The ENSO Cycle and Predictability of US Crop Yields

Jan Alexander Kazimierz Orlowski

2017

The University of Sydney
Faculty of Agriculture and Environment

A thesis submitted in partial fulfilment of requirements for the degree of Doctor of Philosophy in the Faculty of Agriculture and Environment Department of Natural Resource Economics at the University of Sydney
Declaration

A thesis submitted in partial fulfilment of requirements for the degree of Doctor of Philosophy in the Faculty of Agriculture and the Environment Department of Natural Resource Economics at the University of Sydney

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Jan Orlowski

August 2017
Abstract

While the impacts of the El Nino Southern Oscillation (ENSO) are well documented on topics ranging from agricultural production to socio-economic factors, a closer consideration of key interaction terms in this complex relationship is pivotal for better understanding of future welfare impacts and as well as relevant policy implications. The focus this thesis is examining the ENSO link to staple crop production in the United States. The ENSO-production linkages are derived through threshold-like econometric methods, and further scrutinized through an analysis of the predictive content within this link.

Beginning with a review of the topic to date, as well as related topics, the ENSO phenomenon is examined from its roots by considering its measurement, observation, and identification. Research is discussed with regard to weather effects and crop yield influence globally with particular emphasis placed on the continental US. Furthermore, crop growth interactions with key weather variables, as pertaining to ENSO links with local weather, is included. Varied conclusions emerge, yet key aspects relevant to this research such as strong ENSO impacts on US weather and US agriculture reach a consensus. This thesis aims to contribute in the debated areas of narrow geographic influence of ENSO on weather, crops, and the how the two are linked.

Following an econometric approach, ENSO effect on corn and soybean yields in the US is estimated in particular by taking advantage of recent advances in modelling temperature influence on yield, through nonlinear parameterization of climate variables. Temperature variables are included as degree days, facilitating
the measurement of non-linearity in temperature interactions with yield recently shown to define key aspects of the relationship. Via the comparison of competing model specifications, across all major Corn and Soybean producing regions in the United States, the findings of the thesis suggest the ENSO link with crop yields manifests itself primarily via heating degree days. This is true for both El Nino and La Nina phases, which in turn produce spatially heterogeneous effects of different magnitudes and characteristic patterning.

In light of these results this thesis further extends previous literature by examining the effect of ENSO anomalies on agricultural production in an out-of-sample setting. Achieved through running competing model specifications in a form of leave one out analysis and comparing RMSFE forecast accuracy measures. Through this, the above relationship is further scrutinized and discussed from a forecasting standpoint identifying regions with greatest ENSO exposure as well as regions where ENSO offers significant predictive content.

First and foremost, the results present statistically significant ENSO influence over yields, predominantly through suppression of corn yield, driven through the ENSO link to temperatures above a predefined critical temperature threshold unique to each crop. Second, a diverse and unique picture is shown for each ENSO phase with regard to yield as well as key weather variables. Results fall in line with previous research, displaying dramatic sign change of ENSO influence as one moves from the eastern/coastal to the west of the Corn Belt, particularly during La Nina events, but also indicate that in majority of considered counties, the observed effect is not statistically significant, with an exception of two tightly defined geographical
clusters where ENSO is shown to offer considerable predictive content for corn yields. Nonetheless, resulting implications for optimal producer strategies can provide a powerful adaptive measure to anticipated/forecasted ENSO outcomes, predominantly planting date and crop mix. Key results prove value to such strategies, particularly in those regions where the pathway of ENSO influence for production is obvious, and statistically significant in a pseudo-forecasting environment. The thesis is completed with a discussion on ENSO use in risk management for agricultural producers as related to the scope of the results presenter here, predominantly addressing implications for crop insurance.
Acknowledgements

First and foremost, I would like to express my deep gratitude and special appreciation to my supervisor David Ubilava, who has been a remarkable mentor throughout my Ph.D. and prior studies. Through his expertise, encouragement, and support my Ph.D. experience has been productive and stimulating. Your advice on research and all related aspects has been priceless. I would also like to thank and convey my appreciation to my auxiliary supervisor Terry Griffin and Tihomir Ancev for their unfailing support and insight.

The University of Sydney has provided me with exceptional opportunity and a productive environment, in particular the Faculty of Agriculture and Environment and the School of Economics. I have been privileged to work with the staff, colleagues, and friends.

Finally, I am thankful to my family, Joanna, Marius, and Magda for their support and encouragement.
# TABLE OF CONTENTS

**DECLARATION** ............................................................................................................................................ ii

**ABSTRACT** .................................................................................................................................................... iii

**ACKNOWLEDGMENTS** ...................................................................................................................................... vi

**TABLE OF CONTENTS** ................................................................................................................................. vii

**LIST OF FIGURES** ........................................................................................................................................ ix

**LIST OF TABLES** ........................................................................................................................................... x

**LIST OF ABBREVIATIONS** ........................................................................................................................... xi

**CHAPTER 1.**

**INTRODUCTION** ............................................................................................................................................. 1

**CHAPTER 2.**

**ENSO MEASUREMENTS AND APPLICATIONS FOR AGRICULTURE – A REVIEW** ...... 12

2.1 **DEFINING ENSO** ...................................................................................................................................... 14

2.2 **CAPTURING ENSO LINK WITH LOCAL WEATHER AND PRODUCTION** ............................................ 19

2.3 **ENSO IMPACT ACROSS THE US** ......................................................................................................... 23

2.4 **ENSO AND PRODUCER DECISIONS** ..................................................................................................... 25

2.5 **IMPLICATIONS OF ENSO IMPACT** ....................................................................................................... 26

2.6 **ENSO INFLUENCE BEYOND THE US** .................................................................................................... 29
CHAPTER 3.
ESTIMATING THE EFFECT OF ENSO ON U.S. CROP PRODUCTION ........................................... 32

3.1 Conceptual framework ........................................................................................................ 33
3.2 Empirical framework ......................................................................................................... 41
3.3 Data .................................................................................................................................. 49
3.4 Results and interpretation ................................................................................................. 57
3.5 Conclusions ....................................................................................................................... 88

CHAPTER 4.
FORECASTING ENSO INFLUENCE ON PRODUCTION ......................................................... 92

4.1 Conceptual framework ........................................................................................................ 93
4.2 Empirical framework ......................................................................................................... 96
4.3 Data .................................................................................................................................. 104
4.4 Results and interpretation ................................................................................................. 106
4.5 Conclusions ....................................................................................................................... 116

CHAPTER 5.
IMPLICATIONS ....................................................................................................................... 118

CHAPTER 6.
CONCLUSION ......................................................................................................................... 126

REFERENCES ......................................................................................................................... 129
LIST OF FIGURES

FIGURE 3.1. Distribution of corn log-yields across counties and over time .......... 50
FIGURE 3.2. Distribution of soybean log-yields across counties and over time........ 51
FIGURE 3.3. Degree Day distributions........................................................................ 53
FIGURE 3.4. ENSO fluctuations and magnitudes ......................................................... 56
FIGURE 3.5. La Nina impact on corn yields................................................................. 59
FIGURE 3.6. El Nino impact on corn yields ................................................................ 61
FIGURE 3.7. La Nina impact on soybean yields........................................................... 63
FIGURE 3.8. El Nino impact on soybean yields ........................................................... 64
FIGURE 3.9. Histogram of corn and soybean estimates of direct ENSO impact ....... 65
FIGURE 3.10. El Nino and La Nina influence over Heating Degree Days - corn .... 68
FIGURE 3.11. El Nino and La Nina influence over Heating Degree Days - soybean ... 71
FIGURE 3.12. Heating Degree Days and Normal Degree Days for corn and soybeans... 73
FIGURE 3.13. La Nina and El Nino influence over Precipitation: corn and soybean . 75
FIGURE 3.14. La Nina and El Nino influence in corn producing counties (BP) ........... 77
FIGURE 3.15. La Nina and El Nino influence in soybean producing counties (BP) .... 78
FIGURE 3.16. Significant ENSO impact on corn and soybean yields ..................... 79
FIGURE 4.1. El Nino ~ yield relationship for corn and soybean .............................. 102
FIGURE 4.2. ENSO influence on yield, Heating Degree Days, Normal Degree Days for corn and soybeans ............................................................................................................. 104
FIGURE 4.3. RMSFE values for corn and soybean under alternate model specifications ................................................................................................................................. 107
FIGURE 4.4. Significant counties displaying forecast accuracy .............................. 109
LIST OF TABLES

Table 3.1. Basic model significant corn counties at $p<0.05$ ........................................ 82

Table 3.2. Basic model significant soybean counties at $p<0.05$ ........................................ 83

Table 3.3. Full model significant corn counties at $p<0.05$ .............................................. 85

Table 3.4. Full model significant soybean counties at $p<0.05$ .............................................. 86
LIST OF ABBREVIATIONS

AYI – Area Yield Insurance
COAPS – Center for Ocean-Atmospheric Prediction Studies
DD – Degree Day
DF – Degrees of Freedom
DSFW – Days Suitable for Field Work
ENSO – El Nino Southern Oscillation
EPIC – Erosion Productivity Impact Calculator
FE – Fixed Effects
FIPS – Federal Information Processing Standards
GDD – Growing Degree Day
GDP – Gross Domestic Product
GRP – Group Risk Plan
HDD – Heating Degree Day
IV – Instrumental Variable
JMA – Japan Meteorological Agency
MAPE – Mean Absolute Percentage Error
MAE – Mean Absolute Error
NDD – Normal Degree Day
NOAA – National Oceanic and Atmospheric Administration
OLS – Ordinary Least Squares
ONI – Oceanic Nino Index
Prc – Precipitation
RMA – Risk Management Agency
RMSE – Root Mean Squared Error
RMSFE – Root Mean Square Forecast Error
SMAPE – Symmetric Mean Absolute Percentage Error
SOI – Southern Oscillation Index
SST – Sea Surface Temperature
SSTA – Seas Surface Temperature Anomaly
TNI – Trans Nino Index
USDA – United States Department of Agriculture
USDA ERS – USDA Economic Research Service
USDA FAS – USDA Foreign Agricultural Service
USDA NASS – USDA National Agricultural Statistics Service
VPD – Vapor Pressure Deficit
WUE – Water Use Efficiency
Chapter 1.
Introduction

The fundamental factors behind economic progress and our ability to benefit from our environment can be, at the most basic level, attributed to climate and agriculture. It is through modern agricultural practices that nutritional intake is not a time-consuming endeavor, allowing us to direct our attention to other sophisticated tasks. Climate, by its very nature, dictates our decisions and need for adaptation. The far-ranging effects of climate can be seen in earth’s diversity in terms of health, economic growth, and culture across the varied climatic regions of the world. However, essential to our global society is the relationship between climate and agriculture (Oram, 1985).

With the emergence of new climate trends, particularly global climate anomalies, and growing volatility in key climate variables, the relationship between climate and agriculture has never been so significant to humanity as it is today. Here and henceforth, the term “climate anomaly” refers to far-reaching temporal changes in climate conditions, from their long-term norms, over large geographical areas. There are various such anomalies, including, but not limited to, Madden Julian Oscillation, Arctic Oscillation, North Atlantic Oscillation, and the El Nino Southern Oscillation (ENSO). The influence these anomalies, particularly ENSO, have on agriculture is extensive. Innovations in data collection, storage, and use of econometric methods have allowed us to gather a wide and complex variety of information on, as well as analyze this relationship, in order to stabilize, plan, and
improve agricultural output. ENSO is of particular interest, as its effects are vast and play a significant role in global food security. This thesis analyzes ENSO’s teleconnections with local weather and how the resulting correlations affect local agricultural productivity, focusing specifically on United States (US) staple crops like corn and soybeans (Trenberth et al, 1998). The objective of this thesis is to examine key factors of yield uncertainty faced by corn and soybean producers by analyzing ENSO in the context of agricultural production variability and prediction.

Agricultural production plays a vital role in global welfare and economic stability. The significance of this industry varies in the comparison between developed and developing nations. As of 2015, the agricultural sector in the United States accounts for roughly 11% of total employment and a relatively modest 5.5% share of GDP (USDA ERS). For developing nations, the agricultural sector plays an even more meaningful role, often constituting the majority of exports as well as the primary source of employment. One may make the conjecture that risk management and innovation in this sector should focus on aspects directly applicable to developing nations and their respective regions of the world. However, one must consider that developed nations, such as the US, provide the majority of production of many staple crops. These are consumed globally and, in many cases, are vital sources of nutritional intake. For this reason, among others, considering downside yield risk and variation in the US is of global relevance and significant socio-economic importance.

Agricultural commodity producers face a great deal of variability in crop yield levels and received prices, and employ various risk management tools in order
to combat this. Agriculture is unique in that the majority of output is dependent on exogenous and unknown variables, namely, weather conditions. As variability in yields and prices leads to global food instability and economic instability on both the micro (individual producers) and macro levels (the entire production chain and international trade), agricultural risk management tools are a key point of interest. These tools fall into three main categories: price management, farm-level management, and crop insurance. Currently, with regard to price management, a tested approach relies on the use of futures markets to hedge against future price fluctuations. This approach is particularly important in agricultural commodities where supply is determined at an early stage of the production process and difficult to alter throughout the season. Farm-level management focuses on yield risk - or more precisely, downside yield risk – and employs various tools, with limited success. This is due to the intricate nature of yield determinants, which are predominantly biological in nature and range from soil type to local weather to climate anomalies such as ENSO (Rosenzweig and Hillel, 2008). In order to better control these biological determinants, producers implement various farm-level practices such as use of fertilizers, pesticides, and irrigation. Crop insurance provides support to producers by acting as an intermediary between farmers and direct government support, and is a key source of stability in the agricultural sector. However, despite these tools, substantial yield losses often occur, and even these tools display inefficiency and high cost of subsidization. This is because these approaches attempt to mitigate losses after they have occurred, using means such as reimbursing for failed crops, etc. An alternative approach would be to foresee future
problems, and mitigate them as much as possible by implementing changes in current agricultural activity, such as adjusting crop type and shifting planting dates. In order to do so, however, complex information on agricultural systems (such as local weather forecasts, etc.) is needed, as is the ability to analyze and understand this information.

Examining global climate anomalies, such as ENSO, opens the possibility of predicting future weather conditions. With local high frequency weather conditions, one cannot predict, with any accuracy, weather forecasts more than one or two weeks in advance. With medium frequency climate anomalies like ENSO, however, one can predict forecasts up to several years in advance. Not only do climate anomalies associated with extreme weather events pose threats, but the steadily rising global temperatures responsible for these events (including El Nino and La Nina events) are themselves posing significant risks to agricultural producers. A 2003 study quantified the negative effects of climate change, finding that an increase in temperature by 1 degree Celsius could have a potential 17% decrease in yield (Lobell and Asner, 2003). The results are unclear, however, ranging from a decision to shift agricultural production to the north, to hopes that the increasing $CO_2$ levels will in turn stimulate crop growth. Nevertheless, effective ENSO forecasts play an indisputable role in improving producers’ decision-making. Thus, understanding the relationship between local weather and global climate anomalies like ENSO is of utmost importance.

The causal relationship between ENSO and US crop yields is analyzed by considering the correlation between local weather and yields. Of particular interest
are local weather variables, most notably temperatures and precipitation, via which ENSO manifests itself upon local crop production. In general, a robust El Nino carries warm conditions in the northwest and greater precipitation and lower temperatures in the southern US. La Nina on the other hand results in predominantly warm and dry conditions along the southern US (NOAA). The research compiled and analyzed in this thesis measures the relationship between ENSO and yields as well as the predictive content of ENSO for agricultural production, giving us two possibilities. First, the research may show a significant link between ENSO and crop yields in the US, expressed through identifiable weather pathways. Or, alternatively, this link may be unobservable with available data or diluted across a multitude of variables. If ENSO is observed to manifest itself via definable weather pathways, its influence (on crop yields) gains significance within a yield-risk framework through its potential predictive content.

The foundational premise of this research, and previous research in this field, is that there are observable variables which determine the yield of commercial crops. These variables can be easily identified as local climatic conditions driving the supply of essential nutrients and chemicals required for plant development. They interact not only with one another, but also with other factors, such as soil type and geographical location. A complex dynamic soon emerges that explains key development stages in a plant's life cycle. Crop simulation yields, such as the Erosion Productivity Impact Calculator (EPIC), based on crop production variable data, have been able to somewhat fill in the data and knowledge gaps (USDA). While they have
too many degrees of freedom to be truly reliable, they offer great insight into what variables impact crop yields the most.

The two most important known variables to determine crop yields are precipitation and temperature. Which one of these has the most impact, however, is inconclusive, as the last 30 years of research on climate anomalies and weather impacts on yields have shown. At times, temperature plays a more decisive role than precipitation, while other times, precipitation takes the lead. This is further complicated by these conditions varying by region and crop type. Several papers, such as the 1999 papers by Legler et al, and Phillips et al, respectively, identify precipitation as the key yield influencer. More recent papers, however, reach a nearly unanimous consensus of temperature proving most important, partially through its interaction with precipitation (Mourtzinis et al, 2015; Mourtzinis et al, 2016; Schlenker and Roberts, 2008). Additionally, to better understand these interactions, one needs to also scrutinize a crop’s biological processes. Other considerations include water use efficiency (WUE), and water deficit during key stages, which are closely related to a soil’s water retention capacity. These measures further establish temperature as a key variable of interest. The Schlenker and Roberts 2009 paper, which discussed non-linear temperature effects on yields, expanded the field’s understanding of temperature-yield effects tremendously. Schlenker and Roberts include the full range of temperatures a crop may be exposed to throughout its growing season (Schlenker and Roberts, 2009). Their approach accounts for non-linear temperature effects by incorporating a vector of weather variables into the regression framework.
ENSO occurs in a recognizable pattern of several phases, each of considerable length, which means its direct and indirect influence on temperature and precipitation variables is greatly varied, and of great significance to agricultural production. Of particular significance is the strength and magnitude of the unique impacts associated with each phase across various regions of the world. Such events extend past the scope of crop variability, and also directly influence human lives and livelihoods across the globe. ENSO originates from a particular area of the eastern Pacific Ocean, off the coast of Peru. The process is caused by changes in trade winds between Peru and Papa New Guinea, resulting in shifting warmer/colder ocean currents. It is through the teleconnections between sea surface temperatures (SST’s) and atmospheric pressures that ENSO effects have a global reach. Although arguably more complex, ENSO is characterized through three phases: El Nino (warm SST phase), La Nina (cold SST phase), and the Neutral Phase (mild SST). Through historical readings of SST changes and atmospheric pressure changes, researchers can analyze the ENSO link to not only global weather conditions but also to key factors that are closely linked with weather - i.e. agricultural production. Greater detail on ENSO readings, proxies, phase definition, and data interpretation can be found in the literature review.

The ENSO link to crop production in the US is of interest on several levels. The US is a key staple crop producer, and the significant changes and extreme weather conditions caused by ENSO in the US affect the rest of the world. Primarily the impact and cost of ENSO events on US agriculture which under 2001 ENSO strength and projected increased frequency of El Nino and La Nina phases could cost
up to 300-400 million USD in the US alone. Secondary and of particular interest in this thesis are the benefits of a deeper understanding of ENSO influence with regard to mitigating downside yield risk and understating future food security concerns. For example, considering non-linear ENSO effects. If one were to assume a negative impact of La Nina on crop yields in the US, it would not necessarily hold that an El Nino phase will result in positive effects of similar magnitude. Because of this, this thesis measures on non-linear ENSO effects in the estimation modeling framework, rather than taking a purely binary approach.

This non-linearity is particularly important when considering the ENSO-yield relationship. Interestingly, the literature presents a mix of results regarding each ENSO phase's impact on crop yields, even for the same crop. This, however, can be attributed to the varied geographical impact of ENSO. Today, research presents a near unanimous consensus on where La Nina events prove to be the most detrimental to yields, compared to El Nino events. To better understand these relationships, researchers have identified through which pathways ENSO impacts yields most greatly. The vast majority of contemporary research presents temperature as the key variable of interest (Phillips et al, 1999; Legler et al, 1999). Following the conclusions made by Schlenker and Roberts (2009), my research supports temperature as a key yield influencing variable and refers to high temperature levels in particular.

