SUPPLEMENTARY METHODS

Mapping agricultural management: A global dataset of crop-specific nitrogen, phosphate, and potash fertilizer application rates and consumption

To develop our crop-specific dataset of nitrogenous, phosphate, and potash fertilizer application for 138 crops and pasture, we used a spatial disaggregation method building on the work of Potter et al.[1] to fuse both national and, where available, sub-national data from a variety of sources (Table S2).

Data collection

We collected national and sub-national data on fertilizer application rates for crops and crop groupings. A major source of data was the fifth edition of "Fertilizer Use by Crop" (hereafter referred to as the FUBC5 dataset), a joint publication from the International Fertilizer Industry Association (IFA), the International Fertilizer Development Center (IFDC), the International Potash Institute (IPI), the Phosphate and Potash Institute (PPI), and the Food and Agriculture Organization of the United Nations (FAO)[2]. The publication contains national-level application rate data by crop for 42 countries, compiled from the following data sources: FAO questionnaires given to member countries; IFA questionnaires given to industry companies, research institutes, and fertilizer associations; IFDC questionnaires sent to experts attending courses, seminars, and professional meetings; and IPI and PPI communications with experts. Most of the application rate data from the FUBC5 dataset are for the years 1999 or 2000, but data for some countries are as old as 1994 and as recent as 2001.

To expand spatial and sub-national data coverage, we also collected data from national statistical bureaus, FAO reports, and national-level fertilizer industry associations (see later sections for more description on how these datasets were compiled and harmonized). Following Monfreda et al.[3], we established a 7-year data collection window centered on the year 2000 (1997-2003). We calculated averages for countries when data for multiple years was available within this window. When countries did not have data available within our desired timeframe (as was the case for some countries in the FUBC5 dataset), we used data from the year closest to our data collection window. For some countries, the only fertilizer information available was FAO nutrient consumption data[4], which we collected for all countries available. Data sources are listed in Table S2. Sub-national data was all provided at the state/province-level, except for the US AAPFCO data[5], which we aggregated from the county-level to the state-level for consistency. The countries for which we compiled sub-national data represent 45%, 50%, and 55% of total global N, P_2O_5, and K_2O consumption, respectively (FAO nutrient consumption from 1997-2003, 10).

We next identified "data gaps" for each crop category: countries where we had crop areas but no fertilizer application rate data in our database. As fertilizer use is highly

correlated to income-level[6], we chose to use an income-based extrapolation technique to fill these data gaps. Countries were grouped into four economic aggregates based on the World Bank [7] income classifications: low income, lower middle income, upper middle income, and high income (both OECD and non-OECD countries). For each crop, we calculated area-weighted average fertilizer application rates for each economic aggregate. We then identified the economic group of each country missing application rate data and filled gaps using the average application rates. The poorer data quality of extrapolated rates was noted accordingly in a data quality map corresponding to each crop.

For some crops, we lacked observational data on application rates from any country within a particular economic group. In these cases we calculated the area-weighted average application rate across the entire globe and utilized this rate to extrapolate to areas missing data. While clearly not ideal, having an application rate value for each crop, even if it is of low quality, allows us to scale the application rates to match total FAO nutrient consumption in a country. This allows us to gain a first-order approximation of the true application rate.

We collected fertilizer data (from either a crop-specific or crop-group-specific application rate) for 138 crops and pasture. No tabular fertilizer information in any country was available for some minor crops for which we did have harvested area data (the M3 crop area dataset contains data for 175 crops)[3]. We disregarded these crops in our dataset and assumed negligible fertilizer consumption.

