How Suitable are FAO-GAEZ Crop Suitability Indices for Historical Analysis?

Paul W. Rhode

June 28, 2024

Version 7

Preliminary; comments welcome.

Abstract: This study evaluates the suitability of FAO-GAEZ Crop Suitability Indices for the historical analysis of American agricultural development. It focuses on three staple crops—cotton, wheat, and corn—grown in the United States over the nineteenth and early twentieth centuries. This study first “looks under the hood” to inquire how key parameters used in the construction of FAO-GAEZ indices relate to those reported in the plant science literature. The paper then introduces a novel dataset based on manuscript census reports in an exercise to investigate how well the GAEZ indices predict historical US crop yields for cotton, wheat, and corn. Finally, it explores how the GAEZ regions compare with conventional-defined crop belts in the United States.

Acknowledgements: I have benefited from helpful discussions with Robert Allen, Jeremy Atack, Hoyt Bleakley, John Clegg, David Donaldson, Martin Fiszbein, Tim Guinnane, Josh Hausman, Richard Hornbeck, Jennifer K. Kowalski, Naomi Lamoreaux, Trevon Logan, Alan Olmstead, Elyce Rotella, Mel Stephens, Dean Yang, Gavin Wright, and participants at the Michigan Economic History Workshop. I thank Gavin Wright and Vasily Rusanov for assistance with data acquisition.

Author Affiliation: Rhode is Professor of Economics at the University of Michigan and Research Associate at the National Bureau of Economic Research. Contact Address: pwrhode@umich.edu

"But part of the job of economics is weeding out errors. That is much harder than making them, but also more fun." Robert Solow[1]

The paper evaluates the use of the Food and Agriculture Organization Global Agro-Ecological Zone (FAO-GAEZ) crop suitability indices in historical studies of American agricultural development. The use of these indices has become so common that one hardly needs to cite the numerous papers doing so. The premise of these papers is that one can use modern-day remote sensing data in combination with complex expert-based model to infer potential crop yields in the distant past. One simply dials down the input setting from high to intermediate or low. The GAEZ model was designed for current-day policy purposes, namely "to assess the capacity of the world's natural resources to meet the needs for food of a fast-growing global population, particularly in developing countries (https://gaez.fao.org/)." Outside researchers now use the FAO-GAEZ suitability measures in historical studies to capture the exogenous natural conditions dictating where specific crops can be grown.

I have three main concerns. 1. The FAO-GAEZ numbers likely capture technological changes resulting from investments in plant breeding. The classic work of Zvi Griliches (1957, 11959, 1960) on the introduction and spread of hybrid corn in the United States showed innovation was faster in places with higher initial yield potential. The size and scope of the agricultural research systems matter also likely matter (Alston et al 2010; Kantor and Whalley 2019). These forces make crop suitability indices, created with current-day data, inherently endogenous to the growth progress. Not exogenous. Nor are the errors likely white noise.

2. The FAO-GAEZ numbers may not reflect dynamic environment changes. While they do allow for alternative climatic scenarios, water sources (rainfed vs irrigated), and different input regimes (high, intermediate, and low), they do not incorporate the effects of pest invasion, soil deterioration, or other changes in an unstable biological environment. Institutional factors are also ignored.

3. Parts of the model, on closer inspection, appear simply to be wrong, in consequential ways. The model is quite opaque – but arguably it is not complicated enough to provide a true

---

[1] Quoted in Jorgenson and Griliches (1967, p. 249).

mapping of the “proper place” on the globe of every important crop since the advent of agriculture. This was not the modelers’ own original goals, though it is the use to which model has now been put. If one looks under the model’s hood, one finds the extensive use of cut-and-pasted round numbers for key parameters; employed without detailed documentation and changed across versions without explanation (see Fischer et al, 2012, 2021). I have not seen these issues discussed in the literature, nor any attempt to justify the use of GAEZ model for historical inquiries.

In this evaluation, I will focus on three major staple crop--wheat, corn, and cotton – grown in the United States over the nineteenth and twentieth centuries. I will not critique specific research contributions in the academic literature nor cite specific social scientists except to mention my use of suitability indices down-loaded from online replication data sets. I use publicly-available replication data to emphasize that the issues raised in this paper are not the product of some funky data collection procedure at my end.[2] The maps for the United States generated from the replication data (and presented in the published work) do look similar to those displayed by the FAO-GAEZ.

There are pros and cons regarding this paper’s US focus. Many studies using the indices have investigated US agricultural development. Also on the pro side, the United States has abundant historical data – consistently collected over time and space—to use in evaluating the GAEZ indices. On the con side, the GAEZ data were not originally compiled with the US situation in mind. That said, the maps do include North America without any special comment or exceptions noted. Nor do the US-oriented studies suggest any need for adjustment.

This study first “looks under the hood” to inquire how key parameters used in the construction of FAO-GAEZ indices relate to those reported in the plant science literature. It then introduces a novel dataset based on manuscript census reports in an exercise to investigate how well the GAEZ indices predict historical US crop yields for cotton, wheat, and corn. Finally, this study explores how the GAEZ regions compare with conventional-defined crop belts in the United States, most critically assessing the placement of the cotton belt.

---

[2] I use the replication files of Hornbeck and Naidu 2019; Hornbeck and Keskin 2019; and Acharya, Blackwell, and Sen 2016. As I shall argue, if the FAO-GAEZ data are right, the historical issues raised are bigger than any covered by the existing literature.

I will focus on the Version 3 data, which was released in 2012 and has been subject to the most analysis by economists and economic historians. The Version 3 data includes an intermediate level of inputs, which has been most used in historical studies. The Version 4 data, released in 2020, appear to be largely the same, although this version maps yield potential only for low and high input levels.[3]

Scholars use measures for (1) Agro-climatic potential yields," which are based on climatic and soil moisture conditions and (2) "Suitability and Attainable Yield," which combines the agro-climatic potential yields with the results of soil/terrain evaluation. I will focus largely on the latter.

Looking under the Hood

The FAO-GAEZ employs a rather opaque model with sevens submodules (see Fischer et al 2012, 2021). Figure 1 reproduces the schema from Version 3. Social scientists have investigated the model at various levels. But I have seen little or no analysis looking under the hood, at how the potential yield estimates are derived. (Some analysis that I have seen valorizes the fact that the FAO-GAEZ does not attempt to situate the estimates in actual farm practices.)

A key first step involves the biomass and yield parameters introduced in Module II. The FAO-GAEZ generates its estimates of a crop's potential yield as the product of the crop's Harvest Index (HI) and its Leaf Area Index (LAI). The Harvest Index, as conventionally defined, measures how much of the plant's above-ground biomass is devoted to the products that human's value—the grain kernels and cotton lint for the crops considered here. The Leaf Area Index measures the extent of the leaf canopy available to capture the sunlight arriving on the planted acreage and convert its energy into the plant's biomass.

Table 1 reproduces, in abbreviated form, the "Biomass and yield parameters" data from GAEZ Version 3 for wheat (winter and spring), maize, and cotton. (The link in the FAO document is broken, so I am relying on an archived version.) The information of interest includes the "Harvest Index" and "Maximum leaf area index." The parameters in Version 3 differ from

---

[3] First global AEZ assessment was released in 2000 (GAEZ v1) and Global AEZ resource evaluations in 2002 (GAEZ v2).

those in Version 4. (See Table 2, derived from GAEZv4, Appendix table 4.6). The GAEZ documentation does not provide reasons for the changes. Nor were the differences between Version 3 and Version 2 explained. Indeed, the detailed sources of the parameters are not documented in any case. The text cites FAO working papers from the 1970s for estimating African crop yields (Kassam 1979) and consultations with unnamed experts.[4]

The FAO-GAEZ model is not as sophisticated as the start-of-the-art crop simulation models --such as the DSSAT (Decision Support System for Agrotechnology Transfer) models -- commonly used in the agronomy literature.[5] The creators of the FAO-GAEZ model appear to recognize this limitation as the documentation for version 4 regarding "Recommendation for further GAEZ development" discusses extensions to integrate with DSSAT modeling (as in Tian et al., 2020.) Fischer et al. (2021, p. 179) notes "[e]xpanding this approach would be beneficial for enhanced AEZ applications." However desirable, extensions in this direction will not assist in using the model for historical analysis unless older cultivars are included.

The first point that this reader notices about the GAEZ's current approach is that big round numbers appear to be liberally cut and pasted across the table. The number of significant digits reported is low. And there are no error bounds.

The second point that this reader notices is that the numbers appear, in specific cases, to be wrong. The use of the very similar parameters in Version 3 for wheat and corn is surprising in reference to the changes associated with the Green Revolution in temperate zones. The GAEZ has the Harvest Index for winter wheat increase from 0.2 with low inputs, to 0.35 with intermediate inputs, to 0.5 with high inputs across all crop cycles. The changes for spring wheat are from 0.15 to 0.3 to 0.45 across all crop cycles. And for maize, from 0.2, to 0.33, to 0.45 across all crop cycles.

---

[4] See Appendix 1 on the GAEZ text. The GAEZ version 4 website has "Crop profiles" for wheat, maize, and pearl millet. These do not document the parameters in the Table. They provide links to other FAO sites, but their relevance is unclear.

[5] The DSSAT model (version 4.8.2) now covers 42 crops and incorporates the well-known CERES-Maize and CERES-Wheat models. The *Agronomy Journal*, the official joint publication of the American Society of Agronomy (ASA), Crop Science Society of America (CSSA), and Soil Science Society of America (SSSA), has 164 articles using the DSSAT model. This journal has no publication using the GAEZ model. A common practice in the *Agronomy Journal* articles using the DSSAT is to evaluate model performance.

Even a cursory reading of the literature on Green Revolution plant breeding in the temperate countries teach us that the Harvest Index increased significantly for wheat but changed little for maize. Attention in maize breeding has been devoted to increasing plant density. The Harvest Index started high before the advent of hybrid corn breeding—above 0.4—and remained high. Thus, the Harvest Index was well above those used by the GAEZ. For wheat, with the introduction of semi-dwarf varieties, the Harvest Index did increase (although plant density did not greatly change). Wheat and maize are fundamentally different crops and changed via fundamentally different pathways (Evans 1993).

Evans (1993, p. 243) reports the dogma that "wheat breeders have improved only the crop's partition efficiency to the grain without an effect on the total phytomass produced…" By way of contrast, Duvick (1984, p. 15; 2005, pp. 83-84; 96-97) notes the very small changes in the harvest index in corn production. But Duvick (1999) and Evans (1993, pp. 246-47) observe that, in corn breeding, the key issue regarding the Harvest Index is maintenance, to reduce the incidence of rising barrenness at higher plant densities.[6] In analysis of the growth patterns of historical wheat and maize cultivars, Hutsch and Schubert (2017) show yields rose by 59 percent for wheat and 51 percent for maize. In the same comparisons, the Harvest Index for wheat rose by 50 percent whereas that for maize rose by only 11 percent. Total biomass per acre for wheat was "approximate stable" whereas that for maize "increased in the same way as grain yields".[7]

The patterns displayed in the GAEZ table of parameters for the changes associated with different input levels are counter to changes in actual practice in advanced countries in the temperate zone. It is possible that the Harvest Index changed for maize in the tropical zone though the evidence is unclear (Hay and Gilbert 2001).

The FAO-GAEZ's use of the same minimum temperature variables, 10 degrees C, for both maize and cotton is also striking. The GAEZ's treatment of cotton is sharply at variance

---

[6] Neither Duvick (1997) nor Russell (1991) found changes in the Leaf Area Index for maize varieties grown in Iowa after 1930. Tollenaar, McCollough, and Dwyer (1994, p. 192) stress the importance of plant density for increasing maize yields. Slater et al. (1994, pp. 17-19) chart the close relationship between the rise in the harvest index for wheat and grain yields for cultivars released in the United States, Australia, Argentina, and the United Kingdom. But they find little change in above-ground biomass.

[7] The Hutsch and Schubert (2017) analysis is based on Hays (1995) and Russell (1991).

with the US agronomic literature on cotton. Cotton likes it hot and sunny. As G. Collings (1926, p. 4) observed: "Cotton requires six months of warm; weather and four to five months of uniformly high temperature." C. P. Brooks (1898 p. 108) adds: "Cotton flourishes best when the nights are warm; as well as the day, but this advantageous feature, from a cotton raising point of view, is of distinct disadvantage to the human." J-H Chang (1968) commented, on the flip side, that the cotton plant hates cold weather and is highly intolerant of frost.

Standard degree day calculations for cotton use a base temperature of 60°F (15.55°C).

Cotton DD= $DD_{60} = \sum(\text{Ave Temp}-60°) = \sum(((\text{Max Temp}+\text{Min Temp})/2)-60°)$

In words, the cotton degree days are the cumulated number of days that the temperature exceeds 60° weighted by the excess. A day with an average temperature of 61° counts as 1 degree-day; a day with an average temperature of 65° counts as 5 degree-days.

In the Mid-South, the cotton experts tie cotton suitability to acquiring 2200-2600 DD_{60} over the 130-160 days between planting and harvesting (Tharp 1960; National Cotton Council, n.d.). Daily temperatures lower than 60°F result in boll shedding and reduced yields. Cotton degree day base is well above corn (10°C/50°F); or wheat (alternatively placed at 5°C/40°F and at 0°C/32°F).[8]

But the FAO-GAEZ does not calculate $DD_{60.}$ They calculate cumulative degree days using base 50°F, (that is10°C) as with corn degree days (replacing the 60°F in the above formula with 50°F). The FAO-GAEZ also calculate cumulative base 5°C; and base 0°C, the lower DD relevant for wheat. But nothing higher. The GAEZ model has no way to directly capture cotton's love of heat.

This reader's observations are very specific. But there is nothing in the FAO documentation to indicate that more care and attention was devoted to generating the other key parameters used in this module.

Another issue in the GAEZ framework, evident in Tables 1 and 2, is the fixed menu of crops with specific growth cycles. For a given crop, the model includes several variants with

---

[8] Humans are said to like 65°F and dislike deviations from it. Heating degree days are the cumulative degree days below the 65°F threshold; cooling degree days are those above it. Humans like cool nights where cotton likes it hor.

different growing cycles and reports the yields with highest local suitability. The level of technology does not affect the availability of crops of different growing cycles. And yet gains and losses of cultivars with different growing cycles has been a key feature of agricultural development, historically.[9]

For temperate and subtropical cotton, the GAEZ reports on three growing seasons: 135, 150, and 165 days. But we know that before the invasion of boll weevil (1892-1922), farmers in the southern US grew cottons over a season of 180-220 days. The pest had its most devasting effects on the late crop and pushed farmers to adopt earlier maturing varieties. (The pest invasion also inverted the north-south yield gradient.) The adoption of mechanical cotton picking (in the 1950s and 1960s) and especially the spread of one-pass picking (in the 1980s and1990s) also altered—that is, shortened—the relevant growing season. The fixed set of choices on the GAEZ menu does not reflect the historical options available to farmers.

