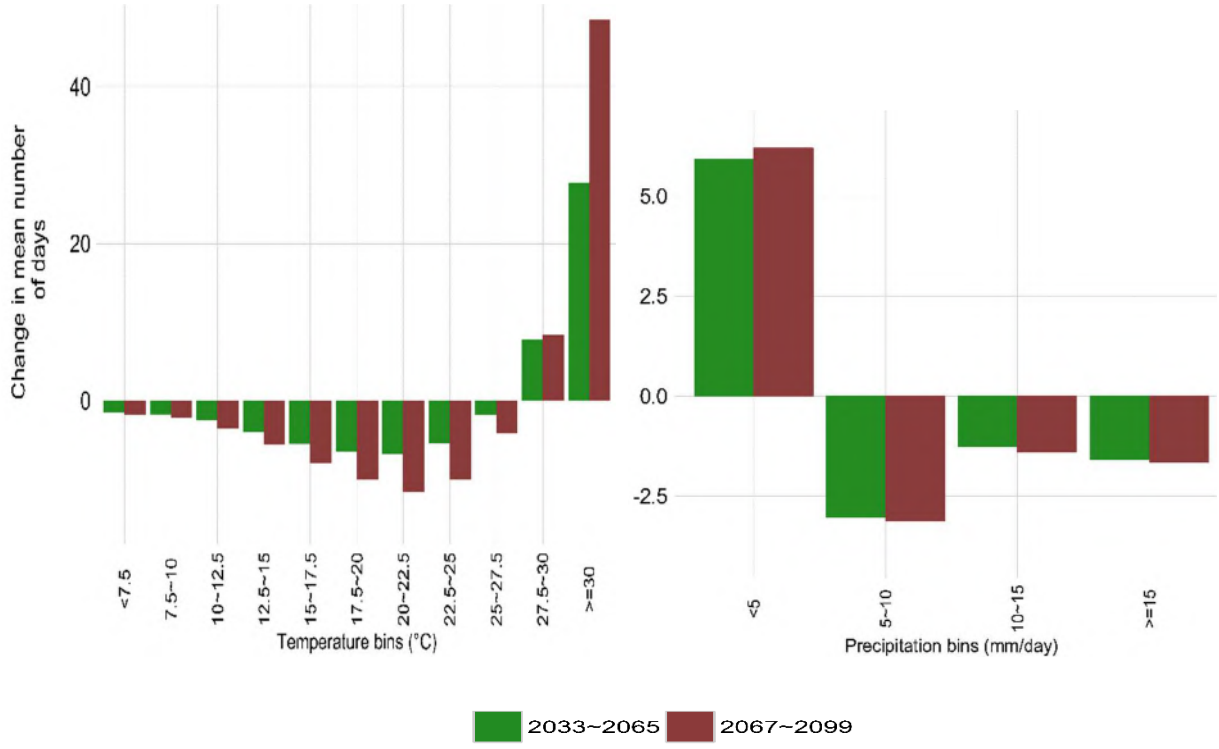


Figure 2C. Change in distribution of HadGEM2-ES temperature and precipitation bins, for two mean future periods in RCP 8.5 scenario.

The change in number of days is computed by first averaging the number of days (in each bin) in each USDA crop county, for the historical and future periods. The average number of days (in each bin) are then computed over all counties. The difference in distribution is calculated as future period – historical period.

Table 5C. Regression summary with Robust S.E. (in parentheses) for rainfed soybeans and wheat. The six regressions models (4.6 – 4.11) follow the specifications (eqs. 4.6 - 4.11) in Chapter 4.