While much of the literature in this field discusses ENSO impact in the US and globally, there has been substantially less discussion on the predictive content of ENSO in a forecasting methodology for staple crop yields. As ENSO displays
increasingly greater volatility and strength, causing increasingly more damage to human lives and global food supply, the ENSO discussion must transition from analyzing the ENSO phenomenon, to a practical applicability of ENSO knowledge. This thesis explores how forecasts of future ENSO events can be used to forecast future crop yields. SST variability in the Pacific can be forecast up to one year in advance (Latif et al, 1994), and more recent studies show that ENSO events can be forecast up to two years in advance (Chen et al, 2004). Predictability of the outcome of ENSO phases has always been of interest, with Phillips et al (1999) noting that yield outcome may be more predictable during the transition period from one phase to another. This comprehensive research can prove invaluable in understanding, analyzing, and forecasting ENSO’s events and impacts, and, by extension forecasting and planning crop yields.

Asking the question “Does ENSO really matter?” not only builds the wealth of research available, but also contributes to the topic in three distinct ways. First, the research combines the non-binary ENSO effects with non-linear temperature effects on 60 years of historical soybean and corn yields. Second, having identified the pathways through which ENSO influences yields, a pseudo-forecasting methodology is employed to explore the predictive content of ENSO anomalies on US staple crop yields. And third, results are considered with regard to current use of ENSO information by producers. Conclusions presented here are discussed in parallel with studies on the utility such information may carry for yield/profit maximization and multiple crop producing regions/farms.
This thesis is concerned with establishing a causal link between ENSO anomalies and crop yields through a series of specifications, and identifying a weather pathway defining this link. Such a link is then exploited for use in forecasting methodology in order to better assess potential utility gains from incorporating ENSO into a risk management framework. Conceptually, this is achieved through a combination of ordinary least squares (OLS) and panel regression models. The estimation exercise utilizes county-level yield data for soybean and corn which are analyzed through panel data regression with time and location fixed effects (FE). Location specific FE’s are altered between state and county level, results of which motivate the subsequent forecasting methodology.

Variable selection occurs both on a theoretical basis as well as on a selection contingent on regression model performance. Such as for example, accounting for trade-offs between concerns over multicollinearity and variable omission, bearing in mind the structure of the causal relationship. In the econometric framework, there is a proclivity for redundant explanatory variables, where, for example, precipitation and humidity may be closely interdependent. As a result, ENSO influence over weather variables is narrowed to precipitation and temperature. Temperature is taken over the range of the crop’s growing season in order to account for non-linear temperature effects. This results in two temperature variables: 1) a measure of the crop’s exposure to a range of temperatures promoting growth, and 2) a measure of the crop’s exposure to temperatures surpassing the plant’s temperature resistance threshold, which results in growth suppression. Subsequently, temperature-yield-ENSO variables are run through various
specifications to test the direct and indirect inter-relationships. Through this, a
causal relationship can be identified between the medium frequency ENSO, the high
frequency local weather conditions, and the corn and soybean yields in the US. Yield
data limitations direct the research to the Corn Belt and the Eastern Seaboard,
which are two major staple crop growing regions in the US. Establishing the link
between ENSO and high frequency local weather facilitates forecasting of crop
yields by extracting and identifying the predictive content of ENSO.

In summary, the key contributions of this thesis center on: 1) ascertaining
weather pathways of ENSO’s non-linear influence over yields, 2) presenting
heterogeneity in causal effects of ENSO, including through these weather pathways,
in major US corn and soybean producing counties, 3) assessing the predictive
content of ENSO for corn and soybean yields, and 4) identifying crop growing
regions particularly vulnerable to ENSO and those likely to benefit from utilizing
ENSO-based forecasts. This thesis concludes with a discussion on the implications of
the results within current downside risk management techniques.
Chapter 2.
ENS0 Measurements and Applications for Agriculture – a Review

Over the past decade, there has been growing interest in the use of climate anomaly data to better understand and predict various weather conditions around the world (Handler, 1990; Legler, 1999; Chen and McCarl, 2000; Chen et al, 2001; Yokoyama, 2002; Tack and Ubilava, 2013b). In particular, through the use of Sea Surface Temperatures (SST) in the mid-Pacific region, researchers can explain a grander story via the El Nino Southern Oscillation weather phenomenon, most commonly referred to as ENSO throughout the relevant literature. Researchers have examined the influential power of these weather patterns in various fields, from agriculture to political change and unrest (Brunner, 2002; Hsiang et al, 2011; Ubilava and Holt, 2013). Some research has taken ENSO “forecasting” a step further by attempting to discern a correlation between ENSO phases and inflation rates/economic change. A 2008 study, however, failed to find such a relationship stating that there is no evidence of co-cyclicality between ENSO fluctuations and inflation rates or economic growth between 1894 and 1999 (Berry and Okulicz-Kozaryn, 2008). The relationship between ENSO and agriculture production responses, which is the focus of this review, is far more discernable (Handler, 1984; Nicholls, 1985; Chen and McCarl, 2000)

The value of a forecasting mechanism derived from a single anomaly with global relevance should not be understated, and therefore a closer look at its potential for improved (or rather informed) management practices is essential. This
point is further amplified via ENSO volatility increasing steadily with climate volatility (e.g., Trenberth and Hoar, 1997; Zhang et al, 2012). A 2001 study by Chen et al, quantifies the damages occurring as a result of severe weather closely associated with ENSO fluctuations. The authors state that if the frequency of warm and cold phases increase (El Nino 0.238 to 0.339 and La Nina 0.250 to 0.351 probability), costs could be 300-400 million USD annually. If the events also intensify in strength these damages could rise to 1 billion USD in the US alone (Chen et al, 2001). Considering costs associated with a La Nina event, major US field crops range between US$2.2-6.5 billion, and El Nino with lower predicted costs between US$1.7-2.2 billion (Adams et al, 1999). Considering a staple crop such as rice, El Nino and La Nina impacts are estimated to range between US$0.7-2.1 billion globally (Chen et al, 2008). Predictions of ENSO induced costs alone may be a sufficient catalyst for the better understanding of the teleconnections at play and how they influence the livelihood of millions around the word, however, one may also look back rather than to the future to see the real costs of ENSO events. In the recent past, one such noteworthy ENSO event occurred between 1982 and 1983, resulting in an estimated US$13 billion in damage (Pfaff et al, 1999). Substantial losses also occurred over the 1997-1998 ENSO (El Nino phase) occurrence (McPhaden 1999; Adams et al, 1999; Changnon 2000). Increasing awareness as to the cost of climate change motivates an in-depth look into the ENSO phenomenon and forecasting possibilities. Utilizing forecasts for crop management can reduce damages associated with severe weather significantly. A proxy for the reduction may be
considered via a predicted drop in crop insurance indemnity payouts of 10-15% through the use of ENSO forecast data (Tack and Ubilava, 2015).

2.1 Defining ENSO

A discussion of ENSO should begin with how the anomaly is measured for use in its various applications. Measurements originate from both sea surface temperature and atmospheric pressure changes along the equatorial segment of the Pacific Ocean between the coast of Peru and Papa New Guinea. Changes in these two variables and their interactions are the foundation of the ENSO phenomenon, summarized as oceanic and atmospheric changes measured in predefined regions of the tropical Pacific Basin (Hayes et al, 1986; Neelin et al, 1998; Allan, 2000; Hanley et al, 2003). For example, Chen et al opt for a monthly time series index measuring oscillations in sea surface temperature in a specific area of the pacific (5,5: N, S – 170,120; W) (Chen et al, 2002). Referring to one of the most commonly used indices, known as the NINO 3.4 SSTA Index (Magrin et al, 2003; Dawe et al, 2009). In this thesis, ENSO is defined as the sea surface temperature (SST) anomaly of the Oceanic Nino Indices (ONI) in the Nino 3.4 region of the Pacific basin. The terminology, as well as alternative measures of ENSO, are described in detail below.

Expectedly the first observations of the ENSO phenomenon did not originate from its global effects carried through atmospheric teleconnections or the relatively stronger impacts in the neighborhood of the tropical Pacific (Trenberth et al, 1998; Kiladis and Diaz, 1989; Allan, 2000). Instead, changes in sea temperatures off the coast of Peru by Ecuadorian and Peruvian fishermen marked the point of the first
ENSO related discovery (Ramage, 1986; Trenberth, 1997). The discovery of this phenomenon from its first observation by fishermen in the 1500's began making appearances with scientific circles as an observed yet not well-explained occurrence, specifically in 1926 and 1942 by the authors Murphy and Lobell respectively. The name by which this occurrence was referred to was derived from the first observers and the timing of warm water currents over the winter period, resulting in the name El Nino or the Christ child (Wang and Fiedler, 2006). The term Southern Oscillation, on the other hand, was the result of another vital contribution to the investigation and defining of ENSO by Sir Gilbert Walker, resulting in a myriad of papers between 1920 and 1930. Most importantly Walker defined the widespread effects on climate anomalies, in particular referring to atmospheric pressure changes and their teleconnections between the Pacific basin and precipitation in Asia (Walker and Bliss 1932). Walker's contributions reside primarily in setting the first definitions of atmospheric pressure systems and the respective interconnections. More important, the author addressed how these systems may be linked to climate on a global scale through his 1923 and 1924 papers (Katz, 2002). In a 1972 paper, Bjerknes links the two key observations described above stating there is an "ocean-atmosphere interaction for inter-annual climate change" (Bjerknes, 1972). The changes described in the variables above are the result of weakening and strengthening trade winds causing changes in water temperature composition and variation in atmospheric pressures (Philander and Rasmusson, 1985).
Although the historical discovery of this phenomenon is significant in its own right, of particular interest to the purposes of this thesis is the dual nature of ENSO measurement. Namely ENSO influence, as well as ENSO measures, are derived both from oceanic changes as well as atmospheric changes, describing a relationship which facilitates the teleconnections spreading ENSO effects globally. A plethora of ENSO related literature began taking place in 1983 after the 1983 El Nino event which grabbed the attention of researchers due to its significant magnitude. From this point forward both ENSO measurement, definition and implications have developed at a tremendous rate.

ENSO is accounted for through a variety of proxy variables measuring deviations from the status quo of both atmospheric pressure variations and sea surface temperature variations. The atmospheric pressure readings, referred to as the Southern Oscillation Index (SOI), are categorized into phases based on the generally accepted ENSO phase definition frequently described as El Nino, La Nina and the Neutral Phase (NOAA). These phases are also used by the Oceanic Nino Indices (ONI) describing variation in sea surface temperatures as a 3-month average between November, December and January. SOI readings are taken in an area between the western and eastern extremities of the Pacific Ocean, more precisely by comparing pressure changes between Tahiti located in the mid-Pacific and Darwin on the western end of the Pacific (Trenberth, 1984; Ropelewski and Jones, 1987). The SOI is often computed as the difference between Darwin and Tahiti, leading to concern about imprecision derived from taking readings from two separate locations (Wolter and Timlin, 1998). The SOI, unlike the Oceanic Index, cannot
directly relate by definition to warm or cold phase but rather deviations in pressure where an atmospheric pressure change of +7 (on the SOI scale) refers to La Nina (Cold Phase) and a drop in pressure of -7 refers to El Nino (Warm Phase) (Australian Bureau of Meteorology). Under this anomaly reading structure on average, the ENSO cycle lasts 4 years, with 5 El Nino and 3 La Nina episodes recorded between 1980 and 2000 (NOAA). This speaks to the variability of ENSO where phases are seldom identical between cycles and vary in magnitude (La Nina vs. El Nino) (Ropelewski and Halpert 1987; Guilderson and Schrag, 1998; Allan, 2000).

The SOI closely mirrors the readings acquired from the ONI (NOAA). The ONI relies on temperature data in the form of sea surface temperatures (SST) and often utilized in research as sea surface temperature anomalies (SSTA). Unlike the SOI the ONI index is composed of several sub-indices, each referring to a specific location in the Pacific basin. These span an area along the equator from the coast of Peru towards Papa New Guinea, where moving in the same direction the first regions index is Nino 1+2, Nino 3, Nino 3.4, and finally Nino 4 representing a region of the Pacific basin east of Papa New Guinea (Trenberth and Stepaniak, 2001; Glantz, 2001; NOAA). One should note that Nino 3.4 is not a total of both Nino 3 and Nino 4 but rather partially contains each region. Amongst these indices, there is also the JMA index (Japan Meteorological Agency) spanning the area between 4°S-4°N, 150°W-90°W, as well as the TNI index which is given as the difference between Nino 1+2 and Nino 4 as normalized anomalies presented by Trenberth and Stepaniak in their 2001 paper (COAPS). ENSO is again defined by anomalous deviations
representing El Nino, La Nina, or in the event of no deviations the Neutral Phase. Temperatures readings over the +0.5 threshold for 5 consecutive 3 month running means defines an El Nino phase, while readings below -0.5 defines a cold or La Nina phase (NOAA).

ENSO cycles typically occur every 4-8 years with the frequency displaying an increasing trend (Brunner 2002; Wang and Fiedler 2006). Furthermore, the increasing frequency and severity have been tied to climate change and global warming trends observed around the globe (Urban et al, 2000; Timmermann et al, 1999; Trenberth and Hoar 1997).

In this manner, anomalous deviations result in either the cool phase (La Nina) or the warm phase (El Nino) (Bartels et al, 2012). These phases can span several years and often result in divergent effects throughout the world. A 5-phase classification provides an alternative to the 3-phase system described above. Due to limitations in data collection, the 3-phase system is the most popular and easiest to use, however, the 5-phase system appears to have considerable benefits. The 5 phases are 1) persistently negative 2) persistently positive 3) rapidly falling 4) rapidly rising 5) neutral (Chen et al, 2002). In the 1998 paper Phillips et al use Stone and Aucliciens 5-phase definition, which according to their study, its use over the 3-phase definition could double the amount of welfare for the agriculture sector. Furthermore, a 5 phase system (or even more) has the ability to reduce the in-phase variability of forecasts and predictions resulting in difficulties in deriving forecasts, which may be solved by more precise definitions of ENSO phases (Phillips et al, 1998). In practice, the 5-phase system is seldom encountered in the literature.
2.2 Capturing ENSO link with local weather and production

ENSO can be thought of as a medium frequency weather event closely correlated with high-frequency local weather events. In general, throughout the literature ENSO proves to influence precipitation above all other climatic conditions, a close second is variations in temperature. A possible combination of both can cause drought, which has been shown to be most detrimental to staple crop production, in particular, rice or other water-dependent crops (Wassmann et al, 2009). The argument, in essence, boils down to via which pathways (or both) does ENSO influence yields. Easily recorded and observed local climate variables are temperature and precipitation, hence a detailed overview at how ENSO interacts with these variables on a global scale is warranted. When discussing such links, one must consider ENSO influence by region as well as by magnitude of influence on particular climate variables. This argument is supported by the simple fact that ENSO impact on weather is by no means homogenous from a global perspective. On the contrary, it may be the case that clusters of ENSO influence can be discerned even on a relatively precise regional basis, such as county or shire level.

Considering the United States cold ENSO events have larger impacts on southern regions in the US, while warm ENSO events have larger impacts on the North (Legler et al, 1999). Additional vectors beyond temperature and precipitation including mean yields measured as a percentage change of the mean during El Nino relative to mean of neutral phase can be far more informative of ENSO impact on yield (Tack & Ubilava, 2013b). These notions are supported by Legler et al (1999),
whose research shows that precipitation totals and medians are not enough to accurately quantify all ENSO related yield deviations. Precipitation measurements around the world have displayed a close relationship with ENSO, with varying effects and magnitudes from one region to another (Ropelewski and Halpert, 1987; Stone et al, 1996). Subsequently, following the ENSO pathway towards yields, the influence on precipitation variability on crop yields has been explored thoroughly mainly focusing on soil water retention and interaction with temperature variables. Moving precipitation aside as the weather variable most influenced by ENSO, researchers must consider which weather variable has the greatest influence over crop yields. The nearly unanimous consensus zeros in on temperature (Wheeler et al, 2000), especially for staple crops such as corn. Specifically, extremely high temperatures proving to carry the most detrimental impact on yield and biological development (Lobell and Burke 2008; Schlenker and Roberts, 2009; Fisher et al, 2012). This is particularity true in key phases of plant development, for instance during flowering (Wheeler et al, 2000). This research is expanded on, stating that in general crops have been observed to become less prone to changes in precipitation (via technological advances for example) yet more sensitive to significantly higher temperatures (Roberts and Schlenker, 2010; Lobell et al, 2013). To better understand the influence of temperature the use of Degree Day Data has shown great statistical power in regression models, considering non-linear effects of temperature on crop yields. Non-linear temperature effects are comprehensively discussed by Schlenker and Roberts (2009). Where rising temperatures promote
biological growth up to a certain threshold, after which there is a steep detrimental effect.

This relationship between yield and temperature introduces the advantage of including innovative climate variables, such as VPD. Vapor Pressure Deficit (VPD) measures the water holding capacity of surrounding air by differencing water holding capacity at full saturation vs nonsaturated (normal). The inclusion of VPD in yield ~ climate regression analysis has resulted in the improved predictive power of the model alongside Degree Days (Fisher et al, 2012). A similar inclusion of VPD by Lobell et al across historical corn yields shows an increasing sensitivity of corn yield to high VPD levels (Lobell et al, 2014). Another paper finds that there may be lower risk to direct heat exposure for yields, with VPD playing a much more significant role (Lobell et al, 2013). These papers highlight the growing literature on methodology and use of appropriate weather variables and proxy variables to establish a link between ENSO and agricultural production. Further reflecting upon interactions, consider the popularly acknowledged correlation between rising global temperatures and $CO_2$ emissions. The resulting negative temperature effects on agricultural productivity could be offset by the positive effects of higher $CO_2$ levels (Wheeler et al, 2000).

Tack and Ubilava delve further into the more precise use of ENSO data with yield, stating that the use of "acreage weighted aggregate" measures for influences of ENSO fails to bring to light heterogeneity within a particular county of study. Instead, mean variance and measures of downside risk for crops are recommended (Tack & Ubilava, 2013b). This is echoed by Paz et al displaying that the aggregation
of yield estimates above the county level may not provide an adequate interpretation of ENSO impacts (Paz et al, 2012). Such descriptive statistics as referred to by Tack and Ubilava are necessary when considering the overall distribution of crop yields in order to analyze tails and skewness, where positive skewness refers to downside risk faced by producers. The lack of such detailed analysis has the potential for misinterpreting ENSO data and failing to draw the most important and relevant conclusions. The methodology used by Tack and Ubilava for arriving at conditional yields for each county was the Moment Based Maximum Entropy Model. Performed by leveraging the model’s parameter estimates and allowing for the measurement of lower tail outcome probabilities (Tack and Ubilava, 2013b).

In 2002, a paper by Chen et al brings to light an issue regarding the use of long time series data and maintaining a consistent probability distribution which can change due to technological progress and various other factors (Chen et al, 2002). This issue is considered by Tack and Ubilava (2013b) through adding trends to ENSO regression models. As a result, allowing the mean of the yield distribution to vary over time and space, which introduces yet another issue of spatial (not only temporal) variability. This is significant because yield data are both temporally and spatially distributed. Legler et al (1999) identified the issue of the “diverse nature of ENSO-related yield deviations” and for this reason (among others) referred to the EPIC (Erosion Productivity Impact Calculation) framework. Further studies refer to spatial and temporal variability such as the recent study by Bartels et al (2013) finding that, for example, ENSO is spatially consistent at the regional level.
Finally Podesta et al find 4 key issues with measuring ENSO impacts on agriculture: 1) Crop records encompass only a limited number of ENSO events, 2) difficulty in assessing current reaction to climate-induced vulnerability due to change in technology or other exogenous factors, 3) difficulty in separating various origins of agriculture risk, and 4) risk estimates from aggregate data may not be appropriate for farm level risk management (spatial aggregation dulls year-to-year variability) (Podesta et al, 2002). Yet another caution lies with shifting yield response to weather types, meaning yield does not always behave similarly to comparable weather conditions. Two papers discussing this aspect of yield forecasting place the reasons for this in new crop varieties, improved technologies, and last but not least varying soil type (Changnon and Winstanley, 2000).

