Non-linear Estimation of the Impacts of Climate Change: the Case of Corn and Soybeans Crop Yields in Brazil

André Corsini Leão

Advisor: Wolfram Schlenker

New York, May of 2008

Abstract

The objective of this paper is to estimate the relationship between weather and yields of two major agricultural cultures in Brazil: corn and soybeans. Based on the methodology of Schlenker and Roberts (2006), it was found that an increase of 5^{o}F in the yearly average temperatures is associated with a decrease of 12.8% for corn yields and 9.3% for soybeans yields. Another important finding is that, similar to the USA, an asymmetric relationship between weather and crop yields was found: after a certain threshold (between 91^{o} F and 96^{o} F for corn and above 96^{o} F for soybeans), temperatures start to be increasingly harmful to yields.

Keywords: Climate change, Non-linear temperature effects, Brazilian agriculture, Soybeans production, Corn production.

TABLE OF CONTENTS:

TABLE OF FIGURES:

1 Introduction

The main focus of this thesis is to estimate the relationship between temperature and the yields of corn and soybeans production in Brazil. Both are amongst the country's main agricultural products, being of central importance not only to the eating habits of Brazilians, but to the whole economy.

It is straightforward to imagine that agricultural production and temperatures are correlated. A fixed-effects model, which accounts for the relevant differences in municipalities, would correctly capture their relationship if correctly specified. To refrain from making wrong assumptions about the functional form of the relationship between yields and temperature, a non-parametric procedure will be used, so different possible functional forms can be captured. As put in Schlenker and Roberts (2006): "accurate estimation of non-linear effects is particularly important when predicting effects on non-marginal changes in temperatures, as is expected under climate change."[1]

It is expected that after a certain critical threshold, higher temperatures become increasingly harmful to corn and soybeans yields in Brazil. This implies that "if climate change shifts the temperature distribution above such a threshold, the impacts would be significant."[2]

Climate change

Global warming is being actively discussed nowadays. It has been getting the attention of professionals from different areas, from economists, to scientists. In the scientific front, consensus seems far from being achieved, as there are currents that believe that the climate

[1] **SCHLENKER, Wolfram and ROBERTS, Michael J.**, "Estimating the impact of climate change on crop yields: the importance of non-linear temperature effects", p.1
[2] Schlenker and Roberts (2006), p.2

change which is being observed is caused by human activity, while others believe this is part of a natural cycle, where human participation is not significant. In the meantime, economists still argue about the magnitude of the economic impacts of climate change. But in one aspect, there seems to be large reliable empirical evidence: "Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice, and rising global average sea level"[3].

Leaving the arguments aside, one thing can be taken for sure: climate change implies in socio-economic impacts. Taking this into consideration, the estimation of its costs is needed in order for the economy to respond optimally.

Sustainable development

Brazil is a developing country, where one of the most important political goals is promoting growth and development. At the same time, attaining it in a sustainable way is of primary importance, as resources are scarce and more economic activity can mean more pollution and health damages to the population.

Economic theory indicates that for the global social utility maximum to be achieved it is required that "the marginal benefit (…) be set equal to its marginal cost"[4]. Then, it will be guaranteed that the correct economic incentives are being put in place.

This is the central point of cost-benefit analysis. One argument in favor of doing it is the following: "Because society has limited resources to spend on regulation, benefit-cost analysis can help illuminate the trade-offs involved in making different kinds of social investments. In

[3] IPCC (2007) – Summary for Policymakers – p.1
[4] **FREEMAN III, A. Myrick**, "Environmental Policy Since Earth Day I: What Have We Gained?", *Journal of Economic Perspectives*, Winter 2002, *16*, 126

this regard, it seems almost irresponsible to not conduct such analyses, because they can inform decisions about how scarce resources can be put to the greatest social good"[5].

The exercise of estimating costs and benefits of environmental regulations has been done extensively in the USA and other countries, but to the author's knowledge, not in Brazil. This is one of this work's main motivations, although it is too broad and complex of an endeavor at the current stage. The focus of this thesis is to estimate the impacts of climate change on crop yields, which is one component of the societal costs.

---

[5] **ARROW, Kenneth J. et al**, "Is There a Role for Benefit-Cost Analysis in Environmental, Health, and Safety Regulation?", *Science*, April 12 1996, *272*, 221

2 Literature review

Schlenker and Roberts (2006) estimate a decrease on the yields of three major crops (corn, soybeans and cotton) in the United States by 25%-44% under the slowest warming scenario and 60%-79% under the most rapid by the end of this century.[6]

As stated in their work: "the key contribution of our study is in identifying a highly non-linear and asymmetric relationship between temperature and yields"[7]. Moreover, "the stability of the estimated relationship across regions, crops and time suggests it may be transferable to other crops and countries". This was a main stimulation for transplanting their model to Brazil.[8]

Schlenker and Roberts (2006) also argue that "there is some consensus that warming will likely be harmful to agriculture in tropic and sub-tropic zones". This means that Brazil could be considerably affected.

Another important finding of their paper is that the potential for adaptation to climate change is somewhat low. The argument is that in their dataset of 55 years, there was "no evidence that crops have become significantly better at withstanding extreme heat above the upper threshold. Moreover, hotter southern states exhibit the same threshold as cooler states in the north, suggesting that there is limited potential for adaptations".

One important difference between the United States and Brazil is the fact that the latter is mostly located at the tropical region, above the Tropic of Capricorn. However, its corn and soybeans cultures are mostly concentrated in the Southern, Southeast and Center-West regions of

---

[6] These results do not take into account the possibility that farmers start new businesses because climate change has yielded good conditions and/or close businesses because they predict losses. The estimates ideally work for an assumption that the same farmers continue producing.
[7] Schlenker and Roberts (2006), abstract
[8] Schelnker and Roberts (2006) go through a more detailed explanation of how the non-linear relationship was defined. At this point, what is important is that this paper goes in the same direction of estimating non-linear impacts under the belief that the true underlying relationship has this form.

the country, which are closer to the tropic[9]. In a rough direct observation from the temperature predictions of the IPCC[10] for North-America and South America, it seems that the most productive areas in Brazil will not be affected as much as in the USA.

[9] Map of the agricultural production of Brazil (IBGE, 2006) <ftp://geoftp.ibge.gov.br/mapas/tematicos/mapas_murais/prod_agricola_e_veget_natural.pdf>
[10] Regional climate projections <http://www.ipcc.ch/graphics/gr-ar4-wg1.htm>