Dependent variable: Difference Estimated Coefficients (GGCMs - USDA) as defined in eqs. (4.6–4.11) of Chapter 4, for Soybeans

	(4.6)	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)
Potential Yield	-0.012**	-0.012*	0.003	-0.010**	-0.020***	0.006
	(0.006)	(0.006)	(0.002)	(0.005)	(0.005)	(0.005)
Change in Cultivar	0.019***	0.020***	-0.0003	0.025***	0.036***	-0.001***
	(0.007)	(0.007)	(0.0003)	(0.005)	(0.006)	(0.000)
Dyn. planting Window	-0.006	-0.007		-0.016***	-0.017***	
	(0.003)	(0.004)		(0.003)	(0.003)	
Heat stress	-0.021***	-0.024***	0.007	-0.021***	-0.034***	-0.001***
	(0.007)	(0.007)	(0.004)	(0.007)	(0.008)	(0.000)
GGCM Calibration (Site)	-0.008	-0.006		0.003	-0.006	
	(0.005)	(0.004)		(0.004)	(0.005)	
I(hi_t * Pot_Yield)				0.001	0.029***	
				(0.006)	(0.004)	
I(hi_t * Cultivar)				-0.025***	-0.053***	
				(0.007)	(0.004)	
I(hi_t * Plant_window_Dyn)				0.032***	0.032***	
				(0.004)	(0.004)	
I(hi_t * H_stress)				0.002	0.032***	
				(0.008)	(0.006)	
I(hi_t * Calib_Site)				-0.026***		
				(0.005)		
I(lo_p * Pot_Yield)				-0.018		-0.004
				(0.015)		(0.005)
I(lo_p * Cultivar)				0.002		0.001***
				(0.015)		(0.00004)
I(lo_p * Plant_window_Dyn)				0.020***		
				(0.003)		
I(lo_p * H_stress)				0.002		0.012***
				(0.016)		(0.002)
I(lo_p * Calib_Site)				-0.034**		
				(0.015)		
Adjusted F Statistic and	4.289***	4.885***	1.553	42.577***	24.052***	2.325*
Degress of Freedom	(df = 4;77)	(df = 4;59)	(df = 2;17)	(df = 14;77)	(df = 8;59)	(df = 5;17)
Observations	78	60	18	78	60	18
Adjusted R^2	0.101	0.147	-0.132	0.421	0.310	-0.355
Residual Std. Error	0.014 (df = 73)	0.014 (df = 55)	0.014 (df = 15)	0.011 (df = 63)	0.012 (df = 51)	0.016 (df = 12)

Note: *p<0.1; **p<0.05; ***p<0.01

Dependent variable: Difference Estimated Coefficients (GGCMs - USDA) as defined in eqs. (4.6 – 4.11) of Chapter 4, for Wheat

	(4.6)	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)
Potential Yield	-0.003	-0.004	0.0001	-0.003	-0.007***	0.003*
	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.001)
Change in Cultivar	0.006*	0.009**	-0.001	0.012***	0.016***	0.004***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.000)
Dyn. planting Window	-0.006**	-0.008***		-0.013***	-0.014***	
	(0.003)	(0.003)		(0.002)	(0.002)	
Heat stress	-0.005	-0.006	0.001	-0.0005	-0.003	-0.005***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.000)
GGCM Calibration (Site)	0.003**	0.004**		0.006***	0.004**	
	(0.001)	(0.001)		(0.001)	(0.001)	
I(hi_t * Pot_Yield)				0.002	0.012***	
				(0.003)	(0.002)	
I(hi_t * Cultivar)				-0.017***	-0.026***	
				(0.004)	(0.003)	
I(hi_t * Plant_window_Dyn)				0.019***	0.021***	
				(0.003)	(0.003)	
I(hi_t * H_stress)				-0.019***	-0.012**	
				(0.004)	(0.004)	
I(hi_t * Calib_Site)				-0.008***		
				(0.002)		
I(lo_p * Pot_Yield)				0.001		-0.004
				(0.004)		(0.003)
I(lo_p * Cultivar)				-0.017***		-0.007***
				(0.005)		(0.002)
I(lo_p * Plant_window_Dyn)				0.016***		
				(0.004)		
I(lo_p * H_stress)				0.002		0.010***
				(0.004)		(0.002)
I(lo_p * Calib_Site)				-0.008***		
				(0.002)		
Adjusted F Statistic and	4.289***	4.885***	1.553	42.577***	24.052***	2.325*
Degress of Freedom	(df = 4;77)	(df = 4;59)	(df = 2;17)	(df = 14;77)	(df = 8;59)	(df = 5;17)
Observations	78	60	18	78	60	18
Adjusted R^2	0.128	0.210	-0.186	0.532	0.527	0.433
Residual Std. Error	0.007 (df = 73)	0.007 (df = 55)	0.005 (df = 15)	0.005 (df = 63)	0.005 (df = 51)	0.003 (df = 12)