A similar problem arises for maize. The GAEZ reports on six growing seasons: 90, 105, 120, 135, 150, and 165 days. But one reads, that in Canada, hybrid corn seed is marketed today that is promised to mature in a season in the low 70s. Articles about the "Corn Belt moving north" comment that the change is not principally due to global warming, but rather to "improvements in germplasm."[10] And as the work of Olmstead and Rhode (2008, 2011) demonstrate, the northward march of the Corn Belt is hardly new. Their mapping of the geography of maize production, linked to the introduction and spread of more hardy varieties, show the northern fringe advanced by three degrees latitude over the 1839 to 1969 period (Olmstead and Rhode 2011, p. 182).

---

[9] The level of technology --high, intermediate, and low in version 3 and just high and low in version 4—enters the model in two main ways. It affects the terrain capable of supporting specific crops. And it affects yield potential through changing the crop's harvest index and leaf cover, but not the menu of available crops.

[10] Robert Arnason, "Monsanto focuses on early maturing corn" Western Producer, July 16, 2015/'
https://www.producer.com/news/monsanto-focuses-on-early-maturing-corn/
Gord Gilmour "The Corn Belt Moves North" https://www.country-guide.ca/crops/the-corn-belt-moves-north/ March 8, 2016.
Ralph Pierce: "New in corn hybrids for the West"
https://www.country-guide.ca/crops/new-in-corn-hybrids-for-the-west/ Oct. 18, 2017
Mark Harsall "New Corn Varieties for 2022"
https://www.grainews.ca/features/new-corn-varieties-for-2022/ Dec 8, 2021.

Yield Tests

Data on historical crop yields can be used as a further test of the “ground truth” of the GAEZ indices. In 1850 and 1860, as part of the US Census of Social Statistics, the federal census asked the local census marshals for information on “usual average crops” in their area of enumeration. The same census schedule solicited information on wealth, taxation, schools, churches, libraries, newspapers, wages, and other social statistics. Coverage was far from complete and the county-level yield statistics were never published. The manuscript forms are available for many states, in archives and libraries scattered across the country. Figure 2 illustrates a sample page. The “usual yield” statistics are assembled here into a county-data set and used for the first time.[11]

The data were collected from manuscript records from National Archives (some on microfilm available at Ancestry.com and Family Search.com) as well as the Washington, DC branch of the National Archives. Records in the holdings of state libraries and archives in California, Connecticut, Indiana, Michigan, Minnesota, Oregon, New Hampshire, New Jersey, Rhode Island, and Vermont were accessed. And Gavin Wright kindly provided records from parts of the South.[12] Data for Ohio are largely missing at present, but it might be possible to create proxies from the abundant data available from state statistical surveys.

The paper will focus on the data for 1860, when there are 155 county-level observations for cotton; 868 for wheat; and 878 for corn. See Figure 4 for a map of the 1860 coverage. The reporting area represents only a fraction of the counties where these crops were grown. One should not form the impression that the “usual average” yields data are sparse. While entries are completely missing for many counties, they are superabundant for other counties. In Connecticut, in 1860, there are observations for 40 different townships across the eight counties. These are averaged into 8 CT observations in the dataset.

---

[11] The section on the schedule inquired about “seasons and crops, the kinds of crops short, to what extent. And he usual average crops.” Wright (1900 p. 46). Some marshals reported the total quantities produced; but the data set includes only the physical quantities of crops per acre.

[12] Toman (2002) used the usual yields figures in combination to the reports on short crops to estimate actual cotton and corn yields in the South in 1850 and 1860.

James D. B. DeBow, superintendent of the 1850 Census, produced a state-level table of crop yields which as “very incomplete, but nothing better can be framed upon the returns, which in general were very carelessly made or entirely neglected.” US Census Office (1854, p. 178). Reports in 1860, the second time that the census marshals collected the data, were more extensive.

The Census of Social Statistics represented the largest and most systematic effort to estimate crop yields in antebellum America. The "Usual Average Crop" estimates were not tied to actual production in specific year, but to the census marshals' assessment of typical local conditions. The US Census of Agriculture collected output data in this period, but did not enumerate acreage by crop until 1880 (see below). The US Department of Agriculture (USDA), formed in 1863, first reported state-level yields for corn and wheat in 1867 and for cotton in the late 1870s.

Summary statistics for the usual yield per acre estimates in 1860 are as follows:

Crop	Obs.	Mean	St. Dev	Min	Median	Max
Cotton	155	870lb	292	100	800	2000
Wheat	868	16.5bu	6.6	5	15	50
Corn	878	34.2bu	12.0	6	35	100

The aggregate numbers for wheat and corn appear in the ballpark of what is reported in Parker and Klein (1966, p. 500) and Atack and Bateman (1987, p. 162-85). The cotton number relates to the yield of seed cotton. It translates to somewhat under 0.75 bales per acre, which is roughly consistent with the yields reported in Gray (1933, pp. 708-09). It appears high, likely because the counties from new areas are overrepresented in 1860 cotton sample. The numbers for each of the crops is on the high side of what J. D. B. DeBow (US Census Office 1854, pp. 176-78) averred where the actual yields for 1850. Note the census marshals reporting Usual yields were typically in locations growing the crops under consideration. This leads to issues of selection, which are examined in Appendix 2.

Histograms for the distribution of 1860 Usual yields per acre appear in Figure 3. There is lumpiness in the distributions, certainly. The yields will be transformed into logs to address scaling issues.

A first pass: using GAEZ indices to predict 1860 Usual Average crop yields

As a first pass at quantitatively evaluating the historical suitability of the GAEZ indices, I ask how well they predict the 1860 Usual yield numbers at the county level. I run the following specifications, in logs, for cotton, corn, and wheat:

(1) $Ln\_Usual\_Crop = \alpha_0 + \alpha_1 \, Ln\_GAEZ\_Crop$

(2) $Ln\_Usual\_Crop = \alpha_0 + \alpha_1 \, Ln\_GAEZ\_Crop + \alpha_2 Lat + \alpha_3 Lon$

Are the elasticities, α_1, close to unity? With a coefficient of unity, the 1860 Usual yields would move in proportion with the GAEZ suitability indices. Are the R-squared's large? Do the estimates stand up to including geographic controls?

Table 3 reports the results of this simple analysis. Panel A reports regressions predicting 1860 Usual yields using the GAEZ own crop indices. The GAEZ indices do not predict the historical data well. The R^2s without the latitude and longitude controls are very low. The coefficients are close to zero, not one. When latitude and longitude controls are included, GAEZ cotton loses all explanatory power. Panel B raises an additional problem. When GAEZ wheat is included with GAEZ maize to explain corn yields, GAEZ wheat wins and the coefficient of GAEZ corn become negative. This is not simply an issue of multicollinearity because the standard errors do not blow up; in some cases, they shrink. Including latitude and longitude controls rescues the situation, but only by robbing explanatory power from the GAEZ variabilities. This is a poor showing, indeed.

A second pass: Horseraces between Usual yields and the GAEZ indices to explain crop density

It is possible that the 1860 Usual Average yield numbers are bad data, that the US census marshals badly gauged the world in which they lived. And that the GAEZ creators knew better. To explore this possibility, I compare the ability of the 1860 Usual yield and GAEZ indices to explain actual US crop yields during the late nineteenth and early twentieth century.

From 1880 on, the Census of Agriculture collected data on both output and acreage for specific crops. These allow the calculation of yields by county (Haines, Fishback, Rhode 2018). With the GAEZ indices and the Usual Yield data, we can run solo sprints as well as head-to-head races, both with and without latitude and longitude controls. And as with the above exercise, we can compare the ability of the GAEZ maize and GAEZ wheat in explaining corn yields. A further point of interest from this exercise is to see how the estimates change over time, from census to census. The 1860 Usual yield data might be more informative in the earlier years than in the later years. Note, of course, the 1860 Usual yields sample is fixed, and cannot shift as overall production does.

Tables 4-10 reports the regression results. The GAEZ indices perform extraordinarily badly, in many and varied ways. In solo sprints (in the whole data set and in that restricted to counties with 1860 Usual yields data), the GAEZ indices do a poorer job than usual yields in explaining actual outcomes. And in head-to-head races, the 1860 Usual yields almost always win. For all their problems, the 1860 Usual yields perform better than the GAEZ counterparts, even into the twentieth century.

Turning first to the solo sprints in cotton (Table 4), one observes that in the full sample without geographic controls (in the upper left corner), the coefficient on GAEZ cotton is negative (but statistically insignificant). Restricting attention to the sample where Usual yields are available (next column over), the coefficient on GAEZ cotton becomes mixed, but is still very small and (but statistically insignificant). In the same sample (the right column), Usual yields start stronger in 1880—with a coefficient of 0.38, positive and statistically significant. Nothing to brag about, except relative to the poor performance of GAEZ cotton. The explanatory power of Usual yields does decline over time, but this is in keeping with the observation that the boll weevil invasion altered yield patterns in the Cotton Belt. It should not be taken as support for the use of the GAEZ indices, but quite to the contrary.

Table 5 reports results from the head-to-head contest between GAEZ cotton and 1860 Usual cotton yields. In 1880, the Usual yields number wins. The bottom panels of Table 4 and 5 includes latitude and longitude controls. These soak up much of the explanatory power. This is another mark against the GAEZ yields as they are designed (or at least used) to capture effects beyond smooth changes in geography.

Turning to wheat (in Tables 6 and 7), one observes that GAEZ wheat index does okay in the solo sprints, both in the whole sample and the sample restricted to counties with 1860 Usual yield data. The coefficients on GAEZ wheat are positive and statistically significant. They remain low, far from a value of unity (moving proportionally with actual yields). Moreover, the 1860 Usual wheat yields perform better—with higher positive coefficients, both in the solo sprints (Table 6) and the head-to-head races (Table 7).

The coefficients on Usual yields do drift down over time, reflecting dynamics attenuating the effects of initial (1860) conditions. (Note this is not due to compositional effects driven by western settlement because the counties in the 1860 sample are fixed.) Even at the end of the sample period, however, 1860 Usual yields out-performs GAEZ wheat. And we have not exhausted the bad news for GAEZ wheat (see the discussion of Table 10 below). The bottom panel of Table 5 includes latitude and longitude controls. These again soak up explanatory power, but leave the overall patterns unchanged.

And, finally, onto predictions of actual historical yields of corn (in Tables 8, 9, and 10), one observes strikingly poor performance for GAEZ maize in combination with very good performance by the 1860 Usual yields for corn. Table 8 reports results from the solo sprints involving GAEZ maize and 1860 Usual corn yield. Table 9 reports on the head-to-head race. The coefficient on 1860 Usual yields for corn starts very high – close to 0.9—drifts down and then re-bounds (in the 1930s when the federal crop programs restrict acreage planted). The GAEZ maize coefficients vary in sign, but are always small and economically insignificant. Adding latitude and longitude controls improves the appearance of GAEZ maize versus 1860 Usual yields. But the changes are only marginal.

Table 10 reveals a further problem for the use of the GAEZ indices. This table reports the results of a horserace between GAEZ maize and GAEZ wheat to explain historical census corn yields. In the head-to-head contest without geographic controls, GAEZ wheat typically wins. Its coefficient is positive and significantly different from zero. The coefficient on GAEZ maize is negative (and statistically significant). The power of GAEZ wheat to explain corn yields undermines its usefulness as an indicator of local suitability for specific crops, a use to which it is commonly employed. The variable is deployed as a measure of comparative advantage when it does the opposite here. The correlation between the GAEZ maize and GAEZ wheat indices is high—above 0.75. Adding geographic controls rescues GAEZ maize, chiefly by robbing both GAEZ variables of explanatory power.

Figure 5 illustrates the results for the top panel of Table 4-10 for the regression excluding latitude and longitude controls.

Appendix 2 includes the results of comparable regressions (a) using the GAEZ high suitability measures for maize and wheat; and (b) using the same variables as in the test but run on annual county-level crop yield data from the National Agricultural Statistical Service for selected states over the 1919-61 period. The findings are in line with what is reported in the text above. In the regressions for wheat using the high indices, GAEZ eventually dominates the 1860 Usual yields, but the elasticities are always well below unity, that is, proportional changes.

Mapping GAEZ Crop Regions

As a further test of the suitability of the GAEZ indices for historical analysis examine the maps in Figure 6 for cotton. Panel A maps the cotton suitability data for the United States with its 1860 county boundaries. Panel B maps the location of the canonic Cotton Belt (Stine and Baker 1918). Panel C maps for cotton production relative to county land area in 1860, so the number of bales per total land area. The GAEZ date has high yield potential for upland cotton in Illinois, Indiana, central Missouri, eastern Kansas and Nebraska, and into Iowa, regions where cotton was never commonly grown. Alabama and parts of Mississippi are assigned low suitability. The GAEZ map is very different from our common understanding of the historical location of the American cotton belt.[13]

The conventional wisdom was that 37 degrees north latitude was the upper limit of the US cotton belt. Milton Whitney (1896 p. 143) observed: "Cotton production in the United States is limited by climatic conditions to that portion of the country south of latitude 37°. A few small areas are cultivated north of this on account of local peculiarities of the climate or market conditions, but for the most part the cultivation does not extend north of the northern boundary of North Carolina." The line "crosses the country a little below latitude 37°, south of which the

---

[13] In the colonial period, farm families in the South, Pennsylvania, New Jersey, Kentucky grew cotton plants in their home gardens. They kept a single or small number of plants to supply mending thread; a few early bolls opened before frost killed the plants. This practice ended once supplies of commercial cotton thread become available. See Gray (1933) and Chaplin (1993).

climatic conditions are generally favorable to the production of cotton, and north of which they are unfavorable on account of the short season and the relatively low mean temperature."

Agronomist Harry Brown (1927, p. 225) stated: "In the United States but very little cotton is grown north of latitude 37°. This is near the northern boundary of North Carolina…The states north of the thirty-seventh parallel have a growing season too short for cotton to mature and the mean temperature is too low." Brown (1927, p. 226) continued: "Along the northern margin of the Cotton Belt the last killing frost in the spring is, on the average, about Apr. 10, and the first killing frost in the fall about Oct. 25, the growing season for cotton being about 200 days. In the southern portion of the Cotton Belt the last killing frost comes about Mar. 10, on the average, and the first killing frost in the fall about Nov. 25. This gives a growing season of about 260 days."

Then suitability of different lands for cotton production was dynamic, not static. Brown (1937, p. 226) noted how the boll weevil invasion changed the relative productivity advantages within the cotton belt: "The extra length of frostless season in the southern part of the belt, however, is of very little consequence, on account of the prevalence of the boll weevil. No cotton is produced after the first of August, as a rule. Larger yields are to be obtained where the growing season is long, but as weevil infestation is heavy during the latter half of the growing season in most areas of the belt, only early developing varieties can be used advantageously — varieties that will mature bolls beyond weevil damage within a period of 110 days." A longer growing season had been advantageous before the boll weevil invasion, but ceased to be so afterwards as the pest population grew during the course of the season (see Lange, Olmstead, and Rhode 2009).

The USDA chief geographer, Oliver Baker (1927b, pp. 65-66) added that the boll weevil pushed the cotton belt two to three counties inland off the coast. These counties became "humid subtropical crop belt" producing vegetables and fruits. He noted the cotton belt's "northern boundary, which closely follows the line of 200-day average frost-free season, and of 77° F. average summer temperature, is remarkably distinct along most of its length…. The proportion of crop land in cotton along this boundary increases from practically nothing to 20 and 30 per cent within the width, usually, of a county. Owing to the spread of the boll weevil and the high price of cotton, this northern border has been pushed north in recent years from 20 miles in the eastern portion to 100 miles in northwestern Texas." The boll weevil pushed the cotton belt north, but not as far north as the GAEZ indices placed it. But also note how sharply Baker defines the northern border. (See also Baker 1936, Figures 90, 107, 116).

