Bowdoin College
Bowdoin Digital Commons

Economics Department Working Paper Series

Faculty Scholarship and Creative Work

9-2016

Measuring the Relative Importance of Different Agricultural Inputs to Global and Regional Crop Yield Growth Since 1975

Erik Nelson
Bowdoin College, enelson2@bowdoin.edu

Clare Bates Congdon
Bowdoin College

Follow this and additional works at: http://digitalcommons.bowdoin.edu/econpapers

Part of the Agricultural and Resource Economics Commons

Recommended Citation

Nelson, Erik and Bates Congdon, Clare, "Measuring the Relative Importance of Different Agricultural Inputs to Global and Regional Crop Yield Growth Since 1975" (2016). *Economics Department Working Paper Series.* Paper 12.
http://digitalcommons.bowdoin.edu/econpapers/12

This Working Paper is brought to you for free and open access by the Faculty Scholarship and Creative Work at Bowdoin Digital Commons. It has been accepted for inclusion in Economics Department Working Paper Series by an authorized administrator of Bowdoin Digital Commons. For more information, please contact jmontgom@bowdoin.edu.

Full Title:
Measuring the relative importance of different agricultural inputs to global and regional crop yield growth since 1975

Short Title:
Explaining growth in global and regional crop yields since 1975

Authors. Erik Nelson,[1]* Clare Bates Congdon[2]

Affiliations.
[1] Bowdoin College, Department of Economics, 9700 College Station, Brunswick, ME 04011-8497, USA.
[2] Bowdoin College, Department of Computer Science, 8650 College Station, Brunswick, ME 04011, USA.
*Corresponding author: enelson2@bowdoin.edu

Abstract
We identify the agricultural inputs that drove the growth in global and regional crop yields from 1975 to the mid-2000s. We find that improvements in agricultural technology, increased fertilizer use, and changes in crop mix around the world explained most of the gain in global crop yields, although impacts varied across the latitude gradient. Climate change over this time period caused yields to be only slightly lower than they would have been otherwise. In some cases cropland extensification had as much of a negative impact on global and regional yields as climate change. To maintain the momentum in yield growth across the globe 1) use of agricultural chemicals and investment in agricultural technology in the tropics must increase rapidly and 2) international trade in agricultural products must expand significantly.

Introduction
A consensus has emerged that recent climate change has had a negative effect on crop yields around the world (e.g., *1–4*). Accelerating climate change is likely to put even more downward pressure on agricultural productivity around the world in coming years. Further, demand for food will grow quickly as the world races to a population of ~12 billion by 2100 (*5*). Therefore, the vital question is: How can the world's farmers increase crop productivity, as necessitated by global population growth, despite the expected drag on yields caused by climate change while leaving the socially desirable amount of forest, grasslands, and other semi-natural land cover around the world (*6*)?

Before suggesting a way forward on this issue, we first have to determine what agricultural inputs are most important to yield growth around the world. Here we use global yield and agricultural input data from 1975 to the mid-2000s to determine what agricultural production inputs were most responsible for the growth in global and regional yields during this time period. The inputs we consider include growing season weather, crop choice, investment in irrigation capability, land, and machinery, agricultural technology, fertilizer use, cropped footprint (*7*), and cropped soil quality. We find that improvements in agricultural technology, increased fertilizer use, and changes in crop mix around the world explained most of the gain in global crop yields from 1975 to the mid-2000s. Technological improvement was a particularly important driver of yield growth in the temperate region and crop mix and fertilizer use were particularly important drivers of yield growth in the tropics. Further, the deleterious impacts of climate change on yield were small compared to the yield-augmenting factors noted above. Finally, cropland extensification over the last 40 years has dragged average global yields down as well, sometimes as much as climate change has.

Our results indicate that 1) transferring technology and other inputs to the tropics 2) encouraging countries to exclusively concentrate on growing the crops most suited to their soil-climate conditions (and trading for the rest of the crops their consumers want), and 3) increasing the productivity of existing cropland in lieu of additional cropland extensification will be the most effective ways to ameliorate climate change's expected drag on global yields.

Results

We used two analytical methods to measure relative importance of agricultural inputs to the growth in global and regional crop yields between 1975 and the mid-2000s.

First, we estimated country-level yield functions with a fixed-effects econometric model using a 1975 to the mid-2000s global panel dataset (Tables S1 and S2 and Data files S1 and S2). We estimated country-level yield functions using both Mg ha^{-1} and M kcals ha^{-1} yield metrics: Mg or M kcal production across *all crops* in a country in year *t* divided by hectares of cropland in the country in year *t*. Second, we used the estimated yield functions and the 1975 to the mid-2000s panel data to obtain annual expected country-level yields, both in Mg ha^{-1} and M kcals ha^{-1}, for the 1975 to the mid-2000s time period. Third, we generated global or regional expected crop yield in year *t* by taking the weighted average of expected country-level yields in year *t* with cropped hectarage in each country in year *t* as weights. This process generated three expected "all-crop" yield curves, one for the globe, one for the temperate region, and one for the tropics region (see Fig. 1 for the global Mg ha^{-1} and M kcals ha^{-1} expected yield functions).

Then to estimate the overall contribution of an agriculture production input or a group of inputs on 1975 to mid-2000s global or regional crop yield trends, we again found the expected global or region yield curve (as explained above) *while holding the input or inputs in question fixed at observed 1975 levels*; all other variables take on observed values. For example, when measuring the impact of soil quality change on yield, the "soil quality" counterfactual yield curves were estimated with the quality of cropped soil around the world remaining fixed at 1975 levels while all other inputs varied as observed. Then by integrating over the gap formed between the expected global or regional yield curve and the counterfactual global or regional yield curve we have measured the relative contribution of that input or group of inputs to 1975 to mid-2000s growth in global or regional yields, all else equal. The larger a counterfactual's integral (in absolute terms), the greater the impact that the input or group of inputs in question had on global or regional yield trends from 1975 to the mid-2000s. A positive (negative) integral means that the 1975 to mid-2000s changes in the input in question had, on net, a positive (negative) impact on average global or regional yield.

When discussing results below, we normalize the size of a counterfactual's integral by measuring its size relative to the size of the integral formed by the numeraire counterfactual. In a numeraire counterfactual all inputs are held at 1975 levels *except* growing season weather over each country's crop production area, which varied as observed (the numeraire counterfactuals always form the largest integrals) We refer to a numeraire counterfactual's integral as the 'Mg gap' or the 'kcals gap' (Fig. 2). For example, the mean global "crop mix" counterfactual when yield is measured in Mg ha^{-1} has an integral of 9.11 over the 1975 to 2007 period (Table 2). The mean global "numeraire Mg" counterfactual when yield is measured in Mg ha^{-1} produces an integral of 30.53. Thus, the mean global "crop mix" counterfactual makes up or explains 9.11 / 30.53 = 29.83% of the 1975 to 2007 global Mg gap. The larger the percentage, positive or negative, the more important the counterfactual's input or group of inputs was to determining the 1975 to mid-2000s global or regional yield trend.

We also used decision tree algorithms, our second analytical method, to obtain a "second opinion" on which agricultural inputs were most important in explaining the growth in global and regional crop yields between 1975 and the mid-2000s. A decision tree segregates a process' outcomes (in our case, changes in observed country-level yields) based on the attributes of a process (in our case, changes in country-level input levels). A tree can be interpreted as the rules that best map attributes of a process to the outcome of the process. In our case we find rules – ranges in annual changes in input levels – that predicted changes in country-level yields best (Figs. S1 – S12 and Data file S3). When using econometric techniques to build a yield function, we make several assumptions regarding the variable-generating process. In decision tree analysis, a machine learning algorithm, we identify key features of the data without committing to statistical assumptions.

For each analytical method we discuss two sets of results. In one case we derive results for the time period 1975 to 2007. However, this set of results does not include fertilizer as a production input. In the other case we derive results for the time period 1975 to 2002. This set of results does include fertilizer use as an explanatory variable. The source of much of our agriculture data changed their fertilizer collection methods beginning in 2003 (*8*). Harmonizing the two fertilizer databases was not practical. Below we will refer to results derived from the 1975 to 2002 dataset as the "wide" results and results derived from the 1975 to 2007 dataset as the "long" results.

Technological and crop-mix change and increased fertilizer use has explained most recent yield growth

When using either the long and wide datasets, time was the largest contributor to crop yield growth (both in terms of Mg ha^{-1} and M kcals ha^{-1}) at the global and temperate region levels (Tables 1 and 2 for the wide and long results, respectively). (Unless otherwise stated, we discuss mean results in the text.) At the global level, the time counterfactual's integral makes up approximately 57% or 72% of the Mg gap (always wide and long results, respectively, unless otherwise stated) and 37% or 47% of the kcal gap. In the time counterfactual we hold the year variable fixed at 1975. In the temperate region the time counterfactual makes up 79% or 90% of the Mg gap and 62% and 67% of the kcal gap. At the other extreme, the time counterfactual only explains -1.5% or 24% and -12.5% or 18% of the tropic's Mg and kcal gaps, respectively.

Our econometric model's time trend jointly captures the impact of several agricultural inputs that are omitted from our global panel database. Between 1975 and the mid-2000s, agricultural technology, agriculture management science, pesticide use, and international trade of agricultural commodities (variables missing from our dataset) have increased around the world (*9*). That greater technology, better management, and more pesticides have increased yield is intuitive. However, the impact of increasing globalization on yields was important as well. Greater liberalization of agricultural domestic policy around the world and advancements in shipping technology meant that farmers were able to access international markets at increasingly lower costs (*10*). Greater market access spurred greater investment in farms (e.g., *11*). Further, as cropland around the world became scarcer relative to the supply of rural labor, farmers increasingly became motivated to maximize yield rather than economize on labor use (e.g., *12*). The time trend crudely accounts for the joint impact of these unobserved factors on global and regional yields (including fertilizer use in the long results but not in the wide results, which explicitly includes fertilizer use). Our results make it clear that the growth in agricultural technology, input use, farm management, globalization, and market liberalization disproportionally benefited the farmers of more developed nations in the temperate region over the last 40 years than it did farmers of tropical countries.

When using either the wide or long datasets, change in crop mix was the largest net contributor to yield growth in the tropics. The tropical region's integral from the crop mix counterfactual, where we keep the relative mix of crop hectarage in each country frozen at 1975 levels, makes up 55% or 61% and 58% or 65% of the tropic's Mg and kcals gaps, respectively. Between 1975 and 2007 oil crops, sugarcane, roots and tubers, and fruit became a larger part of cropped area in the tropic region (Fig. 3). According to the estimated yield models (Tables S1 and S2), replacing wheat and other grain production with sugarcane, roots and tubers, and fruit production was particularly important to improving overall crop yield in the tropics. The gain in yield due to this crop switching can partly be explained by a simple substitution effect: Tropical cropland was increasingly used to grow denser fruits and roots and tubers versus less dense wheat. However, this also reflects a comparative advantage effect, as wheat and most grains are most effectively grown in cooler climates while fruits are most cost-effectively grown in the tropics (*13*). In comparison to its impact in the tropics, change in crop mix in the temperate region had little impact on yield when measured in Mg and only slightly improved yield when measured in M kcals.

The change in a country's crop mix from 1975 to the mid-2000s was most likely driven by changes in global demand for various foodstuffs (e.g., *14*, *15*) and the increasing globalization of crop production and trade (*9*). As an example of the former effect, retail sales of foods with high oil and fat content increased dramatically in many countries from 1983 to 2002. Further, the number of calories that the average global person got from cereals fell while the number of calories they got from fruits and vegetables rose from 1996 to 2002 (*16*). As an example of the globalization effect, consider that the reduction of several trade barriers in the early 1990s was largely responsible for the doubling of soybean production in Brazil (*17*). Other potential explanations for country-level changes in crop mix include farmers adapting to climate change. However, there is little evidence of adaptation being a large driver of crop mix change.

Increasing fertilizer use across the globe from 1975 to 2002 (Table 2) was the next most important contributor to the steady gains in yield over that time period (only the wide dataset includes fertilizer data). When yield is measured in Mg ha^{-1}, the fertilizer counterfactual makes up 23% to 32% to 38% of the Mg gaps (the temperate, global, and tropics Mg gaps, respectively, using the wide dataset). When yield is measured in M kcals ha^{-1}, fertilizer makes up 12% to 23% to 42% of the kcals gaps (again, the temperate, global, and tropics Mg gaps, respectively, using the wide dataset). Further, the time trend no longer has a positive effect on the tropical yield when using the wide dataset. In fact, the time counterfactual produces a negative kcal gap in the tropics. In other words, the positive impact of time on tropical yield when using the long dataset is entirely explained by the time trend's incorporation of fertilizer effects.

Recent climate change slightly dampened yield growth

Compared to time, crop mix, and fertilizer use, the impact of the other agricultural inputs on recent global and regional yield was much less significant in terms of magnitude. When using the long or wide datasets, recent increases in daytime growing season temperatures (DGSTs; Table 4) negatively affected global and regional yields. When yield is measured in Mg ha^{-1}, the DGST counterfactual makes up –4% or –6% of the global Mg gap (as before, the order is always wide and long results, respectively, unless otherwise stated). When yield is measured in M kcals ha^{-1}, the DGST counterfactual makes up –4% or –5% of the global kcals gap. In the DGST counterfactual we fixed DGSTs around the world at 1975-1977 averages. The negative impact of increasing DGSTs on global yield was almost entirely explained by its drag on tropical yields; the impact of increasing DGSTs on temperate region yields was almost non-existent.

All else equal, warm days and cool nights allow for vigorous plant growth during the day and efficient plant respiration at night (*18–21*). In contrast, warmer nighttime temperatures cause more wasteful respiration and less energy for growth during the day, all else equal. Therefore, we were surprised to find that increasing nighttime growing season temperatures (NGSTs) at the global and tropical region scales (Table 4) were associated with a boost in yields. The NGST counterfactual makes up ~10% of tropic's Mg and kcal gaps. However, in the temperate region we find evidence of the expected impact of increasing NGS temperatures on yield: the NGST counterfactual makes up –3% or –4% and –3% or –2% of the temperate region's Mg and kcal gaps, respectively. Changes in growing season precipitation almost had no effect across the globe or in either region.