Mapping of application rate information

As with previous studies, our approach matches spatial data on agricultural land use with tabular application rate data for particular crops or crop groups. Potter *et al.*[1] linked cropland or crop-group maps from the M3 croplands dataset [3] and the M3 pasture dataset[8] to each national-level application rate data entry in the FUBC5 dataset. We used and revised the Potter *et al.* linkages, especially focusing on which Monfreda *et al.* (2) datasets were used for crop groupings. For example, in Morocco, Potter *et al.* distributed FUBC5 application rate data for the category "oil crops, other" onto the Monfreda *et al.* crop map for "oilseeds, other". Since the only oil crop with its own application rate data listed in the IFA/FAO/IFDC report is sunflower, we chose instead to distribute the fertilizer application rates for the "oil crops, other" category onto all the Monfreda *et al.* oil crop maps except sunflower (this includes not only the "oilseeds, other" category, but also soybeans, sesame seed, safflower seed, etc.). The same method was applied to identify constituent crops for all crop groups.

In most cases, national and sub-national fertilizer application rates from our data (Table S2) were first directly applied to the appropriate crop maps. We modified the raw application rates at this step in three cases: 1) if a data source indicated that only a percentage of a particular cropland area was fertilized, 2) if the fertilized pasture area in

a country was less than the pasture area for that country from the M3 pasture dataset, and 3) if we had data for seasonal varieties of barley and wheat. Below are the adjustments made for these three special cases:

1. Consistent with the Potter *et al.*[1] methodology, when only a percentage of a cropland area was fertilized we adjusted application rates downward by the same percentage. For example, the FUBC5 dataset indicates that 85% of Mexico's avocadoes are fertilized at an average rate of 120 kg N/ha, so we applied an application rate of 102 kg N/ha to all of Mexico's avocado area.

2. Similar to case 1), in many cases only a percentage of pastureland in a country was fertilized. While this percentage was not explicit in the FUBC5 dataset, we calculated this number by comparing the FUBC5 fertilized pasture area with the total M3 pasture areas within each country[8]. For areas where the M3 pasture areas were greater than FUBC5 pasture areas, we reduced application rates by the proportion of FUBC5 pasture area to M3 pasture area (i.e. if FUBC5 listed half the pasture area contained in the M3 dataset, we reduced the FUBC5 pasture application rates by half).

3. For seasonal varieties of wheat and barley, we calculated average "wheat" and "barley" application rates, weighting the FUBC5 seasonal crop application rates by the FUBC5 seasonal areas.

Harmonize with FAO consumption dataset

To harmonize our dataset with 1997-2003 FAO national nutrient consumption data (10), we first calculated initial estimates for global consumption of N, P_2O_5, and K_2O by multiplying our crop application rate maps by M3 crop areas. We differentiated between "trusted crops" – crops for which we have sub-national or national-level application rate information – and "untrusted crops" – crops for which application rates were derived through the aforementioned extrapolation procedure. In most cases we trusted the application rates from our trusted crops, and thus we only scaled untrusted crop application rates up or down to match average FAO total national nutrient consumption (note that the same scalar was applied to all untrusted crops). Two special cases led us to have less trust in our "trusted crop" consumption and we altered our scaling procedure:

1. When the scaling correction for untrusted crops required more than a doubling of those application rates within a country, we chose to scale the application rates of all crops to meet FAO consumption levels. In a few small countries, we also capped the scalar for all crops at a doubling of application rates. In these cases our data were not reconciling either due to underreporting of cropland area, crops missing from our dataset, or errors with either the application rate data or the FAO consumption data.

2. When the total fertilizer consumption summed over trusted crops alone already exceeded or nearly exceeded (>95%) the FAO consumption within a country, we

adjusted the scaling procedure by scaling the application rates of all crops to match the FAO consumption. Again, this is another case where our multiple datasets were not reconciling due to one of the above possible complications.

No FAO consumption data was available for Gibraltar, Liechtenstein, Western Sahara, and 37 small island countries and territories. For these locations, consumption was recorded as "not a number". Application rate data remained un-scaled and was noted accordingly in the data quality map.