2.3 ENSO impact across the US

When discussing the measurement of ENSO impact on agricultural production in the US a strong candidate emerges in corn, of which the US is the world-leading producer. Amidst leading studies, there is a clear consensus on ENSO influence across the US. The following section aims to summarize the key findings.

Some studies take a modeling approach based on simulated yields, such as Adams et al (1995), testing the response of simulated agricultural productivity to ENSO related "climate-variability parameters". The authors find that cold ENSO events have a greater impact on southern regions in the US, while warm ENSO events have larger impacts on the North (Legler et al, 1999). More specific results for corn are found in a variety of notable papers, most recently by Tack and Ubilava.
in 2013. El Nino reduces central corn belt means, with positive yield impacts when moving east or west of the Corn Belt. While La Nina reduces west of center Corn Belt means, positive impacts when moving east or further west of the Corn Belt (Tack & Ubilava, 2013b). Below the Corn Belt, southeastern US, La Nina years display positive impacts on corn yields while El Nino generates negative impacts (Hansen et al, 1998). Linking sea surface temperatures with corn yields displays a meaningful relationship in various studies, with up to 15% US corn belt corn yield variability attributed to sea surface temperature fluctuations (Handler, 1990; Phillips et al, 1999).

In the southeastern states of Florida, Georgia, and Alabama El Nino brings lower temperatures and greater precipitation (Martinez et al, 2008). Detrimental effects on corn yield are associated with both, from potential freezing in the germination stages to over saturation and drowning of seedlings. Paz et al studied cotton yields as influenced by ENSO at different planting dates and spatial aggregation levels, showing that planting date is an important variable that has to be adjusted to an anticipated ENSO event. Achieved by looking at ENSO influences on spatial scales lower than state level (Paz et al, 2012). Planting dates before May 9th, show highest yields during La Nina, while for dates after May 23rd El Nino yield is higher. Finally, for cotton producers in the southeastern United States, the highest expected profit can be generated in the neutral phase while the lowest occurs during the La Nina years (Aitsahlia et al, 2011).
2.4 ENSO and producer decisions

Challenges not only exist in use of ENSO from an academic perspective, but also in the translation of ENSO climate data into usable information for improving producer's decision-making as discussed by Podesta et al (2002). Each phase does not signal one specific response, rather it varies across geographical locations and crop type or crop mix. This is one of the reasons why Barrett argues that the use of ENSO data alone, especially raw data, has limited use if the policy environment, technology for the application, and delivery mechanisms for data to producers have not advanced sufficiently (Barrett, 1998). These concerns are shared with Magrin who also argues that information alone is not the chest of panaceas for producers; the proper "decision-support tools" must also be available, as is a change in the decision-making process (Magrin et al, 1999). Furthermore, one must mention the case may be that various regions (globally) may experience indistinguishable ENSO signals (Baigorria et al, 2008).

Producers can utilize forecast data to not only avoid risk but also so significantly increase profits, however variability in inter-phase predictions may deter producers from utilizing ENSO forecasts (Letson et al, 2005). The value of which must exceed the costs, thus providing the potential for welfare gains.

Considering dual-purpose winter wheat production in the southern high planes of the US, Mauget et al (2009) show that use of climate data and optimizing practices to suit future weather conditions has shown to increase profits significantly. Two notable papers offer estimates of such welfare gains. Firstly, via a simulation model, the value of more precise ENSO forecasts is said to be between US$240-323 million
Another study considering maize in the US state of Georgia finds utilizing ENSO forecasts in adapting crop managed results in gains per hectare between US$3-5 (Jones et al, 2000). Additionally, ENSO related weather variability is shown to influence producer management decisions aside from apparent yield effects. In fact, recent findings indicate that precipitation influences producers’ days suitable for field work (DSFW), effectively disrupting crop yield levels regardless of direct weather effects on yield (Mark et al, 2015). Under an El Nino phase, DSFW are shown to decrease, while the opposite may be true for La Nina, resulting in suboptimal use of lobar and machinery (Mark et al, 2014b). Furthermore, through the influence of rainfall, it is shown that ENSO, if forecasted in advance, could influence planting date decisions (Woli et al, 2013).

2.5 Implications of ENSO impact

Having established reasonable evidence of a connection between ENSO occurrences and agricultural activities, various interesting topics emerge from this conjecture. Namely the benefits of using irrigation to more sophisticated water management techniques such as water markets, which are emerging throughout the world. From initial seed germination to the final ripening stage of plant growth, there exist a myriad of influencing as well as interdependent variables. Water stress has been shown to cause the greatest harm to yield, more than all other factors combined (Phillips et al, 1999; Lambers et al, 2008). Furthermore, the lower the water holding capacity of soil the greater is the influence of ENSO (Woli et al, 2013). As Tack and Ubilava mention in their paper, adverse events can result in short-term damages,
however through the use of forecasts and long run planning perspective substantial risk can be avoided by adapting management techniques in response to forecasted changes in weather (Tack and Ubilava, 2013b). From a producers’ perspective, various options emerge in response to informative forecasts. These can range from changing planting date to changing crop type or even adapting location or soil type.

These actions are shown to have substantial risk mitigating power, in the sense of making the best of poor future weather conditions or maximizing profits during positive weather conditions. Roberts et al succinctly summarize the importance of rainfall in the Philippines by stating that a clear relationship emerges in that SSTA influences rainfall which in turn influences production, yield, and area harvested (Roberts et al, 2009). The degree to which water constraint impacts yields clearly does vary among crops. As mentioned earlier, highly water-dependent crops such as rice are influenced by ENSO most severely.

Interestingly some studies have found little influence of ENSO on precipitation, sunshine hours, and annual mean temperatures yet in some ways unexpectedly clear yield differences occur in La Nina and El Nino years compared to Neutral phase years (Liu et al, 2013). Implying ENSO may influence various weather variables of interest in crop growth. This was the case for US winter wheat and summer maize. The study goes on to produce interesting spatial variation in ENSO, with La Nina carrying the strongest and positive influence on yield in the west pacific region and negative influences in the east Pacific region (Liu et al, 2013). Finally, the authors found that Maize production was more valuable than winter
wheat in El Nino and La Nina years, which has significant importance for producers’
crop mix and planting date decisions.

Taking a closer look at soil type and vital stages of a crops growing season,
the middle of the planting window proves to be the least susceptible to ENSO (Woli
et al, 2013). Enforcing the approach of including growing seasons within ENSO-
Yield models. Woli et al study the ENSO effect on peanut yield for the southeastern
US as influenced by soil type, and planting date at regional and sub-regional levels
(9 planting dates and 7 soil types in rain-fed systems). Considering the spatial and
temporal variability of Water Use Efficiency (WUE ratio of cotton seed yield v.
evapotranspiration) as influenced by ENSO by looking at data from 1950 to 2006 for
the states of Alabama, Florida, and Georgia, the study finds significant differences
between ENSO phases. El Nino years show WUE deviations negative for early
planting and positive for later and larger planting date, while in La Nina WUE reacts
in the opposite way to El Nino Years (Garcia y Garcia, 2010). However stronger
spatial dependence is seen in La Nina years, which depends on water retaining
capacity of soil. The differences between WUE variations on ENSO phases can be
summarized as Neutral < El Nino < La Nina. Garcia y Garcia et al 2010 come to the
strong conclusion that ENSO based planting decisions can be beneficial to reducing
rain-fed cotton vulnerability in the south-eastern USA, particularly during La Nina
(Garcia y Garcia et al, 2010).
2.6 ENSO influence beyond the US

Considering other regions, such as India, the earlier point regarding water stress is exemplified by Selvaraju (2003) in that rice is more greatly influenced by ENSO than other crops. Again, irrigation can help weather interphase fluctuations in rainfall. For example, in the northern Chinese planes, Liu et al find that both El Nino and La Nina reduce precipitation, in turn reducing winter wheat yields. While on the other hand, stronger yields occur during neutral years- especially when irrigation is used (Liu et al, 2014). Naturally, the degree to which ENSO influences these changes in rainfall is significant when considering the reliability of such forecasts, the authors of the 2013 paper found that as much as 49-53% of rainfall variation in the northern Chinese planes can be attributed to ENSO events (Liu et al, 2014). Another study by Phillips et al (1998) find a strong correlation between ENSO and rainfall in southern Africa, more specifically Zimbabwe, through considering ENSO influence on maize growth parameterized on soil conditions (Phillips et al, 1998). The 4 regions studied by Phillips in southern Africa where: Karoi, Gweru, Masvingo, Beitbridge which displayed the most significant decrease in rainfall as a result of ENSO during the El Nino phase compared to La Nina and the Neutral Phase. (Phillips et al, 1998). Results produced through considering ENSO influence on maize growth parameterized on soil conditions.

Similar conclusions as to ENSO –precipitation- yield are drawn from a study on grain crops in central-eastern Argentina, using the 3 phase categorical ENSO classification. Resulting in that ENSO anomalies are most vividly seen through precipitation anomalies in November to July (Magrin et al, 1999). More specifically,
Maize and Sorghum show increasing yield deviations occurring more frequently than under normal conditions during warm ENSO events, and less frequently in cold ENSO events. On the other hand, yield increases are greater than those occurring in cold events (Magrin et al, 1999). Soybeans followed this scheme, while winter Wheat displayed no association with ENSO. Maize displayed the clearest association with ENSO. This may be due to sorghum being able to modify its maturity in response to water availability, illustrating the importance of which crops are chosen for studies (Magrin et al, 1999).

Australia has also displayed a close relationship between ENSO and precipitation, in turn affecting crop yields. This relationship is shown to be heavily influenced by the IPO (Inter-decadal Pacific Oscillation). Where when the IPO presents falling SST’s, ENSO via SOI (Southern Oscillation Index) has a close correlation with local weather, particularly precipitation (Power et al, 1999). With regard to ENSO influence over crop yields, a 1985 paper by Nicholls finds evidence of potential meaningful correlation between ENSO and a variety of crops such as wheat and barley. The lack of definitive evidence results from lack of rich crop and weather data, and potentially due to divergent ENSO influences over Australia. A more recent study of ENSO effects on Australian wheat, identifies 3 distinct type of El Nino phases, with divergent influence, depending on variables such as "the timing of onset and location of major ocean temperature and atmospheric pressure anomalies" (Potgieter et al, 2005).

Further considering Latin America and from a nearby geographical region (although coastal), rainfall in central Chile is also found to be correlated with SSTA,
as well as evapotranspiration which is affected by ENSO phases as SSTA rises (warmer temperature resulting in greater precipitation) (Meza, 2004). La Nina years generate greater demand for water, while El Nino years experience low evapotranspiration and thus lower water demand (Meza, 2004).

Motivating the research at hand does not require a great deal of imagination. With the US being a major producer of caloric intake foods globally and producing 40% of global corn, yield risks should not be ignored. More troubling are possible remedies to established climate change models and increase weather volatility. The simplest approach lies in shifting production to more suitable climates. However, this may not be as simple as one imagines, corn production has been migrating for the past 100 years already taking into account climatic changes (Beddow and Pardey, 2015). Accelerating climatic changes point towards an accelerated shift of crop growing regions. The potentially taking the US for example studies have found that the highest quality soil types lie in currently moderate temperature regions (Roberts and Schlenker, 2010). Hence shifting towards cooler locations, as temperatures rise, may not be complete solution. This one example illustrates the complexity and significance of adverse weather occurrences across the United States.
Chapter 3.
Estimating the Effect of ENSO on U.S. Corn and Soybean Production

ENSO influence over our modern world is varied in strength as well as impact. Agriculture, however, is not only directly influenced by ENSO but also a key link in the chain which spreads ENSO’s global influence. Defining this causal relationship has been the topic of numerous papers over a substantial period of time (Handler, 1990; Legler, 1999; Chen et al, 2001; Tack and Ubilava, 2013b). Additionally, there exists a similar body of literature discussing not only ENSO influence but weather and climate variables link with yield in a less ENSO centric perspective (Long et al, 2006; Miao et al, 2015). In establishing the causal link necessary for the purposes of the research at hand the range of topics and issues discussed in previous literature is of great use and provides a solid foundation. In particular, literature discussing interaction between key weather variables and crop growth (Schlenker and Roberts, 2009) as well as literature discussing “indirect impacts” on yield (Miao et al, 2015) and how to account for them are of particular relevance.

In this modeling and estimation exercise, it is of particular importance to define a precise weather pathway through which ENSO influences US crop yields. In order to facilitate the measurement of accurate estimates, the ENSO ~ Yield relationship is described along various dimensions. Where yield is modeling not only directly on ENSO, but rather various specifications described via which weather variables ENSO influences yield to the greatest degree. In order to allow for sufficiently large datasets in the regression analysis two staple crops are chosen, namely corn and
soybean. Major producing regions are shared for both crops allowing for further comparison of ENSO effect not only spatially but also between crops. In the case of the magnitude of ENSO impact between Corn and Soybeans, as the results will present, the differences will be primarily based on each crops sensitivity to particular weather variables.

Both corn and soybean are annual crops, grown over the March-August period. Corn is produced both for the domestic market, making up 95% of fed grain produced, and as much as 20% of corn production is exported abroad. Major corn production regions include Illinois, Iowa, Indiana, South Dakota, Nebraska, Kentucky, Ohio, and Missouri. Illinois and Iowa alone make up 1/3 of US corn production. Soybean is often grown in rotation with corn. Production of soybean occurs predominantly in the upper Midwest, making up 80% of total production. Soybeans have experienced a rising popularity, making up close to 90% of US oilseed production, particularly after success in the field of bioengineering (USDA ERS, 2017).

3.1 Conceptual Framework

On the “fine” scale regional level (county) a yield function or production of soybean and corn must aim to identify those factors responsible for the crops’ biological development. These factors can be categorized into two separate groupings, first those directly affecting the biological development of each crop such as the exogenous inputs for agricultural production (i.e. weather). Through the use of both irrigated and non-irrigated fields, a given crop experiences local weather changes
differently. Nonetheless, the relative impact of short-lived deviations in precipitation is low considering more drastic changes in growing season precipitation exposure. Such changes over the span of the growing season affect both irrigated and non-irrigated crops, hence alleviating the issue of defining such "direct" factors for a mixed dataset.

The key purpose is aggregating yields from both irrigated and non-irrigated sources into a single mixed dataset, has to do with data availability and the preferred time interval over which yields for both crops should be observed. Choice of time interval varies among studies and rightly so, as the definition of climate for various research purposes may take different forms. Certain climatic conditions may change over a 3-year time span while others may only alter after 20 or more years have passed. With regard to the focus on this study, ENSO (as described in greater detail in the literature review section of this thesis) is a reoccurring phenomenon composed of 3 phases occurring on average every 3-7 years to complete a cycle. However, this alone is merely anecdotal, as it is not sufficient to observe one occurrence of the ENSO phenomenon. Simply observing how yields shift during a particular La Nino or El Nino phase does not provide sufficient variability in both local weather response as well as ENSO magnitude to warrant the strong definition of such a relationship. Therefore in the interest of the study as long a time interval as possible was necessary to achieve desired statistical significance in any relations between ENSO, local weather, and local corn and soybean production. Taking into consideration issues regarding the magnitude of weather variable impact on the two groups of yield, an aggregated yield set was optimal.
The yield or production function must not only consider those variables of specific interest to the topic of this study, but also all variables responsible for influencing yield levels for any given year. Returning back to the concept of direct and indirect factors influencing yield levels is at the center of measuring ENSO influence over local US staple crop production.

Direct factors, as previously noted, are described as weather variables resulting in uncontrollable factors necessary for the biological development of the crop at key growth stages. Key growing stages are aggregated together and defined as the crops’ growing season. For both corn and soybean, the growing seasons spans a period of 5 months from March to August. Over this period local weather variables at a county level were aggregated on an annual basis. Aggregation causes a lack of accounting for intra-year variability, or rather growing season variability, in these variables which is a cause for concern as field crops have shown to be affected by variability in key weather variables (Wu 2008, Miao et al 2015). This issue is accounted for by considering the relative magnitude of each weather variable on yield suppression. Two such fundamental weather variables are precipitation and temperature, both fundamental in the biological processes of plant development. Precipitation presents various challenges and difficulties in precisely measuring not only precipitation on its own but rather water supply and access to the plants’ stomata. Thus, precipitation and variation in precipitation does not serve as a comprehensively accurate proxy for water availability for a plant over its growing season. This is due to various factors including soil water retention, water use efficiency of the plant under various conditions and regions, and evapotranspiration.
Due to the challenges and relative complexity of accurately measuring such factors (i.e. WUE, evapotranspiration, soil water content), such measures are beyond the scope of this research resulting in the inclusion of precipitation within the yield function as a simple monthly average over growing season months. This provides a relatively sufficient account of water availability during the plants growing season, although lacking vital variability. Naturally, overexposure to water, hence higher precipitation, result in damaging effects on crops yield. To account for a known nonlinear relationship between rainfall and yield, the precipitation enters in a quadratic form in the regression.

Aggregating monthly averages of precipitation over the growing season on a county level is further justified by the relative impact of precipitation compared to temperature. Recent work, as detailed in the literature review, promotes the arguments that temperature variability particularly at the higher threshold results in the greatest yield suppression. Even more so, temperature proves to promote biological development and plant growth to the greatest degree, under so called “normal” temperature intervals, defined below, relative heating degree days (above a crops “comfort temperature” threshold). This in turn requires a sophisticated procedure for accounting for this vital variability in temperatures. As previous studies have taken the above approach applied to precipitation, the effect has resulted in under estimated impacts of temperature by failing to account for the full distribution of temperatures in a crops’ growing season. To account for this vital non-linear response of yields to temperature, the use of spatially and temporally fine scale temperature data at monthly intervals is used to measure temperature at...
a county level. A range of so called Growing Degree Days (GDD) is acquired for each county of interest per month, where exposure to temperatures ranging from 0 degrees Celsius to 34 degrees Celsius is measured. These GDD are included in the crop production function as two separate variables accounting for the non-linear temperature effect. These variables should not be confused with energy use concepts by similar name, for example used in measuring energy consumption during heating and cooling of households. Relative to each crops preferred temperature thresholds, monthly temperatures are divided into so called Normal Degree Days (NDD) and Heating Degree Days (HDD) given as,

\[ NDD_{jit} = dd0^\circ C_{it} - ddx^\circ C_{jit} \] \hspace{1cm} [3.1]

\[ HDD_{jit} = ddx^\circ C_{jit} \] \hspace{1cm} [3.2]

where \( NDD_{jit} \) for crop \( j \), county \( i \), and year \( t \) is defined by the difference between \( dd0^\circ C_{it} \) and \( ddx^\circ C_{jit} \). Where \( dd0^\circ C_{it} \) presents a measure of the exposure in a given county and year (aggregated over crops growing season) to temperatures above 0°C. In turn \( ddx^\circ C_{jit} \) is a measure for exposure to temperatures above \( x \), defined as the critical temperature threshold for each crop, \( x = 29^\circ C \) and \( x = 30^\circ C \) for corn and soybeans respectively (Schlenker et al, 2009). In this manner \( HDD_{jit} \) represent exposure to temperatures above each crops’ given threshold. In summary, Degree Days measure exposure to temperatures in length of time, using daily minimum and maximum temperatures. For corn (soybeans) NDD are defined as exposure to
temperatures 0(0)-28(29)°C, and HDD exposure to temperatures of 29(30)°C and above. Degree Days for each 1°C interval are available between 0°C and 40°C, with each interval value representing the length of time a county is exposed to the given temperature interval. For example, in year 2000 in Adams (IL) the HDD (corn) is 17.25. Meaning, 17.25 is the cumulative exposure for corn crops in Adams (IL), during March-August 2000 (growing season), to temperatures above 29°C.