Recent change in cropped soil quality and cropland footprint had a negligible effect on yield growth

Recent changes in the quality of land cropland around the world have had a mixed effect on yield growth. One way we measure a country's change in cropland quality over time is by calculating the change in its cropped soil's nutrient availability and retention capacity (*22*). We also measure a country's extensive change in footprint by tracking its net areal change in cropland over time. The extensive change in cropped area is a catch-all for the change in land quality conditions not measured by the change in the nutrient availability and retention capacity of cropped soils. We assume that a country's most productive land has long been used for crops and net growth in cropland extent since 1975 will have had a negative impact on yield as only more marginal lands were available for cropping after 1975. Most of the 1975 to mid-2000s growth in cropland extent has occurred in the tropics (Table 4). Further, the decline in the overall quality of cropped soil has been more dramatic in the tropics as more and more tropical forest area and their poor soils have been used for crops since 1975 (*23*).

A general worsening in the nutrient availability and retention capacity of cropped soils across the globe was associated with slightly lower yields (Tables 1 and 2). However, the extent of the loss was very small (the soil quality counterfactual makes up –0.2% to –1.2% of global Mg and kcal gaps). As expected, net growth in cropped area was associated with a decline in global and tropical Mg yields. Again, however, the extent of the negative impact is relatively minor (the area cultivated counterfactual makes up –13% or –2% to of global Mg gaps and –7% or –5% of tropical Mg gaps). In contrast, and contrary to expectations, net growth in cropped area was associated with an increase in global and temperate region yields when measured in M kcals ha^{-1}. Again, however, the extent of the gap created by net change in cropped area in these cases is relatively small (the area cultivated counterfactual makes up 5% or 16% of global kcals gaps and 12% or 19% of temperate region kcals gaps).

The counterintuitive positive relationship between net cropland expansion and higher M kcal ha^{-1} yield in the temperate region may hold for several reasons. First, it may be that land that was marginal for crops grown earlier in the 20th century became more suitable for the more kcal-denser crop mixes grown over the last 40 years. Second, land that was marginal given earlier technology and cultivars may have become increasingly productive, especially for kcal-rich crops, with emerging technology. Third, cropland across the world has generally become better connected to transportation infrastructure, thereby encouraging farmers to invest in their operations and potentially more than compensating for their land's quality shortcomings (*11, 24*). Finally, we note that these counter intuitive results are less noticeable when using the wide dataset. In other words, the yield curves estimated with the long dataset may be biased upwards with respect to the area cultivated variable due to the omitted fertilizer variable.

Investment in land, machinery, and irrigation had little impact on recent yield growth

Surprisingly, we found investment in irrigation capacity and investment in land and equipment and machinery from 1975 to the mid-2000s (Table 4) to have had very little effect on global and regional yields (see the irrigation capability and investment in land and equipment counterfactuals in Tables 1 and 2). Increases in irrigation capacity had a positive effect on Mg and kcal yield across the globe and in both regions but no irrigation capacity counterfactual produced an integral larger than 4% of a gap. Further, investment in land and farm machinery and equipment appears to have contributed little to yield growth over time. Part of investment's lack of impact on yield is explained by the fact that land development investment per cropped hectare only increased by 10% around the globe between 1975 and 2007 and actually fell over this time period in the tropics (Table 4). However, the lack of investment in land in the tropics was countered by a 60% increase in the value of farm machinery and equipment per cropped hectare in the region over the same time period. The large increase in machinery use in the tropics vis-à-vis the temperate region may explain why the tropical integrals for the investment in land and equipment counterfactual are larger than the analogous integrals for the temperate region. The investment in land and equipment counterfactual makes up 6% of the tropic's Mg gap (with both the wide and long model estimates) and 8% or 1% of the tropic's kcal gap (with the wide and long model estimates, respectively).

Robustness analysis with decision trees

Before we analyzed our panel dataset with decision trees, we first transformed the wide and long datasets into annual change wide and long datasets. These annual change datasets begin with each country's 1975 to 1976 changes and end with each country's 2001 to 2002 changes (wide dataset) or 2006 to 2007 changes (long dataset). We transform continuous distributions of annual change in country-level crop yield into a discrete distribution of three tertiles; low annual change (L), moderate annual change (M), and high annual change (H) (see Table 5 for an exact numerical definition of these categories).

The decision tree algorithm recursively partitions the dataset, eventually settling on *n* sets of decision sequences that predict outcomes of L, M, and H (*n* traversals of a tree, from the "root" that contains all the data to a "leaf" that contains a subset of the data) (*25–27*). The partitioning of the data can be constrained by one or more pruning rules. We prune a tree to make it easier to interpret and to increase our confidence in its predictive power. Here, we prune trees by mandating that each leaf node in a tree has at least 50 records that support the decision sequence leading to the leaf node. In other words, sets of country-level year-to-year changes in inputs could not be mapped as a branch unless at least 50 instances of that set were observed in the data. After meeting the pruning rules, the decision tree algorithm has produced the sets of annual changes in agricultural inputs that best predict whether a country had an L, M, or H categorical change in annual yield.

Unique combinations of yield metric {Mg ha^{-1}, M kcals ha^{-1}}, scale {globe, temperate, tropics}, and dataset {wide dataset, long dataset} mean that we create 12 unique trees (see Figs. S1 - S12). We summarize the 12 decision trees in several ways. First, we report on the accuracy and complexity of each tree (Table 5 and Data file S3). Second, we list all of the inputs that are found in the first 3 levels of a tree. We highlight these inputs because they do the most towards explaining or predicting annual change in a country's yield. Third, we highlight the traversal in each tree with the highest number of records. These traversals indicate the annual changes in agricultural inputs that are most common across space and time. Finally, we indicate the traversals that generate the greatest proportion of high (H) and low (L) annual country-level yield changes in a tree. These traversals give the ranges in annual input change that, respectively, best predict a high and low annual yield change in a country.

We find that the trees constructed from the wide dataset are simpler (fewer traversals) than those constructed from the long dataset and the trees constructed with the change in Mg ha^{-1} yield metric are simpler than those constructed with the change in M kcal ha^{-1} yield metric. (The econometric analysis also indicates that the wide dataset with yield measured in Mg ha^{-1} has the best fit.) In terms of accuracy of prediction, the trees constructed over the temperate countries are better than the trees generated over all countries and tropical countries only, and the trees generated with yield measured in M Kcals ha^{-1} are better than the trees generated with yield measured in Mg ha^{-1}. In summary, annual yield changes in the temperate countries are explained by a narrower set of annual input changes than annual yield changes in the tropics. In other words, explanations of changes in tropical yields are messier.

Next we describe the inputs found closest to the roots of trees where the root of the tree contains all the data and we define "close to the root" as the first three levels of a tree from its root (the first three decisions). Changes in a country's crop mix – change in relative area devoted to sugarcane, roots and tubers, and wheat – appear close to the roots of all 12 trees. In particular, sugarcane is found close to the root of all 12 trees and the roots and tubers crop category is found close to the root of all 3 trees formed with the long dataset when yield is measured in Mg ha^{-1}. The annual change in DGSTs is close to the root of 3 of the 4 trees estimated over the tropical countries. Finally, change in cultivated area is found close to the root of the two trees estimated over the temperate countries when yield is measured in Mg ha^{-1}. In summary, the decision trees indicate that recent annual changes in yield across the globe were most associated with changes in crop mix and that each region had idiosyncratic drivers of yield change as well.

(In the decision tree analysis we de-trend the data by using annual changes; in the fixed-effects analysis we de-trended the data by including time as an explanatory variable. This means the decision tree analysis cannot account for the various unobserved inputs that are correlated with time.)

A gain in the proportion of a country's crop mix devoted to sugarcane was the best predictor of high (H) yield change in five of the six trees created with the wide dataset and four of the six trees created with the long dataset. Prediction of H became a bit more complicated in the global trees estimated with the long dataset. In trees estimated with the long dataset, gains in wheat and roots and tubers in the proportional mix of a country's crop profile, modest changes in sugarcane's contribution to the proportional mix, and growing seasons that had cooler daytime temperatures than the previous growing season were most likely to have led to a high annual gain in a country's yield.

The best set of predictors for a negative change in annual yield (the L yield category) is a bit more expansive than the sets of best predictors for H annual-yield change. Not surprisingly, losses in proportion of a country's crop mix devoted to sugarcane are found in all tree branches with the highest proportion of L observations. In the tropics, a one-year gain in DGSTs and NGSTs were also associated with yield losses from one year to the next. Finally, a gain in a country's cultivated area from one year to the next was associated with a negative change in a temperate country's Mg ha^{-1} yield.

Comparing econometric results to decision tree results

When we compare the decision trees (Table 5) to the counterfactual analyses (Tables 1 and 2) several similarities and differences emerge. First, both analyses highlight that changes in crop mix have has been one of the most important contributions to the gain in crop yields over the last 40 years. The decision tree analysis also reinforces the econometric evidence that gains in DGSTs dampened gains in yields more in the tropics than in the temperate region. The trees, like the counterfactual analysis, also suggest that investment in irrigation, land, and machinery and equipment, and the quality of cropped soil have had little effect on yield change. The counterfactual and the decision tree analyses disagree on the importance of fertilizer use in explaining yield gains over the last 40 years, however; the counterfactual analysis deems this input more important than the decision tree analysis.

Discussion

Improvements in agricultural technology, management, and science, changes in crop mix, and increased fertilizer use were responsible for the lion's share of yield improvement around the world from 1975 to 2007. The negative yield impacts associated with increases in growing season temperatures were smaller. In some cases the change in cropland soil quality and cropland footprint were just as detrimental to yields as changes in climate.

Suggestions for maintaining yield growth momentum

The downward pressure on crop yields due to climate change will worsen in the future (e.g., *28*). We see two paths to continued yield improvements despite this growing drag on yields. First, investment in agricultural technology, chemical inputs, management, and science in the tropics is vitally important (the so-called closing of "yield gaps;" *12*). As indicated by the "time" counterfactuals, the tropics have not yet experienced the technological revolution that the temperate region has. Second, if each country can increasingly specialize in the crops best suited for their (changing) climate and trade for the rest of their crop needs, then the spatial allocation of crops will become more efficient. For example, our results indicate the continued divestment in grain production in the tropics and greater investment in grain production in the temperate zone would do much to boost food production in the future. Further, greater fruit and sugarcane production in the tropics relative to the temperate zone would also help accelerate food production (*29*). More trade liberalization and the reduction or even elimination of national crop subsidy programs will make it easier for each country to grow the crops best suited for their soil-climate conditions (*10*).

Several suggested paths to greater food production are not supported by our analysis. Cropland extensification contributed little to yield gains in the immediate past and are not likely to do so in the future (*24*). Instead, switching to more climate-appropriate crops, using more fertilizers and chemicals and improved cultivars, and improving the nutrient retention capability of already *existing* cropland appears to be a more effective strategy for increasing worldwide food production (i.e., land sparing versus land sharing; *30*). This strategy would also leave more land for nature in an increasingly populated world. Further, we are also skeptical that an emphasis on investment in infrastructure in of itself (i.e., machinery and irrigation capacity) will significantly increase yields in the future; these investments did not do much to boost crop production in the recent past. Machinery that is compatible with precision agriculture (i.e., technology) is likely to be more effective than just more tractors and other machinery. Of course, the recommendation on investment in irrigation could change if climate change severely disrupts current rainfall patterns.

Analysis limitations

This analysis is limited by several data issues. First, our treatment of weather data (see Materials and Methods below) does not allow us to decompose changes in growing season weather into spatial reallocation of cropland on the landscape effects and atmospheric climate change effects. Separating these trends would help us better understand the effect of recent climate change on crop yields around the world. Another shortcoming of this analysis is that it does not specifically account for farmer reaction to climate change; this omission could bias our results. For example, if the changes in the spatial pattern of production and crop choice were partially affected by climate change, then we have underestimated the impact of climate change and overestimated the impact of crop choice and cropped-footprint change on recent yield trends. In addition, we are missing data for all countries that were in the Soviet Union and many Warsaw Pact countries (e.g. Poland, Hungary, etc.). The panel database we use does not contain a consistent set of data back to 1975 for these countries. Most of these countries are in the temperate region. Therefore, our analysis, especially the temperate region analysis, could be biased due to the omission of these countries from the dataset. Further, the source of our gridded crop maps stopped updating annual global grid cell maps of cropped land after releasing the 2007 data (*31*). Thus our dataset ends with 2007 data and cannot be extend into the early 2010s. Finally, to conduct this analysis, we either had to summarize the native grid-level data on cropped soil quality and growing season weather at the country level or we had to decompose the native country-level data on production, crop mix, and investment to the grid-cell level. We used the former approach.

A limitation of our decision tree analysis is that trees are constructed in a "greedy" fashion, iteratively splitting on the most powerful agricultural inputs (in a predictive sense) as the branches are built; this can lead to suboptimal trees when there are nonlinear interactions among the variables. Quinlan's C4.5 algorithm (*25*) for the decision tree approach strives to mitigate the biasing effect of the iterative tree-building approach by repeatedly building a tree with a subset of the data and assessing its quality on the held-out data to find the most robust trees; the RWeka decision-tree packaged used for this analysis is a slightly updated version of C4.5. Additionally, more work could be done to explore the results using different transformations of the data, for example, whether the trees would have greater explanatory power if change in yield outcomes were transformed to a discrete distribution of four categories instead of three.

