ABSTRACT

Title of Document: ESSAYS ON CLIMATE CHANGE IMPACTS
AND ADAPTATION FOR AGRICULTURE

Ariel Ortiz Bobea, PhD, 2013

Directed by: Richard E. Just, Distinguished University Professor
Department of Agricultural and Resources Economics

Over the past twenty years economists have developed econometric approaches for estimating the impacts of climate change on agriculture by accounting for farmer adaptation implicitly. These reduced-form approaches are simple to implement but provide little insights into impact mechanisms, limiting their usefulness for adaptation policy. Recently, conflicting estimates for US agriculture have led to research with greater emphasis on mechanisms including renewed interest in statistical crop yield models. Findings suggest US agriculture will be mainly and severely affected by an increased frequency of high temperatures with crop yield suggested as a major driver.

This dissertation is comprised of three essays highlighting methodological aspects in this literature. It contributes to the ongoing debate and shows the preeminent role of extreme temperature is overestimated while the role of soil moisture is seriously underestimated. This stems from issues related to weather data quality, the presence of time-varying omitted weather variables, as well as from modeling assumptions that inadvertently underestimate farmers' ability to adapt to seasonal aspects of climate change. My work illustrates how econometric models of climate change impacts on crop production can be improved by structuring them to admit some basic principles of agronomic science.

The first essay shows that nonlinear temperature effects on corn yields are not robust to alternative weather datasets. The leading econometric studies in the current literature are based on a weather dataset that involves considerable interpolation. I introduce the use of a new dataset to agricultural climate change research that has been carefully developed with scientific methods to represent weather variation with one-hour and 14 kilometer accuracy. Detrimental effects of extreme temperature crucially hinge upon the recorded frequency at the highest temperatures. My research suggests that measurement error in short amounts of time spent at extreme temperature levels has

disproportionate effects on estimated parameters associated with the right tail of the temperature distribution. My alternative dataset suggests detrimental temperature effects of climate change over the next 50-100 years will be half as much as in leading econometric studies in the current literature.

The second essay relaxes the prevalent assumption in the literature that weather is additive. This has been the practice in most empirical models. Weather regressors are typically aggregated over the months that include the growing season. Using a simple model I show that this assumption imposes implausible characteristics on the technology. I test this assumption empirically using a crop yield model for US corn that accounts for differences in intra-day temperature variation in different stages of the growing season. Results strongly reject additivity and suggest that weather shocks such as extreme temperatures are particularly detrimental toward the middle of the season around flowering time, which corrects a disagreement of empirical yield models with the natural sciences. I discuss how this assumption tends to underestimate the range of adaptation possibilities available to farmers, thus overstating projected climate change impacts on the sector.

The third essay introduces an improved measure of water availability for crops that accounts for time variation of soil moisture rather than season-long rainfall totals, as has been common practice in the literature. Leading studies in the literature are based on season-long rainfall. My alternative dataset based on scientific models that track soil moisture variation during the growing season includes variables that are more relevant for tracking crop development. Results show that models in the literature attribute too much variation in yields to temperature variation because rainfall variables are a crude and inaccurate measure of the moisture that determined crop growth. Consequently, I find that third of damages to corn yields previously attributed to extreme temperature are explained by drought, which is far more consistent with agronomic science. This highlights the potential adaptive role for water management in addressing climate change, unlike the literature now suggests.

The fourth essay proposes a general structural framework for analyzing the mechanisms of climate change impacts on the sector. An empirical example incorporates some of the flexibilities highlighted in the previous essay to assess how farmer adaptation can reduce projected impacts on corn yields substantially. Global warming increases the length of the growing season in northern states. This gives farmers the flexibility to change planting dates that can reduce exposure of crops during the most sensitive flowering stage of the crop growth cycle. These research results identify another important type of farmer adaptation that can reduce vulnerability to climate change, which has been overlooked in the literature but which becomes evident only by incorporating the principles of agronomic science into econometric modeling of climate change impact analysis.

ESSAYS ON CLIMATE CHANGE IMPACTS AND ADAPTATION FOR AGRICULTURE

By

Ariel Ortiz Bobea

Dissertation submitted to the Faculty of the Graduate School of the

University of Maryland, College Park, in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

2013

Advisory Committee:
Professor Richard E. Just, Chair
Robert G. Chambers
Maureen L. Cropper
Kenneth E. (Ted) McConnell
Roberton C. Williams III

Dedication

A Macú, Fahelo, mis abuelos, mis hermanos y a Patricia.

A Pepé y a Memé, ahora empiezo una nueva aventura con sueños de ser un Nolan Ryan de otra Liga.

A Julia.

Acknowledgements

I am extremely grateful to my advisor Richard E. Just for his unyielding support throughout this project. He simply could not have provided a better balance of freedom and guidance. I cannot thank him enough for his impressive responsiveness to my concerns and inquiries. Prof. Just has not only been a source of encouragement and motivation for my research but also a source of humility and prudence. I feel extremely fortunate to benefit from the guidance of such a thoughtful and well-rounded economist, which also happens to be the most efficient person I have ever known.

I'm very indebted to Prof. Robert G. Chambers for his thoughtful advice and support. I am deeply appreciative of his unceasing striving to challenge his students and push them into questioning pre-conceived ideas or conventional ways of thinking about problems. I'm also thankful for instilling in me and other students that applied economists must have higher standards and responsibilities than theoreticians, not the opposite. As odd as it may sound, I have never laughed so hard as I did during his applied production analysis course.

I am also grateful to many faculty members who have contributed both to my training and my research through comments on manuscripts or during seminars. I want to thank Marc Nerlove, Ted McConnell, Erik Lichtenberg, Rob Williams, Maureen Cropper, Lars Olson, Ramón López, Anna Alberini, Ken Leonard, Charles Towe, Jeanne Lafortune, as well as many of my classmates and participants to our environmental economics seminar series.

I want to thank Alfredo Ruiz-Barradas, a Research Scientist in the Department of Atmospheric and Oceanic Science, for his admirable patience and readiness to help me navigate climate data. I also thank Prof. Sumant Nigam from the same department for inviting me to present my research to his non-economist students.

Finally, I would like to thank many professors, teachers and distinct individuals in the US, France and the Dominican Republic who in one way or another have contributed to making graduate school in agricultural and resource economics a reality. I want to thank Peter Wilcoxen, Mary Lovely and John McPeak at Syracuse University, as well as Marc Dufumier, Hubert Cochet and Laurence Roudart at AgroParisTech in France. I want to thank my former boss and Minister of the Environment of the Dominican Republic, Max Puig, for his encouragement to pursue doctoral studies.

Table of Contents

Essay I

Nonlinear Temperature Sensitivities of Crop Yields in Climate Change Research

Abstract

Several recent studies suggest US agriculture will be negatively affected by climate change through more frequent exposure to extremely high temperatures. For example, the most respected of these studies finds that temperature is particularly harmful beyond a threshold of around 30ºC and, as a result, predict that yields of major US crops will decrease by 30-82% under climate change. In this paper I show that these estimated detrimental effects of extreme temperature are highly dependent on the climate dataset they use. I find that detrimental effects of extreme temperature are reduced by 50% in a replication for US corn yields based on an alternative high-quality climate dataset. This alternative dataset is derived from the North American Land Data Assimilation System (NLDAS) which is a joint project of four major US agencies and universities. The results stems from the fact this alternative dataset has a thicker right tail of the temperature distribution ($\geqslant$35ºC). My findings highlight the need for future studies to conduct thorough cross-validation with gold standard climate datasets to ensure external validity.

1 Introduction

Recent studies point to extreme temperature as the main driver of large negative projected impacts of climate change on agriculture. The evidence is based on hedonic studies of land prices (e.g. Schlenker et al., 2006) as well as on statistical crop yield models for the US and elsewhere (Schlenker and Roberts, 2009a, Lobell et al., 2011). In this line of research, Schlenker and Roberts (2009, henceforth SR) make an important contribution by showing temperature effects on crop yields are nonlinear and increasingly detrimental beyond crop-specific thresholds. They find crop yield losses in the range of 30–82% under climate change, driven primarily by the increase in temperature exposure above threshold values of 29°C (corn), 30°C (soybean), and 32°C (cotton). Meerburg et al. (2009) find these large damages pessimistic and argue that corn is successfully grown in countries like Brazil where temperature above these thresholds is more common. Schlenker and Roberts (2009b) argue, however, that results for Brazil are in line with their findings for the US.

The estimation of detrimental effects of extreme temperature obviously hinges on the correlation of low crop yields with extremely high temperatures. However, extreme temperatures remain rare and have very short durations. In a 27-year sample of 800 Midwest counties explored in this paper, exposure beyond 35°C represents, on average, about 0.5% of the March-August period. Because of the waveform shape of the temperature curve, measurement error in temperature levels lead to larger measurement errors in exposure to extreme temperature. A seemingly small difference in a handful of hours over the growing season can lead to substantially different estimated effects of extreme temperature on yield and therefore on climate change impact projections.

In order to explore the role of measurement error in the right tail of the temperature distribution, I replicate the model in SR for US corn using an alternative weather dataset and a 27-year corn yield panel of 800 Midwest rainfed counties representing 70% of US corn production. The alternative weather dataset I use is based on the North American Land Data Assimilation System (NLDAS) which is a joint project of the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), Princeton University and the University of Washington. The NLDAS forcing weather dataset features hourly and 14-km resolution and has been shown to closely match observations of highly precise weather stations in the Great Plains (Cosgrove et al., 2003). In contrast, the dataset in SR is interpolated by the authors by combining daily weather station data with monthly data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) from Oregon State University and simply assumes a daily temperature sine-curve.

My findings suggests that nonlinear effects are highly sensitive to very small differences in the right tail of the temperature distribution ($\geqslant 35$ °C). Specifically, the NLDAS dataset I use has a slightly thicker right tail resulting in a reduction of roughly 50% of the estimated detrimental effects of extreme temperature on yield. Both weather datasets fit well the production data and I was not able to statistically discriminate between them. The comparison, however, highlights that the large impacts in these studies are significantly dependent on the weather dataset, and particularly on how fat the tails of temperature distributions are. This work thus serves as a word of caution and hopefully will lead to the development of thoroughly cross-validated and externally validated climate datasets for climate change impact analysis.

The paper is organized as follows. Section 2 describes data sources and compares the NLDAS dataset to the SR dataset. Section 3 briefly discusses the crop yield model developed in SR. I present regression results based on both datasets and discuss implications in section 4 before concluding in section 5.

2 Data

2.1 Data sources and dataset construction

This study seeks to explore the role measurement error of extreme temperature exposure in the estimation of nonlinear temperature effects on crop yields. To do so I replicate the statistical crop yield models developed in SR using two weather datasets: the original SR dataset which was kindly provided by the authors and an alternative dataset derived from the NLDAS and is publicly available for download.[1]

SR provide the details of how the weather dataset is constructed in the appendix of their paper. In summary, the authors generate daily weather data by interpolating daily but spatially sparse weather station data using monthly but spatially detailed model-generated weather data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) developed at Oregon State University. The interpolation yields daily precipitation as well as minimum, maximum and average temperature for each 2.5×2.5 mile grid over the contiguous US for the 1950-2005 period. County-level weather data is obtained by aggregating grids covered with agricultural land as measured by a LandSat satellite image. As the following section will show, the semi-parametric crop yield model relies on variables representing the exposure to individual temperature bins during the growing season. To replicate their results I use daily county-level weather data to derive the distribution

---

[1]See http://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing_download.php for directions.

of temperature exposure for each degree bin under the assumption of a sine-shaped temperature curve.[2]

The NLDAS weather dataset is derived from the North American Regional Reanalysis (NARR). Details about the NLDAS and NARR are provided in Mitchell et al. (2004) and Mesinger et al. (2006), respectively. The NLDAS was developed to account for the role of soil water in surface energy fluxes which are important for weather forecasting. NARR was developed to provide a time and spatially consistent weather dataset over North America. NLDAS uses a spatially-downscaled and time-disaggregated version of the NARR dataset as input and features hourly observations for each 6.8×8.6 mile grid over the contiguous US since January 1, 1979.[3] Cosgrove et al. (2003) show the NLDAS dataset closely matches the highly precise weather observations of the Department of Energy's Atmospheric Radiation Measurement (ARM) program in the Great Plains.

Agricultural data were obtained from the U.S. Department of Agriculture's National Agricultural Statistics Service (USDA-NASS). Because corn production data, which is the same as used by SR, is available only at the county-level, the hourly gridded weather data must be spatially aggregated for each county. I do so by weighting each NLDAS data grid within a county by the amount of cropland within each grid. The cropland area is derived from the USDA-NASS's 2011 Crop Data Layer which has a 30-meter resolution. This allows weighting NLDAS data grids according to the amount of farmland they include in constructing the county-level observations. Hourly observations were subsequently used to construct exposure to individual degree-bins for the March-August period for each year and county.

Because rainfed and irrigated corn yields are expected to respond differently to exogenous environmental conditions, their respective parameters must be estimated separately. For this purpose, I restrict the sample to counties where at least 75% of the acreage, on average, is rainfed. Figure 1 illustrates where the sample counties are located. The dataset corresponds to a balanced corn yield panel of 800 Midwest rainfed counties for 1979-2005, which represents 70% of US corn production. This sample period is restricted to the overlapping periods of the two weather datasets.

2.2   Dataset comparison

Aggregation leads to loss of information. I thus compare daily weather observations rather than pluri-monthly aggregate measures as weather regressors in the crop yield model. Because of the

---

[2]In the appendix in SR, the authors insist that the distribution of temperature for each degree must be performed for each individual grid prior to aggregating at the county level. This procedure is more computationally demanding than the one I describe and yields virtually identical results. The likely reason is that there is almost perfect correlation of temperature for bins falling within the agricultural land of a given county.

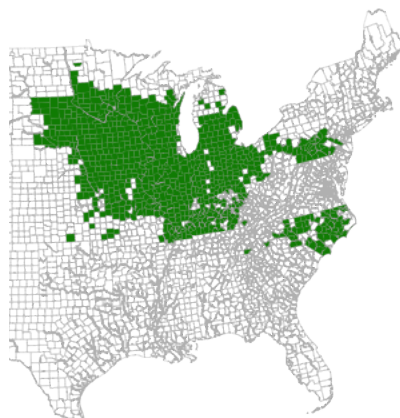[3]The source NARR dataset features 32-km and 3-hour resolution.

Figure 1: Rainfed counties in the sample

central role of temperature in the results in this literature, I restrict the comparison to this variable.

Figure 2 shows the pairwise comparisons of daily minimum, average and maximum temperature by month for the full sample (1979-2005, 800 counties). The red line represents a smooth spline curve associated with the regression of each variable in the SR dataset on the same variable in the NLDAS dataset. The dashed black line represents the bisector along which observations across datasets are identical. In each quadrant I provide a blue (green) density function representing the empirical distribution for each variable in the SR (NLDAS) dataset.

The two datasets are highly similar for the most common values of each variable. This is indicated by the close connection of the red fitted line with the bisector around the highest density concentrations. However, the red fitted line tends to be flatter than the bisector indicating a lower range of variation in the SR dataset. This pattern is apparent for minimum temperature but is particularly strong for maximum temperature during the warmer months of the summer. This suggests that extremely hot events, which typically occur during these months, register lower temperature in the SR dataset than in the NLDAS dataset. The obvious implication is that less exposure to extremely high temperature is reflected in the SR dataset.

A systematic bias between datasets would be reflected as parallel shifts of the red fitted line in relation to the bisector. This is not the case. Rather, a bias toward the mean of each variable is apparent in the SR dataset, which is particularly apparent for maximum temperature in the summer.

To assess how these differences affect weather regressors, I represent the average time spent in each degree bin for various months in figure 3. The first two panels (A-B) are box-plots representing temperature variation within each temperature bin for various months.[4] Panel C shows variation

---

[4]Each box provides the level of the median, the 25th and 75th percentiles. The whiskers extend to the most extreme data point which is no more than the interquartile range from the box.
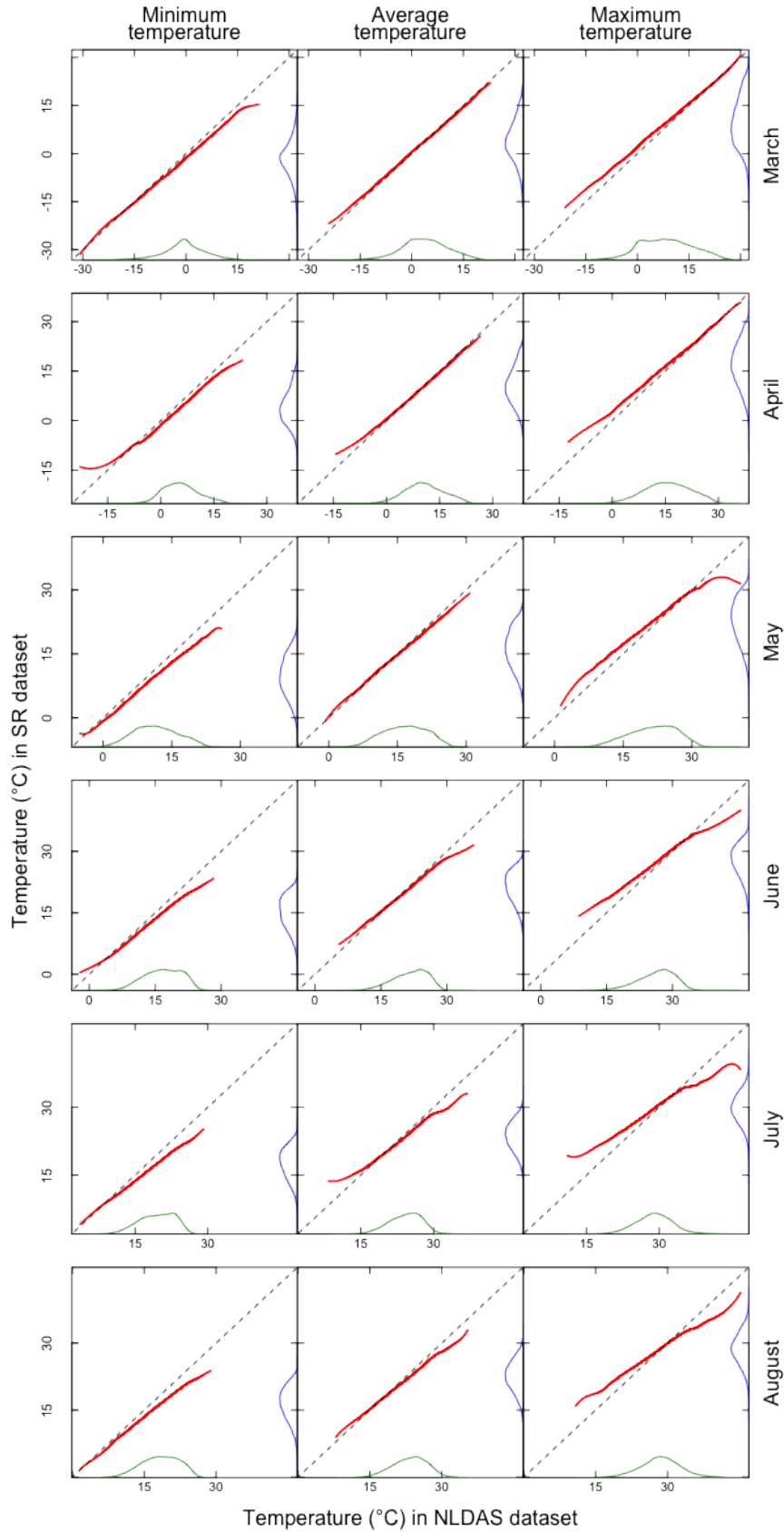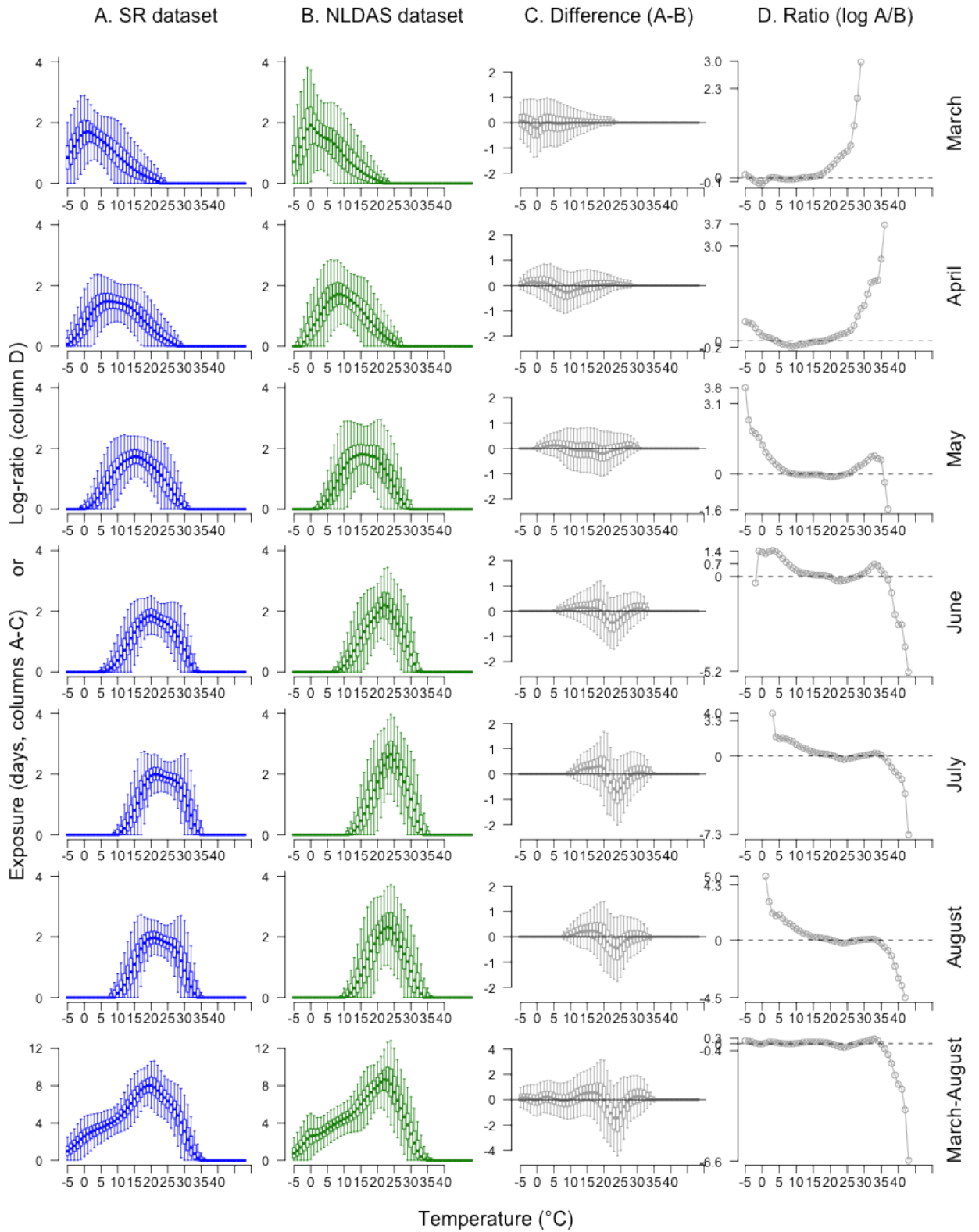
Figure 2: Temperature comparison

6

Figure 3: Comparison of temperature distributions

of the difference of daily exposures to each bin between datasets. Positive (negative) values for a bin indicate that the SR dataset records more (less) time spent within that particular bin than the NLDAS dataset. Panel D represents the log-ratio of time spent in a bin between datasets. Positive values indicate that the SR dataset has relatively higher exposure to a particular bin than the NLDAS dataset.[5]

Panels A through C reveal fairly small absolute differences between the datasets for all bins. The NLDAS dataset exhibits a more distinctive mode for most months, particularly for the warmer months of June-August. This is represented in panel C as a dip of boxes in the 20-25ºC range. This pattern is also visible for the full March-August period in the last row. Notably, the differences are small in absolute terms for the right tail of the distribution. Indeed, the boxes are virtually centered around zero for high temperature bins for all months.

Panel D, on the other hand, reveals particularities of how temperature exposure differs at the tails in both datasets. For the cooler months of March and April, the SR dataset shows higher relative levels of exposure to high temperature (25-30ºC). However, for the warmer months of May-August, the SR dataset exhibits a much lower relative prevalence of exposure to high temperature ($\geqslant$35ºC). In other words, the SR has higher (lower) exposure levels to high temperatures during the colder (hotter) months.

In summary, the SR dataset records a lower level of extremely high temperature as a representation of aggregate exposure for the March-August period (last row). In other words, the NLDAS dataset has a thicker right tail of the temperature distribution. As an illustration, table 1 shows the average exposure in the March-August period to temperature above 35ºC is 14.4 hours and 22.3 hours in the SR and NLDAS datasets, respectively. While the difference is a mere 7.9 hours in a six-month window, this represents a hefty 55% decrease in observations that are responsible for the major predicted yield differences.

The next section demonstrates that the nonlinear effects of extreme temperature are based on the recorded exposure at these extreme values.

3   The model

Statistical models that regress crop yields on weather variables have traditionally relied on monthly or pluri-monthly average temperature and precipitation data. Early examples can be traced back to the early part of the last century (Wallace, 1920; Hodges, 1931). Since then, the

---

[5]This calculation is based on the ratio of the average total monthly exposure for each bin, rather than the monthly average of the daily exposure ratios for each bin. This approach avoids issues related to zero exposure to a bin in a day. In addition, bins for which the average total monthly exposure is zero were obviously omitted from the graph.

Table 1: Average exposure to various temperature intervals for March-August

Interval	SR dataset (hours)	NLDAS dataset (hours)	Difference (SR-NLDAS)	% change (SR-NLDAS)/SR
25-29ºC	638.7	685.9	-47.2	-7.4
30-34ºC	211.9	180.6	31.4	14.8
≥35ºC	14.4	22.3	-7.9	-55.2

convention has been to include linear and quadratic variables based on temporally aggregated data to capture the nonlinear effects of both temperature and precipitation on yield. Marginal effects of these variables are typically expected to exhibit an "inverted U" shape, suggesting diminishing marginal effects of each weather variable with a unique optimum.

Schlenker and Roberts (2009a) made an important contribution by recognizing that daily average temperature fails to convey the consequences of exposure to extreme temperature and, thus, may not be adequate for capturing nonlinear effects of temperature on corn yields. Hypothetically, two days with equal average temperature may represent very different exposures to very high temperatures. This suggests that the shape of the daily time curve matters.

To address this needed refinement, SR developed an innovative approach that estimates the effect of exposure to different levels of temperature on yield separately. They compute the amount of time spent during the season (March-August for corn) in each of many temperature bins. The exposure to each degree bin is then adapted to various specifications. This is an extension of previous hedonic work on land prices in Schlenker et al. (2006), but with greater flexibility in the specification of the temperature response function.

Here, I replicate their model for corn yields for purposes of comparison. I restrict the sample period to the 1979-2005 period of overlapping data, which is shorter than the 1950-2005 period used by SR. However, their results are reported to be similar for temporal subsets of the sample.

Their general model assumes that temperature effects on yield are cumulative and substitutable over time. The nonlinear effects of temperature on yield are captured by the function $g(h)$ representing "yield growth" that depends on temperature h. Logged corn yield y_{it} in county i and year t are represented as:

$$y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}(h)d(h) + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i + \epsilon_{it} \tag{1}$$

where $\phi_{it}(h)$ is the time distribution of temperature (i.e., the temperature-time path) for March-August, p_{it} is precipitation, z_{it} is a state-specific quadratic time trend and the c_i are county fixed-effects. The maximum likelihood estimation procedure accounts for spatial correlation of the errors. Over thirty different spatial weight matrices were evaluated by comparing models that only differ by the weight matrix. The weight matrix based on the inverse distance of the seven nearest neighboring counties yielded the highest value of the likelihood function at the optimum parameter values and thus was selected.[6]

Equation (1) cannot be estimated directly because of the integral. Therefore, I follow SR and consider different specifications to approximate the integral as a sum: a step function allowing different effects at each 1ºC interval (SR1), another step function allowing different effects at 3ºC intervals (SR2), an eighth-degree polynomial (SR3), and a cubic B-spline with eight degrees of freedom (SR4).[7] The specification for SR1 is:

$$y_{it} = \sum_{h=0}^{40} g(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)] + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i + \epsilon_{it}$$

where $\Phi_{it}(h)$ is the cumulative distribution of temperature in county i and year t. Specifications for SR2, SR3, and SR4 and more detailed results for each specification are provided in the appendix.

For the SR1 specification, the effects of the data differences explored in section 2 are clear. A thicker right tail for temperature distributions in the NLDAS dataset results in higher values of $\Phi_{it}(h+1) - \Phi_{it}(h)$ for very high temperature (e.g. $h \geqslant 35$ºC). All else equal, this should result in lower estimated effects of $g(h+0.5)$. In other words, the detrimental effects of temperature would seem lower at such high values. A similar reasoning applies to the other specifications, although these assume a certain degree of dependence on effects of neighboring temperature bins.

4   Results and discussion

Figure 4 shows nonlinear effects of temperature on corn yield based on the SR (left) and NLDAS (right) datasets.[8] The temperature distributions are represented with red histograms under the corresponding graph. Temperature effects exhibit similar thresholds around 29°C, but detrimental

---

[6]The weighting matrices included eight neighboring structures and four weighting schemes. The neighboring structures are: 5 through 10 nearest neighbors, neighbors within 200km, and neighbors using the Delaunay triangulation. The weighting schemes are: binary, inverse distance, inverse squared distance, and inverse square root of distance.

[7]Only SR2 and SR3 are part of the original SR study. In addition, SR developed a piecewise linear model which yields similar results to the other specifications. The SR1 was included to assess the effects of narrow temperature bins and SR4 to allow for a more flexible less susceptible to extreme polynomial curvature near the end points specification.

[8]The precipitation response function is omitted but remains very flat. I also rely on the precipitation data from the SR dataset only so that I isolate the effect of the temperature difference. Nevertheless, results based on season-long precipitation variables from the NLDAS dataset are virtually identical.

Figure 4: Nonlinear temperature effects using different datasets

A. SR dataset   B. NLDAS dataset
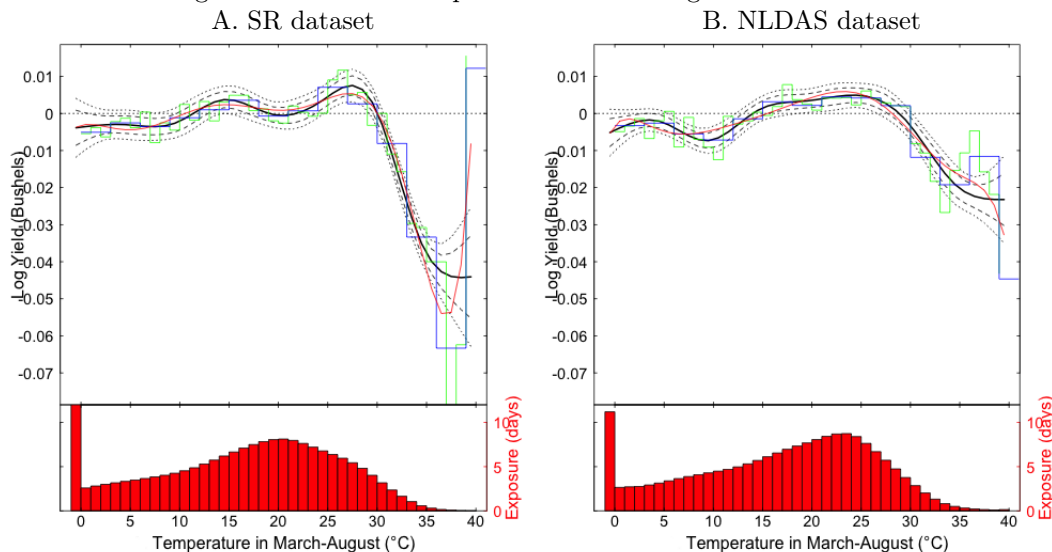
effects of high temperature under the SR dataset are roughly twice those obtained with the NLDAS dataset. As expected, the result based on the SR dataset is very similar to that reported in SR. Because the impacts are driven by extreme temperature, the NLDAS dataset suggests projected climate change impacts would be roughly reduced by half in this model.[9]

Thus, the lower recorded exposure to very high temperature in the SR dataset explains why extreme temperatures appear more detrimental when the analysis is based on their dataset. The average exposure difference in the March-August period to temperature above 35°C (7.9 hours) amounts to less than 0.2% of the length of the growing season. Yet, this small difference clearly seems to drive this result.