Enhance sub-national resolution

Sub-national consumption and aggregate application rate data, when available, was used to add spatial resolution to our national application rate data. Consumption data came in three main forms: 1) total nutrient consumption in each sub-national unit, 2) fertilizer consumption by type (i.e. "nitrogenous" or "compound") in each sub-national unit, and 3) average nutrient application rates (across all crops) in each sub-national unit. We multiplied average application rates by the number of potentially fertilized hectares (as defined by the sum of the crop proxy and pasture maps) to obtain nutrient consumption in each sub-national unit. Then, for all countries except the US, we harmonized the sum of the sub-national consumption data for each nutrient (including compound fertilizers when available) by scaling it to match the FAO national consumption data. In the US, sub-national consumption data was already listed in units of N, P_2O_5, and K_2O, but it could not be compared to FAO consumption because the data did not have national coverage. Due to this complication, we used the US sub-national consumption data directly without calibration to FAO.

Next, we added up consumption according to our application rate and area maps in each sub-national unit. Application rates for all crops, except those for which we had sub-national application rate data, were scaled so that the sum of all consumption in the sub-national unit matched the sub-national consumption data. Scalars were allowed to vary ±25%, since we observed that variation from the median rate commonly varied ±25% in countries where we had sub-national data. Sub-national application rates were not scaled using the sub-national consumption data.

The sub-national scaling cap of ±25% can slightly affect consistency with the FAO consumption dataset. Thus, for countries where we calculated and used sub-national consumption scalars, we once again scale all application rates – except those originally from sub-national data sources – to match FAO consumption data.

Record data quality

The quality of application rate data varies substantially across the globe due to the availability of input data. For example, an application rate may come directly from unaltered sub-national data, it could be a national-level application rate scaled by sub-

national consumption data, an extrapolated rate from similar-income countries normalized to FAO consumption, etc. Thus, we recorded data quality in a data type map for each nutrient and crop combination that details the quality of the input data and the manipulations made (if any) to record or estimate fertilizer application rate at every location where that crop is cultivated. Data type is indicated in each map through a unique numerical code.

Calculating climate bins and yield gaps

To calculate attainable yields (AYs) and yield gaps for our 17 major crops (wheat, rice, maize, soybean, barley, sorghum, millet, cotton, rapeseed, groundnut, sunflower, sugarcane, potato, cassava, oil palm, rye, and sugar beet), we build on recently developed climate analog techniques[9,10]. Crop and climate data used in these analyses are all at the 5 arc-minute by 5 arc-minute resolution.

We calculate unique climate bins for each crop by establishing 100 zones of similar annual precipitation and growing degree-day (GDD) characteristics, where each of the 100 zones contains equal harvested area (see section "Comparison to previous yield gap analyses" for a discussion about using annual precipitation as the moisture variable). We use interpolated mean daily temperatures from the WorldClim dataset[11] and crop-specific base temperatures from Licker et al.[9] to derive growing degree-days using methods described by Licker et al.[9]. Mean annual precipitation (P) is directly derived from the WorldClim dataset[11].

Using these two variables, we discard grid cells that are climate outliers by defining a compact contour in precipitation-GDD space containing 95% of a crop's harvested area (crop area and yield data is from Monfreda et al.)[3]. This contour is derived in several steps. First, we define a 2-dimensional histogram of harvested area in a precipitation-GDD space with 300x300 bins. A smoothed distribution is then constructed by convolution of the 2-D histogram with a Gaussian distribution G, defined in Eqn. S1, where the smoothing lengths L_P and L_{GDD} are chosen as $1/10^{th}$ of the span of the crop-specific domain of precipitation and GDD values, respectively. N equals the normalization constant.

$$G(GDD,P) = Ne^{\left(-\frac{P^2}{L_P^2}-\frac{GDD^2}{L_{GDD}^2}\right)} \qquad \text{(Eqn. S1)}$$

A contour is defined which includes 95% of the smoothed area distribution. If this step results in multiple contours, the smoothing convolution is recomputed with the smoothing lengths L_P and L_{GDD} 10% larger. This convolution is recalculated with progressively larger smoothing lengths until a single contour is found that contains 95% of the crop's harvested area. This contour represents a climatic envelope in which we feel comfortable calculating attainable yields from our dataset. Attainable yields and

intensification potential are only calculated for the grid cells characterized by climates within this climate envelope.