As noted above, agronomists today cumulate DD_{60} to measure the suitability of locations and crop seasons for growing cotton. Figure 6 presents a map of Cumulated DD_{60} above 2200 threshold from the USPest.org website. It correlates with the canonical cotton belt very well. Soil scientists consider cotton a crop that thrives in thermic soils (as opposed to mesic soils, where maize and spring small grains grow, and frigid soils where winter small grains grow.) Maps of thermic soils match the canonical cotton belt as well (D'Avello, Bathgate, and Thompson 2021).[14]

A glance back at the Figure 5, Panels A and C show, by way of contrast, there is little relation between the cotton suitability index and where cotton was grown historically, certainly not in 1860. In that year, the cotton suitability measure (using rainfed water supplies and intermediate levels of inputs) and the quantity of cotton produced per county acreage are extremely weakly correlated: 0.065 for the nation as a whole. If the sample is restricted to the southern states (as defined by the legality of slavery), the correlation falls to 0.003, basically nil. If one weighs by county's land area, the correlation rises to 0.132 for the nation as a whole. But for the South, it falls to -0.033, that is, negatively correlated![15]

If the cotton suitability zone were as far north as the GAEZ data place it, there are major implications for US history. Many major issues require rethinking.[16] A key conflict in the antebellum period was the territorial expansion of slavery. Why, if the heart of the cotton zone

---

[14] Thermic soils have a mean annual temperature between 15 and 22°C. Mesic soils have mean temperatures between 8 and 15°C; frigid soils have mean temperatures below 8°C. "Soil Temperature Regime." Plant & Soil Sciences eLibrary.https://passel2.unl.edu/view/lesson/69c7561e50b3/11#:~:text=The%20thermic%20soil%20temperature%20regime,50%20cm%20below%20the%20surface.

[15] If one uses the Deep Roots measures of cotton suitability, the correlation is 0.054 in the nation as a whole and 0.089 in the slavery legal region. The correlations weighing for county land area are 0.173 and 0.064, respectively. Very weak association. Wawro and Katznelson (2020, pp. 308-09) criticize the use of GAEZ cotton suitability indices by Acharya, Blackwell, and Sen on different grounds, namely the indices reflect potential endogeneity and authors' selective treatment of the data.

[16] One should, of course, keep an open mind. An enterprising soul, perhaps one of the researchers using the GAEZ indices, might move to Iowa, establish a commercially successful cotton growing operation at scale there, and prove the many generations of naysayers wrong. Of course, the climate has changed over recent decades.

was Missouri, Illinois, Indiana, southern Iowa, eastern Kansas, and eastern Nebraska, why was cotton production in the region restricted to the Bootheel region in southeastern Missouri and the southern-most counties of Illinois and Indiana? Why was so much corn and wheat grown elsewhere in the region? Cotton, after all, produced more value per acre than corn or wheat. And cotton had high value to weight, making shipment more feasible than say corn. Midwestern farmers could not have been stuck in their old, traditional ways, as they had just moved these new lands and started to put them under grain cultivation. And southern planters with experience with cotton growing were also on the move. Why did they ignore these economic opportunities?

There are political implications as well. If the region on the GAEZ map was highly suitable for cotton production, why did southern planters accept the Missouri Compromise of 1820? This allowed Missouri to enter the Union as a slave state but set the boundary for the westward expansion of slavery at 36°30' latitude, Missouri's border with Arkansas. Why did Illinois voters reject the introduction of slavery in the early 1820s? In the contest between material self-interest and ideological dispositions, were material self-interests so weak? In the late 1830s and early 1840s, why did southern expansionists seek to acquire Texas to maintain their region's monopoly on good cotton lands, but not seek the areas to the north? Why, when Kansas and Nebraska opened up for settlement in the 1850s, did so few cotton-growing slaveholders move there? Why, when advancing his popular sovereignty argument, did Stephen Douglas (and others) claim that Congress need not forbid plantation slavery from Kansas and Nebraska, God already did so by creating an unsuitable environment? In the late 1850s, how could southern planters have engaged in King Cotton thinking, based on the idea they had a monopoly on suitable cotton land? How could they base their effort to secede from the Union so heavily on such thinking if so much suitable land lie to the north?

Placing the cotton suitability zone where the GAEZ does creating unfathomable difficulties for standard interpretations of US history. We need not tarry and quibble with any existing article or book using the data. If the GAEZ cotton map is right, we should move onto directly to re-asking all the big questions in US history about economic and political attitudes and behavior.

Before proceeding in the grand exercise in re-interpretation, however, let us consider the test provided by the experiences during the American Civil War. The cutoff of southern cotton supplies led to a quadrupling in the real price of raw cotton; in wholesale markets, cotton price rose by 7.1 times relative to wheat over the 1860-64 period. This induced experiments with cotton cultivation worldwide. Cotton cultivation was tried across the US North. But was there a sustained cotton increase in, say, Iowa or Illinois during and after the War?

The US Census reported no cotton output in Iowa in 1850 and 1860 (Haines, Fishback, and Rhode 2018). What happened later? The 1870 Census reported… no cotton output in Iowa. The 1880 Census, no cotton; and the two-volume special report on cotton cultivation (which included California and Utah) left Iowa out. The 1890 Census, no cotton. The Census reported no cotton in Illinois in 1850, then 1,482 bales in 1860. After the war, there were 465 bales produced in 1870, but none reported in 1880 or 1890.

In 1865, Iowa State Agricultural Society conducted a survey of county agricultural societies, including questions about the responses of local farmers to the American Civil War (Schaffer 1866, pp. 369-547). Question 11 inquired directly about raising southern staples such as cotton, flax, hemp, and tobacco. The survey correspondents uniformly answered the inquiry about cotton negatively, stating that either that, if attempted, the crop was not a success, or that anticipating failure, local farmers did not even bother to try.

Here is a sampling of the responses to Question 11 by county in Iowa: Clinton county, p. 389: cotton was "not raised"; Cerro Gordo, p. 392: "But little attention has been paid to cultivation of hemp, falx, cotton, or tobacco."; Crawford, p. 396: "almost untried"; Davis, p. 414 :"Very little raised… Cotton is raised by some for home consumption, but will not pay as a general thing."; Dubuque, p. 423: "could not raise cotton if I wanted to…"; Jefferson, pp. 479-81: "Cotton has matured, and is grown as a curiosity; not to supply the most trifling demand of the household. Large quantities of prime seed from North Carolina were distributed last spring, but produced no results worthy of note."; Linn, p. 498 "I do not know of any attempt"; Marshall, p. 505: "… has not been tested"; Poweshiek, p. 516: "There is no cotton"; Taylor, p. 521: "only been raised by way of experiment"; Union, p. 525: "Cotton, none raised."; Van Buren, pp. 530-31: "The cultivation of cotton, hemp, flax and tobacco are articles our farmers do not waste time with, especially cotton and hemp"; Washington, p. 534: "Hemp, flax, cotton, and tobacco have not been cultivated."[17]

In Illinois, farms in the counties at the southern tip had long grown cotton crops on a small scale. During the American Civil War, Illinois Central Railroad provides experimental cottonseed to other farmers in the state. In 1864, the Illinois State Agricultural Society conducted a survey similar to that conducted the next year in Iowa (Reynolds 1865). Here is a sampling of the responses by county: Adams "cotton not grown"; Bureau "little raised"; Hardin "About 10,000 pounds of cotton raised. Previous to this time there was not exceeding 500 pounds of cotton raised in the county." Jefferson "increased production" Jo Daviess: the head of the country society was going to offer prize "but I have not time"; La Salle "Flax, hemp and cotton are not raised extensively" Randoph: "More than twenty years ago cotton gins were common in, and cotton was shipped to a considerable extent, from the county; but its cultivation, except in patches for home use, had been abandoned, until the extreme high prices of cotton during the past four years caused our enterprising farmers to resume its cultivation on an extended scale. But the late reduction in price and the ravages of the worm that injured very much of it last year (1864) has caused a great decrease in the amount planted this spring." Small amounts of cotton continued to be grown in southern Illinois during the next decade, without lasting effects. A 1924 State Experiment State publication found it reasonable to begin: "Cotton growing in a new enterprise in Illinois (Evans, Hackelman, and Bauer,1924)."

The implication of this discussion is that the GAEZ cotton map is simply wrong.

The placement of the FAO-GAEZ suitability zones for wheat and corn are also questionable. The zone for corn suitability appears too far south, in hotter places, and that for wheat too far east. The Iowa corn suitability ranking index, (CSR2), produced by Iowa State University, ranks the northern counties of that state higher than the southern counties whereas the FAO-GAEZ does the reverse.

The FAO-GAEZ wheat numbers are akin to generic index for all the non-southern crops. According to the GAEZ, the best area for growing wheat is the classic corn belt. The

---

[17] A year earlier (in Shaffer 1865, p. 364), the correspondent for Taylor county, Iowa reported that recent experiments showed cotton yielded "a good one half crop can be had in ordinary years".

wheat map is a lot like the mid-nineteenth century observer, John Klippart, wrote about-- before the introduction of the hard reds (see Olmstead and Rhode 2011). There is not much power to differentiate wheat from corn, or from anything else.

Commentary

I am not suggesting the USDA historical crop belt maps or Census of Social Statistics 1860 Usual Average crop data can be used in place of the GAEZ data or even employed to derive a reliable substitute. I would argue that differential levels of investments in the improvement of specific crops and the inherent instability of the agricultural environment make it well-nigh impossible to define for all time what crops can be grown where. Breeders can modify some crop characteristic more easily than others. For example, cotton can be turned from short-day perennial scrub to a day-light neutral plant grown as an annual; but it has proved far more difficult to breed in resistance to frost. Neither I nor the GAEZ creators, frankly, possess this knowledge. One of the inhibitions to making the present investigation has been recurrent demands to provide a replacement. I cannot at present.

I instead simply assert the evidence present here indicates the GAEZ indices are not suitable for use in historical studies of American agricultural development.[18] In cross section, the GAEZ makes the regions suitable for wheat and corn barely distinguishable. It situates the cotton belt in the wrong place, too far north by several hundred miles. In time series, the GAEZ does not reflect important changes in how crops are grown, and fails to correctly model the changing constraints on crop production. GAEZ fails, for cotton, even to incorporate the most relevant climatic data in the model. And the GAEZ indices fail to predict historical US yields for the major staple crops coherently or consistently.

A large body of literature in economics treats the GAEZ crop suitability indices as measures of the extrinsic potential of global locations to produce specific crops, unaffected by human investments in plant breeding or by shocks due to the inherent instability of the biological environment. The data, which were created for entirely different purposes, are treated as an undistorted lens into the past. In the words of Ernest Hemingway at the close of *The Sun Also Rises*: "Isn't it pretty to think so?" Sadly, it isn’t so.

---

[18] For a measure of the extend to land suitable for arable farming, Ramankutty et al. (2002) and related work is an alternative.

References

Acharya, Avidit; Blackwell, Matthew; Sen, Maya. 2016. "Replication Data for: The Political Legacy of American Slavery", https://doi.org/10.7910/DVN/CAEEG7, Harvard Dataverse, V1UNF:6:ZKw2EJymJa5N2yuM3bxjog== [fileUNF]

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2018. *Deep Roots: How Slavery Still Shapes Southern Politics*. Princeton, NJ: Princeton Univ. Press.

Alston, Julian M.; Matthew A. Andersen; Jennifer S. James; and Philip G. Pardey. 2010. *Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R&D Spending*. New York: Springer.

Atack, Jeremy, and Fred Bateman. 1987. *To Their Own Soil: Agriculture in the Antebellum North*. Ames, IA: Iowa State Univ. Press.

Baker, Oliver E. 1927a. "Agricultural Regions of North America. Part II--The South." *Economic Geography* 3, no. 1 (Jan.): 50-86.

Baker, Oliver E. 1927b. "Agricultural Regions of North America. Part IV--The Corn Belt." *Economic Geography* 3, no. 4 (Oct.): 447-65.

Baker, Oliver E. 1936. *Atlas of American Agriculture. Physical Basis including Land Relief, Climate, Soils, and Natural Vegetation of the United States*. Washington, DC: GPO.

Brooks, C. P. 1898. *Cotton: its uses, varieties, fibre structure, culture, and preparation for the market*. New York: Spon & Chamberlain.

Brown, Harry. 1927. *Cotton: history, species, varieties, morphology, breeding, culture, diseases, marketing, and uses.* New York: McGraw-Hill.

Chaplin, Joyce. 1993. *An Anxious Pursuit: Agricultural Innovation and Modernity in the Lower South, 1730-1815*. Chapel Hill: Univ. of North Carolina Press.

Collings, Gilbeart. 1926. *Production of Cotton*. New York: Wiley.

Chang, Jen-Hu. 1968. *Climate and Agriculture: An Ecological Survey.* Chicago: Aldine.

D'Avello, Tom; Jon Bathgate; and James A. Thompson. 2021. "Modeling the thermic soil temperature regime boundary of the eastern United States." *Soil Science Society of America Journal* 85: 2100-2114. https://doi.org/10.1002/saj2.20320

Duvick, Donald N. 1984. "Genetic Contributions to Yield Gains of U.S. Hybrid Maize, 1930 to 1980." in *Genetic Contributions to Yield Gains of Five Major Crop Plants*. Madison, WI: ASA-CSSA.

Duvick, Donald N. 1997. "What is Yield?" in *Developing Drought- and Low N-Tolerant Maize*. G. O. Edmeades, M. Banziger, H. R. Mickelson, and C. B. Pena–Valdivia, eds. Mexico, DF: CIMMYT, pp. 332–35.

Duvick, Donald N. 1999. "Heterosis: Feeding People and Protecting Natural Resources." in *Genetics and Exploitation of Heterosis in Crops*. Madison, WI: ASA-CSSA-SSSA.

Duvick, Donald N. 2005. "The Contributions of Breeding to Yield Advances in Maize (*Zea Mays* L.)" Advances in Agronomy 86, pp. 83-145.

Evans, Lloyd T. 1993. *Crop Evolution, Adaptation and Yield.* New York: Cambridge Univ. Press.

Evans, J. A.; J. C. Hackleman; and F. C. Bauer. 1924. *Cotton growing in Illinois*. Univ. of Illinois Agricultural College and Experiment Station Circular No. 279. Urbana, IL.

Fischer, Günther; Freddy O. Nachtergaele, Sylvia Prieler, Edmar Teixeira, Géza Tóth, Harrij van Velthuizen, Luc Verelst, David Wiberg. 2012. *Global Agro-Ecological Zones (GAEZ v3) – Model documentation*. Rome, FAO.

Fischer, G., Nachtergaele, F.O., van Velthuizen, H.T., Chiozza, F., Franceschini, G., Henry, M., Muchoney, D. and Tramberend, S. 2021. *Global Agro-Ecological Zones v4 – Model documentation*. Rome, FAO.