In order to confirm this claim, I construct two hybrid datasets for which I swap the exposure to extreme temperature (⩾35°C) across the SR and NLDAS datasets. I re-ran the models with the two hybrid datasets and results are shown in figure 5. Panel A (B) shows the models based on temperature bin exposure in the 0-34°C from the SR (NLDAS) dataset combined with bin exposure above 35°C from the NLDAS (SR) dataset. A comparison with figure 4 shows this swap of extreme temperature bins results in an exchange of the portion of the temperature response curves above 30°C.

---

[9] A non nested J-test based on the two weather datasets was inconclusive. This suggests that both weather datasets have distinctive explanatory power in explaining yield variation. Because exposure to extreme temperature is limited in the sample, an inability to statistically discriminate between datasets is unsurprising when comparing how well alternative datasets explain yield variation over the full sample period.

Figure 5: Results with hybrid datasets

A. SR (0-34ºC) + NLDAS (⩾35ºC) dataset

B. NLDAS (0-34ºC) + SR (⩾35ºC) dataset
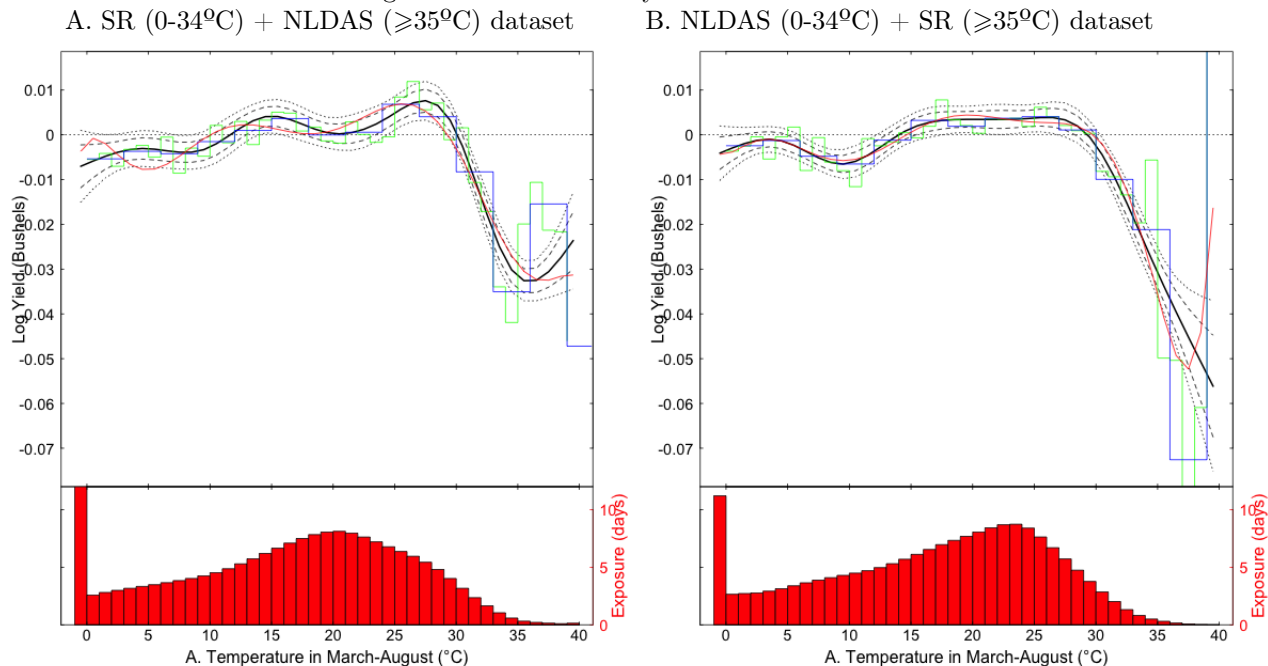
This result stems from the local flexibility given to each temperature bin in its effects on yield. These results illustrate that substantial differences in the detrimental effects of temperature can be driven by seemingly small absolute differences in the right tail of temperature distribution (⩾35ºC), well beyond the threshold level (29ºC). This explains how the two datasets can fit the yield data essentially equivalently well overall, but yet suggest remarkably different effects of very high temperatures on yield.

In section 2 I emphasized that the difference between datasets was not a systematic bias in temperature. Such bias would simply shift the temperature response function horizontally in figure 4, which it does not. Rather, the evidence clearly implies that the high detrimental effects of temperature obtained by SR stems from the "thickness" of the right tail of the temperature distribution in their dataset.

I have undertaken further analysis to verify that the difference in the right tail of the temperature distribution does not originate from the assumption of a sine-shaped temperature curve, which I do not report here. Similar results are obtained when a daily sine curve assumption is imposed to the NLDAS dataset as well.

5   Conclusion

SR have made an important methodological contribution by showing with large-scale weather

and production observational data that temperature has highly nonlinear effects on crop yield. They find major US crops would be dramatically affected under climate change with yield losses in the range of 30–82% depending on the crop and scenario.

However, this study shows their quantitative result is highly sensitive to the weather dataset. More precisely, the replication of their model based on an alternative high quality weather dataset suggests detrimental effects of roughly half those found with the SR dataset.

A closer look at the differences in temperature distributions between the SR and the NLDAS datasets reveals that the critical difference is that the latter dataset has a slightly "thicker" right tail. The difference is small in absolute terms, but it is large relative to the total exposure at the critical temperature levels. Such differences are shown to be the source of the discrepancy in the regression results.

The implications of these results for climate change impact assessments on agriculture are important. The work in SR and related studies suggest that extreme temperature is the greatest threat to agriculture under climate change. The evidence they present highlights the highly negative effects of extreme temperature on crop yields, which is also reflected in hedonic models of agricultural land prices. In parallel, they find that precipitation changes play a small role, even for rainfed areas. Thus, accepted wisdom is that climate change impacts will be driven primarily by the shift of the temperature distribution to the right. This would increase the frequency of exposure to detrimental temperature levels and generate large losses to the sector.

However, the simple replication carried in this study, based on an alternative dataset from a highly respected source employing advanced interpolation techniques, indicates that estimated impacts on corn yields could be half of those reported by SR based on the same model. Perhaps more importantly, results here raise questions regarding the robustness of the model to alternative weather datasets as results seem highly dependent on small differences in extreme temperature recordings.

Appendix

The specification for the model with a step function allowing different effects at each 3°C interval (SR2) is:

$$y_{it} = \sum_{h=0,3,6,9...}^{39} \gamma_h \underbrace{\left[\Phi_{it}(h+3) - \Phi_{it}(h)\right]}_{x_{it,h}} + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i + \epsilon_{it}$$

The model effectively regresses yield on the time spent within each interval in a given county and year $x_{it,h}$.

Model SR3 assumes that the "yield growth" function $g(h)$ is an eighth-degree polynomial of the

form $g(h) = \sum_{j=1}^{8} \gamma_j T_j(h)$ where where $T_j()$ is the jth order Chebyshev polynomial. Replacing $g(h)$ with this expression yields:

$$
\begin{aligned}
y_{it} &= \sum_{h=-1}^{39} \sum_{j=1}^{8} \gamma_j T_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)] + p_{it}\delta + z_{it}\tau + c_i + \epsilon_{it} \\
&= \sum_{j=1}^{8} \gamma_j \underbrace{\sum_{h=-1}^{39} T_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)]}_{x_{it,j}} + p_{it}\delta + z_{it}\tau + c_i + \epsilon_{it}
\end{aligned}
$$

The model effectively regresses yield on eight temperature variables $x_{it,j}$ which represent the jth-order Chebyshev polynomial evaluated at each temperature bin.

In a similar fashion, model SR4 assumes that $g(h) = \sum_{j=1}^{8} \gamma_j S_j^3(h)$ where $S_j^3()$ is the piece-wise cubic polynomial evaluated for each jth interval defined by eight control points.

$$
y_{it} = \sum_{j=1}^{8} \gamma_j \underbrace{\sum_{h=-1}^{39} S_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)]}_{x_{it,j}} + p_{it}\delta + z_{it}\tau + c_i + \epsilon_{it}
$$

14

Essay II

Is weather really additive in agricultural production? Implications for climate change impacts

Abstract

Recent reduced-form econometric models of climate change impacts assume climate is additive. This is reflected in climate regressors which are aggregated over several months that include the growing season. In this paper I develop a simple model to show how this assumption imposes implausible characteristics on the production technology which are in serious conflict with the agricultural sciences. The additivity assumption implies equal marginal productivities of weather as well as equal interactive effects of weather with endogenous inputs across all season stages. I test this assumption using a crop yield model of US corn that accounts for variation in weather at various times of the growing season. Results strongly reject additivity and suggest that weather shocks such as extreme temperatures are particularly detrimental toward the middle of the season around flowering time, in agreement with the natural sciences. I discuss how the additivity assumption tends to underestimate the range of adaptation possibilities available to farmers, thus overstating projected climate change impacts on the sector.

JEL Classification Codes: Q54, Q51, Q12

Keywords: climate change, agriculture, production, additivity

1 Introduction

Agriculture is arguably one of the most researched sectors in the climate change impacts literature. Statistical and econometric approaches have become increasingly popular among economists as an alternative to their earlier biophysical process-based counterpart. These empirical approaches exploit cross-sectional or time variation of observational data to recreate hypothetical counter-factual changes in local climate based on the revealed preference paradigm (e.g. Schlenker, Hanemann and Fisher, 2005, and Deschênes and Greenstone, 2007). The typical approach consists in estimating a reduced-form model capturing the sensitivity of agricultural production or welfare to changes in monthly or pluri-monthly climate variables and multiplying the resulting parameters by the predicted changes of climate variables under climate change to derive potential impacts on the sector.

A crucial challenge in this line of research is to choose the right climate variables. This is a difficult task because of the complex interactions of farmer behavior with crop growth and environmental conditions. Weather fluctuates and affects crop growth throughout the growing season and informed farmers adjust the timing and level of inputs accordingly. Attempts to capture too many of these bio-physical and behavioral complexities statistically quickly become subject to multicolinearity and spurious correlations (see Kaufmann and Snell, 1997, for a discussion).

Somewhat dichotomous approaches have developed in the literature. In the econometric literature, researchers have made somewhat arbitrary choices of variable types (e.g., precipitation, temperature, soil moisture) and time-frames of aggregation (pluri-monthly or monthly averages or totals) with little basis for discrimination other than model fit and parsimony. A more parsimonious model, i.e., with less parameters to estimate, may be chosen because it offers comparable predictive power despite violating agronomic wisdom. This seems to be the case regarding the choice of time-frame of aggregation for climate variables in this literature. Alternatively, in the agronomic and agricultural science-based literature, models have been grounded in agronomic principles and agricultural production experiments without considering behavior and revealed preferences of farmers. In practice, these objectives conflict. While agronomic science suggests that environmental conditions have varying effects throughout the season, some of the most influential econometric studies have relied on climate variables aggregated over several months. For instance, the econometric studies of Schlenker, Hanemann and Fisher (2006) and Deschênes and Greenstone (2007) aggregate climate variables over the April-September period while Schlenker and Roberts (2009, henceforth SR) regress crop yields on climate variables aggregated over the March-August (corn and soybeans) and April-October (cotton) periods.

Agriculture is well-known to be a time-sensitive activity and pluri-monthly aggregation of weather over the season seems at odds with this fundamental characteristic. This feature is not only known to farmers and agronomists, but to agricultural economists who have developed production models accounting for the sequential nature of its decision-making process (e.g., Mundlak and Razin, 1971; Antle, 1983). Season-long pluri-monthly windows for weather aggregation imply serious assumptions about the technology and farmer's ability to adapt in the long run. In particular, it imposes neutral technical change as well as implausible interactions of weather with endogenous farmer inputs and decisions. It also conceals the potential for farmer adaptation through changes in the timing of the growing season. Although this practice might be innocuous for short run forecasting purposes, it can have serious consequences for long-term climate change impact analysis.

The purpose of this paper is to explore whether the temporal additivity of weather assumption is valid for agricultural production. This is a prevalent premise in the econometric climate change literature for which consequences have received little attention. In my exposition, I develop a simple theoretical model to explore the implicit assumptions stemming from the adoption of a reduced form approach and the use season-long weather variables. By a reduced form model, I refer to a model that not only excludes the accompanying structure of how decision processes interact with changing technology, but also a model that aggregates some of the processes temporally for the purposes of empirical implementation. For clarity of exposition I focus on reduced form crop yield models, such as SR, that regress crop yields on weather variables. I then explore this question empirically and test for weather additivity using a 31-year balanced panel of US county-level corn yields representing 70% of US production. Results suggest that weather effects are not additive, and rather that extreme temperatures are particularly detrimental during the middle of the growing season. I then discuss some of the implications of these findings for adaptation and the related shortcomings of assuming additivity of weather in the context of climate change impact analysis.

The paper is organized as follows. In section 2 I develop a simple theoretical model to illustrate the implicit assumptions of time aggregation in reduced-form crop yield models which are widely used in this literature. In section 3 I present an empirical model to explore effects of these assumptions and discuss the results and implications for climate change impact analysis. Section 4 concludes.

2 Implicit assumptions of weather additivity

To explore the implications of imposing additivity of weather inputs in reduced form crop yield models, consider an underlying optimization model in which the farmer makes decisions sequentially during the season. Input decisions in a stage of the growing season $s = \{1, ..., S\}$ are made with uncertainty about remaining weather and are conditioned on decisions already made as well as weather already observed. Assume a risk-neutral farmer. The expected profit maximization problem is:

$$\max_{x_{ts}} E_s \left[p_t y_t - \sum_{j=1}^{s} r_{tj} x_{tj} - \sum_{j=s+1}^{S} r_{tj} x_{tj}^*(x_{ts}) \mid w_{t1}, ..., w_{t,s-1}, x_{t1}, ..., x_{t,s-1} \right] \tag{1}$$

subject to

$$y_t = f_t \left(x_{t1}, ..., x_{tS}, w_{t1}, ..., w_{tS} \right) + \epsilon_t \tag{2}$$

where E_s represents an expectation at the beginning of crop stage s, input price vectors r_j apply to input vectors x_{tj} chosen at each growing stage j, output price p_t applies to yield y_t which depends on variation in weather w_{ti} as well as input decisions x_{tj} that apply through the various stages of the growing season. $x_{tj}^*(x_{ts})$ represents the optimal input decision vector at future stages given input decisions at all prior production stages $x_{t1}, ..., x_{t,s-1}$ and all prior observed weather $w_{t1}, ..., w_{t,s-1}$ during the growing season as well as the current decision x_{ts} at stage s assuming all future decisions will be made optimally given further weather realizations. Yield can be represented generally as shown in (2) where ϵ_t is a random error in production.

Changing Technology

The technology denoted by f_t must be considered as changing over time t for the long-run nature of climate change analysis. If output price and yield are considered uncorrelated at the individual farmer level and the output price expectation does not vary with the crop stage (for conceptual simplicity), then the first-order conditions for (1) after substituting (2) are:

$$E(p_t) \frac{\partial}{\partial x_{ts}} E \left[f_t \left(x_{t1}, ..., x_{t,s-1}, x_{ts}, x_{t,s+1}^*, ..., x_{tS}^*, w_{t1}, ..., w_{tS} \mid w_{t1}, ..., w_{t,s-1}, x_{t1}, ..., x_{t,s-1} \right) \right] - r_{ts} = 0 \tag{3}$$

Clearly, this optimization process, which is solved by backwardation, causes input decisions at stage s to depend on weather variables at prior stages $w_{t1}, ..., w_{t,s-1}$. This is a direct theoretical

18

reason why interactions among weather and input decisions arises. Further, input decisions and weather variables could also be correlated because some input decisions can affect vulnerability of crops to future weather during the growing season (see Just and Pope, 1979).

Correlation of Input Choices and Weather

Yield is a central component of farmer profit and therefore an important channel for analyzing climate change impacts. The literature readily recognizes that accuracy of climate change impact assessments for agriculture depend on representing yield impacts correctly. While representing the full complexity of this decision model is impractical, a first approximation of the crop yield model implied by this optimization problem might be represented as:

$$y_t = T_t\alpha + X_t\beta + W_t\gamma + Z_t\delta + \epsilon_t \tag{Model 1}$$

where $T_t = [1, t]$, $X_t = [g(x_{t1}), \ldots, g(x_{tS})]$, $W_t = [h(w_{t1}), \ldots, h(w_{tS})]$, and Z_t is a vector of functions of applicable interaction terms among and between input decisions and weather. The g and h functions represent nonlinear effects of inputs and weather variables, respectively. For instance, the functional form of h could capture well-known detrimental effects on crop yield of high temperatures and extreme precipitation levels.

In contrast, however, standard practice in the literature omits farmer inputs and interactions, which reduces Model 1 to a further simplified form,

$$y_t = T_t\alpha^* + W_t\gamma^* + u_t \tag{Model 2}$$

where the constant term in α^* is implicitly modified by $\bar{X}\beta + \bar{Z}\delta$ and the error term implicitly represents $u_t = X_t\beta - \bar{X}\beta + Z_t\delta - \bar{Z}\delta + \epsilon_t$. This simplification, however, can severely bias estimates of the γ^* parameters that are used to assess the impacts of climate change. The reason is that weather variables are correlated with the omitted input variables and with the interactions of terms among and between input and weather variables.

One reason economists in this field have been willing to use highly approximating specifications is that the interest is not in unbiased estimates of γ, but in unbiased long-run yield forecasts $\Delta y = \Delta W\gamma^*$. However, such forecasts based on Model 2 assume that the correlation between W_t and u_t remains unchanged as climate change occurs. In other words, W_t functions as a proxy not only for weather conditions but also for correlated farmer behavior. This assumption

is violated if the conditioning of optimal input choices on previous input decisions and weather $x_{ts}^* \left(x_{t1}, ..., x_{t,s-1}, w_{t1}, ..., w_{t,s-1} \right)$ change. This means that if the correlation of the proxy with the omitted variable of interest changes in the forecasting period then forecasts are biased. This can occur for various reasons.

Non-neutral Technical Change

First, Model 2 imposes neutral technical change. This is problematic for assessing climate change effects because the parameters in γ^* are likely functions of t. This would alter the optimal choice x_{ts}^* and thus change the correlation between W_t and u_t which would bias forecasts. This also occurs when technical change is induced by climate change, which would make γ a function of long run climate.[1] Second, changes in relative prices can lead to changes in optimal input use as shown in (3). Such changes may be major over a long time period, and thus introduce an additional source of bias for long-run yield projections.

Weather Aggregation Bias

The simplification that is also taken in most econometric studies of climate change analysis is to use season-long measures of weather. For US corn a common period is March to August because it spans the full growing season for most producing regions. This practice imposes the same γ parameters on weather variables in all stages of crop growth in the same season:

$$y_t = T_t \alpha^* + h \left(\sum_{j=1}^{S} w_{tj} \right) \bar{\gamma} + v_t \qquad \text{(Model 3)}$$

where the number of parameters in γ^* is reduced by a factor of S to obtain $\bar{\gamma}$ and the error term is implicitly further modified to $v_t = \sum_{j=1}^{S} \left(h(w_{tj}) - h(\bar{w}_j) \right) \gamma + X_t \beta - \bar{X}\beta + Z_t \delta - \bar{Z}\delta + \epsilon_t$.[2] This third model assumes $\left[h(w_{t1}), \ldots, h(w_{tS}) \right] \gamma^* = h \left(\sum_{j=1}^{S} w_{tj} \right) \bar{\gamma}$, which assumes h is factor-wise separable. This implies that weather realizations $w_{t1}, ..., w_{tS}$ are perfect substitutes within the growing season. However, this is assumption is in serious conflict with evidence from the agricultural sciences.

Aggregation of weather effects throughout the growing season can be very misleading. Extreme weather events are not equally likely across stages of the growing season. Many crops across the Midwest are planted in the spring and harvested in the fall when temperatures are cooler. The

---

[1] It is possible to allow γ^* to vary over time (e.g. Roberts and Schlenker, 2011) but estimates are inevitably confounded with potential time-trends in β. For instance, if inputs are becoming more productive (increasing β) but make crops more vulnerable to weather shocks ($\delta < 0$), γ^* may well appear becoming more detrimental over time despite γ and δ actually remaining unchanged.

[2] The expression in (Model 3) can be easily adapted to allow season-long averages of weather variables.

Table 1: Characteristics of yield models

Model	Specification	Error term	Inputs β	Weather γ	Interactions δ
1	$y_t = T_t\alpha + X_t\beta + W_t\gamma + Z_t\delta + \epsilon_t$	$\epsilon_t \overset{iid}{\sim} \mathcal{N}(\mu,\sigma^2)$	U	U	U
2	$y_t = T_t\alpha^* + W_t\gamma^* + u_t$	$u_t = X_t\beta - \bar{X}\beta + Z_t\delta - \bar{Z}\delta + \epsilon_t$	C	U	C
3	$y_t = T_t\alpha^* + h\left(\sum_{j=1}^S w_{jt}\right)\bar{\gamma} + v_t$	$v_t = \sum_{j=1}^S \left(h(w_{tj}) - h(\bar{w}_j)\right)\gamma + u_t$	C	E	C+E

U: Parameters are unrestricted; C: Parameters are constant over time but may differ across stages;
E: Parameters are equal across stages.

middle of the season, which includes the sensitive flowering stage, typically occurs in the summer months when extreme temperatures are more prevalent. As a result temperature shocks aggregated over the entire season may appear to be detrimental to *all* crop stages, rather than to the most sensitive stages of crop growth.

Moreover, the model also presumes that the correlation between season-long weather variable $h\left(\sum_{j=1}^S w_{tj}\right)$ and the unexplained residual v_t would remain constant under climate change. Given the underlying model in (1) is sequential this presumes the first order conditions in (3) remain unchanged for any sequence of weather variables $w_{t1}, ..., w_{t,s-1}$ with the same sum, that is,

$$\frac{\partial}{\partial x_{ts}} E\left[f_t\left(x_{t1}, ..., x_{t,s-1}, x_{ts}, x^*_{t,s+1}, ..., x^*_{tS}, w_{t1}, ..., w_{tS} \mid w_{t1}, ..., w_{t,s-1}, x_{t1}, ..., x_{t,s-1}\right)\right]$$

$$= \frac{\partial}{\partial x_{ts}} E\left[f_t\left(x_{t1}, ..., x_{t,s-1}, x_{ts}, x^*_{t,s+1}, ..., x^*_{tS}, w_{t1}, ..., w_{tS} \mid \sum_{j=1}^{s-1} w_{tj}, x_{t1}, ..., x_{t,s-1}\right)\right]$$

which suggests no interactions among and between weather and endogenous inputs across stages. As an illustration, this suggests that farmers would time fertilizer, pesticide, and irrigation applications independently from weather. This is obviously incorrect. The preeminent role of weather forecasts in agricultural production constitutes a clear counterexample.

Summary

Table 1 summarizes the key points of this section. Yield forecasts under climate change based on all three models assume constant relative prices, a likely artifact of the unpredictability of relative prices far in the future. Model 1 can accommodate more general forms of technical change than model 2, but both models 1 and 2 impose neutral technical change. In that sense, model 1 has wider applicability as it allows the exploration of interactive effects of weather and endogenous farmer inputs. This could include analysis of input uses that attenuate vulnerability to weather shocks.

However, such farmer behavior is generally poorly observed. Model 2 offers an alternative spec-

ification that omits farmer inputs and interaction effects. This model presents a somewhat flexible form to explore implicitly the potential changes in yield with varying effects throughout the season. However, because the functional form assumes neutral technical change, input and interactive parameters are assumed to remain constant over the sample and forecasting periods.

Model 3, which is the widely used specification in the literature, imposes additive weather and thus implies equal marginal productivities of weather as well as equal interactive effects of weather with endogenous inputs across all season stages. This is in additional to assuming constant weather and interactive parameters as for model 2. To explore the validity of the additivity assumption in model 3, I rely on model 2 and test for equality of parameters across stages. I develop an empirical model for this purpose in the next section.

3   Empirical exploration

Data

To explore the assumption of weather additivity in crop yield models requires matching data on weather conditions with crop progress at various times over the growing season in each location. The highly detailed weather data used in this paper is based on the North American Land Data Assimilation System (NLDAS) which is a joint project of the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), Princeton University and the University of Washington. The NLDAS weather dataset features hourly and 14-km resolution and has been shown to closely match observations of highly precise weather stations in the Great Plains (Cosgrove et al., 2003). These data thus allow considerable local specificity. For instance, Indiana, which has the lowest average county size in the Midwest (1,025 km^2), includes over five NLDAS grids per county on average.

Agricultural data are obtained from the U.S. Department of Agriculture's National Agricultural Statistics Service (USDA-NASS). Because corn production data is available only at the county-level, the hourly gridded weather data must be spatially aggregated for each county. I do so by weighting each NLDAS data grid within a county by the amount of cropland within each grid. The cropland area is derived from the USDA-NASS's 2011 Crop Data Layer which has a 30-meter resolution. This allows weighting NLDAS data grids according to the amount of farmland they include in constructing the county-level observations. Hourly observations were subsequently used to construct exposure to individual degree-bins for the March-August period for each year and county.
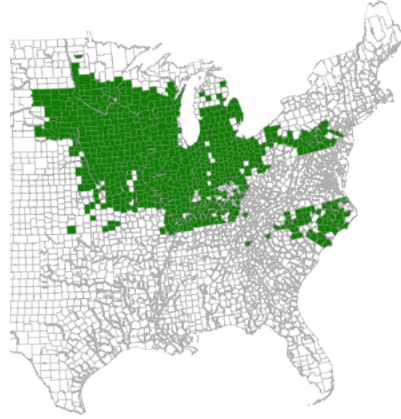
Figure 1: Rainfed counties in the sample

Because rainfed and irrigated corn yields are expected to respond differently to exogenous environmental conditions, their respective parameters are estimated separately. For this purpose, I restrict the sample to counties where at least 75% of the acreage, on average, is rainfed. Figure 1 illustrates where the sample counties are located. The dataset corresponds to a balanced corn yield panel of 800 Midwest rainfed counties for 1981-2011, which represents 70% of US corn production.

My major focus in this paper is to evaluate the validity of time-aggregation of weather variables throughout the growing season. In order to do so, I account for variation in weather conditions throughout the growing season. This allows estimation of possible varying effects from intra-seasonal environmental conditions on crop yield. Accounting for the timing effect using standard agronomic principles requires information on crop stages. I thus rely on the Crop Progress and Conditions weekly survey by USDA/NASS which provides state-level data on farmer activities and crop phenological stages from early April to late November. Reporting across states and years is not balanced. Although state reports date back to 1979, reporting for corn that includes both the onset (planting/emergence) and the end of the season (maturation/harvesting) begin in 1981 for the major producing states.

Specifically, this survey reports the percentage of a state's corn acreage undergoing certain farming practices and reaching specific crop stages.[3] As a consequence, it does not offer clear "boundary" dates between stages because of the timing variations within states.[4] For the purpose of defining

---

[3]The report includes progress of farming activities (planting and harvesting) and of corn phenological stages (emerged, silking, doughing, dented and mature). The USDA defines these crop stages as follows. Emerged: as soon as the plants are visible. Silking: the emergence of silk-like strands from the end of corn ears, which occurs approximately 10 days after the tassel first begins to emerge from the sheath or 2-4 days after the tassel has emerged. Doughing: normally half of the kernels are showing dent with some thick or dough-like substance in all kernels. Dented: occurs when all kernels are fully dented, and the ear is firm and solid, and there is no milk present in most kernels. Mature: plant is considered safe from frost and corn is about ready to harvest with shucks opening, and there is no green foliage present.

[4]Visual inspection of district-level crop progress reports, which are available for only a few states, surprisingly

such boundaries of the growing season for each county, I obtain stage median acreage dates. These correspond to the dates at which 50% of the acreage in a given state has reached each stage in a given year.[5]

Crop stages reported by the USDA are not equally spaced in the growing season. They arguably correspond to visible markers that can be easily verified to simplify data collection. Some past studies (e.g.,Kaufmann and Snell, 1997) have relied on weather variables matched to precise crop stages. However, results are sometimes difficult to interpret, especially for non-agronomists. In order to convey a more accessible crop advancement metric, I divide the growing season into eight segments centered around flowering (i.e. silking), which is considered the midpoint of the season. Four equally-spaced periods occur in the vegetative phase (between planting and silking) and four equally-spaced periods occur in the reproductive or grain-filling phase (between silking and maturation). For simplification, the crop advancement division is converted into percentages with intervals of 12.5%. Thus, the 0-12.5% and 87.5-100% stages correspond, respectively, to the first and last segments just after planting and just before maturation, and 37.5-50% and 50-62.5% correspond, respectively, to the segments just before and just after flowering.

Natural scientists have found that crop development or phenology is proportional to accumulated Growing Degree-Days (GDD, see e.g. Hodges, 1991; Smith and Hamel, 1999; Fageria et al., 2006; Hudson and Keatley, 2009). This variable is defined by the area under the temperature-time curve that falls between two temperature thresholds (10 and 30ºC for corn) during a given period of time. Warmer conditions generally lead to faster GDD accumulation and more rapid crop development. This concept can be used to split the growing season into equally-spaced segments.

Following this approach, I compute a cumulative GDD variable starting at planting for each state and year and use it to represent the eight segments of the season. Figure 2 illustrates how these seasonal segments are located in the 2001 calendar for Illinois. Although the segments have a different number of days, segments 1-4 and 5-8 are equally spaced in terms of GDD. Thus, wider segments signal slower development due to cooler conditions.

Exposure to temperature bins is aggregated within each of these segments. As a result, the temperature variables account for exposure to different temperature levels during the eight individual segments of the growing season. This allows assessment of how sensitivity to temperature varies with crop advancement. Finally, because exposure to some temperature levels is low or nonexistent for

reveals variation similar to overall state progress for most years.

[5]For a few states and years, crop progress reporting began too late (the state had already surpassed the 50% acreage level) or stopped too early (the state had not yet reached the 50% acreage level). For these cases, which represent less than 5% of the cases, I obtained the median acreage date by extrapolation.
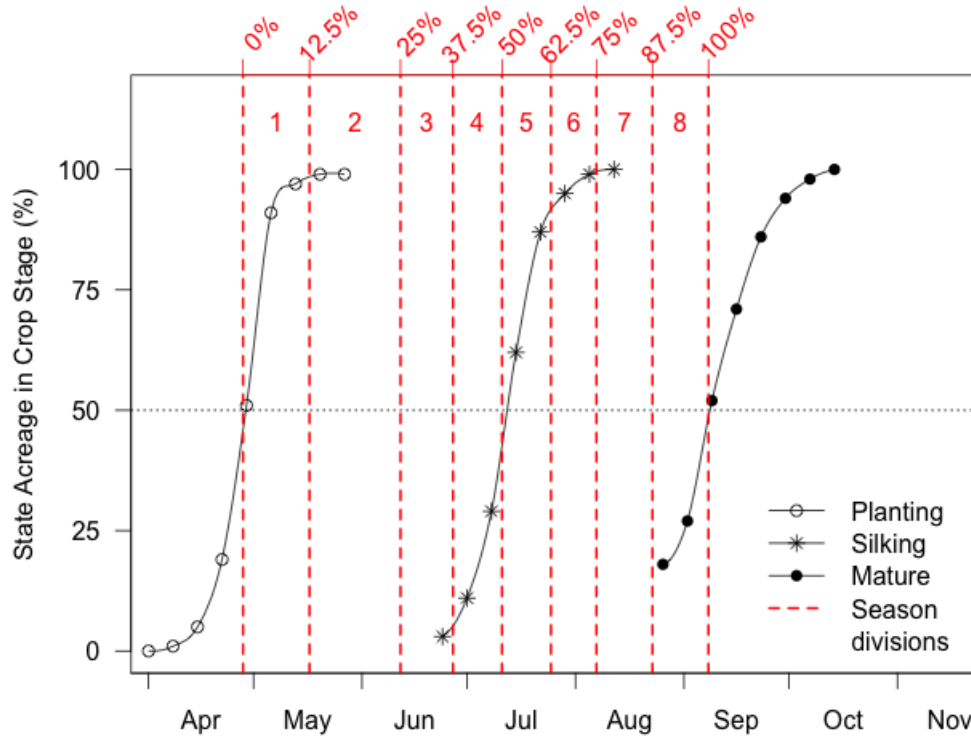
Figure 2: Season divisions for Illinois corn in 2001.

some crop stages, I aggregate bins at the extremes that do not represent more than 0.15% of the growing season on average over the sample period.