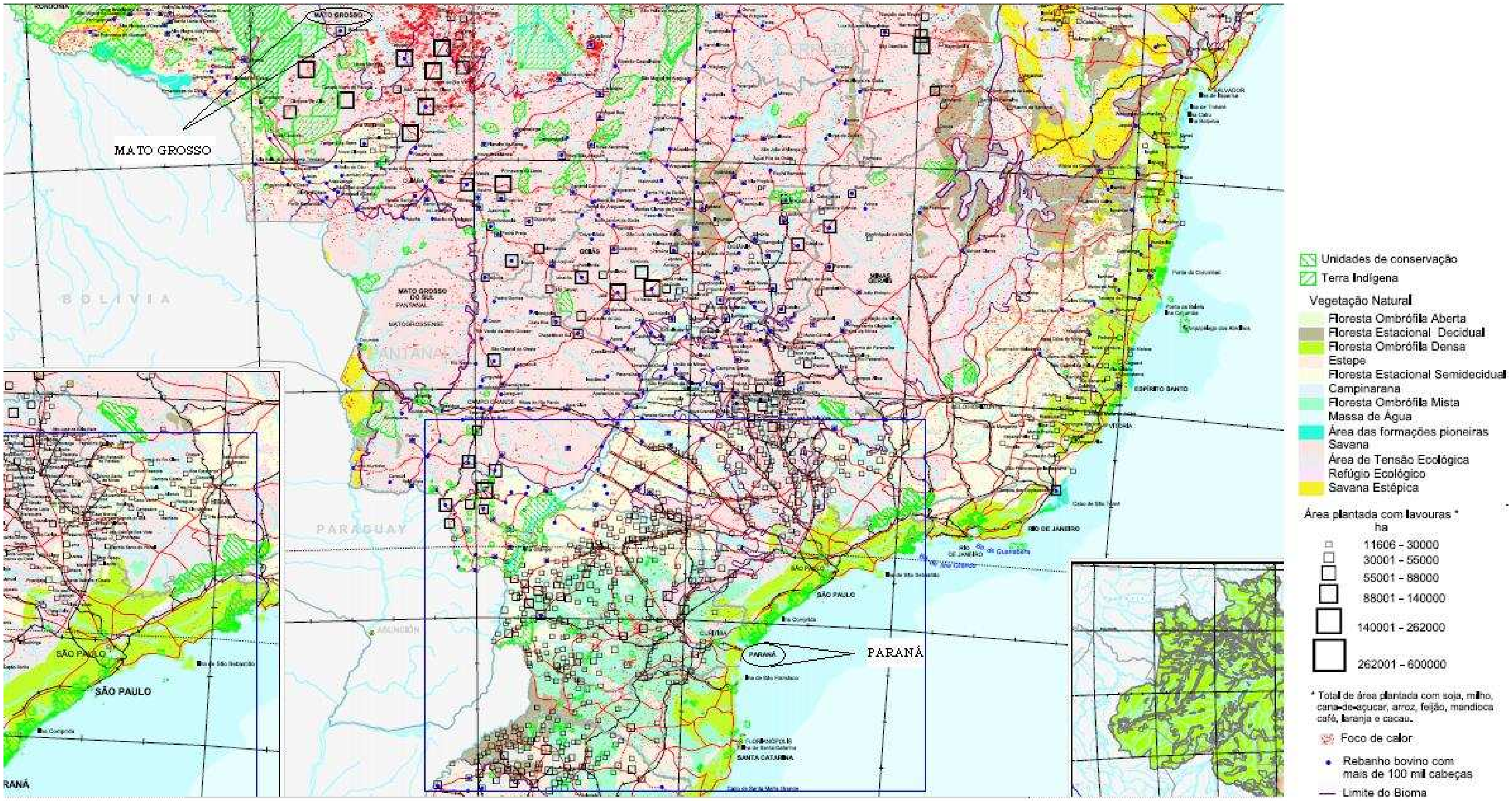


Figure 1 – Agricultural Map of Brazil: This map focuses on the most representative states for soybeans production – Mato Grosso, and for corn production – Paraná. The black squares represent the areas occupied by agriculture (in hectares). It is also noticeable the difference between the vegetation and climate on each state – Mato Grosso is dominated by savannas and seasonal tropical climate, with arid winters and humid summers, whereas Paraná has mixed forests of *Araucaria* and humid subtropical climate.

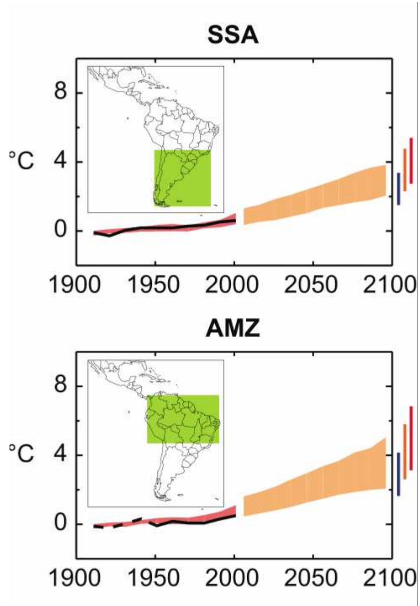


Figure 2.1 - Climate Change: These pictures indicate that climate change will be slightly stronger over the AMZ area, which contains Mato Grosso than in the SSA area, which contains Paraná.

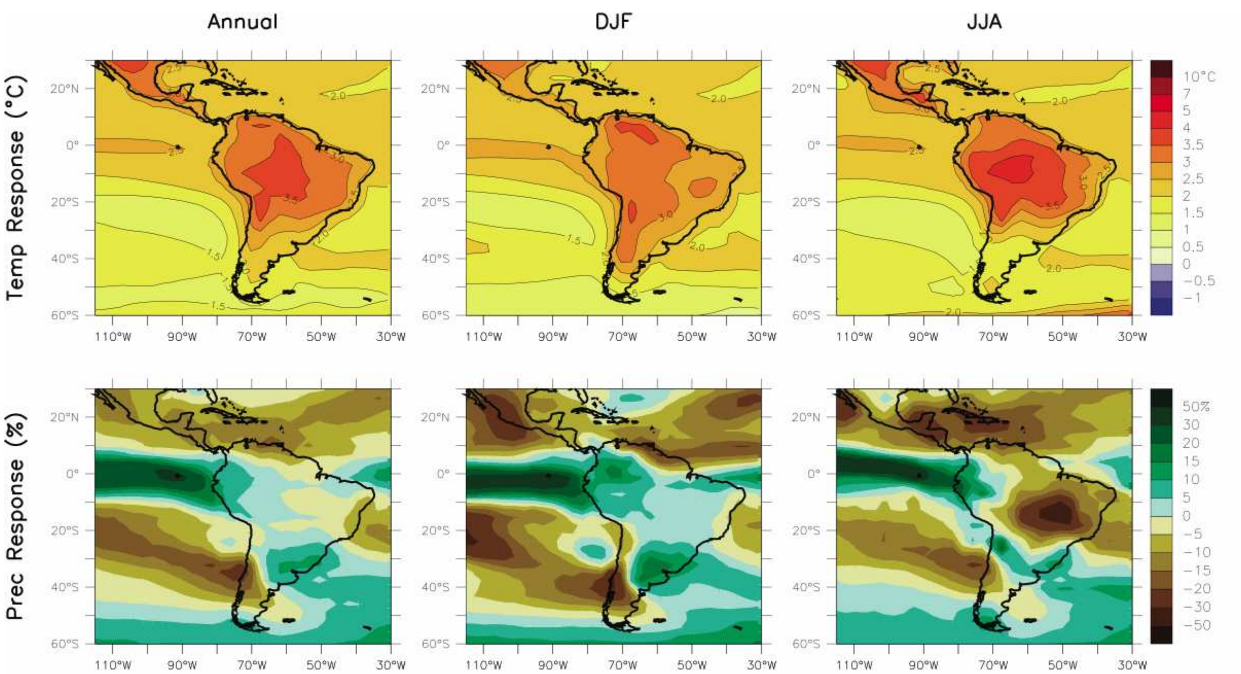


Figure 2.2 – Climate Change: Again, it is depicted that the state of Paraná should suffer less from the impacts of global warming than Mato Grosso. These are predictions of annual mean, DJF and JJA temperature change between 1980 to 1999 and 2080 to 2099, averaged over 21 models. Second row: same as top, but for fractional change in precipitation (DJF and JJA are two specific procedures of measurement).

The Summary for Policymakers of the IPCC states that "in temperate zones soybean yields are projected to increase"[11] when it refers to Latin America. So, it is possible that in the long term, there will be no significant loss on the soybeans culture in Brazil, as losses in the Center-West (that has tropical climate) can possibly be matched by gains in the Southern region (temperate). There were no specific statements for corn production.

The Summary for Policymakers of the IPCC also takes into consideration the possibilities for adaptation to climate change. In the agriculture sector, "adjustment of planting dates and crop variety; crop relocation; improved land management, e.g. erosion control and soil protection through tree planting"[12] are suggested as possible strategies to be adopted. The policy role of promoting R&D and the technological constraint to adaptation are also underlined.

Taking into consideration the mechanism used on the focused exercise of estimating the impact of climate change in crop yields in the USA and the broader regional perspective given by the IPCC, this work intends to elucidate the relationship between weather and corn and soybeans yields in Brazil in an attempt to determine the likely impacts that warming may have.

Spatial distribution of corn and soybeans production

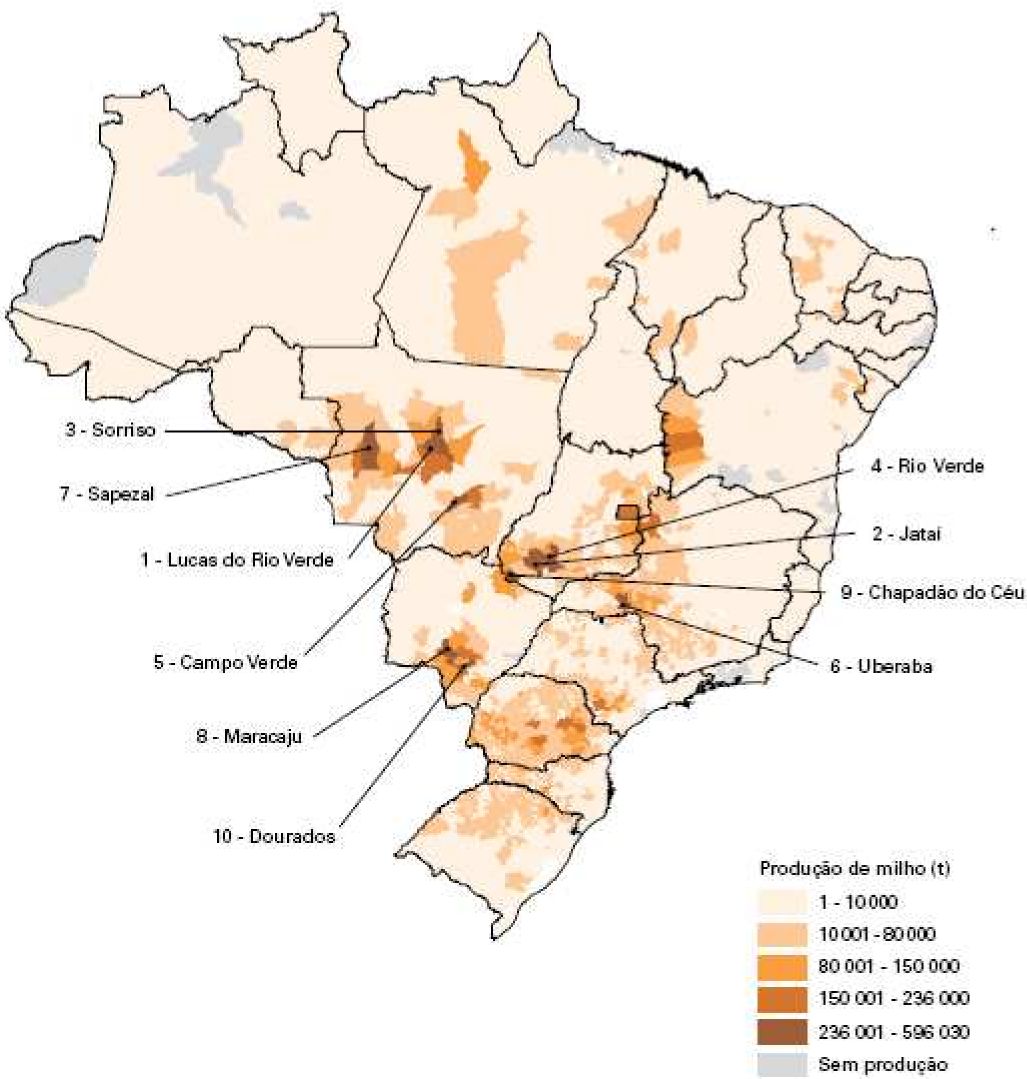
Every year, IBGE[13] publishes the report *Produção Agrícola Municipal* (Municipal Agricultural Production).