Note: p<0.1; **p<0.05; ***p<0.01

Bibliography

Anderson D B 1936 Relative Humidity or Vapor Pressure Deficit *Ecology* **17** 277–82 Online: http://dx.doi.org/10.2307/1931468

Auffhammer M, Hsiang S M, Schlenker W and Sobel A 2013 Using Weather Data and Climate Model Output in Economic Analyses of Climate Change *Rev. Environ. Econ. Policy* **7** 181–98 Online: http://reep.oxfordjournals.org/content/7/2/181.abstract

Bassu S, Brisson N, Durand J L, Boote K, Lizaso J, Jones J W, Rosenzweig C, Ruane A C, Adam M, Baron C, Basso B, Biernath C, Boogaard H, Conijn S, Corbeels M, Deryng D, De Sanctis G, Gayler S, Grassini P, Hatfield J, Hoek S, Izaurralde C, Jongschaap R, Kemanian A R, Kersebaum K C, Kim S H, Kumar N S, Makowski D, M??ller C, Nendel C, Priesack E, Pravia M V, Sau F, Shcherbak I, Tao F, Teixeira E, Timlin D and Waha K 2014 How do various maize crop models vary in their responses to climate change factors? *Glob. Chang. Biol.* **20** 2301–20

Baylis K, Paulson N D and Piras G 2011 Spatial Approaches to Panel Data in Agricultural Economics : A Climate Change Application **3** 325–38

Blanc É 2016 *Statistical Emulators of Maize , Rice , Soybean and Wheat Yields from Global Gridded Crop Models* Online: http://globalchange.mit.edu/

Blanc E and Sultan B 2015 Agricultural and Forest Meteorology Emulating maize yields from global gridded crop models using statistical estimates *Agric. For. Meteorol.* **214–215** 134–47 Online: http://dx.doi.org/10.1016/j.agrformet.2015.08.256

BONDEAU A, SMITH P C, ZAEHLE S, SCHAPHOFF S, LUCHT W, CRAMER W, GERTEN D, LOTZE-CAMPEN H, MÜLLER C, REICHSTEIN M and SMITH B 2007 Modelling the role of agriculture for the 20th century global terrestrial carbon balance *Glob. Chang. Biol.* **13** 679–706 Online: http://doi.wiley.com/10.1111/j.1365-2486.2006.01305.x

Burke M and Emerick K 2015 Adaptation to Climate Change: Evidence from US Agriculture *Am. Econ. J. Econ. Policy* **0**

Cai R, Yu D and Oppenheimer M 2012 Estimating the Effects of Weather Variations on Corn Yields using Geographically Weighted Panel Regression *Agric. Appl. Econ. Assoc. 2012 AAEA Annu. Meet.* **39** 230–52

Challinor a. J and Wheeler T R 2008 Crop yield reduction in the tropics under climate change: Processes and uncertainties *Agric. For. Meteorol.* **148** 343–56

De Cian E, Lanzi E and Roson R 2013 Seasonal temperature variations and energy demand *Clim. Change* **116** 805–25

Conley T G 1999 GMM estimation with cross sectional dependence *J. Econom.* **92** 1–45