The summary statistics for each stage are presented in table 2. As expected, the early and late parts of the season are slightly longer in terms of days. Total average precipitation by stage is fairly even when adjusting for crop stage length. Also, exposure to low (high) temperatures are more likely in the extreme (middle) parts of the season. This point should be kept in mind when assessing the effects of extreme temperature on crop yields. It is clear that exposure to high temperature ($\geqslant 30^{\circ}$C) is considerably higher on average for stages 4 through 7, spanning over the middle of the season

Table 2: Summary statistics of weather variables by stage

Stage	1	2	3	4	5	6	7	8	1 - 8
Advancement (%)	0-12.5	12.5-25	25-37.5	37.5-50	50-62.5	62.5-75	75-87.5	87.5-100	0-100
Length (days)	25.1	17.2	15.1	13.9	13.1	13.0	13.9	18.9	130.0
Precipitation (mm)	90.4	65.6	52.6	48.9	42.4	40.4	43.7	55.9	439.8
Exposure (hours)									
0-10°C	58.8	0.0	0.0	0.0	0.0	0.0	0.0	35.5	94.2
10-20°C	330.9	147.8	85.0	61.9	58.6	71.7	96.0	203.2	1055.1
20-30°C	208.4	245.5	246.0	232.7	214.0	201.5	200.3	189.2	1737.8
$\geqslant$30°C	9.5	31.2	45.9	53.9	55.6	53.0	49.7	34.2	332.9

when corn flowering occurs.

Model

In this section I present an empirical version of Model 2 to test for weather additivity throughout the growing season. The model approximates the reduced form crop yield model in SR, which introduced an innovative approach to estimate the effect of exposure to different levels of temperature on yield separately using temperature bins. A key characteristic of their study is that the exposure to different temperature levels is computed during the entire season (March-August for corn). Thus, their estimated model uses season-long weather variables as in Model 3. To generalize to the case of Model 2, I relax the additivity assumption and explore different response functions throughout the growing season.

While the SR model assumes that temperature effects on yield are cumulative and substitutable over time, the nonlinear effects of temperature on yield are captured by the function $h(w)$ representing "yield growth" that depends on temperature w. Function h is obviously homologous to the function by the same name presented in the theoretical section. Logged corn yield y_{it} in county i and year t are represented as:

$$y_{it} = \int_{\underline{w}}^{\overline{w}} h(w)\phi_{it}(w)dw + p_{it}\gamma_1 + p_{it}^2\gamma_2 + z_{it}\alpha + c_i + \epsilon_{it} \tag{4}$$

where $\phi_{it}(w)$ is the time distribution of temperature (i.e., the temperature-time path) for March-August, p_{it} is precipitation, z_{it} is a state-specific quadratic time trend and the c_i are county fixed-effects. In order to relax the additivity assumption I allow h, γ_1 and γ_2 to vary within the growing season. The following model introduces this flexibility:

$$y_{it} = \int_{\underline{s}}^{\overline{s}} \int_{\underline{w}}^{\overline{w}} h(w,s)\phi_{it}(w,s) + p_{it}(s)\gamma_1(s) + p_{it}^2(s)\gamma_2(s)dwds + z_{it}\alpha + c_i + \epsilon_{it} \tag{5}$$

where $\phi_{it}(w,s)$ is the time distribution of temperature at each stage of the season s. Note that s indicates the advancement of the growing season for a given year and location. Equation (5) cannot be estimated directly because of the double integral. Therefore, I follow an approach similar to SR and approximate the integral as a sum according to four alternative approaches: a step function allowing different effects at each 1°C interval (S1), a step function allowing different effects at 3°C intervals (S2), an eighth-degree polynomial (S3), and a cubic B-spline with eight degrees of freedom

(S4). This is done for each of eight stages of the season. The specification for S1 is:

$$y_{it} = \sum_{s=1}^{8} \sum_{w=0}^{40} \underbrace{h(w+0.5,s)}_{\gamma^h_{ws}}[\Phi_{it}(w+1,s) - \Phi_{it}(w,s)] + p_{its}\gamma_{1s} + p_{its}^2\gamma_{2s} + z_{it}\alpha + c_i + \epsilon_{it} \quad (6)$$

where $\Phi_{it}(w,s)$ is the cumulative time in temperature bin w in county i and year t in stage s.[6] Specifications for S2, S3, and S4 are provided in the appendix. Because unobserved explanatory factors are likely to be spatially correlated, I account for spatial correlation of the errors in the estimation.

Results

Figure 3 presents the temperature response function for each stage of the growing season. There is agreement between the four specifications for each individual stage, although some discrepancies can be perceived at extreme temperature values. The most important result is that the response functions clearly differ by stage of the growing season. For instance, exposure to temperatures exceeding 30ºC have detrimental effects on crop yield toward the middle of the growing period (e.g., 37.5-50%) but not toward the end of the season (e.g., 87.5-100%). This confirms wisdom from the agricultural sciences that crop growth around the flowering stage of corn is the most sensitive to environmental stress. Note that temperature response functions do not extend over the same range of temperature for all stages (e.g. they do not extend to temperatures higher than $30ºC$ for the 0-12.5% stage). This is due to lack of observations of extreme values at some stages of the growing season.

I performed a simultaneous test of equality of weather parameters across stages for both temperature and precipitation effects to confirm the visual differences.[7] An asymptotic chi-square test rejects equality of parameters across stages with p-values below 2×10^{-16}. Thus there is strong evidence that both the temperature and precipitation response functions vary throughout the season.

These generalizations have important implications for climate change impact analysis. If climate change is particularly detrimental through an increased exposure to very high temperatures, then this type of weather shock is more likely to occur in the summer, which is roughly around the middle

---

[6]Note that because some temperature levels do not occurred in some stages, the associated parameters are not estimated. For instance, $h(0.5, s = 5) = 0$, because temperature around 0ºC never occur in the fifth stage, thus the associated parameter $\gamma^h_{0,5}$ is not estimated.

[7]The test was conducted for both step function specifications which offer clear parameter equivalence across stages. The spline and polynomial specifications have eight parameters per stage but are defined over different temperature ranges which does not allow comparing these eight parameters directly across stages in a meaningful way.
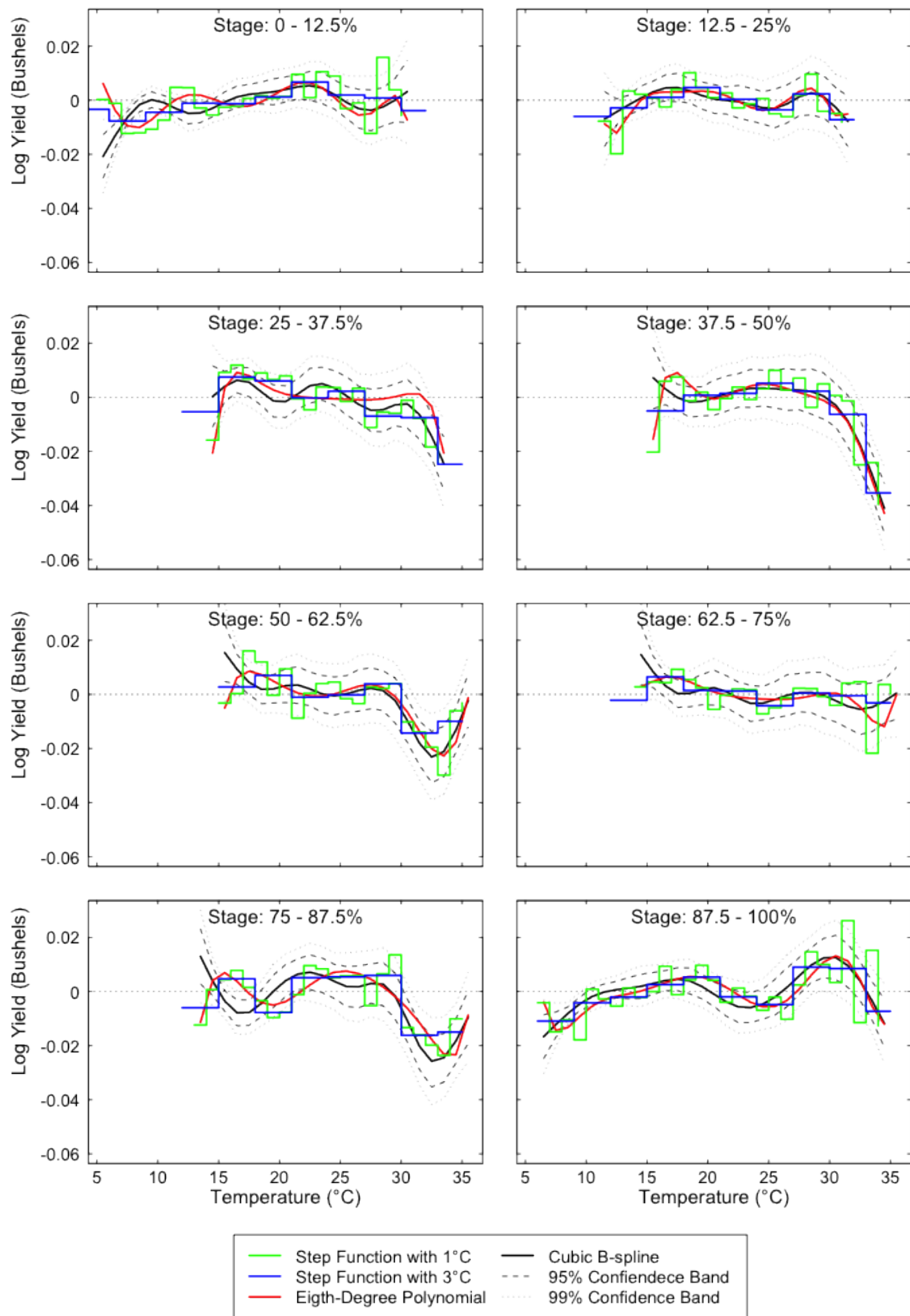
Figure 3: Temperature response by stage

28

of the current growing season for corn in the Midwest. As figure 3 reveals, extreme temperature is more damaging toward the middle of the season. As a consequence, in the presence of climate change with an increase in the frequency of high temperatures, farmers would be able to shift the growing season. A shift of, even a few weeks can be sufficient to reduce exposure during the most heat-sensitive stages of growth. Some part of the growing season would still be affected by extreme temperature, but the detrimental effects would be reduced. This type of adaptation scenario is made possible when weather is treated as non-additive. Ortiz-Bobea and Just (2013) explore this possibility and show sizable benefits for farmers of changing the timing of the growing season through changes in planting dates.

On the other hand, models that impose additivity of weather imply much more limited possibilities for adaptation. The reason is that shifting the growing season does not lead to any sizable estimated benefit because weather shocks are assumed to have the same effects over all parts of the growing season. This is particularly relevant when considering that climate change is estimated to change intra-seasonal climate patterns.

4   Conclusion

One of the crucial challenges in empirical studies that assess the potential impacts of climate change on agriculture is the choice of the right climate variables. A common practice in this literature is to rely on season-long variables because it leads to parsimonious models with relatively high levels of statistical fit. In contrast, this paper I shows that the underlying assumption of this approach is invalid. While a reduced-form model with non-additive weather may provide some insights in how production might change in response to a change in climate, the additional assumption of weather additivity introduces strong restrictions which are at odds with the accepted wisdom of agronomic science. It not only assumes that all marginal productivities of weather variables are equal across stages of crop growth, and that weather variables are perfect substitutes among states, but also that the interaction with time-sensitive endogenous inputs (e.g., fertilizer and pesticides) is constant no matter when the weather input is realized. This latter point would imply that farmers do not rely on weather forecasts for adjusting the timing of input decisions, which is obviously incorrect.

Based on an empirical analysis of US corn yields, I show that both temperature and precipitation effects statistically differ across stages. Results strongly reject the additivity assumption. Crop yield is shown to be especially sensitive to temperatures exceeding 30⁰C toward the middle of the season

during the flowering period. The same temperature levels do not seem to affect yield when they occur close to the end of the season at maturation time.

Relaxing the additivity assumption in this literature can open the door to assessment of a richer set of adaptation possibilities. Climate change impact studies that restrict the range of farmer adaptation will inevitably overestimate potential costs. In that sense results stemming from current yield models are pessimistic.

Appendix

The specification for the model with a step function allowing different effects at each 3°C interval (S2) is:

$$y_{it} = \sum_{s=1}^{8} \sum_{w=0,3,6,9...}^{39} \gamma_{ws}^{h} \underbrace{[\Phi_{it}(w+3,s) - \Phi_{it}(w,s)]}_{x_{its,h}} + p_{its}\gamma_{1s} + p_{its}^{2}\gamma_{2s} + z_{it}\alpha + c_{i} + \epsilon_{it}$$

The model effectively regresses yield on the time spent within each interval in a given county and year $x_{its,h}$.

Model S3 assumes that the "yield growth" function $h(w)$ is an eighth-degree polynomial of the form $h(w,s) = \sum_{k=1}^{8} \gamma_{ks}^{h} T_{ks}(w)$ where where $T_{k}()$ is the kth order Chebyshev polynomial. The h superscript in γ simply differentiates temperature parameters from precipitation parameters. Replacing $g(h)$ with this expression yields:

$$
\begin{aligned}
y_{it} &= \sum_{s=1}^{8} \sum_{w=-1}^{39} \sum_{k=1}^{8} \gamma_{ks}^{h} T_{ks}(w+0.5,s)[\Phi_{it}(w+1,s) - \Phi_{it}(w,s)] + p_{its}\gamma_{1s} + p_{its}^{2}\gamma_{2s} + z_{it}\alpha + c_{i} + \epsilon_{it} \\
&= \sum_{s=1}^{8} \sum_{k=1}^{8} \gamma_{ks}^{h} \underbrace{\sum_{w=-1}^{39} T_{ks}(w+0.5,s)[\Phi_{it}(w+1,s) - \Phi_{it}(w,s)]}_{x_{its,k}} + p_{its}\gamma_{1s} + p_{its}^{2}\gamma_{2s} + z_{it}\alpha + c_{i} + \epsilon_{it}
\end{aligned}
$$

The model effectively regresses yield on eight temperature variables $x_{its,k}$ which represent the kth-order Chebyshev polynomial evaluated at each temperature bin.

In a similar fashion, model S4 assumes that $h(w,s) = \sum_{k=1}^{8} \gamma_{ks} S_{ks}^{3}(w)$ where $S_{k}^{3}()$ is the piece-

wise cubic polynomial evaluated for each jth interval defined by eight control points.

$$y_{it} = \sum_{s=1}^{8} \sum_{k=1}^{8} \gamma_{ks}^{h} \underbrace{\sum_{w=-1}^{39} S_{ks}(w+0.5, s)[\Phi_{it}(w+1, s) - \Phi_{it}(w, s)]}_{x_{its,k}} + p_{its}\gamma_{1s} + p_{its}^{2}\gamma_{2s} + z_{it}\alpha + c_i + \epsilon_{it}$$

Essay III

Understanding Temperature and Moisture Interactions in the Economics of Climate Change Impacts and Adaptation on Agriculture

Abstract

Growing econometric and statistical evidence points to high temperature as the main driver of large negative effects of climate change on US agriculture. This literature also suggests a limited role for precipitation in overall impacts. This paper shows this finding stems from the widespread use of calendar precipitation variables, which poorly represent water availability for rainfed crops. I rely on a state-of-the art dataset with very high spatial (14km) and temporal (1h) resolution to develop a statistical model and unpack the effects of temperature and drought stress and analyze their interactions. Using a 31-year panel of corn yields covering 70% of US production, I account for nonlinear effects of soil moisture with varying effects throughout the growing season, in addition to nonlinear temperature effects. I show that yield is highly sensitive to soil moisture toward the middle of the season around flowering time. Results show that omission of soil moisture leads to overestimation of the detrimental effects of temperature by 30%. Because climate change affects intra-seasonal soil moisture and temperature patterns differently, this omission also leads to very different impacts on US corn yields, with a much greater role for water resources in overall impacts. Under the medium warming scenario (RCP6), models omitting soil moisture overestimate yield impacts by almost 100%. The approach shows a more complete understanding that climate change impacts on agriculture are likely to be driven by both heat and drought stresses, and that their relative role can vary depending on the climate change scenario and farmer ability to adapt.

JEL Classification Codes: Q54, Q15, Q51, R15

Keywords: climate change, agriculture, impacts, adaptation, drought, temperature stress, nonlinear effects, omitted variable bias, spatial error panel model.

1   Introduction

Agriculture is arguably one of the most vulnerable sectors to climate change. Much economic work has focused on developing econometric approaches to evaluate overall impacts of climate change on the sector implicitly (Mendelsohn, Nordhaus and Shaw, 1994; Schlenker, Hanemann and Fisher, 2005; Deschênes and Greenstone, 2007). Controversy even on the sign of these impacts persists and remains unresolved because of the inherent vulnerability of these highly-reduced-form approaches to various forms of omitted variable bias (see Deschênes and Greenstone, 2007; Fisher et al., 2012). Although these approaches differ by the structure of the underlying data (cross-sectional or panel), they share an important common characteristic in how they capture water availability and heat effects through the use of precipitation and temperature variables.

Most innovation in econometric climate change impact studies regarding climate variables concerns the measurement of heat exposure. In their seminal hedonic paper, Mendelsohn, Nordhaus and Shaw (1994) regressed US land price data on linear and quadratic terms of average monthly precipitation and temperature for the months of January, April, July and October. However, Schlenker et al. (2006) triggered a small revolution by suggesting that monthly averaging eliminates valuable information regarding daily exposure to very high temperatures. They proposed accounting separately for the cumulative exposure to moderate (8-32ºC) and high (34ºC) temperatures over the entire growing season. This approach has been found to improve the fit of the hedonic model, and can be found in leading studies such as Schlenker et al. (2005) and Deschênes and Greenstone (2007).

Further work on this area has been carried out by Schlenker and Roberts (2009, henceforth SR), who have developed the most advanced approach to date for capturing the nonlinear effects of temperature on crop yields. They make use of highly detailed weather data and flexible semi-parametric techniques that allow each temperature bin to have separate effects on yield.

Although econometric and crop yield studies have attempted to account for heat in increasingly flexible ways, little attention has been given to how these studies treat water availability. Most studies simply rely on monthly or pluri-monthly precipitation variables. A possible explanation is the growing consensus across econometric and statistical crop yield studies that precipitation plays a limited role in climate change impacts. My word shows that improving the representation of water availability has been undervalued.

For instance, based on worldwide observational data, Lobell and Burke (2008) explore the relative role of temperature, precipitation, and choice of climate model on climate change impact uncertainty. They find for most crops and regions that uncertainties related to temperature, in particular yield

sensitivity to temperature, represents a greater contribution to climate change impact uncertainty related to precipitation. They conclude that understanding crop responses to temperature is one of the most important needs for climate change impact assessments and adaptation efforts for agriculture.

The growing consensus from econometric models (e.g. Schlenker et al., 2005) and statistical yield models is that climate change impacts will be largely driven by exposure to heat. For instance, SR find that substituting a single full day of the growing season at 29°C with a full day at 40°C translates into a predicted decline of 7% for corn yields holding all else constant. According to this study, corn, soybean and cotton yields would decrease by 30–46% before the end of the century under the slowest warming scenario, and by 63–82% under the most rapid warming scenario, if current growing regions and seasons remain fixed. A surprising result is that a hypothetical drop of 50% in precipitation reduces corn yield by just over 10%.

The evidence that changes in precipitation may have only a marginal role in overall climate change impacts presages a dire future for US agriculture. Indeed, it implies that water management practices that provide greater control of soil moisture, such as irrigation, would not offer a significant counterbalancing effect to yield losses from heat stress.

However, this evidence is difficult to reconcile with agronomic experimental evidence. For sentence, yield reductions in excess of 90% for corn can occur when water-deficits span key stages of the season (NeSmith and Ritchie, 1992). A possibility is that heat and drought stresses are statistically confounded in the modeling efforts to date. This is plausible for three major reasons. First, heat waves and drought have a well-known interconnection. Second, drought significantly affect crop yields. Third, variables used to capture water availability for rainfed crops, such as precipitation aggregated over several months, are a poor representation of water supply. in the form of soil moisture, which is the form in which it matters for crop production.

Relying on season-long precipitation as a measure of water availability to crops has potentially crucial shortcomings. A pivotal concern is the implicit assumption that rainfall is a perfectly substitutable input over time within a season. This implies that it does not matter when it rains as long as it rains within the season. The agronomic literature suggests otherwise and, specifically, that crop sensitivity varies considerably throughout the season. For instance, Fageria et al. (2006, p.89, 93, 157, 180) argue that water deficiency and extreme temperatures during the mid-season flowering period of cereal and leguminous crops has greater implications for yield than any other period.

Another potential issue is that water availability for crop growth should arguably be more closely related to the *stock* of water in the soil (soil moisture) at any given point in time than to the *inflow*

of water to the ground over a long period (such as measured by pluri-monthly precipitation for rainfed crops). For instance, the soil is quickly saturated during intense rainfall and additional rain runs off and is no longer available to crops. Thus, the same amount of rainfall spread over time yields greater availability of water to crops because it allows rain to seep in the soil. Indeed, the fraction of rain that infiltrates the soil and becomes available for crop growth depends on how wet the soil is initially. Also, rain water evaporates more rapidly during hot, dry and windy conditions. Thus, a given rainfall event in the summer is not as effective in keeping the soil wet as in the cooler spring or fall. Precipitation also seeps to deeper soil layers out of root reach in more porous soil (e.g. sandy soil). As a consequence, factors such as recent rainfall, temperature, humidity, soil type, slope or crop stage affect the extent to which precipitation can be effectively available for crop growth. In summary, precipitation is only a part of the equation of water availability to crops whereas soil moisture itself is arguably a more appropriate metric.

Unpacking the relative contributions of heat and drought stress in climate change impact scenarios is a major priority for econometric analysis because it should improve understanding of potential impact and adaptation mechanisms. **?** emphasize that clarifying the structure of adaptation mechanisms facilitates the assessment of potential welfare impacts. Informed assessment of adaptation possibilities depends fundamentally on capturing the mechanisms that facilitate farmers' abilities to adapt to new climatic inputs and constraints. As a consequence, the timing of environmental conditions within the season may matter if farmers can choose to limit their exposure to adverse intra-seasonal conditions by shifting planting times, changing the crop mix, or making other counteracting production decisions.

In essence, the choice of climate variables is intimately related to the structure of the farmer's optimization problem. For instance, choosing fixed calendar periods for climate variables assumes a fixed growing season. **?** show that such a restriction overestimates corn yields damages by 30 to 70% under a 5°F warming scenario in the Upper Midwest. They rely on the fact that a warmer climate results in a longer non-freezing period that provides greater flexibility in the choice of planting date. Because yield sensitivity to high temperatures is stronger around the middle of the season (when corn is flowering), earlier planting by two to three weeks shifts the sensitive period away from the most detrimental summer heat. Thus, under-representation of adaptation possibilities leads to overestimation of impacts.

In this paper, I expand the horizons of literature by unpacking the effects of heat stress and drought stress, and identifying their interactions. I build on previous frontier work by SR on nonlinear effects of temperature and explore the nonlinear effects of soil moisture on yields at different

points during the growing season. The emphasis on timing has the ultimate purpose of improving representations of farmer adaptation possibilities given changes in relevant environmental conditions forecasted by accepted climate change models. This should facilitate econometric adaptation analysis based on revealed preference data that accounts for intra-seasonal changes of environmental conditions associated with climate change.

To develop my model of the role of soil moisture as well as other climate variables, I rely on a state-of-the-art soil moisture and weather dataset from the North American Land Data Assimilation System (NLDAS), which offers very high resolution in both space (14km) and time (1h). I replicate the SR panel model for corn yield and contrast it with a model that accounts for the timing and level of soil moisture using various flexible semi-parametric specifications. To demonstrate these issues clearly, I focus only on corn production in the Upper Midwest, which is the most productive area for high-valued field crops in the US.

Results suggest sizable nonlinear effects of soil moisture on crop yields that are particularly large toward the middle of the season around the flowering period. Results show that failure to account for soil moisture not only significantly reduces model fit, but leads to overestimation of the detrimental effects of heat stress by about 30%. This stems from the fact that soil moisture has so far been confounded with high temperatures leading to omitted variable bias. Because soil moisture and temperature change patterns differ within the season, this omission also leads to an overestimation of overall impacts by almost 100% by the end of the century under the medium warming scenario (RCP6). These results imply that water resources will play a major role in the overall impacts, in stark contrast to models based on calendar precipitation variables.

This paper is organized as follows. Section 2 describes data sources and how regressors and climate change scenarios were constructed. Section 3 presents estimates of the leading reference model using this refined dataset and contrasts these results with a model augmented with the soil moisture possibilities facilitated by this dataset. Section 4 projects climate change impacts for these models under various climate change scenarios and compares results in terms of overall impacts with focus on the relative roles played by heat and drought stress. Section 5 presents discussion about the implications of this model for climate change impact assessment and outlines an agenda for the future research it motivates. I conclude in section 6.

2 Data sources and variables

2.1 Soil moisture and weather Data

This paper seeks to improve understanding of drought stress for the purposes of climate change impact assessment by improving the representation of water availability to crops. Rather than using precipitation variables, I rely directly on measures of water content in the soil. While disaggregated weather data can be obtained with relative ease, this is not the case for soil moisture. Detailed soil moisture measurements are typically confined to experimental fields in some states. The feasible alternative for broad-based geographic models is to rely on the latest model-generated soil moisture estimates which serve as proxies.

The North American Land Data Assimilation System (NLDAS, Mitchell et al., 2004) is a joint project by the National Oceanic and Atmospheric Administration's National Centers for Environmental Prediction (NOAA/NCEP), the National Aeronautics and Space Administration (NASA), Princeton University, and the University of Washington. It offers state-of-the-art gridded weather and soil moisture datasets. The NLDAS uses weather station, satellite, radar and reanalysis data together with four different land surface models to generate estimates of soil moisture across North America.[1] These estimates account for parameters such as soil type, land cover, and slope with a 1km resolution. Specifically, the second stage of the NLDAS project, or NLDAS-2, provides model output data in the form of water mass for several soil layers as well as the model input weather data at an impressive level of detail. The large dataset contains hourly observations in near real time with a spatial resolution of 14km over North America since January 1979.[2]

The NLDAS project team is particularly attentive to the accuracy of its forcing weather data (precipitation and temperature, etc.) as well as of its model output (soil water content, etc.). Cosgrove et al. (2003) describe the techniques used to generate the hourly NLDAS weather data. They perform a cross-validation for the 1/1/1998 to 11/30/1999 period against observed data from U.S. Department of Energy's Atmospheric Radiation Measurement program Clouds and Radiation Testbed (ARM CART) sites. For instance, the cross-validation regression for hourly temperature exhibits an R^2 of 0.980 with a small bias of -0.479ºC. This is accomplished with an ARM CART site in the Southern Great Plains that covers hundreds of thousands of square kilometers and contains the world's largest collection of advanced remote sensing instruments, which is considered one of the best outdoor laboratories in the world. Its purpose is to serve as a gold standard to cross-validate the

---

[1]The soil model names are Noah, Mosaic, Sacramento Soil Moisture Accounting (SAC-SMA), and Variable Infiltration Capacity (VIC).

[2]The size of the NLDAS-2 weather and soil moisture hourly dataset for 1979-2011 for the North American domain in gridded format exceeds 2,000GB for only one of the four soil models.

output of global climate models. Needless to say, the measurements from this facility are superior to that of any standard weather station. Although this validation spans less than a year of observations, this is an impressive level of precision for cross-validation of hourly NLDAS weather data.

At the time of this writing, the cross-validation for NLDAS-2 model output (soil water storage) was under submission and unavailable.[3] However, Schaake et al. (2004) carried out a cross-validation for the first phase of the NLDAS project, NLDAS-1, which might provide a hint on how these soil models compare for NLDAS-2. They show that simulated water storage values from both the SAC and Noah soil models agree well with the measured values in several sites across Illinois, one of the major producing states in the sample of this paper. Their study also shows that the ranges of variability of SAC-SMA, Noah, and VIC water storage are close to the observed range. Expectations are that simulated water storage has been improved further in the NLDAS-2 data that I use in this paper.

The NLDAS dataset provides several advantages. First, it arguably offers the most reliable proxy of soil moisture across North America. Second, it offers spatial and temporal resolutions that allow a high level of detail in constructing county-level variables with the temporal detail necessary to match environmental conditions in critical parts of the growing season. Third, its hourly resolution eliminates the need to make assumptions about the temperature-time curve within a day (often assumed to follow a sine curve) which could provide more accurate measures of the distribution of temperature exposure.

The dataset could also present some shortcomings. First, it offers four different soil models. Although they yield qualitatively similar soil moisture contents, which one provides the best estimates is still unclear. However, the cross-validation for the NLDAS-1 project could provide a hint into which models perform better. Second, the NLDAS does not account for actual soil depth. The models apply over a fixed 2 meter soil column divided into 4 layers (0-10cm, 10-40cm, 60-100cm, 100-200cm). Locations with shallow soils have a lower water holding capacity and become saturated or dry out more quickly, which would interfere with correct estimation. However, my study region has some of the deepest soils in the US and this should not be a concern. Third, the NLDAS soil moisture estimates only account for water supplied through precipitation. As a result, they do not offer an accurate representation of soil moisture in irrigated areas. At best, they estimate the soil moisture deficit that is made up by irrigation in these locations.

For the above reasons, not withstanding the shortcomings, I rely on the NLDAS-2 dataset to extract hourly weather (precipitation and temperature) and water soil content for the upper soil

---

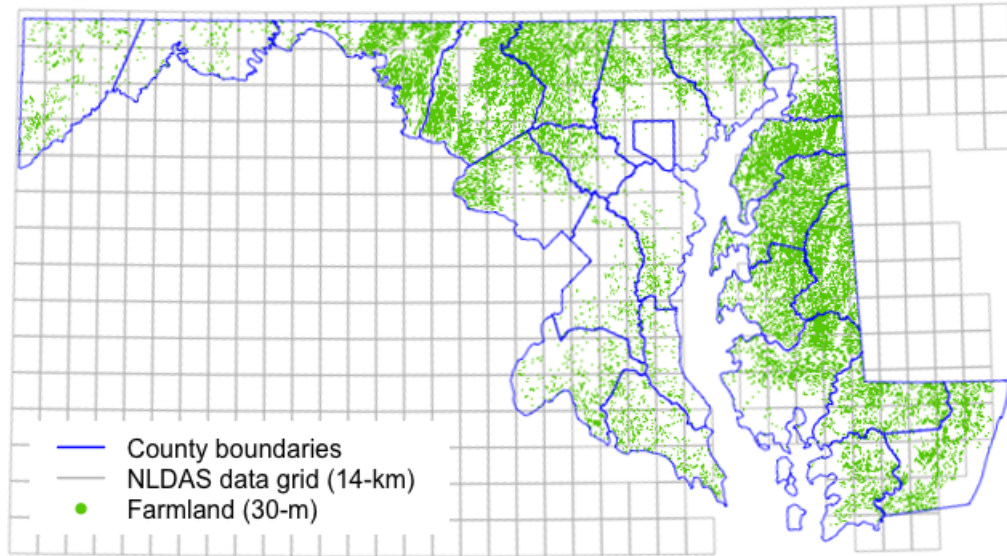[3]David Mocko (NASA), personal communication, November 21, 2012

Figure 1: The construction of county-level observations.

layer based on the Noah soil model.[4] To my knowledge, this is the first study to use the NLDAS dataset in this literature.

To construct county-level observations, I account for the amount of cropland within each NLDAS soil moisture and weather data grid. I proceed by overlaying USDA's 2011 Crop Data Layer (with 30m resolution) over the NLDAS data grid (with 14km resolution) to compute the total amount of cropland falling within each data grid. I then overlay the NLDAS grid over US county boundaries and compute the share of each grid falling within each county. I finally generate the hourly county-level observations by weighting each NLDAS data grid within a county by the amount of cropland it contains. Figure 1 offers a representation of crop cover, the NLDAS data grid, and county boundaries for the state of Maryland.

For illustrative purposes, figure 2 presents hourly soil moisture, precipitation, and temperature for in a midwest county. Panel A illustrates how soil moisture (shown in blue in the upper part of the graph) suddenly increases after a precipitation event (shown in green in the lower part of the graph) and then gradually decreases as the soil dries out. Panel B illustrates how soil moisture varies rather slowly over time (aside from the spikes at precipitation events) when compared to daily fluctuations in temperature (shown in red).