Once a contour is identified, we identify 10 zones along the GDD axis that each contain 10% of the remaining harvested area. Within each GDD zone, we divide the area into 10 equal-harvested-area precipitation bins. Continuing this method in each GDD zone results in 100 crop-specific equal-area bins with similar GDD and annual precipitation characteristics (Fig. S8).

Within each of the 100 bins, we analyze the yield distributions to determine AYs. We first temporarily discard the smallest-area grid cells (for a total of 5% of the bin area), in order to remove potential outliers from the yield dataset. An "attainable yield" is then defined as the area-weighted 95th percentile observed yield within a climate bin. Intensification potential (Fig. 1, Fig. 2) and average yields (online supplement) are defined as closing yield gaps in the worst performing regions up to different levels (50%, 75%, 90%, and 100%) of AYs.

Analyzing drivers of yield gaps

To analyze drivers of yield gaps, we used a nonlinear least-squares algorithm to fit input-response models to yield distributions within each climate bin for every crop.

Explanatory variables

Nitrogen, phosphate, and potash fertilizer application rates used as inputs for the models are directly from the fertilizer dataset developed in this study.

We calculate the maximum proportion of crop growing area irrigated in each grid cell in order to establish the spatial extent of irrigation technology and infrastructure for each crop (this variable is listed as *IRR* in the equations below and mapped for major cereals in Fig. 3b). We utilize the MIRCA2000 dataset[12] for this calculation, which contains monthly rainfed and irrigated areas for our crops of interest. We restrict our search for maximum irrigated proportion to months for which the reported crop growing area is at least 75% of the maximum in order to exclude anomalous growing conditions (e.g. a small area of a particular crop may be cultivated beyond of the normal growing season and be 100% irrigated, but this would not reflect the extent of irrigation capacity within the main growing season when only 50% of the area is irrigated).

Model functional form

While debate exists about the ideal functional form for input-yield models, there is general agreement in limited substitutability between inputs and a yield plateau at high inputs. To calculate our empirically derived crop yield models, we use a nonlinear least-

squares algorithm (trust region reflective) to fit input-response models to yield distributions within each climate bin. We utilize a standard functional form (Mitscherlich-Baule, Eqn. S2) for yield response to nutrient (nitrogen, phosphate, and potash) application rates[13,14]. We follow the von Liebig "law of the minimum"[13] to assess the combined effects of inputs. Thus, grid cell crop-specific yield (Y_{modGC}) is modeled as in Eqn. S2 for 100% rainfed grid cells, where Y_{max} is the maximum yield possible within the climate bin; b_{NP} and b_K describe the y-intercepts for each nutrient-yield response curve (the potash y-intercept is unique from the other nutrients as described in the *parameters* section below); c_N, c_P, and c_K are response coefficients that describe the percent of Y_{max} achieved at a given nutrient level; N_{GC}, P_{GC}, and K_{GC} are kg/ha of N, P$_2$O$_5$, and K$_2$O fertilizer applied to the grid cell.

$$Y_{modGC} = \min \left(\begin{array}{c} Y_{max}\left(1 - b_{NP}e^{(-c_N N_{GC})}\right), Y_{max}\left(1 - b_{NP}e^{(-c_P P_{GC})}\right), \\ Y_{max}\left(1 - b_K e^{(-c_K K_{GC})}\right) \end{array} \right) \quad \text{(Eqn. S2)}$$

When grid cells contain a mixture of rainfed and irrigated areas, rainfed yields may be limited by nutrient application (as in Eqn. S2) or a climate-specific rainfed yield maximum (Y_{maxRF}). When grid cell nutrient application rates exceed those required to achieve the rainfed yield maximum, we assume nutrients in excess of those required to achieve the rainfed yield maximum are applied preferentially to irrigated areas. For example, nitrogen requirements for the rainfed yield maximum (N_{reqRF}) are calculated as in Eqn. S3 and nitrogen application rates for irrigated lands are calculated as in Eqn. S4. Similar calculations are done for phosphate and potash.