Deryng D, Sacks W J J, Barford C C C and Ramankutty N 2011 Simulating the effects of climate and agricultural management practices on global crop yield *Global Biogeochem. Cycles* **25**n/a-n/a Online: http://www.agu.org/pubs/crossref/2011/2009GB003765.shtml

Deschênes O and Greenstone M 2007 The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather *Am. Econ. Rev.* **97** 354–85 Online: http://www.aeaweb.org/articles?id=10.1257/aer.97.1.354

Deschênes and Greenstone 2012 The Economic Impacts of Climate Change : Evidence from in Weather Output and Random Fluctuations Agricultural *Am. Econ. Rev.* **97** 354–85

Egbendewe-Mondzozo A, Musumba M, McCarl B A and Wu X 2011 Climate change and vector-borne diseases: An economic impact analysis of malaria in Africa *Int. J. Environ. Res. Public Health* **8** 913–30

Elhorst J P 2010 Applied Spatial Econometrics: Raising the Bar *Spat. Econ. Anal.* **5** 9–28

Elliott J, Kelly D, Best N, Wilde M, Glotter M and Foster I 2013 The Parallel System for Integrating Impact Models and Sectors (pSIMS) *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery* XSEDE '13 (New York, NY, USA: ACM) p 21:1--21:8 Online: http://doi.acm.org/10.1145/2484762.2484814

Elliott J, Müller C, Deryng D, Chryssanthacopoulos J and Boote K J 2014 The Global Gridded Crop Model intercomparison : data and modeling protocols for Phase 1 (v1 . 0) 4383–427

Elliott J, Müller C, Deryng D, Chryssanthacopoulos J, Boote K J, Büchner M, Foster I, Glotter M, Heinke J, Iizumi T, Izaurralde R C, Mueller N D, Ray D K, Rosenzweig C, Ruane A C and Sheffield J 2015 The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0) *Geosci. Model Dev.* **8** 261–77 Online: http://www.geosci-model-dev.net/8/261/2015/

Frieler K, Levermann A, Elliott J, Heinke J, Arneth A, Bierkens M F P, Ciais P, Clark D B, Deryng D, Döll P, Falloon P, Fekete B, Folberth C, Friend A D, Gellhorn C, Gosling S N, Haddeland I, Khabarov N, Lomas M, Masaki Y, Nishina K, Neumann K, Oki T, Pavlick R, Ruane A C, Schmid E, Schmitz C, Stacke T, Stehfest E, Tang Q, Wisser D, Huber V, Piontek F, Warszawski L, Schewe J, Lotze-Campen H and Schellnhuber H J 2015 A framework for the cross-sectoral integration of multi-model impact projections: Land use decisions under climate impacts uncertainties *Earth Syst. Dyn.* **6** 447–60

Gaure S 2013 lfe: Linear Group Fixed Effects *R J.* **5** 104–117 Online: http://journal.r-project.org/archive/2013-2/gaure.pdf

Hatfield J L, Boote K J, Kimball B A, Ziska L H, Izaurralde R C, Ort D, Thomson A M and Wolfe D 2011 Climate impacts on agriculture: Implications for crop production *Agron. J.* **103** 351–70

Hempel S, Frieler K, Warszawski L, Schewe J and Piontek F 2013 A trend-preserving bias correction – The ISI-MIP approach *Earth Syst. Dyn.* **4** 219–36

Holzkämper A, Calanca P and Fuhrer J 2012 Statistical crop models: predicting the effects of temperature and precipitation changes *Clim. Res.* **51** 11–21 Online: http://www.int-res.com/abstracts/cr/v51/n1/p11-21/

Hsiang S M 2010 Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proc. Natl. Acad. Sci. U. S. A.* **107** 15367–72 Online: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2932627&tool=pmcentrez&rendertype=abstract

Iglesias A, Rosenzweig C and Pereira D 2000 Agricultural impacts of climate change in Spain : developing tools for a spatial analysis **10** 69–80