Panel A in figure 3 shows the temperature variation within each bin for the March-August time

---

[4]The moisture in this superficial layer (0-10cm) is highly correlated with moisture in deeper layers although the correlation weakens with depth and varies throughout the year. Because simulating climate change impacts consists in multiplying estimated parameters by the projected change in the associated variables, assessing the effect of deeper soil moisture changes would require climate change data on these layers. Unfortunately, data is only available for the superficial layer and, therefore, I cannot directly assess the contribution of deeper soil moisture changes.
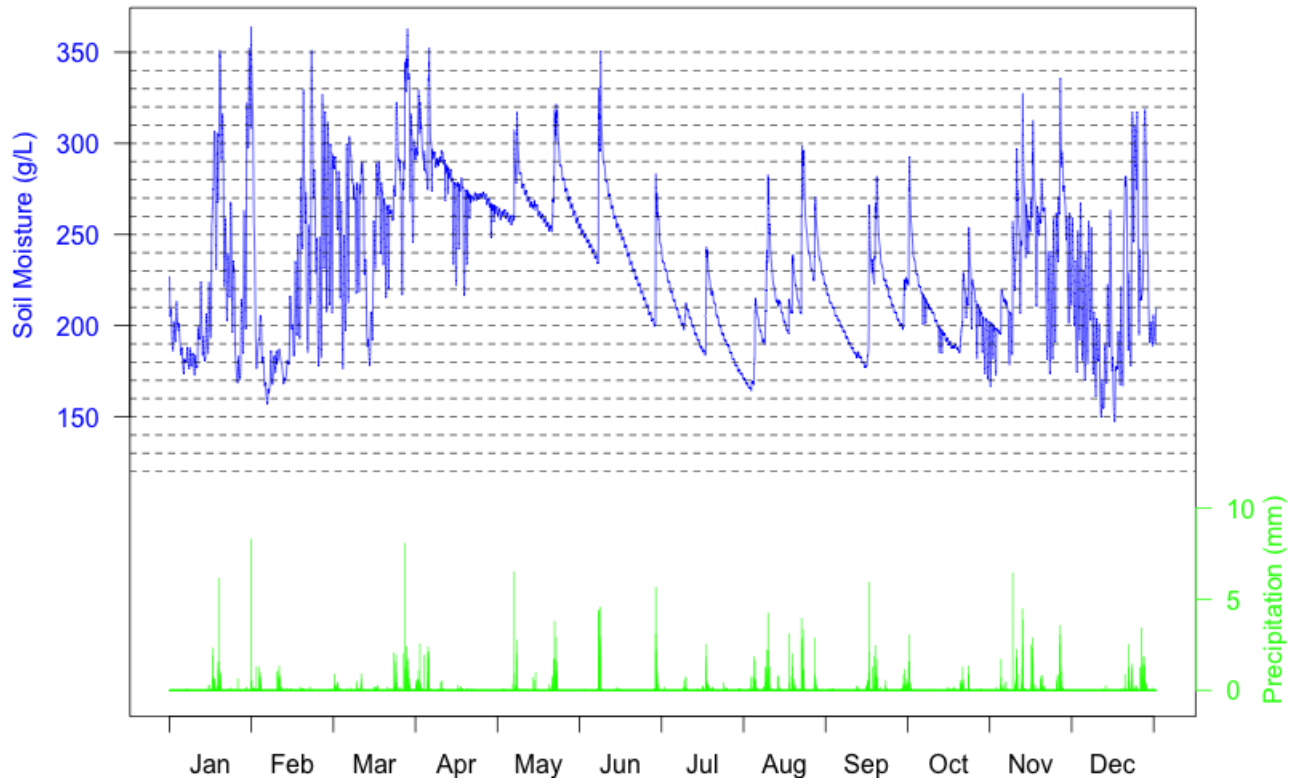
A. Precipitation and resulting soil moisture
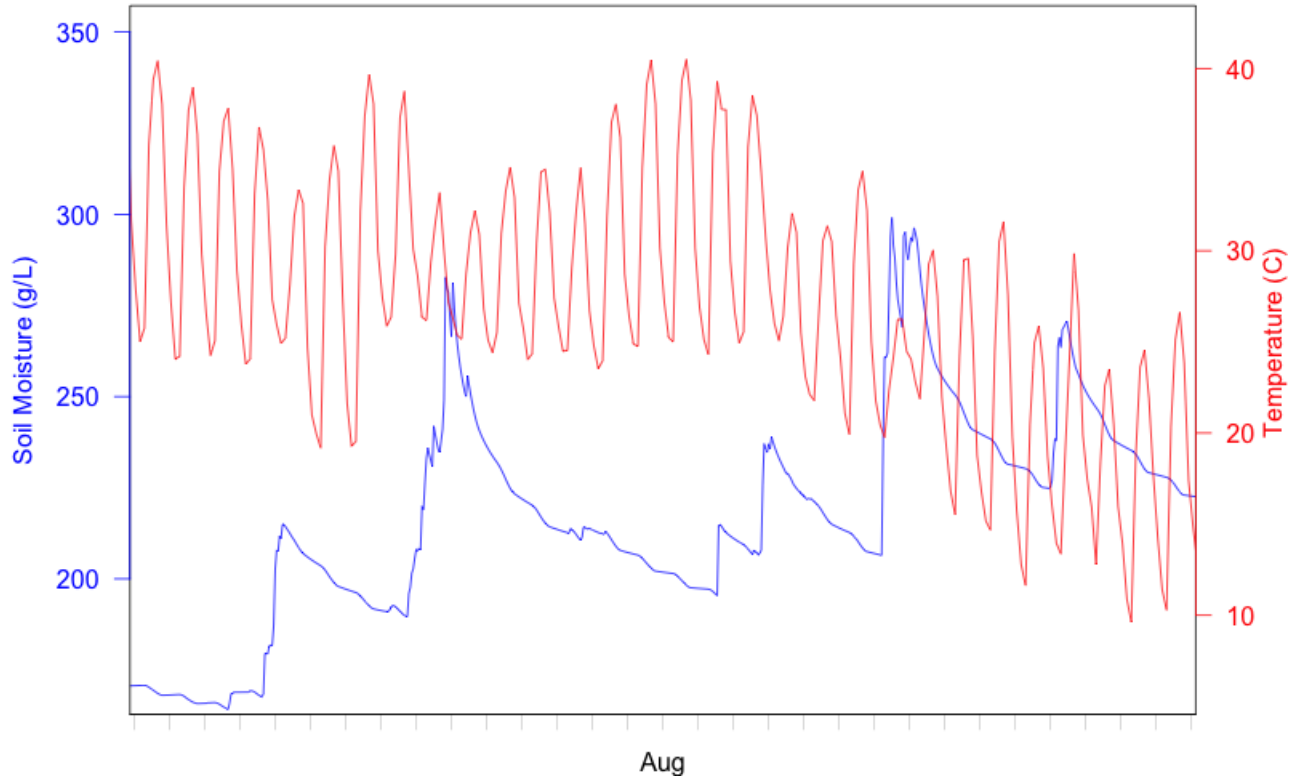
B. Temperature and soil moisture in August

Figure 2: Environmental variables for 1988 in Adams county, Illinois
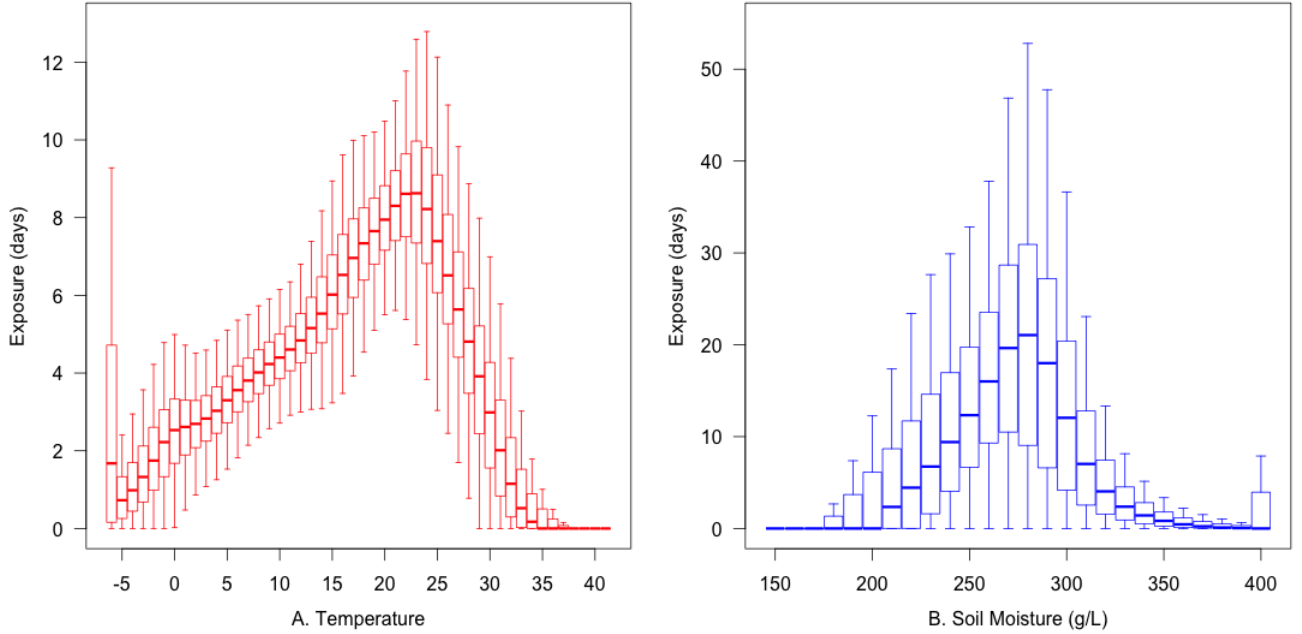
40

Figure 3: Temperature and soil moisture exposure distributions

window within the sample. For each temperature bin, the central line, box edges, and whiskers, represent the median, quartiles, and extremums, respectively. The most frequent temperatures fall between 20 and 25$^{\text{o}}$C. Temperature exposure under -5$^{\text{o}}$C was collapsed to the same bin and explains the tall bar and whiskers on the left. This is mainly driven by northern counties in the sample for which exposure to sub-freezing temperatures in march is not uncommon.

In a similar fashion, panel B of 3 illustrates the soil moisture variation within each bin for March-August. The most frequent soil moisture level is 280 grams of water per liter of soil (g/L). Soil moisture at or above 400g/L is collapsed to a single bin which explains the taller bar on the right. It is worth emphasizing that exposure to high levels of moisture, say above 350g/L, are often short-lived and typically correspond to exposure driven by moisture "spikes" after rainfall events (as illustrated in figure 2).

In order to assess the non-linear effects of soil moisture, I construct variables corresponding to the time spent within each 10g/L soil moisture interval in the 120-350 g/L range. These moisture bins are represented by the dashed lines in figure 2A. Because moisture outside this interval occurs, on average, less than 8 days in the March-August period, I aggregate exposure to these extreme levels to its closest moisture bin. In a similar fashion to SR, I also construct variables for the exposure to temperature bins used to account for heat stress. In particular, I collapse temperature exposure above 40$^{\text{o}}$C to the same bin.

The dataset developed in this paper compares to the dataset generated by SR, which is the most sophisticated weather dataset previously used for this type of analysis. They developed a daily weather dataset by interpolating daily but spatially sparse data from weather stations, with monthly but spatially detailed (4km) data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset developed by Oregon State University. According to their cross-validation, the spatio-temporal interpolation yields fairly accurate values for daily temperature but not for daily precipitation. Although this dataset has a longer time coverage (1950-2005), its obvious limitation for the purpose of this paper is the lack of soil moisture information.

As a way to verify the existence of a meaningful difference between the SR dataset and the data I derived from the NLDAS, labelled as "OB", I illustrate temperature exposure and precipitation densities from both datasets in figure 4. The figure shows data for the overlapping period across datasets (1979-2005) and for 800 counties in the rainfed sample of this study.

Panel A shows that the relative frequency of temperatures are somewhat different. The most common temperature range in the SR dataset is around 17-20ºC while it is 20-23ºC in the OB dataset. Also, in the OB dataset, the decrease in exposure around the most frequent temperatures is steeper toward higher temperatures (>23ºC) than toward lower ones (<20ºC). This is not exactly the case for the SR data. The graph on the right in panel A, illustrates the difference in exposure between both datasets and shows that the frequency of temperatures in the 20-27ºC range is lower in the SR data, but higher for lower and higher ranges. Particular attention should be given to the higher frequency of 28-35ºC temperatures in the SR dataset because observations in this range are used to estimate the effects of extreme temperature on yield. These differences are possibly due, in whole or in part, to the assumption of a daily sine curve in the temperature variation, or the spatio-temporal interpolation used to generate the temperature exposure data used by SR.

Finally, panel B shows that precipitation distributions are similar and differences for the large majority of cases do not exceed 50mm, or 2 inches, over the March-August period.

2.2   Accounting for timing of soil moisture conditions

My major contribution is to account for the nonlinear effects of soil moisture and timing in the growing season, which permits putting the role of temperature variation in context. This should facilitate more accurate econometric analysis of adaptation possibilities to climate change that accounts for changes in intra-seasonal environmental conditions.

Accounting for the timing effect requires information on crop stages. I thus rely on the Crop
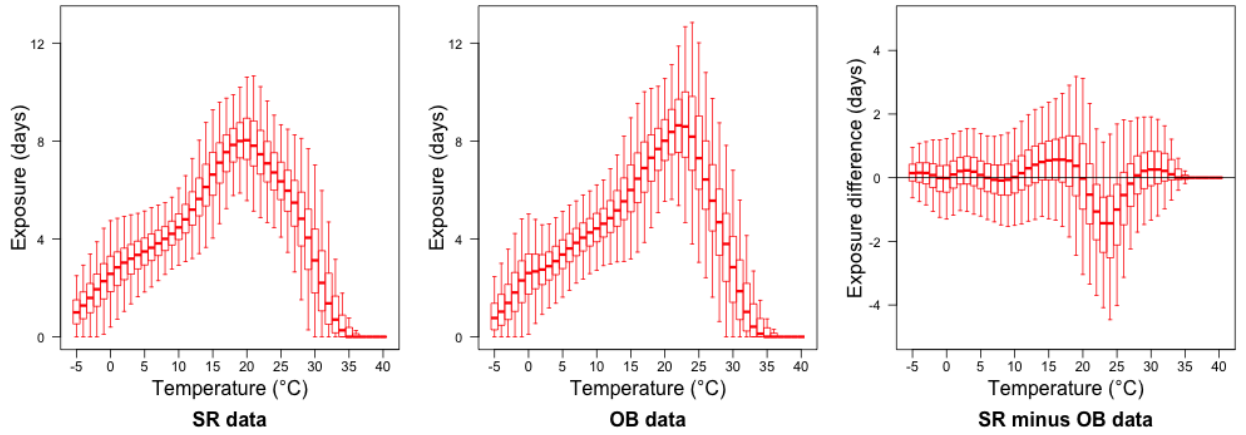
A. Temperature exposure for March-August
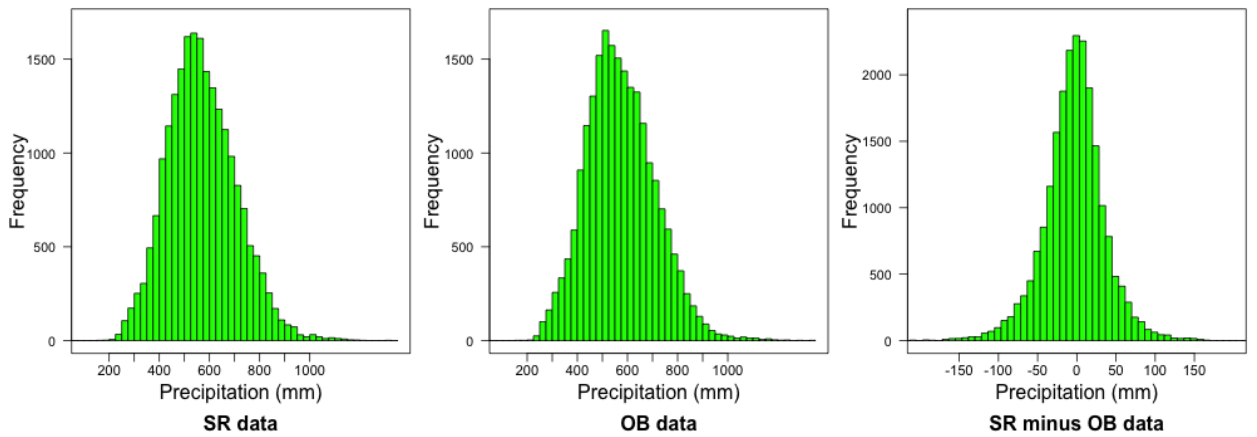
B. Precipitation distribution for March-August

Figure 4: Comparison of SR and OB datasets (sample counties, 1979-2005).

Progress and Conditions weekly survey by USDA/NASS which provides state-level data on farmer activities and crop phenological stages from early April to late November. Reporting across states and years is not balanced. Although state reports date back to 1979, reporting for corn that includes both the onset (planting/emergence) and the end of the season (maturation/harvesting) begin in 1981 for the major producing states.

Specifically, this survey reports the percentage of a state's corn acreage undergoing certain farming practices and reaching specific crop stages.[5] As a consequence, it does not offer clear "boundary" dates between stages because of the timing variations within states.[6] For the purpose of defining such boundaries of the growing season for each county, I obtain stage median acreage dates. These correspond to the dates at which 50% of the acreage in a given state has reached each stage in a given year.[7]

Crop stages reported by the USDA are not equally spaced in the growing season. They arguably correspond to visible markers that can be easily verified to simplify data collection. Some past studies (e.g.Kaufmann and Snell, 1997) have relied on weather variables matched to precise crop stages. However, results are sometimes difficult to interpret, especially for non-agronomists. In order to convey a more accessible crop advancement metric, I divide the growing season into eight segments centered around flowering (i.e. silking), which is considered the midpoint of the season. Four equally-spaced periods occur in the vegetative phase (between planting and silking) and four equally-spaced periods occur in the reproductive or grain-filling phase (between silking and maturation). For simplification, the crop advancement division is converted into percentages with intervals of 12.5%. Thus, the 0-12.5% and 87.5-100% stages correspond, respectively, to the first and last segments just after planting and just before maturation, and 37.5-50% and 50-62.5% correspond, respectively, to the segments just before and just after flowering.

Natural scientists have found that crop development or phenology is proportional to accumulated Growing Degree-Days (GDD, see e.g. Hodges, 1991; Smith and Hamel, 1999; Fageria et al., 2006; Hudson and Keatley, 2009). This variable is defined by the area under the temperature-time curve

---

[5]The report includes progress of farming activities (planting and harvesting) and of corn phenological stages (emerged, silking, doughing, dented and mature). The USDA defines these crop stages as follows. Emerged: as soon as the plants are visible. Silking: the emergence of silk-like strands from the end of corn ears, which occurs approximately 10 days after the tassel first begins to emerge from the sheath or 2-4 days after the tassel has emerged. Doughing: normally half of the kernels are showing dent with some thick or dough-like substance in all kernels. Dented: occurs when all kernels are fully dented, and the ear is firm and solid, and there is no milk present in most kernels. Mature: plant is considered safe from frost and corn is about ready to harvest with shucks opening, and there is no green foliage present.

[6]Visual inspection of district-level crop progress reports, which are available for only a few states, surprisingly reveals variation similar to overall state progress for most years.

[7]For a few states and years, crop progress reporting began too late (the state had already surpassed the 50% acreage level) or stopped too early (the state had not yet reached the 50% acreage level). For these cases, which represent less than 5% of the cases, I obtained the median acreage date by extrapolation. More details are provided in the appendix.
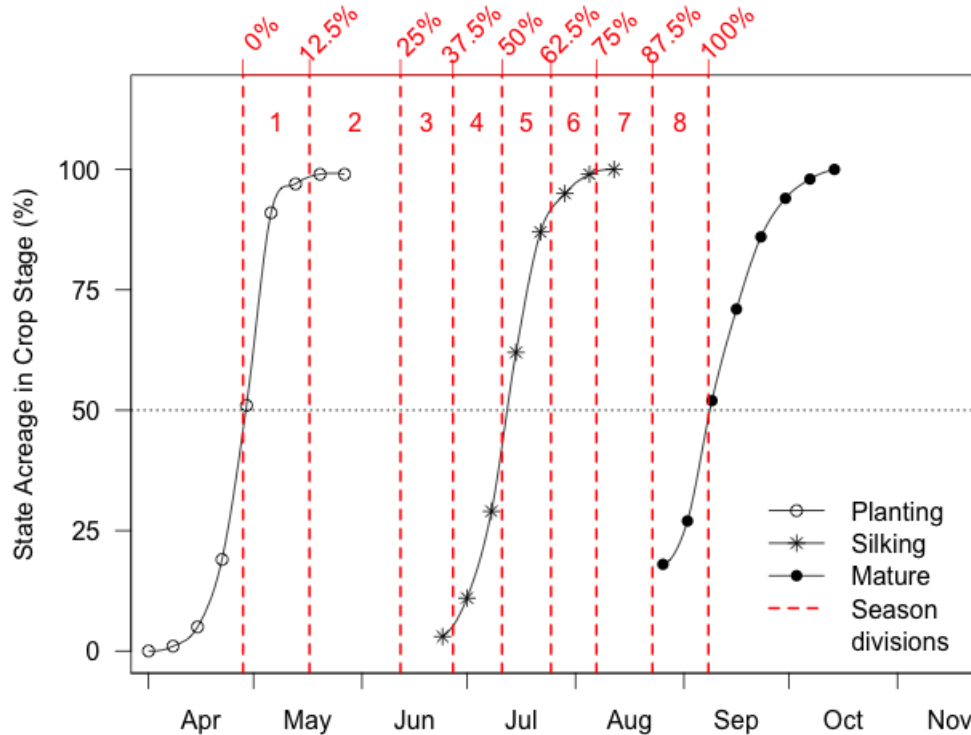
Figure 5: Season divisions for Illinois corn in 2001.

that falls between two temperature thresholds (10 and 30ºC for corn) and two time periods. Warmer conditions generally lead to faster GDD accumulation and more rapid crop development. This concept can be used to split the growing season into equally-spaced segments.

Following this approach, I compute a cumulative GDD variable starting at planting for each state and year and use it to represent the eight segments of the season. Figure 5 illustrates how these season segments are located in the 2001 calendar for Illinois. Although the segments have a different number of days, segments 1-4 and 5-8 are equally spaced in terms of GDD. Thus, wider segments signal slower development due to cooler conditions.

Exposure to moisture bins is aggregated within each one of these segments. As a result, the moisture variables account for exposure to different moisture levels during each one of the eight segments of the growing season. This allows assessment of how drought sensitivity varies with crop advancement.

2.3   Agricultural data and sample counties

Agricultural data were obtained from publicly available USDA/NASS sources and include county-level corn yield and acreage. Yield is the dependent variable in estimation models and acreage is
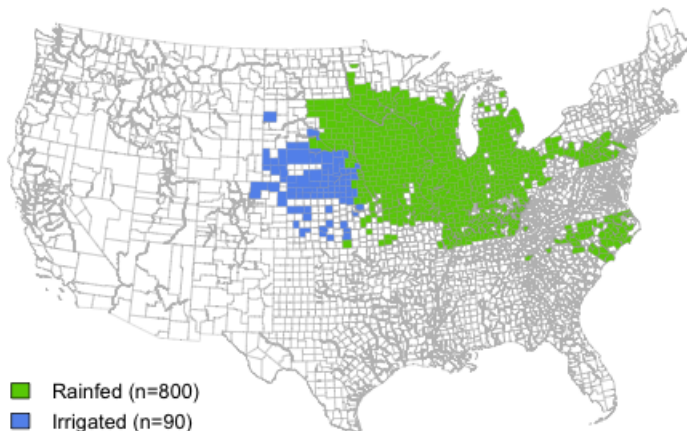
Figure 6: Rainfed and irrigated counties in the sample

used to weight county-level climate change impacts to obtain aggregate estimates for the sample. Because rainfed and irrigated corn yields are expected to respond differently to exogenous environmental conditions, their respective parameters must be estimated separately. For this purpose, I split the sample into rainfed and irrigated counties where a county is considered rainfed if at least 75% of its acreage, on average, is non-irrigated. Figure 6 illustrates how the sample is divided. The dataset corresponds to a balanced panel of 800 rainfed and 90 irrigated counties for 1981-2011. Although this paper focuses on rainfed counties located in 14 different states, results for irrigated counties are reported in the appendix for illustrative and falsification purposes.

2.4 Climate Change Data and Scenarios

Climate change data were obtained from the second version of the Hadley Centre Global Environment Model (HadGEM2). The HadGEM2 is one of the latest and most advanced climate models. It has a higher spatial resolution and improved representation of the atmosphere compared to the earlier HadCM3 model which is commonly used in the literature (Collins et al., 2008). The HadGEM2 model is also being used in the preparation for the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5), scheduled for publication in late 2013.

In the upcoming AR5 report, the nature of climate change scenarios has been modified. They no longer represent "emission scenarios" but are "representative concentration pathways" (RCPs). Instead of describing economic scenarios and their resulting emissions (e.g., the familiar A1B, A1, B1 scenarios), they now represent sets of a wide range of projections for the main drivers of climate change, which are greenhouse gases, air pollutants and land use change. These scenarios are classified in terms of their "radiative forcing", which roughly represents the strength of different human and

natural agents in causing climate change (See IPCC 2007, p.131 for a detailed definition). The convention is to associate the radiation level by 2100 to the scenario name. For instance, the most severe RCP8.5 scenario represents a rising radiative forcing pathway leading to 8.5 watts/m^2 in 2100.[8] The higher the radiative forcing, and the greater is the resulting warming.

Because the crop stage time windows do not correspond to calendar periods, I cannot rely on widely used monthly data. Instead, I obtain daily data corresponding to the RCP2.6, RCP6, and RCP8.5 scenarios for average temperature, precipitation, and soil moisture for the superficial soil layer (0-10cm) for periods 1985-2005, 2039-2059, and 2079-2099. The first period serves as a reference period for current climate. The others represent the mid-century and end-of-century climates.

The variable changes for each grid are obtained by subtracting the mid-century and end-of-century means from the current climate reference period. These changes are then matched to the corresponding counties. However, regression models are based on nonlinear transformation of these variables and, thus, the original level of the variable for the reference period 1985-2005 matters. Accordingly, I add the change in the untransformed variables to the NLDAS variables before performing nonlinear transformations. As explained in Fisher et al. (2012), this approach maintains the spatial smoothness of projected climate changes.

Figure 7 presents projected changes in temperature, precipitation, and soil moisture for the three scenarios for the mid-century and end-of-century periods. Panel A shows that the frequency of temperatures below 20-25°C will almost uniformly decrease while the frequency increases would be clustered around 30-35°C. This is a manifestation of the nonlinear changes in exposure to high temperature from an increase in temperature.

Panel B shows that precipitation changes are mixed, although most counties will see their March-August precipitation decrease in most scenarios. With a mean precipitation around 550mm (see figure 4B), mean precipitation reductions hover around 0-7% except for the most severe scenario, which has mean precipitation reductions in the 10-25% range.

Panel C illustrates how soil moisture is expected to vary for each of the eight seasonal segments (using current average segment windows). The lower (upper) part of each graph represents early (late) segments of the season. The general pattern is that more humid soils will be more frequent at the beginning of the season while decreases in their frequency occur towards the latter stages. This is represented by blue (red) areas located toward the bottom right (left) corner, and red (blue) areas located toward the upper right (left) corner. Only the more severe RCP8.5 scenario does not follow this pattern with almost universal decreases in the frequency of humid soils. This is represented by

---

[8]A watt is the standard unit of power, which is a transfer of energy per unit of time.

blue (red) areas toward the left (right) side of the graph.

A interesting pattern arises in panel C that is highly meaningful for econometric adaptation analysis. A moisture "inversion" occurs during the season. The early season becomes more humid, while the end of the season becomes drier. This suggests that farmers may be able to adapt to this intra-seasonal change by altering planting dates to limit their exposure to detrimental parts of the season. This pattern is not perceptible in the March-August precipitation changes that solely suggest modest season-long decreases.

3 Models for heat and drought stress

3.1 Replication of a reference model for heat stress

Statistical models that have regressed crop yields on weather variables have traditionally relied on monthly or pluri-monthly average temperature and precipitation data. Early examples can be traced back to the early part of the last century (Wallace, 1920; Hodges, 1931). Since then, the convention has long been to include linear and quadratic variables based on temporally aggregated data to capture the nonlinear effects of both temperature and precipitation on yield. Marginal effects of these variables are typically expected to exhibit an "inverted U" shape, suggesting diminishing marginal effects of each weather variable with a unique optimum.

Schlenker et al. (2006) made an important contribution by recognizing that daily average temperature fails to convey the consequences of exposure to extreme temperature and, thus, may not be adequate for capturing nonlinear effects of temperature on farm prices. Hypothetically, two days with equal average temperature may represent very different exposures to very high temperatures. This suggests that the shape of the daily time curve matters.

To address this needed refinement, SR developed an innovative approach that estimates the effect of exposure to different levels of temperature on yield separately. They compute the amount of time spent during the season (March-August for corn) in each of many temperature bins. The exposure to each degree bin is then adapted to various specifications.

Here, I replicate their model for corn for purposes of comparison. As stated in the data section, I restrict the sample period to 1981-2011. This is shorter and later than the 1950-2005 period used by SR. However, their results are reported to be similar for temporal subsets of the sample. The balanced panel dataset in this paper represents over 70% of US corn production annually.

Their general model assumes that temperature effects on yield are cumulative and substitutable over time. The nonlinear effect of temperature on yield are captured by the function $g(h)$ representing "yield growth" that depends on temperature h. Logged corn yield y_{it} in county i and year t are
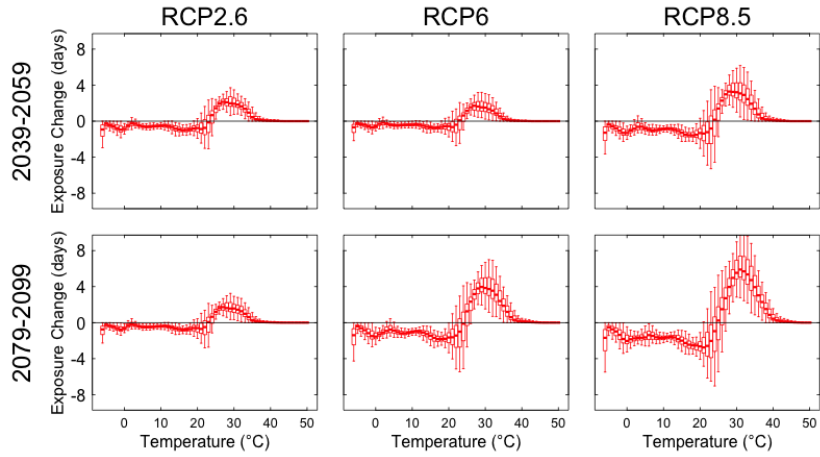
A. Temperature exposure change for March-August
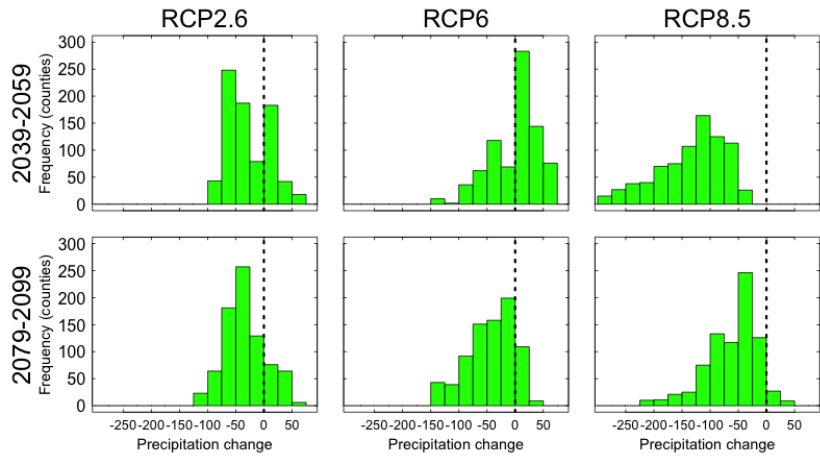
B. Precipitation change for March-August
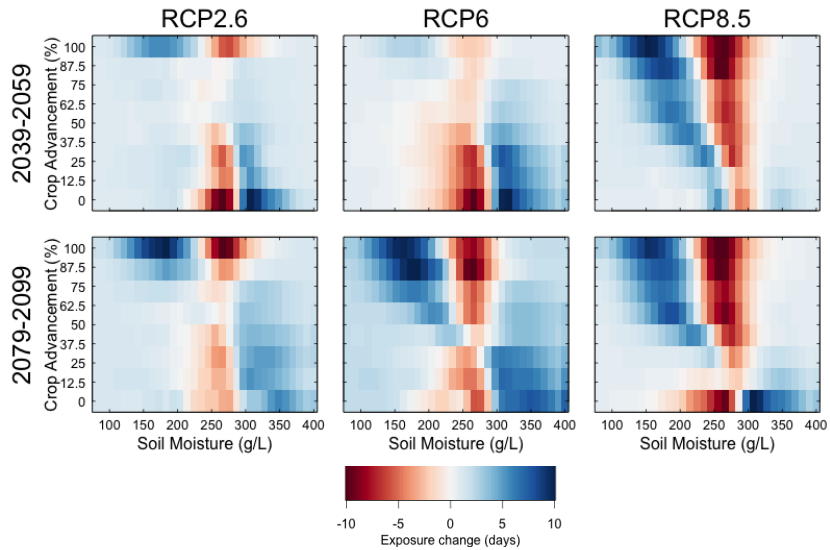
C. Soil moisture change for each season stage

Figure 7: Changes in environmental variables with climate change scenarios

represented as:

$$y_{it} = \int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}(h)d(h) + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i + \epsilon_{it} \qquad (1)$$

where $\phi_{it}(h)$ is the time distribution of temperature (i.e., the temperature-time path) for March-August, p_{it} is precipitation, z_{it} is a state-specific quadratic time trend and the c_i are county fixed-effects. The maximum likelihood estimation procedure accounts for spatial correlation of the errors. Over thirty different spatial weight matrices were evaluated by comparing models that only differ by the weight matrix. The weight matrix based on the inverse distance of the seven nearest neighboring counties yielded the highest value of the likelihood function at the optimum parameter values and thus was selected.[9]

Equation (1) cannot be estimated directly because of the integral. Therefore, I follow SR and consider different specifications to approximate the integral as a sum: a step function allowing different effects at each 1⁰C interval (SR1), another step function allowing different effects at 3⁰C intervals (SR2), an eighth-degree polynomial (SR3), and a cubic B-spline with eight degrees of freedom (SR4).[10] The specification for SR1 is:

$$y_{it} = \sum_{h=0}^{40} g(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)] + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i + \epsilon_{it}$$

where $\Phi_{it}(h)$ is the cumulative distribution of temperature in county i and year t. Specifications for SR2, SR3, and SR4 and more detailed results for each specification are provided in the appendix.