Gray, Lewis C. 1933. *History of Agriculture in the Southern United States to 1860*. Washington, DC: Carnegie Institute.

Griliches, Zvi. 1957. "Hybrid corn: An exploration in the economics of technological change." *Econometrica* pp. 501–22.

Griliches, Zvi. 1958. "Research costs and social returns: Hybrid corn and related innovations." *Journal of Political Economy* 66, no. 5, pp. 419–31.

Griliches, Zvi. 1960. "Hybrid corn and the economics of innovation." *Science 132* (3422), pp. 275–80.

Haines, Michael, Price Fishback, and Paul Rhode, 2018. *United States Agriculture Data, 1840-2012* (ICPSR 35206). Inter-university Consortium for Political and Social Research.

Hay, RKM. 1995. "Harvest index: a review of its use in plant breeding and crop physiology." *Annals of Applied Biology* 126, no. 1, pp. 197-216.

Hay, RKM, and R. A. Gilbert. 2001. "Variation in the harvest index of tropical maize: evaluation of recent evidence from Mexico and Malawi." *Annals of Applied Biology* 138, no. 1, pp. 103-09.

Hemingway, Ernest. 1926. *The Sun Also Rises*.

Hornbeck, Richard, and Pinar Keskin. 2014. "The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought." *American Economic Journal: Applied Economics* 6, no. 1 (Jan.): 190–219. https://doi.org/10.1257/app.6.1.190.

Hornbeck, Richard, and Pinar Keskin. 2019. Replication data for: The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought. Nashville, TN: American Economic Association [publisher], 2014. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-12. https://doi.org/10.3886/E113876V1

Hornbeck, Richard, and Suresh Naidu. 2014. “When the Levee Breaks: Black Migration and Economic Development in the American South.” *American Economic Review* 104, no. 3 (March 2014): 963–90. https://doi.org/10.1257/aer.104.3.963.

Hornbeck, Richard, and Suresh Naidu. 2019. Replication data for: When the Levee Breaks: Black Migration and Economic Development in the American South. Nashville, TN: American Economic Association [publisher], 2014. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-11. https://doi.org/10.3886/E112749V1

Hutsch, Birgit, and Sven Schubert. 2017. “Harvest Index of Maize (Zea mays L.): Are There Possibilities for Improvement?“ *Advances in Agronomy* (Jan.): 37-82.

IIASA/FAO, 2012. *Global Agro-ecological Zones (GAEZ v3.0).* IIASA, Laxenburg, Austria and FAO, Rome, Italy.

IIASA/FAO, 2021. *Global Agro Ecological Zones version 4* (GAEZ v4). IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Jorgenson, D. W., and Z. Griliches. 1967. “The Explanation of Productivity Change.” *Review of Economic Studies* 34, no. 3 (July): 249-83.

Kantor, Shawn, and Alexander Whalley. 2019. “Research Proximity and Productivity: Long-Term Evidence from Agriculture.” *Journal of Political Economy*. 127, no, 2 (April): 817-54.

Lange, Fabian, Alan L. Olmstead, and Paul W. Rhode. 2009. “The Impact of the Boll Weevil, 1892-1932” *Journal of Economic History* 69, no. 3 (Sept.): 685-718.

National Cotton Council of America. n.d. “Advancing Cotton Education: Growth and Development of a Cotton Plant.” Downloads Feb. 25, 2024 https://www.cotton.org/tech/ace/growth-and-development.cfm?csSearch=123100_1

Olmstead, Alan L., and Paul W. Rhode 2008. *Creating Abundance: Biological Innovation and American Agricultural Development*. Cambridge Univ. Press.

Olmstead, Alan L., and Paul W. Rhode 2011. "Responding to Climatic Challenges: Lessons from U.S. Agricultural Development," in Gary D. Libecap and Richard H. Steckel, eds. *Climate Change Past and Present: Uncertainty and Adaptation*. Chicago, Univ. of Chicago Press, pp. 169-94.

Parker, William N., and Judith V. Klein. 1966, “Productivity Grown in Grain Production in the United States, 1840-1860 and 1900-1910.” in *Output, Employment, and Productivity in the United States after 1800*. Studies in Income and Wealth Vo. 30 Princeton, NJ: Princeton Univ. Press.

Reynolds, John P. 1865. *Transactions of Illinois State Agricultural Society, Vol. V, 1861-64*. Springfield, IL: Baker & Phillips.

Ramankutty, N., J.A. Foley, J. Norman, and K. McSweeney. 2002. “The global distribution of cultivable lands: current patterns and sensitivity to possible climate change.” *Global Ecology and Biogeography* 11, no. 5 (Sept.): 377-92.

Russell, W. A. 1991. “Genetic Improvement in Maize Yields.” *Advances in Agronomy*. 46 pp. 245-98.

Slater, Gustavo A., Emilio H. Satorre, and Fernado H. Andrade. 1994. "Increases in Grain Yield in Bread Wheat from Breeding and Associated Physiological Changes" in *Genetic Improvement in Field Crops*, Gustavo A Slater, ed. New York: Marcel Dekker, pp. 1-68

Schaffer, J. M. 1866. *Report of the Secretary of the Iowa State Agricultural Society, for the year 1865*, vol. 11. Des Moines, IA: F. W. Plamer, State Printer.

Schaffer, J. M. 1865. *Report of the Secretary of the Iowa State Agricultural Society, for the year 1864*, vol. 10. Des Moines, IA: F. W. Plamer, State Printer.

Stine, O. C., and O. E. Baker. 1918. *Atlas of American Agriculture, Part V. The Crops. Section A. Cotton*. Advance Sheets 4. Washington, DC: GPO.

Tharp, William H. 1960. *The cotton plant: how it grows and why its growth varies*. Agricultural Research Service Agricultural Handbook No. 178. Washington, DC: U.S. Dept. of Agriculture.

Tian, Zhan, Hanqing Xu, Laixiang Sun, Dongli Fan, Günther Fischer, Honglin Zhong, Peiqun Zhang, Edward Pope, Chris Kent, Wei Wu. 2020. "Using a cross-scale simulation tool to assess future maize production under multiple climate change scenarios: An application to the Northeast Farming Region of China" *Climate Services* 18, 100150.

Tollenaar, Matthijs, Desmond E. McCullough, and Lianne M. Dwyer. 1994. "Physiological basis of the Genetic Improvement in Corn" in *Genetic Improvement in Field Crops*, Gustavo A Slater, ed. New York: Marcel Dekker, pp. 183-236.

Toman, Jane. 2002. "Plantation Labor Management on United States Cotton Plantations in the Antebellum South." Yale Univ. Economics. PhD dissertation.

US Census Office. 1854. *Statistical View of the United States… Being a Compendium of the Seventh Census*. Washington, DC: GPO.

US Census Office. 1884. *Tenth Census, Vol, 5-6, Report on Cotton Production in the United States: Also Embracing Agricultural and Physico-geographical Descriptions of the Several Cotton States and of California*. Washington, DC: GPO.

US Department of Agriculture. National Resource Conservation Service. 2022. *User Guide for National Commodity Crop Productivity Index (NCCPI) Version 3.0*. Washington, DC: USDA.

USPest. n.d. OSU IPPC GRASSLinks 3.5beta: A web interface for GRASS GIS version 5.x https://uspest.org/glinks35/

Wawro, Gregory J., and Ira Katznelson. 2020. "American political development and new challenges of causal inference." *Public Choice* 185, no. 3 (Dec.): 299-314.

Whitney, Milton. 1896. "Climatology and Soils" chapter in A. True, ed., *Cotton Plant*" *Its History, Botany, Chemistry, Culture, Enemies, and Uses*. Washington, DC: USDA.

Wright, Carroll D. 1900. *The History and Growth of the United States Census*. Washington, DC: GPO.

Appendix 1

The italicized test below reproduces the GAEZ language about validating the parameters:

Various modes have been pursued for "ground-truthing" and verifying results of the GAEZ suitability analysis. Apart from consulting expert knowledge and agricultural research institutes, results have been systematically compared with research data and agricultural statistics. In particular the following activities have been conducted intensively by IIASA and staff of FAO's Economic and Social Department and its Agricultural Department.

Confirmation of estimated potential crop distribution and yields against quantitative and qualitative occurrence of these crops in national and subnational agricultural statistics.

Comparison of limits of AEZ potential crop distribution with limits to actual distribution of agricultural land (e.g., by comparison with spatial land use/land cover databases and crop distribution maps).

Various modules in GAEZ are from well tested origin such as (i) the estimation of crop water stress and related yield losses (CROPWAT), (ii) the robust biomass and yield calculation procedures (Kassam 1977) which have been applied tested and scrutinized in case studies in more than twenty countries and (iii) the agro-edaphic suitability procedures and results have benefitted of extensive verification with documented expert knowledge, resulting from numerous sub-national, national, regional and global AEZ assessments over the past three decades.

It should, however, be understood that in the light of improved knowledge, any part of the GAEZ suitability procedures and the model parameters will be scrutinized and may be subject to updating by FAO and IIASA. Also, the model and model parameters are expected to benefit from refinement as a result of follow-up applications.

The italicized text below reproduces verbatim, the description from the GAEZ documentation.

Three input level selection options are available: high level inputs, intermediate level inputs, and low level inputs.

Low-level inputs/traditional management

Under the low input, traditional management assumption, the farming system is largely subsistence based and not necessarily market oriented. Production is based on the use of traditional cultivars (if improved cultivars are used, they are treated in the same way as local cultivars), labor intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control and minimum conservation measures.

Intermediate-level inputs/improved management

Under the intermediate input, improved management assumption, the farming system is partly market oriented. Production for subsistence plus commercial sale is a management objective. Production is based on improved varieties, on manual labor with hand tools and/or animal traction and some mechanization. It is medium labor intensive, uses some fertilizer application and chemical pest, disease and weed control, adequate fallows and some conservation measures.

High-level inputs/advanced management

Under the high input, advanced management assumption, the farming system is mainly market oriented. Commercial production is a management objective. Production is based on improved high yielding varieties, is fully mechanized with low labor intensity and uses optimum applications of nutrients and chemical pest, disease and weed control.

Appendix 2

Section 1: Selective Reporting of the 1860 Usual Yield Data

The 1860 Usual yield data was selectively available. It is helpful to relate an indicator variable for availability to the GAEZ crop suitability index and the measure of crop density in 1860—production per county area. For cotton, both matters when entered together. For wheat and corn, the crop density matters.

Table A2.1: Predicting the availability of the Usual Yield data with GAEZ Crop Suitability Indices and 1860 Crop Density

	coeff. [se]	coeff. [se]	coeff. [se]
D_1860_Usual_Cotton			
Constant	0.109 [0.009]	0.195 [0.0169	0.287 [0.024]
Ln_GAEZ_Cotton_Int	0.019 [0.005]		0.071 [0.015]
Ln_Cotton_per_area		0.044 [0.006]	0.044 [0.006]
#obs	1,819	696	689
R_sq	0.007	0.082	0.109
D_1860_Usual_Wheat			
Constant	0.146 [0.021]	0.284 [0.020]	0.241 [0.032]
Ln_GAEZ_Wheat_Int	0.103 [0.012]		0.032 [0.019]
Ln_Wheat_per_area		0.049 [0.004]	0.0453 [0.005]
#obs	2,779	1,870	1,868
R_sq	0.027	0.056	0.057
D_1860_Usual_Corn			
Constant	0.235 [0.016]	0.324 [0.031]	0.313 [0.032]
Ln_Maize_GAEZ_Int	0.048 [0.008]		0.014 [0.014]
Ln_Corn_per_area		0.021 [0.005]	0.018 [0.006]
#obs	2,738	1,961	1,957
R_sq	0.014	0.009	0.009

Section 2: Regressions using GAEZ high suitability indices.

	Coeff	SE		Coeff	SE
Solo Sprints for Wheat Yields					
	GAEZ wheat high			1860 Usual wheat	
1930	0.058	0.014		0.429	0.033
1940	-0.039	0.001		0.418	0.029
1949	0.135	0.013		0.228	0.030
1964	0.168	0.011		0.177	0.024
1997	0.134	0.013		0.043	0.027
Horse Race for Wheat Yields					
	GAEZ wheat high			1860 Usual wheat	
1930	0.067	0.022		0.436	0.033
1940	0.111	0.018		0.421	0.028
1949	0.171	0.018		0.221	0.028
1964	0.170	0.016		0.168	0.023
1997	0.144	0.019		0.049	0.026
Solo Sprints for Corn Yields					
	GAEZ maize high			1860 Usual corn	
1930	-0.045	0.013		0.574	0.031
1940	-0.006	0.010		0.856	0.040
1949	0.003	0.014		0.600	0.031
1964	0.034	0.014		0.454	0.033
1997	0.070	0.012		0.320	0.021
Horse Race for Corn Yields					
	GAEZ maize high			1860 Usual corn	
1930	-0.101	0.023	high	0.603	0.031
1940	0.017	0.300		0.857	0.041
1949	-0.026	0.023		0.610	0.031
1964	0.061	0.026		0.446	0.033
1997	0.144	0.018		0.287	0.021
Horse Race for Corn Yields					
	GAEZ maize high			GAEZ wheat high	
1930	-0.386	0.017		0.652	0.024
1940	-0.250	0.015		0.457	0.021
1949	-0.379	0.018		0.718	0.024
1964	-0.383	0.020		0.655	0.025
1997	-0.198	0.018		0.393	0.021

Section 3: NASS data

The National Agricultural Statistical Service produced a dataset of annual county-level crop yields for selected states from the USDA crop reporting service for the 1919-61 period. Cotton is commodity number 12121999; wheat 10199999; and corn 11199199. The results in these series are largely similar to what is reported in the text.