NDD refer to the range of temperatures under which plant growth is likely promoted or at the least progressed at a normal pace. Potential extreme cold weather shocks may not be properly accounted for in the modelling framework, particularly temperatures below 0°C. However, detrimental impacts of the NDD temperature range are unlikely given temperature behavior over the data set’s 60 years during the March-August growing season. It is shown that as temperatures increase towards a critical temperature threshold, biological development is increased both in magnitude and speed. However once this critical temperature threshold is surpassed there is a sharp and steep decline in not only plant growth but also final yield count. Further enforcing this point is that the decline in final yield count is much steeper than the growth in yield count under the upper bound normal degree days. Hence Heating Degree Days (HDD) are those temperatures beyond the critical threshold. These temperature thresholds vary between various crops, where corn for example has a marginally higher critical temperature than soybeans. These differences result in the inclusion of unique HDD and NDD for the two crops considered in this thesis. In summary both NDD and HDD are county level variables, annualized over each crops’ growing season, allowing for the measurement of non-
linear temperature effects. These result from the plant experiencing the full
distribution of temperatures, and HDD allow the production function to account for
the extent of the critically higher temperatures each crop was exposed to.

Fundamental to the research at hand is an understanding of how
temperature and precipitation interact with local yields, the above describes the
conceptual framework for capturing this relationship. However, this relationships’
explicit purpose is to capture ENSO effect on corn and soybean yields in the US. This
refers back to a primary goal of the research to define through which pathways
ENSO influences these staple crop yields. The pathway to be defined may very well
be unique to ENSO and via some unobservable pathway which can only be
attributed to the ENSO phenomenon, however it may be the case that ENSO impacts
various observable weather variables which in turn proceed to result in significant
impacts on yields in the United States.

The ENSO phenomenon is captured and expressed through a multitude of
proxies. In recent studies it has been shown that sea surface temperatures present
the highest correlation to magnitude of ENSO deviations, and this approach is
undertaken in this study. ENSO anomalies based on these surface temperatures
represent the phase and magnitude of ENSO events. It is key that El Nino and La
Nina as well as Neutral phases are not simply of a binary nature, meaning that for
example 1 represents El Nino year while 0 represents a La Nina year. This approach
does capture local weather changes under alternate phase definition, however it
imposes an assumption of El Nino and La Nina events being of identical magnitude
at each occurrence and furthermore fails to account for the magnitude of ENSO
cycles relative to such cycles in previous years. Hence in order to account for potential non-linearity in La Nina and El Nino events as well as magnitude of the ENSO cycle as a whole and each event individually, ENSO is expressed as anomalies from neutral years in a non-binary nature.

Having established a suitable proxy for the ENSO phenomenon, the proxy can be related to more than local yields directly. The modeling exercise will also include ENSO influence on local weather conditions to establish the relationship between ENSO and those forces most influencing yields (i.e. key weather variables), particularly those included in the initial production function. Finally ENSO influence on yield is to be measured in the production function controlling for such weather conditions as temperature and precipitation. This brings about the subject of indirect factors influencing yield levels, those which may be controlled for or rather under the control of farm level decision makers.

Due to price data limitations, as a result of the considerably large data set (both temporally and spatially) of yield and weather variables, the inclusion of a price or price proxy variable has been omitted. Lack of a price variable influences interpretation of indirect factors on yield levels. These may take many forms, particularly when considering yields as both biological output and economic output. Certainly, yield level would be dependent on expectations of future supply, potential risks associated with severe weather, expected demand, trends in costs of agricultural inputs, and a multitude of other factors affecting the pricing of the farm level commodity. The above yield effects results from a producers’ expectations and hence decisions at time of planting and field preparation prior to the growing
season. These include crop choice, crop mix, and use of farm level risk management tool such as fertilizers and pesticides.

With regard to addressing these concerns, the inclusion of a trend variable partially accounts for these factors. Furthermore, in this study yield is taken in the form of yield per planted acre. Meaning that increasing or decreasing acreage from year to year need not be addressed nor estimated in the model, additional benefits include model simplicity. By omitting changes in acreage and considering yield per acre rather than planted acre one is able to omit from the discussion various ambiguities regarding cross price elasticities between different crops, fertilizer use and or price. Hence model effectiveness is benefited by retaining degrees of freedom and avoiding over specification. In other words, limiting the number of parameters which provides greater degrees of freedom for estimating each parameters variability. This is particularly important when the number of observations may be limited. The subsequent sections describe the modeling framework used to achieve the isolation and identification of the desired relationship between ENSO and local yields via observable weather pathways as well as unique ENSO effects which may prove less tangible.

3.2 Empirical Framework

Capturing the unique effects of ENSO on yield as well as its relationship with key yield influencing weather variables requires a sophisticated process for measuring and more importantly estimating relationships among variables. In most basic
terms, the effect of a treatment should be captured via the research design, where observable factors are hypothesized to influence an outcome variable. In this fairly broad category of research design one is interested in determining causal inference (Hsiang, 2016). The simplest of which is the comparison of two samples and expose them to two different treatments. Considering the research problem this would be comparing ENSO in sample 1 to ENSO + Δ ENsO in sample 2. Evidently, recreating the complexity of local weather correlated with ENSO and its impacts on yields would be impossible to achieve in the required controlled environment. Hence, rather than extracting an unbiased estimator from such experiments an attractive alternative takes advantage of econometric analysis to approximate an estimator for the impact of a treatment (Hsiang, 2016). Through the mathematical process of regression analysis, one is able to identify which factors matter most and for whom they matter. Furthermore, one is able to identify those factors which do not carry weight in a relationship and which do, in other words identifying statistically significant parameters.

Building upon the simple example presented, there are significant challenges to dealing with complex variables such as weather which are closely linked with each other. These complexities arise from the measurement stage and follow through to the modeling and interpretation stage of the research design. Deriving an unbiased estimator from observations, however, is unlikely predominantly due omitted variable bias as well as the violation of the unit homogeneity assumption (Hsiang, 2016). The true parameters are unknown due to inherent model uncertainty as well as parameter uncertainty. Where, following the frame work
presented by Hsiang (2016), ENSO + Δ ENSO in the treatment group \( t \) to ENSO in control group \( t - 1 \), and \( x \) represents a vector of yield influencing factors which are observable yet not associated with ENSO directly. Finally, \( Y_t \) represent yield in each sample group. Thus, building a set of \( Y_t \) “such that” it is exposed to the treatment or standard treatment. Denoting ENSO as \( e \), the estimation of parameters takes place via approximation and statistical inference through regression analysis. In this manner, we build upon a simple multiple regression model to achieve parameters (beta coefficients) via the following generalized expression for ENSO impact on yields.

\[
\hat{\beta} = E[Y_{it}|e_{it} + \Delta e, x_{it}] - E[Y_{it-1}|e_{it-1}, x_{it-1}]
\]  

[3.3]

where, as before \( Y \) refers to yields and \( e \) refers to ENSO. Importantly, in the presentation of ENSO above, the variable functions as an aggregate of all the pathways through which ENSO affects production. Meaning that not only solely unique ENSO based influence, if any, is being measured but also influences through the observable pathways by which ENSO manifests its impact on production. These pathways for example include various forms of temperature and precipitation interactions with crops throughout their growth and development stages.

Following a regression based approach the research design takes on a hybrid of cross sectional analysis over various time periods, frequently referred to as fixed effects panel analysis. Cross sectional analysis as a research design is not appropriate due to strong reliance on the unit homogeneity assumption. Moreover,
cross-sectional analysis would not capture the fundamental variation in ENSO.
Papers such as Lobell and Asner (2003) take a long differences approach for crop
yields, while papers have considered climate influence over a wider economic
perspective such as Burke et al (2015) who measured the relationship between
climate and civil unrest.

The structure of the data used in the research takes the form of longitudinal
data or data collected from various points in time and across many units.
Considering this structure where points in time are annual data points and units
refer to major corn and soybean producing counties in the United States, the data
structure is referred to as panel data. This refers to many observations across many
units allowing for an effective measurement of change over time, in particular
fluctuations over time. As all panel data sets, key features include the use of FIPS
(Federal Information Processing Standards) code as time-invariant unit or location
identifiers (county identifier), time varying variables referencing key explanatory
variables for yield development discussed in the preceding section, and finally a
time indicator which throughout the research takes annual form regardless of being
averaged over a total 12 months or over the crops growing season months. These
basic characteristics form the panel data set used within the modeling framework.
Raw data however undergoes a considerable degree of manipulation.

Taking a basic econometric model where corn or soybean yield is simply a
function of a vector of weather variables including an ENSO proxy for peak ENSO
period events. An econometric regression of this sort clearly may suffer from
endogeneity bias. Endogeneity bias occurs when there is correlation between one of
the explanatory variables (vector of weather variables) and the error term. This is a concern within the modelling framework, as it may be the case that variables correlated with ENSO could be omitted. Soil type, for example, may affect this relationship and through the use of fixed effects the issue of omitting a country specific variable can be mitigated. One way to deal with inherent endogeneity bias within the model is the inclusion of instrumental variables (IV), or similarly and more effectively construct a fixed effects panel data model. One must note however that fixed effects is not sufficient for all forms of endogeneity bias, such as simultaneity, time varying measurement errors and unobserved effects. In essence, all time constant effects are removed, thus using only within unit change rather than between unit change as well.

The modeling structure in this thesis takes a fixed effects panel model approach for various reasons. Firstly, considering ENSO effect as well as local weather influence over yields as a control, the modeling framework aims to measure and quantify the influence of those explanatory variables over time. Fixed effects allow for controlling for characteristics within each county or unit which may influence the impact that the explanatory variables may exert on yield. In this way, fixed effects accounts for correlation between the error term and the explanatory variable, thus providing the net effect of predictors without the need for including IVs limiting the degrees of freedom (df) and potentially causing issues with over specification of the model. This is of particular value since the model specification estimates county specific effects, hence all the parameters within the model are assumed heterogeneous across counties. A model specification was tested for state
specific effects; however, results were inclusive and provided no statistical significance. Issues to consider here are those related to temporal as well as spatial dependence, as some individual characteristics (ex. farm management techniques or soil type) may be shared among counties.

As noted previously, the fixed effect panel model specifications below assume unique county characteristics which are invariant over time. First, we model corn and soybean yield per acre planted for a mix of irrigated and non-irrigated farm land, $Y_{it}$ for $t$ the relevant time series, $i$ counties and $j^{th}$ crop (in what follows, the crop-specific subscript is omitted for simplicity). $Y_{it}$ is modeled directly as a function of ENSO over its peak period in a fixed effects panel data regression,

$$Y_{it} = \beta_{1i}e_t1(e_t \ge 0) + \beta_{2i}e_t1(e_t < 0) + \mu_i + \theta_i t + \epsilon_{it}$$  \[3.4\]

where $\mu_i$ is a county-specific fixed effect, and $\theta_i$ is a county-specific trend parameter. This specification (including [3.5] and [3.6]) accounts for heterogeneity in the panel with regard to ENSO effect by deriving ENSO estimates, $\beta_i$, for each factor (county) specified, in contrast to running a series of independent county level OLS regressions. $e_t$ can be a scalar or a vector of ENSO variables (e.g., La Nina and El Nino), where $1(.)$ is an indicator function that takes value of 1 if the condition within the parentheses is satisfied, and 0 otherwise. With regard to model specification, the key characteristic lies in the error term, $\epsilon_{it}$, varying non-stochastically over time and county. County specific estimates are achieved by the ENSO variable interacting with a factor, namely county fixed effects.
The above model captures a variety of ENSO related yield effects, which are hypothesized to decrease once variables correlated with ENSO are included in the regression. As stated previously the ENSO variable is not included in the common format of a binary variable, instead it takes into account the magnitude of positive (El Nino) and negative (La Nina) ONI (Oceanic Nino Index) deviations.

Controlling for temperature, as a proven weather variable most damaging to corn and soybean yields during their respective growing seasons, is included in the model taking into account non-linear temperature effects. This has been discussed in detail by Schlenker and Roberts (2009) with a clear argument stating that temperature effects on yield are severely underestimated if a monthly or growing season average is only taken into consideration. Structuring the model for temperature non-linearity allows for control of the variability of temperatures in the data. This model includes two individual temperature variables,

\[
Y_{it} = \beta_1 e_t I(e_t \geq 0) + \beta_2 e_t I(e_t < 0) + \delta T_{it} + \mu_i + \theta_i t + \epsilon_{it}
\]  

[3.5]

where \(\beta_i, \theta_i, \epsilon_{it}\) remain the same as before however \(\beta_i\) will lose a degree of its explanatory power to the new temperature explanatory variables. This is due to ENSO, at least to some degree, channeling its effects via observable and measurable local temperature conditions. These conditions, defined by \(T_{it}\), are presented as a vector of temperature variables \(NDD_{it}\) and \(HDD_{it}\) (Normal and Heating Degree Days, respectively). Furthermore, as temperature is a key yield promoter and yield suppressant the impact of controlling for these factors is expected to be significant.
The temperature values below the critical threshold and those above the critical threshold, $NDD_t$ and $HDD_t$ respectively, are acquired via degree days and are time as well as region dependent. Temperature variables are aggregated over the growing season via degree days and their respective impact on county level yield is captured through the parameter $\delta$ net of any vectors via which ENSO impacts yields not attributed to temperature. In turn, by introducing the temperature variables in the equation, the interpretation of the parameters associated with ENSO has changed. Now they measure the production effect of ENSO net of the temperature pathway.

Building upon this model a complete weather vector is included by the additional precipitation explanatory variable.

\[
Y_{it} = \beta_1 e_t I(e_t \geq 0) + \beta_2 e_t I(e_t < 0) + \varphi_i W_{it} + \mu_i + \theta_i t + \epsilon_{it} \tag{3.6}
\]

Here $W_{it}$ presents a vector of weather variables including temperature and additionally county level precipitation, aggregated over the respective annual growing season as a 5-month mean. Finally, $\beta_i$ and $\varphi_i$ represent parameters to be estimated; again $\epsilon_{it}$ is an independent and identically distributed error term.

To complete the modeling framework ENSO must be included in two manners. First, as seen above, ENSO is included as an explanatory variable. Second, ENSO’s inclusion is vital for deriving parameter estimates to ascertain which observable weather pathways are most closely linked to the ENSO phenomenon. This is of particular relevance in capturing the intuition behind the links between
the medium frequency ENSO and high frequency local weather conditions responsible for yield growth or suppression. Additionally, capturing the above intuition will be integral in a forecasting framework. Particular emphasis was placed on the relationship between ENSO and local temperatures given as

\[ W_{it} = \beta_1 e_t \mathbb{I}(e_t \geq 0) + \beta_2 e_t \mathbb{I}(e_t < 0) + \mu_t + \theta_t + \epsilon_{it} \]  

[3.7]

Where \( W_t \) is a vector of weather variables (i.e., \( NDD_{it}, HDD_{it}, \) and/or \( Prc_{it} \)), \( \beta_i \) refers to temperature and precipitation impacts of ENSO under alternate phase definition, El Nino and La Nina, for the normal temperatures, non-critical range of temperatures, and precipitation each crop is exposed to.

### 3.3 Data

Yield data was acquired in annual form from the USDA National Agricultural Statistics Service for both corn and soybeans. For corn, corn grain yield measured in bushels per acre from mixed irrigated and non-irrigated farm land. Soybean data was taken as bushels per acre from mixed irrigated and non-irrigated sources. Yield variables for both corn and soybeans were included in the various models as annual county level observations. Furthermore, for additional robustness in the model only those counties with complete data sets were used over the 59-year period from 1951 – 2010. Filtering for only complete yield resulted in matching 900 major corn and soybean producing counties to a comprehensive weather data set. Both soybean
and corn yields display positive upward trends in yield across the years included in the set, with corn yields displaying greater variability yet a stronger upwards trend. For both crops, outliers are concentrated at the lower tail of the distribution, with only a handful of outliers presenting dramatically outperforming yield levels. This produced an elongated asymmetric tail for both distributions. Yield values across the 59 years over which data has been acquired can be neatly summarized via box and whisker plots grouped in 5 year intervals (Fig 2.1, Fig 2.2).

![Figure 3.1: Distribution of corn log-yields across counties and over time](image)

Corn Yield Producing Counties (N=900)

Annual log(Yield) (Bushels per Acre)
Soybean data set yields present various outliers in the upper tail while the upper tail of corn yield outliers on the other hand are far less sporadic and contained within the upper quartiles.\(^1\) Keeping in the difference in scale between crops, several other notable characteristics may be discussed. These include the apparent widening (as time progresses) of the interquartile range, as a measure of spread holding 50% of the full range of yield values regardless of the shape of distribution. Yields as expected were not normally distributed in raw form, hence the data undergoes a logarithmic transformation (as seen in figures 1 and 2) for each observation, an action taken in order to linearize the data set, resulting in the

---

\(^1\) Noticeably for soybean yields there was an outlier associated with the year 1973 (raw value 93). In the interest of clarity an outlier with a yield of 230 (raw value, bushels per acre) registered in Lauderdale, Alabama (1077) has been removed from the data set prior to producing the above summary.
natural logarithm of the original observation. Upward movement in yields over the 
years can be seen by the steadily rising median values, with special interest taken 
when observing those years where particularly large changes in yield levels can be 
seen.

A rich weather data set was provided by Wolfarm Schlenker of Columbia 
University. The data set provides county level yields across the continental United 
States between 1950 and 2010, including precipitation, average temperature, 
minimum temperature, maximum temperature, and a full range of degree day 
temperatures. Degree day weather data measures the length of exposure within 
each given month of the year that a particular county was exposed to temperatures 
above a certain value. The set offers a range from 0 to 40 °C, comfortably 
encompassing the critical thresholds for both corn and soybeans. To illustrate, a 
day40 observation for a given county in a given month and year measures the 
length of exposure that the county experienced temperatures above 40 °C. For this 
reason a range of degree days ending at 40 degrees is sufficient for illustrating 
extreme heat exposure. Even if there were far higher temperatures to which the 
cy County was exposed these would be captured by the final day term in the data set 
and the length of exposure would not warrant more precise temperature 
measurements above 40 (as the majority of day40 measures are zero throughout 
the major crop producing counties).

As discussed in some detail in the conceptual framework temperature effects 
are assumed to be non-linear and therefore modeled in a manner to take advantage 
of this relationship. Two-degree day temperature variables were included in each
model allowing for degree day variables to enter the appropriate equations in a piecewise linear form, producing a kink at the each crops respective temperature threshold. This relationship was accounted for by estimating both Normal Degree Days (NDD) and Heating Degree Days (HDD) within the modeling framework. These can be seen in the following representations of their respective distributions (Fig 2.3).

![NDD and HDD Distributions](image)

**Figure 3.3: Degree Day Distributions**

The NDD distribution represent the length of exposure to normal and beneficial temperature ranges for plant growth, while HDD represent the length of exposure to temperatures detrimental to plant development and shown to cause yield suppression. The critical temperature threshold setting the HDD range lower limit is 30 degrees Celsius and 29 degrees Celsius for corn and soybeans, respectively. NDD
as expected fit the bell curve of a normal distribution rather well, while HDD represent the upper tail of the full temperature distribution and therefore displays steadily declining levels of heat exposure. This is due to the simple fact the HDD represent those temperatures at which yield has been shown to decrease dramatically and for the majority of corn crops results in an interruption (if not end) to the production of biological matter.

Precipitation data was given as a monthly total, and included in the respective models as total precipitation for each crop growing season. It is worth mentioning that total precipitation as a variable in the modeling framework, for a given county, would be identical for both corn and soybeans due to both crops experiencing parallel growing seasons. This however would not be the case for the two temperature variables, NDD and HDD, as the crops have slightly different critical temperatures for biological development. Further implications and interpretation of results will return to this difference in threshold, in particular discussing the considerably minimal difference of 1 degree Celsius.