Results are summarized in figure 8. The effects of exposure to various levels of temperature vary considerably. Exposure to temperature in the 12-30⁰C range are beneficial while exposure is increasingly detrimental above 30⁰C. These results are qualitatively similar to what SR report. However, replication suggests that extreme temperature is considerably less damaging. While SR report that exchanging a single day at 29⁰C with a day at 40⁰C reduces yield by approximately 7%, none of the specifications in this replication suggest a yield reduction exceeding 3%.

To verify this discrepancy, I compare all specifications (SR1-SR4) applied to the OB and SR datasets for the overlapping 1979-2005 period. Results are shown in figure 9. Surprisingly, estimates of the same model used in the original SR study in panel B, show twice the sensitivity to high

---

[9]The weighting matrices included eight neighboring structures and four weighting schemes. The neighboring structures are: 5 through 10 nearest neighbors, neighbors within 200km, and neighbors using the Delaunay triangulation. The weighting schemes are: binary, inverse distance, inverse squared distance, and inverse square root of distance.

[10]Only SR2 and SR3 are part of the original SR study. In addition, SR developed a piecewise linear model which yields similar results to the other specifications. The SR1 was included to assess the effects of narrow temperature bins and SR4 to allow for a more flexible less susceptible to extreme polynomial curvature near the end points specification.
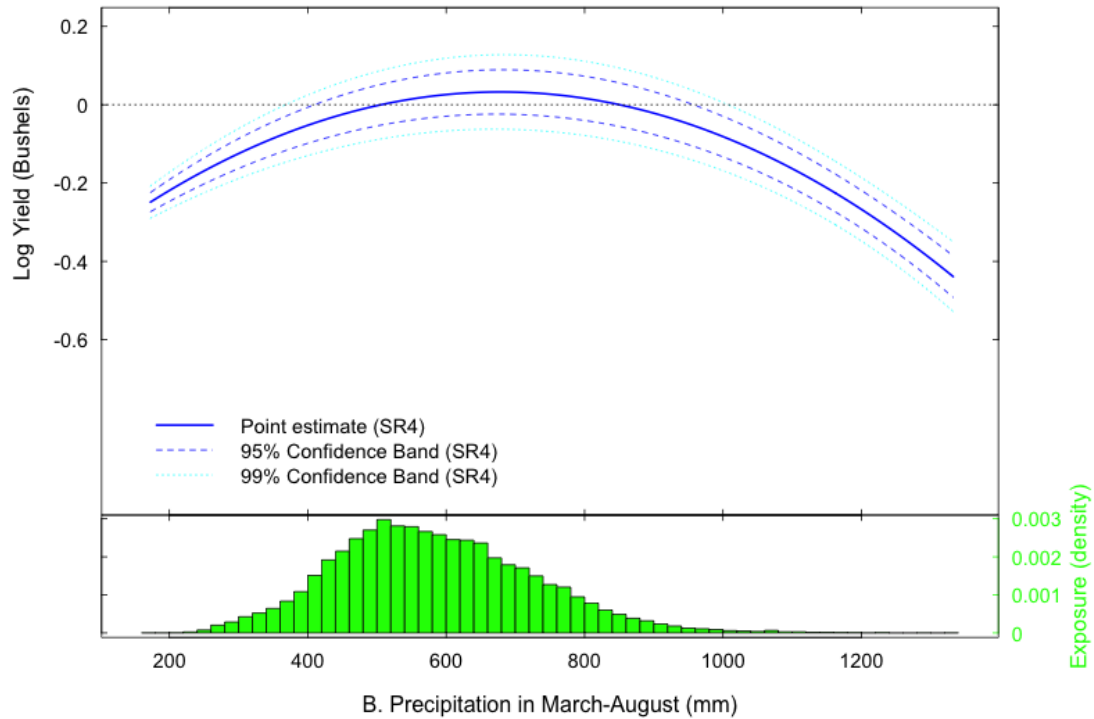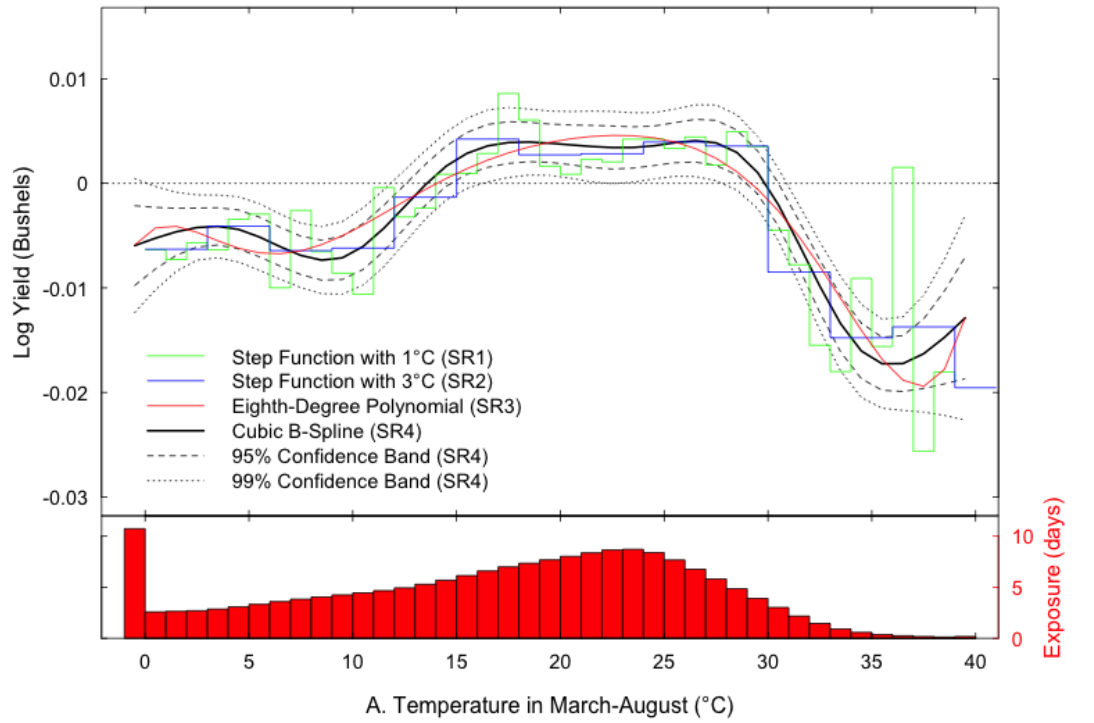
Response curves are centered around zero and weighted by temperature bin exposure or precipitation density. As a result areas above zero correspond to the most beneficial half of occurrences. Confidence bands for the temperature curve correspond to SR4.
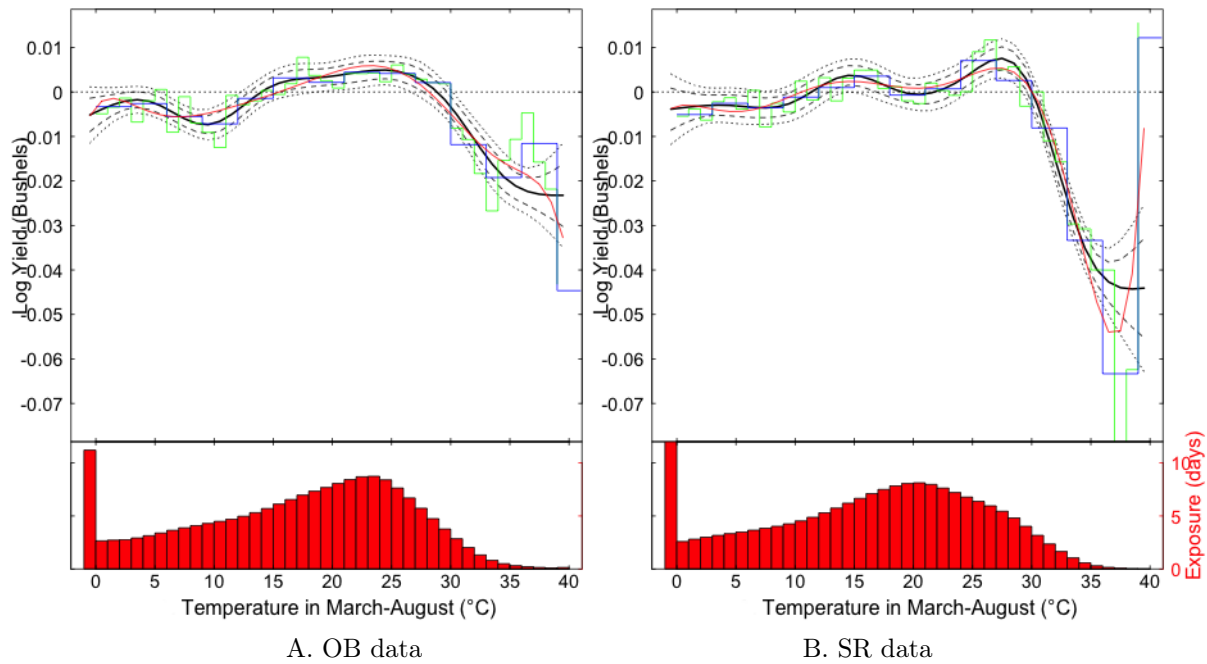
Figure 8: The SR model

51

Figure 9: Comparison of the spline specification using OB and SR data (1979-2005)

temperature when based on the SR data as when based on the OB data as shown in panel A. This is particularly striking given the seemingly small differences between the temperature distributions shown in figure 4. Figure 4 reveals the datasets exhibit relatively small differences for most temperature bins. However, these differences can exceed be relatively large for the very high temperatures. The average exposure in the March-August period to temperature above 35ºC is 14.4 hours and 22.3 hours in the SR and OB datasets, respectively. These 7.9 hours represent a 55% difference. The lower exposure to very high temperature recorded in the SR dataset is consistent with extreme temperature appearing more damaging.

In an attempt to discriminate between the OB and SR datasets, I performed a J-test between models based on these datasets. However, the test is inconclusive with t-statistics for the fitted values of the alternative model in excess of 10. Although the test is not conclusive, the implicit damaging effects of extreme temperatures are highly sensitive to nature of the weather data, particularly to small absolute differences in recorded exposure to very high temperature.[11]

On the other hand, this replication suggests an optimal level of precipitation of 678 mm for March-August, higher than the sample mean of 584 mm. Reaching the optimal precipitation level through a 15% increase, implies an insignificant yield gain of just 1%. Similarly, a dramatic 50% drop

---

[11]In analysis not shown in the paper, I swapped the exposure to extreme temperature (>35ºC) across SR and OB datasets and re-ran the models with the hybrid datasets. This resulted in an exchange of the shape of the temperature response curve at these extreme temperature levels. This confirms that the difference in the slope of the temperature response function for very high temperature mainly stems from the difference in recorded exposure to temperature exceeding 35ºC between both datasets.

in precipitation only represents a 15% yield reduction. Given that most climate change scenarios predict mean decreases ranging from 0 to 10% (see figure 7), these precipitation changes are expected to generate small to modest changes in yield according to this model. This is consistent with the small role attributed to precipitation in SR and other studies such as Schlenker et al. (2005), Schlenker et al. (2006) and Deschênes and Greenstone (2007). These results are at odds with agronomic evidence that emphasizes the pivotal role of water in crop production (NeSmith and Ritchie, 1992; Blum, 1996; Barnabás et al., 2008).[12]

3.2   A model accounting for soil moisture

The models of Section 3.1 that mirrors prior methodology attempt to capture water availability to crops with a season-long precipitation variable. My hypothesis is that this is an inappropriate measure of water availability for crops because it does not represent soil moisture conditions and their timing of these conditions in the growing season.

To address this potential shortcoming, I develop a model that assumes that crop yield also responds to soil moisture in possibly nonlinear and varying magnitudes throughout the season. Effectively, my model pools the SR model and the new soil moisture variables I introduce. The new model, which I label "OB", assumes that the effects of soil moisture m on yield are cumulative but non- substitutable over time in the season. The nonlinear effects of soil moisture on yield are captured by the function $f(m,s)$ representing the dependence of yield growth on soil moisture m at each stage of the season s. Logged corn yield y_{it} in county i and year t are represented as:

$$y_{it} = \underbrace{\int_{\underline{h}}^{\overline{h}} g(h)\phi_{it}(h)d(h) + p_{it}\delta_1 + p_{it}^2\delta_2 + +z_{it}\tau + c_i}_{\text{SR model}} + \underbrace{\int_{\underline{s}}^{\overline{s}} \int_{\underline{m}}^{\overline{m}} f(m,s)\psi_{it}(m,s)d(m)d(s)}_{\text{Moisture effects}} + \epsilon_{it} \quad (2)$$

where $\psi_{it}(m,s)$ is the distribution of soil moisture (i.e., the soil moisture-time path illustrated in figure 2) at each stage of the season s.

As in the SR model, equation (2) cannot be estimated directly. The objective is to approximate the double integral on $f(m,s)$ as a double sum. The first sum is over different moisture levels. I consider the same four approximation specifications I used to estimate the SR model: a step function allowing different effects for each 10g/L soil moisture interval (OB1), a step function allowing different effects at 30g/L intervals (OB2), an eighth-degree polynomial (OB3) and a cubic B-spline

---

[12]In the appendix I also present results for irrigated counties. Although results are not as clear, temperatures above 30ºC appear as detrimental as for rainfed counties. This is in contrast to the findings in SR that show in their appendix that temperatures above 30ºC are more than twice as damaging for eastern and mostly rainfed counties. The main discrepancy between my results and theirs concerns rainfed counties.

with eight degrees of freedom (OB4).

The second sum is over different season stages. For this purpose the season is split into eight segments as described in the data section so that $\underline{s}$ and $\bar{s}$ correspond, respectively, to planting and maturation.[13]

Note each SR specification is nested in the corresponding OB specification such that SR1 is nested in OB1, SR2 in OB2, etc. The specification for OB1, for example, is:

$$
\begin{aligned}
y_{it} &= \sum_{h=0}^{40} g(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)] + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i \\
&+ \sum_{s=1}^{8} \sum_{m=120}^{350} f(m+5,s)[\Psi_{it}(m+10,s) - \Psi_{it}(m,s)] + \epsilon_{it}
\end{aligned}
$$

where $\Psi_{it}(m,s)$ is the cumulative distribution of moisture for the s-th season segment in county i and year t. Specifications for OB2, OB3, and OB4 and more detailed results for each specification are provided in the appendix.

Results are summarized in figures 10 and 11. The 3-dimensional graph in panel A of figure 10 corresponds to the soil moisture effects on yield based on the cubic B-spline specification (OB4). It shows that yield effects vary considerably over the season.[14] Early in the season (at low crop progress) the yield response function is fairly flat, suggesting that deficient levels of moisture at this stage do not affect yield very much. In fact, high levels of moisture (>300g/L) at this stage are slightly detrimental, which is consistent with well-known damages from water-logging to young plants.

As the season advances, soil moisture levels around 265g/L imply crop yields on the trend, but lower or higher levels of moisture lead, respectively, to low and high yields. Yield damages are the most severe, as expected, right around the middle of the season when corn flowering occurs. Replacing a single day at 265 g/L with a day at 125g/L represents approximately a 1% yield decrease.

Although this result is 3 times lower than for the hypothetical exchange of a full day at 29°C with a day at 40°C (see the previous section), comparisons should be considered with care. Figure 2 illustrates that soil moisture deviations are much more persistent than temperature deviations, suggesting that the potential exposure to detrimental levels of moisture are likely to last days or

---

[13]Soil moisture conditions after maturation do not have an impact on yield although they might affect other quality characteristics such as kernel humidity.

[14]The tessellation is obtained by joining the stage-specific soil moisture yield responses (presented individually with confidence bands in the appendix) at regular intervals of soil moisture. A look at the individual stage-specific yield responses and their confidence bands (in the appendix) shows that this pattern is statistically significant.
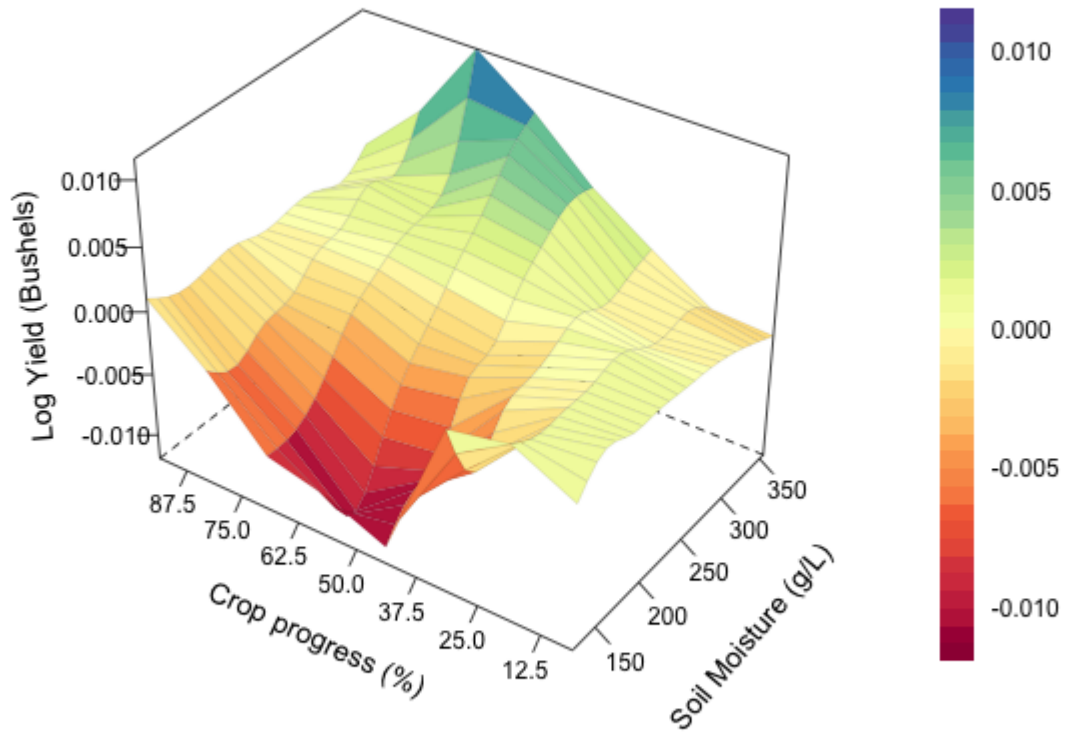
A. Soil moisture effects at different season stages
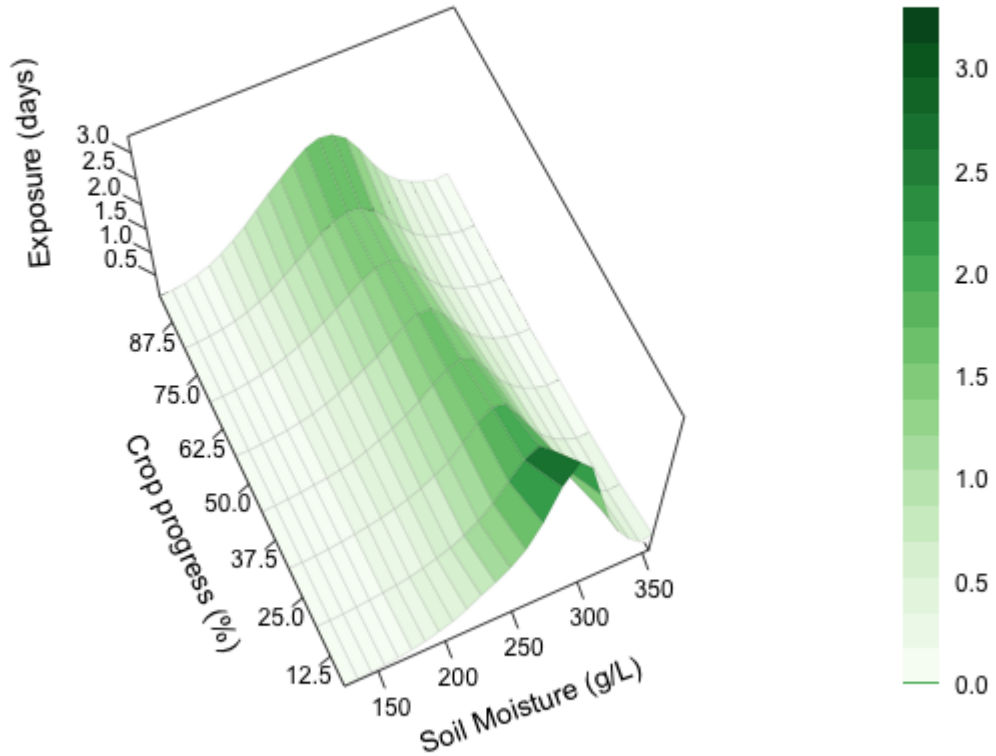
B. Distribution of soil moisture at different season stages

Figure 10: Soil moisture effects for the OB model

even weeks. On the other hand, the daily fluctuations of temperature require several days to build up to an extreme day or two of exposure to high temperatures ($>30^{\circ}$C).

Higher than normal levels of moisture, on the other hand, seem beneficial to yield. This is particularly the case in the second half of the season. Replacing a full day at 265 g/L with a day at 355g/L causes a yield increase in the range of $0.4 - 1\%$. This is consistent with the high water demand during flowering and grain-filling stages in corn.

At the end of the season the yield response flattens. Variations in soil moisture still make a difference but not as much as in the middle of the season. Because the statistical model and the climate change impact scenarios only consider the superficial 10cm soil layer, these results may overlook the fact that adult plants extract water from deeper soil layers late in the season.

A somewhat puzzling result is that very high soil moisture is virtually always found to be beneficial except for very early stages in the season. Extreme events such as flooding are undoubtedly detrimental, but these are not captured by the soil moisture variables. This is likely due to the division of the growing season into relatively short segments that do not account for cumulative exposure to very high levels of moisture spanning several segments. Perhaps these extreme events are captured by a season-long precipitation variable, which exhibits a significant role only for very high levels of precipitation as shown in the bottom of figure 11.

Interestingly, the precipitation response curve in figure 11 is similar to that of irrigated counties (see the appendix), which suggests that, after accounting for moisture, season-long precipitation captures only extreme events such as flooding in both rainfed and irrigated areas. This provides additional evidence that season-long precipitation is a rather poor measure of water supply for rainfed crops because it fails to account for the timing of soil moisture levels throughout the season.

A crucial finding is that the temperature response in the OB model, on the top of figure 11, is flatter for high temperatures than in the SR model. I superimpose the temperature response in both models in figure 12. In particular, high temperatures appear to be about 30% less detrimental to yield when soil moisture is considered. This difference is consistent with omitted variable bias and can occur if low soil moisture is both correlated with high temperature and is a good predictor of yield.

The correlation between dry soil and maximum daily temperature does not come as a surprise because this phenomena is well understood and documented in the climate science literature.[15]

---

[15]The reason is that water in the soil plays a crucial role in the partition of energy transfers between "latent heat" and "sensible heat." When the soil is wet, solar energy is spent evaporating this water without generating a temperature change (latent heat). However, when the soil is dry, no water is available to evaporate so solar energy is directly spent heating up the surroundings (sensible heat).
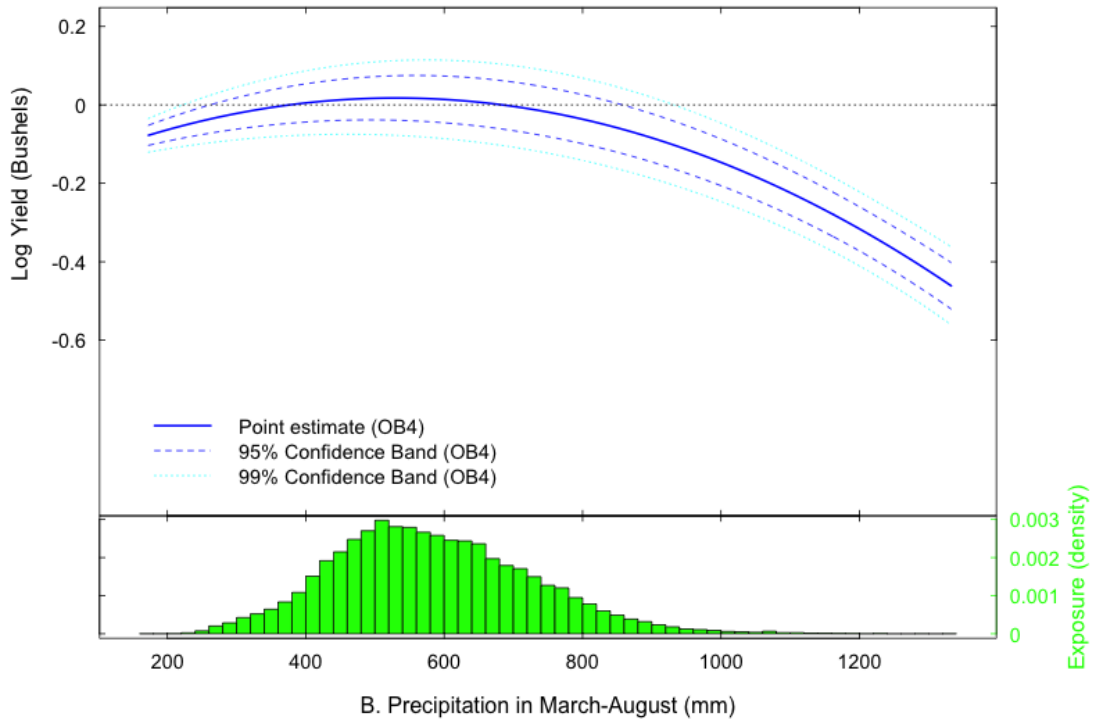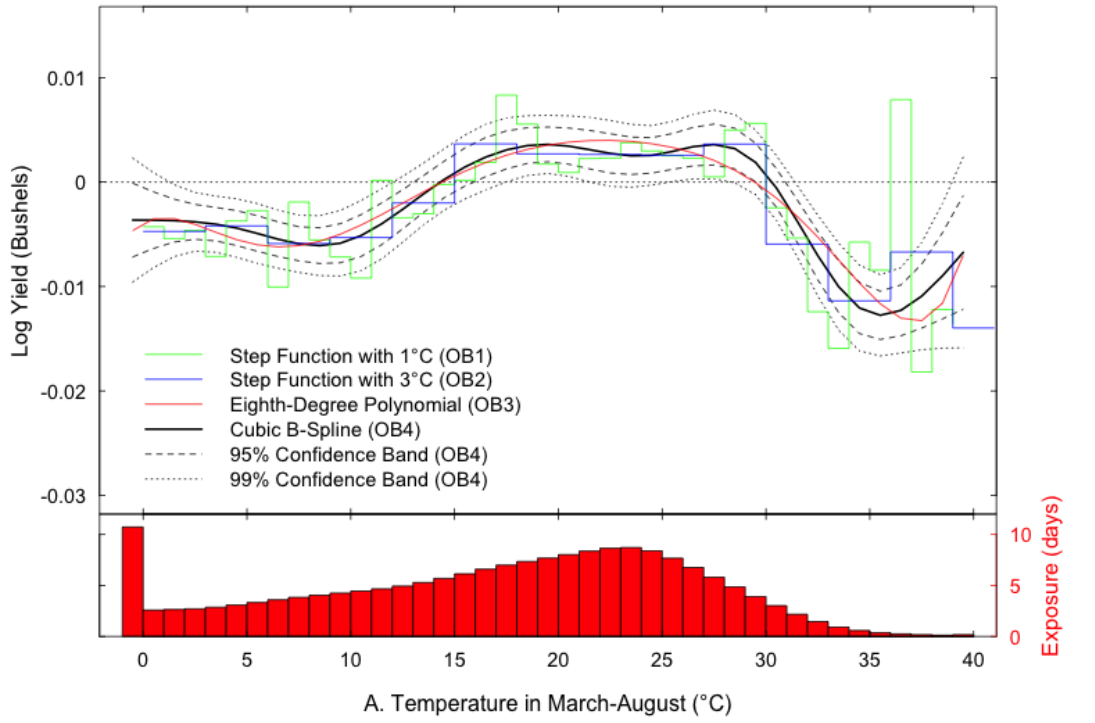
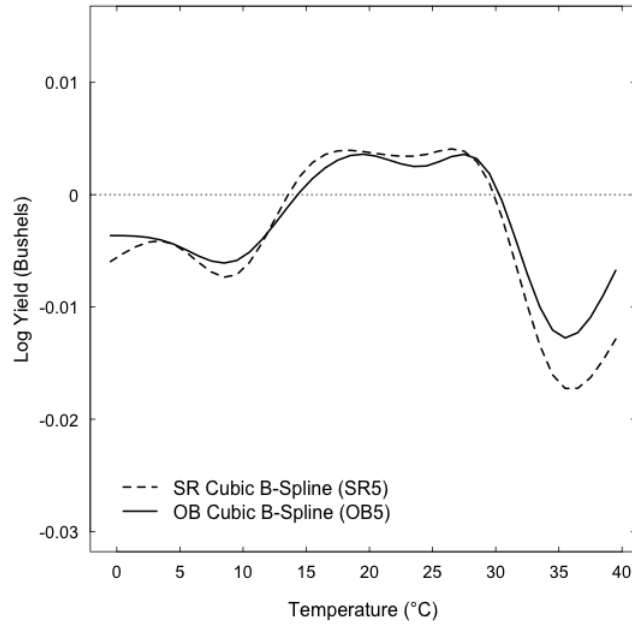Figure 11: Temperature and precipitation effects for the OB model

Figure 12: Comparing temperature effects between the SR and OB models.