Iizumi T, Yokozawa M, Sakurai G, Travasso M I, Romanenkov V, Oettli P, Newby T, Ishigooka Y and Furuya J 2014 Historical changes in global yields: major cereal and legume crops from 1982 to 2006

Glob. Ecol. Biogeogr. **23** 346–57 Online: http://doi.wiley.com/10.1111/geb.12120

IPCC 2014 *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J.* (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press)

Jones C D, Hughes J K, Bellouin N, Hardiman S C, Jones G S, Knight J, Liddicoat S, O'Connor F M, Andres R J, Bell C, Boo K O, Bozzo A, Butchart N, Cadule P, Corbin K D, Doutriaux-Boucher M, Friedlingstein P, Gornall J, Gray L, Halloran P R, Hurtt G, Ingram W J, Lamarque J F, Law R M, Meinshausen M, Osprey S, Palin E J, Parsons Chini L, Raddatz T, Sanderson M G, Sellar A A, Schurer A, Valdes P, Wood N, Woodward S, Yoshioka M and Zerroukat M 2011 The HadGEM2-ES implementation of CMIP5 centennial simulations *Geosci. Model Dev.* **4** 543–70

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 and Ritchie J T 2003 The {DSSAT} cropping system model *Eur. J. Agron.* **18** 235–65 Online: http://www.sciencedirect.com/science/article/pii/S1161030102001077

Lee H L, Hertel T W, Sohngen B and Ramankutty N 2005 Towards an integrated land use data base for assessing the potential for greenhouse gas mitigation *GTAP Tech. Pap. No. 25* Online: https://www.gtap.agecon.purdue.edu/resources/download/2375.pdf

LeSage J P 2008 An Introduction to Spatial Econometrics *Rev. d'économie Ind.* **123** 19–44 Online: http://rei.revues.org/3887

LeSage J P and Pace R K 2014 The Biggest Myth in Spatial Econometrics *Econometrics* **2** 217–49

Liu J, Williams J R, Zehnder A J B and Yang H 2007 GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale *Agric. Syst.* **94** 478–93

Lobell D B and Burke M B 2010 On the use of statistical models to predict crop yield responses to climate change *Agric. For. Meteorol.* Online: http://dx.doi.org/10.1016/j.agrformet.2010.07.008

Lobell D B and Field C B 2007 Global scale climate–crop yield relationships and the impacts of recent warming *Environ. Res. Lett.* **2** 14002

Lobell D B, Hammer G L, McLean G, Messina C, Roberts M J and Schlenker W 2013 The critical role of extreme heat for maize production in the United States *Nat. Clim. Chang.* **3** 497–501 Online: http://www.nature.com/doifinder/10.1038/nclimate1832%5Cnhttp://www.nature.com/nclimate/journal/v3/n5/full/nclimate1832.html?WT.ec_id=NCLIMATE-201305

Lobell D B, Schlenker W and Costa-Roberts J 2011 Climate trends and global crop production since 1980. *Science* **333** 616–20

Lobell D B, Sibley A and Ivan Ortiz-Monasterio J 2012 Extreme heat effects on wheat senescence in India *Nat. Clim. Chang.* **2** 186–9 Online: http://dx.doi.org/10.1038/nclimate1356

Morell F J, Yang H S, Cassman K G, Wart J Van, Elmore R W, Licht M, Coulter J A, Ciampitti I A, Pittelkow C M, Brouder S M, Thomison P, Lauer J, Graham C, Massey R and Grassini P 2016 Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? *F. Crop. Res.* **192** 1–12 Online: http://dx.doi.org/10.1016/j.fcr.2016.04.004

Moss R H, Edmonds J A, Hibbard K A, Manning M R, Rose S K, van Vuuren D P, Carter T R, Emori S, Kainuma M, Kram T, Meehl G A, Mitchell J F B, Nakicenovic N, Riahi K, Smith S J, Stouffer R J,

Thomson A M, Weyant J P and Wilbanks T J 2010 The next generation of scenarios for climate change research and assessment *Nature* **463** 747–56 Online: http://dx.doi.org/10.1038/nature08823