In 2006[14], it stated that the region with the largest corn production was the South in the first crop and the Center-West for the second crop[15]. Moreover, corn was found to be the culture

[11] IPCC (2007) – Summary for Policymakers – p.10
[12] IPCC (2007) – Summary for Policymakers – p.15
[13] Brazilian government's statistical office <www.ibge.gov.br>
[14] **IBGE,** *"Produção Agrícola Municipal 2006 – Comentário"* – p.18-23 <http://www.ibge.gov.br/home/estatistica/economia/pam/2006/comentario.pdf>
[15] This is an important issue for the determination of a specific growing season for the whole country, as will be seen later in this work. Ideally, the analysis would have to be done separately for regions with different growing seasons.

most widely distributed around the country. The states of Paraná, Minas Gerais, Rio Grande do Sul, São Paulo, Mato Grosso and Goiás were the six largest producers.

Figure 3: Corn production distribution, top 10 municipalities - 2006[16]



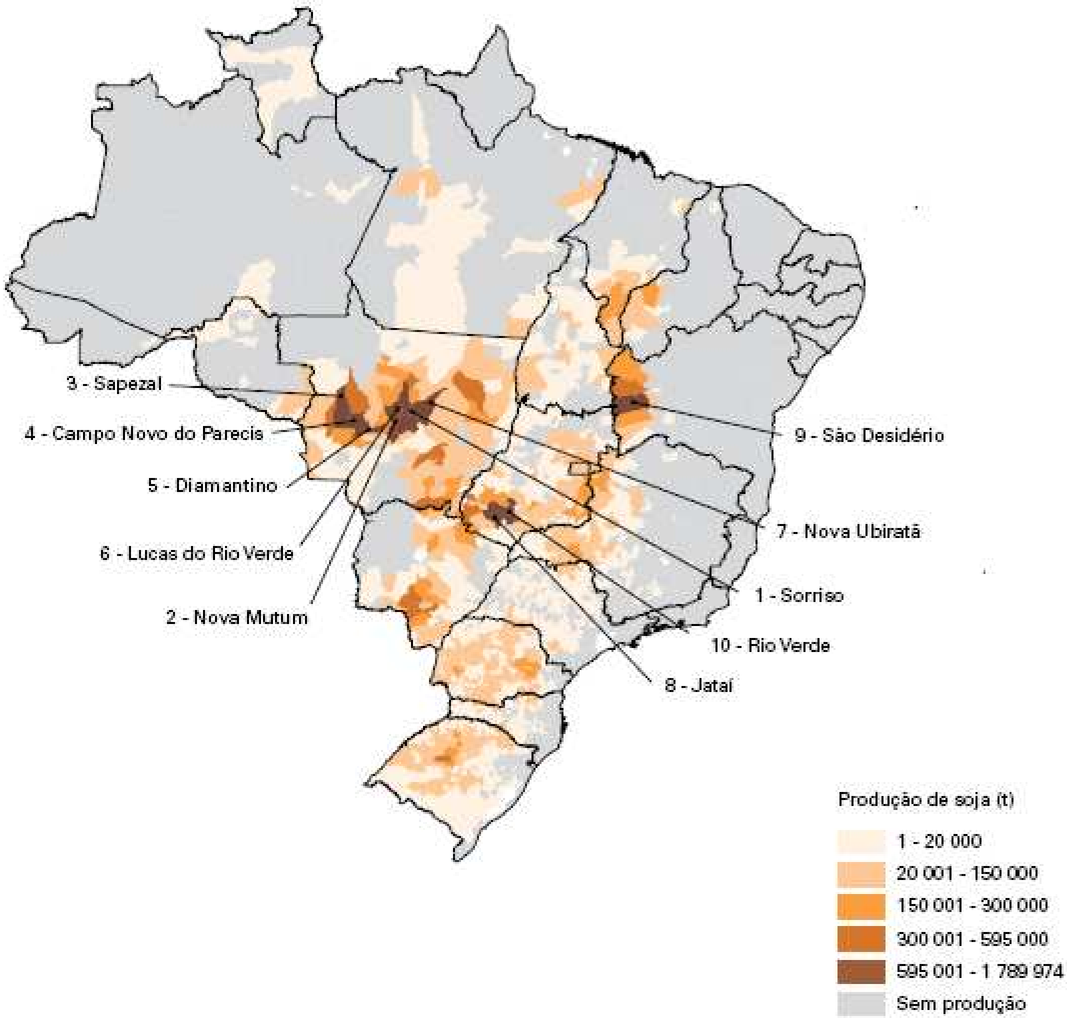
Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal 2006.

Now considering soybean production, in 2006, it is clear on the picture below that it is not nearly as widely distributed throughout the country as corn. The six largest producers were

[16] Grey areas in the map mean no production. Numbers in the legend are production in tons.

the states of Mato Grosso, Paraná, Rio Grande do Sul, Goiás, Mato Grosso do Sul and Minas Gerais.

Figure 3.2: Soybeans production distribution, top 10 municipalities - 2006



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Agropecuária, Produção Agrícola Municipal 2006.

3 Conceptualization

To estimate the effects of climate change on crop yields, the following framework is implemented:

- Harvests are directly affected by heat and humidity. Thus, good proxies in this case would be the temperatures and precipitation measured at the same place as the crop[17].
- With a database of past temperatures, precipitation levels and crop yields (crop yield=production/area harvested), one can study the relationship between the first two variables and the latter, estimating the impacts of an exogenous increase in temperature, which is exactly what climate change represents in this case.

To be able to estimate a precise model, municipal data will be used. IBGE will be the source for all agricultural data in the municipal level. It also offers us GIS data that will allow us to pair crop yields to weather data.

The dependent variable: crop yields

Data for crop yields was obtained on SIDRA[18] (IBGE's system of automatic recovery of data). The data available is production in tons and area in acres. It is collected on the LSPA survey (Systematic Survey of Agricultural Production), which is monthly, but only yearly data from 1990 to 2006 are publicly available[19].

[17] In cases where the data was not available for the specific municipality, the closest weather station was used. Distances were minimized using GIS software.
[18] <www.sidra.ibge.gov.br>
[19] Only yearly data makes sense to be used, since the time unit of analysis is the crop.

The author has transformed production data from tons to bushels and area from acres to hectares, and then divided the former by the latter to obtain the crop yields for each year in each municipality.

The independent variables: time spent at each 5 degree temperature intervals and precipitation levels

All the weather data was obtained through the National Climatic Data Center website[20]. The publicly available data for download goes from 1973 to 2005. 445 weather stations were considered[21], where each one is identified by latitude and longitude coordinates.

The datasets had the following daily variables: minimum temperature, maximum temperature (in degrees Fahrenheit) and level of precipitation (in inches).

To construct the yearly temperature variable that was used in this paper's models, the time spent at each 5 degree temperature interval, it was assumed that temperatures grow from their minimum to their maximum following the path of a cosine curve, where the argument of the function is the time spent at each degree Fahrenheit. This way, the unit of time being considered is a percentage of the day summed up over the growing season.

5 degree temperature intervals were used in order to increase the number of observations, which would be too little if the time spent was calculated for each degree of temperature, and reduce the number of independent variables (which logically would be too large if every degree of temperature was taken into account). With the daily time spent on each interval in hand, this data was summed up over the growing season, resulting in the independent variable used.[22]

---

[20] <http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html>

[21] Some of these stations were outside the Brazilian territory. However, they had to be considered because they were the closest stations to some of the municipalities in Brazil that will be analyzed in this study.