In a recent paper, Mueller and Seneviratne (2012) show evidence on a global scale that dry soil is correlated with high temperatures, particularly during the hot months of the year. This phenomenon is also acknowledged by climate scientists and weather forecasters in their models. Global Climate Models (GCMs), such as the HadGEM2 used in this paper, include soil moisture modules precisely to account for the role of soil water in atmospheric energy balances. As an illustration, the original motivation for developing soil moisture estimates by the NLDAS was to improve weather forecasts: *"specifically, this system is intended to reduce the errors in the stores of soil moisture and energy which are often present in numerical weather prediction models, and which degrade the accuracy of forecasts"* (NLDAS website).

Figure 13 shows the empirical joint-density of soil moisture and temperature using the hourly NLDAS data for the March-August window. The shape of the density and iso-density curves clearly show that high temperatures are more likely when soil moisture is low. This pattern is even more salient during hot periods of the day (e.g., 4:00PM), suggesting that the high temperatures of the day are particularly correlated with low levels of soil moisture.

The fact that high temperatures are correlated with dry soil and that dry soil is a good explanatory variable for yield gives clear evidence that soil moisture is an omitted variable in models using season-long precipitation variables such as the SR model, Schlenker et al. (2005) and Deschênes and Greenstone (2007). Because dry soil negatively affects yield, the direction of the bias is downward,

A. Joint density over all hours of the day
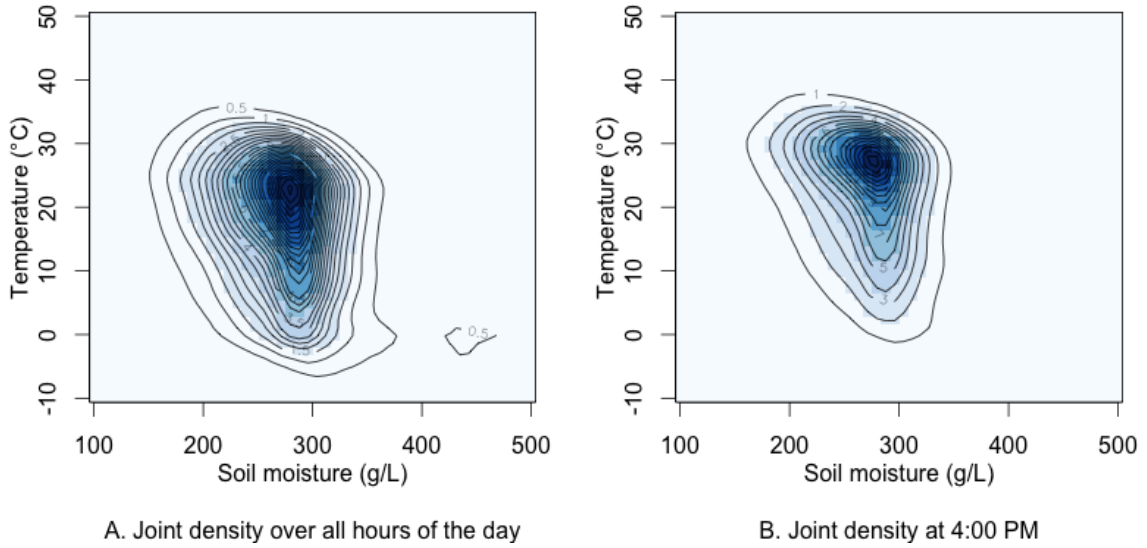
B. Joint density at 4:00 PM

Figure 13: Empirical joint density of hourly soil moisture and temperature for rainfed counties during the March-August period (1979-2011).

toward greater damages from extreme temperature. Thus, the results in this section imply to an overestimation of extreme temperature effects of about 30% in models that use only a season-long measure of precipitation.

3.3 Robustness analysis

Aside from the qualitative implications of accounting for soil moisture in the OB model, the inclusion of soil moisture also yields improved statistical fit. However, the OB model introduces a relatively large number of parameters.[16] A genuine concern is that the improvement is only artificial. To test this possibility, I rely on the fact that the SR model is nested within the OB model to run likelihood ratio tests of whether the improved fit is statistically significant given the number of additional parameters. The tests strongly reject the hypotheses ($p < 0.000001$) that the improved fit is random.

Figure 14 shows out-of-sample reductions of root mean squared error (RMSE) with respect to a model that regresses log yield only on a county time-trend. Years were sampled 1000 times at random for sample splits representing 20, 50 and 80% of the observations in the sample. Estimated parameters at each round were used to forecast out-of-sample observations. The reductions in average RMSE are reported. The RMSE reductions range from 40 to 75%.

The OB model outperforms the SR model in out-of-sample predictions for all sample splits. If

---

[16]Specifications OB1, OB2, OB3, and OB4 introduce an additional 192, 71, 64 and 64 parameters ,respectively, with respect to the corresponding SR specifications.
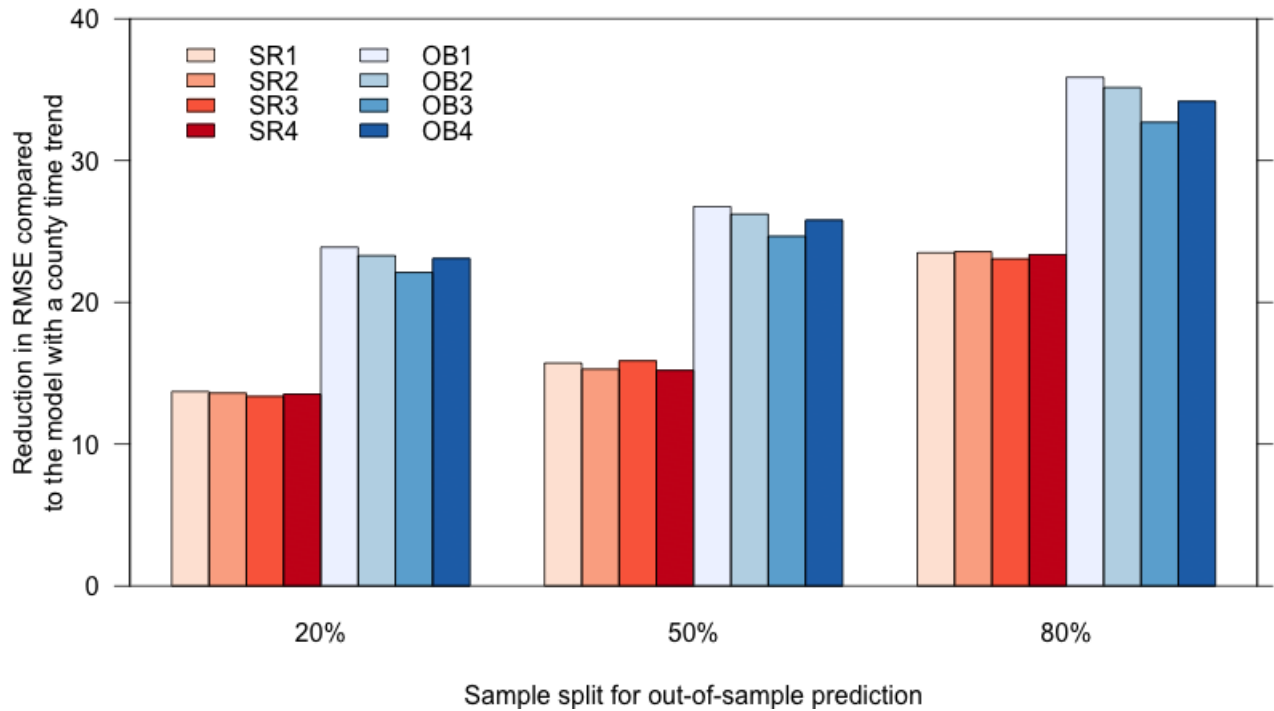
Figure 14: Model fit and out-of-sample predictions.

the model is over-fitting observations, the out-of-sample superiority would be expected to deteriorate as larger splits of the data are used for out-of-sample prediction. However, this is not the case.

4 Climate Change Impacts

Statistical models are commonly used to assess the potential impacts of climate change on agriculture. The conventional approach is to multiply the estimated parameters by the projected mean changes in regressors under alternative climate scenarios. Effectively, this approach relies on the estimated yield sensitivity to environmental conditions during the sample period, and predicts yield changes conditional on the projected changes in those conditions.

The changes in temperature, precipitation, and soil moisture conditions are presented in figure 7. The general pattern for temperature is an increase in the frequency of high temperatures, particularly around 30 to 35°C. The general pattern for precipitation is toward a slight to moderate decrease under all scenarios. Soil moisture changes, on the other hand, present a rich picture of seasonal dependence because predicted changes vary at different stages of the season. The general pattern points to an increase in soil moisture earlier in the season and a drying-out towards the end of the current growing season such as could not be captured by a season-long variable.

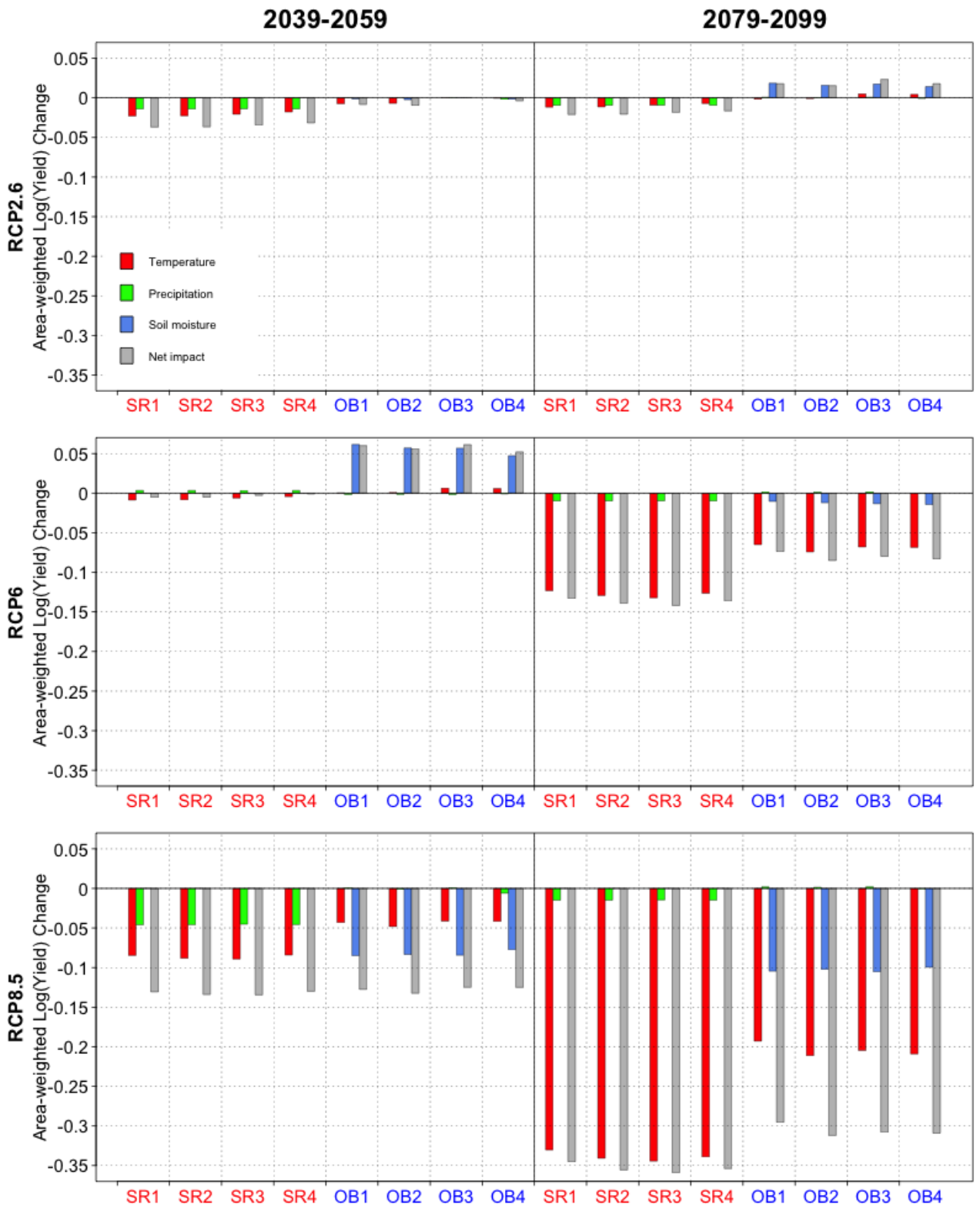The contribution of each variable (temperature, precipitation and soil moisture) as well as their

Figure 15: Climate change impacts and individual variable contributions

61

joint net effect on yield are presented in figure 15 for all models and climate change scenarios for the 2039-2059 and 2079-2099 periods.

Impacts for the low warming scenario RCP2.6 are close to zero (in the top row). The SR model predicts mid-century yield reductions of about 3% while the OB model predicts even smaller yield effects of about 1%. For the end of the century, the SR model predicts even smaller yield reductions than in mid-century, while the OB model predicts slight yield increases of about 2%. Under the RCP2.6 scenario, the temperature stabilizes around the middle of the century, which explains why detrimental temperature effects do not increase over the century. However, soil moisture patterns change as shown in figure 7C. In particular, soil moisture increases over much of the season with the exception of sharp moisture reductions in the last two season stages when lower soil moisture is less damaging. While the SR model predicts small damages from lower precipitation, the OB model predicts small gains from increases in soil moisture during key stages, particularly at the end of the century. This underscores the importance of accounting for the timing of climate changes within the growing season.

Impacts for the most severe scenario RCP8.5 (in the bottom row) are negative for both models although they are somewhat smaller for the OB model at the end of the century (right column in figure 15). However, the impact channels, as shown by the relative role of variables in each model, differ considerably. In the SR model, temperatures overwhelmingly drive negative impacts as indicated by the long red bars. The reduction in season-long precipitation plays a slightly negative and relatively small role.

The OB model suggests a very different relative role of variables in the most severe scenario. At mid-century, negative effects from soil moisture exceed negative effects from temperature. These negative impacts of soil moisture stem from detrimental decreases of soil moisture during the middle and the end of the growing season. At the end of the century, temperature damages exceed soil moisture damages, but their relative role is much lower than in the SR model. Damages from temperature are over 30% lower when soil moisture is considered. This stems from the lower damages from high temperatures in the OB model as illustrated in the model section by figure 12.

The medium warming scenario RCP6 presents the most interesting and contrasting impacts for both mid-century and end-of-century periods. For the mid-century period, the SR model implies virtually no impacts while the OB model suggests positive effects of about 5%, driven by increases in soil moisture during the first half of the growing season (see the central column of figure 15C). Toward the end of the century, the SR model predicts impacts of about -13% while the OB model predicts damaging effects of little more than half as much at about -8%. Again, temperature effects

in the OB model play a relatively smaller role (about half) and end-of-season soil drying explains the remaining part of the damage.

In summary, these findings suggest that accounting for soil moisture changes both overall impacts for some scenarios, but especially the relative role of variables driving these impacts. While precipitation is found to have a very small role in the SR model, soil moisture is a major factor in explaining impacts in the OB model. Furthermore, accounting for soil moisture reduces the share of the impacts attributable to heat stress by half in scenarios with the largest damages.

Finally, it is crucial to emphasize that these results assume a fixed growing season. A warmer climate, for instance, lengthens the growing season and provides added planting flexibility. In this context, the added intra-seasonal soil moisture representations of this model provide unique insights. A shift toward earlier planting dates, which is the direction found to be possible and beneficial in ?, would undoubtedly move the growing season in the direction where critical soil moisture levels are increased under the more severe climate change scenario. In other words, these findings show that much of the negative impacts found in these simulations are due to detrimental conditions that can be avoided, even more than could be accounted for by models ignoring intra-seasonal moisture changes.

5   Discussion

Statistical yield models are and will be a critical component of econometric climate change impact assessment models for agriculture as an alternative to biophysical process-based impact models. The fundamental strength of the structural econometric approach will be the ability to include farmer adaptation behavior grounded in the revealed preference paradigm. Observed yield fluctuations reflect optimal decisions based on within-season adaptations to cope with a changing and exogenous weather.

Climate change impacts will depend critically on the ability of farmers to adapt to changing environmental situations. For reliable estimates of adaptation possibilities and assessment of plausibility, the role of major variables must be unpacked in overall estimates. The current widespread approach of relying on season-long precipitation variables for capturing water availability underestimates the role of drought stress in climate change impact studies. This underestimation leads to an almost doubling of overall implicit damages for the middle warming scenario RCP6 at the end of the century. Because low soil moisture negatively affects crop yield but is correlated with high temperatures, the exclusion of soil moisture variables leads to omitted variable bias that suggests an even higher detrimental effect of temperature.

Models that suggest that water supply plays a limited role in climate change impacts in contrast to the central detrimental role of high temperature have suggested a dire future for US agriculture. These models suggest that access to water management practices, such as changing planting dates, or changing irrigation or no-till farming practices that help control the timing or keep moisture in the soil, would play only a marginal role if the overwhelming impacts are driven by heat stress alone.

On the contrary, however, accounting for soil moisture and its timing throughout the season shows that water availability is and could be a major factor in explaining potential impacts. For the mid-century projections, soil moisture appear to be the most determining factor in explaining yield impacts. This offers a more complete picture of agricultural impacts, and makes clear the fact that both, heat and drought stress will play major roles.

Turning to policy implications, agricultural adaptation policy should be concerned not only about resilience to heat but also to drought. Better modeling of channels is crucial to attribute effects to interrelated environmental variables. Relying on simple variables such as total precipitation can omit factors that are correlated with other variables in the model and thus generate bias in predicted patterns of climate change impact channels.

Because soil moisture data is difficult to obtain, some might be tempted to justify the use of models that omit soil moisture conditions, suggesting instead that temperature effects serve as a valid proxy for both heat and drought related stress. However, the results of this paper show that this justification is flawed. The validity of a proxy depends not only on its good correlation with the variables of interest during the estimation sample period, but also on whether this correlation is *maintained* during the projection period, which in this case is many decades into the future.

That temperature and soil moisture conditions will maintain the same correlation in the future is a cavalier assumption. For instance, an important implication is that the patterns emerging from the HadGEM2 point to a wetter early season but dryer late season. This is not the same pattern found for temperature. Thus, the correlation justifying extreme temperature as an appropriate proxy is not warranted for climate change analysis.

Moreover, given that the non-freezing period will be longer with warmer temperatures, farmers will very likely have greater flexibility in choosing planting dates. Given that the most sensitive period to drought is toward the middle of the season, earlier planting would possibly lead to substantial yield damage reductions through summer and fall drought avoidance. **?** show this mechanism is important in avoiding heat stress during the sensitive flowering period in corn. Their results suggest that earlier planting, ranging from 2 to 3 weeks depending on the state, reduces corn yield impacts of a uniform 5°F warming scenario by 30 to 70% in the Upper Midwest. Interestingly, this is the

same direction of change in planting dates that would tend to increase soil moisture in the critical time of crop development under all climate change scenarios analyzed in this paper. In summary, shifting the growing season earlier in the calendar will plausibly lead to substantial gains both from heat and drought stress avoidance.

6   Conclusion

This paper develops a statistical crop yield model that accounts for both nonlinear temperature effects and nonlinear soil moisture effects throughout the crop season. Because soil moisture is not recorded over large areas, the model makes use of the state-of-the-art NLDAS dataset with hourly and 14km resolution observations of environmental conditions. I contrast this model with a leading model in the literature by Schlenker and Roberts (2009a) that accounts for water availability through a season-long precipitation variable.

Findings suggests that water availability plays a much greater role than previously suggested by the competing model. Yields are found to be very sensitive to soil moisture conditions particularly toward the middle of the season, precisely when high water demand and sensitivity to drought are expected.

Because of well-known correlations between soil moisture and high temperatures, omitting soil moisture conditions from statistical models overestimates damages by almost 100% by the end of the century for the medium warming scenario (RCP6). This is also reflected in the projected climate change impacts. Temperature effects play a substantially smaller role, ranging from a third to a half less in overall impacts, than in models omitting soil moisture. On the other hand, patterns in climate change projections from the HadGEM2 model suggest that temperature alone should not be considered as an appropriate proxy to capture dry soil conditions because the correlation between these two variables is not warranted in the climate change forecasts (although it might serve as a good proxy in the sample period).

The inclusion of soil moisture conditions also substantially and significantly improves model fit. Results indicate that the improved fit is not the result of over-fitting as out-of-sample predictions do not deteriorate as smaller shares of the sample are used for prediction.

This paper suggests that precipitation, and more precisely soil moisture, is a crucial aspect of climate change impact assessment for agriculture. It also warns that the omission of soil moisture conditions can lead to overestimation of heat related stress. This counters the prevailing view in the statistical literature that future impacts and adaptation possibilities would primarily hinge upon crop resilience to heat stress. These results point to a more complete understanding that both

heat and drought stress will have fairly large roles in driving impacts, and these roles might change depending on the scenario under consideration.

The empirical model validated by this paper can have a number of useful applications both within and beyond climate change impact assessment. Most importantly, a model with the richness of soil moisture conditions is needed to add assessment of farmer adaptation possibilities using revealed preference data and models. However, in the short run, extreme weather can jeopardize harvests and lead to drastic increases in food prices with serious economic and social implications. This model coupled with the rising availability of remote sensing data for weather and phenological information could be an important part of an early-season warning system for regional or global food crises.

Another related application could be in improving early-season crop yield forecasts. For example, the USDA produces early season forecasts of crop production based on extensive survey data obtained by expensive agronomic sampling techniques requiring localized quantification of crop yield components (plant density, number of kernels per ear, kernel weight, etc). By using highly detailed remote sensing data, the approach of this paper could yield competing estimates at a fraction of the cost. These early-season forecasts could eventually compete with heavily parametrized process-based crop models used by traders in agricultural commodity futures markets.

A final word of caution applies to this form of climate change impact assessment. Because greenhouse gas concentrations do not vary significantly during 1981-2011, the approach cannot possibly account for the effects of CO_2 fertilization. In addition, the approach generally accounts for changes in mean climate and ignores the potentially crucial impacts of change in climate variability. However, this approach offers a first approximation of potential damages if the overall sensitivity of yields, growing regions, and seasons remain unchanged. Complementary studies can account for other additional sources of adaptation and yield more nuanced climate change impact scenarios.

Appendix

A1. Determining growing season boundaries

As indicated on footnote 7, crop progress reporting began too late (the state had already surpassed the 50% acreage level) or stopped too early (the state had not yet reached the 50% acreage level) for a few states and years. For these cases, I obtain the median acreage date by extrapolation.

For this purpose I estimate, through non-linear least squares, a 2-parameter logistic model for all observations for a given stage and state. The model has a common slope or discrimination parameter a and year-specific threshold or difficulty parameters b_{year}. The model regresses stage

progress $PROG$ on day of the year DOY for a given state and stage. The model is:

$$PROG_{DOY,year} = (1 + \exp(-a(DOY - b_{year})))^{-1} + \epsilon_{DOY,year}$$

Figure 16, on the left, shows the fit of the model for North Carolina (red line) corresponding to the silking stage for the average year, i.e. for $b = \bar{b}$. The common parameter a assumes that silking progress has a similar "shape" from year to year. The year-specific threshold parameter b_{year} allows the curve to be shifted horizontally. Allowing for b to vary over years is important because it is precisely for unusually early-planting and late-harvesting years that progress data lacks median progress dates. The fit of the model for the incomplete year of 1995 (shown on the right) is shown in a red dotted line and the extrapolated progress observations are shown as red dots. The crop stage boundary is obtained when the extrapolated curve reaches 50% of the state's acreage.

The paper requires stage boundary dates for 14 states, 3 crop stages (planted, silking, mature) and 31 years (1981-2011), or a total of 1302 stage boundaries. The interpolation procedure was necessary for only 55 cases or less than 5% of the cases. This concerned the states of North Carolina (42 cases), Pennsylvania (3), Missouri (2), Illinois (1), Indiana (1), Kansas (1), Kentucky (1), Michigan (1), Ohio (1), South Dakota (1) and Wisconsin (1).

A2. More on the SR model

The specification for the model with a step function allowing different effects at each 3ºC interval (SR2) is:

$$y_{it} = \sum_{h=0,3,6,9...}^{39} \gamma_h \underbrace{[\Phi_{it}(h+3) - \Phi_{it}(h)]}_{x_{it,h}} + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i + \epsilon_{it}$$

The model effectively regresses yield on the time spent within each interval in a given county and year $x_{it,h}$.

Model SR3 assumes that the "yield growth" function $g(h)$ is an eighth-degree polynomial of the form $g(h) = \sum_{j=1}^{8} \gamma_j T_j(h)$ where where $T_j()$ is the jth order Chebyshev polynomial. Replacing $g(h)$ with this expression yields:
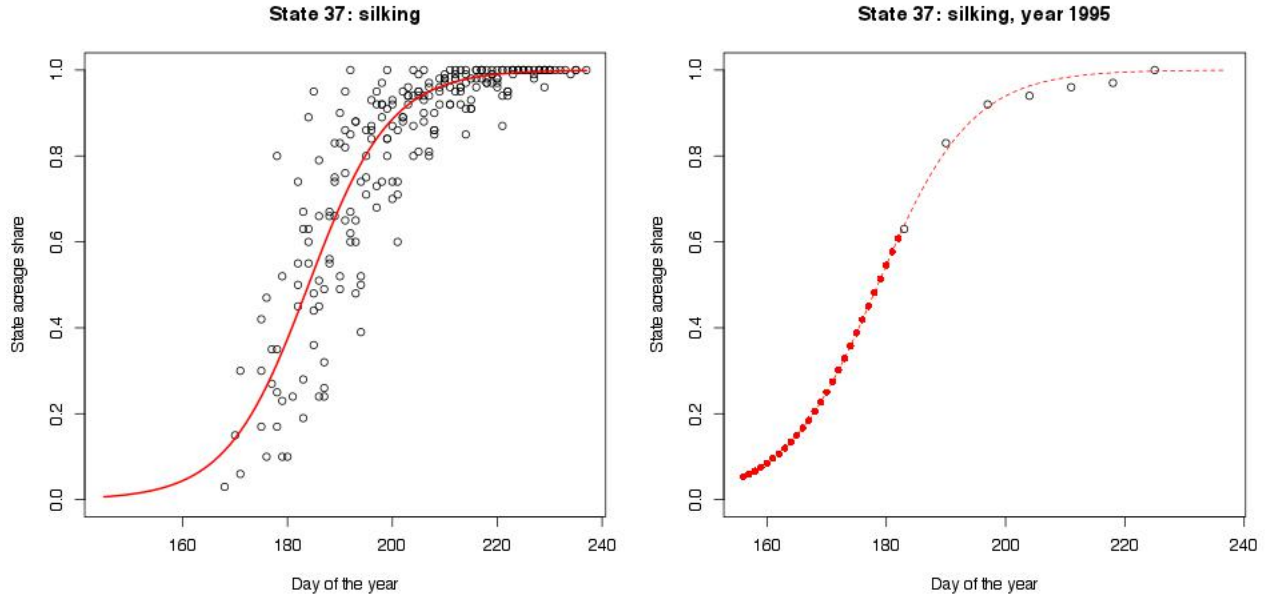
The graph on the left shows the 2-parameter logistic fitted model for the average year in red. On the right, the dotted red line represents the model for year 1995 and the red dots are the extrapolated progress levels. The boundary date for year 1995 is obtained from the extrapolated progress reaches 50% of the state's acreage that year.

Figure 16: Stage boundaries for years with incomplete progress data

$$
\begin{aligned}
y_{it} &= \sum_{h=-1}^{39} \sum_{j=1}^{8} \gamma_j T_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)] + p_{it}\delta + z_{it}\tau + c_i + \epsilon_{it} \\
&= \sum_{j=1}^{8} \gamma_j \underbrace{\sum_{h=-1}^{39} T_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)]}_{x_{it,j}} + p_{it}\delta + z_{it}\tau + c_i + \epsilon_{it}
\end{aligned}
$$

The model effectively regresses yield on eight temperature variables $x_{it,j}$ which represent the jth-order Chebyshev polynomial evaluated at each temperature bin.

In a similar fashion, model SR4 assumes that $g(h) = \sum_{j=1}^{8} \gamma_j S_j^3(h)$ where $S_j^3()$ is the piece-wise cubic polynomial evaluated for each jth interval defined by eight control points.

$$
y_{it} = \sum_{j=1}^{8} \gamma_j \underbrace{\sum_{h=-1}^{39} S_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)]}_{x_{it,j}} + p_{it}\delta + z_{it}\tau + c_i + \epsilon_{it}
$$

Figure 17 and 18 presents the results for all specification for rainfed and irrigated counties.

Figure 17 shows, for rainfed counties on the left, a close agreement across specifications for the

damaging effects of temperatures above 30ºC. The polynomial (SR3) and spline specifications (SR4) show a peculiar upward bent which is not significant due to the low number of observations over that extreme range. For irrigated counties on the right, results are not as clear. Temperatures around 15 and 30 appear beneficial but temperatures around 10 and 22 and above 30 are detrimental. This repeated inversion on the sign of temperature effects is odd and has no clear physical underpinning. A possibility is that this pattern reflects mixing effects of day-time and night-time temperature exposure. However, the damaging nature of temperatures over 30ºC is of similar magnitude to rainfed counties.

Regarding precipitation in figure 18, the response curve for rainfed counties is very similar across specifications, with very low ($<$400mm) and very high ($>$800mm) precipitation levels reducing yield. However, this response curve is almost flat for irrigated counties on the right column, as expected. Indeed, Farmers in irrigated areas control the water supply for very dry years. However, very high levels of precipitation seem to reduce yield, and this could be consistent with damages from flooding events.