	Coeff	SE	#obs	R-sq
Panel A: Ln_cotton_yields				
ln_cotton_inter	0.0193	0.0095	2,056	0.2539
ln_cotton_1860_usual	0.1626	0.0267		
Year	0.0205	0.00082		
Constant	-35.51	1.642		
Panel B: Ln_wheat_yields				
ln_wheat_inter	0.1747	0.0066	13,959	0.402
ln_wheat_1860_usual	0.1323	0.0046		
Year	0.0213	0.00023		
Constant	-39.21	0.4458		
Panel C: Ln_corn_yields				
ln_corn_inter	0.0769	0.0058	16,097	0.328
ln_corn_1860_usual	0.4935	0.0092		
Year	0.0207	0.00029		
Constant	-38.98	0.5662		
Panel D: Ln_corn_yields				
ln_corn_inter	-0.1654	0.0044	38,123	0.283
ln_wheat_inter	0.4791	0.0056		
Year	0.0174	0.00019		
Constant	-31.07	0.3622		

Figure 1: GAEZ Schema

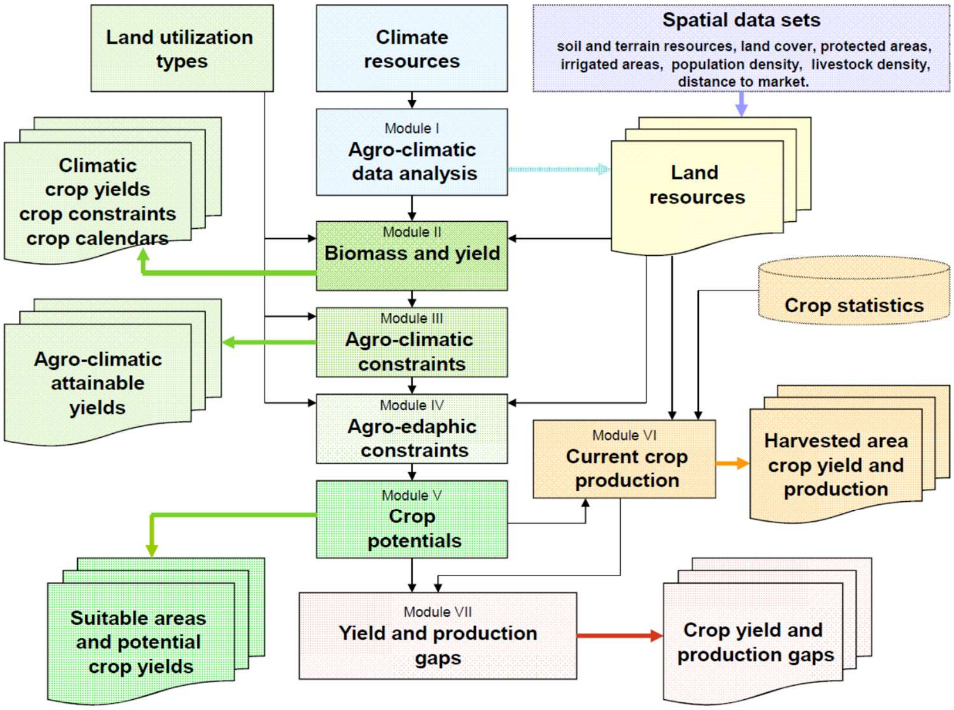
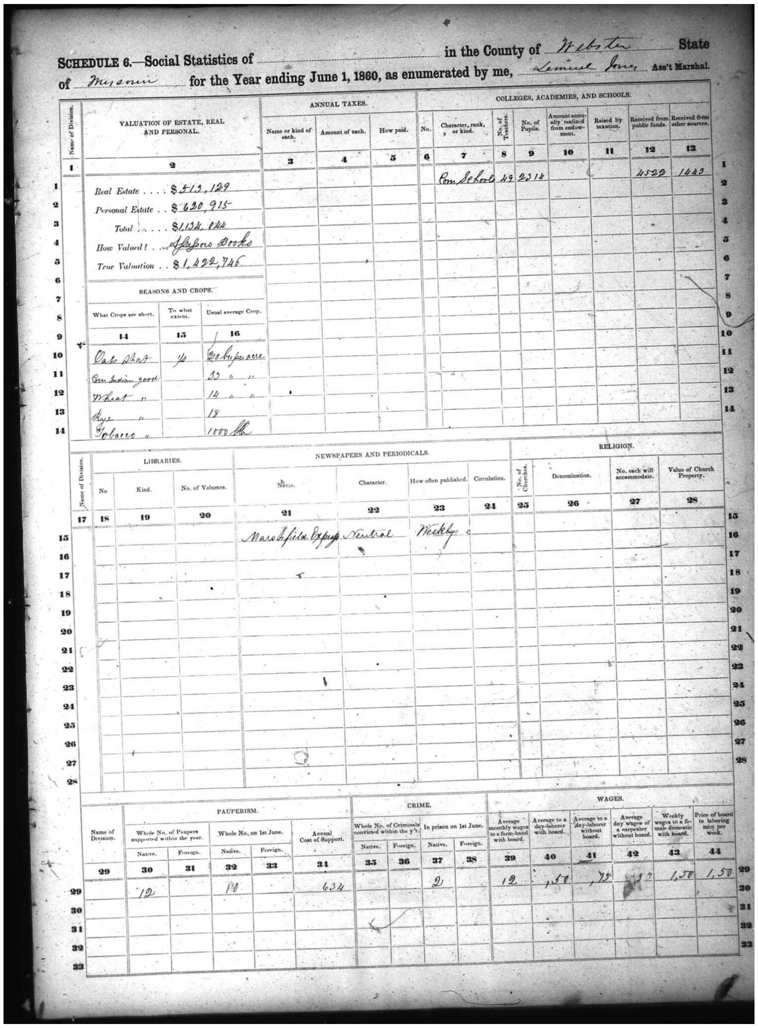


Figure 1-1 Overall structure and data integration of GAEZ v3.0 (Module I-VII)

Figure 2: Manuscript page of Census of Social Statistics, 1860



SCHEDULE 6.—Social Statistics of .... in the County of Webster State of Missouri for the Year ending June 1, 1860, as enumerated by me, Lemuel Jones Ass't Marshal.

VALUATION OF ESTATE, REAL AND PERSONAL.

Real Estate $513,129
Personal Estate . . $620,915
Total $1,134,044
How Valued? . . Assessors Books
True Valuation . . $1,422,746

COLLEGES, ACADEMIES, AND SCHOOLS.

No.	Character, rank, or kind.	No. of Teachers.	No. of Pupils.	Amount annually realized from endowment.	Raised by taxation.	Received from public funds.	Received from other sources.
6	7	8	9	10	11	12	13
	Com Schools	49	2314			4522	1443

SEASONS AND CROPS.

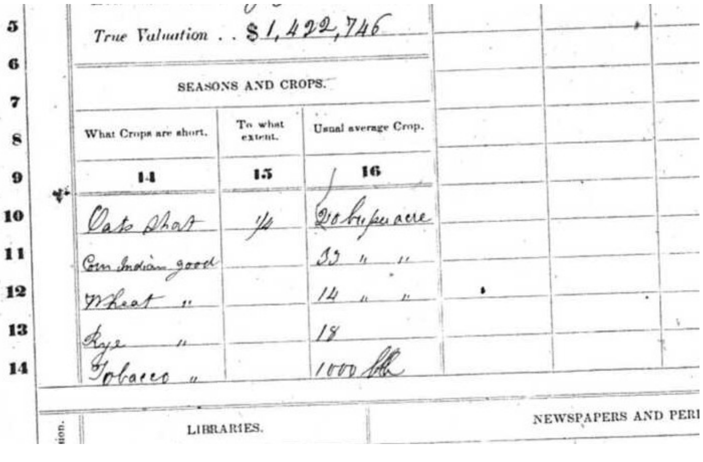
What Crops are short.	To what extent.	Usual average Crop.
14	15	16
Oats Shot	1/4	20 bu per acre
Corn Indian good		33 " "
Wheat "		14 " "
Rye "		18
Tobacco "		1000 lbs

NEWSPAPERS AND PERIODICALS.

Name.	Character.	How often published.	Circulation.
21	22	23	24
Marshfield Express	Neutral	Weekly	

PAUPERISM. CRIME. WAGES.

Name of Division.	Whole No. of Paupers supported within the year. Native.	Foreign.	Whole No. on 1st June. Native.	Foreign.	Annual Cost of Support.	Whole No. of Criminals convicted within the y'r. Native.	Foreign.	In prison on 1st June. Native.	Foreign.	Average monthly wages to a farm-hand with board.	Average to a day-laborer with board.	Average to a day-laborer without board.	Average day wages of a carpenter without board.	Weekly wages to a female domestic with board.	Price of board to laboring men per week.
29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
	12		10		634			2		12	,50	,75	[illegible]	1,50	1,50



True Valuation . . $1,422,746

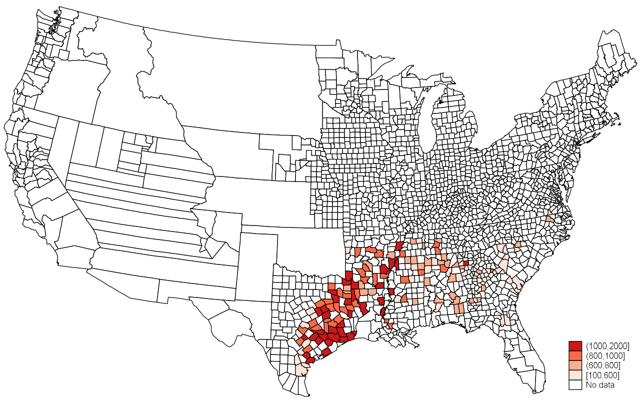
SEASONS AND CROPS.

What Crops are short.	To what extent.	Usual average Crop.
14	15	16
Oats Shot	1/4	20 bu per acre
Corn Indian good		33 " "
Wheat "		14 " "
Rye "		18
Tobacco "		1000 lbs

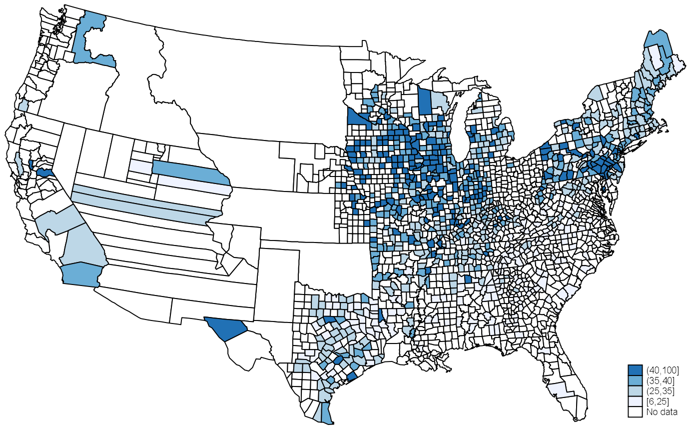
LIBRARIES. NEWSPAPERS AND PERI

Figure 3: Coverage of 1860 Usual Yield Statistics

Cotton



Corn



Wheat

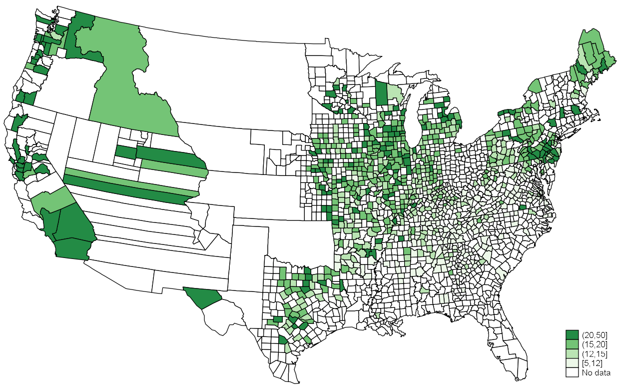
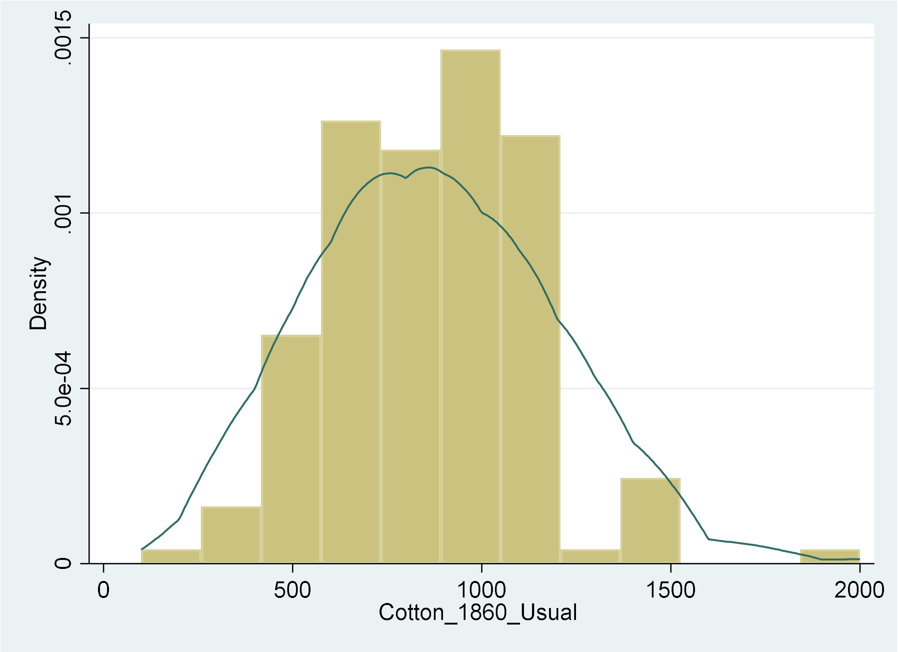


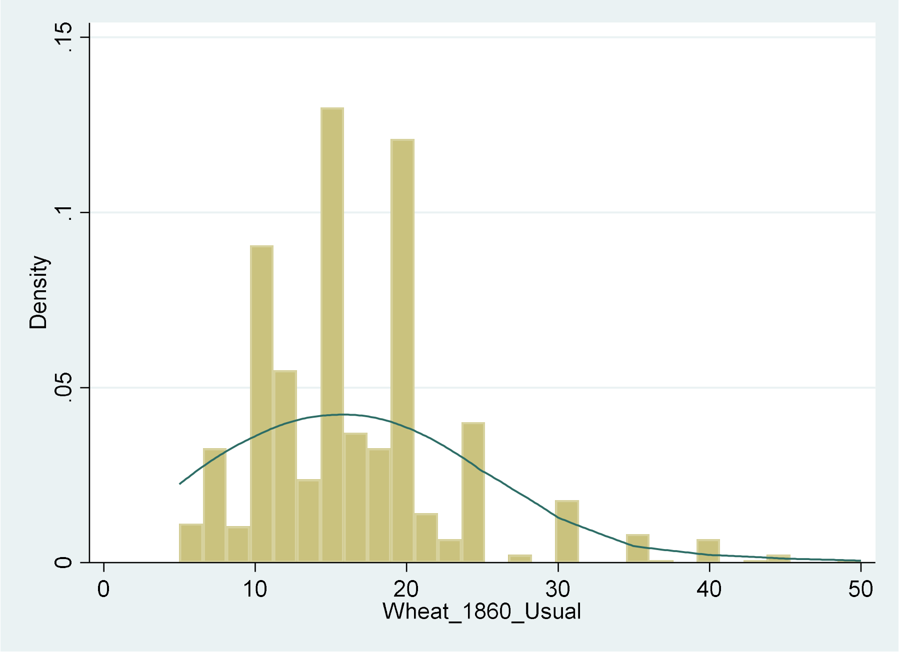
Figure 4: Histograms of Usual Yields, 1860.