[22] When summing up the times spent at each degree over the entire growing season, there was the need to deal with the issue of the varying number of observations each month. Even if it is assumed that this variation is random,

Observing the distribution of maximum and minimum temperatures, cutoff points were established on 101°F for maximum temperatures and on 36°F for minimum temperatures. All the observations below 36°F were dropped and the ones above 101°F were lumped[23] into the highest temperature interval. Since the low temperatures were dropped, the temperature independent variables are relative values.[24]

The other weather variable used was total precipitation over the growing season. This one did not require all the transformations as the temperature component; it is a simple rescaled (see footnote 22) sum of the precipitation level over the growing season.

---

months with less observations would be taken as colder. For this reason, the observations had to be rescaled (time*obsinmonth/daysinmonth). The most precise description of the independent variable would be: extrapolated amount of time spent at 5 degree temperature intervals assuming that the observations in the given month are representative of the whole month.

[23] Dropped and lumped values represent less than 2% of the data available.

[24] For example, the variable time spent at the 91°F -96°F interval is actually time spent at the 91°F -96°F interval in comparison to time spent below 36°F.

4 Research design

This thesis uses available historical data in order to estimate the relationship between crop yields, temperatures and precipitation. This way, the impact of a hypothetical weather scenario on yields will be predicted[25].

All the data publicly available (there is no sampling) is taken into account[26]. This way, the analysis is limited to municipalities that have data for both variables. To the author's knowledge, there is no systematic reason for differences in the number of municipalities covered by each source.

A non-linear model

In order to estimate the impacts of climate change on corn and soybeans in Brazil, a non-linear model will be utilized. Previous research has shown the importance of non-linearity in the estimation of the impact of temperature on crop yields in the United States[27]. This seems to be reasonable. Heat is logically essential to agriculture, but it is not hard to imagine that extreme temperatures, either too high or too low, may be associated with low outcomes.

From this observation we arrive to the conclusion that at some point, the increase in temperatures will present positive returns to crop yields, but after a certain threshold, it will show negative ones. But then there rises a new issue: which is the path that it is likely to follow?

---

[25] A good source for possible temperature scenarios would be IPCC's AR4 report, released in 2007 <http://www.ipcc.ch/>.

[26] Sensitivity checks will be run, where the models are tested only for the most relevant states for corn and soybeans production. This is done to reduce the amount of noise induced by the very large amount of missing values in the dataset. Also, because growing seasons vary from different regions.

[27] Schlenker and Roberts (2006)

When using standard linear regression, the statistician is arbitrarily assuming a functional form for the relationship between the dependent and independent variables being analyzed. This is a limitation. Ideally, one would like to avoid making unnecessary assumptions.

Fortunately, in this present work, it was not necessary to assume a functional form for the influence of temperature on crop yields. It will not be needed to know the path that it follows. A non-parametric procedure was used to estimate the link between temperatures and yields. Considering the amount of time spent at each five degree interval during the growing season of each crop as the set of temperature regressors, virtually any possible relationship between yield and temperature is being accounted for.

Unfortunately, this work is far from being free of assumptions. Only daily minimum and maximum temperatures were available. To be able to generate the temperature interval variables, it was assumed that temperatures grow from its minimum to its maximum, following the path of an inverse cosine curve. It is also assumed that the effect of temperature is perfectly substitutable over the growing season.

After regressing the crop yields on temperatures, average precipitation during the growing season was then included in the model (with a positive expected relationship), with a quadratic component to penalize for excessive precipitation (e.g. floods).

It is also worth thinking about whether temperatures influence the level of production or the percent change of the level of production. The argument that the yield level is a function of temperature and other variables is defensible, but in this study we are going to use the natural logarithm of the yields as our dependent variable because temperature can have a different effect in the level of the yield for each specific soil quality, for example, but arguably the same effect in the percent change for different types of terrain.

Considerations about validity – usage of fixed-effects and a time trend

Brazil occupies a very large territory, encompassing areas with diverse climates. This makes some areas more adequate for agriculture than others. This is one possible factor that may influence the production in different municipalities. In fact, there is a large amount of variables that could be relevant to explain differences in production amongst different locations. However, these are not the focus of this work.

Unfortunately, if any source of time-invariant heterogeneity is not accounted for, the results will be biased. This is why the models will be estimated using fixed-effects for each municipality.

Supposedly, there is also the problem of temporal dependence. A time trend (year and year squared) will be added to the regression in order to control for that. This way, the effect of farmers reacting to past weather events is captured. Adaptation and possible technological changes that happened were taken into account[28].

Another possible threat would be that of heteroskedasticity in the errors, as the municipalities may have systematic differences in the measurement of their production. This will not be controlled, under the assumption that these differences are randomly distributed throughout the country and over time and that IBGE has already normalized the data.

However, two threats to validity persist:

1) As the different climate change scenarios get more distinct from past events, the probability of new responses increases;

[28] A possibly better way of controlling for time dependency is using an auto-regressive model. However, this thesis will stick to the more simplistic approach of just adding the time trend. It is believed that the gains of adding auto-regressive components would be small.

2) Farmers may get out of the business because of overly severe conditions and at the same time new ones may enter in different municipalities that, because of climate change, become better-suited for agriculture – the model does not capture this accordingly.

These threats compromise the generalizability of the future predictions of this work. Up to this point, there are no suggestions of how to control for these factors, but it is assumed that the conclusions for the most conservative scenarios (and closest future) still hold, as they may not be sufficiently strong to significantly draw farmers out of business and/or bring new ones in to business.

5 Results

Determining the growing season

One of the challenges of this thesis was to define a specific growing season for each culture. It would not be realistic to take into consideration all the weather data points for temperatures and precipitation throughout the year, since the crops are only influenced by them from planting to harvest. However, different producers start their crop at different dates, possibly because of regional differences in the climate. It is even possible that while one is already harvesting, the other is in the first stages of seeding.

The ideal way of controlling this would be defining different growing seasons for each area. Unfortunately this was not possible due to limitations in the data and lack of information about growing seasons on the different regions of the country.

The solution found was to run the regressions with different reasonable growing seasons and choose the best fit according to the Adjusted R-Squareds.

In the following, two tables are presented where the preferred model[29] for each culture was used for selected reasonable[30] growing seasons:

[29] The preferred model is determined in the Nested models section of this chapter.
[30] It was not possible to find a definitive growing season from a reliable source. The selected periods are an arbitrary choice of the author, based on unofficial information from diverse websites.

Table 1– Estimates for corn and soybeans – different growing seasons: [31]

CORN

Temperature Interval	Sep-Feb	Sep-Mar	Sep-Apr	Oct-Feb	Oct-Mar	Oct-Apr	Nov-Feb	Nov-Mar	Nov-Apr
	Growing Season / Coefficients								
36-41	0.023	0.014	0.011	0.038 *	0.022	0.016	0.060 **	0.030	0.019
41-46	-0.014	-0.006	-0.014	-0.045 **	-0.031 **	-0.027 **	-0.087 **	-0.041 *	-0.030 **
46-51	0.000	-0.003	0.001	0.012	0.004	0.008	0.027	0.014	0.017 *
51-56	-0.019 **	-0.018 **	-0.012 **	-0.016 *	-0.008	-0.004	-0.017	-0.014	-0.005
56-61	0.006 **	0.007 **	0.005 **	0.003	0.000	-0.002	0.011 *	0.006	0.002
61-66	-0.001	-0.002 *	-0.002 *	-0.002	-0.002	-0.001	-0.005 **	-0.005 **	-0.003 **
66-71	-0.002 **	-0.002 **	-0.002 **	-0.003 **	-0.003 **	-0.003 **	-0.003 **	-0.003 **	-0.003 **
71-76	0.001	0.001	0.001	0.002 **	0.002 **	0.002 **	0.002	0.002 **	0.002 **
76-81	0.007 **	0.007 **	0.007 **	0.007 **	0.007 **	0.007 **	0.008 **	0.007 **	0.007 **
81-86	-0.012 **	-0.011 **	-0.010 **	-0.013 **	-0.012 **	-0.010 **	-0.014 **	-0.012 **	-0.011 **
86-91	0.003 **	0.003 **	0.003 **	0.003 **	0.003 **	0.003 **	0.003 **	0.003 **	0.003 **
91-96	-0.002	-0.003 **	-0.003 **	-0.002	-0.003 *	-0.003 *	-0.001	-0.002	-0.002
96-114	-0.012 **	-0.013 **	-0.014 **	-0.014 **	-0.018 **	-0.019 **	-0.030 **	-0.035 **	-0.035 **
Other Variables									
Year	4.773 **	5.562 **	6.231 **	4.503 **	6.028 **	5.713 **	3.590 **	4.725 **	4.353 **
Year Squared	-0.001 **	-0.001 **	-0.002 **	-0.001 **	-0.002 **	-0.001 **	-0.001 **	-0.001 **	-0.001 **
Precipitation	0.001	0.001 *	0.002 **	0.002 **	0.002 **	0.003 **	0.002 **	0.003 **	0.003 **
Constant	-4784.875 **	-5572.619 **	-6240.619 **	-4512.289 **	-6036.536 **	-5720.653 **	-3602.031 **	-4736.127 **	-4363.661 **
R-squared	**0.803**	**0.800**	0.798	0.799	0.798	0.796	**0.800**	0.799	0.798
Adj R-squared	**0.753**	**0.753**	0.750	0.749	0.750	0.748	**0.753**	0.751	0.749
Number of Observations	15225	16202	16694	13619	14750	15334	13164	14342	14938