A3. More on the OB model

The specification for the model with a step function allowing different effects at each 30 g/L interval (OB2) is:

$$
\begin{aligned}
y_{it} \quad = \quad & \sum_{h=0,3,6,9...}^{39} \gamma_h \underbrace{\left[\Phi_{it}(h+3) - \Phi_{it}(h)\right]}_{x_{it,h}} + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i \\
& + \sum_{s=1}^{8} \sum_{m=100,130...}^{340} f(m+15,s) \underbrace{\left[\Psi_{it}(m+30,s) - \Psi_{it}(m,s)\right]}_{z_{it,m}} + \epsilon_{it}
\end{aligned}
$$

Model OB3 assumes that the "yield growth" function $g(h)$ is an eighth-degree polynomial of the form $f(m,s) = \sum_{j=1}^{8} \lambda_{js} M_j(m,s)$ where where $M_j()$ is the jth order Chebyshev polynomial. Replacing $f(m,s)$ with this expression yields:
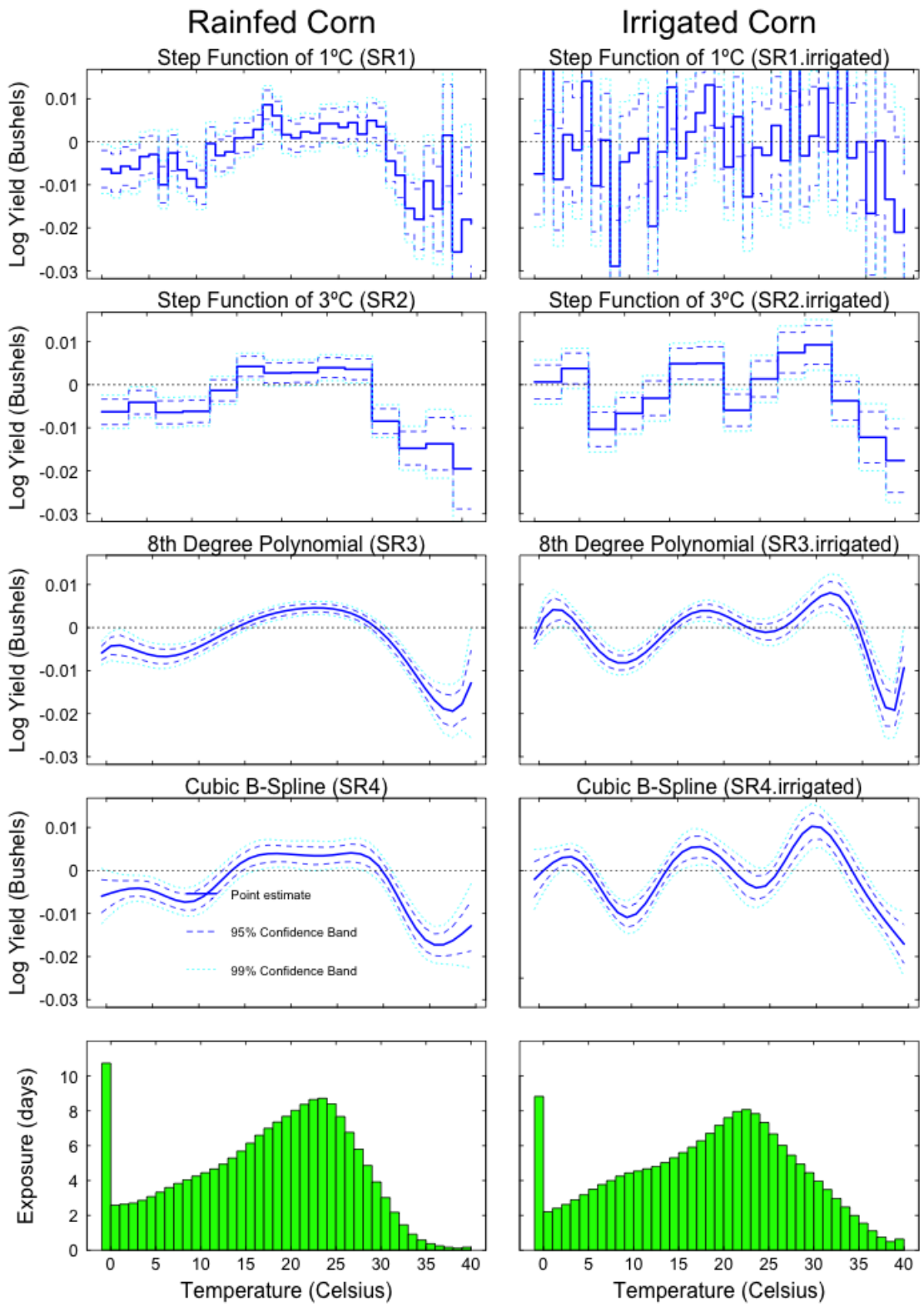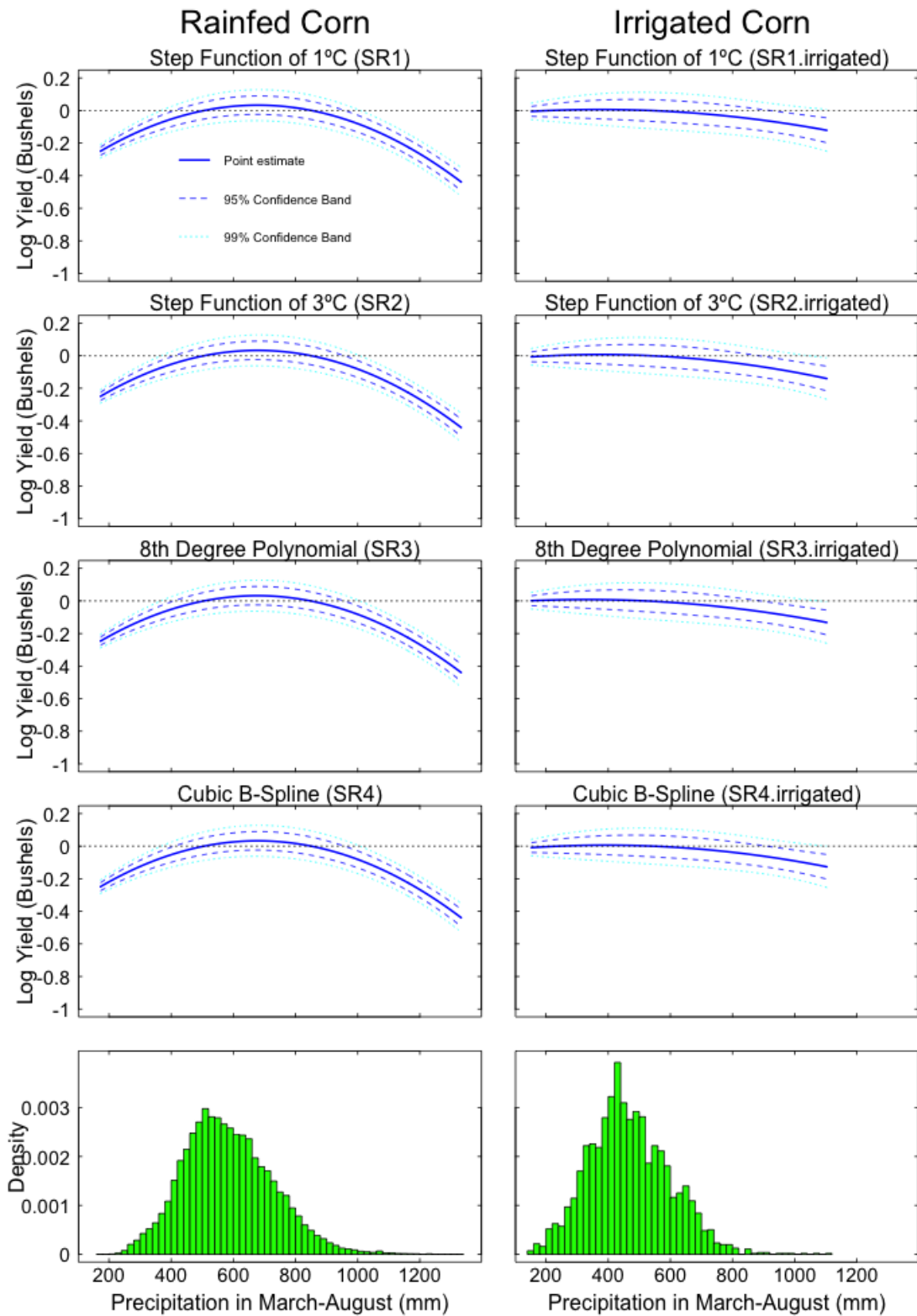
Figure 17: Temperature effects for the SR model

Figure 18: Precipitation effects for the SR model

$$y_{it} = \sum_{h=-1}^{39} \sum_{j=1}^{8} \gamma_j T_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)] + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i$$

$$+ \sum_{s=1}^{8} \sum_{m=120}^{350} \sum_{j=1}^{8} \lambda_{js} M_j(m+5,s)[\Psi_{it}(m+10,s) - \Psi_{it}(m,s)] + \epsilon_{it}$$

$$= \sum_{j=1}^{8} \gamma_j \underbrace{\sum_{h=-1}^{39} T_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)]}_{x_{it,j}} + p_{it}\delta_1 + p_{it}^2\delta_2 + z_{it}\tau + c_i$$

$$\sum_{s=1}^{8} \sum_{j=1}^{8} \lambda_{js} \underbrace{\sum_{m=120}^{350} M_j(m+5,s)[\Psi_{it}(m+10,s) - \Psi_{it}(m,s)]}_{z_{it,m}} + \epsilon_{it}$$

The model effectively regresses yield on eight temperature variables $x_{it,j}$ and eight moisture variables for eight different stages $z_{it,m}$. Each variable represents the jth-order Chebyshev polynomial evaluated at each temperature and moisture bin.

In a similar fashion, model OB4 assumes that $f(m,s) = \sum_{j=1}^{8} \lambda_{js} Z_j^3(m,s)$ where $Z_j^3()$ is the piece-wise cubic polynomial evaluated for each jth interval defined by eight control points:

$$y_{it} = \sum_{j=1}^{8} \gamma_j \underbrace{\sum_{h=-1}^{39} S_j(h+0.5)[\Phi_{it}(h+1) - \Phi_{it}(h)]}_{x_{it,j}} + p_{it}\delta + z_{it}\tau + c_i$$

$$+ \sum_{s=1}^{8} \sum_{j=1}^{8} \lambda_{js} \underbrace{\sum_{m=120}^{350} Z_j(m+5,s)[\Psi_{it}(m+10,s) - \Psi_{it}(m,s)]}_{z_{it,m}} + \epsilon_{it}$$

Figures 19 through 23 present the results for all specifications for rainfed and irrigated counties.

Figure 19 shows the temperature response functions for the OB model. The left column on rainfed counties shows that agreement over all for damages above 30⁰C. The confidence bands become much wider for extreme temperature and the polynomial (OB3) and spline specifications (OB4) show the same peculiar upward bent than in SR3 and SR4. However, this quirk is not significant. The right column for irrigated counties exhibit a rather different response function, although it also suggests negative effects of high temperature.

Figure 20 shows the precipitation response functions for the OB model. Aside from the width of confidence bands, all response functions are extremely similar across specifications and for both rainfed and irrigated counties. They are also very similar to the precipitation response function for

the SR model over irrigated areas (on the right column of figure 18). This is evidence that once soil moisture is accounted for in rainfed areas, precipitation only captures yield variation for very high precipitation levels (e.g. flooding).

Figure 21 shows the soil moisture response functions for the polynomial (OB3) and spline (OB4) specifications. The exhibit fairly similar results for rainfed counties, with the strongest yield responses toward the middle of the season. These response functions are statistically significant as shown in the left columns of figures 22 and 23, which show confidence bands.

The soil moisture response functions for irrigated counties were included in figure 21 as a falsification exercise. Because soil moisture data do not account for irrigation, we should expect the variable to explain yield variation much. Indeed, the surfaces are rather flat, with the exception of very high moisture values which happen to be insignificant, as shown on the left columns of figures 22 and 23. This clearly shows that very low levels of predicted moisture do not explain yield in irrigated counties, as expected.
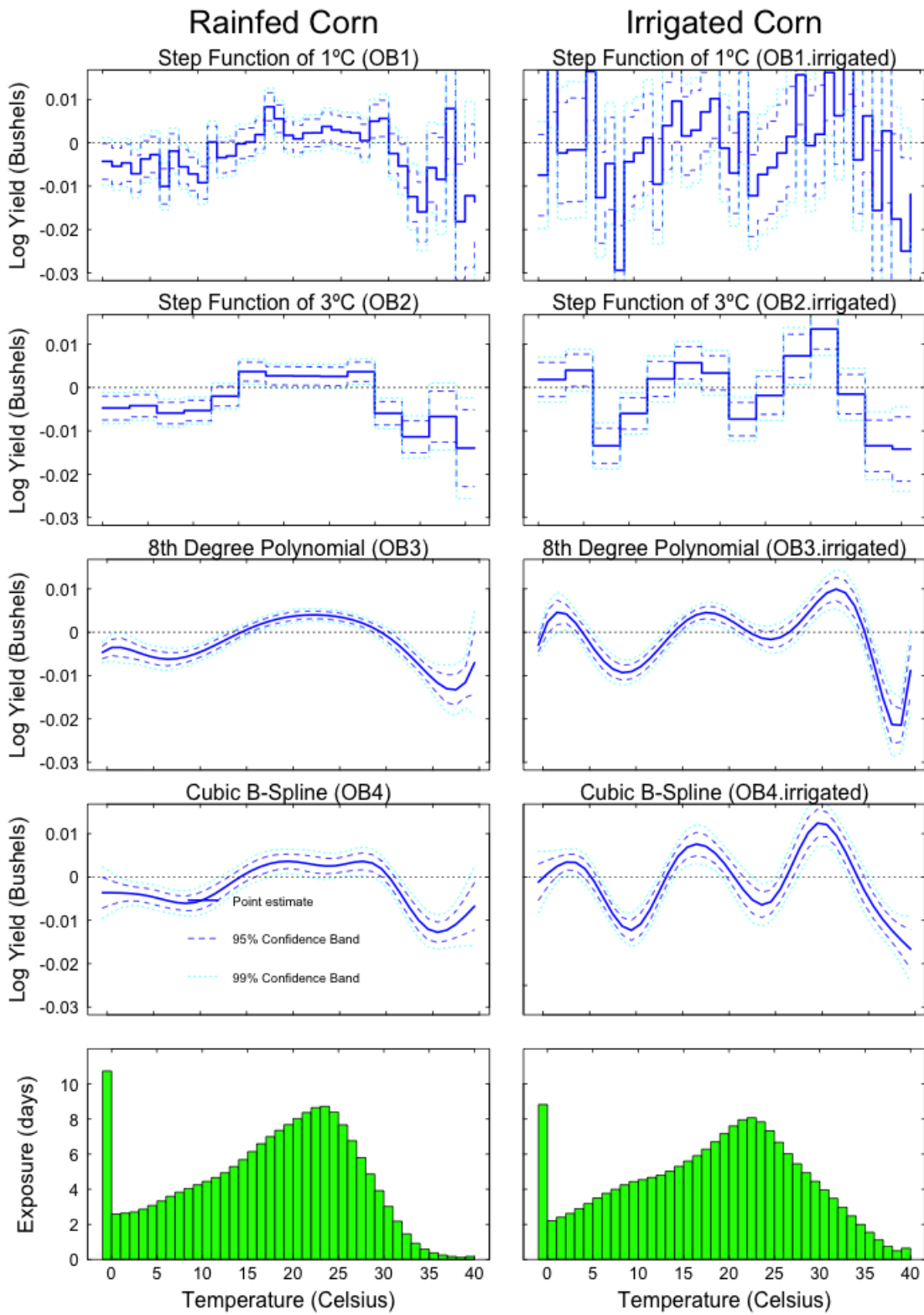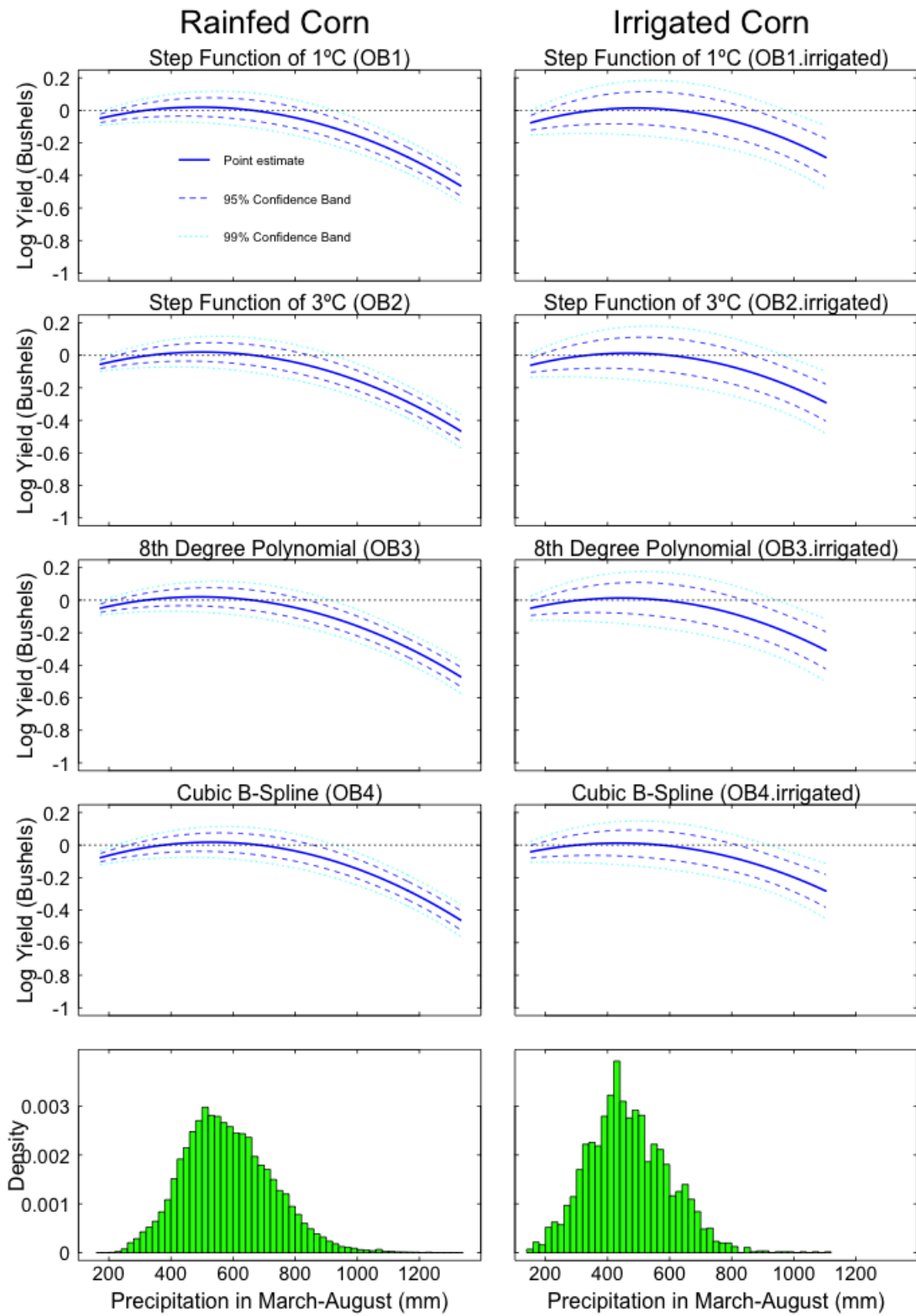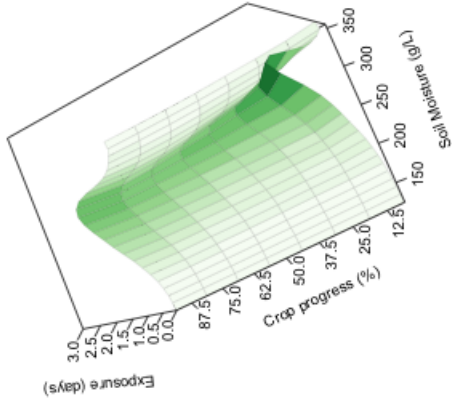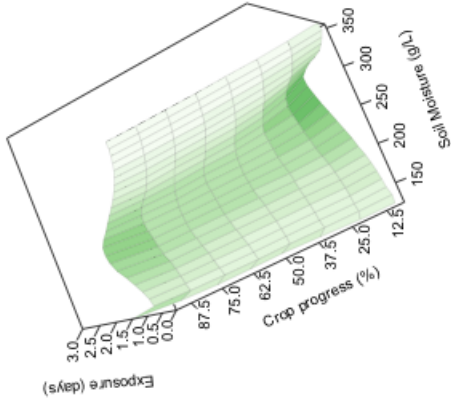
Figure 19: Temperature effects for the OB model

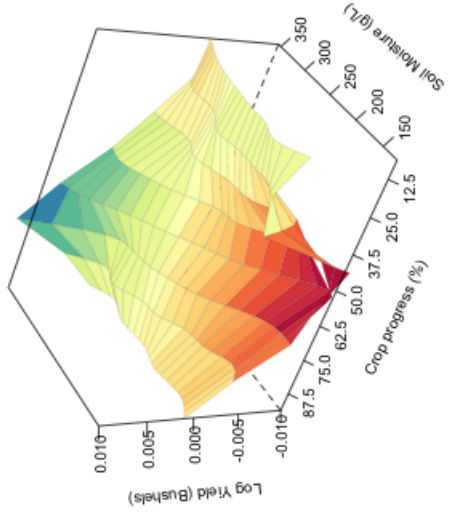Figure 20: Precipitation effects for the OB model
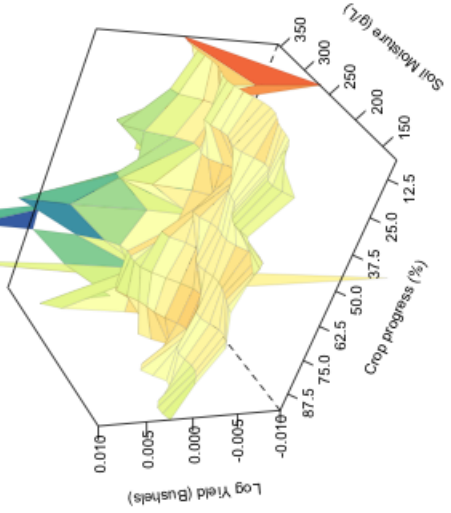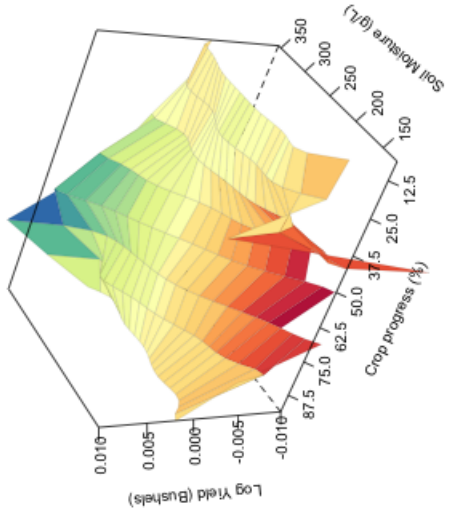
Figure 21: Soil moisture effects for the OB model

Figure 22: Soil moisture effect for the 8th Degree Polynomial specification (OB3)

Figure 23: Soil moisture effect for the Cubic B-Spline specification (OB4)

Essay IV

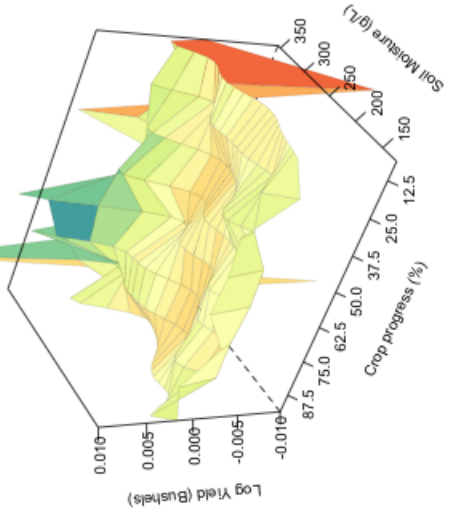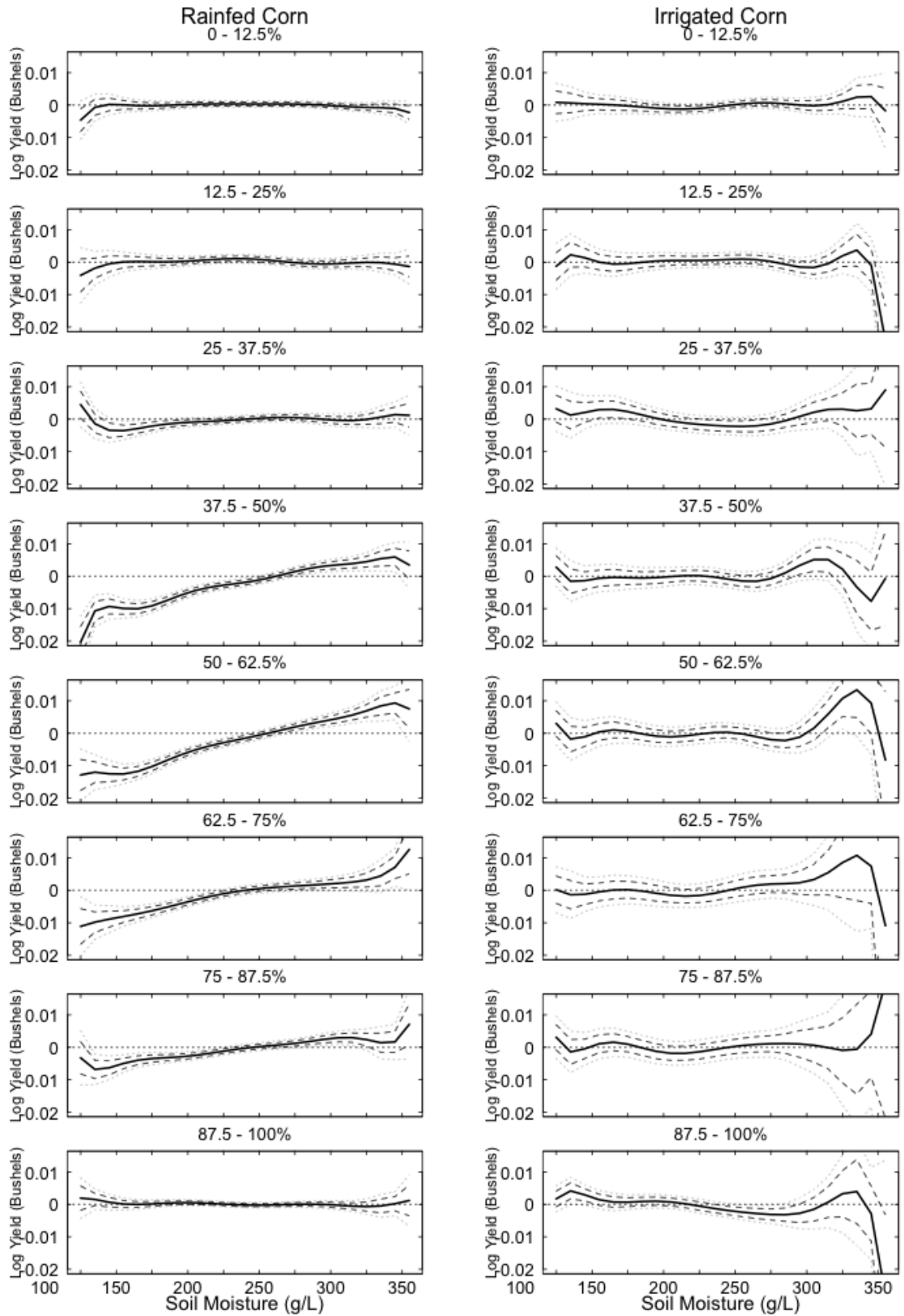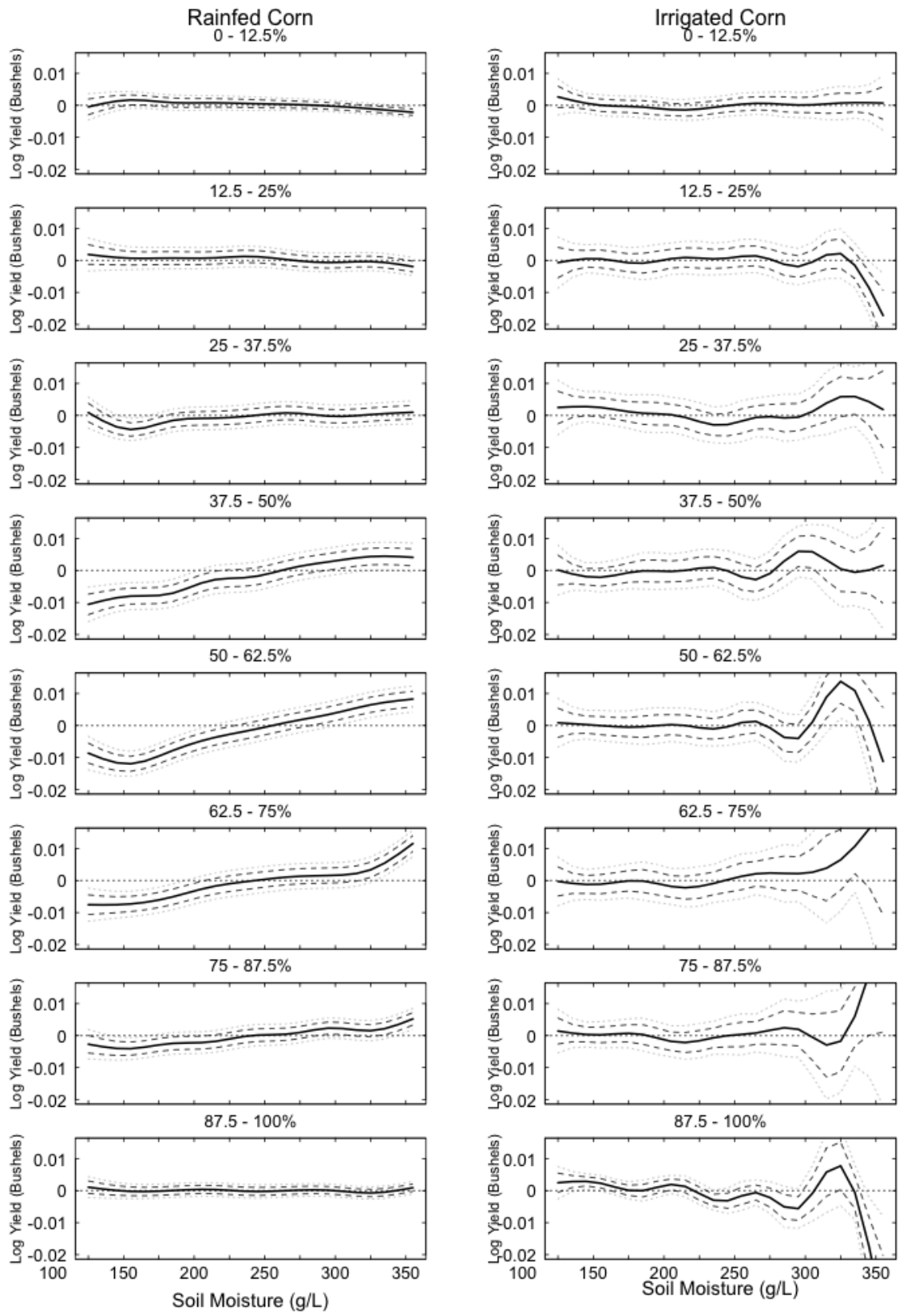# Modeling the Structure of Adaptation in Climate Change Impact Assessment [1]

Abstract

The paper proposes a structural approach for elucidating the mechanisms through which climate change will affect agriculture. Clarifying specific adaptation possibilities facilitates not only the assessment of potential welfare impacts, but also offers the possibility of evaluating policies for improved adaptation. This is in stark contrast with prevalent reduced form approaches in the literature that provide impact estimates without identifying adaptation mechanisms. An empirical illustration of the crop yield impact channel showcases how this analysis provides insights about adaptation possibilities. The example shows how Midwest corn producers from 8 states could reduce yield damages from a 5°F warming by as much as 30 to 70 percent through earlier planting, thus saving an estimated \$3.4 billion annually. The adaptation is made possible by the lengthening of the growing season which provides greater flexibility to farmers for reducing the exposure of sensitive periods of the growing season to detrimental conditions in the summer.

JEL Classification Codes: Q54, Q15, Q51

Keywords: climate change, agriculture, adaptation

---

1 Introduction

While a major focus of econometric climate impact assessments on agriculture has been prediction of overall impacts, future research should identify impact mechanisms and adaptation possibilities. Clarifying specific adaptation possibilities facilitates not only the assessment of potential welfare impacts, but also offers the possibility of evaluating policies for improved adaptation. This depends on capturing mechanisms that provide farmers' abilities to adapt to new climatic constraints in counter-factual conditions.

These impact mechanisms are represented with elaborate detail in agronomic crop models that convey the science of crop production. However, the agronomic models are not well integrated with revealed preferences (e.g., Adams 1989, Adams et al. 1990, Easterling et al. 1992, Rosenzweig and Parry 1994). Thus, congruence of agronomic adaptation possibilities with economic behavior that might be observed in counterfactual circumstances is open to question.

Econometric methods have attempted to represent adaptation implicitly by estimating reduced-form relationships between economic variables and arbitrary forms of aggregate weather measures. Leading examples include the Ricardian approach based on cross-section regression of land prices on weather variables (Mendelsohn, Nordhaus and Shaw 1994 and Schlenker, Hanemann and Fisher 2005, henceforth MNS and SHF) and the profit panel approach consisting of fixed-effects regressions of net annual revenue on weather variables (Deschênes and Greenstone 2007, henceforth DG). Thus, modeling shortcuts have been used to assess potential impacts of exogenous weather variation without modeling decision-making and adaptive innovation explicitly, and without consideration of the specific weather variables of importance in the science of crop production. Therefore, land prices and observed net revenues may capture farmers' optimal adaptive behavior with an unknown degree of imperfection.

While these highly reduced-form approaches have provided first-cut estimates of climate effects, they do not reflect the mechanisms through which impacts occur, which calls into question the feasibility of predicted adaptive behavior as well as robustness to omitted variables bias. Aggregated approaches also prevent identification of structural relationships necessary to consider adaptation policy assessment and cross-validation.

Recently, research using the econometric approach has focused increasingly on impact mechanisms partly as a means of validating results from reduced-form approaches. This includes renewed interest in statistical yield models (e.g. Schlenker and Roberts 2009a, Lobell and Burke 2010) because crop yields represent major mechanisms through which higher temperatures may affect pro-

ducer welfare. However, most yield models rely on season-long weather variables that overlook the varying sensitivity of crops during the growing cycle, and implicitly assume that growing seasons remain fixed.

Under-representation of flexibility causes overestimation of yield impacts. An example is estimation of heat effects ignoring the flexibility offered by lengthening of the growing season. Conventional agronomic wisdom established through field trials on annual crops is that stress during the relatively short flowering period reduces yield more than in any other stage of growth (Fageria, Baligar, Clark and Clark 2006, p.89). This phenomenon is substantial and statistically significant in US county corn yields (Ortiz-Bobea 2011). Thus, a longer season may allow flexibility to shift the flowering period away from a hotter traditional flowering period. This flexibility is ignored by typical econometric approaches, although common in the agronomic models. Other research has considered the agronomic analysis of agricultural zones (Newman 1980, Adams et al. 1990, Kaiser et al. 1993) by considering potential changes in crop mix using multinomial logit models (Mendelsohn and Dinar 2009, ch.5). However, these possibilities are typically considered separately rather than jointly. Of course, capturing all adaptation possibilities is a daunting task given the diversity of agriculture. But accounting for major and obvious adaptation strategies based on revealed preferences provides a critical foundation for adaptation policy analysis. Preliminary work, including Ortiz-Bobea (2012) and an example in this paper, implies that climate change assessment should not stop short of exploring these possibilities.