Cotton: Pounds of Lint per Acres



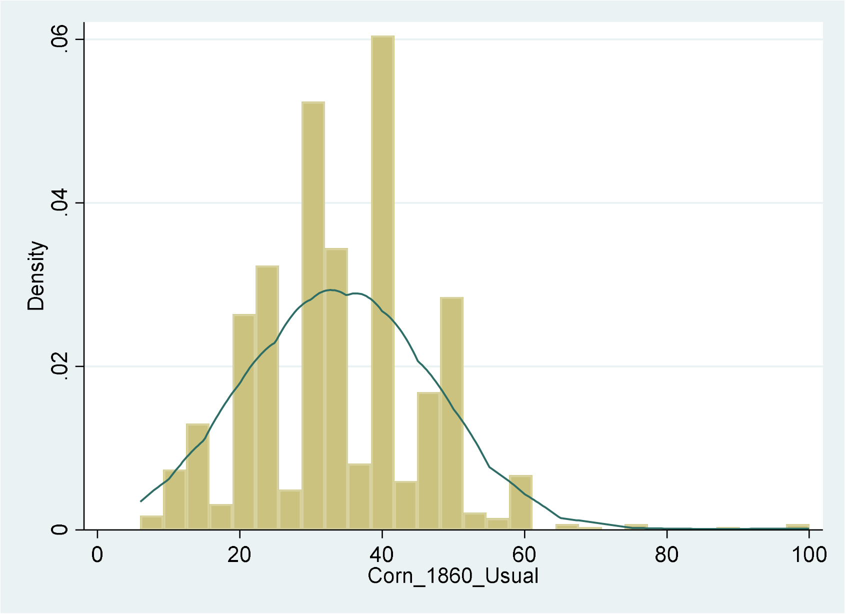
histogram cotton_1860_usual , kdensity kdenopts(width(400) epan2)

Wheat: Bushels of Grain per Acre



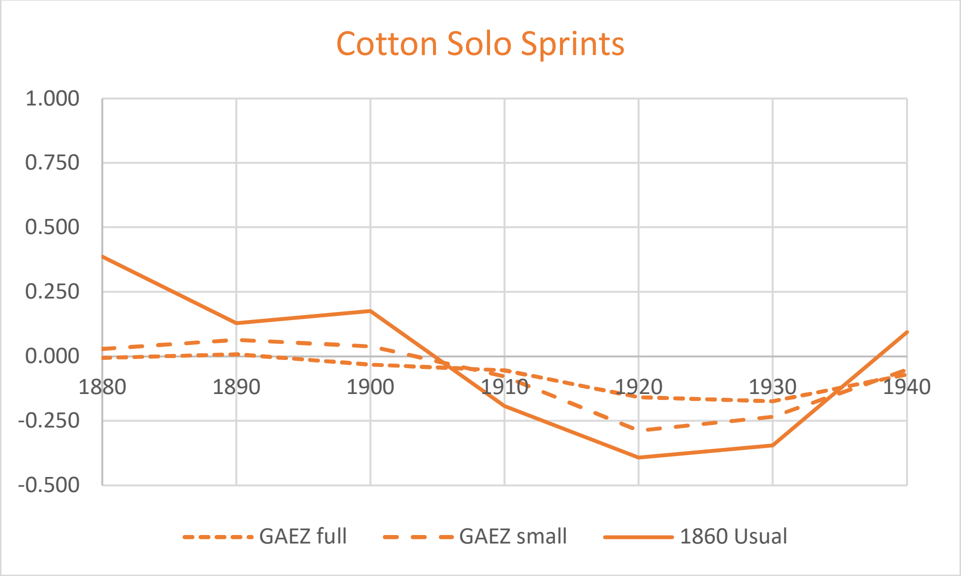
histogram wheat_1860_usual , kdensity kdenopts(width(15) epan2)

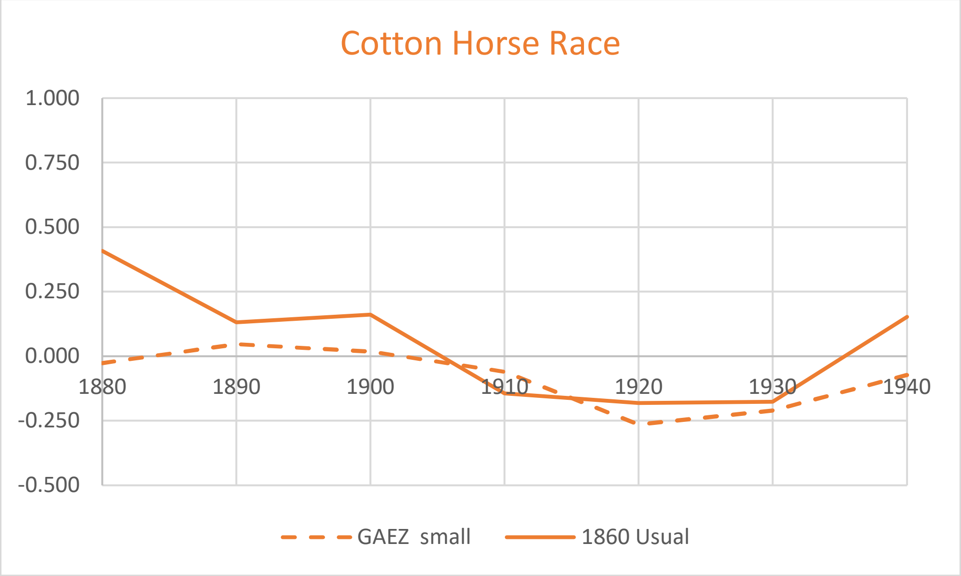
Corn: Bushels of Grain per Acre

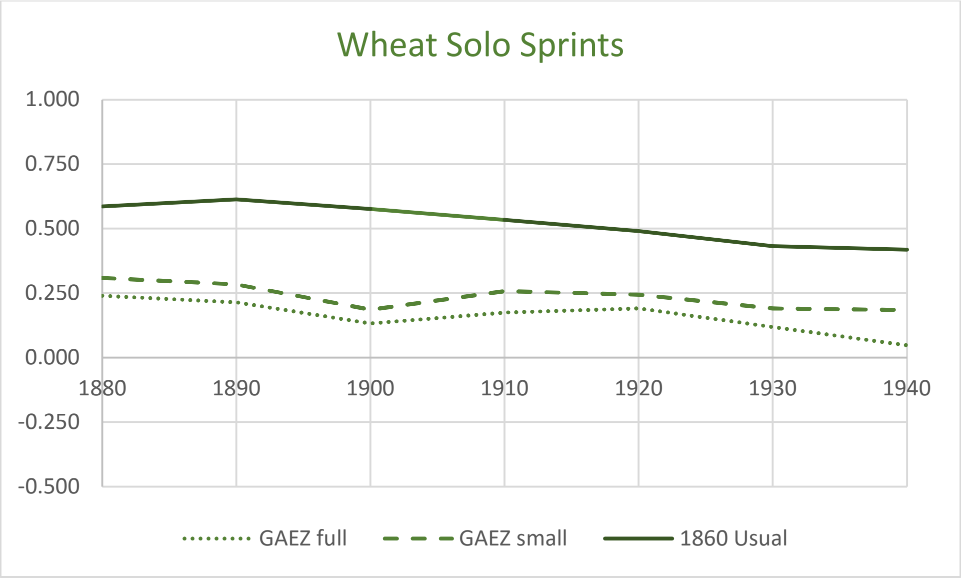


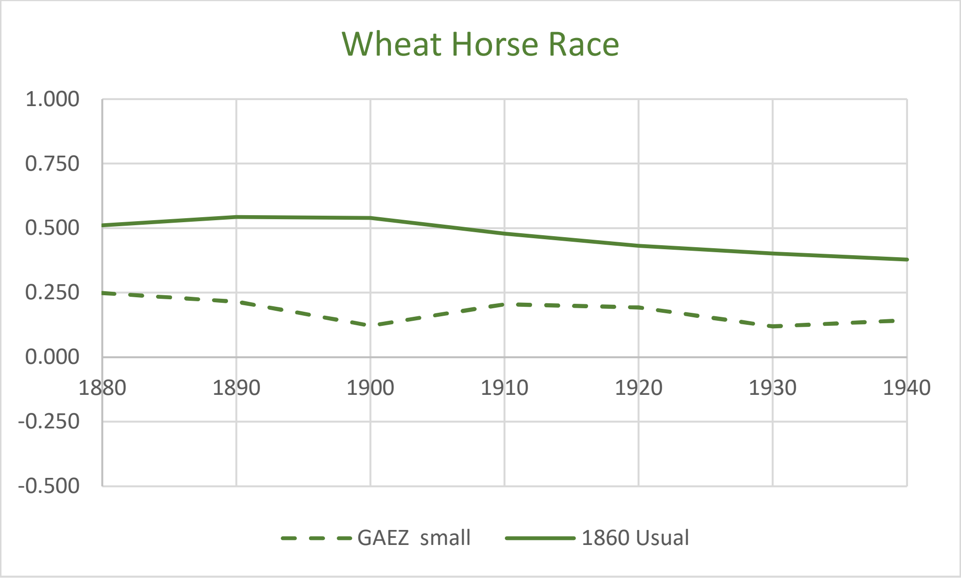
histogram corn_1860_usual , kdensity kdenopts(width(15) epan2)

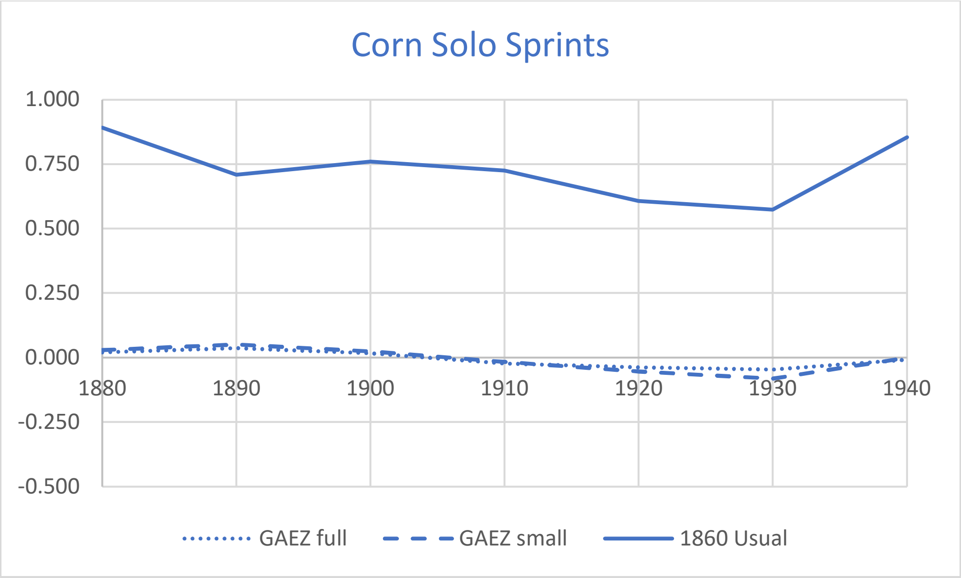
Figure 5: Graphing coefficients from historical yield regressions (without geographic controls)

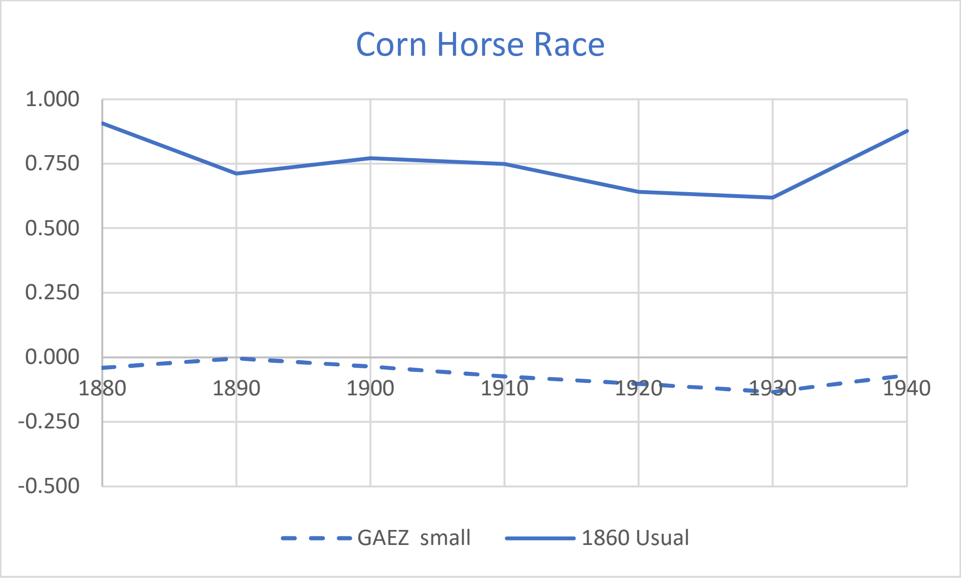





Wheat Solo Sprints
1.000
0.750
0.500
0.250
0.000
-0.250
-0.500
1880
1890
1900
1910
1920
1930
1940
GAEZ full
GAEZ small
1860 Usual


Wheat Horse Race
1.000
0.750
0.500
0.250
0.000
-0.250
-0.500
1880
1890
1900
1910
1920
1930
1940
GAEZ small
1860 Usual


Corn Solo Sprints
1.000
0.750
0.500
0.250
0.000
-0.250
-0.500
1880
1890
1900
1910
1920
1930
1940
GAEZ full
GAEZ small
1860 Usual


Corn Horse Race
1.000
0.750
0.500
0.250
0.000
-0.250
-0.500
1880
1890
1900
1910
1920
1930
1940
GAEZ small
1860 Usual