$$N_{reqRF} = - \ln \left(\frac{1 - (Y_{maxRF}/Y_{max})}{b_{NP}} \right) / c_N \quad \text{(Eqn. S3)}$$

$$N_{IRR} = \frac{N_{GC} - \left(N_{reqRF}(1 - IRR) \right)}{IRR} \quad \text{(Eqn. S4)}$$

Irrigated modeled yield is determined using the nutrient response curves and the nutrient application rates for irrigated area (Eqn. S5). Grid-cell modeled yield is then a simple weighted average of the maximum rainfed yield and the modeled yield on irrigated land (Eqn. S6).

$$Y_{modIRR} = \min \left(\begin{array}{c} Y_{max}\left(1 - b_{NP}e^{(-c_N N_{IRR})}\right), Y_{max}\left(1 - b_{NP}e^{(-c_P P_{IRR})}\right), \\ Y_{max}\left(1 - b_K e^{(-c_K K_{IRR})}\right) \end{array} \right) \quad \text{(Eqn. S5)}$$

$$Y_{modGC} = \left((1 - IRR)Y_{maxRF} \right) + \left(IRR\, Y_{modIRR} \right) \quad \text{(Eqn. S6)}$$

Parameters

We define Y_{max} as the 98th percentile yield within a climate bin and b_{NP} (which defines the y-intercept for the nitrogen and phosphorus curves) is calculated using the 2nd percentile yield within a climate bin. We do not use the maximum and minimum within each bin due to observed outliers within the dataset. The decision to define the y-intercept and yield maxima using climate bin yield information keeps the models grounded in the empirical data, reduces the dimensionality of the nonlinear fitting routine, and provides an *a priori* default model when the input-yield relationships lack explanatory power within some climate bins for some crops.

The parameters b_K, Y_{maxRF}, c_N, c_P, and c_K are defined by the nonlinear regression. We allow b_K to float in order to reflect the higher soil availability of potash relative to nitrogen and phosphorus. The parameter describing the rainfed maximum yield (Y_{maxRF}) is not fixed, since this value is climate-specific. Nitrogen, phosphate, and potash response coefficients (c_N, c_P, and c_K) are constrained to be within 5x of the global response coefficient for a particular crop.

When inclusion of a particular input in the combined equation doesn't contribute to minimizing error in the residuals, that input is removed as an explanatory variable and we do not calculate a bin-specific response to that input. For example, although there may be a small amount of irrigation within a high-precipitation bin, irrigation may not influence yields within this climate and thus we throw out Y_{maxRF} as an explanatory variable (in this case yields are modeled solely as a function of nutrients as in Eqn. S2).

Model evaluation

Sensitivity to soil quality and slope

Soil quality and slope can impact yields at the field scale[15]. While currently available global soils data has many limitations[16,17], we analyzed whether our yield models and yield gap estimates were sensitive to the influence of soil organic carbon (SOC) from the ISRIC-WISE database[16] (aggregated among within-grid-cell soil types according to method C in Batjes *et al.*[16]) and the workability soil quality indicator (as a proxy for soil texture) from the FAO-IIASA Harmonized World Soil Database[18]. Workability categories are defined as 1) no or slight constraints, 2) moderate constraints, 3) severe constraints, 4) very severe constraints, 5) mainly non-soil, and 6) permafrost area. We also examined sensitivity to slope constraints as quantified by FAO-IIASA[19]. Slope categories are defined as 1) no constraints, 2) very few constraints, 3) few constraints, 4) partly with constraints, 5) frequent severe constraints, and 6) very frequent severe constraints, and 7) unsuitable for agriculture.

We first examined raw correlation between climate-normalized global yield gaps and the additional variables (e.g. Fig. S2abc). SOC displayed surprisingly little correlation to percent of attainable yield achieved. R-squared statistics for the correlations were ≤

0.01 for all crops except for the weak positive correlations of oil palm, maize, and soybean (r^2 = 0.04, 0.02, and 0.03, respectively). Workability correlations were generally weak and directionally inconsistent, with the strongest relationships for soybean (a negative relationship with r^2 = 0.08) and sugarbeet (a positive relationship with r^2 = 0.06). Slope correlations were generally stronger than the other variables examined, although they remained directionally inconsistent between crops. Maize and soybean had the strongest negative relationships (r^2 = 0.13 and 0.09, respectively) between percent of attainable yield achieved and slope, while oil palm had the strongest positive association (r^2 = 0.07).