Müller C, Elliott J, Chryssanthacopoulos J, Arneth A, Balkovic J, Ciais P, Deryng D, Folberth C, Glotter M, Hoek S, Iizumi T, Izaurralde R C, Jones C, Khabarov N, Lawrence P, Liu W, Olin S, Pugh T A M, Ray D, Reddy A, Rosenzweig C, Ruane A C, Sakurai G, Schmid E, Skalsky R, Song C X, Wang X, de Wit A and Yang H 2016 Global Gridded Crop Model evaluation: benchmarking, skills, deficiencies and implications *Geosci. Model Dev. Discuss.* 1–39 Online: http://www.geosci-model-dev-discuss.net/gmd-2016-207/

Müller C, Elliott J, Chryssanthacopoulos J, Deryng D, Folberth C, Pugh T A M and Schmid E 2015 Implications of climate mitigation for future agricultural production *Environ. Res. Lett.* **10** 125004 Online: http://stacks.iop.org/1748-9326/10/i=12/a=125004

Nelson G C, van der Mensbrugghe D, Ahammad H, Blanc E, Calvin K, Hasegawa T, Havlik P, Heyhoe E, Kyle P, Lotze-Campen H, von Lampe M, Mason d'Croz D, van Meijl H, Müller C, Reilly J, Robertson R, Sands R D, Schmitz C, Tabeau A, Takahashi K, Valin H and Willenbockel D 2014a Agriculture and climate change in global scenarios: Why don't the models agree *Agric. Econ. (United Kingdom)* **45** 85–101

Nelson G C, Valin H, Sands R D, Havlík P, Ahammad H, Deryng D, Elliott J, Fujimori S, Hasegawa T, Heyhoe E, Kyle P, Von Lampe M, Lotze-Campen H, Mason d'Croz D, van Meijl H, van der Mensbrugghe D, Müller C, Popp A, Robertson R, Robinson S, Schmid E, Schmitz C, Tabeau A and Willenbockel D 2014b Climate change effects on agriculture: economic responses to biophysical shocks. *Proc. Natl. Acad. Sci. U. S. A.* **111** 3274–9 Online: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3948295&tool=pmcentrez&rendertype=abstract

Ortiz-Bobea A 2013 Is Weather Really Additive in Agricultural Production?

Oyebamiji O K, Edwards N R, Holden P B, Garthwaite P H, Schaphoff S and Gerten D 2015 Emulating global climate change impacts on crop yields *Stat. Model.* Online: http://smj.sagepub.com/content/early/2015/01/18/1471082X14568248.abstract

Parry M L, Rosenzweig C, Iglesias a., Livermore M and Fischer G 2004 Effects of climate change on global food production under SRES emissions and socio-economic scenarios *Glob. Environ. Chang.* **14** 53–67

Parry M, Rosenzweig C, Iglesias A, Fischer K and Livermore M 1999 Climate change and world food security : a new assessment **9** 51–67

Porter J R, Xie L, Challinor A J, Cochrane K, Howden S M, Iqbal M M, Lobell D B and Travasso M I 2014 Food security and food production systems *Clim. Chang. 2014 Impacts, Adapt. Vulnerability. Part A Glob. Sect. Asp. Contrib. Work. Gr. II to Fifth Assess. Rep. Intergov. Panel Clim. Chang.* 485–533

Portmann F T, Siebert S and Döll P 2010 MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling *Global Biogeochem. Cycles* **24** 1–24

Ray D K, Ramankutty N, Mueller N D, West P C and Foley J A 2012 Recent patterns of crop yield growth and stagnation *Nat Commun* **3** 1293 Online: http://dx.doi.org/10.1038/ncomms2296

Rodell M, Houser P R, Jambor U, Gottschalck J, Mitchell K, Meng C-J, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin J K, Walker J P, Lohmann D and Toll D 2004 The Global Land Data Assimilation System *Bull. Am. Meteorol. Soc.* **85** 381–94