Materials and Methods
Statistical analysis
First, we use the method of least squares to estimate a fixed effects model of annual per hectare crop yield at the country level from years $\underline{t}$ through $\overline{t}$.

$$Y_{ct} = \alpha_c + \beta_0 + \boldsymbol{\beta}_1 \mathbf{X}_{ct} + \boldsymbol{\beta}_2 \mathbf{K}_{ct} + \beta_3 A_{ct} + \beta_4 S_{ct} + \beta_5 I_{ct} + \beta_6 \mathbf{Z}_{ct} + \beta_7 t + \beta_8 F_{ct} \tag{1}$$

where Y_{ct} is the production of all crops grown in country *c* in harvest year *t* measured either in metric tons (Mg) or millions of kilocalories (M kcals) divided by harvested hectares in country *c* in harvest year *t* (harvest year *t* refers to crops harvested in year *t* but not necessarily planted in year *t*; for example, grain can be planted in October and harvested the next March in many southern hemisphere countries). Further, α_c is the fixed effect intercept for country *c*, $\mathbf{X}_{ct}$ is a vector of harvested hectare percentages across crop or crop groups in country *c* in harvest year *t* (collectively $\mathbf{X}_{ct}$ gives a country's "crop mix" in harvest year *t*; see the supplementary methods for more on $\mathbf{X}_{ct}$, *8*), $\mathbf{K}_{ct}$ contains variables that measure investment in agricultural land and agricultural machinery and equipment per harvested hectare *c* in harvest year *t* (*8*), A_{ct} is the harvested or cropped hectarage in country *c* in year *t* (*8*), S_{ct} summarizes the quality of soil used to grow crops in country *c* in harvest year *t* (*22*), I_{ct} is the percentage of harvested area equipped for irrigation in *c* in harvest year *t* (*8*), $\mathbf{Z}_{ct}$ is a vector of statistics that summarize the weather that occurred over country *c*'s cropland during the growing season of harvest year *t* (*32*, *33*), and F_{ct} measures kg ha^{-1} of fertilizers used in country *c* in year *t* (*8*).

The land investment variable in vector $\mathbf{K}_{ct}$ measures major improvements in the quantity, quality or productivity of land or prevention of deterioration. Activities such as land clearance, land contouring, creation of wells and watering holes are integral to the land improvement. The concept of land improvement includes 1) field improvements undertaken by farmers (e.g., making boundaries, irrigation channels) and 2) other activities undertaken by government and other local bodies such as irrigation works, soil-conservation works, and flood-control structure. The machinery and equipment investment variable in vector $\mathbf{K}_{ct}$ measures the value of tractors, harvesters and thrashers, milking machines and hand tools in a country.

See below for more information on how we constructed the variables in the vector $\mathbf{Z}_{ct}$.

In the estimate of model (1) using the "long" dataset (data file S2) F_{ct} is not included and time $\underline{t}$ equals 1975 and time $\bar{t}$ equals 2007. In the estimate of model (1) using the "wide" dataset (data file S1) F_{ct} is included and time $\underline{t}$ equals 1975 and time $\bar{t}$ equals 2002. We estimate the long and wide versions of model (1) with all countries, tropical countries only, and temperate countries only. A country's regional affiliation is defined by the latitude of the country's capital and the Tropics of Cancer and Capricorn. Model (1) was estimated with the `reg` commend in Stata 12.1. See Tables S1 and S2 for estimates of model (1), including estimated standard errors and p-values. Stata code is available upon request from the authors.

Estimating the overall contribution of an agriculture production input on 1975 to mid-2000s global or regional crop yield

We build expected yield curves for country *c*, $\hat{Y}_{ct}$ for years $\underline{t}$ through $\bar{t}$, by running the country's input data from years $\underline{t}$ to $\bar{t}$ through an estimate of model (1),

$$\hat{Y}_{ct} = \hat{\alpha}_c + \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1\mathbf{X}_{ct} + \hat{\boldsymbol{\beta}}_2\mathbf{K}_{ct} + \hat{\beta}_3 A_{ct} + \hat{\beta}_4 S_{ct} + \hat{\beta}_5 I_{ct} + \hat{\beta}_6\mathbf{Z}_{ct} + \hat{\beta}_7 t + \hat{\beta}_8 F_{ct} \qquad (2)$$

where a "^" indicates an estimate (see Tables S1 and S2 for estimated coefficients). Each country has eight expected yield curves, one for each unique combination of yield measure {Mg ha^{-1}, M kcals ha^{-1}}, scale {globe, appropriate region}, and dataset {long, wide}. Using these country-level yield curves we calculated four expected global yield curves, one for each unique combination of yield {Mg ha^{-1}, M kcals ha^{-1}} and dataset {long, wide}) and eight expected regional yield curves, one for each unique combination of yield measure {Mg ha^{-1}, M kcals ha^{-1}}, scale {temperate, tropics}, and dataset {long, wide}. To construct a global or regional yield curve, $\hat{Y}_{rt}$ for years $\underline{t}$ through $\bar{t}$, we average $\hat{Y}_{ct}$ for each year *t* across all *c* in *r* (globe, temperate, tropics) weighed by each country's cropped hectarage in year *t*,

$$\hat{Y}_{rt} = \sum_{c \in r} \frac{A_{ct}\hat{Y}_{ct}}{A_{ct}} \qquad (3)$$

In Fig. 1 we present the global $\hat{Y}_{rt}$ for years 1975 through 2007 (the long dataset) where yield is measured in Mg ha^{-1} (black solid curve in Fig. 1A) and M kcals ha^{-1} (black solid curve in Fig. 1B).

We build counterfactual yield curves for country c, $\tilde{Y}_{ct}$ for years $\underline{t}$ through $\bar{t}$, by running the country's input data from years $\underline{t}$ to $\bar{t}$ through an estimate of model (1), holding one or more of c's inputs fixed at 1975 levels (the exception is a growing season weather counterfactual; in those cases, we fix the appropriate input at the 1975-1977 annual average). Each country has 84 counterfactual yield curves for the years $\underline{t}$ through $\bar{t}$, one for each unique combination of yield measure {Mg ha^{-1}, M kcals ha^{-1}}, scale {globe, appropriate region}, and 10 counterfactuals with the long dataset and 11 counterfactuals with the wide dataset. Using these country-level counterfactual yield curves, we calculated 42 counterfactual global-yield curves, one for each unique combination of yield measure {Mg ha^{-1}, M kcals ha^{-1}} and 10 counterfactuals with the long dataset and 11 counterfactuals with the wide dataset and 84 expected regional yield curves, one for each unique combination of yield measure {Mg ha^{-1}, M kcals ha^{-1}}, scale {temperate, tropics}, and 10 counterfactuals with the long dataset and 11 counterfactuals with the wide dataset. To construct a global or regional counterfactual yield curve, $\tilde{Y}_{rt}$ for years $\underline{t}$ through $\bar{t}$, we average $\tilde{Y}_{rt}$ for each year t across all c in r, weighed by each country's cropped hectarage in year t,

$$\tilde{Y}_{rt} = \sum_{c \in r} \frac{A_{ct}\tilde{Y}_{ct}}{A_{ct}} \tag{4}$$

where $A_{ct} = A_{c,1975}$ for all t in the numeraire and "area cultivated" counterfactuals. In Fig. 1 we present the global $\tilde{Y}_{rt}$ for the numeraire counterfactual (all inputs other than weather inputs are fixed at 1975 levels) for years 1975 through 2007 (the long dataset) where yield is measured in Mg ha^{-1} (blue solid curve in Fig. 1A) and M kcals ha^{-1} (blue solid curve in Fig. 1B).

In the mean columns of Tables 1 and 2 we present the counterfactual integrals,

$$\lambda_{qmrd} = \sum_{t=\underline{t}}^{\bar{t}} \hat{Y}_{tmrd} - \tilde{Y}_{qtmrd} \tag{5}$$

where q indexes the counterfactual, m indicates yield measure {Mg ha^{-1}, M kcals ha^{-1}}, r indicates scale {globe, temperate, tropics}, and d indicates dataset {long, wide} (Fig. 2). To normalize these integrals we also present the fraction of the numeraire counterfactual integral, $\lambda_{counterfactual,m,r,d}$, that counterfactual q's integral "explains,"

$$\lambda_{qmrd} \Big/ \lambda_{counterfactual,mrd} \tag{6}$$

where we call $\lambda_{counterfactual,mrd}$ r's "m" gap using dataset d.

The counterfactual analyses were conducted with MATLAB R2013a. MATLAB code is available upon request from the authors.

Sensitivity analyses

We generated the "low" and "high" results for each q, m, r, and d counterfactual combination in the following manner (Tables 1 and 2). First, we created 1000 unique vectors of model (1) coefficients by randomly drawing from the multivariate normal distribution with a mean of

$\left[\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2 \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_7, \hat{\beta}_8\right]$ (the estimated vector of beta coefficients) and a covariance matrix of,

$$\left(\sigma\sqrt{\frac{N}{\chi_N^2}}\right)^2 \text{vcov} \tag{7}$$

where σ is estimated model (1)'s root mean square error, N is the number of observations in the dataset, χ_N^2 is a random variable with a chi-square distribution with N degrees of freedom, and *vcov* is estimated model (1)'s variance-covariance matrix for all β's. (We do not vary the estimated α_c coefficients do not vary.)

Second, using the 1000 randomly generated **β** coefficient vectors, we generated 1000 values of $\hat{Y}_{ctmd}$ for all c and t for each unique m and d combination and 1000 values of $\tilde{Y}_{qctmd}$ for all c and t for each unique q, m, and d combination. Third, we generated expected 25th and 75th percentile yield curves for each country and each unique m and d combination by selecting the 25th percentile and 75th percentile values of $\hat{Y}_{ctmd}$ at each t. Fourth, we generated counterfactual 25th and 75th percentile yield curves for each country and each unique q, m, and d combination by selecting the 25th percentile and 75th percentile values of $\tilde{Y}_{qctmd}$ at each t. Fifth, we calculated a region or the globe's expected percentile yield in year t with,

$$\hat{Y}_{tmrd}^{25} = \sum_{c \in r} \frac{A_{ct}\hat{Y}_{ctmd}^{25}}{A_{ct}} \tag{8}$$

$$\hat{Y}_{tmrd}^{75} = \sum_{c \in r} \frac{A_{ct}\hat{Y}_{ctmd}^{75}}{A_{ct}} \tag{9}$$

for each unique m and d combination where the superscripts "25" and "75" indicate the 25th and 75th percentile, respectively. Sixth, we calculated the globe or region's counterfactual percentile yield in year t with,

$$\tilde{Y}_{qtmrd}^{25} = \sum_{c \in r} \frac{A_{ct}\tilde{Y}_{qctmd}^{25}}{A_{ct}} \tag{10}$$

$$\tilde{Y}_{qtmrd}^{75} = \sum_{c \in r} \frac{A_{ct}\tilde{Y}_{qctmrd}^{75}}{A_{ct}} \tag{11}$$

for each unique q, m and d combination. Finally, in the low and high columns of Tables 1 and 2 we present the percentile counterfactual integrals for a given region r,

$$\lambda_{qmrd}^{25} = \sum_{t=\underline{t}}^{\bar{t}} \hat{Y}_{tmrd}^{25} - \tilde{Y}_{qtmrd}^{25} \tag{12}$$

$$\lambda_{qmrd}^{75} = \sum_{t=\underline{t}}^{\bar{t}} \hat{Y}_{tmrd}^{75} - \tilde{Y}_{qtmrd}^{75} \tag{13}$$

Decision tree analysis

Decision trees were constructed using the RWeka package in R and J48 classifiers in particular. These are a reimplementation of Quinlan's C4.5 algorithm (*25*). Trees were evaluated for prediction accuracy using a 10-fold cross-validation strategy. Decision trees are given in Figs. S1 – S12, and results are summarized in Table 5. In the analysis reported here, "leaf nodes" (the resulting subsets of the data after the branching of the tree on decision variables) were required to contain at least 50 observations, using the M option to control the minimum number of instances per leaf. This approach was used to yield trees with higher human interpretability as well as higher prediction accuracy. While 50 is somewhat arbitrary, we explored other values and empirically found it to lead to high prediction accuracy and greater interpretability in the resulting trees. (Interestingly, this approach also worked better for this data than using the C option to control the "confidence" in the pruned trees.)

Creating country-level data for crop yield model and decision tree analysis

To create country-level summary statistics of the quality of cropped soil (S_{ct}) and growing season weather over cropland (contained in vector $\mathbf{Z}_{ct}$) in each country in each harvest year t we used annual global grid cell maps of cropped land (*31*) along with gridded global maps of soil quality (*22*), monthly weather (*32*), and growing season months (*33*). (Ramankutty and Foley stopped updating annual global grid-cell maps of cropped land after releasing the 2007 data. Thus, our dataset ends with 2007 data.) By combining the gridded maps on soil, weather, and growing season months with gridded cropland maps we were able to create summary statistics that preserved the observed spatial heterogeneity in agronomic conditions across a county in any given year. For example, consider the landscape in Fig. 4. Suppose the square landscape represents a country. Assume the large number in each grid cell in Fig. 4A represents the number of cropland hectares in that cell in harvest year t (the small number in the corner of a cell is its ID number). In Fig. 4B each cell's nutrient availability score is given where a 1 indicates 'No or slight nutrient constraint', 2 indicates 'moderate nutrient constraint', 3 indicates 'severe nutrient constraint', 4 indicates 'very severe nutrient constraint', and 5 indicates 'mainly non-soil' (in other words, lower scores mean better soil quality; see *22*). Nutrient availability (N_{ct}) is decisive for successful low-level-input farming and, in some cases, intermediate-input-level farming. A country's composite nutrient availability score on cropland in harvest year t is the weighted average of the nutrient availability scores across all cropland area in the country in harvest year t or,

$$N_{ct} = \sum_{j \in c} A_{jt} N_j \Big/ \sum_{j \in c} A_{jt} \qquad (14)$$

where $j \in c$ is the set of grid cells in country c, N_j is grid cell j's nutrient availability score, and A_{jt} is grid cell j's cropland area in harvest year t (*31*). In the illustrative country represented in Fig. 4 N_{ct} is equal to,

$$N_{ct} = \frac{1 \times 100 + 2 \times 1000 + 3 \times 500 + 2 \times 100 + ... + 3 \times 700}{100 + 1000 + 500 + 100 + ... + 700} = 2.28 \qquad (15)$$

We use the same method to calculate a country's nutrient retention score, given by U_{ct}. Nutrient retention capacity is of particular importance for the effectiveness of fertilizer applications and is therefore of special relevance for intermediate and high input level cropping conditions. The explanatory soil statistic used in the model, S_{ct}, is the average of N_{ct} and U_{ct}.