ENSO may be gathered through various proxies composed of two main categories, atmospheric pressures and sea surface temperatures. Both metrics provide valuable insight and offer unique benefits to the alternate phase classifications system. The research at hands follows in the steps of previous papers by selecting sea surface temperatures, or SST’s, as a measure of ENSO phase definition as well as magnitude (Magrin et al, 2003; Dawe et al, 2009). Sea surface temperatures have proven to provide more accurate measurements, in the sense of superior accuracy in measuring water temperatures across the Pacific Ocean has
been displayed in comparison to atmospheric pressure changes. Furthermore, recent literature has shown, that SST readings provide equal if not greater ENSO phase definition than atmospheric readings, with the nino3.4 index proving among the most sensitive El Nino identifiers (Hanley et al, 2003). Finally, it is of considerable value for the field as well as this research specifically to use the same proxy for ENSO for the benefit of comparing study results and implications. This is of great importance particularly when considering ENSO and agricultural production as there is a growing expanse of related literature delving into increasingly precise foci of the subject. Sea surface temperature readings are taken from the Nino3.4 (5S-5N and 170-120W) region of the pacific basin as average monthly temperature readings. The ONI index classifies a +0.5°C ONI as El Nino and -0.5°C ONI as La Nina, shown in figure 3.4. Note that anomalies in this study are considered as any positive or negative deviation. The monthly data was aggregated and included in the modeling framework with emphasis on the peak ENSO period, the period during which defining SST as well as atmospheric pressure change readings take place. Defining in the sense of phase definition as well as most telling as to the magnitude of the ENSO phase and its corresponding impacts globally. This peak period that spans across November, December, January was transformed into the ENSO proxy variable via a simple average. Thus, for a crop year t, the adequate ENSO measure was obtained by averaging SST anomalies in November and December of year t-1, and January of year t. Which, in effect, is the December Oceanic Nino Index (ONI). While the raw data are taken as anomaly data, hence readings of zero coincide with no anomalous temperature changes signaling a
neutral year. Positive anomalies, or equivalently warmer temperatures, signal an El Nino phase and negative anomalies signal a La Nina Phase. In summary, Fig 2.4, following this phase definition, shows that the most recent ENSO anomaly was a relatively strong El Nino. From the 1950-2010 data set the El Nino phase of greatest magnitude took place between 1997-1998, comparable to the most recently observed El Nino event of 2015.

Figure 3.4: Annual December ONI fluctuations and magnitudes (source: NOAA)

Utilizing this type of data allows the model to not only consider ENSO as a binary variable representing one of the two phases, but also takes into consideration the magnitude of each phase. Capturing the magnitude of phases and their anomalies
allows the models to account for the non-linear effects of ENSO, as recent literature on the subject has led various researches to believe.

3.4 Results and Interpretation

The results describe a spatially diverse ENSO effect, presenting both seemingly sporadic influence as well as clear patterns in particular regions. On the other hand, statistical significance on the county level is seldom observed. Characteristic clustering of significance is visible among a handful corn counties (under 25% of counties studied), and no discernible pattern is visible with regard to soybean. Nonetheless the results support previous findings on the directions of ENSO effect. As well as presenting a contribution in an identifiable link between ENSO effect and significance of that effect with regard to temperature influence over corn yields.

The discrepancy in ENSO effects may be due to a variety of factors, some of which should be identified to some degree of certainty through model specification. As discussed above, corn and soybean yields may be modelled directly [Model 3.4] on the constructed ENSO proxy variable. Through this model specification, ENSO effects were not represented in isolation net of any observable and predefined weather pathways. In this model specification ENSO effects on yields would be expected to be strong as well as significant. This strength of ENSO influence stems from it acting as a “catchall” variable, meaning that a variety of endogenous variables including weather as well as producer decisions. Therefore, the inclusion of HDD, NDD, and Prc was expected to detract from ENSO’s magnitude while maintaining that these variables were also influenced by ENSO themselves. From
this initial modeling exercise, further models were introduced and the results provided further detail in dissecting the relationship between ENSO and yields. Of particular interest is the hypothesis that ENSO produces considerable yield impacts, having controlled for observable weather conditions. In this case a specific ENSO impact would be composed of a multitude of potentially heterogeneous factors, too negligible on their own to warrant individual modeling, yet significant in the aggregate. These may include storms, hail, extreme winds and other unaccounted-for variables such as humidity. More importantly this aggregate of individual factors would be closely correlated with ENSO phases and their respective magnitudes.

Taking the direct effect of ENSO on yields, both crops experience greater and more geographically uniform yield losses under the El Nino phase than under the La Nina phase. Furthermore, the losses under El Nino for both corn and soybean are lower in magnitude than those experienced under La Nina in particular areas. There can be several explanations for such an initial interpretation and generalized theme in the ENSO effect. Primarily such results would point towards relatively more detrimental yet uniform climatic conditions for both crops in their key producing regions. Importantly, although key corn and soybean producing regions are not identical, for all intents and purposes both crops cover a shared area allowing for additional cross-crop comparisons. By comparison La Nina years appear to have varied impacts on yields, including yield growth as well as suppression, and additionally these impacts come in a wider range of magnitude than those in El Nino years.
Beginning with corn, where the closest interaction between ENSO and yields [Model 3.4] is seen, there is support for observing that La Nina phases result in stronger positive and negative yield impacts. Taking a closer look at a La Nina phase a few key features come to mind immediately. First, coastal regions as well as the southern Appalachian regions generally experience negative yield effects. As one moves further east towards the coast as well as further south within the sample this negative impact becomes far stronger. A La Nina phase may result in a minimum of negative yield effect of 6% to a 22%. Regions experiencing the upper tail of the negative influence range have a characteristic clustering effect.

Figure 3.5: La Nina (ONI decrease) impact on corn yields

Note: Model 3.4 results are presented, where yield is modeled directly as a function of ENSO.

Similar conclusions can be drawn when observing the positive impacts of La Nina on yields. Positive impacts are associated with a different region, and appear to
not take place in any eastern or Appalachian regions. Moving from these negatively impacted regions north-west they dissipate into neutral influences and gradually transitioning into a relatively weak positive impact of +6% change in yields for much of the north and northwestern segment of the corn belt, with Nebraska and South Dakota displaying geographical clusters of up to +20% yield promotion. Clustering is again present and displays pockets of strong positive impact on yields. Additionally, the regions of strong positive yield impact are slightly larger and, more importantly, larger players in the total production of corn in the US. However, it can be observed that when taking the Corn Belt as a whole, there is a considerably wide range of variation. As the eastern tier of the Corn Belt experiences negative growth, moving west transitions into positive growth of 20% in certain areas. A transition south to north, from the northern border of Kentucky towards the north of Indiana, Illinois, and Ohio displays positive yield effects. This observation may be partially associated with superior soil fertility in regions above the northern most point of Kentucky.

A far less discernable pattern emerges when measuring the impact of El Nino on corn yields. As before, beginning with the coastal and eastern regions clearly there is predominantly negative El Nino influence on yields, although small pockets of neutral to positive influence exist. However, with the exception of a small geographical cluster of roughly +6% yield growth in Georgia, these pockets of positive yield growth are contained within one or two counties at most. Negative effects are present in larger clusters, such as those observed in North Carolina near
the border with Virginia. For this particular cluster negative influence of El Nino on yields was around -18%.

![Map of US Corn Yield with El Nino Influence](image)

**Figure 3.6: El Nino (ONI increase) impact on corn yields**

*Note: Model 3.4 results are presented, where yield is modeled directly as a function of ENSO.*

Moving north-west the interpretation of El Nino impact does not change dramatically, as it had previously under the La Nina phase definition. A few key features do stand out however. First and foremost, geographical clustering of pockets of strong negative yield impact are much larger and dramatic, however only in comparison to the eastern and coastal regions under the same phase definition. Much further west a few smaller and one large cluster of positive yield influence is visible however the magnitude of positive influence is negligible. The magnitude of negative clusters in the north-west region is certainly not negligibly. Finally, the impact of an El Nino phase on corn can be summarized as predominantly negative
with various clusters displaying strong negative impact, only a few negligible pockets of positive impact can be observed. This is in strong contrast to La Nina phase on corn yields, where there is a great deal of variation as well as clustering of strong positive and negative effects.

Soybean yields on the other hand under both El Nino and La Nina phase definition lack the geographical clustering characteristics of corn, such as cluster size and defined borders. This key difference between corn and soybean offers interesting insights, primarily with respect to ENSO interaction with local weather and each crops’ unique requirement for spurring yield development.

Again, La Nina results in creating geographical clusters of positive influence stronger than those resulting from an El Nino phase. Although not nearly as pronounced there are some similarities between soybean response to La Nina and corn response to La Nina. Soybeans display negative yield effects on the eastern and coastal regions and increasingly positive effects on soybean yields as one progresses not only north but in most particular west. Much of the so called Corn Belt appears to have neutral to slightly positive soybean yield impact under La Nina. The similarity, although slight, between corn and soybean yields under a La Nina phase is of considerable interests.
Figure 3.7: La Nina (ONI decrease) impact on soybean yields

Note: Model 3.4 results are presented, where yield is modeled directly as a function of ENSO.

El Nino further defines the trend of ENSO influence on soybeans, where by comparison to corn yields patterns are less discernable. Positive impacts under either phase definition are rare and, if occurring, not of a noticeable magnitude. Finally, the overall magnitude of ENSO impact is not amplified as it is for corn yields. Noticeably the counties where strong negative El Nino impact is observed, values are scattered and present a weaker tendency towards geographic clustering. The weaker El Nino impact of +/- 7% displays a clearer clustering characteristic, particularly in Illinois and northern Kansas.
Similar to corn yields, El Nino resulted in a greater quantity of negative yield impacts than La Nina. An exception arises in the eastern and coastal soybean producing counties which remain predominantly negative under La Nina to even stronger negative effects during El Nino (Fig 3.8). Furthermore soybeans lack a dramatic change in ENSO influence and its “geographical pattern” between La Nina and El Nino, as it did for corn. In the case of soybean yields El Nino mimics the La Nina pattern of influence, which a key characteristic in that positive yield impacts are far more prevalent and have replaced weak positive yield impacts under La Nina with stronger negative impacts. Weak impacts is a fitting summary to soybean yields under El Nino where percentage change in yield ranges from roughly -8% to +10% change in yield levels. While under a La Nina phase, soybean yield influence ranges primarily between -22% and +9% change in yield levels, in response to a 1 degree warming of sea surface temperature anomalies under El Nino.
Figure 3.9: Histogram of %Δ in yield due to 1°C ONI deviation estimates for corn (top) and soybean (bottom)

Note: Model 3.4 results are presented, where yield is modeled directly as a function of ENSO.
Summarizing the ENSO effect discussed thus far, the results convey a spatially diverse representation of ENSO demonstrating varying influence over geographical locations. Such a relationship between ENSO and corn yields is particularly visible through clustering of ENSO intensity and significance over crop yields. Clustering is evident in soybean yields, yet less pronounced. Clustering can be thought of as regions of homogenous characteristics – at the least homogenous behavior in response to ENSO. As one moves west across the Corn Belt the intensity of ENSO influence changes as discussed above. These results are aligned with previous studies considering the spatial nature of ENSO influence (e.g., Tack and Ubilava, 2013b). The negligible ENSO influence over soybeans should be primarily attributed to the crops temperature resistance and mix of necessary factors for yield growth, likewise the mix of yield suppressing factors is equally important. However as noted above, there is a similarity in results for both corn and soybeans under the El Nino phase definition. Such a result requires closer attention as to the weather trends observed during a warming of sea surface temperatures. Cooling temperatures create a variable Pacific Jetstream, as opposed to El Nino marking a clearly definable change in Jetstream behavior. The variable Jetstream, under La Nina, creates warm temperatures around much of the east coast, with wet conditions as one moves north and northwest. As one moves south towards Florida, crop growing regions experience dry conditions. In this way, La Nina results in a delicate mix of weather conditions around the area of interest. These effects, are in contrast to the climatic conditions observed over these major crop producing states during El Nino; where dry and warm conditions are experienced in the north and
north west. The east coast, however, does not experience dramatic changes. These conditions coincide with the results above stating that under El Nino these regions experience predominately negative yield effects however overall influence is scattered and weak. It may be that the mix of climatic conditions wet, warm, dry as well as a strengthened polar set stream moving across much the Midwestern US results in varied and noticeable effects under La Nina. Although unclear to some degree, the discrepancy between El Nino and La Nina influence way direct towards the notion of a weaker magnitude of El Nino for key crop developing variables in these key production regions. Overall ENSO effect for all corn (soybean) counties is -1.8541(+0.3349) % change in yield, as seen in Figure 3.9.

Before discussing in which regions ENSO effects are statistically significant it is of interest to see how ENSO influences temperatures, in particular those above the designated thresholds for each crop. One would expect similar pattern and pronounced influence of ENSO on the higher ranges of temperatures as these have been shown to influence crop growth and ultimately final yield levels to the greatest extent. Beginning with corn it is important to be reminded of the clear impacts and pattern presented above for corn yields under La Nina, where predominately negative effects were visible in the east and positive effects as one moves west and North West. In the impact of a La Nina episode is therefore of significant interest when considering HDD, or the higher temperature threshold. This can be seen in Figure 3.10.
At first glance the similarity in the patterns are apparent, especially under a La Nina phase definition. These results support the starting hypothesis of the modelling section, with ENSO influencing yields mainly via extreme temperature. In turn,
supported by numerous previous research most significant of which is the Schlenker and Roberts (2008) paper on non-linear temperature effects.

Moving beyond the visible resemblance of the spatial distribution of ENSO impact on yields and HDD (for both phase definitions), it is clear that the negative impacts of La Nina on the length of time that corn yields are exposed to temperatures above 30 degrees Celsius (Figure 3.10) result in positive yield effects (Figure 3.5). In other terms La Nina reduces the negative impacts on yield by negatively affecting HDD in certain regions. Negative impacts of extreme temperatures experienced by corn over its growing season go as high as a 15% decrease in crop exposure to Heating Degree Days due to a 1 degree variation in the La Nina phase. It is of interest that this relatively strong effect in this region does indeed result in positive yield effects. However, it is noteworthy that the negative effects on extreme heat exposure for the majority of the major crop producing regions (those further west) is milder, thus milder in reducing those high temperatures. Yet the effect of this milder reduction of these temperatures results in comparatively weak positive yield effects, than the relatively strongly extreme temperature promoting effects found on the east coast. This further emphasizes corn’s poor adaption and resilience to lengthy exposure to temperatures above the 30 degree threshold. Finally, under El Nino a similarity in patterns between ENSO-Yield and ENSO-HDD is visible, and predominantly results in promoting HDD’s across the major crop producing regions. Thus, it follows that under El Nino we have observed yield suppression rather than yield growth across the vast majority of counties.
For corn the results promote the statement where extreme temperatures influence yields to the greatest degree. For soybean, a similar story emerged however as seen with the ENSO to yield relationship the impacts were weaker. More importantly the impacts of ENSO influence over HDD do not carry over soybean yield suppression or growth on the same scale as it did with corn. This is primarily due to differences in terms of plant growth requirements and greater resilience to extreme temperatures than corn.
Figure 3.11: El Nino (ONI increase) and La Nina (ONI decrease) influence over Heating Degree Days – Soybean

Note: Model 3.7 results are presented, where HDD is modeled directly as a function of ENSO.

Again, positive impact of La Nina on HDD is translated to negative (although weak) effects of La Nina on soybean yields in Figure 3.11. This is particularly evident for eastern counties of Arkansas, southern Illinois and western Tennessee. In these regions exposure to extreme heat is increased by up to 20% and maintained
at roughly a 13% decrease, where highest positive impact on temperature during La Nina is seen in Arkansas. Looking back to ENSO influence on soybean yields in these states in particular yield growth is suppressed under La Nina however to a negligible degree, interestingly in the coastal regions of soybean production experience weak promotion in extreme temperature as a result of La Nina results in more pronounced yield suppression. These variations are most likely due to factors outside of the model parameters and may include both local variables as well as other low frequency weather phenomena which coastal regions may have greater exposure to. Furthermore, although quite weak, suppression of HDD under La Nina results to stronger positive impacts of La Nina on soy yields. With regard to the El Nino phase the story remain parallel to that of corn, where extreme temperatures are on the rise during El Nino throughout all the major soy producing counties and as a result translate to nearly uniform yield suppression of soy crop during an El Nino episode.

Before moving forward, presenting the model results for NDD and HDD on yields, may hold further valuable information to support the interpretation of the results thus far. As stated, first such a discussion will shed light on temperature effects on corn, offering yet another dimension of this relationship. Second the results will put greater scrutiny on differences in temperature threshold between corn and soybeans.
Figure 3.12: Heating Degree Days and Normal Degree Day influence on corn (L) and soybean (R) yields: HDD-Top; NDD-Bottom, controlling for ENSO and precipitation

Note: Model 3.6 results are presented, where yield is modeled as a function of HDD, NDD, ENSO and Precipitation.

The effect of NDD and HDD net of ENSO and precipitation are described in Fig. 3.12. The results offer a reassuring picture where temperatures below the critical threshold for both crops unanimously for all counties offer positive yield effects, particularly so for corn. While temperatures, or rather exposure to temperatures, above the critical threshold offer varied influence. It is important to keep in mind that controlling for ENSO may have captured some key influence from
what would otherwise be attributed solely to HDD. Nonetheless there a strong negative yield impacts visible in the northern counties of the Midwest. Moreover the pockets of slight (yet negative) temperature impact closely mirror those seen when modeling ENSO effect directly on yield from both crops. This is particularly true under the La Nina phase definition. In the following section many of these pockets shared by both ENSO direct effect on yields as well as HDD effect net of ENSO and precipitation present statistically significant results of ENSO influence at the P<0.05.

On the other hand, results appear to show ENSO influence over precipitation at a high magnitude, however statistical significance is limited to a single cluster of counties centered around Minnesota and northern Iowa.
Figure 3.1: El Nino (ONI increase) and La Nina (ONI decrease) influence over Prc:

Corn (R) Soybean (L)

Note: Model 3.7 results are presented, where precipitation is modeled as a function of ENSO

These clusters are in regions of strong reduction in precipitation, as can be seen in Figure 3.15. Interestingly both El Nino (warm) and La Nina (cold) phases offer similar results, and more importantly do not coincide with direct ENSO effect on yield [Model 3.4]. On the contrary regions with strong (and significant) precipitation reduction under La Nina, display significant yield promotion for both corn and soybeans (Figure 3.16). Unlike HDDs, ENSO impact on precipitation does not mirror ENSO impact on yields, instead precipitation results appear to run counter to ENSO
impact on yield. These results display that, under the model framework presented, ENSO effect on yields is predominantly characterized by its influence over HDDs.

Finally, returning to the direct relationship of ENSO on corn and soy yields [Model 3.4], it is important to identify those counties which hold a statistically significant link to ENSO fluctuations under each phase. The spread of significant counties and clustering are key factors which provide further insight into ENSO influence over yield as well as through which pathway it is linked to yields most directly. In Fig. 3.13 total corn county parameters are summarized under both phase definitions by state, while in Fig. 3.14 soybean county parameters are compared. Subsequently, Fig. 3.15 maps values of significant county estimates for both crops.
Figure 3.14: La Nina and El Nino influence in corn producing counties
Figure 3.15: La Nina and El Nino influence in soybean producing counties
Figure 3.16: The ENSO effect on yields U.S. corn (L) and soybean (R)-producing regions. From top-bottom: Cold Phase-Warm Phase. Highlighted counties represent statistical significance at $p<0.05$

Note: Model 3.4 results are presented, where yield is modeled directly as a function of ENSO.

For both crops La Nina phases present the most significant link (for most counties) between ENSO and yield. For soybean, there is no clear trend of clustering with the exception that under La Nina those regions with positive yield effects on the outer rim of the area under study retain statistical significance. This observation is also true of corn yields where clustering is far more evident, however La Nina has at the least two major clusters unique to its self which do not occur under the El
Nino Phase definition. Clusters of significant La Nina effect occur in Nebraska and South Dakota, where yield promotion occurs. And large negative La Nina influence clusters are seen in North Carolina and South Carolina, mainly in South Carolina and spilling over the state boundary into North Carolina. Furthermore, it can be said that those counties experiencing La Nina and El Nino effects of the greatest magnitude (both positive and negative) offer the strongest statistical significance. Tables 3.1 and 3.2 display a summary of statistically significant counties for corn and soybean by state derived from the basic model. Corn and soybean yield’s relationship with La Nina results in the largest number of statistically significant counties, 20% vs. 10% and 17.5% vs. 4% of available counties are significant under La Nina for soybeans and corn respectively. Furthermore, corn yields 154 significant counties out of 881 and soybean yield 127 out of 668 counties. Note that counties may over-lap in significance, meaning some counties may experience statistical significance under both La Nina and El Nino.