2   A structural approach

This paper proposes an econometric framework for assessing potential impacts of climate change on agriculture that tractably unpacks some of the major impact mechanisms. Following the implicit definition in other empirical work, we define climate as the probability distribution of all aspects of weather relevant to a particular period of time, but (i) define the relevant weather variables for our problem based on scientific knowledge of the underlying mechanisms of production, and (ii) use a behavioral model as an empirical underpinning to capture adaptation given those mechanisms. The key element is the explicit treatment of climate change within a classical constrained optimization framework given potential adaptive private and public actions. In this paper, we consider only the simple behavioral model of profit maximization, but more general applications based on revealed preferences are planned.

As an example, higher temperatures lead to the detrimental effects of hotter summers, but also

lengthen the frost-free period, offering farmers the option of longer-season cultivars, different crops, or even relay cropping. Only by a disaggregated approach can the potentially dominating mechanisms of both the detrimental and beneficial aspects of climate change be revealed. And only by combining agronomic knowledge with revealed preferences can these potential counter-factual mechanisms be properly assessed. In addition, this approach can serve to cross-validate qualitatively conflicting results of current leading econometric approaches (see SHF and DG).

For structural modeling, our proposed approach, like most others, presumes prior knowledge of the distribution of climatic inputs and the major climatic constraints imposed by climate change. Specifically, climatic inputs are characterized by the timing and level of their exogenous supply. Climatic constraints arise when the supply of climatic inputs render production infeasible. An example is the time of onset of the growing season which is driven by the last spring frost in the American Midwest. Adequate models must determine whether each constraint is binding in each locality and how its variation contributes to welfare (i.e., to each constraint's shadow price). A disaggregated approach can determine the significance of individual aspects of climate change and their geographic distribution. Shadow prices can then guide investment in adaptation research and related public policy, both topically and financially.

We focus on careful treatment of the physical role of weather variables in production as understood in the production sciences. For example, farmers in temperate US regions choose cultivars that reach maturity before fall frosts because freezing temperatures damage non-mature crops and result in significant yield loss. Reduced-form models attempt to capture this effect through correlations with weather variables such as average October temperature (MNS) or April-to-September growing degree-days (SHF, DG). However, arbitrary calendar variables are likely correlated (imperfectly) with relevant omitted factors, blurring their interpretation for adaptation policy analysis. In contrast, our approach is to rely on variables directly related to the probability distribution of the first fall frost date whereby the benefit of reaching maturity only a week or two later would be reflected in a simulation.

Estimates of such adaptation mechanisms can provide a transparent framework to assess the diverse effects of climate change on the agriculture. Preliminary results we exemplify below call for a fertile research agenda to estimate effects of individual climate change constraints. Such models hold promise for bridging the gap between the econometric and agronomic modeling families by developing a common ground for analysis. Further, structural modeling in a theoretical framework where relationships are qualitatively understood at the outset can reduce omitted variable bias and potential misinterpretation of reduced-form counterparts. For example, reduced-form approaches

can provide little basis for determining expected qualitative relationships. Structural approaches, on the other hand, can answer questions in terms of the estimated strength of qualitatively clear components necessary to facilitate welfare and policy analysis.

3 An optimization-based Approach

Our conceptual framework of behavior is a constrained optimization model where the farmer determines a vector of choice variables or weather-dependent choice rules, $x(\cdot)$, including choice of crop mix and cultivars, other technology choices such as machinery and irrigation/drainage investments, planting dates, and factor input levels. The optimization problem for a risk-neutral farmer with opportunity cost π_0 is

$$\max_{x(\cdot) \geq 0} \pi\left(x(\cdot), p, w, \theta\right) \equiv p(Q(\theta)).q(\theta, x(\theta)) - w(\theta) \cdot x(\theta) \quad \text{s.t. } \pi \geqslant \pi_0 \qquad (1)$$

where p and w are output and input price vectors, Q and q are market and farmer output vectors, and the vector θ describes the timing and level of the exogenous weather inputs.

Applying the envelope theorem to the profit function associated with (1) yields a decomposition of the long-run change in profit from a change in climate,

$$\left.\frac{\partial \pi}{\partial \theta}\right|_{x=x^*} = \frac{\partial p}{\partial Q}\frac{\partial Q}{\partial \theta}q(\theta, x^*(\theta)) + \left(\frac{\partial q}{\partial \theta} + \frac{\partial q}{\partial x^*}\frac{\partial x^*}{\partial \theta}\right)p(Q(\theta)) - \left(\frac{\partial w}{\partial \theta}x^*(\theta) + w(\theta)\frac{\partial x^*}{\partial \theta}\right) \qquad (2)$$

where x^* is the optimal decision vector. The first term represents the effect of output price on profit stemming from the large-scale effect of climate on aggregate supply given aggregate demand. This term is potentially significant if climate change affects production and product mix over broad areas. Grasping its magnitude requires estimating the correlation of heterogeneous regional climate change impacts and how they aggregate into global agricultural output and consequent local price impacts. This difficulty likely explains why climate change studies typically assume fixed prices (e.g. MNS, SHF, DG). However, this effect is likely to have attenuating implications because equilibrium price adjustments tend to spread economic effects across a broad array of markets and, thus, soften impacts on the most affected markets through product substitution.

The second term represents the contribution of climate change to profit through its effect on the individual farmer's output. Crop output can be expanded as the product of acreage $a(\theta)$ and yield $y(\theta, x(\theta))$ where

$$\frac{\partial q(\theta, x(\theta))}{\partial \theta} = \frac{\partial a(\theta)}{\partial \theta} \cdot y(\theta, x(\theta)) + \left(\frac{\partial y(\theta, x(\theta))}{\partial \theta} + \frac{\partial y(\theta, x(\theta))}{\partial x} \frac{\partial x(\theta)}{\partial \theta} \right) a(\theta)$$

and $a(\theta)$ is a subvector of $x(\theta)$. This expression highlights the importance of focusing carefully on the response of both the optimal crop mix and the yields of alternative crops to climate change, and how particular climate-dependent farmer responses affect each.

The third term in (2) measures the cost effect of climate change associated with climate-induced changes in input prices and input use. The former might stem from changing demand pressure on input markets. The latter arises from a wide range of possibilities for changing cultivation practices and crop mix. For example, a farmer might purchase more irrigation water on a given crop to compensate for reduced rainfall or adopt mitigating measures to maintain arable land in the event of an increase in farm-wide flooding, drought, or consequent adverse pest populations.

Such an optimization model thus provides a framework in which to analyze separate climate change impact mechanisms. Major adaptive behaviors within each channel can be explored separately to identify policy-relevant insights for improved adaptation. The model can obviously be expanded to consider additional mechanisms that affect adaptation subject to the limitations of econometric identification. For example, allowing risk aversion can facilitate welfare analysis of changing climate variability, in which case an analysis of reservation utility can shed light on agricultural regions that might no longer farm, or other regions that might begin farming.

4   An empirical illustration

Obviously, an empirical illustration of specific modeling of each of the mechanisms delineated in this model is beyond the scope and space limitations of this paper. Alternatively, we present an empirical example that explores the yield impact channel, $\partial y(\theta, x(\theta))/\partial x \cdot \partial x(\theta)/\partial \theta$, to illustrate subtle but important potential for plausible adaptive behavior and opportunities that tend to remain unexplored in reduced-form models. In particular, we explore the potential effect on corn yield of a 5°F uniform warming through both extreme heat during different crop production stages and the widening of the frost-free period. We explore how the climate-dependent choice of planting date, represented by $\partial x(\theta)/\partial \theta$, may affect yield.

The example is fundamentally based on a statistical corn yield model with weather regressors matched to key stages of the corn production cycle, namely the vegetative, flowering, and the grain-filling periods (see Ortiz-Bobea 2011). This allows estimation of phenological regression coefficients

that are disconnected from fixed calendar periods (Dixon et al. 1994). The advantage of this approach is that the regression results can be easily employed in simulations that allow for shifting growing seasons. Geographic variation in growing seasons can then be used to project future variation in growing seasons under climate change. The model specification is

$$y_{it} = \sum_s \left(\beta_{1,s} Prec_{its} + \beta_{2,s} Prec_{its}^2 + \beta_{3,s} GDD_{its} + \beta_{4,s} DDD_{its} \right) + \psi(t) + \alpha_i + \epsilon_{it}$$

where y_{it} is yield in county i in year t; s is the set of key stages of corn production; $Prec$, GDD and DDD are precipitation, growing degree-days (8-32°C) and damaging degree-days (>34°C); $\psi(t)$ is a quadratic time trend; and the α_i are county effects.

We use a county-level corn yield panel dataset (1985-2005) from a mostly rainfed area covering 8 states, which represents over 65 percent of US corn production. County production and state crop progress data were obtained from USDA-NASS and daily weather data for the 1950-2005 period is from Schlenker and Roberts (2009a).

Following the literature, simulated impacts are obtained by multiplying estimated parameters by the projected mean climate change for the corresponding variables and time frame. However, the phenological approach allows shifting time frames for crop stages. In this context, climate differences are obtained by subtracting current mean climate for the current time frame of a crop stage from the projected mean climate for its simulated time frame. This contrasts with models based on monthly variables that keep the time frame, and therefore the growing season, fixed.

Results from the baseline corn yield model, assuming a fixed growing season (table 1), reflect estimated average damages of 26.3 percent for the sample. Of course, these damages have substantially heterogeneous geographic distribution, which raises significant issues of potential geographic adaptation of crop mixes, but we leave that analysis to another paper. Interestingly, over two-thirds of this damage is associated with high temperature during the flowering period, which is a short period in the full growing season (approximately 2-3 weeks in the June-August period, depending on location). This sharp sensitivity during the flowering period coincides with agronomic findings, but contrasts sharply with econometric models that use season-long weather variables.

The resulting optimization problem is represented in figure 1. The objective is to assess whether relaxation of the freezing constraints in the spring and fall provide sufficient flexibility of planting dates for farmers to reduce exposure to extreme heat during the sensitive flowering period.

If farmers in the Midwest are constrained by the length of the frost-free period, then we should expect a spatial pattern of earlier planting dates coupled with earlier frost-free days in the spring to

Table 1: Yield Sensitivity by Corn Growth Stage

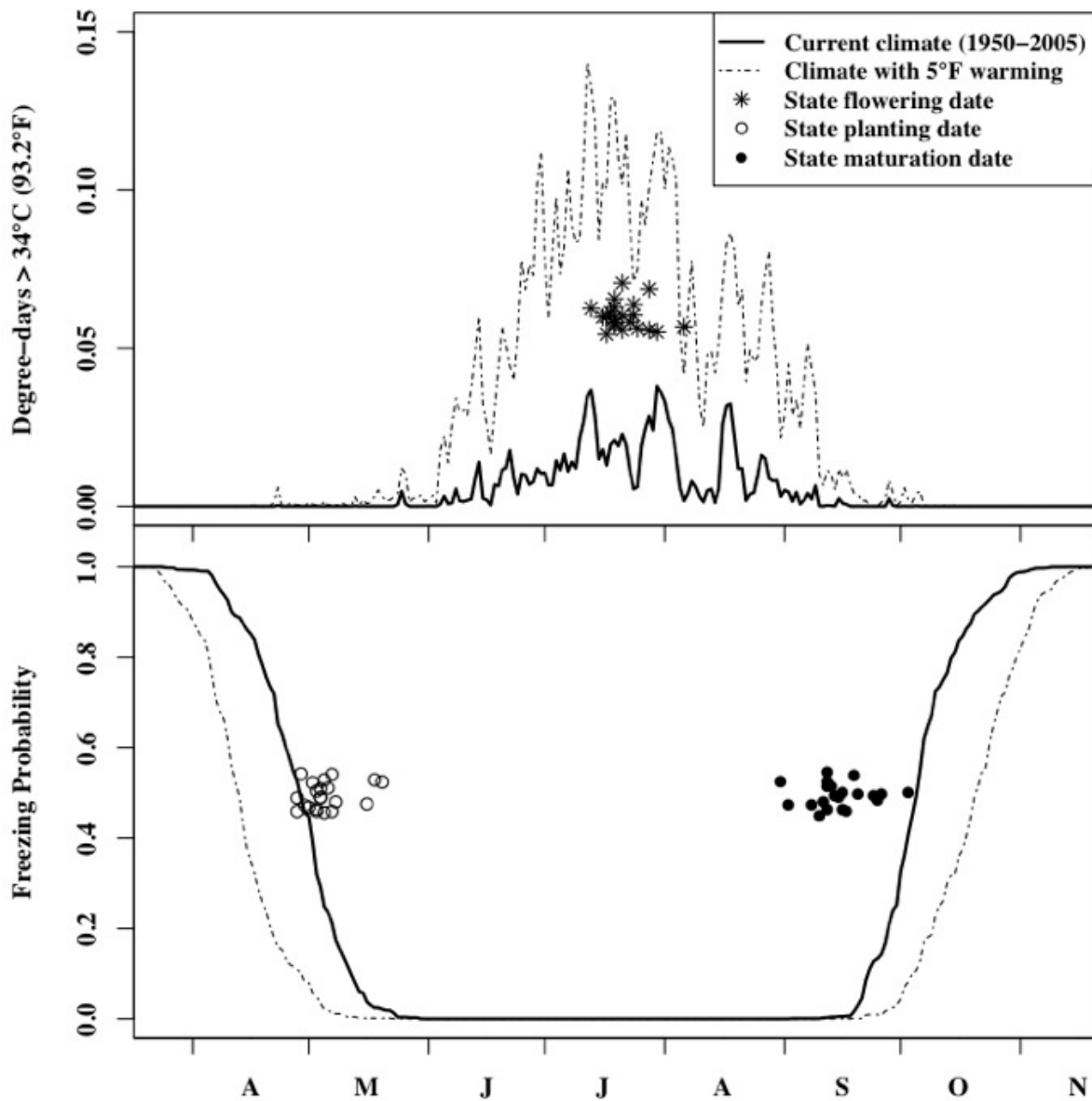| Corn growth stage	Yield Impact from a 5°F warming with a fixed growing season (%) / (bu/acre)	Share of stage influence in total yield impact (%)
Vegetative (Planting to flowering)	-8.9/ -11.7	33.3
Flowering (4 weeks around silking)	-17.6/-23.0	65.8
Grain-filling (Flowering to maturity)	0.3/0.3	0.9
Full cycle	-26.3/34.4	100

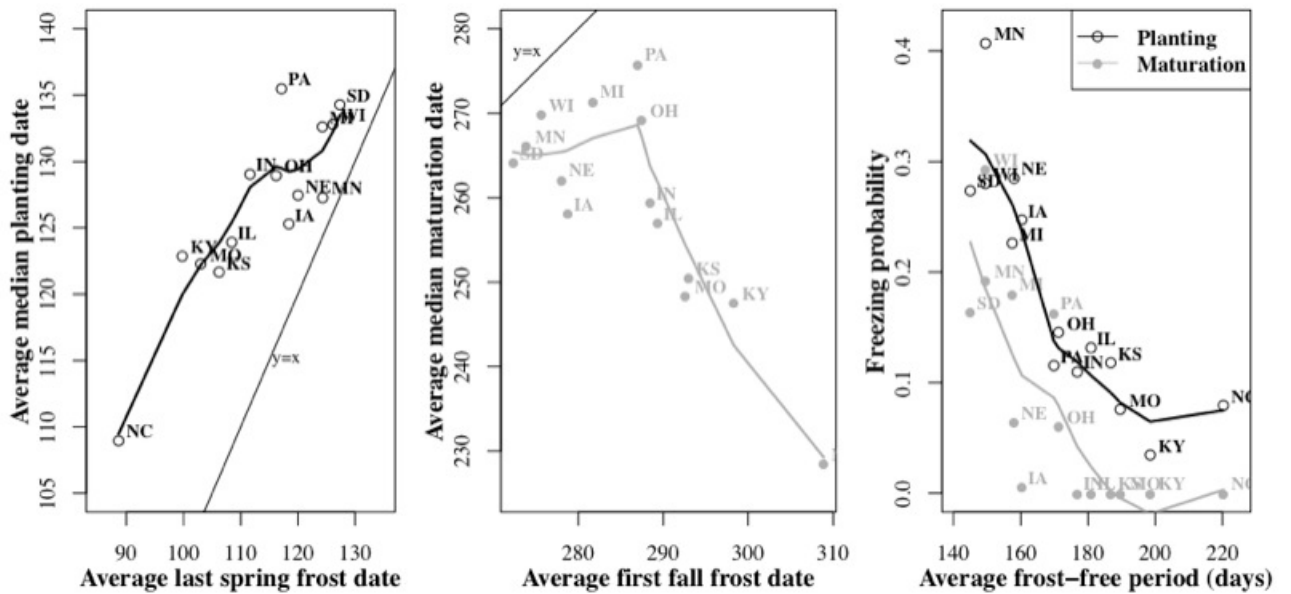Figure 1: Key corn stages and climatic constraints in Iowa

86

Figure 2: Corn growing season and freezing dates

emerge. This pattern is indeed verified in the first panel of figure 2. A similar pattern might also be expected in the fall, with later maturation dates associated with later first fall frosts if the fall frost date is constraining. The second panel of figure 2 shows this pattern up to a point (around day 290 of the year) after which there is a clear disconnect. Clearly, extending the frost-free period as depicted in the lower part of figure 2 shows that states with a narrower frost-free period tend to plant and reach maturity systematically at dates with higher probabilities of freezing. Only when the frost-free period reaches 180 days does the probability of frost at maturation decline to zero. These data suggest that the spring frost threshold is binding for all states, but the fall threshold is only binding for states with less than 180 frost-free days.

To explore the potential effect of shifting the growing season in the year by altering the planting date, we simulated earlier planting dates by shifting the planting date earlier in one-day increments until the planting date coincides with the new spring freezing threshold under a 5°F uniform warming. We also shifted the planting date later until the maturation date coincides with the new fall freezing threshold. At each increment, a new climate dataset was constructed for the time windows corresponding to the vegetative, flowering, and grain-filling periods. The spring and fall thresholds were set to maintain the current probability of freezing levels at planting and maturation for each state. This confines the simulation to a plausible range.

Simulation results are presented in figure 3 where each line represents an acreage-weighted state-
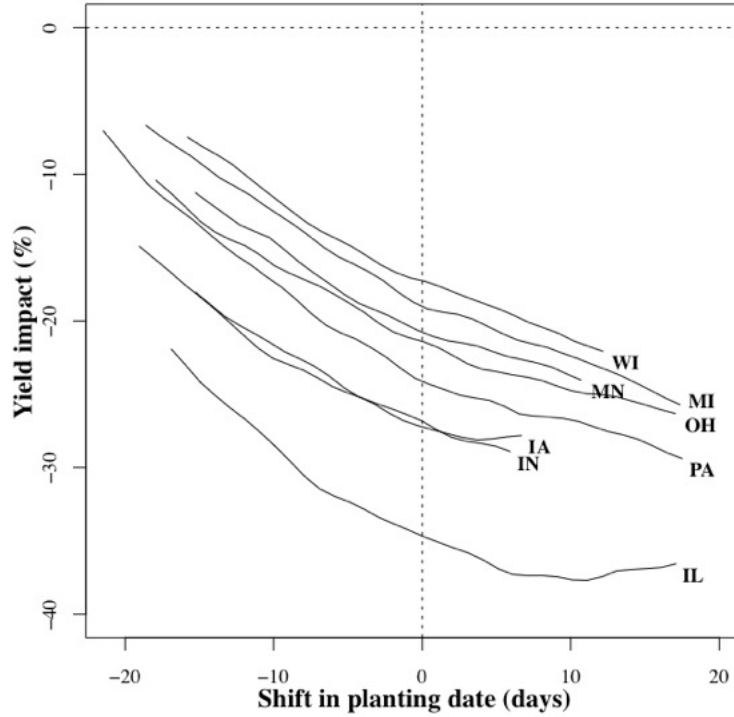
Figure 3: Overall state yield response and planting adaptation

level yield response to the shift in planting date. All states show considerable yield losses without planting adaptation as shown by the negative intercepts. The downward sloping curves, however, show that earlier planting under a 5°F warming scenario reduces damages from higher temperatures. This is the result from shifting the most sensitive period of the production cycle away from higher temperatures in the summer months. Table 2 shows that earlier planting by around 2 weeks results in a significant reduction of damages, ranging from 30 to 70 percent depending on the state. In terms of value, this represents around $3.4 billion for the 8 states combined, or 14 percent of the region's $24 billion annual average production for the 2000-2010 period. For comparison, SHF estimate $5.0 billion in annual damages for the entire US agricultural sector with the same warming scenario together with an 8 percent increase in precipitation; DG estimate $1.3 billion in annual benefits for an alternative scenario.

5   Conclusion

In this paper we propose, and provide evidence on, the need for a model that elucidates some of the major mechanisms through which both the damaging effects and adaptation possibilities from climate change impact agriculture. We submit that a transparent structural econometric approach can open

Table 2: Corn Yield Impacts From a Uniform 5°F Warming

	Without change in planting date (%)/ (bu/acre)	With change in planting date (%)/ (bu/acre)	Impact mitigated with adaptation (%)	Optimal change in planting date (days)	Savings from adaptation (Million 2010 US$)
Illinois	-34.7/-47.3	-21.9/-29.9	36.9	-16	$1,371
Indiana	-26.8/-35.3	-14.9/-19.6	44.4	-18	405
Iowa	-27.1/-37.2	-18.0/-24.7	33.6	-14	848
Michigan	-19.2/-21.6	-6.6/-7.5	65.3	-18	168
Minnesota	-20.6/-26.8	-11.2/-14.6	45.5	-14	330
Ohio	-21.4/-27.0	-10.4/-13.1	51.4	-17	116
Pennsylvania	-23.9/-24.7	-7.0/-7.2	70.6	-20	61
Wisconsin	-17.3/-20.8	-7.4/-8.9	56.8	-15	102
Full sample	-26.3/-34.4	-14.0/-18.5	44.1	-15.8	$3,401

the door to more detailed adaptation policy analysis grounded in revealed preferences. A structural approach grounded in the science of crop production should also allow cross-checking the plausibility of overall reduced-form estimates. Our empirical example shows that plausible adaptation strategies with little extra cost could significantly reduce projected corn yield damages for the 8 states in the sample. The results are demonstrated by a yield model that introduces disaggregated phenological weather variables matched to the production cycle.

Several limitations of our proposed approach should be borne in mind. Crop progress data is available for major producing states only at the state level, obscuring some variations within states. Also, we have not considered other agronomic aspects, such as the accelerating effect of higher temperature on the crop cycle, the potential to adopt different cultivars, or the influence of lower solar radiation on crop photosynthesis during shorter spring days. Our model also assumes three distinct crop stages where weather inputs are separable.

The complexity of the effect of environmental conditions on yield typically leaves researchers with a choice between imperfect proxies. Some of these may imply strong restrictions on farmer flexibility, as our example shows. Weather variables that better capture the effect of environmental conditions and their interaction, such as water balance measures tied to relevant crop stages, also offer new tools for more transparent methods of econometric climate change assessment that may help bridge the gap between alternative methods used for assessing impacts. Indeed, better capturing of the effect of weather variables on production allows better assessment of the physical constraints farmers face but, more importantly, facilitates assessment of the possibilities available for adaptation that have received relatively less attention to date.

References

Adams, Richard M., "Global Climate Change and Agriculture: An Economic Perspective," *American journal of agricultural economics*, 1989, *71* (5).

__ , Cynthia Rosenzweig, Robert M. Peart, Joe T. Ritchie, Bruce A. McCarl, J. David Glyer, R. Bruce Curry, James W. Jones, Kenneth J. Boote, and L. Hartwell Allen**, "Global climate change and US agriculture," *Nature*, May 1990, *345* (6272), 219–224.

Antle, John M., "Sequential Decision Making in Production Models," *American Journal of Agricultural Economics*, May 1983, *65* (2), 282–290.

Barnabás, Beáta, Katalin Jäger, and Attila Fehér, "The effect of drought and heat stress on reproductive processes in cereals," *Plant, Cell & Environment*, 2008, *31* (1), 11–38.

Blum, A., "Crop responses to drought and the interpretation of adaptation," *Plant Growth Regulation*, 1996, *20* (2), 135–148.

Collins, W. J., N. Bellouin, M. Doutriaux-Boucher, N. Gedney, T. Hinton, C. D. Jones, S. Liddicoat, G. Martin, F. O'Connor, and J. Rae, "Evaluation of the HadGEM2 model," *Hadley Cent. Tech. Note*, 2008, *74.*

Cosgrove, Brian A., Dag Lohmann, Kenneth E. Mitchell, Paul R. Houser, Eric F. Wood, John C. Schaake, Alan Robock, Curtis Marshall, Justin Sheffield, Qingyun Duan, Lifeng Luo, R. Wayne Higgins, Rachel T. Pinker, J. Dan Tarpley, and Jesse Meng, "Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project," *Journal of Geophysical Research*, October 2003, *108* (D22), 8842.

Deschênes, Olivier and Michael Greenstone, "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather," *The American Economic Review*, 2007, *97* (1), 354–385.

Dixon, B.L., S.E. Hollinger, P. Garcia, and V. Tirupattur, "Estimating corn yield response models to predict impacts of climate change," *Journal ofAgricultural and Resource Economics*, 1994, *19* (1), 58–68.

Easterling, William E., Mary S. McKenney, Norman J. Rosenberg, and Kathleen M. Lemon, "Simulations of crop response to climate change: effects with present technology and no

adjustments (the ['] dumb farmer' scenario)," *Agricultural and Forest Meteorology*, April 1992, *59* (1-2), 53–73.

Fageria, N. K., V. C. Baligar, Ralph B. Clark, and R. B. Clark, *Physiology of crop production*, Routledge, May 2006.

Fisher, A., M. Hanemann, M. Roberts, and W. Schlenker, "The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment," *American Economic Review*, 2012.

Hodges, J. A., "The Effect of Rainfall and Temperature on Corn Yields in Kansas," *Journal of Farm Economics*, April 1931, *13* (2), 305–318.

Hodges, T., *Predicting crop phenology*, CRC, 1991.

Hudson, I. L and M. R Keatley, *Phenological research: methods for environmental and climate change analysis*, Springer Verlag, 2009.

Intergovernmental Panel on Climate Change, *Climate Change 2007 - The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* 2007.

Just, Richard E. and Rulon D. Pope, "Production Function Estimation and Related Risk Considerations," *American Journal of Agricultural Economics*, May 1979, *61* (2), 276–284.

Kaiser, Harry M., Susan J. Riha, Daniel S. Wilks, David G. Rossiter, and Radha Sampath, "A Farm-Level Analysis of Economic and Agronomic Impacts of Gradual Climate Warming," *American Journal of Agricultural Economics*, May 1993, *75* (2), 387–398.

Kaufmann, Robert K. and Seth E. Snell, "A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants," *American Journal of Agricultural Economics*, February 1997, *79* (1), 178–190.

Lobell, David B and Marshall B Burke, "Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation," *Environmental Research Letters*, July 2008, *3* (3), 034007.

Lobell, David B. and Marshall B. Burke, "On the use of statistical models to predict crop yield responses to climate change," *Agricultural and Forest Meteorology*, October 2010, *150* (11), 1443–1452.

_ , **Marianne Bänziger, Cosmos Magorokosho, and Bindiganavile Vivek**, "Nonlinear heat effects on African maize as evidenced by historical yield trials," *Nature Climate Change*, April 2011, *1* (1).

Meerburg, B. G., A. Verhagen, R. E. E. Jongschaap, A. C. Franke, B. F. Schaap, T. A. Dueck, and A. van der Werf, "Do nonlinear temperature effects indicate severe damages to US crop yields under climate change?," *Proceedings of the National Academy of Sciences*, October 2009, *106* (43), E120–E120.

Mendelsohn, Robert O. and Ariel Dinar, *Climate change and agriculture: an economic analysis of global impacts, adaptation and distributional effects*, Edward Elgar Publishing, 2009.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw, "The Impact of Global Warming on Agriculture: A Ricardian Analysis," *The American Economic Review*, September 1994, *84* (4), 753–771.

Mesinger, Fedor, Geoff DiMego, Eugenia Kalnay, Kenneth Mitchell, Perry C. Shafran, Wesley Ebisuzaki, Dušan Jović, Jack Woollen, Eric Rogers, Ernesto H. Berbery, Michael B. Ek, Yun Fan, Robert Grumbine, Wayne Higgins, Hong Li, Ying Lin, Geoff Manikin, David Parrish, and Wei Shi, "North American Regional Reanalysis," *Bulletin of the American Meteorological Society*, March 2006, *87* (3), 343–360.

Mitchell, Kenneth E., Dag Lohmann, Paul R. Houser, Eric F. Wood, John C. Schaake, Alan Robock, Brian A. Cosgrove, Justin Sheffield, Qingyun Duan, Lifeng Luo, R. Wayne Higgins, Rachel T. Pinker, J. Dan Tarpley, Dennis P. Lettenmaier, Curtis H. Marshall, Jared K. Entin, Ming Pan, Wei Shi, Victor Koren, Jesse Meng, Bruce H. Ramsay, and Andrew A. Bailey, "The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system," *Journal of Geophysical Research*, 2004, *109* (D7), D07S90.

Mueller, Brigitte and Sonia I. Seneviratne, "Hot days induced by precipitation deficits at the global scale," *Proceedings of the National Academy of Sciences*, July 2012.

Mundlak, Yair and Assaf Razin, "On Multistage Multiproduct Production Functions," *American Journal of Agricultural Economics*, August 1971, *53* (3), 491–499.

NeSmith, D.S. and J.T. Ritchie, "Effects of soil water-deficits during tassel emergence on development and yield component of maize (Zea mays)," *Field Crops Research*, January 1992, *28* (3), 251–256.

Newman, J.E., "Climate Change Impacts on the Growing Season of the North American Corn Belt," *Biometeorology*, 1980, *7*, 128–142.

Ortiz-Bobea, Ariel, "Improving Agronomic Structure in Econometric Models of Climate Change Impacts," *Unpublished*, 2011.

_ , "Endogenizing the Growing Season in Climate Change Impact Studies.," *Unpublished*, 2012.

_ **and Richard E. Just**, "Modeling the Structure of Adaptation in Climate Change Impact Assessment," *American Journal of Agricultural Economics*, 2013, *95*, 244–251.

Roberts, Michael J. and Wolfram Schlenker, "The Evolution of Heat Tolerance of Corn: Implications for Climate Change," in "the Economics of Climate Change: Adaptations Past and Present," University of Chicago Press, 2011, p. 225–251.

Rosenzweig, Cynthia and Martin L. Parry, "Potential impact of climate change on world food supply," *Nature*, 1994, *367*, 133–138.

Schaake, John C., Qingyun Duan, Victor Koren, Kenneth E. Mitchell, Paul R. Houser, Eric F. Wood, Alan Robock, Dennis P. Lettenmaier, Dag Lohmann, Brian Cosgrove, Justin Sheffield, Lifeng Luo, R. Wayne Higgins, Rachel T. Pinker, and J. Dan Tarpley, "An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS)," *Journal of Geophysical Research*, January 2004, *109* (D1), D01S90.

Schlenker, Wolfram and Michael J. Roberts, "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change," *Proceedings of the National Academy of Sciences of the United States of America*, September 2009, *106* (37), 15594–15598.

_ **and** _ , "Reply to Meerburg et al.: Growing areas in Brazil and the United States with similar exposure to extreme heat have similar yields," *Proceedings of the National Academy of Sciences*, October 2009, *106* (43), E121–E121.

_ **, W. Michael Hanemann, and Anthony C. Fisher**, "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach," *The American Economic Review*, March 2005, *95* (1), 395–406.

_ , _ , and _ , "The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions," *Review of Economics and Statistics*, February 2006, *88* (1), 113–125.

Smith, D. L and C. Hamel, *Crop yield: physiology and processes*, Springer, 1999.

Wallace, H. A., "Mathematical inquiry into the effect of weather on corn yield in the eight corn belt states," *Monthly Weather Review*, 1920, *48*, 439.