The weather vector $\mathbf{Z}$ includes weather statistics that summarize the weather conditions over a country's cropland during the growing season. We summarize each weather variable at the country level in year t with a procedure very similar to that used to find the country-level cropland soil statistic S. Let $DGST_{jmt}$ and $NGST_{jmt}$ indicate the average daytime high and nighttime low temperature in grid cell j in month m of harvest year t (measured in degrees Celsius) (*32*). Let $DGST_{jt}$ and $NGST_{jt}$ indicate the average of $DGST_{jmt}$ and $NGST_{jmt}$, respectively, across grid cell j's growing season months of harvest year t where we use a grid cell's growing season months for maize to define growing season. Let P_{jt} be the total precipitation in grid cell j during the cell's growing season in harvest year t (measured in millimeters). If a crop was harvested in the spring of year t then some of the weather that contributes to $DGST_{jt}$, $NGST_{jt}$, and P_{jt} occurred in the final months of year $t - 1$. Let $DGST_{ct}$, $NGST_{ct}$, and P_{ct} measure the average monthly daytime high, monthly nighttime low, and growing season precipitation, respectively, over c's cropland during the course of growing season t where weather data is weighted by cropland density in grid cell j.

$$DGST_{ct} = \sum_{j\in c} A_{jt} DGST_{jt} \Big/ \sum_{j\in c} A_{jt} \tag{16}$$

$$NGST_{ct} = \sum_{j\in c} A_{jt} NGST_{jt} \Big/ \sum_{j\in c} A_{jt} \tag{17}$$

$$P_{ct} = \sum_{j\in c} A_{jt} P_{jt} \Big/ \sum_{j\in c} A_{jt} \tag{18}$$

where A_{jt} is the area of grid cell j that was cropped in year t. The weather vector $\mathbf{Z}_{ct}$ in model (1) also incudes the squares of $DGST_{ct}$, $NGST_{ct}$, and P_{ct}.

MATLAB code was used to construct S_{ct}, $DGST_{ct}$, $NGST_{ct}$, and P_{ct} and is available from the authors upon request.

Maps of country-level change in agricultural inputs

Maps of 1975 – 1977 to 2005 – 2007 country-level changes in various model (1) inputs are given in Figs. S13 – S21.

Supplementary Materials

Fig. S1. Decision tree for globe, yield measured in Mg ha^{-1}, using the "long" dataset.

Fig. S2. Decision tree for temperate region, yield measured in Mg ha^{-1}, using the "long" dataset.

Fig. S3. Decision tree for tropics, yield measured in Mg ha^{-1}, using the "long" dataset.

Fig. S4. Decision tree for globe, yield measured in M kcals ha^{-1}, using the "long" dataset.

Fig. S5. Decision tree for temperate region, yield measured in M kcals ha^{-1}, using the "long" dataset.

Fig. S6. Decision tree for tropics, yield measured in M kcals ha^{-1}, using the "long" dataset.

Fig. S7. Decision tree for globe, yield measured in Mg ha^{-1}, using the "wide" dataset.

Fig. S8. Decision tree for temperate region, yield measured in Mg ha^{-1}, using the "wide" dataset.

Fig. S9. Decision tree for tropics, yield measured in Mg ha^{-1}, using the "wide" dataset.

Fig. S10. Decision tree for globe, yield measured in M kcals ha^{-1}, using the "wide" dataset.

Fig. S11. Decision tree for temperate region, yield measured in M kcals ha^{-1}, using the "wide" dataset.

Fig. S12. Decision tree for tropics, yield measured in M kcals ha^{-1}, using the "wide" dataset.

Fig. S13. Percentage change in 1975-1977 to 2005-2007 growing season daytime temperature by country.

Fig. S14. Percentage change in 1975-1977 to 2005-2007 growing season nighttime temperature by country.

Fig. S15. Percentage change in 1975-1977 to 2005-2007 growing season precipitation by country.

Fig. S16. Percentage change in 1975-1977 to 2005-2007 soil score by country.

Fig. S17. Percentage change in 1975-1977 to 2005-2007 hectares of irrigation capacity per cropped hectare by country.

Fig. S18. Percentage change in 1975-1977 to 2005-2007 equipment investment ($ 2005) per cropped hectare by country.

Fig. S19. Percentage change in 1975-1977 to 2005-2007 land investment ($ 2005) per cropped hectare by country.

Fig. S20. Percentage change in 1975-1977 to 2005-2007 all crop M kcals per hectare yield by country.

Fig. S21. Percentage change in 1975-1977 to 2005-2007 all crop Mg per hectare yield by country.

Table S1. Econometric estimates of fixed effects model (1) with the "long" global, tropics, and temperate datasets.

Table S2. Econometric estimates of fixed effects model (1) with the "wide" global, tropics, and temperate datasets.

Supplementary Methods: Crop groups

Data file S1. "Wide" dataset.

Data file S2. "Long" dataset.

Data file S3. Accuracy of decision trees.

References

1. W. Schlenker, W. M. Hanemann, A. C. Fisher, The impact of global warming on US agriculture: an econometric analysis of optimal growing conditions. *Rev. Econ. Stat.* **88.1**, 113-125 (2006).

2. W. Schlenker, M. J. Roberts, Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proc. Natl. Acad. Sci. U.S.A.* **106.37**, 15594-15598 (2009).

3. O. Ashenfelter, K. Storchmann, Using hedonic models of solar radiation and weather to assess the economic effect of climate change: the case of Mosel valley vineyards. *Rev. Econ. Stat.* **92.2**, 333-349 (2010).

4. D. B. Lobell, W. Schlenker, J. Costa-Roberts, Climate trends and global crop production since 1980. *Science* **333.6042**, 616-620 (2011).

5. D. Tilman, C. Balzer, J. Hill, B. L. Befort, Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20260-20264 (2011).

6. J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, Solutions for a cultivated planet. *Nature* **478**, 337-342 (2011).

7. J. M. Beddow, P. G. Pardey, Moving matters: the effect of location on crop production. *J. Econ. Hist.* **75**, 219-249 (2015).

8. FAOSTAT (Food and Agriculture Organization of the United Nations). FAOStat database, available at http://faostat3.fao.org/home/E. (2011).

9. J. M. Alston, P. G. Pardey, Agriculture in the global economy. *J. Econ. Perspect.* **28.1** 121-146 (2014).

10. K. Anderson, Globalization's effects on world agricultural trade, 1960–2050. *Philos. T. Roy. Soc. B.* **365**, 3007-3021 (2010).

11. M.D.C. Vera-Diaz, R. K. Kaufmann, D. C. Nepstad, P. Schlesinger, An interdisciplinary model of soybean yield in the Amazon Basin: the climatic, edaphic, and economic determinants. *Ecol. Econ.* **65.2** 420-431 (2008).

12. D. B. Lobell, K. G. Cassman, C. B. Field, Crop yield gaps: their importance, magnitudes, and causes. *Annu. Rev. Environ. Resour.* **34.1** 179 (2009).

13. A. Costinot, D. Donaldson, Ricardo's theory of comparative advantage: old idea, new evidence (National Bureau of Economic Research, No. w17969, 2012).

14. S. L. Pollack, Consumer demand for fruit and vegetables: the US example. *Changing Structure of Global Food Consumption and Trade* **6** 49-54 (2001).

15. P. Pingali, Westernization of Asian diets and the transformation of food systems: implications for research and policy. *Food Policy* **32**, 281-298 (2007).

16. A. Regmi, M. Gehlhar, Eds. New Directions in Global Food Markets. AIB-794. USDA/ERS (2005).

17. R. D. Schnepf, E. Dohlman, C. Bolling, Agriculture in Brazil and Argentina: Developments and Prospects for Major Field Crops. Market and Trade Economics Division, Economic Research Service, U.S. Department of Agriculture, Agriculture and Trade Report, WRS-01-03, (2001).

18. S. Peng, J. Huang, J. E. Sheehy, R. C. Laza, R. M. Visperas, X. Zhong, G. S. Centeno, G. S. Khush, K. G. Cassman, Rice yields decline with higher night temperature from global warming. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9971-9975 (2004).

19. T. Fulu, M. Yokozawa, Y. Xu, Y. Hayashi, Z. Zhang, Climate changes and trends in phenology and yields of field crops in China, 1981–2000. *Agr. Forest Meteorol.* **138**, 82 (2006).

20. P. Thomison, Can warm nights reduce grain yield in corn? C.O.R.N. Newsletter- Ohio State University, Issue 2010, 22 (2010).

21. W. R. L. Anderegg, A. P. Ballantyne, W. K. Smith, J. Majkut, S. Rabin, C. Beaulieu, R. Birdsey, J. P. Dunne, R. A. Houghton, R. B. Myneni, Y. Pan, J. L. Sarmiento, N. Serota, E. Shevliakova, P. Tans, S. W. Pacala, Tropical nighttime warming as a dominant driver of variability in the terrestrial carbon sink. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15591-15596 (2015).

22. G. Fischer, F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, D. Wiberg, Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome, Italy (2008).

23. P. C. West, H. K. Gibbs, C. Monfreda, J. Wagner, C. C. Barford, S. R. Carpenter, J. A. Foley, Trading carbon for food: Global comparison of carbon stocks vs. crop yields on agricultural land. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19645-19648 (2010).

24. W. F. Laurance, J. Sayer, K. G. Cassman, Agricultural expansion and its impacts on tropical nature. *Trends Ecol. Evol*. **29.2** 107-116 (2014).

25. J. R. Quinlan, C4.5: Programs for machine learning. Morgan Kaufmann Publishers (1993).

26. W-Y. Loh, Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **1.1** 14-23 (2011).

27. H. R. Varian, Big data: New tricks for econometrics. *J. Econ. Perspect.* **28.2** 3-27 (2014).

28. A. P. K. Tai, M. V. Martin, C. L. Heald, Threat to future global food security from climate change and ozone air pollution. *Nature Climate Change* **4**, 817-821 (2014). http://dx.doi.org/10.1038/nclimate2317.

29. W. G. K. Mauser, F. Zabel, R. Delzeit, T. Hank, B. Putzenlechner, A. Calzadilla, Global biomass production potentials exceed expected future demand without the need for cropland expansion. *Nature Communications* **6**, 8946 (2015). doi:10.1038/ncomms9946.

30. A. Balmford, R. Green, B. Phalan, What conservationists need to know about farming, *Philos. T. Roy. Soc. B.* **279**, 2714-2724 (2012).

31. N. Ramankutty, J. Foley, Estimating historical changes in global land cover: croplands from 1700 to 1992. *Global Biogeochem. Cy.* **13**, 997-1028 (1999).

32. I. Harris, P.D. Jones, T.J. Osborn, D. H. Lister, D.H., Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol*. **34**, 623–642 (2014). doi:10.1002/joc.3711.

33. W. J. Sacks, D. Deryng, J. A. Foley, N. Ramankutty, Crop planting dates: an analysis of global patterns. *Global Ecol. Biogeogr.* **19(5)**, 607-620 (2010).

Acknowledgments: The authors wish to thank Jae Bradley and Clarissa Hunnewell, undergraduates at Bowdoin College, for help with putting datasets together and analyzing data.

Funding: This work was done with no funding.

Author contributions: E.J.N. did everything other than construct the decision trees. C.B.C constructed the decision trees. C.B.C. also wrote and edited portions of the text.

Competing interests: The authors declare that they have no competing interest.

Figures and Tables

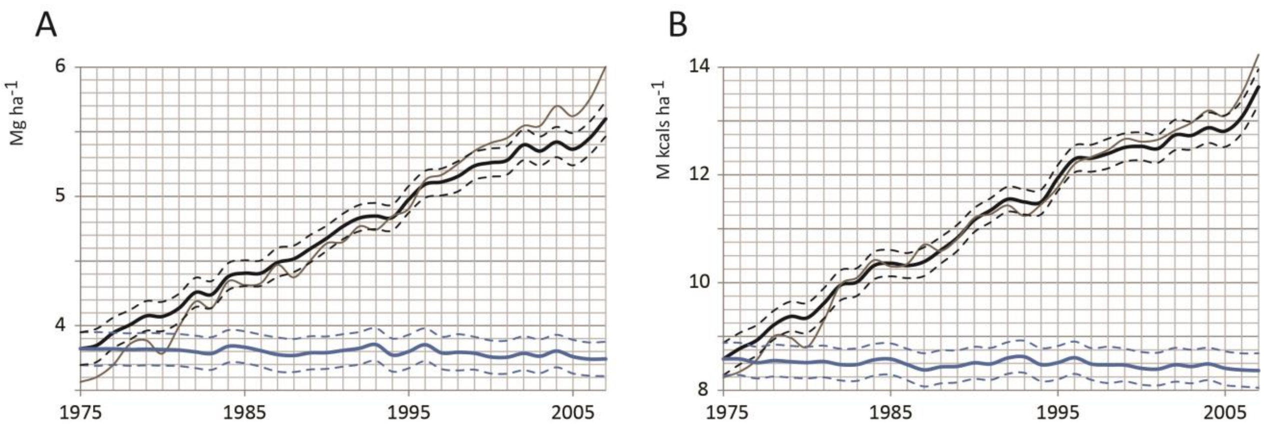

Fig. 1. Expected global yield given 1975-2007 spatiotemporal data (black lines where dashed lines indicate +/- one standard deviation) and numeraire counterfactual global yield (blue line where the dashed lines indicate +/- one standard deviation). The counterfactual global yield curves were constructed by holding all country-level agricultural inputs at 1975 levels explanatory except growing season weather. These graphs are based on "long" model results (the dataset with 1975 to 2007 data and does not include the fertilizer variable). Expected global yield grew 46.5% when measured in Mg ha^{-1} (**A**) and 58.8% when measured in M kcals ha^{-1} (**B**) between 1975 and 2007. Under the counterfactual global yield fell by 2.1% when measured in Mg ha^{-1} (**A**) and 2.5% when measured in M kcals ha^{-1} (**B**). The light gray line indicates observed global yields.