As with any form of hypothesis testing, no hypothesis can be rejected or accepted with unequivocal certainty. Meaning that the null hypothesis may be incorrectly rejected resulting in biased estimates and erroneous statistical significance (Type I error). Such a form of Type I errors should be considered when reviewing the results, particularly due to the complexity of weather and plant growth interactions. In this way, spurious correlation may arise through the omission of an unidentified variable correlated with both yields and ENSO. However, it is unlikely for such a variable to cause both ENSO and yields, hence the direction of causality moves from ENSO and through the omitted variable. In this
manner, omitted variable bias in unlikely since such a variable would act as a mediator rather than a confounder. Furthermore, the model specifications presented here aim to identify unique ENSO influence, if present, beyond the fundamental yield influencing weather variables (precipitation and temperature).
### Table 3.1 Basic model significant corn counties at p<0.05

<table>
<thead>
<tr>
<th>State</th>
<th># Available Counties</th>
<th>% Significant Counties (Under El Nino and La Nina)</th>
<th>Average Estimate %</th>
<th>Phase</th>
<th>% Significant Counties</th>
<th>Average Estimate %</th>
<th>Significant counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>5</td>
<td>30%</td>
<td>-17.3%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>60%</td>
<td>-17.3%</td>
<td>3</td>
</tr>
<tr>
<td>Illinois</td>
<td>97</td>
<td>6.2%</td>
<td>-12.6%</td>
<td>El Nino</td>
<td>12.4%</td>
<td>-12.6%</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Indiana</td>
<td>86</td>
<td>0.6%</td>
<td>-11.3%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>1%</td>
<td>-11.3%</td>
<td>1</td>
</tr>
<tr>
<td>Iowa</td>
<td>99</td>
<td>10.6%</td>
<td>-8%</td>
<td>El Nino</td>
<td>15.2%</td>
<td>-16.2%</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>6%</td>
<td>12.3%</td>
<td>6</td>
</tr>
<tr>
<td>Kentucky</td>
<td>69</td>
<td>8%</td>
<td>-12.9%</td>
<td>El Nino</td>
<td>5.8%</td>
<td>-12.2%</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>10.1%</td>
<td>12%</td>
<td>7</td>
</tr>
<tr>
<td>Maryland</td>
<td>16</td>
<td>6.3%</td>
<td>-11.1%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>12.5%</td>
<td>-11.1%</td>
<td>2</td>
</tr>
<tr>
<td>Michigan</td>
<td>41</td>
<td>3.7%</td>
<td>-5.2%</td>
<td>El Nino</td>
<td>4.9%</td>
<td>-13.5%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>2.4%</td>
<td>11.5%</td>
<td>1</td>
</tr>
<tr>
<td>Minnesota</td>
<td>58</td>
<td>0.9%</td>
<td>11.5%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>1.7%</td>
<td>11.5%</td>
<td>1</td>
</tr>
<tr>
<td>Missouri</td>
<td>77</td>
<td>11%</td>
<td>-14%</td>
<td>El Nino</td>
<td>15.6%</td>
<td>-14.3%</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>6.5%</td>
<td>-12.9%</td>
<td>5</td>
</tr>
<tr>
<td>Nebraska</td>
<td>82</td>
<td>19.5%</td>
<td>16.2%</td>
<td>El Nino</td>
<td>1.2%</td>
<td>-14.3%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>37.8%</td>
<td>17.2%</td>
<td>31</td>
</tr>
<tr>
<td>North Carolina</td>
<td>58</td>
<td>12.9%</td>
<td>-13.8%</td>
<td>El Nino</td>
<td>3.5%</td>
<td>-12.5%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>22.4%</td>
<td>14%</td>
<td>13</td>
</tr>
<tr>
<td>Ohio</td>
<td>68</td>
<td>0.7%</td>
<td>-12.4%</td>
<td>El Nino</td>
<td>1.5%</td>
<td>-12.4%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>South Carolina</td>
<td>22</td>
<td>34.1%</td>
<td>-15.2%</td>
<td>El Nino</td>
<td>4.6%</td>
<td>-13.3%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>63.6%</td>
<td>-15.4%</td>
<td>15</td>
</tr>
<tr>
<td>South Dakota</td>
<td>46</td>
<td>19.6%</td>
<td>12.8%</td>
<td>El Nino</td>
<td>2.2%</td>
<td>-13.7%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>37%</td>
<td>14.3%</td>
<td>17</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>57</td>
<td>0.9%</td>
<td>11.2%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>1.8%</td>
<td>11.2%</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>881</td>
<td>11%</td>
<td>-5.5%</td>
<td>El Nino</td>
<td>4%</td>
<td>-13%</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>17.5%</td>
<td>0.6%</td>
<td>103</td>
</tr>
</tbody>
</table>

*Note: Average estimate % refers to the percentage change in yield as a result of a 1°C deviation in ONI. Column 4 displays state averages for overall ENSO effect, while column 7 displays El Nino and La Nina effect individually.*
Table 3. 2 Basic model significant soybean counties at p<0.05

Note: Average estimate % refers to the percentage change in yield as a result of a 1°C deviation in ONI. Column 4 displays state averages for overall ENSO effect, while column 7 displays El Nino and La Nina effect individually.

Exceptions to La Nina’s dominance present themselves in the Eastern Corn Belt for corn, namely the majority of counties in Illinois and Iowa are influenced by El Nino rather than La Nina (12.37% vs. 0% and 15.15% vs 6%, respectively).

Furthermore, in Iowa the negative impact resulting from El Nino outweights La Nina in terms of magnitude, approximately -16% yield suppression in El Nino years compared to approximately +12% yield promotion in La Nina years. Michigan and
Missouri present similar results. With regard to soybeans, Virginia stands out with 32.14% of counties experiencing significant El Nino impacts (-12.58% yield suppression) compared to only 3.57% of counties significant under La Nina, also resulting in negative yield effects of -9.80%.

Comparing the direct effect of ENSO on yields to the full model (controlling for NDD, HDD, Prc), overall ENSO influence over corn is weaker under the full model. In other words, ENSO effect net of precipitation and temperature. While for soybean, the negligible results have strengthened by number of significant counties and strength of ENSO (overall). This could signal that for soybean ENSO influence is not through temperature or precipitation. Furthermore, soybean carry a higher temperature threshold and have displayed greater resistance to heat compared to corn (Lobell et al, 2013). This result should be expected, particularly due to the clear connection between HDDs and the ENSO yield relationship. Tables 3.3 and 3.4 display a summary of counties with significant ENSO impacts under the full model.
<table>
<thead>
<tr>
<th>State</th>
<th># Available Counties</th>
<th>% Significant Counties (Under El Nino and La Nina)</th>
<th>Average Estimate %</th>
<th>Phase</th>
<th>% Significant Counties</th>
<th>Average Estimate %</th>
<th>Significant counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>5</td>
<td>20%</td>
<td>-10.8%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>40%</td>
<td>-10.8%</td>
<td>2</td>
</tr>
<tr>
<td>Illinois</td>
<td>97</td>
<td>5%</td>
<td>-9.5%</td>
<td>El Nino</td>
<td>9%</td>
<td>-9.5%</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Iowa</td>
<td>99</td>
<td>5%</td>
<td>-0.1%</td>
<td>El Nino</td>
<td>7%</td>
<td>-10.4%</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>3%</td>
<td>10.1%</td>
<td>3</td>
</tr>
<tr>
<td>Kentucky</td>
<td>69</td>
<td>4%</td>
<td>-10.1%</td>
<td>El Nino</td>
<td>9%</td>
<td>-10.1%</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Michigan</td>
<td>41</td>
<td>13%</td>
<td>0.04%</td>
<td>El Nino</td>
<td>5%</td>
<td>-9.9%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>22%</td>
<td>10%</td>
<td>9</td>
</tr>
<tr>
<td>Minnesota</td>
<td>58</td>
<td>2%</td>
<td>8.7%</td>
<td>EL Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>3%</td>
<td>8.7%</td>
<td>2</td>
</tr>
<tr>
<td>Missouri</td>
<td>77</td>
<td>14%</td>
<td>-1%</td>
<td>El Nino</td>
<td>9%</td>
<td>-9.9%</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>19%</td>
<td>7.8%</td>
<td>15</td>
</tr>
<tr>
<td>Nebraska</td>
<td>82</td>
<td>20%</td>
<td>2.2%</td>
<td>EL Nino</td>
<td>1%</td>
<td>-11.1%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>39%</td>
<td>15.5%</td>
<td>32</td>
</tr>
<tr>
<td>North Carolina</td>
<td>58</td>
<td>3%</td>
<td>-5.4%</td>
<td>El Nino</td>
<td>3%</td>
<td>1.1%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>3%</td>
<td>-12%</td>
<td>2</td>
</tr>
<tr>
<td>North Dakota</td>
<td>12</td>
<td>4%</td>
<td>9%</td>
<td>EL Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>8%</td>
<td>9%</td>
<td>1</td>
</tr>
<tr>
<td>Ohio</td>
<td>68</td>
<td>1%</td>
<td>-9.9%</td>
<td>EL Nino</td>
<td>1%</td>
<td>-10%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>South Carolina</td>
<td>22</td>
<td>11%</td>
<td>-9%</td>
<td>EL Nino</td>
<td>5%</td>
<td>-13.7%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>18%</td>
<td>-4.3%</td>
<td>4</td>
</tr>
<tr>
<td>South Dakota</td>
<td>46</td>
<td>17%</td>
<td>8.3%</td>
<td>EL Nino</td>
<td>15%</td>
<td>7.2%</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>20%</td>
<td>9.4%</td>
<td>9</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>57</td>
<td>3%</td>
<td>10.1%</td>
<td>EL Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La Nina</td>
<td>5%</td>
<td>10.1%</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>791</strong></td>
<td><strong>8.8%</strong></td>
<td><strong>-1.3%</strong></td>
<td><strong>EL Nino</strong></td>
<td><strong>5%</strong></td>
<td><strong>-8%</strong></td>
<td><strong>43</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>La Nina</strong></td>
<td><strong>13%</strong></td>
<td><strong>4.9%</strong></td>
<td><strong>82</strong></td>
</tr>
</tbody>
</table>

Table 3.3 Full model significant corn counties at p<0.05

Note: Average estimate % refers to the percentage change in yield as a result of a 1°C deviation in ONI. Column 4 displays state averages for overall ENSO effect, while column 7 displays El Nino and La Nina effect individually.
<table>
<thead>
<tr>
<th>State</th>
<th># Available Counties</th>
<th>% Significant Counties (Under El Nino and La Nina)</th>
<th>Average Estimate %</th>
<th>Phase</th>
<th>% Significant Counties</th>
<th>Average Estimate %</th>
<th>Significant counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>4</td>
<td>88%</td>
<td>1.3%</td>
<td>El Nino</td>
<td>75%</td>
<td>-12.1%</td>
<td>3</td>
</tr>
<tr>
<td>Arkansas</td>
<td>27</td>
<td>19%</td>
<td>10.5%</td>
<td>La Nina</td>
<td>100%</td>
<td>14.6%</td>
<td>4</td>
</tr>
<tr>
<td>Illinois</td>
<td>98</td>
<td>5%</td>
<td>0.2%</td>
<td>El Nino</td>
<td>3%</td>
<td>-8.9%</td>
<td>3</td>
</tr>
<tr>
<td>Indiana</td>
<td>87</td>
<td>18%</td>
<td>0.3%</td>
<td>La Nina</td>
<td>6%</td>
<td>9.3%</td>
<td>6</td>
</tr>
<tr>
<td>Iowa</td>
<td>98</td>
<td>13%</td>
<td>9.1%</td>
<td>El Nino</td>
<td>5%</td>
<td>-8.6%</td>
<td>4</td>
</tr>
<tr>
<td>Kansas</td>
<td>43</td>
<td>45%</td>
<td>1.6%</td>
<td>La Nina</td>
<td>32%</td>
<td>9.2%</td>
<td>28</td>
</tr>
<tr>
<td>Michigan</td>
<td>32</td>
<td>34%</td>
<td>0.6%</td>
<td>El Nino</td>
<td>0%</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Minnesota</td>
<td>63</td>
<td>13%</td>
<td>0.1%</td>
<td>La Nina</td>
<td>26%</td>
<td>9.1%</td>
<td>25</td>
</tr>
<tr>
<td>Missouri</td>
<td>77</td>
<td>25%</td>
<td>10.5%</td>
<td>El Nino</td>
<td>77%</td>
<td>13.5%</td>
<td>33</td>
</tr>
<tr>
<td>Ohio</td>
<td>56</td>
<td>9%</td>
<td>9.5%</td>
<td>El Nino</td>
<td>10%</td>
<td>-10.7%</td>
<td>6</td>
</tr>
<tr>
<td>South Dakota</td>
<td>18</td>
<td>11%</td>
<td>9.8%</td>
<td>La Nina</td>
<td>17%</td>
<td>10.9%</td>
<td>11</td>
</tr>
<tr>
<td>Tennessee</td>
<td>18</td>
<td>56%</td>
<td>1.7%</td>
<td>El Nino</td>
<td>51%</td>
<td>10.5%</td>
<td>39</td>
</tr>
<tr>
<td>Virginia</td>
<td>28</td>
<td>20%</td>
<td>1.4%</td>
<td>La Nina</td>
<td>18%</td>
<td>9.5%</td>
<td>10</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>42</td>
<td>13%</td>
<td>8.8%</td>
<td>La Nina</td>
<td>22%</td>
<td>9.8%</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>691</td>
<td>26.3%</td>
<td>4.5%</td>
<td>El Nino</td>
<td>40.70%</td>
<td>10.5%</td>
<td>35</td>
</tr>
</tbody>
</table>

**Table 3.4 Full model significant soybean counties at p<0.05**

*Note: Average estimate % refers to the percentage change in yield as a result of a 1°C deviation in ONI. Column 4 displays state averages for overall ENSO effect, while column 7 displays El Nino and La Nina effect individually.*
On the state level, with a few exceptions, both El Nino and La Nina influence is weakened under the full model, albeit modestly. Furthermore, the overall characteristics remain unchanged with predominantly positive La Nina yield effects and negative El Nino effects. State level characteristics also mirror the direct model, where for example soybean yields in Virginia remain more sensitive to El Nino rather than La Nina. The weaker role played by HDDs and soybean yields, discussed previously, may be responsible for the dramatic increase in soybean producing counties with significant yield impacts due to La Nina.

In summary, considering key weather variables, the relationship between ENSO and US corn and soybean yields appears to be driven by changes in Heating Degree Days, as evidenced throughout the initial discussion of results. Regions experiencing significant HDD promotion under ENSO, similarly experience significant yield reduction. One should note, the results also present distinct ENSO related yield effects not captured via temperature or precipitation. It is of interest to note how ENSO effect on yield, as portrayed by any of the competing models, mirrors the relationship between ENSO and Heating Degree very closely. In this manner, describing the primary variable responsible for influencing yields via ENSO. Intuitively this rings true, as it is those temperatures that impact yields most greatly, more than normal degree days may promote positive growth. As a result, a clear pattern and relationship emerges, however the significance and reliability of this relationship (so to speak) must considered in closer detail. The significance of this relationship is discussed in greater detail in the forecasting exercise.
3.5 Conclusions

The discussion above relies on the model referred to as the basic model, describing ENSO effect directly on yields and hence allowing the model to capture the full ENSO effect including the many pathways through which it manifests itself. As expected the full model including degree days sways results, this is due to ENSO variable capturing yield impacts net of these vital weather pathways which promote or suppress yields. Comparison of the full model and the basic model further enforce the evidence for ENSO primarily influencing yield through heating degree days. This is particularly evident when geographically mapping county ENSO estimates.

Precipitation measures and outcomes may lay some doubt on the level of certainty in the above statements, however. As mentioned in the discussion, precipitation shocks due to ENSO do not translate into yield shocks. These changes in precipitation appear strong (up to 70% reduction) and in some areas significant. A possible reason for limited yield effect may be adaptation mechanism, such as irrigation, that help mitigate yield reduction as a result of ENSO influence over precipitation. Hence, it would appear that it is possible to adjust for precipitation changes, and not for temperature changes. One may argue, for example, that it is thanks to fine scale temperature data and modeling of non-linear temperature effects that this relationship between ENSO and temperature presents itself as evident. In similar fashion, the lack of proper precipitation measurement, accounting for variability within the growing season, and inclusion within the model may be to blame for the poor performance of precipitation proxies within the model and results. Furthermore, it is important to mention the limitations on time span
and data quality when considering irrigated vs non-irrigated county yields. The resulting use of mixed data very well may prevent any meaningful discussion on the implications.

In summary, the results for corn and soybean display minimal statistical significance, and hence minimal ENSO impact on yields as per the model specifications. Nonetheless, certain counties display significant ENSO impacts, with corn displaying a more meaningful response than soybean. These previously mentioned "pockets" or clusters of significance are likely caused by varying soil types across the area of study and are discussed in closer detail below.

Beginning with corn, the “eastern corn belt” experiences relatively neutral La Nina effects compared to the east and west boundaries of the study area. El Nino on the other hand displays concentrated negative yield impacts in this region where corn is typically grown. The most dramatic geographical shifts in ENSO impact are displayed from east to west. This is particularly true in a La Nina Phase where Nebraska and South Dakota (major US corn producing states) display strong positive impacts, while the eastern tier of the Corn Belt experiences yield suppression. Furthermore, corn yields in regions with a positive impact, such as the north/northwest of the corn belt, display lower variance in yield compared to yields experiencing suppression along the east coast. Overall a La Nina phase presents strong geographical clustering of both negative and positive effects for corn. El Nino, on the other hand, displays predominantly negative impacts, with only small pockets of positive impact in the west of the Corn Belt and southern coastal regions (i.e. Georgia). With regard to geographical clustering, there is considerable
clustering of negative effects with some clusters reaching the same magnitude as counties in the eastern tier of the Corn Belt under La Nina. One should note, geographical clustering presented here should be interpreted with consideration of previously mentioned Type I errors. The false rejection of the null hypothesis (rejecting the lack of ENSO influence on yield) may be a result of spurious correlation, potentially occurring through omission of region-specific variable(s).

Soybeans mirror corn’s relationship with ENSO but in a varied manner. What is meant by this is that under La Nina soybeans present closely similar results to corn, however at a universally lower magnitude. While under El Nino once again the predominant effect is negative, by comparison, the effect is much stronger for soybeans than it was for corn. More specifically under La Nina, there is a strong degree of clustering with predominantly neutral to weakly positive yield promotion. As stated, the geographical transition of negative to positive impacts relates to corn under La Nina however at a lower magnitude of ENSO impact throughout. Finally, El Nino producers nearly entirely negative effects with numerous clusters displaying up to -15% reductions in yield levels.

Finally, producer expectations and subsequent behavior should be given consideration. Producers may very well have ENSO related expectations which are taken into account when planning acreage, crop mix and or choice. A scenario under which producers have access and act upon ENSO forecasts would result in potentially unreliable estimates. However, the reliability of ENSO forecasts preceding the growing season for both soybeans and corn is questionable. The inaccuracy of ENSO forecasts prior to a corn or soybeans producers’ decision-
making window for crop choice and acreage is related to the spring barrier. The spring barrier describes the distinct increase in variability explained by both statistical and dynamic models after the northern hemisphere spring, particularly after April. Hence the likelihood of yield values reflecting significant ENSO adaptation is unlikely.
Chapter 4.
Forecasting ENSO Influence on Production

Measuring further benefits and uses derived from ENSO estimation leads to evaluating the predictive content of ENSO. In this forecasting exercise it is necessary to utilize knowledge of ENSO teleconnections, climate anomaly interactions over long distances, with local weather and how key variables composing local weather relate to crop growth. In this manner, the source of ENSO influence displays apparent importance, may it be through a predefined weather pathway or mix of unknown variables closely linked to ENSO volatility. In either case, the strength and confidence of such forecasts is equally imperative and will speak to the benefits of ENSO based forecasts for staple crop production in the US.