The yield gap correlations above may be misleading when SOC, workability, or slope are also correlated with management practices. Thus, we constructed added variable plots (e.g. Fig. S2def) to assess the unique explanatory power of the additional variables when controlling for both climate and management. Each additional variable was regressed onto the modeled yields, and the residuals from these regressions were used as explanatory variables to explain the residuals from the yield model described by Eqns. S2-S6. The added variable plots show little unique yield variability explained by SOC, workability, and slope. No r-squared statistics > 0.01 were observed for the SOC added variable plots. All added variable plots for workability showed r-squared statistics ≤ 0.02, with the exception of the negative relationship for sunflower (r^2 = 0.03). All added variable plots for slope showed r-squared statistics ≤ 0.02, with the exception of the negative relationships for barley, sunflower, maize, and soybean (r^2 = 0.03, 0.03, 0.05, and 0.04, respectively). Added variable plot results suggest that slope may be the most important of the additional variables considered, and we highlight slope as a potentially important topic for future work. However, given the inconsistency in the impact (and direction) of slope on yield for different crops, as well as its relatively weak explanatory power when accounting for management practices, we chose not to incorporate slope into our analysis.

Given the agronomic knowledge about the importance of SOC, workability, and slope at the field scale, it is surprising we did not see greater sensitivity to these variables. One possible reason for lack of sensitivity is the quality of the global soils data, which is known to need improvement[16,17]. Another possible explanation is the landscape scale of this analysis. Variability of growing conditions can impact yields within a single field, and the aggregation that occurs with landscape-level yield and soil/slope statistics may remove much of the yield signal. Moreover, farm management practices also heavily influence soil characteristics, and this variability is not captured in the global soils data. An updated soils dataset utilizing the state-of-the-art data[17] could enable better quantification of the role of soils in determining both observed and attainable yields.

Cross-validation

Model prediction error was assessed using 5-fold cross-validation. For maize, wheat, and rice, we divided unique census-unit yield observations within each climate bin into

five 20% (by area) samples. For each sample, yields were predicted using model coefficients calibrated using the four other samples. This technique allows us to test how well the model performs on independent validation samples.

Cross-validation modeled yields (predicted yields for all five samples using the cross-validation technique) were compared with full-sample modeled yields (model output generated when all the data is used to calibrate the coefficients). Across all three crops, cross-validation modeled yields had ~0.1 t/ha higher root mean squared error (RMSE) than the full-sample modeled yields (full-sample modeled yield RMSE was 1.38, 0.83, 0.94 t/ha for maize, wheat, and rice, respectively, compared to cross-validation modeled yield RMSE of 1.47, 0.91, and 1.05 t/ha). Cross-validation modeled yield r-squared statistics were 0.76, 0.68, and 0.68 for maize, wheat, and rice, respectively, compared to full-sample modeled yield r-squared statistics of 0.79, 0.74, and 0.74. These results suggest that the modeling approach is robust and is not overly dependent upon the data utilized to calibrate the model.

Analyzing yield-limiting factors and input tradeoffs

Utilizing the crop- and climate-specific input-yield models, we are able to assess input requirements for achieving a given yield. We apply these tools to assess input requirements for current yields (to determine possible decreases in input use) and input requirements for various scenarios of closing yield gaps (to determine possible increases in input use).