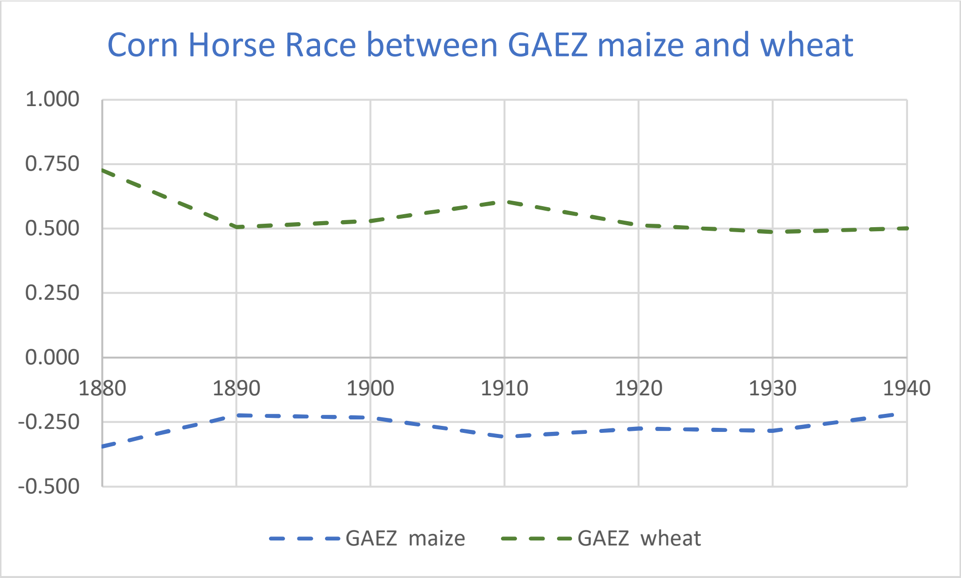
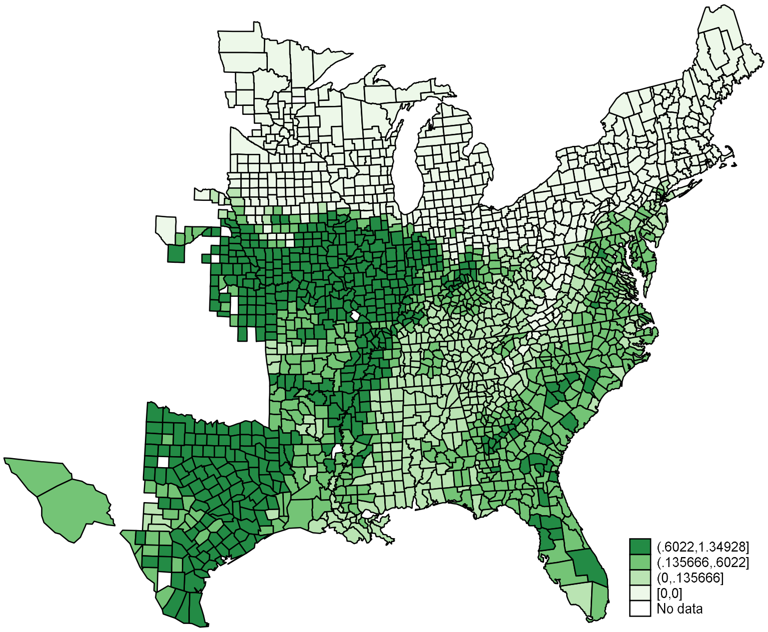
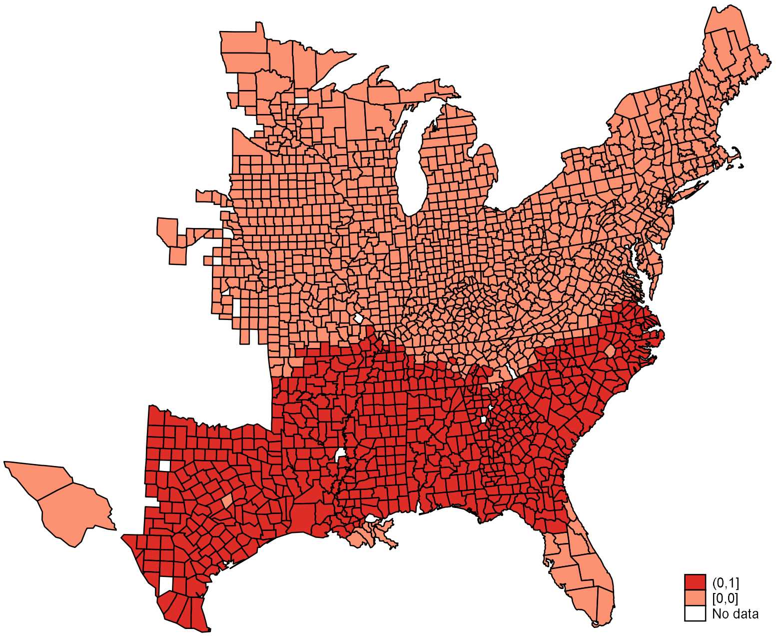

Corn Horse Race between GAEZ maize and wheat
1.000
0.750
0.500
0.250
0.000
-0.250
-0.500
1880
1890
1900
1910
1920
1930
1940
GAEZ maize
GAEZ wheat

Figure 6: Cotton

A. GAEZ cotton suitability



B. Canonical USDA Cotton Belt



C. 1860 Cotton per County Acre

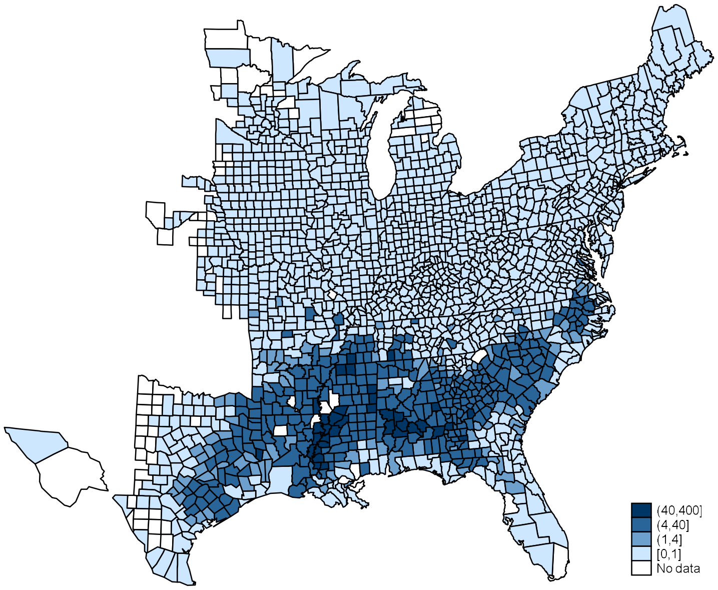
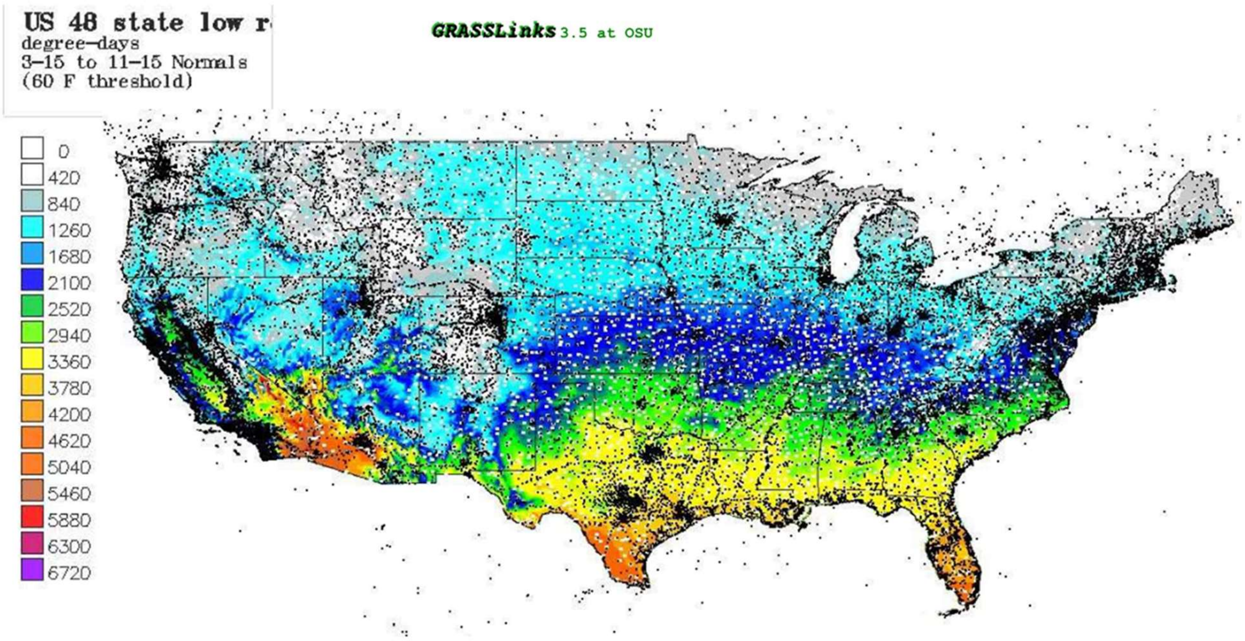
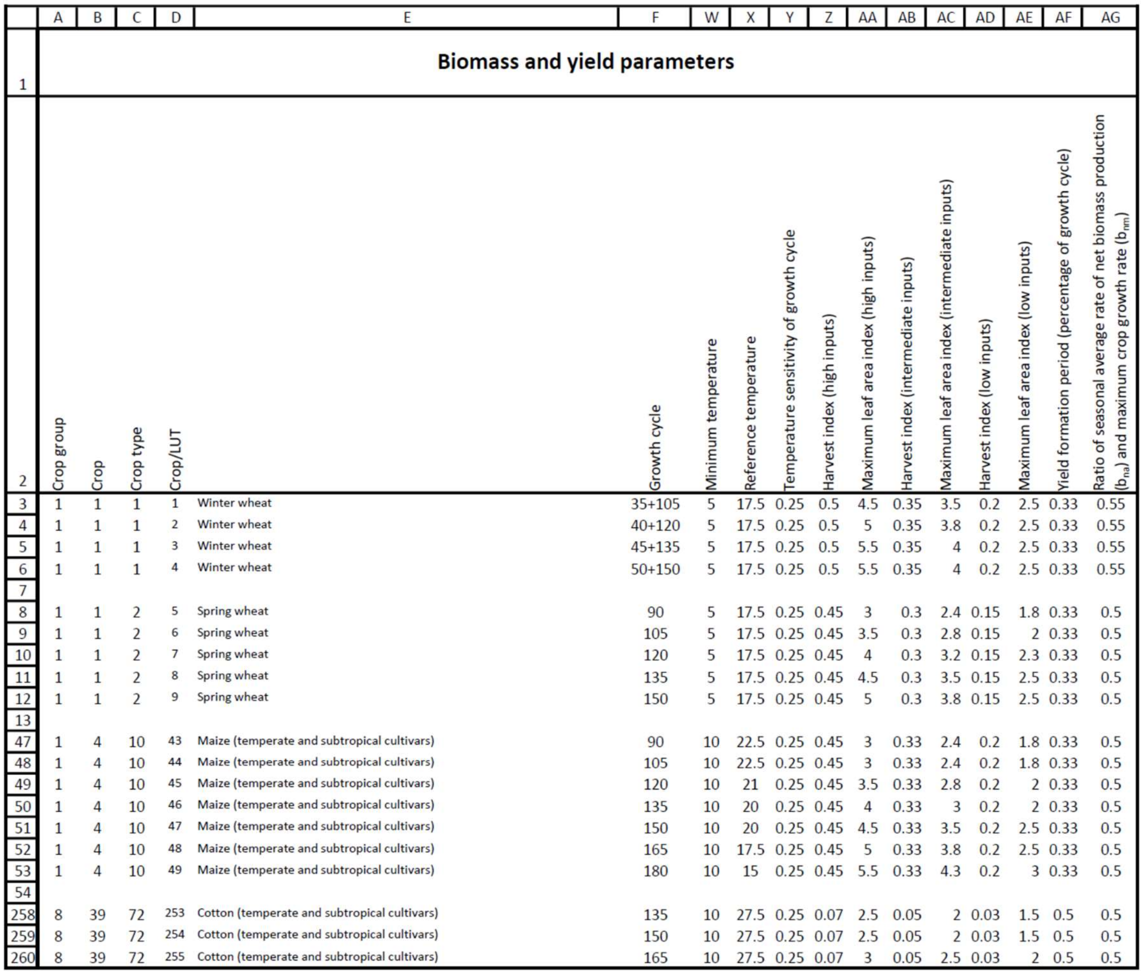


Figure 7: DD60s



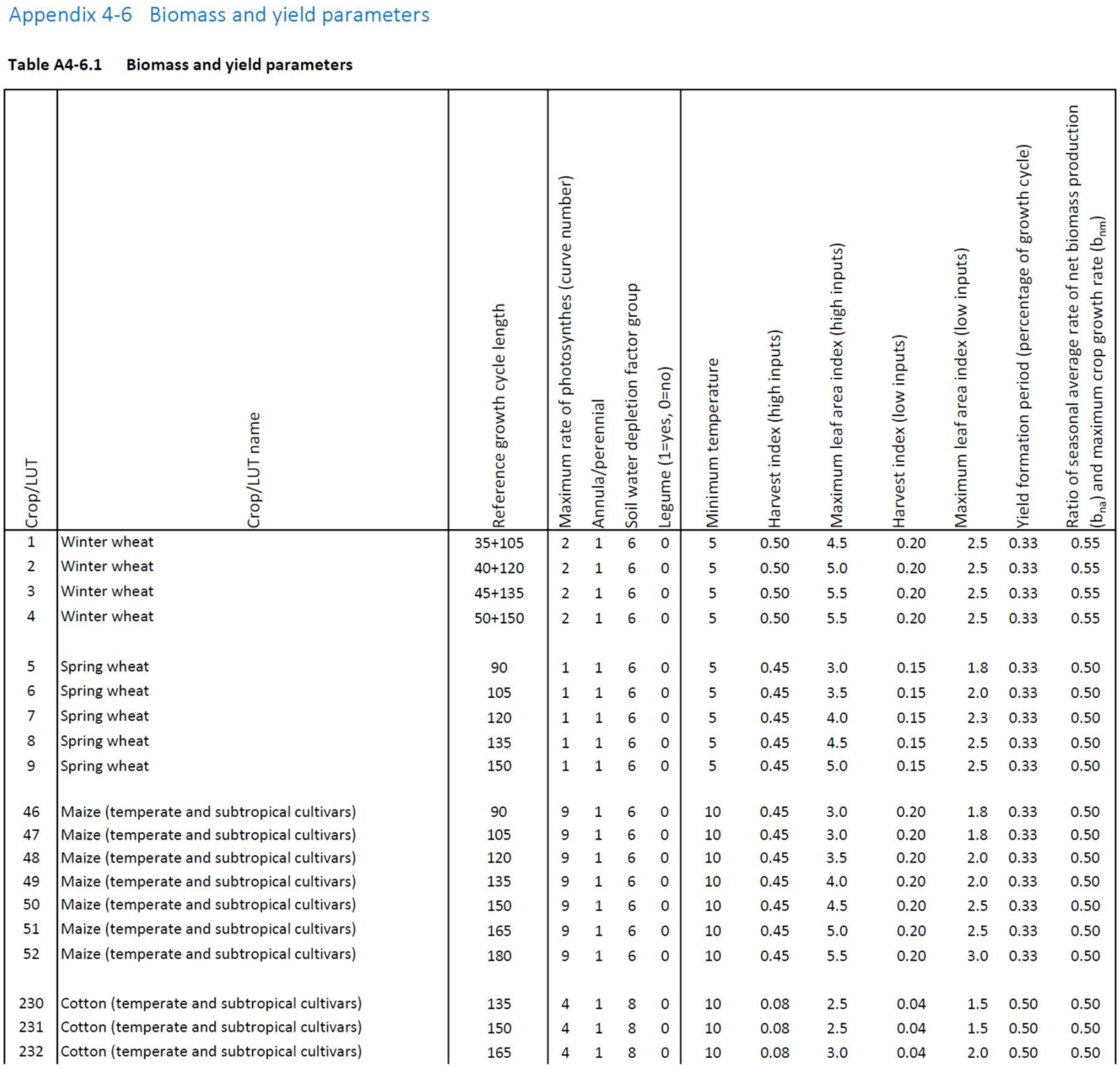
Source: USPest.org OSU IPPC GRASSLinks 3.5beta: A web interface for GRASS GIS version 5.x