SOYBEANS

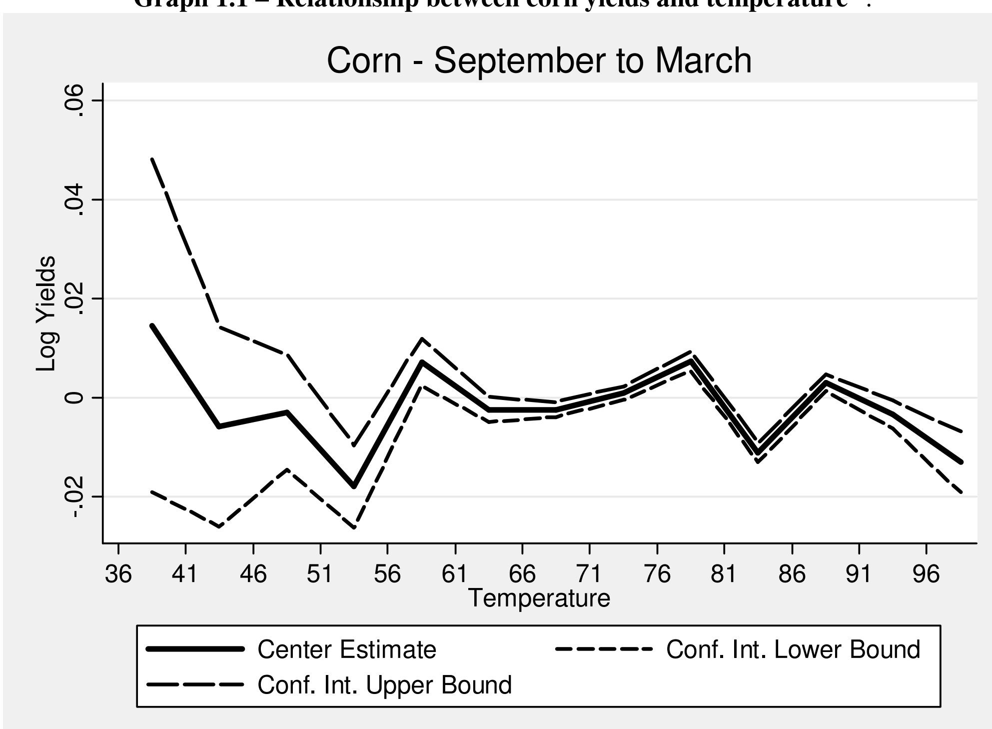
Temperature Interval	Sep-Feb	Sep-Mar	Sep-Apr	Oct-Feb	Oct-Mar	Oct-Apr	Nov-Feb	Nov-Mar	Nov-Apr
	Growing Season / Coefficients								
36-41	0.029 *	-0.034 **	-0.045 **	0.073 **	-0.015	-0.034 **	0.108 *	-0.033 *	-0.060 **
41-46	-0.013	-0.014	0.008	-0.065 **	-0.048 **	-0.005	-0.149	-0.043 **	0.023
46-51	-0.009	-0.008	-0.009 **	0.009	-0.001	-0.005	0.003	-0.004	-0.012 *
51-56	-0.002	-0.002	-0.012 **	-0.019 *	-0.008	-0.019 **	-0.009	-0.001	-0.020 **
56-61	-0.004 *	-0.005 **	0.004	0.010 **	0.004	0.007 **	0.029 **	0.017 **	0.016 **
61-66	-0.006 **	-0.004 **	-0.005 **	-0.014 **	-0.009 **	-0.007 **	-0.019 **	-0.013 **	-0.010 **
66-71	-0.002 **	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	-0.001
71-76	0.001	-0.002 *	-0.002 *	0.001	-0.001	-0.001	0.002 *	-0.001	-0.001
76-81	0.006 **	0.007 **	0.006 **	0.007 **	0.008 **	0.007 **	0.004 **	0.007 **	0.006 **
81-86	-0.004 **	-0.004 **	-0.005 **	-0.007 **	-0.006 **	-0.005 **	-0.005 **	-0.005 **	-0.005 **
86-91	0.003 *	0.000	-0.002	0.003 *	-0.001	-0.003 **	0.000	-0.003 **	-0.004 **
91-96	-0.007 **	-0.002	0.002	-0.010 **	0.000	0.003 *	-0.008 **	0.002	0.004 **
96-114	-0.001	-0.007 *	-0.009 **	-0.001	-0.016 **	-0.018 **	-0.004	-0.024 **	-0.022 **
Other Variables									
Year	14.780 **	15.960 **	18.576 **	15.895 **	15.472 **	19.051 **	15.238 **	15.089 **	18.544 **
Year Squared	-0.004 **	-0.004 **	-0.005 **	-0.004 **	-0.004 **	-0.005 **	-0.004 **	-0.004 **	-0.005 **
Precipitation	0.015 **	0.014 **	0.017 **	0.015 **	0.013 **	0.016 **	0.024 **	0.018 **	0.021 **
Precipitation Squared	-0.001 **	-0.001 **	-0.001 **	-0.001 **	-0.001 **	-0.001 **	-0.001 **	-0.001 **	-0.001 **
Constant	-14776.690 **	-15955.650 **	-18564.240 **	-15890.610 **	-15467.310 **	-19038.210 **	-15235.330 **	-15086.680 **	-18533.170 **
R-squared	0.612	0.618	0.617	0.620	0.617	0.618	**0.626**	**0.625**	**0.624**
Adj R-squared	0.481	0.499	**0.500**	0.488	0.495	0.496	0.496	**0.503**	**0.503**
Number of Observations	4246	4629	4847	3872	4361	4590	3758	4266	4501

[31] A * sign means significant at the 10% level, ** at the 5% level. Coefficients reported as 0.000 may be either positive or negative.

The highest Adjusted R-Squareds[32] were obtained in the models September-March (corn) and November-April (soybeans). These were chosen not only for presenting the highest Adjusted R-Squareds but also because they have the expected signs with significant coefficients at high temperature intervals. Also, both seemed more realistic choices when taking into account the information obtained through the internet from relevant agricultural associations and other related institutions in Brazil.

The following graphs show the predicted changes in the log yields of each crop for a unit increase at the time spent at each degree interval.

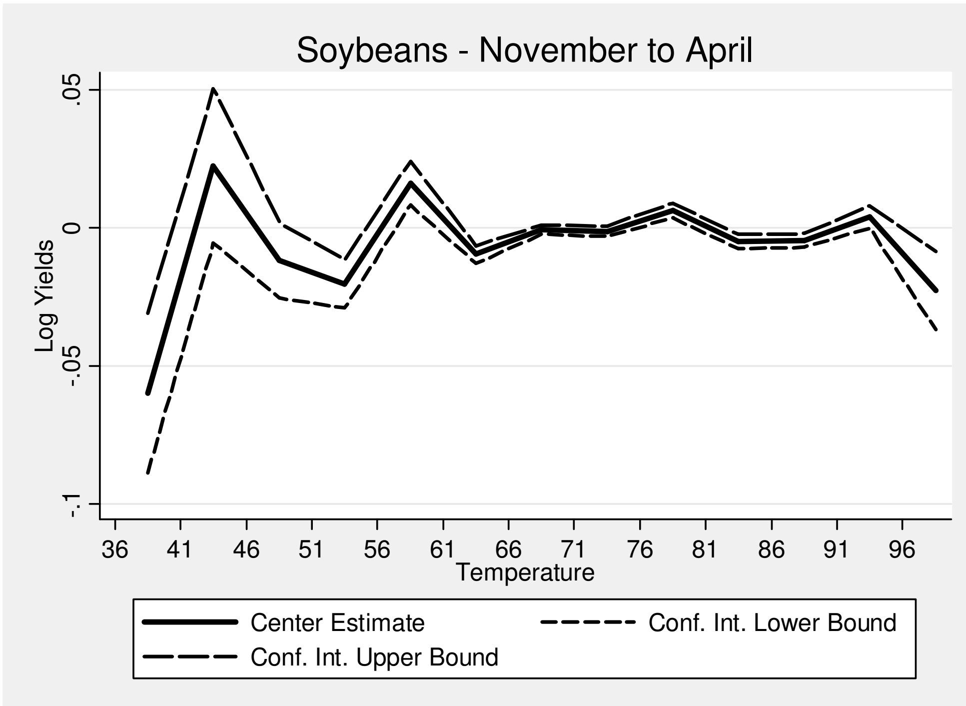
Graph 1.1 – Relationship between corn yields and temperature[33]:



[32] In this case, the Adjusted R-squared (capturing the loss of degrees f freedom) is a better measure of goodness of fit then the R-squared because of the different number of observations in each model.