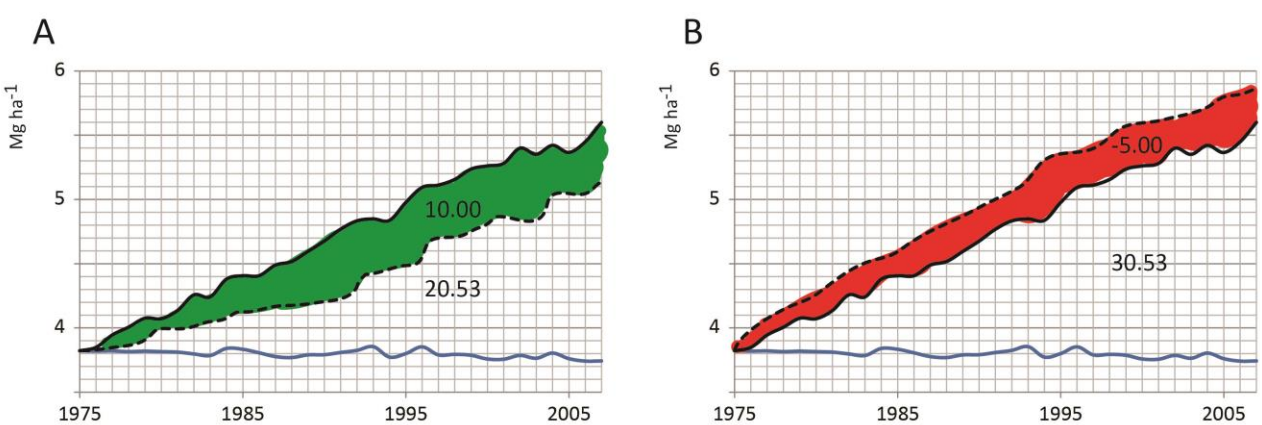

Fig. 2. Measuring the impact of an agricultural input on 1975 to mid 2000s global or regional yields. In (**A**) counterfactual *q* (one or more inputs are held fixed at 1975 levels in each country) produces an estimated global yield function, measured in Mg, given by the dotted black line. Assume the Riemann integral of the area between the expected global or regional yield curve (the solid black line) and the counterfactual global or region yield curve is 10.00. Further, assume the Riemann integral of the area between the expected global or regional yield curve (the solid black line) and the numeraire counterfactual yield curve (the solid blue line) is 30.53. Then counterfactual *q* explains 10 / 30.53 or 33% of the "global Mg gap." In (**B**) counterfactual *y* explains –5 / 30.53 or –16% of the "global Mg gap."

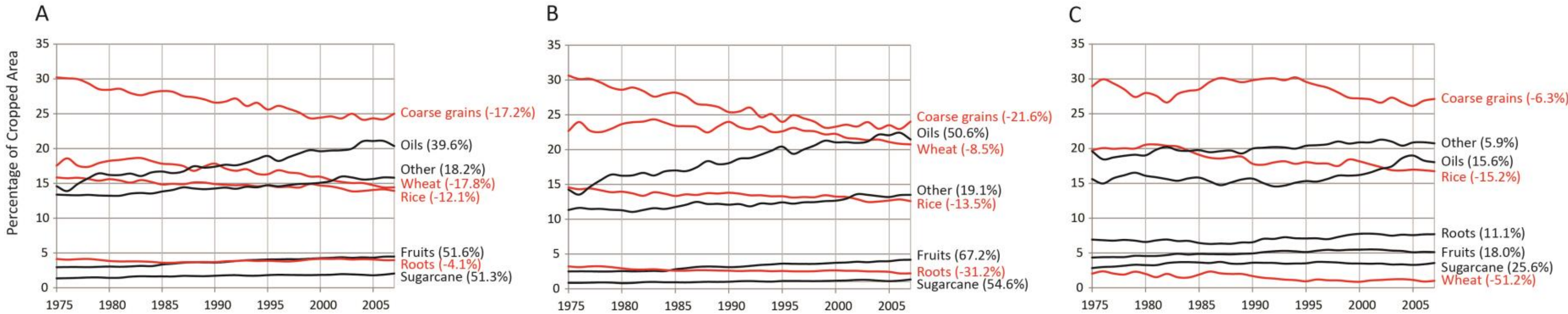

Fig. 3. Cropped area by crop type (crop mix) across the globe (A), across countries in the temperate region (B), and across countries in the tropical region (C). These graphs give the weighted average of area planted in each crop group across the globe or region over time. We use cropped hectarage in country *c* in year *t* as weights. Red (black) indicates a decrease (increase) in the crop or crop group's share in the overall mix between 1975 and 2007. The percentage change indicates the change between 1975 and 2007.

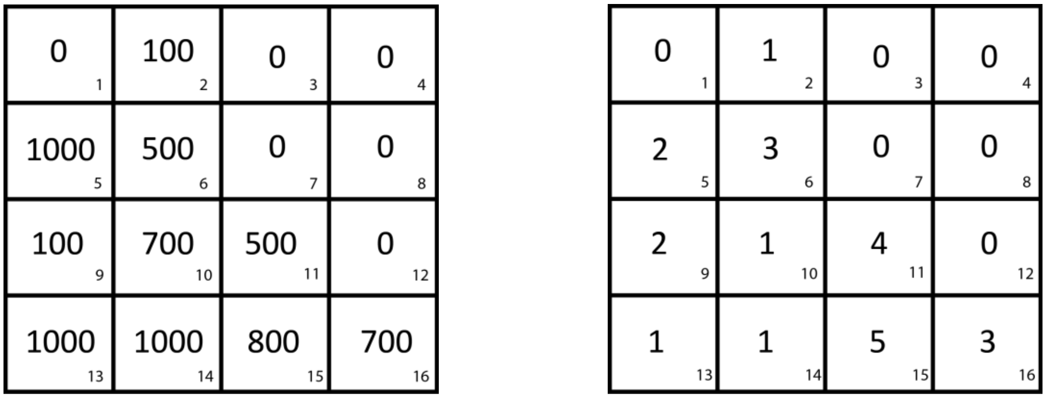

Fig. 4. Illustration of the calculation of the soil score for a country. Harvested hectares in each grid cell in an illustrative country (**A**) where the small numbers in the corner of a grid cell indicate cell ID. Nutrient availability score (N_{ct}) in each grid cell **(B)** where 1 indicates 'No or slight nutrient constraint', 2 indicates 'moderate nutrient constraint', 3 indicates 'severe nutrient constraint', 4 indicates 'very severe nutrient constraint', and 5 indicates 'mainly non-soil' (Fischer et al. 2008).

Table 1. The size of the area between the expected yield curve and a counterfactual's yield curve when fertilizer is included as an input ("wide" model results). The global model uses all countries while the regional models only use countries in the given region. The "Low" estimates are calculated with the 25th percentile annual yield estimates in each country. The "High" estimates are calculated with the 75th percentile annual yield estimates in each country. The cells in black indicate the integral if all agricultural inputs other than weather are fixed at 1975 levels (the numeraire counterfactuals; see Figs. 1 and 2). All other cells have an increasingly dark shade of green (red) as the integrals get more positive (negative). Pure white occurs at 0.

		Mg ha^{-1}			M kcals ha^{-1}		
Counterfactual	**Model**	**Low**	**Mean**	**High**	**Low**	**Mean**	**High**
No change other than weather	Globe	25.00	25.49	25.63	67.17	65.72	71.60
	Temperate	33.46	33.51	32.74	64.95	65.24	64.97
	Tropics	27.62	28.13	31.43	90.60	92.04	103.76
Soil quality of cropland	Globe	0.39	-0.08	-0.17	-9.55	-0.14	3.61
	Temperate	-0.14	0.07	-0.30	0.78	0.23	0.69
	Tropics	1.68	-0.03	0.04	5.51	0.54	-1.23
Area cultivated	Globe	-3.46	-3.35	-3.16	-3.21	3.51	5.33
	Temperate	-2.99	-1.95	-2.61	6.38	7.89	7.30
	Tropics	-0.37	-1.92	-2.06	3.11	-3.85	-4.93
Daytime growing season temp.	Globe	-1.02	-1.12	-0.72	-1.67	-2.86	3.50
	Temperate	-0.37	-0.05	0.24	-0.47	0.24	-0.28
	Tropics	-1.72	-2.95	-1.24	-12.72	-9.55	-8.68
Nighttime growing season temp.	Globe	1.56	1.24	1.98	-4.14	3.25	9.06
	Temperate	-2.29	-0.99	-0.91	-2.00	-1.94	-2.77
	Tropics	3.96	2.80	2.73	11.39	9.74	35.54
Growing season precipitation	Globe	-0.03	-0.01	0.51	2.63	-0.02	2.61
	Temperate	-0.95	-0.10	0.69	-0.43	-0.21	-1.16
	Tropics	0.75	0.01	0.29	-6.17	0.01	15.42
Crop mix	Globe	6.40	5.91	6.30	18.42	22.49	25.00
	Temperate	1.01	0.54	1.71	5.99	7.09	5.49
	Tropics	15.82	15.52	16.10	53.80	53.71	61.18
Irrigation capability	Globe	0.61	0.27	0.25	0.77	0.67	3.54
	Temperate	1.31	1.14	1.70	-0.32	1.31	1.22
	Tropics	0.77	0.50	0.87	-10.48	0.27	4.08
Investment in land and equipment	Globe	0.51	0.01	-0.23	3.43	-0.25	3.91
	Temperate	-1.23	-0.24	0.48	-1.69	-0.45	-2.62
	Tropics	3.10	1.71	3.30	9.14	7.03	-2.96
Time	Globe	14.87	14.46	14.96	24.55	24.07	21.91
	Temperate	25.99	26.40	27.03	40.93	40.63	39.84
	Tropics	0.30	-0.43	0.61	-10.13	-11.54	-16.39
Fertilizer	Globe	7.70	8.17	8.04	9.02	15.09	20.12
	Temperate	6.81	7.54	8.50	6.45	7.88	5.67
	Tropics	11.20	10.56	10.88	35.74	38.72	29.91

Table 2. The size of the area between the expected yield curve and a counterfactual's yield curve when fertilizer is not an input ("long" model results). See Table 1's legend for more details.

		Mg ha^{-1}			M kcals ha^{-1}		
Counterfactual	**Model**	**Low**	**Mean**	**High**	**Low**	**Mean**	**High**
No change other than weather	Globe	30.17	30.53	31.34	78.18	87.65	84.88
	Temperate	42.53	41.81	40.66	78.07	81.12	83.10
	Tropics	37.19	36.76	37.21	115.27	121.09	99.66
Soil quality of cropland	Globe	-1.80	-0.38	-1.84	-7.80	-0.34	-9.57
	Temperate	1.10	0.46	-0.59	-0.58	0.44	0.65
	Tropics	0.91	1.87	2.56	4.79	6.55	7.59
Area cultivated	Globe	-0.25	-0.70	0.80	2.56	14.28	16.32
	Temperate	1.93	2.17	1.31	13.30	15.00	17.67
	Tropics	-1.97	-1.82	-2.23	10.49	2.79	-0.45
Daytime growing season temp.	Globe	-2.36	-1.83	-1.06	-8.90	-4.09	-7.88
	Temperate	-1.62	-0.91	-2.46	-0.84	-1.36	-0.06
	Tropics	-3.32	-3.76	-3.55	-17.06	-12.24	-25.26
Nighttime growing season temp.	Globe	0.73	1.28	0.87	-2.43	2.93	9.62
	Temperate	-2.59	-1.63	-3.11	-1.34	-1.34	1.30
	Tropics	4.02	3.61	4.10	-0.49	9.06	-3.58
Growing season precipitation	Globe	-0.76	-0.01	-1.07	-4.28	-0.03	-4.74
	Temperate	-1.12	-0.03	-2.72	-2.29	-0.20	0.14
	Tropics	0.43	0.02	0.95	0.21	-0.01	-4.49
Crop mix	Globe	8.50	9.11	9.51	31.68	32.34	27.45
	Temperate	-1.29	-0.27	-0.68	11.16	8.50	9.03
	Tropics	23.27	22.49	23.21	69.19	78.88	66.39
Irrigation capability	Globe	-0.07	0.05	-0.61	-10.56	0.31	0.59
	Temperate	1.64	1.73	0.50	-0.48	1.43	0.96
	Tropics	1.90	0.82	2.16	-1.82	1.71	4.67
Investment in land and equipment	Globe	-1.02	0.16	-0.04	-5.77	-0.20	-4.22
	Temperate	-1.19	-0.30	-3.34	2.69	-0.54	2.08
	Tropics	2.79	2.03	1.87	-14.40	1.56	-16.07
Time	Globe	21.57	22.07	21.18	38.36	40.97	33.59
	Temperate	38.19	37.48	34.97	52.58	54.62	54.17
	Tropics	8.52	8.78	8.41	13.49	21.49	25.82

Table 3: Mean fertilizer values at the global and tropical and temperate regions levels (kg / cropped ha). All averages are weighted by cropped area in each country in each year.

	1975 – 77 average	2000 – 02 average	% Change
Globe	84.17	128.56	52.73%
Temperate	99.64	152.82	53.37%
Tropics	34.37	68.46	99.16%

Table 4. Mean values at the global and tropical and temperate regions levels. All averages are weighted by cropped area in each country in each year.

	1975 – 77 average	2005 – 07 average	% Change	1975 – 77 average	2005 – 07 average	% Change
	Hectares (Millions)			**Irrigation (Equipped ha / cropped ha)**		
Globe	7.23	8.86	22.54%	0.199	0.253	26.87%
Temperate	11.39	12.54	10.10%	0.233	0.318	36.55%
Tropics	3.57	5.63	57.55%	0.099	0.122	23.91%
	Soil score (a lower score means better nutrient availability and retention capacity)			**Equipment investment ($ M (2005) / 10,000 cropped ha)**		
Globe	1.51	1.56	2.82%	8.41	9.19	9.30%
Temperate	1.39	1.38	-0.38%	10.90	12.82	17.60%
Tropics	1.89	1.92	1.63%	1.22	1.95	59.54%
	Growing season daytime temp. (Celsius)			**Land development investment ($ M (2005) / 10,000 cropped ha)**		
Globe	27.68	29.06	4.98%	10.84	11.85	9.32%
Temperate	26.88	28.06	4.40%	9.81	11.24	14.58%
Tropics	29.87	30.90	3.45%	14.08	13.40	-4.84%
	Growing season nighttime temp (Celsius)			**Growing season precipitation (mm)**		
Globe	16.87	18.31	8.53%	115.16	113.09	-1.80%
Temperate	15.90	17.05	7.23%	125.83	128.69	2.27%
Tropics	19.64	20.77	5.75&	158.76	162.10	2.11%

Table 5. Summary of 12 decision trees.