Table 1: GAEZ Version 3



	A	B	C	D	E	F	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	**Biomass and yield parameters**																
2	Crop group	Crop	Crop type	Crop/LUT		Growth cycle	Minimum temperature	Reference temperature	Temperature sensitivity of growth cycle	Harvest index (high inputs)	Maximum leaf area index (high inputs)	Harvest index (intermediate inputs)	Maximum leaf area index (intermediate inputs)	Harvest index (low inputs)	Maximum leaf area index (low inputs)	Yield formation period (percentage of growth cycle)	Ratio of seasonal average rate of net biomass production (b_{na}) and maximum crop growth rate (b_{nm})
3	1	1	1	1	Winter wheat	35+105	5	17.5	0.25	0.5	4.5	0.35	3.5	0.2	2.5	0.33	0.55
4	1	1	1	2	Winter wheat	40+120	5	17.5	0.25	0.5	5	0.35	3.8	0.2	2.5	0.33	0.55
5	1	1	1	3	Winter wheat	45+135	5	17.5	0.25	0.5	5.5	0.35	4	0.2	2.5	0.33	0.55
6	1	1	1	4	Winter wheat	50+150	5	17.5	0.25	0.5	5.5	0.35	4	0.2	2.5	0.33	0.55
7																	
8	1	1	2	5	Spring wheat	90	5	17.5	0.25	0.45	3	0.3	2.4	0.15	1.8	0.33	0.5
9	1	1	2	6	Spring wheat	105	5	17.5	0.25	0.45	3.5	0.3	2.8	0.15	2	0.33	0.5
10	1	1	2	7	Spring wheat	120	5	17.5	0.25	0.45	4	0.3	3.2	0.15	2.3	0.33	0.5
11	1	1	2	8	Spring wheat	135	5	17.5	0.25	0.45	4.5	0.3	3.5	0.15	2.5	0.33	0.5
12	1	1	2	9	Spring wheat	150	5	17.5	0.25	0.45	5	0.3	3.8	0.15	2.5	0.33	0.5
13																	
47	1	4	10	43	Maize (temperate and subtropical cultivars)	90	10	22.5	0.25	0.45	3	0.33	2.4	0.2	1.8	0.33	0.5
48	1	4	10	44	Maize (temperate and subtropical cultivars)	105	10	22.5	0.25	0.45	3	0.33	2.4	0.2	1.8	0.33	0.5
49	1	4	10	45	Maize (temperate and subtropical cultivars)	120	10	21	0.25	0.45	3.5	0.33	2.8	0.2	2	0.33	0.5
50	1	4	10	46	Maize (temperate and subtropical cultivars)	135	10	20	0.25	0.45	4	0.33	3	0.2	2	0.33	0.5
51	1	4	10	47	Maize (temperate and subtropical cultivars)	150	10	20	0.25	0.45	4.5	0.33	3.5	0.2	2.5	0.33	0.5
52	1	4	10	48	Maize (temperate and subtropical cultivars)	165	10	17.5	0.25	0.45	5	0.33	3.8	0.2	2.5	0.33	0.5
53	1	4	10	49	Maize (temperate and subtropical cultivars)	180	10	15	0.25	0.45	5.5	0.33	4.3	0.2	3	0.33	0.5
54																	
258	8	39	72	253	Cotton (temperate and subtropical cultivars)	135	10	27.5	0.25	0.07	2.5	0.05	2	0.03	1.5	0.5	0.5
259	8	39	72	254	Cotton (temperate and subtropical cultivars)	150	10	27.5	0.25	0.07	2.5	0.05	2	0.03	1.5	0.5	0.5
260	8	39	72	255	Cotton (temperate and subtropical cultivars)	165	10	27.5	0.25	0.07	3	0.05	2.5	0.03	2	0.5	0.5

Table 2: GAEZ Version 4



Appendix 4-6 Biomass and yield parameters

Table A4-6.1 Biomass and yield parameters

Crop/LUT	Crop/LUT name	Reference growth cycle length	Maximum rate of photosynthes (curve number)	Annula/perennial	Soil water depletion factor group	Legume (1=yes, 0=no)	Minimum temperature	Harvest index (high inputs)	Maximum leaf area index (high inputs)	Harvest index (low inputs)	Maximum leaf area index (low inputs)	Yield formation period (percentage of growth cycle)	Ratio of seasonal average rate of net biomass production (b_{na}) and maximum crop growth rate (b_{nm})
1	Winter wheat	35+105	2	1	6	0	5	0.50	4.5	0.20	2.5	0.33	0.55
2	Winter wheat	40+120	2	1	6	0	5	0.50	5.0	0.20	2.5	0.33	0.55
3	Winter wheat	45+135	2	1	6	0	5	0.50	5.5	0.20	2.5	0.33	0.55
4	Winter wheat	50+150	2	1	6	0	5	0.50	5.5	0.20	2.5	0.33	0.55
5	Spring wheat	90	1	1	6	0	5	0.45	3.0	0.15	1.8	0.33	0.50
6	Spring wheat	105	1	1	6	0	5	0.45	3.5	0.15	2.0	0.33	0.50
7	Spring wheat	120	1	1	6	0	5	0.45	4.0	0.15	2.3	0.33	0.50
8	Spring wheat	135	1	1	6	0	5	0.45	4.5	0.15	2.5	0.33	0.50
9	Spring wheat	150	1	1	6	0	5	0.45	5.0	0.15	2.5	0.33	0.50
46	Maize (temperate and subtropical cultivars)	90	9	1	6	0	10	0.45	3.0	0.20	1.8	0.33	0.50
47	Maize (temperate and subtropical cultivars)	105	9	1	6	0	10	0.45	3.0	0.20	1.8	0.33	0.50
48	Maize (temperate and subtropical cultivars)	120	9	1	6	0	10	0.45	3.5	0.20	2.0	0.33	0.50
49	Maize (temperate and subtropical cultivars)	135	9	1	6	0	10	0.45	4.0	0.20	2.0	0.33	0.50
50	Maize (temperate and subtropical cultivars)	150	9	1	6	0	10	0.45	4.5	0.20	2.5	0.33	0.50
51	Maize (temperate and subtropical cultivars)	165	9	1	6	0	10	0.45	5.0	0.20	2.5	0.33	0.50
52	Maize (temperate and subtropical cultivars)	180	9	1	6	0	10	0.45	5.5	0.20	3.0	0.33	0.50
230	Cotton (temperate and subtropical cultivars)	135	4	1	8	0	10	0.08	2.5	0.04	1.5	0.50	0.50
231	Cotton (temperate and subtropical cultivars)	150	4	1	8	0	10	0.08	2.5	0.04	1.5	0.50	0.50
232	Cotton (temperate and subtropical cultivars)	165	4	1	8	0	10	0.08	3.0	0.04	2.0	0.50	0.50

Table 3: Predicting 1860 Usual Yields using GAEZ Crop Suitability Indices

	coeff. [se]	#obs	R_sq
Panel A: Using own crop indices			
Ln_1860_Usual_Cotton			
Ln_GAEZ_Cotton_Int	0.131 [0.031]	155	0.104
(w/ Lat, Lon controls)	0.001 [0.034]	155	0.332
Ln_1860_Usual_Wheat			
Ln_GAEZ_Wheat_Int	0.115 [0.021]	865	0.033
(w/ Lat, Lon)	0.077 [0.019]	865	0.334
Ln_Usual_Corn			
Ln_GAEZ_Maize_Int	0.077 [0.016]	876	0.025
(w/ Lat, Lon)	0.073 [0.014]	875	0.364
Panel B: Horserace between GAEZ Maize and GAEZ Wheat to explain Usual Corn yield			
Ln_Usual_Corn			
Ln_GAEZ_Maize_Int	-0.094 [0.020]	876	0.173
Ln_GAEZ_Wheat_Int	0.323 [0.026]		
w/ Lat, Lon			
Ln_GAEZ_Maize_Int	0.059 [0.028]	875	0.364
Ln_GAEZ_Wheat_ Int	0.021 [0.038]		

Table 4: Predicting Census Cotton Yields using GAEZ Cotton and 1860 Usual yields separately

	Solo Sprint		Solo Sprint		Solo Sprint	
	Gaez		Gaez small sample		1860 Usual	
	Coef	SE	Coef	SE	Coef	SE
1880	-0.006	0.013	0.029	0.028	0.387	0.062
1890	0.008	0.014	0.064	0.028	0.128	0.070
1900	-0.032	0.011	0.039	0.020	0.175	0.049
1910	-0.054	0.014	-0.078	0.024	-0.193	0.060
1920	-0.158	0.021	-0.288	0.038	-0.393	0.107
1930	-0.174	0.016	-0.234	0.034	-0.346	0.092
1940	-0.071	0.020	-0.052	0.037	0.094	0.091

W/ lat and lon controls

	Solo Sprint		Solo Sprint		Solo Sprint	
	Gaez		Gaez small		1860 Usual	
	Coef	SE	Coef	SE	Coef	SE
1880	0.007	0.013	0.007	0.030	0.363	0.067
1890	-0.057	0.014	-0.036	0.033	0.063	0.078
1900	-0.007	0.012	0.011	0.025	0.157	0.060
1910	0.074	0.013	0.039	0.027	0.014	0.064
1920	-0.044	0.017	-0.062	0.033	0.018	0.080
1930	-0.093	0.160	-0.059	0.034	-0.008	0.083
1940	-0.016	0.019	0.079	0.041	0.321	0.095

Table 5: Predicting Census Cotton Yields using GAEZ Cotton and 1860 Usual yields together

	Horserace Gaez small		1860 Usual	
	Coef	SE	Coef	SE
1880	-0.026	0.026	0.409	0.065
1890	0.047	0.030	0.131	0.074
1900	0.018	0.210	0.161	0.520
1910	-0.060	0.025	-0.144	0.062
1920	-0.264	0.040	-0.181	0.100
1930	-0.210	0.035	-0.177	0.087
1940	-0.072	0.039	0.152	0.096

W/ lat and lon controls

	Horserace Gaez small		1860 Usual	
	Coef	SE	Coef	SE
1880	0.006	0.027	0.362	0.068
1890	-0.036	0.033	0.063	0.078
1900	0.107	0.025	0.157	0.061
1910	0.039	0.027	0.014	0.064
1920	-0.062	0.033	0.018	0.080
1930	-0.059	0.034	-0.007	0.082
1940	0.079	0.040	0.321	0.094

Table 6: Census Wheat Yields using GAEZ Wheat and 1860 Usual yields separately

	Solo Sprint Gaez		Solo Sprint Gaez small sample		Solo Sprint 1860 Usual	
	Coef	SE	Coef	SE	Coef	SE
1880	0.240	0.017	0.309	0.024	0.586	0.037
1890	0.214	0.016	0.284	0.023	0.613	0.033
1900	0.132	0.015	0.186	0.023	0.576	0.034
1910	0.175	0.016	0.258	0.023	0.534	0.034
1920	0.191	0.014	0.244	0.020	0.490	0.031
1930	0.119	0.014	0.191	0.022	0.432	0.033
1940	0.048	0.009	0.185	0.019	0.418	0.029

W/ lat and lon controls

	Solo Sprint Gaez		Solo Sprint Gaez small sample		Solo Sprint 1860 Usual	
	Coef	SE	Coef	SE	Coef	SE
1880	0.103	0.014	0.161	0.018	0.294	0.033
1890	0.119	0.012	0.163	0.018	0.342	0.033
1900	0.024	0.013	0.068	0.020	0.324	0.034
1910	0.059	0.013	0.131	0.019	0.328	0.033
1920	0.104	0.012	0.150	0.018	0.309	0.033
1930	0.023	0.012	0.053	0.019	0.208	0.033
1940	0.084	0.012	0.108	0.017	0.258	0.031

Table 7: Census Wheat Yields using GAEZ Wheat and 1860 Usual yields together

	Horserace			
	Gaez small		1860 Usual	
	Coef	SE	Coef	SE
1880	0.249	0.022	0.511	0.034
1890	0.216	0.020	0.544	0.032
1900	0.122	0.021	0.539	0.033
1910	0.206	0.021	0.479	0.033
1920	0.193	0.019	0.432	0.030
1930	0.120	0.021	0.402	0.033
1940	0.144	0.018	0.379	0.028

W/ lat and lon controls

	Horserace			
	Gaez small		1860 Usual	
	Coef	SE	Coef	SE
1880	0.141	0.018	0.258	0.032
1890	0.139	0.017	0.302	0.032
1900	0.043	0.019	0.330	0.034
1910	0.110	0.018	0.303	0.033
1920	0.129	0.018	0.276	0.032
1930	0.038	0.018	0.199	0.033
1940	0.090	0.017	0.234	0.030

Table 8: Predicting Census Corn Yields using GAEZ maize and 1860 Usual yields separately

	Solo Sprint		Solo Sprint		Solo Sprint	
	Gaez		Gaez small		1860 Usual	
	Coef	SE	Coef	SE	Coef	SE
1880	0.021	0.014	0.029	0.023	0.891	0.036
1890	0.037	0.011	0.051	0.017	0.708	0.024
1900	0.017	0.011	0.024	0.019	0.759	0.029
1910	-0.023	0.012	-0.016	0.021	0.725	0.035
1920	-0.038	0.011	-0.053	0.019	0.607	0.034
1930	-0.046	0.011	-0.081	0.019	0.574	0.031
1940	-0.008	0.009	0.000	0.024	0.854	0.040

W/ lat and lon controls

	Solo Sprint		Solo Sprint		Solo Sprint	
	Gaez		Gaez small		1860 Usual	
	Coef	SE	Coef	SE	Coef	SE
1880	0.107	0.009	0.138	0.014	0.499	0.030
1890	0.104	0.009	0.089	0.014	0.510	0.027
1900	0.096	0.007	0.089	0.013	0.470	0.028
1910	0.064	0.008	0.097	0.014	0.389	0.030
1920	0.039	0.008	0.034	0.015	0.363	0.033
1930	0.024	0.007	0.002	0.013	0.297	0.027
1940	0.103	0.009	0.112	0.016	0.403	0.034

Table 9: Predicting Census Corn Yields using GAEZ maize and 1860 Usual yields together

	Horserace Gaez small		1860 Usual	
	Coef	SE	Coef	SE
1880	-0.041	0.018	0.907	0.036
1890	-0.004	0.012	0.711	0.025
1900	-0.035	0.014	0.771	0.029
1910	-0.074	0.017	0.749	0.035
1920	-0.102	0.016	0.641	0.033
1930	-0.134	0.015	0.619	0.030
1940	-0.069	0.020	0.877	0.041

W/ lat and lon controls

	Gaez small		1860 Usual	
	Coef	SE	Coef	SE
1880	0.105	0.013	0.458	0.029
1890	0.053	0.012	0.489	0.027
1900	0.056	0.012	0.448	0.028
1910	0.070	0.013	0.361	0.030
1920	0.008	0.014	0.360	0.034
1930	-0.022	0.012	0.306	0.027
1940	0.086	0.014	0.370	0.034

Table 10: Predicting Corn Yields using GAEZ Maize and GAEZ Wheat indices together

Corn

	Horserace			
	Gaez Maize		Gaez Wheat	
	Coef	SE	Coef	SE
1880	-0.344	0.014	0.726	0.020
1890	-0.223	0.014	0.506	0.018
1900	-0.232	0.012	0.530	0.017
1910	-0.307	0.013	0.605	0.018
1920	-0.275	0.012	0.514	0.017
1930	-0.283	0.012	0.487	0.016
1940	-0.214	0.010	0.501	0.017

W/ lat and lon controls

	Horserace			
	Gaez Maize		Gaez Wheat	
	Coef	SE	Coef	SE
1880	0.044	0.016	0.108	0.023
1890	0.058	0.017	0.077	0.024
1900	0.042	0.014	0.098	0.021
1910	0.043	0.014	0.040	0.021
1920	0.028	0.014	0.020	0.022
1930	0.030	0.013	-0.011	0.019
1940	0.078	0.017	0.046	0.025