[33] These graphs represent the change in the log yields associated to one extra unit of time spent at each 5 degree temperature intervals. Their purpose is to depict the harmful effects to yields after the threshold.

Graph 1.2 – Relationship between soybeans yields and temperature:



Nested models

The purpose of this section is to show the changes implied by each added component to the model. Growing seasons September to March and November to April were used for corn and soybeans, respectively[34]. Remember that a fixed-effects component had to be used to control for time-invariant heterogeneity and a time-trend was added to account for time dependency. Precipitation data was also used, since it is supposed to be another important weather factor related to agriculture production.

[34] This point may be misleading. It is not clear if the nested models determined the preferred one to be used for finding the growing season or if growing season was determined before finding the preferred model. In fact, the same results would be obtained if we found the growing season and then determined the preferred model or vice-versa. The author tested all the possible combinations, but showing them is not relevant.

In the following a table is presented, where each model to the right has the addition of a new factor[35]:

Table 2 – Nested models:

CORN

Model #	1	2	3	4	5
Temperature Interval			Coefficients		
36-41		0.026	0.017	0.014	0.015
41-46		-0.001	-0.009	-0.006	-0.007
46-51		-0.003	-0.001	-0.003	-0.002
51-56		-0.013 **	-0.017 **	-0.018 **	-0.018 **
56-61		0.005 **	0.007 **	0.007 **	0.007 **
61-66		-0.003 **	-0.004 **	-0.002 *	-0.002 *
66-71		-0.002 **	-0.001 *	-0.002 **	-0.002 **
71-76		0.001	0.001 *	0.001	0.001
76-81		0.009 **	0.007 **	0.007 **	0.007 **
81-86		-0.012 **	-0.010 **	-0.011 **	-0.011 **
86-91		0.004 **	0.004 **	0.003 **	0.003 **
91-96		-0.004 **	-0.004 **	-0.003 **	-0.003 **
96-114		-0.009 **	-0.011 **	-0.013 **	-0.013 **
Other Variables					
Year			3.228 **	5.562 **	5.587 **
Year Squared			-0.001 **	-0.001 **	-0.001 **
Precipitation				0.001 **	0.000 **
Precipitation Squared					0.000
Constant	3.071 **	3.154 **	-3242.943 **	-5572.619 **	-5598.215
R-squared	0.723	0.787	0.794	0.800	0.800
Adjusted R-squared	0.703	0.738	0.746	0.753	0.753
Number of Observations	84332	16684	16684	16202	16202

SOYBEANS

Model #	1	2	3	4	5
Temperature Interval			Coefficients		
36-41		-0.023	-0.031 **	-0.035 **	-0.060 **
41-46		-0.028 **	-0.021	-0.013	0.023
46-51		0.031 **	0.012 **	0.001	-0.012 *
51-56		-0.010 **	-0.027 **	-0.026 **	-0.020 **
56-61		0.012 **	0.015 **	0.016 **	0.016 **
61-66		-0.008 **	-0.009 **	-0.010 **	-0.010 **
66-71		-0.001	-0.001	-0.001	-0.001
71-76		0.000	-0.001	-0.001	-0.001
76-81		0.008 **	0.007 **	0.007 **	0.006 **
81-86		-0.006 **	-0.006 **	-0.006 **	-0.005 **
86-91		-0.004 **	-0.004 **	-0.005 **	-0.004 **
91-96		0.007 **	0.006 **	0.006 **	0.004 **
96-114		-0.037 **	-0.027 **	-0.026 **	-0.022 **
Other Variables					
Year			17.782 **	18.687 **	18.544 **
Year Squared			-0.004 **	-0.005 **	-0.005 **
Precipitation				0.004 **	0.021 **
Precipitation Squared					-0.001 **
Constant	3.365 **	3.336 **	-17770.900 **	-18675.310 **	-18533.170 **
R-squared	0.402	0.555	0.607	0.620	0.624
Adjusted R-squared	0.341	0.417	0.485	0.497	0.503
Number of Observations	25256	4709	4709	4501	4501

[35] A model using only a constant was estimated to show how much of the R-Squared is being accounted for by the fixed effects component itself.

The preferred models are 4 for corn and 5 for soybeans. Precipitation squared did not seem to fit the data for corn well. Its coefficient is not statistically significant at the 10% level and the after its addition, the constant in the model turned statistically insignificant at the 10% level.

All the models show the expected negative effect of high temperatures to the log yields of each culture, after a certain critical point. For corn production, this threshold appears to be in the 91-96 degree interval, where for soybeans, it is located in the above 96 degrees interval.

Moreover, the preferred corn model shows that each extra unit of time spent at the 91-96 degree interval is associated with a 0.3% average loss on yields, whereas the effect above the 96 degrees interval increases to a 1.3% loss. Soybeans apparently have a higher temperature threshold, but at the same time, the negative effect is more pronounced. An extra unit spent after in the above 96 degrees interval is associated with a 2.2% average loss on yields on the preferred model.

These estimates were presented just to illustrate the magnitude of the effect of extreme high temperatures on both cultures. Comprehensive climate change scenarios will be drawn in the conclusion section.

In Schlenker and Roberts (2006), the critical threshold temperature estimated was considerably lower, 29^{o}C (84.2^{o}F) for both corn and soybeans, however the effect over yields (at the threshold) was similar: approximately -0.5% for the former and -1% for the latter. Unfortunately it is not possible to test if the findings from their paper were statistically significantly different from this thesis'.[36] [37]

---

[36] Schlenker and Roberts (2006) estimate a variety of models. The comparable one to this thesis' was built with 1^{o}C temperature intervals and did not have its standard errors reported because there were 50 temperature intervals.
[37] Bootstrapping the residuals of the models would be an alternative way of testing if the parameter estimates are statistically significantly different and more correct since it does not assume the distribution of the parameters is

Even in the case the critical threshold proves to be statistically significantly different, one could hypothesize this happens because of different climate or agricultural techniques that can make the crops more resistant to higher temperatures. Given that Brazil has, in the average, higher temperatures in the corn and soybeans productive regions, it is possible that over time production has adapted to this difference. This explains the difference in the thresholds. But being it Brazil or USA, after overcoming the thresholds, the effects over the yields should be similar[38], which indicates the robustness of the model over different countries. This is understandable, since it is probably a characteristic more intrinsic to the plant (corn or soybeans) then to environment where it was planted.

known, but this goes beyond the scope of this thesis. Moreover, the data from Schlenker and Roberts (2006) is not available so it could not be done.

[38] Here it is meant similarity in sign and with a growing harmful effect, but not necessarily with the same coefficients, since this may also be a factor influenced by the local environment in question.

6 Sensitivity checks – the top six states in production for each crop

The quality of the data used in this thesis is certainly not one of its strongest points. IBGE is an institution that has had scarcity of resources to run its surveys sometimes in the past. Also, the amount of missing data obtained from the weather stations is very large. It is assumed that these problems represent random sources of error, meaning it only induces noise and lack of precision on the estimation, but no bias.

In an attempt to reduce this amount of noise, the top six states on corn and soybeans production[39] were sampled and the same models were run taking only these into consideration. It is expected that different growing seasons and critical thresholds will be found, exactly because we are restricting the analysis to a different area. However, the same asymmetric effect when time spent at the 5 degree temperature intervals overcomes the thresholds shall be found.

Determining the growing season

First, the growing season has to be determined. The following table shows the estimates for different seasons using only data from the top 6 states in production for each culture:

[39] These represented 76.9% of the total corn production in the country and 86.0% of the total soybeans production in 2006. **IBGE,** *"Produção Agrícola Municipal 2006 – Comentário"* – p.18-23

Table 3 – Estimates for corn and soybeans – different growing seasons (top 6 states):