The ENSO is a quasi-cyclical phenomenon that repeats itself every three-to-seven years. While the patterns of this cycle are irregular, recent advancements in climatology have suggested improved predictability of this climate anomaly (Lou et al, 2008). This, of course, is valuable from an economic standpoint (National Research Council, 1997; Chang et al, 2000; Jones et al, 2000). It therefore follows that these benefits should be extended to the ENSO ~ Yield relationship in the effort of improving the forecast accuracy of staple crop yields, in particular in moments of extreme weather or climatic phase shifts. Forecasts allow for the elimination of various inefficiencies, such as promoting improved farm level management practices and planning as well as improvements in external risk mitigation strategies such as hedging through commodity markets. Possibly the benefits of
greatest interest lie in improved crop insurance policies through improved accuracy of premiums, coverage levels, and indemnity payouts. The applications are wide and merit a closer discussion of their own, however these benefits in terms of a crop producer begin with exploring the predictive content of ENSO. Along this line a basic question emerges: Does ENSO really matter? That is, while there can be a statistically significant linkage between ENSO shocks and crop yields, the economic importance of such relationship will likely lie in an out-of-sample predictability of yields due to these climate anomalies. A question of significant importance as ENSO phases appear to be increasing in frequency and severity, where under such conditions annual costs associated with ENSO could rise to 1 billion USD in the US alone (Chen et al, 2001).

To achieve these end results ENSO influence must be forecastable within a degree of certainty and under a clearly established framework. Hence, ENSO analysis should extend to the task of assisting in the improved predictability of crop yields. The mechanism which justifies ENSO as a valuable predictor and the methodology applied is presented and discussed below.

4.1 Conceptual Framework

Prediction of staple crop yields has gained the attention of researches for many years, most recently tremendous gains have been seen with new methods of data gathering and analyses. Crop yield simulators are often used in the field of agricultural economics to replace incomplete data sets, with notable success (Legler et al, 1999; Mauget et al, 2009; Liu et al, 2013). The focus of this research rather
than simulating yields was to explore the predictive content of ENSO and capture the predictive power of local weather conditions for corn and soybeans yields. In this line of thought, lack of variables such as soil quality and precise modeling of each crops biological functions are less of a missing variable issue but rather irrelevant to the model objectives. Hence the foundation lies on the teleconnections between the global ENSO phenomenon and local weather conditions on a county level in major corn and soybean producing regions in the US.

This relationship is further emphasized when considering the relative frequencies. ENSO can be described as a medium frequency global weather event, closely correlated with various local weather variables. In turn these local weather variables, county level weather conditions, in relation to ENSO are of a high frequency. The high frequency weather conditions, although most influential to local yield levels, are difficult to forecast within any degree of certainty more than several weeks ahead. This however is not the case for ENSO which can be forecast up to several months rather than days, particularly in the initial stages of a phase. It is this relationship between the medium frequency ENSO and high frequency local weather which facilitates ENSO based yield forecasts. Building upon this foundation, the identification of the precise weather pathways through which ENSO manifests itself on crop yields allows for efficient model specification and achieved through a focus on ENSO’s predictive power rather than solely on ENSO’s influence over yields from an in-sample methodology. Recent studies on this topic show that if modeled correctly key weather variables such as precipitation and temperature have a dominating effect on ultimate yield level outcomes. In recent years the focus has
shifted away from precipitation as a key driver of yield levels with a growing interest on temperature levels as well as interactions between temperature and water retention and availability in the surroundings. In this study temperature is considered in greater detail through non-linear temperature effects via degree days. Moreover, the previous chapter finds strong evidence for extreme temperatures acting as the driving force of ENSO’s influence over both corn and soybean yields. These results bring confidence to the forecasting methodology and are in line with previous studies. For a more detailed discussion on non-linear temperature effects the reader may refer to the literature review section of this thesis, as well as Schlenker and Roberts (2009) influential paper on this topic.

Finally, although considerable interest lies in the predictive content of ENSO via Heating Degree Days (temperatures above a given tolerance threshold) the results from the ENSO estimation exercise may not directly translate to the forecasting exercise presented here. This pre-defined weather pathway via which ENSO manifests itself on crop yields will be placed under greater scrutiny with a methodology of competing econometric models under alternate specifications, allowing for comparison and selection of those with the greatest forecast accuracy. The results may present a preferred model specification or a combination of models performing best. This is particularly important when isolating regions where predictions are statistically significant, not necessarily the same regions displaying statistical significance from the standpoint of an estimation methodology. Furthermore, if such statistical significance is present for definable regions or clusters, the benefits of utilizing such additional forecasting power could apply to
improved crop mix and farm level management as well as improvements in crop insurance premiums and indemnity payouts, with estimates stating that with ENSO forecast information indemnity payout burden for crop insurers could drop by 10-15% (Tack and Ubilava, 2015). Although not within the scope of the research at hand, the use of ENSO derived forecasts with respect to crop insurance presents considerable issues with regard to moral hazard from both the insured parties and insurers perspectives.

4.2 Empirical Framework

The modeling framework underlying the forecasting exercise examines the out of sample performance in a methodology similar to leave-one-out cross validation. Beginning with a discussion on the modeling framework and moving on the forecasting framework, consider a basic econometric regression model presented as a system of two equations capturing ENSO impact on yields through local weather variables.

\[ y_{it} = \alpha_i(t) + \beta' w_{it} + \epsilon_{it} \]  \hspace{1cm} [4.1]

\[ w_{it} = \delta' e_t + \nu_{it} \]  \hspace{1cm} [4.2]

Where both models present a time series OLS analysis for a single county, with \( y_{it} \) denoting corn or soybean yield for county \( i \) in year \( t \). The explanatory variable as
defined by $w_{it}$ presents a vector of weather variables or alternatively a single weather variable, namely precipitation and/or temperature. $\alpha_t(t)$ Denotes the time varying deterministic component (i.e., intercept and trend). $e_i$ variable is defined as a continuous variable representing both positive and negative deviations in SST anomalies. Hence yields are assumed to respond linearly to ENSO shocks. This may be a restrictive assumption, particularly for a subset of counties in consideration, as suggested in the previous chapter. But such modelling approach is desirable when working with a relatively small sample to ensure that outliers, if any, are not driving the results (both in identifying the parameters using the estimation sub-sample, and assessing forecasts using the hold-out observations). $\varepsilon_{it}$ And $\nu_{it}$ are independent and identically distributed error terms.

Weather variables are aggregated over the crop growing season, with precipitation being the sum of precipitation of each crop growing season and temperature captured through exposure to a range of temperatures from 0-40 degrees Celsius. This measure is the same as in the previous chapter, where temperature is defined via two variables; one denoting “normal” temperatures and the second denoting those temperatures above each crop’s yield temperature thresholds. Temperature thresholds are taken as given, with extreme or heating degree days being those above 29 degrees Celsius and in the case of soybeans 30 degrees Celsius. The full range of temperatures below the critical threshold are encompassed in the normal degree days threshold. No variable is produced for cooler to cold temperatures as these do not play a significant role for both the crops under consideration as well as the regions and growing seasons. Growing season
aside, both crops risk most significant damage within the first 48 and 24 hours, respectively, after planting (CropWatch University of Nebraska). Such an acute shock would be difficult to account for within the modelling framework. Instead it is more prudent to focus on the likely extreme heat variable which has shown to affect summer crop yields significantly (Schlenker and Roberts, 2009). Considering the vector of weather variables $e_t$ defines ENSO through a proxy of sea surface temperatures (SST’s), describing ENSO during its peak period from November to January. Importantly this peak period is the period prior to the crop growing season in each given year, hence serving as a lagged ENSO measure. Furthermore $\alpha_t(t)$ encompasses time varying deterministic components and $\nu_{lt}$ and $\epsilon_{lt}$ are independent and identically distributed error terms. Finally $\beta$ and $\delta$ are both in vector form, storing the parameters to be estimated for use later in the forecasting exercise. In this way the above simultaneous equations facilitate the estimation of the links between ENSO and corn and soybean yields through a predefined weather pathway, composed of multiple weather variable or alternatively a single weather variable describing the so called pathway.

The second specification instead of including a weather pathway omits weather variables all together to estimate the direct effects of ENSO on yields, offering the following reduced form specification,

$$y_{lt} = \alpha_t(t) + \eta'e_t + \nu_{lt}$$  \hspace{1cm} [4.3]
where again $\alpha_i(t)$ is a vector of time varying deterministic components not accounted for in the model, and $v_{it}$ is an independently and identically distributed error term. The middle term $e_t$ again is a proxy for ENSO impact during its peak trimester prior to the crops growing season. As compared to $\beta$ however in the previous specification utilizing a system of two equations to $\eta$ there are key differences in interpretation. The difference lies in that $\eta$ acts as a catch all vector of combined effects from every pathway which ENSO may influence, including those accounted for in the basic model as well as those unaccounted-for pathways through which ENSO manifests its influence over crop yields. This is very well the case for ENSO influence over crop yields, where its influence reaches crop yields through multiple vectors. As noted these vectors may not only not be accounted for in the basic model but can extend beyond “conventional” weather variables such as the precipitation and temperature presented above. For more discussion on uncongenial weather pathways through which ENSO may influence yields see Tack and Ubilava (2013b).

With this multitude of unaccounted for weather pathways an alternate specification is presented where, as in the basic model, weather is modeled as a function of ENSO. As a result of this approach one can identify ENSO related links net of the weather effects considered in previous models.

$$y_{lt} = \alpha_i(t) + \beta' w_{lt} + \eta' e_t + \xi_{lt} \quad \text{[4.4]}$$
Through this specification, unnamed and unaccounted-for weather variables (or equivalently pathways), which may include occurrences such as hail storms, strong winds, high humidity or non-weather related factors such as prices. The listed factors may very likely have a close relationship with the ENSO phenomenon. Although related to the key weather variables of temperature and precipitation, they may be lost within the aggregation of these generalized weather terms.

It is important to note that these three alternative specifications should not be expected to always give similar results. This notion is founded on the basic principle of correlation, namely that it is not necessarily transitive by nature. Referring to the direction of the relationship, where if X is correlated to Y it does not necessarily stand that if Y is correlated to Z then Z must be correlated with X. This may cause issues with respect to ENSO, which is correlated with local weather (as described in the previous chapter), and clearly yield is closely correlated with local weather however it may stand that statistically significant links are lacking between ENSO and yields directly.

Furthermore, differentiating the approach undertaken for the estimation exercise to the forecasting model specification a few points are worth mentioning. Namely the trade-off between model uncertainty and parameter uncertainty can be seen. Where it may well be that a more precise representation of the ENSO –Yield relationship is acquired through a predefined weather pathway as a vector of multiple weather variables. However the additional parameters compared to the direct ENSO effect model may cause inefficiencies from a forecasting stand point, simply due to the necessary estimations of a larger set of parameters on which
forecasts are based. Due to this as mentioned in the opening statements, either of the above specifications may yield the strongest forecast accuracy. Possibly a combination of the above models may yield the greatest accuracy.

The forecasting methodology takes place a pseudo forecasting exercise, more akin to a form of leave one out analysis where out of sample forecasts result from estimation model parameters. Beginning with the estimation of ENSO influence over corn and soybean yields through three alternate specifications the forecasting exercise follows utilizing the respective parameter estimates. Three main concerns arise within a forecasting methodology as pertains to three sources of uncertainty, namely model parameter estimates, data used in models, and uncertainty as to the structure of the model itself.

For each crops yield modeled under a given specification the out of sample forecast and forecast evaluation follows a 5 step methodology. First, from the 59 year data set of yields, six observations were drawn at random without replacement. Next the chosen model was estimated utilizing the remaining observations in the data set. From which point parameter estimates were gathered and the ENSO variable acquired along with the remaining parameters is used to forecast yields for the omitted annual yields. Having acquired forecasts, or pseudo forecasts, in this manner the accuracy of forecasts was obtained by capturing forecast errors. Forecast error is given simply by

\[ \epsilon_{is} = y_{is} - \hat{y}_{is}, \quad s = 1, \ldots, S \]
where $\hat{y}_{is}$ denotes the predicted yield in a given year for a given crop yield, and $S$ is the number of observations left-out in the estimation stage. These forecast errors are then used in calculating the measure of forecast accuracy. Forecast accuracy may be measured through various methods, each with its pros and cons. These include, but are not limited to, scaled and percentage errors including Mean Absolute Error, Root Mean Squared Error and Mean Absolute Percentage Error or Symmetric Mean Absolute Percentage Error respectively. Root Mean Square Forecast Error (RMSFE), as an accuracy measure, the out-of-sample equivalent of the residual standard deviation, which allows assessing and comparing different models for their ability to predict variables of interest. Percentage errors primarily down fall is the assumption of a meaningful zero, although not an issue within the current framework, other issues include asymmetric penalties for positive and negative errors. Scaled errors on the other hand are appropriate within the current framework as comparisons across different data sets are not required and the scale remains constant, in this sense it is a relative measure for comparison within one time series. A potential issue with the use of RMSFE as a measure of forecast accuracy is that the variance of the errors captured may not be constant, significant due to the RMSFE being the standard deviation of the forecast errors. With the above description RMSFE can be given in the following function form:

$$ RMSFE = \sqrt{\frac{1}{S} \sum_s \epsilon_{is}^2} $$  \hspace{1cm} [4.6]
hence providing insight on the spread of forecast errors through the mean of the squared difference between observation and forecast, or stated differently the square of the error term.

The final step, or Step 5, was to repeat the above methodology many times to generate the RMSFE distribution. In summary, the four preceding steps are:

i) randomly draw 6 years without replacement from the complete set of years available for the analysis;
ii) estimate the parameters of the unrestricted and restricted models (i.e., with and without ENSO variables in the equation) using observations from the remaining years;
iii) forecast yields using the parameter estimates derived in step ii; and
iv) obtain forecast errors, and calculate the measures of forecast accuracy (RMSFE_U and RMSFE_R, for unrestricted and restricted models respectively).

Interpretation of the relative RMSFE values is straightforward. In a two-model scenario if on average RMSFE_U < RMSFE_R, or RMSFE_U - RMSFE_R < 0 it would follow that the unrestricted model (RMSFE_U) improves over the restricted one, and as such will serve as evidence of the important role of ENSO in predicting yields. Welch's t-test was used to ascertain the statistical significance of the above inequalities.

The restricted, or the benchmark model, is simply annual county level yield modeled as a function of a trend variable given as,
\[ y_{it} = \alpha_i(t) + \varepsilon_{it} \]  

Where \( y_{it} \) denotes corn or soybean yield for county \( i \) in year \( t \); \( \alpha_i(t) \) and \( \varepsilon_{it} \) are the trend variable and an independently and identically distributed error term, respectively. In this way the 3 separate models as well as a combined model of the 3 is compared to the benchmark model. Where the combined model is a convex combination of the forecasts from the three models, each of which accounts for ENSO information. Hence each county’s yield time series provides a county specific preferred model based on the lowest value of their respective RMSFE’s.

**4.3 Data**

The composition of the data set was not organized in a panel form as it was in the previous chapter, instead county level regressions are run for each county independently. Given the purpose of the exercise is out-of-sample forecasting, and not the in-sample inference, it will suffice to run county-by-county regressions, which are equivalent to estimating the parameters of the model in the panel regression setting with every right-hand-side variable interacted with the county-specific fixed effect. Further enforced due to ENSO effect being assumed to vary across counties and hence by model specification exclusively capturing county variation (i.e., considering county-specific heterogeneity in the data), motivated by findings in the previous chapter. With this exception in mind, the data used is identical to the data utilized in the estimation exercise and categorized by three sets:
corn and soybean yields, key weather variables, and ENSO measures. Each set spans 59 years from 1951 to 2010, with all model variables taking on annual form either through growing season aggregation or raw annual data. Production data for corn and soybean was downloaded from the USDA NASS Quickstats website, as bushels per acre and filtering for only complete data sets. Resulting in data for 699 soybean producing counties and 900 corn producing counties between 1951 and 2010.

Weather data was provided in monthly county level form as a rich weather data set providing precipitation totals as well as fine scale (county) degree day (DD) temperature data. Furthermore through this fine scale temperature data the modeling approach was able to build upon the analysis presented by Schlenker and Roberts (2009) in capturing non-linear temperature effects. Monthly DD data was transformed into two annualized temperature variables, Normal Degree Days (NDD) and Heating Degree Days (HDD). Where NDD represent exposure to temperatures below the critical threshold of 29 and 30, for corn and soybean respectively. Finally HDD measure the exposure of each crop to heating or Heating Degree Days above the temperature threshold. Both variables were annualized by growing season from March to August for each given year.

Anomaly data are represented through sea surface temperature (SST) readings rather than atmospheric pressure changes, from the Nino3.4 region of the mid-Pacific ocean. SST's are taken in anomaly form signaling either a warming of sea surface temperatures (El Nino) or a cooling of sea surface temperatures (La Nina) by more than +/- 0.5 degrees Celsius. Data was obtained from the Climate

\footnote{I thank Prof Wolfram Schlenker of Columbia University for providing these data.}
Prediction Center of the National Oceanic and Atmospheric Administration in the US. The SST's are recorded as a monthly time series which in turn were transformed into an annual variable for inclusion in the above models by average monthly readings during the peak trimester spanning from November to January. This 3-month average also utilized in previous research is referred to as the Oceanic Nino Index (ONI). Finally, ENSO information was combined with annual yield and weather data for a crop year \( t \) by calculating the mean SST anomalies in November and December of year \( t-1 \), and January of year \( t \) in order to form a single ENSO proxy variable.

4.4 Results and Interpretation

Results from the first stage of the forecasting process reflect the results presented in the estimation section of this thesis. Again, only a few counties display statistically significant results, implying that ENSO is not meaningful for the majority of corn and soybean producers in the US. Nonetheless, again for corn, several geographical clusters of significance emerge. Furthermore, the resulting forecast accuracy measures for all corn and soybean producing counties offer additional insights discussed below. Noting that model specification and functional form is different in the two approaches, a brief review of ENSO influence over both crop yields was in order. Presentation of the estimation results below is predicated on a) the OLS approach undertaken in this chapter and specification of ENSO as binary, and b) providing background to the forecasting accuracy results based on the above models. The estimation results are the first step in the forecasting framework and in
this sense, form the foundation for the parameters used to acquire out-of-sample forecasts.

In line with previous research, ENSO displays spatially diverse influences for both corn and soybean yields with a tendency for geographical clustering. A general trend of ENSO is the gradual change in sign of influence as one moves from east to west across the continental US. Considering the El Nino phase it can be seen that corn yield shift from a 10% decrease in yields along the east coast towards positive yield effects as one moves west. The western most areas of the study, i.e. South Dakota and Nebraska, display roughly positive 6% yield promotion on average. Positive clusters of El Nino influence were observed on the northern border of South Dakota and North Dakota with the largest and most visible positive cluster in terms of geographic expanse and magnitude occurring at the north-eastern tip of Nebraska and along the Nebraska-Iowa border. Yield suppression of a significant magnitude occurs across the majority of the Corn Belt particularly in the southern tier of the corn belt as well as the eastern Appalachian region and large areas of North Carolina and South Carolina. These negative impact regions display clustering, with the most notable occurring in inland North Carolina. These results of El Nino estimates along with soybean estimates can be seen in Fig. 4.1.
For both crops El Nino influence was estimated through the model specification incorporating HDD, NDD and ENSO. Considering this specification with regard to soybean yields a similar spatial influence of El Nino can be observed, with
yield suppression visible along the east coast, specifically in the state of Virginia, and transitioning into positive yield growth as one moves west. The western border of the considered counties displays the highest yield levels under an El Nino event, with particularly strong yield promotion in Kansas reaching roughly 9% yield increase due to a 1 degree Celsius deviation in the peak trimester (November-December-January) of SST readings. Another area which forms a weak cluster of positive yield influence can be seen in northern Ohio and Michigan. Finally, a weak yet well-defined cluster of yield suppression under El Nino can be seen at the southern tip of Illinois. The combined, direct, and indirect model specification produce similar results particularly the direct effect of ENSO specification, signaling the importance of a predefined weather pathway imposing ENSO influence.