	Dataset	Global		Temperate		Tropical	
		Annual change in Mg ha^{-1}	**Annual change in M kcals ha^{-1}**	**Annual change in Mg ha^{-1}**	**Annual change in M kcals ha^{-1}**	**Annual change in Mg ha^{-1}**	**Annual change in M kcals ha^{-1}**
<u>Tree accuracy</u>: **Percentage of predictions that are correct / cross-validation accuracy**	Long	56.8 / 50.8	57.3 / 50.4	61.9 / 58.2	62.9 / 59.7	51.3 / 46.1	51.6 / 43.5
	Wide	56.6 / 50.7	57.1 / 49.4	60.7 / 59.1	63.0 / 59.7	50.0 / 42.4	51.2 / 40.5
Number of branches on tree	Long	25	29	9	10	13	19
	Wide	21	23	4	5	11	18
Annual change explanatory variables in first 3 levels of a tree	Long	Sugarcane (Sugar); roots & tubers (R&T)	Sugar; wheat	Sugar; R&T; Area cultivated (A); Investment in land (Land)	Sugar; R&T	Sugar; R&T ; Daytime growing season temperature (DGST)	Sugar; DGST
	Wide	Sugar; Irrigation capability (I)	Sugar; fertilizer (F)	Sugar; A	Sugar; I	Sugar; DGST; Land	Sugar; A
<u>Heaviest branches</u>: **Percentage of all observations in tree on that branch and all predictive "rules" on the branch**	Long	21.8% -0.17 < Sugar ≤ 0.24 -0.67 < R&T ≤ 1.08 -0.35 < Wheat ≤ 0.23 -1.04 < DGST NGST ≤ 0.53 I ≤ 0.05 Land > 0	19.0% 0.05 < Sugar	47.3% -0.17 < Sugar ≤ 0.06 -0.67 ≤ R&T	29.8% -0.19 < Sugar ≤ 0.06 -0.66 < R&T -0.16 < Rice ≤ 0.82	30.9% Sugar ≤ 0.0 -0.68 < R&T ≤ 0.97 -0.06 < Oil -0.45 < Fruit -1.04 < DGST I ≤ 0.05 0.00 < Land ≤ 0	20.8% -0.13 < Sugar ≤ 0.31 -0.06 < Oil Rice ≤ 0.22 -1.09 < DGST ≤ 1.07 A ≤ -3662
	Wide	26.1% -0.17 < Sugar ≤ 0.29 Fruits ≤ 2.57 Wheat ≤ 1.25 -0.93 < R&T I ≤ 0.04 -7.75 < F ≤ 1.77	12.7% 0.18 < Sugar	54.8% -0.16 < Sugar ≤ 0.16	36.2% -0.16 < Sugar ≤ 0.09 0.00 < I	26.7% 0.00 < Sugar ≤ 0.01 -0.06 < Oil -0.97 < Other -0.46 < Fruit -1.00 < DGST Land ≤ 0 0.00 < Equipment	9.7% -0.13 < Sugar ≤ 0.20 -0.94 < Wheat ≤ 3.89 -0.79 < NGST -6.88 < Growing season precipitation -2354 < A ≤ 128055 0 < Land -8.13 < F

	Dataset	Global		Temperate		Tropical	
		Annual change in Mg ha^{-1}	**Annual change in M kcals ha^{-1}**	**Annual change in Mg ha^{-1}**	**Annual change in M kcals ha^{-1}**	**Annual change in Mg ha^{-1}**	**Annual change in M kcals ha^{-1}**
<u>Branch with greatest proportion of 'H'</u>: **Percentage of all observations in tree on that branch and all predictive "rules" on the branch**	Long	1.4% -0.17 < Sugar ≤ 0.24 -0.67 < R&T 0.66 < Wheat DGST ≤ -1.04 I ≤ 0.05	2.3% -0.17 < Sugar ≤ 0.05 0.55 < Wheat DGST ≤ -0.82	14.7% 0.29 < Sugar	27.8% 0.06 < Sugar	3.3% 0.22 < Sugar	2.5% 0.31 < Sugar
	Wide	9.8% 0.29 < Sugar	12.7% 0.18 < Sugar	23.4% 0.16 < Sugar	29.1% 0.09 < Sugar	6.0% 0.00 < Land	3.5% 0.20 < Sugar
<u>Branch with greatest proportion of 'L'</u>: **Percentage of all observations in tree on that branch and all predictive "rules" on the branch**	Long	11.7% Sugar ≤ -0.17	11.7% Sugar ≤ -0.17	15.3% Sugar ≤ -0.17 -9095 < A	18.2% Sugar ≤ -0.19	3.0% Sugar ≤ 0.00 R&T ≤ -0.68 -1.04 < DGST Irr ≤ 0.05 0.00 < Land	4.4% Sugar ≤ -0.13
	Wide	1.0% -0.17 < Sugar ≤ 0.29 2.57 < Fruits I ≤ 0.04	12.3% Sugar ≤ -0.16	17.5% Sugar ≤ -0.16 -8309 < A	21.7% Sugar ≤ -0.16	3.5% Sugar ≤ -0.16 -1.00 < DGST Land ≤ 0.00	7.1% -0.13 < Sugar ≤ 0.20 Wheat ≤ -0.33 Rice ≤ 0.03 0.13 < NGST A ≤ -2354

Notes: A high yield change ("H") in a country is given by a one year change of (0.158,10.1] Mg ha^{-1} or (0.354,30.2] M kcals ha^{-1} with the long dataset and (0.17,7.66] Mg ha^{-1} or (0.401,30.2] M kcals ha^{-1} with the wide dataset. A low yield change ("L") in a country is given by a one year change of ([-10.2,-0.0647] Mg ha^{-1} or [-30.7,-0.197] M kcals ha^{-1} with the long dataset and [-10.2,-0.0703] Mg ha^{-1} or [-30.7,-0.208] M kcals ha^{-1} with the wide dataset. Input names in black refer to crop mix inputs, names in red refer growing season weather inputs, and names in blue refer to other input types.

Supplementary Materials

Fig. S1. Decision tree for globe, yield measured in Mg ha^{-1}, using the "long" dataset.

Fig. S2. Decision tree for temperate region, yield measured in Mg ha^{-1}, using the "long" dataset.

Fig. S3. Decision tree for tropics, yield measured in Mg ha^{-1}, using the "long" dataset.

Fig. S4. Decision tree for globe, yield measured in M kcals ha^{-1}, using the "long" dataset.

Fig. S5. Decision tree for temperate region, yield measured in M kcals ha^{-1}, using the "long" dataset.

Fig. S6. Decision tree for tropics, yield measured in M kcals ha^{-1}, using the "long" dataset.

Fig. S7. Decision tree for globe, yield measured in Mg ha^{-1}, using the "wide" dataset.

Fig. S8. Decision tree for temperate region, yield measured in Mg ha^{-1}, using the "wide" dataset.

Fig. S9. Decision tree for tropics, yield measured in Mg ha^{-1}, using the "wide" dataset.

Fig. S10. Decision tree for globe, yield measured in M kcals ha^{-1}, using the "wide" dataset.

Fig. S11. Decision tree for temperate region, yield measured in M kcals ha^{-1}, using the "wide" dataset.

Fig. S12. Decision tree for tropics, yield measured in M kcals ha^{-1}, using the "wide" dataset.

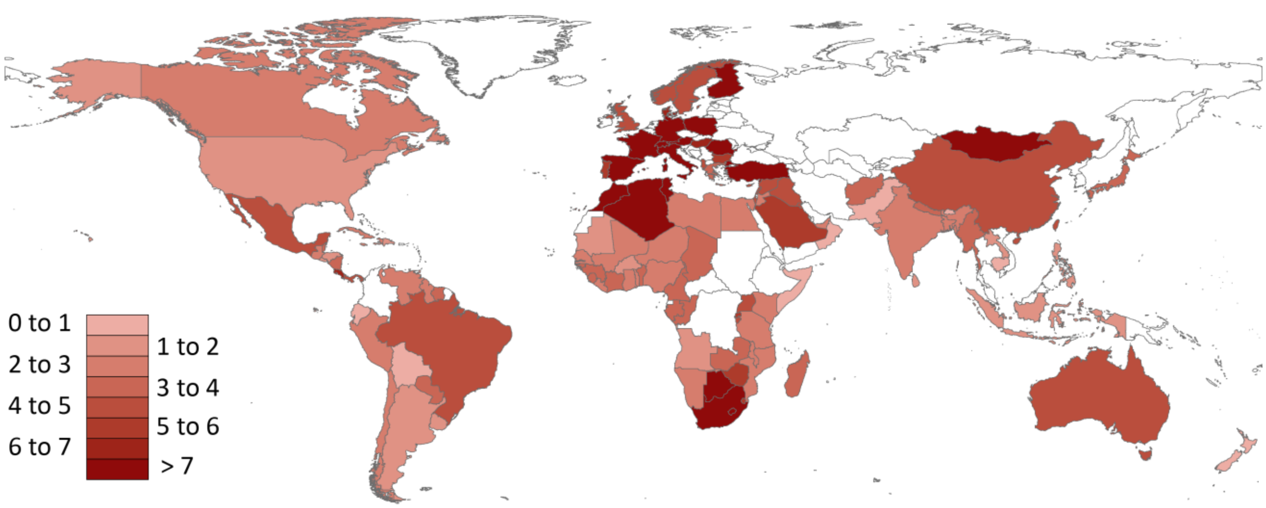

Fig. S13. Percentage change in 1975-1977 to 2005-2007 growing season daytime temperature by country.

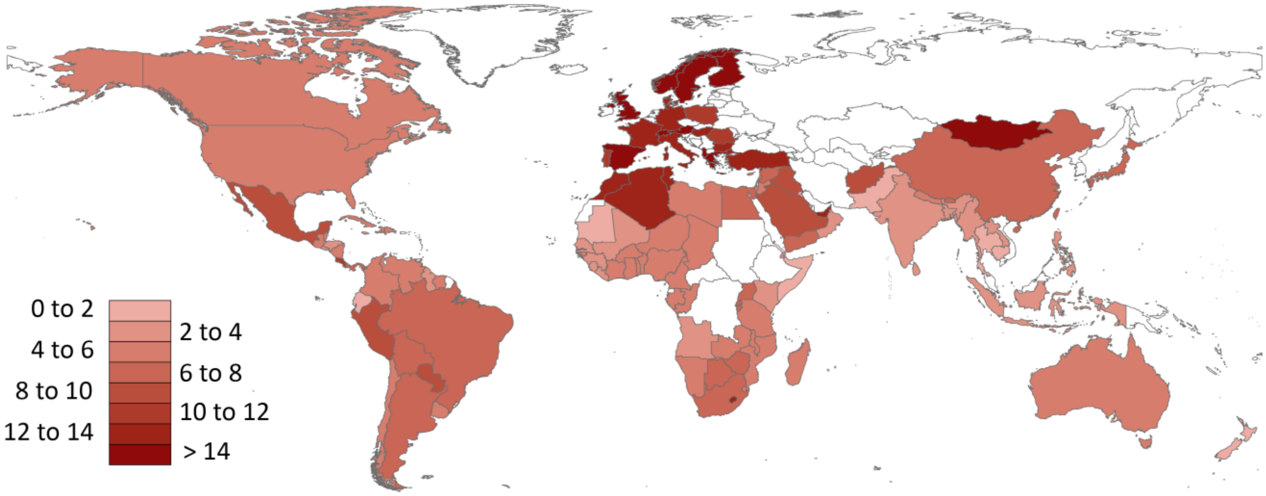

Fig. S14. Percentage change in 1975-1977 to 2005-2007 growing season nighttime temperature by country.

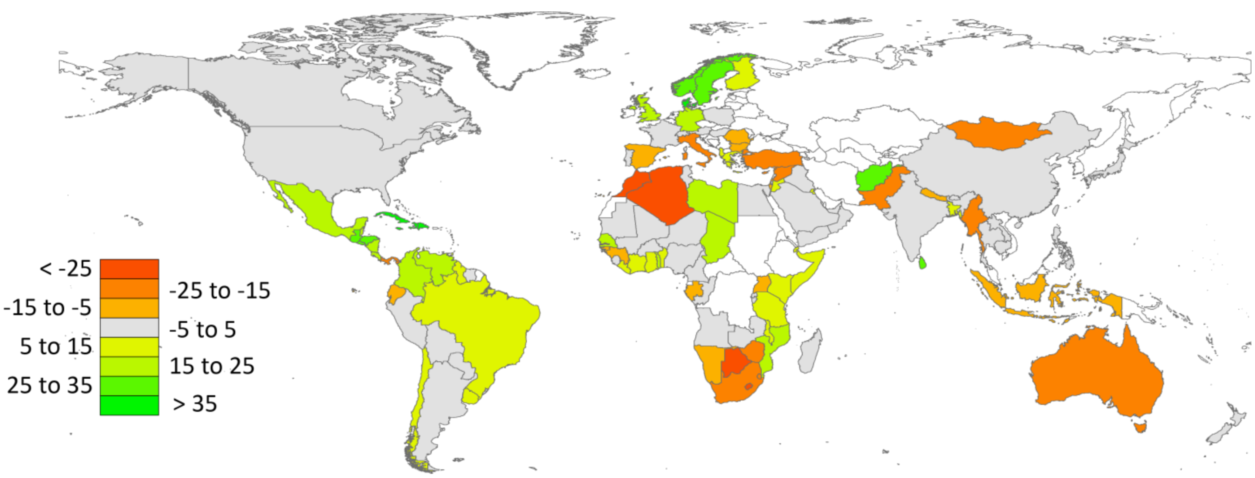

Fig. S15. Percentage change in 1975-1977 to 2005-2007 growing season precipitation by country.