CORN

	Growing Season								
	Sep-Feb	Sep-Mar	Sep-Apr	Oct-Feb	Oct-Mar	Oct-Apr	Nov-Feb	Nov-Mar	Nov-Apr
Temperature Interval	Coefficients								
36-41	0.022 *	0.008	0.001	0.037 **	0.017	0.009	0.155 **	0.025 *	0.014
41-46	-0.010	-0.001	-0.004	-0.039 **	-0.031 **	-0.024 **	-0.257 **	-0.037 **	-0.023 **
46-51	-0.004	-0.005	-0.005	0.010 *	0.006	0.003	0.018	0.008	-0.007
51-56	-0.018 **	-0.016 **	-0.012 **	-0.018 **	-0.009 *	-0.007 *	-0.016 *	-0.014 *	-0.010 **
56-61	0.008 **	0.007 **	0.004 **	0.005 *	0.001	-0.002	0.012 **	0.007 **	0.000
61-66	-0.002 **	-0.003 **	-0.003 **	-0.003 **	-0.004 **	-0.003 **	-0.005 **	-0.006 **	-0.004 **
66-71	-0.002 **	-0.001 **	-0.001	0.000	-0.001	0.000	0.000	0.000	0.000
71-76	-0.001	-0.001 *	-0.002 **	0.002 *	0.000	-0.001	0.003 **	0.000	-0.001
76-81	0.008 **	0.008 **	0.008 **	0.007 **	0.008 **	0.008 **	0.006 **	0.007 **	0.007 **
81-86	-0.007 **	-0.007 **	-0.007 **	-0.008 **	-0.008 **	-0.007 **	-0.008 **	-0.008 **	-0.008 **
86-91	-0.002	-0.001	-0.001	-0.003	-0.002	-0.001	0.000	-0.001	0.000
91-96	-0.002	0.003	0.003 *	-0.001	0.004 *	0.004 **	0.002	0.007 **	0.006 **
96-114	0.004	-0.009 **	-0.008 **	0.001	-0.017 **	-0.016 **	0.001	-0.031 **	-0.025 **
Other Variables									
Year	7.532 **	8.836 **	9.414 **	8.077 **	9.508 **	10.110 **	7.412 **	9.308 **	10.611 **
Year Squared	-0.002 **	-0.002 **	-0.002 **	-0.002 **	-0.002 **	-0.003 **	-0.002 **	-0.002 **	-0.003 **
Precipitation	0.009 **	0.005 **	0.005 **	0.011 **	0.006 **	0.007 **	0.013 **	0.008 **	0.008 **
Precipitation Squared	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000
Constant	-7544.914 **	-8847.215 **	-9424.635 **	-8087.037 **	-9517.466 **	-10117.950 **	-7423.348 **	-9318.399 **	-10619.160 **
R-squared	0.637	0.632	0.647	0.631	0.625	0.641	0.633	0.629	0.646
Adj R-squared	0.547	0.547	0.564	0.533	0.532	0.550	0.537	0.536	0.554
Number of Observations	7578	8182	8465	6889	7501	7812	6571	7218	7529

SOYBEANS

	Growing Season								
	Sep-Feb	Sep-Mar	Sep-Apr	Oct-Feb	Oct-Mar	Oct-Apr	Nov-Feb	Nov-Mar	Nov-Apr
Temperature Interval	Coefficients								
36-41	0.020	-0.043 **	-0.045 **	0.047	-0.031 *	-0.033 **	0.251 **	-0.029	-0.035 **
41-46	0.010	-0.002	0.001	-0.026	-0.028	-0.017	-0.426 **	-0.043 *	-0.021
46-51	-0.004	-0.005	0.000	0.018 *	0.004	0.009	0.021	-0.004	0.004
51-56	-0.011 *	-0.008	-0.012 **	-0.018	-0.007	-0.018 **	-0.007	-0.002	-0.023 **
56-61	0.007 *	0.003	0.003	0.009 *	0.001	0.005	0.032 **	0.019 **	0.016 **
61-66	-0.011 **	-0.008 **	-0.006 **	-0.015 **	-0.011 **	-0.008 **	-0.018 **	-0.015 **	-0.011 **
66-71	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.000	0.001
71-76	-0.001	-0.004 **	-0.004 **	0.003	-0.003 **	-0.003 **	0.005 **	-0.002	-0.003
76-81	0.013 **	0.013 **	0.011 **	0.012 **	0.013 **	0.011 **	0.011 **	0.011 **	0.010 **
81-86	-0.008 **	-0.008 **	-0.007 **	-0.009 **	-0.008 **	-0.008 **	-0.007 **	-0.007 **	-0.007 **
86-91	0.000	-0.002	-0.001	0.002	-0.003	-0.001	0.000	-0.004 *	-0.002
91-96	-0.004	0.006 *	0.003	-0.011 **	0.005	0.002	-0.009	0.008 **	0.003
96-114	-0.005	-0.021 **	-0.019 **	-0.001	-0.030 **	-0.025 **	0.004	-0.033 **	-0.024 **
Other Variables									
Year	18.250 **	19.678 **	20.170 **	18.433 **	20.299 **	21.140 **	17.392 **	20.079 **	21.124 **
Year Squared	-0.005 **	-0.005 **	-0.005 **	-0.005 **	-0.005 **	-0.005 **	-0.004 **	-0.005 **	-0.005 **
Precipitation	0.012 **	0.012 **	0.008 **	0.012 **	0.009 **	0.006 *	0.014 **	0.009 **	0.006
Precipitation Squared	0.000 **	0.000 **	0.000 **	0.000 **	0.000 **	0.000	0.000 **	0.000 **	0.000
Constant	-18239.030 **	-19667.680 **	-20159.720 **	-18419.940 **	-20286.900 **	-21127.840 **	-17380.910 **	-20068.420 **	-21112.530 **
R-squared	0.635	0.630	0.618	0.638	0.626	0.612	0.647	0.629	0.615
Adj R-squared	0.496	0.505	0.488	0.499	0.494	0.473	0.510	0.494	0.472
Number of Observations	2308	2615	2754	2100	2408	2555	2002	2328	2475

The highest Adjusted R-Squareds were obtained in the models September-April (corn) and November-February (soybeans).

In this case, the growing season that fitted the data better for soybeans did not show the expected harmful effects after the critical temperature threshold. Since it was also the model with the smaller number of observations, it will be disregarded. One interesting finding is that the second highest Adjusted R-Squared was found in a model that has a growing season considerably different from the estimated using all the states, September-March. The non-linear effects are still similar, a loss of 2.1% after the threshold, compared to 2.2% considering all the states.

As for corn, the results were not the expected. It was stated in previous sections, that two of the states that are amongst the top 6 states for both cultures, Paraná and Mato Grosso are the top producers in different corn crops (the first and the second respectively). It may be the case that the growing seasons are different amongst them for the majority of producers, making the estimation inconsistent. Two states that have clearly different growing seasons should not be grouped together.

Nested models

For the growing seasons September-April (corn) and September-March (soybeans), different specifications of the models were tested, as seen in the table below:

Table 4 – Nested models (top 6 states):

CORN

Model #	1	2	3	4	5
Temperature Interval			Coefficients		
36-41		0.016	0.004	0.003	0.001
41-46		0.003	-0.008	-0.006	-0.004
46-51		-0.008 **	-0.002	-0.005	-0.005
51-56		-0.005 *	-0.015 **	-0.012 **	-0.012 **
56-61		0.001	0.004 **	0.004 **	0.004 **
61-66		-0.002 **	-0.003 **	-0.003 **	-0.003 **
66-71		-0.002 **	-0.001	-0.001	-0.001
71-76		0.000	-0.002 **	-0.002 **	-0.002 **
76-81		0.009 **	0.008 **	0.008 **	0.008 **
81-86		-0.009 **	-0.008 **	-0.007 **	-0.007 **
86-91		0.001	0.000	-0.001	-0.001
91-96		0.004 **	0.003 *	0.003 *	0.003 *
96-114		-0.007 *	-0.011 **	-0.009 **	-0.008 **
Year			8.744 **	9.417 **	9.414 **
Year Squared			-0.002 **	-0.002 **	-0.002 **
Precipitation				0.001	0.005 **
Precipitation Squared					0.000 **
Constant	3.685 **	3.689 **	-8755.085 **	-9426.834 **	-9424.635 **
R-squared	0.512	0.589	0.641	0.647	0.647
Adjusted R-squared	0.478	0.497	0.560	0.563	0.564
Number of Observations	43265	8933	8933	8465	8465

SOYBEANS

Model #	1	2	3	4	5
Temperature Interval			Coefficients		
36-41		-0.006	-0.028 **	-0.037 **	-0.043 **
41-46		-0.027 **	-0.021 *	-0.008	-0.002
46-51		0.000	0.007	-0.001	-0.005
51-56		-0.009	-0.015 **	-0.009 *	-0.008
56-61		0.006	0.007 **	0.003	0.003
61-66		-0.007 **	-0.010 **	-0.009 **	-0.008 **
66-71		-0.001	0.001	0.001	0.001
71-76		-0.003 **	-0.005 **	-0.004 **	-0.004 **
76-81		0.015 **	0.014 **	0.014 **	0.013 **
81-86		-0.008 **	-0.010 **	-0.009 **	-0.008 **
86-91		0.000	0.002	-0.002	-0.002
91-96		-0.006 *	0.001	0.006 *	0.006 *
96-114		-0.011 *	-0.021 **	-0.022 **	-0.021 **
Year			16.529 **	19.748 **	19.678 **
Year Squared			-0.004 **	-0.005 **	-0.005 **
Precipitation				0.001	0.012 **
Precipitation Squared					0.000 **
Constant	3.359 **	3.361 **	-16522.290 **	-19737.170 **	-19667.680 **
R-squared	0.404	0.532	0.596	0.628	0.630
Adjusted R-squared	0.349	0.387	0.470	0.502	0.505
Number of Observations	17683	2837	2837	2615	2615

In contrast with the models that include all the states, the one with precipitation squared (5) was the chosen one for corn. Also, after restricting the analysis only to the top 6 states for each culture, the weather variables and the time trend are explaining a larger proportion of the variance in the dependent variable. This is observable by subtracting the Adjusted R-Squared of model 5 from model 1 (where the variance in the log yields is being explained only by time invariant factors relative to each municipality). In the restricted state analysis, it corresponds to 8.6%. Considering all the states, it was 5.0%.