Once again ENSO estimates for yields, as presented above for the model including both degree day effects and ENSO effects, can be compared to ENSO estimates for both NDD and HDD. Understanding the importance of temperatures, especially those above the upper resistance threshold, one would expect to find similarities between ENSO influence over HDD and ENSO influence over both corn and soybean yields. Again, the relationships Fig. 4.2 are presented as the effects of an El Nino or warm phase of the ENSO phenomenon. The direct effect ENSO model as displayed on top row of Fig. 4.2 for comparison.
Figure 4.2: Top to Bottom - ENSO influence on Yield, Heating Degree Days, Normal Degree Days for corn (L) and soybean (R)

Note: El Nino phase is displayed. La Nina displays the same effects, but in opposing direction.

Model 4.1 results are shown, describing ENSO impact on yield, HDD and NDD.
The results provide a depiction of the hypothesis stated above. Normal degree days are uncorrelated with ENSO yield effects, or at the very least, not a driving force behind the ENSO ~ yield relationship for either corn or soybean yields. Heating of Heating DD’s however outline the ENSO influence over yields almost perfectly, with some incongruences visible with respect to soybean.

How all this translates to the forecasting results is the focus of this exercise. First it is important to observe the geographical distribution of forecasting accuracy and how it interacts with ENSO’s spatially heterogeneous effects seen in Fig. 4.2. By nature this interaction would also need to consider the pattern visible of HDD as well as possibly NDD. Precipitation related estimates have been omitted from the results, however precipitation models were run for comparison. Omission of these estimates arises from two issues. First, precipitation is measured as a sum of monthly means, and therefore does not account for precipitation fluctuations and non-linearity. As a result, true precipitation impact is diluted. Finally, the second issue acts more as a further motivation, as little correlation between precipitation and temperature was seen and additionally well documented lower sensitivity of yields to variations in precipitation compounds hesitation as to its inclusion.

As described in the methodology RMSFE or root mean squared forecast errors were used to examine the predictive power of the models in consideration, in this way gaining an understanding of the predictive content of ENSO anomalies. The results below present forecast accuracy as RMSFE ratios by comparing forecast accuracy of competing models to the base model where yield was a function of a
trend variable. Those counties with RMSFE ratios below 1 suggest that ENSO facilitates better prediction of yields. RMSFE ratio is the given models RMSFE over the base model (only trend) RMSFE. Therefore, it follows that ratios below 1 suggest greater accuracy. Having omitted precipitation from the model’s weather vectors where appropriate, the competing models were those with ENSO incorporated directly or through degree days or a combination of both.
Figure 4.3: RMSFE values for corn (L) and soybean (R) under alternate model specifications. Values below 1 suggest improved accuracy.
Figure 4.3 above shows the forecast accuracy for all counties under consideration for both crops under the three competing models. The direct ENSO model specification produces the greatest forecast accuracy over all, however this statement cannot be generalized for all crop producing counties. Where for example under the direct specification for corn yields much of Nebraska displays an RMSFE ratio of 1 or greater while under the indirect model produces slightly more accurate forecasts. Soybean, however, experienced clusters of differing accuracy which unlike for corn were maintained more or less throughout the competing models. However, highly accurate clusters for corn producing counties maintain integrity throughout the competing models as well, particularly the cluster seen in the southern tier of the corn belt spanning the states of southern Illinois, Kentucky, and Missouri.

Furthermore, there was a notable discrepancy between regions under competing model specifications. Most interestingly the inclusion of degree day variables within the framework improves forecast accuracy in eastern regions, while lower accuracy was observed in the northwest regions. These discrepancies should be considered with the heterogeneous ENSO effects over the counties considered, resulting in a pattern under which forecast accuracy is important in those regions negatively affected by an El Nino like event. With regard to soybean, forecasting accuracy improved dramatically compared to much of the northwest and corn belt. In this way, soybeans struck a similarity with corn, as both crops experienced strong yield suppression under an El Nino event in exactly those east coast regions showing considerably stronger forecast accuracy.
Finally, forecast accuracy may be presented in terms of where there was statistically significant predictive content of ENSO on corn and soybean yields. Those counties displaying statistical significance are shown in the Figure 4.4.

Figure 4.4: Statistically significant counties (highlighted) displaying forecast accuracy.

Note: Figure 4.4 units describe % change in yield due to a 1°C deviation in ONI. El Niño phase is displayed. La Niña displays the same effects, but in opposing direction.
The results present negligible predictive content of ENSO for soybean, with one cluster appearing in northern Virginia where yields experienced up to 4% reduction under an El Nino phase. However, there are scattered counties which experience statistically significant yield promotion under El Nino, particularly in southern Michigan. Corn yields on the other hand present far more concrete results with ENSO producing statistically significant predictive content for definable clusters, the southern tier of the Corn Belt as well as the eastern Appalachian region in North Carolina. Both regions had distinct parameters in the plotting of RMSFE ratios as well as ENSO estimates due to strong forecast accuracy as well as strong yield suppression under an El Nino Phase, up 10% yield reduction in both clusters. These results further enforce the observation under which forecast accuracy, or in other words ENSO relevance, improved in those regions negatively influenced by ENSO under an El Nino phase definition. In summary accuracy improved under the indirect ENSO model and under two additional stipulations. First, for those counties with negative El Nino impact on yields, and second, accuracy improved predominantly for those counties located on the east coast.

4.5 Conclusion

With regards to results model selection and specification was of particular interest, more specifically if a combination or one model stood out as providing most reliable predictive content. In this respect, the direct ENSO model outperformed the indirect model by gaining a few regions of significance, with a few caveats. Interestingly those clusters of counties which displayed greatest forecast accuracy tend to retain
accuracy between competing models, further stressing the relevance of ENSO forecasts in such regions. It should be added that comparing coastal/eastern counties and those further west (predominantly negative effect and positive effect under El Nino, respectively) varied in preferred model selection. Eastern counties forecast accuracy benefited through the inclusion of DD variables, while the opposite was true for western counties which experienced mostly yield promotion.

With regard to the spatial influence of ENSO, and how it varied between the two crops of interest, the discussion above was based on El Nino influence, with modelling framework presenting opposite effects when considering La Nina. Again, the spatially diverse nature of ENSO was seen for both crops, broadly defined as a transition from negative effects on the east and coastal regions to positive effects moving through and towards the west of the Corn Belt. Corn displays geographical clusters of particularly strong yield promotion occurring on the border of north and South Dakota, as well as northern Nebraska. This closely mirrored ENSO influence over extreme temperatures, again enforcing the hypothesis where ENSO influence over yields was predominantly driven by Heating Degree Days. Several clearly defined clusters of ENSO influence were presented for soybeans. Although relatively weak magnitude, the clearly defined borders imply clearly defined regions on ENSO interaction with yields and local weather in a unique fashion. Once such example for soybeans can be found in counties centred on the southern end of Illinois. Finally, statistically significant forecasts were found in the southern tier of the Corn Belt and the eastern Appalachian region.
Chapter 5. Implications

The nature of ENSO as a medium frequency weather phenomena, predictable up to two years in advance considering the last decade, has been emphasized throughout the text and is essential to its relevance to producers (Lou et al, 2008). High-frequency local weather patterns may be, at best, forecasted a couple weeks ahead hence marking them as impractical for producer decisions on the farm level. Coupling ENSO with local weather events allows for extended forecast ability of those variables influencing yields. The greater the lead time the more value is received by the producer, with 3 to 6 months laying out the minimum yet effective requirement for sufficient planning and planting. Under such a scenario, forecasts based on ENSO, adjusting corn planting and crop mix based on ENSO phase designation can lead to benefits of up to 15 USD per hectare in the southeast United States (Jones et al, 2000). An earlier paper estimated the added benefits of ENSO based predictions, under a scenario of perfect forecasts, would add 323 million USD in value to the US agricultural sector (Solow et al, 1998). Additionally, a 1995 paper found the economic value of perfect ENSO forecasts in the southeastern US to be 145 million USD and imperfect forecasts to be 96 million USD (Adams et al, 1995). These may be divided into short-term and long-term benefits of ENSO forecasts, where short run benefits may be less clear and damages may be unforeseeable while long-term benefits include substantial risk and "unanticipated occurrence" mitigation (Tack and Ubilava, 2013b).
Key benefits of utilizing ENSO and ENSO-based forecasts lie in farm level management primarily referencing optimal crop mix choice as well as other factors such as fertilizer and pesticide use. In this way, the use of ENSO data provides an avenue for mitigating impacts of adverse conditions and also make the most of positive weather conditions before they occur, meaning prior to the planting season (Letson et al, 2005). Such a risk management approach not only relies on minimizing losses under adverse conditions but also making the best of conditions forecasted by ENSO in selecting optimal crop type, planting date, and land use. The notions are repeated throughout the relevant literature, with a study measuring the benefits of ENSO based land allocation decisions in Argentina as ranging between 5 and 15 USD per hectare (Messina et al, 1999). Finally, as the authors also state, ENSO based forecasts not only predict negative scenarios but may also let producers make the best of optimal conditions depending on crop choice, allowing for a bespoke decision process on the farm level. Risk management strategies on the farm level are also dependent on DSFW, primarily driven by precipitation variability and in turn influence planting date and employment of resources (Mark et al, 2014a). Planning for DSFW may be considerably improved taking ENSO forecasts into account. Improvements to the decision-making process are information based and through this greater detail and information content may only increase the added value of ENSO forecasting and ENSO based forecasts for yields.

ENSO information also interacts with effectiveness and actuarial fairness of crop insurance policies, another indisputably key risk management tool at US producer's disposal.
Tack and Ubilava (2013a) take note of data availability, enrollment and deadlines for RMA (Risk Management Agency) setting crop insurance rates. A 2007 paper confronted this issue as well in their study of index insurance using ENSO for Piura, Peru stating that a serious concern to insurance providers is the potential of ENSO data being predicted before the flood season and before the insurance subscription date (Khalil et al, 2007). Forecasts do not have to undermine insurance contracts if properly designed, in the latter case positive synergies can arise between forecasts, insurance and effective input use (Carriquiry and Osgood, 2012). Adjusting insurance products price based on fluctuations in forecasts assists insurers in managing fluctuations and costs of re-insurance. Furthermore, these forecasts can improve the actuarial fairness of schemes such as Area Yield Insurance, and have positive attributes for producers. These positive attributes, however, only occur if insurance rates are adjusted to reflect the information available from ENSO forecasts (Nadolnyak et al, 2008).

Considering a more specific crop insurance scheme, in the use of ENSO in Area Yield Insurance several important facts arise, these include that crop yield distributions in different phases are statistically different from each other and there are consistently different patterns of losses for different ENSO phases (Nadolnyak et al, 2008). For example, highest losses are seen in Corn and Cotton during El Nino and lowest in La Nina years regardless of coverage level chosen (Nadolnyak et al, 2008). However, this differs among crops, as for example peanut losses are greatest during neutral years.
Dispersion and heterogeneity of risk (or losses) are key to a sustained and effective crop insurance scheme. Tack and Ubilava (2013a) show that crop and weather shocks derived from ENSO can result in losses across wide production regions within the same time period. If their regions are small enough this could be disconcerting news for Area Yield Insurance implementation and sustainability. Another study states that ENSO is spatially consistent at the regional level with strong ENSO effects, and larger yields being found in La Nina years than in EL Nino years (Fraisse et al, 2013). Concerning for Area Yield Insurance, since the regional level remains consistent. However, the fact that the sub-regional level is not consistent is troublesome for basis risk in AYI. Khalil et al (2007) discuss the potential for clustering of proxy payoffs creating significant risk and potential losses for insurance providers through the use of indexed insurance based on ENSO data, however not referring to Area Yield insurance.

As mentioned at the start, the producer also has the ability to cause adverse selection through his/her actions. For example, in the application of weather-indexed insurance to the agricultural region of Piura, Peru in order to ensure the integrity of ENSO data both from the producers’ perspective and insurers perspective the data are collected and provided by a third party before use in the crop insurance scheme in Piura, Peru (Khalil et al, 2007). Positive aspects of the Piura case include minimal evidence for the possibility of predicting index, however, there is substantial evidence of payout event clustering which puts the sustainability of the scheme into question (Khalil et al, 2007). However, even the use of a 3rd party supplier and distributor of data are not always sufficient for eliminating abuse. A
study by Ker and McGowan (2000), show that in the United States producers would have the ability to generate excess rents through adverse selection on their behalf by the use of ENSO and other weather-related data (Ker and McGowan, 2000). This is possible through reinsurance decisions being set up in such a way that extra profits can be gained by insurers through adverse selection based on ENSO knowledge. Such information may also be available to producers however the expertise and resources available to insurers allows them to make the most use of the information and the producers have a very limited potential use, especially for small-scale farmers. Insurer performing adverse selection is possible, to some degree, possible thanks to the Standard Re-insurance Agreement (Ker and McGowan, 2000). ENSO data can be used to decrease adverse selection through the use of forecast conditional premiums, however through the use of ENSO data inter-temporal adverse selection can take place (Nadolnyak and Vedenov, 2013). Producers take advantage of this data when premiums are based on an unconditional distribution and when expected losses are high, resulting in an “income transfer” from insurers to producers. The increased accuracy of forecasts results in higher inter-temporal adverse selection. The way around excessive inter-temporal adverse selection is for the RMA to use forecast conditioned premiums to keep program actuarially sound. Such a step transfers the uninsurable (or known ahead) portion of the risk to the producers (Nadolnyak and Vedenov, 2013).

Finally, the manner by which ENSO forecast data are delivered as well as proper implementation methods for change in management techniques presents its own unique challenges. These challenges focus on translating ENSO data into usable
information for improving producers’ decision-making (Podesta et al, 2002). The value of this must exceed the obstacles, which has the potential for welfare gains as Mauget et al (2009) show that use of climate data and optimizing practices to suit future weather conditions has shown to increase profits significantly. The study considers the value of ENSO forecast information to dual purpose winter wheat production in the southern high plains of the US through the use of simulated data. The authors show that greater accuracy in regional perception forecasts may not lead to more forecast value for individual farmers, in other words, more accuracy is not necessarily better if accuracy is defined by regional precipitation averages (Mauget et al, 2009). However, another study finds that although increased regional accuracy may not necessarily be more beneficial, the use of sub-annual ENSO indicators allows from discernment between "continuity and change" before or during the crop season (Royce et al, 2011). Furthermore, Mauget et al use two different wheat prices to measure the risks and benefits of season forecast, interestingly showing that the best management practices for specific forecast conditions changed as the relative value of grain changed (Mauget et al, 2009). This shows that many variables must be considered when defining the value of ENSO data, and also they must be considered before a producer decides to implement and use ENSO data.

In regards to accessibility, the "trickle-up" approach resonates throughout the literature in that emphasizing that ENSO data has great opportunities for improving crop yields and welfare of small-scale producers, especially in underdeveloped regions of the world. A significant degree of work has been
performed on the use of ENSO data in weather-indexed insurance as well as the use of ENSO data as a "proxy index" as proposed by Khalil et al (2007). The data would act as a proxy for traditional climate data especially if data are missing or only short time series are available. Issues arising from using ENSO data as a proxy index (in Peru) include the reliability of different probability levels for exceeding max seasonal rainfall, and more importantly the strength of ENSO data's relationship with “local” outcome which may arise due to sampling uncertainties (Khalil et al, 2007). Keil et al study the tools available to producers in central Sulawesi, Indonesia for greater resilience to ENSO related droughts (Keil et al, 2007). Shows that one of these tools is increased access and use of ENSO forecast data by local producers. However it is this increase in availability that is a substantial obstacle to overcome and it is important to share expertise and knowledge on the collection and use of ENSO data from Western nations to underdeveloped nations in order to assist in the development and improvement of ENSO forecasts (Candel, 2007).

Lastly, results presented in this thesis may be expanded by future studies in several directions. One direction building upon the implications discussion would entail analyzing producers crop mix choice based their use of ENSO based forecasts, and measuring their resulting utility levels. A similar topic of interest would be closer integration of insurance premiums and coverage levels with ENSO based forecasts. Finally coupling the relationship between ENSO, HDD and yield with climate warming models would open a discussion on future geographical shifts in major crop growing regions.
In conclusion, Rubas et al (2008) imply a sense of urgency stating that developing countries should invest in ENSO data gathering and analysis technology as soon as possible, as the authors view this as the single dominant strategy available. Furthermore, they expand on this sense of urgency by finding that the best time to adapt and adjust to ENSO information is early on. Once 60-95% of producers have adopted the information there is no incentive to do so. Referring to the “ceiling” on the value of ENSO data adoption at alternative timing schemes (Rubas et al, 2008).
Chapter 6.
Conclusion

The El Nino Southern Oscillation produces far-reaching impacts across the world’s major crop producing regions. Given the possibility of this weather phenomenon’s rising volatility and frequency (Chen et al, 2001), a deeper understanding of its impacts and implications is vital, with the most basic of impacts is its influence on staple crop yields as well as consideration of ENSO’s predictive content with regard to implications. The results presented in both the estimation and forecasting exercise aim to shed light on these key questions and unknowns with the help of fine scale weather data and utilizing the latest in statistical advances for model composition. The US is a natural choice when considering staple crop yields with the US producing the majority of world corn, 13,601 million bushels in 2015/16 (USDA FAS, 2016), as well as being the major producer of soybean, 106.86 million metric tons between 2015/16 (USDA FAS, 2017). Considering these numbers of such staple goods, serving as a major source of caloric intake globally, large external shocks would carry consequences not only to the respective industries but also communities around the world.

The case presented finds itself on an upward trajectory of urgency with global food production growing in relevancy rather than being replaced or deemed immaterial to the many global challenges we face today. Agriculture is arguably a bedrock of civilization and social stability. With growing world populations, particularly in developing nations, the need for efficient agricultural production is
more important than ever to a study finding that in order to keep up with food demand crop production alone will need to double by 2050 (Ray et al., 2013). The authors go on to say that given current trends in agricultural production the necessary target of 2.4% annual growth will not be achieved, this is with the data pointing to corn as the crop with the highest growth potential of 1.6% per year. Such figures are particularly gripping when paired with the results presented in both the estimation and forecasting exercises, furthermore, the crop with second highest expected growth rate is soybean (1.3% annually) (Ray et al., 2013).

As presented in the results of this thesis, soybean yields are negligibly forecastable under an ENSO model framework. Nonetheless, a statistically significant link exists between ENSO and soybean yields in dozens of major soybean producing counties in the US. More importantly, corn displays stronger links to ENSO both solely from an estimation standpoint (i.e., in-sample fit) as well as from a forecasting perspective. The empirical chapters of this thesis indicate temperature as the key channel through which ENSO most likely influences staple crop yields in the US to the greatest degree. ENSO appears to carry considerable influence over precipitation, however, these changes do not translate into yield shocks. For both crops included in the models, extreme temperatures, those above each crop’s respective temperature threshold, inflict ENSO’s most detrimental impact on yields.

Further significance of a pathway through which ENSO influences US yields has to do with the rise in research arguing that global temperatures are on the rise. Various global heating models predict significant temperature increases in key crop producing regions. Rising temperatures as a result of climate change have shown to
jeopardize not only crops reliant on rainwater but even those planted in irrigated fields, with decreasing water yields in the central and southern US (Rosenberg et al, 2003). The results presented here echo concerns on extreme temperatures, mainly through the rising volatility of ENSO and justified through the apparent pathways via which ENSO manifests itself over crop yields.
References


Murphy, R. C. (1926). Oceanic and climatic phenomena along the west coast of South America during 1925. Geographical Review, 16(1), 26-54.