Rosenzweig C, Elliott J, Deryng D, Ruane A C, Müller C, Arneth A, Boote K J, Folberth C, Glotter M, Khabarov N, Neumann K, Piontek F, Pugh T a M, Schmid E, Stehfest E, Yang H and Jones J W 2014 Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl. Acad. Sci. U. S. A.* **111** 3268–73 Online: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3948251&tool=pmcentrez&rendertype=abstract

Rosenzweig C, Jones J W, Hatfield J L, Ruane A C, Boote K J, Thorburn P, Antle J M, Nelson G C, Porter C, Janssen S, Asseng S, Basso B, Ewert F, Wallach D, Baigorria G and Winter J M 2013 The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies *Agric. For. Meteorol.* **170** 166–82 Online: http://dx.doi.org/10.1016/j.agrformet.2012.09.011

Schlenker W and Lobell D B 2010 Robust negative impacts of climate change on African agriculture *Environ. Res. Lett.* **5** 1–8 Online: http://stacks.iop.org/1748-9326/5/i=1/a=014010

Schlenker W and Roberts M J 2006 Estimating the impact of climate change on crop yields: The importance of nonlinear temperature effects Online: http://www.nber.org/papers/w13799

Schlenker W and Roberts M J 2009 Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc. Natl. Acad. Sci. U. S. A.* **106** 15594–8

Seager R, Hooks A, Williams A P, Cook B, Nakamura J and Henderson N 2015 Climatology, Variability, and Trends in the U.S. Vapor Pressure Deficit, an Important Fire-Related Meteorological Quantity *J. Appl. Meteorol. Climatol.* **54** 1121–41 Online: http://dx.doi.org/10.1175/JAMC-D-14-0321.1

Sitch S, Smith B, Prentice I C, Arneth A, Bondeau A, Cramer W, Kaplan J O, Levis S, Lucht W, Sykes M T, Thonicke K and Venevsky S 2003 Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model *Glob. Chang. Biol.* **9** 161–85 Online: http://doi.wiley.com/10.1046/j.1365-2486.2003.00569.x

Tack J, Barkley A and Nalley L L 2015 Effect of warming temperatures on US wheat yields *Proc. Natl. Acad. Sci.* **112** 6931–6 Online: http://www.pnas.org/content/112/22/6931.abstract%5Cnhttp://www.pnas.org/content/112/22/6931.full.pdf

Urban D W, Sheffield J and 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 *Environ. Res. Lett.* **10** 45003 Online: http://stacks.iop.org/1748-9326/10/i=4/a=045003?key=crossref.58e5b11cc0b47418b44a35b4a231f5fc

Vicente-Serrano S M, Beguer??a S and L??pez-Moreno J I 2010 A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index *J. Clim.* **23** 1696–718

van Vuuren D P, van Ruijven B, Hoogwijk M M, Isaac M and de Vries H J M 2006 *TIMER 2: Model description and application*

Wang E, Smith C J, Bond W J and Verburg K 2004 Estimations of vapour pressure deficit and crop water demand in APSIM and their implications for prediction of crop yield, water use, and deep drainage *Aust. J. Agric. Res.* **55** 1227–40 Online: http://dx.doi.org/10.1071/AR03216

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2013 The Inter-Sectoral Impact Model Intercomparison Project (ISI – MIP): Project framework 1–5

Watson J, Challinor A J, Fricker T E and Ferro C A T 2015 Comparing the effects of calibration and climate errors on a statistical crop model and a process-based crop model *Clim. Change* **132** 93–109

Wing I S, Monier E, Stern A and Mundra A 2015 US major crops' uncertain climate change risks and greenhouse gas mitigation benefits *Environ. Res. Lett.* **10** 115002 Online: http://stacks.iop.org/1748-9326/10/i=11/a=115002?key=crossref.2d9b882bb362c58e29b86169963a44de