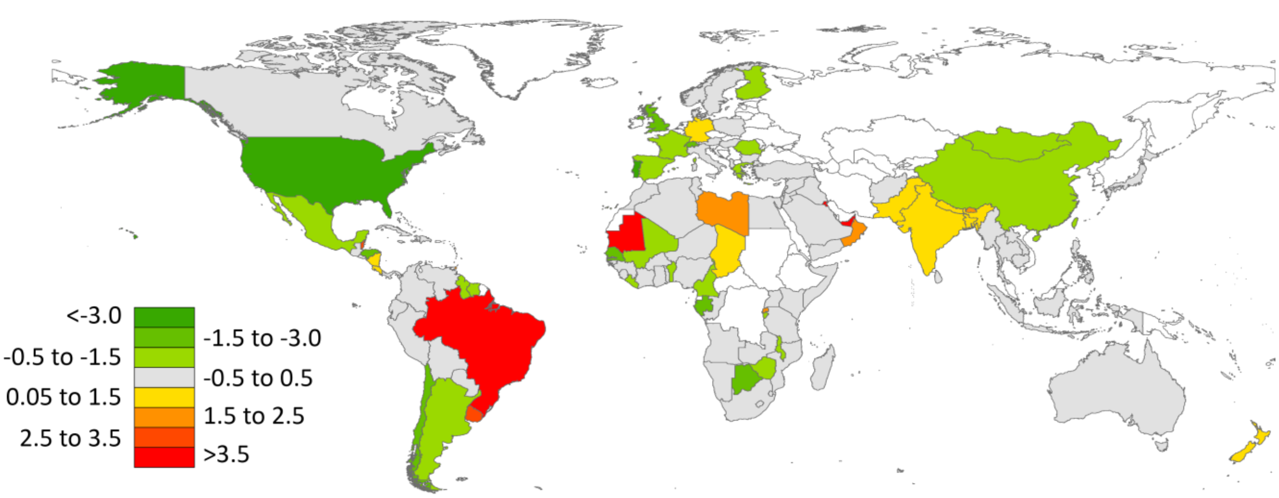

Fig. S16. Percentage change in 1975-1977 to 2005-2007 soil score by country.

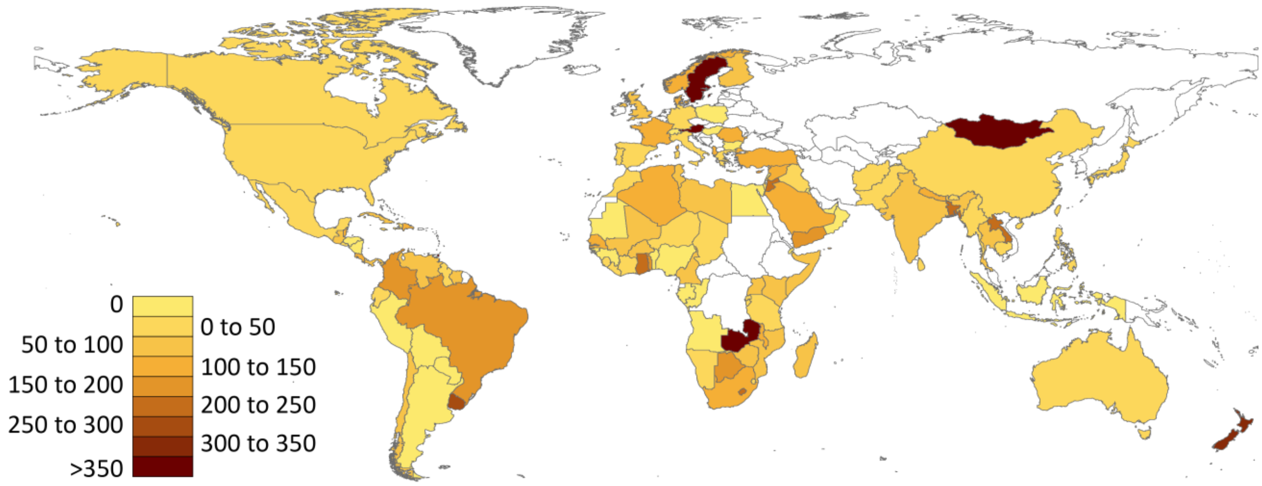

Fig. S17. Percentage change in 1975-1977 to 2005-2007 hectares of irrigation capacity per cropped hectare by country.

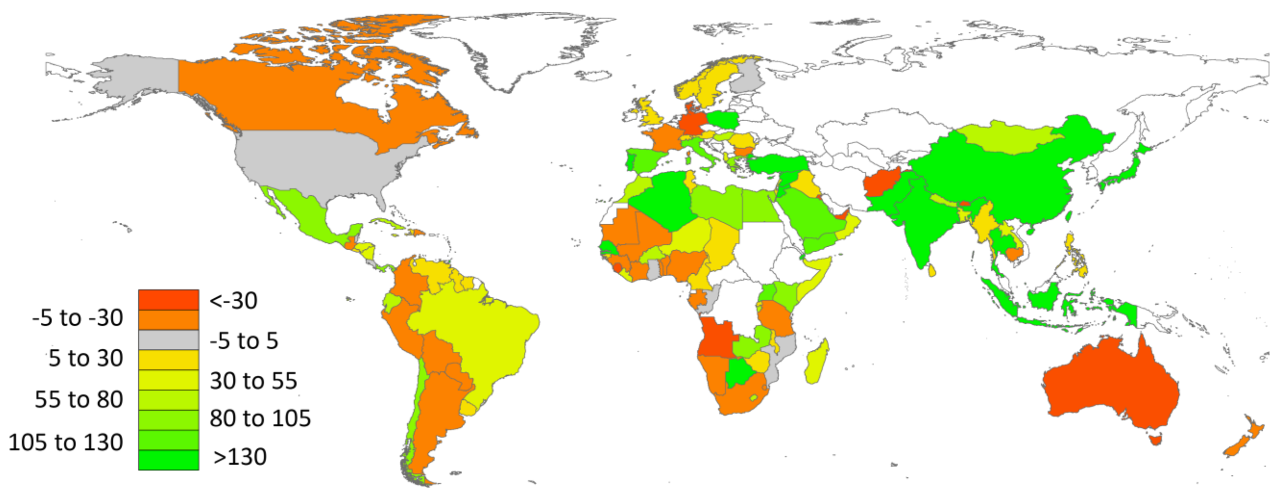

Fig. S18. Percentage change in 1975-1977 to 2005-2007 equipment investment ($ 2005) per cropped hectare by country.

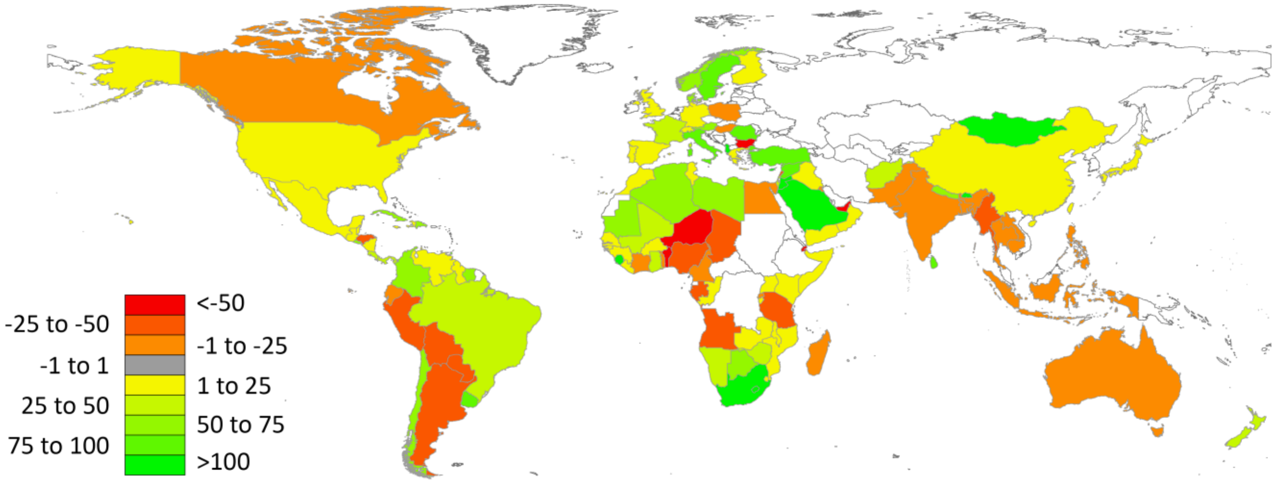

Fig. S19. Percentage change in 1975-1977 to 2005-2007 land investment ($ 2005) per cropped hectare by country.

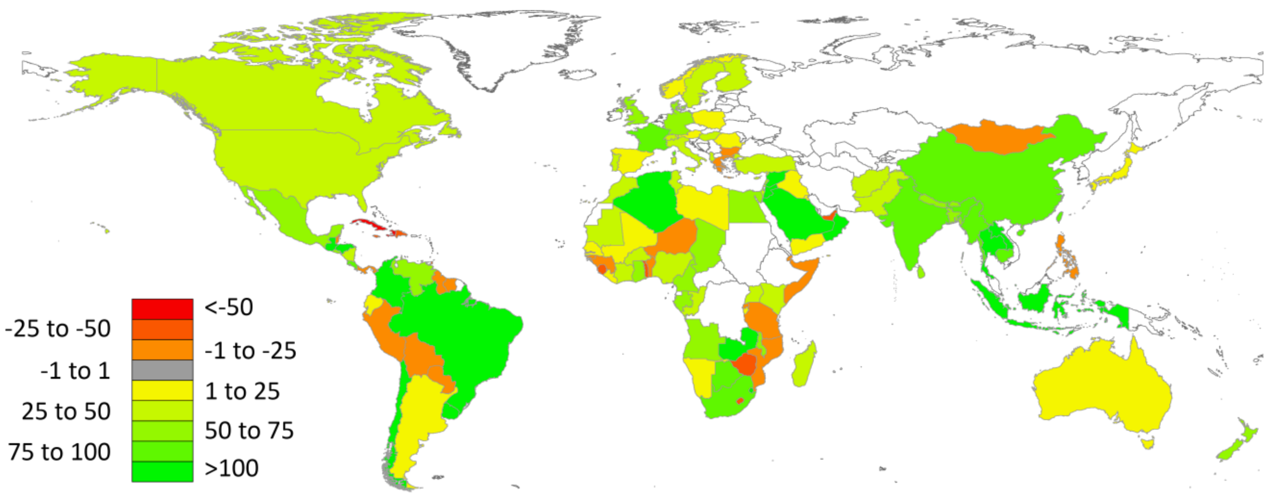

Fig. S20. Percentage change in 1975-1977 to 2005-2007 all crop M kcals per hectare yield by country.

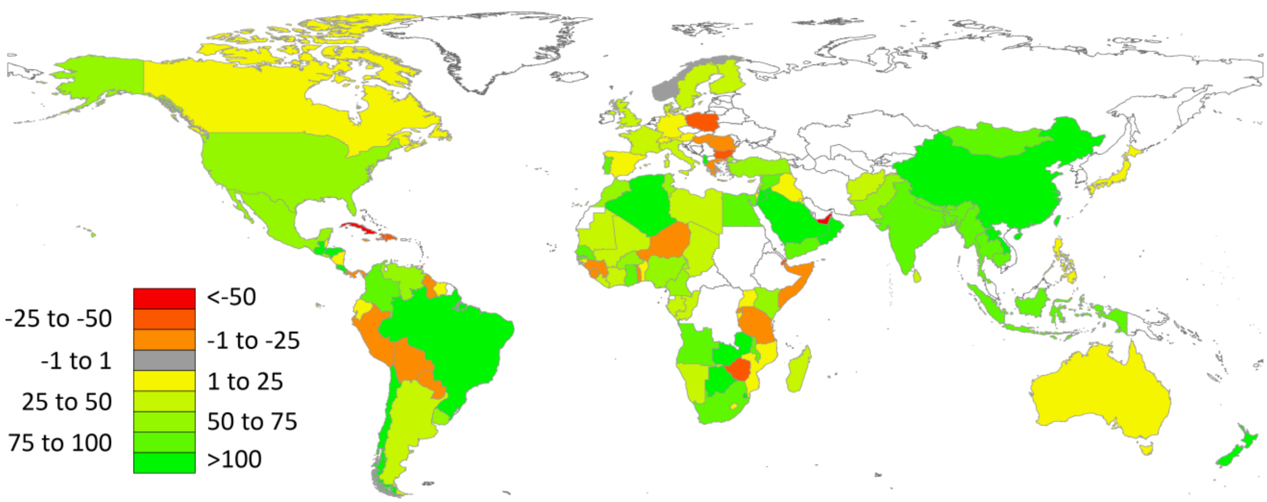

Fig. S21. Percentage change in 1975-1977 to 2005-2007 all crop Mg per hectare yield by country.

Table S1. Econometric estimates of fixed effects model (1) with the "long" global, tropics, and temperate datasets. Estimated coefficients with standard errors in parentheses. Standard errors are robust standard errors. '***' indicates statistical significance at p = 0.01, '**' indicates statistical significance at p = 0.05, and '*' indicates statistical significance at p = 0.10. Country fixed effect coefficients and SE are available upon request.