These results underline the importance of defining the correct growing seasons for each state and the influence it can have on the estimates. This is an issue that future research should try to solve in the case more reliable data is available.

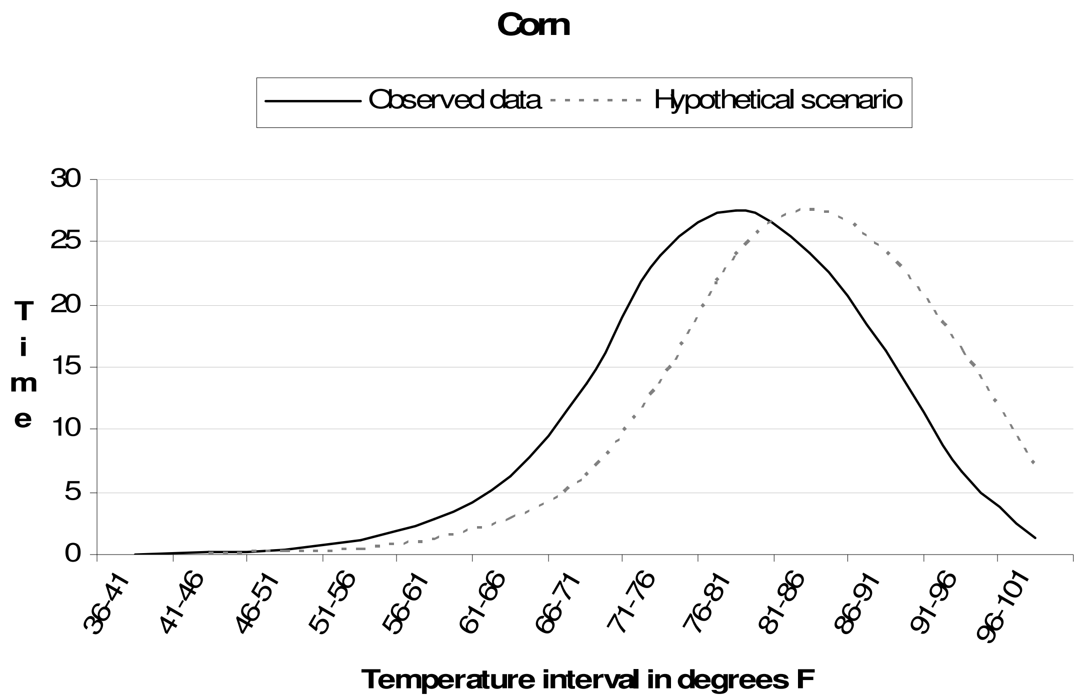
7 Conclusion

The coefficients estimated indicate the predicted change in the log yields for a unit increase in time spent at each five degree temperature interval. With these in hand, it is possible to predict the change in the yield of future crops as a function of a hypothetical temperature scenario.

Since the models estimated are only sensitive to temperature changes higher than 5^{o}F, it is possible to predict the loss in yields for each crop if the average temperatures increased by 5^{o}F[40].

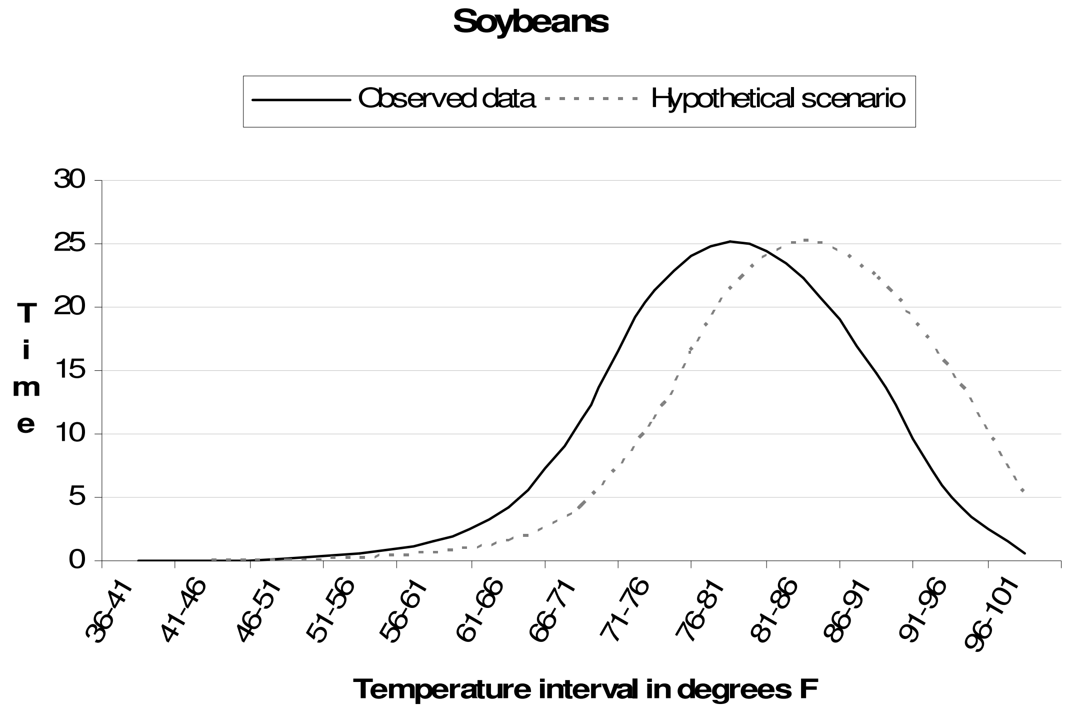
The models that are going to be used will be the same that were selected in chapter 5: September-March for corn and November-April for soybeans.

Graph 2.2 – Time spent at each temperature interval over the September-March growing season:



[40] This for the thesis' model is equivalent to affirming that the distribution of country average time spent at each five degree temperature interval uniformly shifted one interval to the right (the next higher interval).

Graph 2.3 – Time spent at each temperature interval over the November-April growing season:



Using the differences between the hypothetical scenario and the observed data for each temperature interval (change in time spent at each 5^{o}F interval), it is possible to calculate the predicted change in the log yields using the coefficients already estimated. The results are a loss of 9.3% for soybeans yields and 12.8% for corn.

It is worth remembering that the predictions found are static, meaning that the possible technological changes are not taken into account, since only the temperature coefficients from the estimated model were used. This also implies that the predictions are valid for producers that stay in the same location (these predictions are not valid if the spatial distribution of production changes a lot). In other words, an assumption is being made: that technological and spatial distribution of production changes will not be significant.

Also, the hypothetical scenario is taking into account temperatures averaged over the entire country. However, where productive regions have clearly different climates and growing seasons, the ideal way would be to estimate the models for each region and predict the changes in yields with the corresponding scenario. One last remark that should be made is that the precipitation levels were assumed not to change in the hypothetical scenario.

This work had to go through severe data limitations; however, the relationship between weather and crop yields found for Brazil is somewhat similar to what has been found for the USA by Schlenker and Roberts (2006), which is a good indicator. The methodology used is meant to be applicable for other countries. Future research could follow a similar path, possibly with more reliable data, observing the importance of correctly defining the growing season for each culture in different areas.

8 Bibliography

ARROW, Kenneth J. et al, “Is There a Role for Benefit-Cost Analysis in Environmental, Health, and Safety Regulation?”, *Science*, April 12 1996, *272*, 221

FULLERTON, Don and STAVINS, Robert, “How economists see the environment”, *Nature*, October 1 1998, *395*, 433-434

FREEMAN III, A. Myrick, “Environmental Policy Since Earth Day I: What Have We Gained?”, *Journal of Economic Perspective*s, Winter 2002, *16*, 126

IBGE, *“Produção Agrícola Municipal 2006 – Comentário”* – p.18-23 <http://www.ibge.gov.br/home/estatistica/economia/pam/2006/comentario.pdf>

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE FOURTH ASSESMENT REPORT, “Climate Change 2007: Synthesis Report – Summary for Policymakers”, 2007

SCHLENKER, Wolfram and ROBERTS, Michael J., “Estimating the impact of climate change on crop yields: the importance of non-linear temperature effects” (non-published working paper)

SCHLENKER, Wolfram and ROBERTS, Michael J. "Nonlinear Effects of Weather on Corn Yields," *Review of Agricultural Economics (Proceedings of the American Agricultural Economics Association Sessions at the 2006 Allied Social Science Associations Meeting, Boston, MA)*, Fall 2006, *28*(3), 391-398

Data sources:

<www.ibge.gov.br>

<www.sidra.ibge.gov.br>

<http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html>

<http://www.ipcc.ch/>