	(1)	(2)	(3)	(4)	(5)	(6)
Yield Measure	Mg/ha	Mg/ha	Mg/ha	M Kcals/ha	M Kcals /ha	M Kcals /ha
Scale	Globe	Tropics	Temperate	Globe	Tropics	Temperate
Soil Score	-7.389*** (2.855)	5.759*** (1.541)	-7.288** (3.356)	-6.613*** (2.425)	20.13*** (5.472)	-6.940*** (1.653)
Cropped area (1,000,000 ha)	0.017** (0.007)	0.037*** (0.007)	0.028*** (0.008)	0.095*** (0.002)	0.165*** (0.004)	0.009*** (0.001)
Percentage of cropped area in fruit production	0.052*** (0.017)	0.145*** (0.026)	-0.0214 (0.027)	0.132*** (0.032)	0.319*** (0.081)	0.032* (0.018)
Percentage of cropped area in grain (less wheat) production	-0.050*** (0.007)	-0.034*** (0.006)	-0.066*** (0.013)	-0.076*** (0.012)	-0.127*** (0.022)	-0.030*** (0.010)
Percentage of cropped area in oil crop production	-0.033*** (0.007)	2.61×10^{-5} (0.007)	-0.080*** (0.013)	-0.019 (0.013)	-0.005 (0.021)	-0.031*** (0.012)
Percentage of cropped area in rice crop production	0.006 (0.007)	0.007 (0.008)	0.002 (0.015)	0.077*** (0.019)	0.047* (0.028)	0.201*** (0.023)
Percentage of cropped area in roots and tubers production	0.036*** (0.013)	0.022 (0.014)	0.093*** (0.029)	-0.061** (0.029)	-0.088** (0.043)	-0.057* (0.031)
Percentage of cropped area in sugarcane production	0.558*** (0.022)	0.551*** (0.027)	0.719*** (0.032)	2.338*** (0.077)	2.221*** (0.101)	3.158*** (0.092)
Percentage of cropped area in wheat production	-0.043*** (0.008)	-0.250*** (0.033)	-0.066*** (0.014)	0.037*** (0.013)	-0.778*** (0.120)	0.057*** (0.011)
Growing season daytime temperature (degrees Celsius)	-0.668*** (0.255)	-1.422** (0.638)	-0.641** (0.295)	0.301 (0.351)	-5.101*** (1.880)	0.169 (0.327)
The square of growing season daytime temperature	0.009* (0.005)	0.0184* (0.011)	0.011* (0.006)	-0.013* (0.007)	0.068** (0.030)	-0.006 (0.006)
Growing season nighttime temperature (degrees Celsius)	0.152 (0.188)	0.483 (0.372)	-0.087 (0.252)	-0.709** (0.336)	2.362** (1.138)	-0.983*** (0.283)
The square of growing season nighttime temperature	-0.002 (0.006)	-0.006 (0.010)	-0.001 (0.010)	0.027** (0.012)	-0.045 (0.030)	0.029*** (0.010)
Growing season precipitation (mm)	0.001 (.001)	0.001 (0.002)	0.001 (0.002)	0.002 (0.005)	0.001 (0.006)	0.003 (0.005)
The square of growing season precipitation	-1.51×10^{-6} (1.48×10^{-6})	-1.22×10^{-6} (1.43×10^{-6})	-2.50×10^{-6} (3.61×10^{-6})	-4.16×10^{-6} (4.57×10^{-6})	-2.14×10^{-6} (5.01×10^{-6})	-1.4×10^{-5}* (7.90×10^{-6})
Hectares equipped for irrigation / cropped area	0.047 (0.235)	1.174* (0.612)	1.269*** (0.294)	0.279* (0.160)	2.460 (2.261)	1.049*** (0.349)
Equipment investment ($ M (2005) / cropped ha)	22.26*** (7.259)	840.3** (423.0)	1.265 (7.47)	-32.60** (14.39)	652.8 (1206)	-41.14*** (14.68)

	(1)	(2)	(3)	(4)	(5)	(6)
Yield Measure	Mg/ha	Mg/ha	Mg/ha	M Kcals/ha	M Kcals /ha	M Kcals /ha
Scale	Globe	Tropics	Temperate	Globe	Tropics	Temperate
Land development investment ($ M (2005) / cropped ha)	5.52*** (1.002)	2.301 (2.176)	-115.9*** (20.83)	5.063*** (0.831)	-3.620 (7.461)	-71.69** (28.76)
Year	0.042*** (0.002)	0.017*** (0.003)	0.071*** (0.004)	0.078*** (0.006)	0.041*** (0.010)	0.103*** (0.006)
Constant	-55.87*** (6.874)	-27.49*** (9.236)	-112.2*** (8.766)	-136.0*** (12.85)	-63.45** (28.81)	-185.9*** (11.49)
N	4533	2182	2224	4533	2182	2224
F-value		757.76***		639.33***	644.00***	675.67***
Root MSE	1.134	1.124	1.018	3.146	4.007	1.451

Table S2. Econometric estimates of fixed effects model (1) with the "wide" global, tropics, and temperate datasets

	(1)	(2)	(3)	(4)	(5)	(6)
Yield Measure	Mg/ha	Mg/ha	Mg/ha	M Kcals/ha	M Kcals /ha	M Kcals /ha
Scale	Globe	Tropics	Temperate	Globe	Tropics	Temperate
Soil Score	-2.605 (3.583)	-0.142 (1.972)	-1.477 (3.937)	-4.526* (2.517)	2.315 (7.071)	-4.862*** (1.776)
Cropped area (1,000,000 ha)	-0.009 (0.006)	0.030*** (0.008)	-0.001 (0.008)	0.036** (0.015)	0.112*** (0.034)	0.069*** (0.011)
Percentage of cropped area in fruit production	0.037* (0.021)	0.117*** (0.025)	-0.029 (0.028)	0.122*** (0.039)	0.213** (0.088)	0.037** (0.017)
Percentage of cropped area in grain (less wheat) production	-0.035*** (0.007)	-0.025*** (0.007)	-0.050*** (0.011)	-0.042*** (0.013)	-0.094*** (0.023)	0.002 (0.011)
Percentage of cropped area in oil crop production	-0.021*** (0.007)	-0.001 (0.007)	-0.063*** (0.012)	-0.0004 (0.014)	-0.009 (0.025)	-0.007 (0.012)
Percentage of cropped area in rice crop production	0.004 (0.009)	0.017 (0.011)	-0.024 (0.016)	0.066*** (0.024)	0.084 (0.039)	0.174*** (0.022)
Percentage of cropped area in roots and tubers production	0.035** (0.014)	0.059*** (0.015)	0.055 (0.036)	-0.054 (0.034)	0.089* (0.048)	-0.121** (0.033)
Percentage of cropped area in sugarcane production	0.535*** (0.034)	0.524*** (0.034)	0.747*** (0.046)	2.217*** (0.129)	2.048*** (0.129)	3.206*** (0.133)
Percentage of cropped area in wheat production	-0.030*** (0.008)	-0.221*** (0.037)	-0.054*** (0.013)	0.051*** (0.014)	-0.712*** (0.138)	0.074*** (0.011)
Growing season daytime temperature (degrees Celsius)	-0.361 (0.237)	-2.413*** (0.703)	-0.095 (0.273)	0.467 (0.431)	-4.448* (2.359)	0.576* (0.310)
The square of growing season daytime temperature	0.004 (0.005)	0.035*** (0.012)	0.002 (0.006)	-0.016* (0.008)	0.056 (0.040)	-0.010* (0.006)
Growing season nighttime temperature (degrees Celsius)	0.566*** (0.177)	1.03** (0.418)	0.031 (0.233)	0.279 (0.398)	2.354 (1.452)	-0.883*** (0.265)
The square of growing season nighttime temperature	-0.013** (0.006)	-0.020* (0.012)	-0.004 (0.009)	0.001 (0.014)	-0.038 (0.041)	0.022** (0.009)

	(1)	(2)	(3)	(4)	(5)	(6)
Yield Measure	Mg/ha	Mg/ha	Mg/ha	M Kcals/ha	M Kcals /ha	M Kcals /ha
Scale	Globe	Tropics	Temperate	Globe	Tropics	Temperate
Growing season precipitation (mm)	0.001 (0.001)	0.001 (0.002)	0.006*** (0.002)	-0.001 (0.005)	0.001 (0.006)	0.008* (0.004)
The square of growing season precipitation	$-1.15x10^{-6}$ $(1.27x10^{-6})$	$-9.1x10^{-7}$ $(1.39x10^{-6})$	$-8.94x10^{-6}$** $(3.83x10^{-6})$	$-7.67x10^{-7}$ $(4.35x10^{-6})$	$-7.00x10^{-7}$ $(5.06x10^{-6})$	$-1.74x10^{-5}$** $(7.19x10^{-6})$
Hectares equipped for irrigation / cropped area	0.332** (0.166)	0.943 (0.850)	1.220*** (0.374)	0.844*** (0.166)	0.521 (3.14)	1.41* (0.789)
Equipment investment ($ M (2005) / cropped ha)	10.595 (6.674)	870.081 (672.122)	-7.621 (7.056)	-34.212** (16.038)	3515.753 (2452.381)	-42.54*** (14.815)
Land development investment ($ M (2005) / cropped ha)	-25.804*** (2.564)	-24.913 (19.505)	-104.215*** (28.419)	-34.024*** (5.374)	-47.715 (69.102)	-92.282 (66.691)
Year	0.038*** (0.003)	-0.001 (0.004)	0.070*** (0.004)	0.064*** (0.008)	-0.031** (0.015)	0.107*** (0.007)
Fertilizer (kg / ha)	0.008*** (0.001)	0.017*** (0.002)	0.007*** (0.001)	0.016*** (0.002)	0.062*** (0.009)	0.007*** (0.001)
The square of fertilizer	$-1.39x10^{-6}$*** $(1.84x10^{-7})$	$-4.93x10^{-6}$*** $(1.67x10^{-6})$	$-1.12x10^{-6}$*** $(1.74x10^{-7})$	$-3.54x10^{-6}$*** $(4.74x10^{-6})$	$-2.08 x10^{-5}$*** $(6.03x10^{-6})$	$-1.46 x10^{-6}$*** $(1.80x10^{-7})$
Constant	-64.962*** (7.698)	32.256*** (11.44)	-128.605*** (9.254)	-123.322*** (16.228)	110.702*** (38.928)	-206.105*** (13.691)
N	3611	1731	1816	3611	1731	1816
F-value		901.38***			601.84***	
Root MSE	0.994	1.051	0.837	3.051	3.864	1.307

Supplementary Methods: Crop groups used to define crop mix
Rice (FAO code 27).

Wheat (FAO code 15).

Sugarcane (FAO code 156).

Coarse grains includes crops with FAO codes 44, 677, 56, 79, 75, 71, 83, 97.

Oil crops includes crops with FAO codes 265, 249, 242, 336, 263, 333, 299, 292, 254, 339, 260, 296, 270, 280, 328, 289, 236, 267, 275, 311, and 329.

Fruits (not including melons) includes crops with FAO codes 515, 526, 486, 552, 461, 591, 531, 512, 554, 550, 577, 569, 619, 603, 549, 507, 560, 592, 497, 571, 490, 600, 534, 521, 587, 574, 489, 536, 523, 547, 530, 541, 544, 495, and 558.

Roots and tubers includes crops with FAO codes 125, 116, 149, 122, 136, 137, and 135.

All other crops.

See the FAOStat website for crop codes and their crop names.

Data file S1. "Wide" dataset.

1. ID: UNFAO Country Code
2. Year
3. Tropical: a 1 indicates that that country is a tropical country and a 0 indicates that the country is a temperate country
4. tons / ha: a country's crop yield in year t in metric tons / ha (I summed all tons of crops produced in a country and divided by total cropped hectares in a country)
5. million kcals / ha: a country's crop yield in year t in millions of kcals / ha (I summed all kcals of crops produced in a country and divided by total cropped hectares in a country)
6. soilscore: The composite soil quality score of the land that was cropped in year t in country k (on a 1 to 5 scale with lower numbers indicating better soil).
7. ha: total cropped hectares in year t in country k
8. rice: percentage of cropped area in rice in year t in country k
9. wheat: percentage of cropped area in wheat in year t in country k
10. sugar: percentage of cropped area in sugarcane in year t in country k
11. grains: percentage of cropped area in coarse grains in year t in country k
12. oil: percentage of cropped area in oil crops in year t in country k
13. fruits: percentage of cropped area in fruits in year t in country k
14. roots: percentage of cropped area in roots and tubers in year t in country k
15. other: percentage of cropped area in all other crops in year t in country k
16. davg: The composite average daytime temperature over cropped lands during the growing season year t in country k (Celsius)
17. navg: The composite average nighttime temperature over cropped lands during the growing season year t in country k (Celsius)
18. pavg: The total rainfall over cropped lands during the growing season year t in country k (mm)
19. irr: Fraction of cropped lands that are equipped for irrigation in year t in country k
20. land: total money invested in agricultural land development divided by cropped hectares in year t in country k (2005 constant US $ / ha)
21. eqp: total money invested in agricultural equipment divided by cropped hectares in year t in country k (2005 constant US $ / ha)
22. fert: kilograms of fertilizer used in the country divicde by cropped hectares in year t in country k.

Data file S2. "Long" dataset.

1. ID: UNFAO Country Code
2. Year
3. Tropical: a 1 indicates that that country is a tropical country and a 0 indicates that the country is a temperate country
4. tons / ha: a country's crop yield in year t in metric tons / ha (I summed all tons of crops produced in a country and divided by total cropped hectares in a country)
5. million kcals / ha: a country's crop yield in year t in millions of kcals / ha (I summed all kcals of crops produced in a country and divided by total cropped hectares in a country)
6. soilscore: The composite soil quality score of the land that was cropped in year t in country k (on a 1 to 5 scale with lower numbers indicating better soil).
7. ha: total cropped hectares in year t in country k
8. rice: percentage of cropped area in rice in year t in country k
9. wheat: percentage of cropped area in wheat in year t in country k
10. sugar: percentage of cropped area in sugarcane in year t in country k
11. grains: percentage of cropped area in coarse grains in year t in country k

12. oil: percentage of cropped area in oil crops in year t in country k
13. fruits: percentage of cropped area in fruits in year t in country k
14. roots: percentage of cropped area in roots and tubers in year t in country k
15. other: percentage of cropped area in all other crops in year t in country k
16. davg: The composite average daytime temperature over cropped lands during the growing season year t in country k (Celsius)
17. navg: The composite average nighttime temperature over cropped lands during the growing season year t in country k (Celsius)
18. pavg: The total rainfall over cropped lands during the growing season year t in country k (mm)
19. irr: Fraction of cropped lands that are equipped for irrigation in year t in country k
20. land: total money invested in agricultural land development divided by cropped hectares in year t in country k (2005 constant US $ / ha)
21. eqp: total money invested in agricultural equipment divided by cropped hectares in year t in country k (2005 constant US $ / ha)

Data file S3. Accuracy of decision trees.