Assessing possible input reductions

Possible input reductions are calculated by estimating necessary input application given yield limitation by other inputs. First, yields are modeled using the suite of input-yield models for maize, wheat, and rice. For each crop, we then assessed the nitrogen, phosphate, potash, and irrigation levels necessary to achieve current modeled yields. On each grid cell, one of the aforementioned inputs will be limiting, and the others will be more or less in balance with that limiting nutrient. We calculate "required" nutrients as the amount of other inputs needed when all inputs are in balance. This approach explicitly examines nutrient imbalances, but implicitly also examines inefficiencies in use of particular inputs. For example, we may quantify possible nitrogen reductions in a given area; this overuse of nitrogen may be a function of imbalanced nutrient supply and/or a function of widespread inefficiencies in nitrogen application and uptake.

Input increases to close yield gaps

To assess input increases to close yield gaps, we first identify grid cells where we quantify yield gaps at a given level (i.e. 50% of attainable yields or 75% of attainable yields) using our empirical crop yield data. In climate bins where we do not have an

irrigation response parameterized (i.e. including a value for Y_{maxRF} did not aid in minimizing regression residuals), we can simply assess input requirements to close yield gaps using Eqn. S3. In climate bins with an irrigation response, closing yield gaps may be achieved through a combination of nutrient and irrigation changes to the cropland within these grid cells (e.g. Fig. S5). We explicitly explore the complexities of this multidimensional yield response surface by considering only changes to nutrient application (1), only changes to irrigated area (2), or joint changes to nutrient application and irrigated area (3).

1. To assess whether yield gaps could be closed with nutrient-only intervention, we first fixed irrigation levels to those observed in the MIRCA2000 dataset. We then calculated areas where yield gaps could be closed with increasing nutrients utilizing the models of yield response to nutrients given a certain level of irrigation (Eqns. S3-S6). Given the asymptotic nature of the yield response to nutrients, some grid cells are modeled as able to close the yield gap but utilize unrealistic nutrient application rates. To account for this issue, we calculate the 95th percentile of globally observed N, P_2O_5, and K_2O application rates for the crop of interest. We only categorize grid cells as able to close yield gaps with nutrients only when projected nutrient requirements are within these 95th percentile application rate limits.

2. To assess whether solely increasing irrigated area could close yield gaps, we again utilize our models of yield response to nutrients and irrigation (Eqns. S3-S6). Fixing nutrient application, we solve for the irrigation levels needed to close yield gaps. Grid cells are classified as able to close yield gaps with irrigation only when irrigated proportion needed to achieve the desired yield level is ≤1.

3. To project input changes under joint irrigation and nutrient intervention, we calculate the nitrogen by irrigated area (NxI) response surface for each climate bin as in Fig. S5. Given limitation by other inputs, we first determine the "effective" nitrogen fertilization rate. Using the effective nitrogen fertilization rate and the current irrigated area proportion, we determine the current placement of the grid cell on the NxI yield response surface. We calculate the contour on the NxI surface corresponding to our desired yield level (50% or 75% of attainable yields) and normalize the nitrogen axis to a 0-1 scale using the 95th percentile of crop-specific nitrogen application. We then determine the minimum-distance change in nitrogen and irrigation to meet this contour. Phosphate and potash requirements for the desired yield are then calculated with the new irrigated area proportion.

For the above analyses, a special case arose when we lacked a climate bin-specific response for a particular nutrient input. Such a situation arises when the inclusion of that nutrient in the combined model did not contribute to minimizing error in the residuals, and thus was dropped as an explanatory variable. In these cases, we estimate the input-yield curve using the bin-specific yield maximum (Y_{max}) and the average response coefficient of interest (c_N, c_P, or c_K) across all the bins in which we had

parameterized that response coefficient. Likewise, the y-intercept for the potash curve (defined by b_K as the proportion of Y_{max} achievable with no additional potash inputs) is estimated using the average b_K across all bins where the response is parameterized.

Yield-limiting factors

We categorically assess yield-limiting factors (Figs. 3, S3-S4) by comparing current input use against projected required inputs needed to close yield gaps (when both nutrients and irrigated area are allowed to change). Grid cells are categorized having achieved the target yield when either the observed yields [3] or the modeled yields exceed the target yield.

SUPPLEMENTARY DISCUSSION

Comparison to previous yield gap analyses

Climate bins used in this analysis are defined by empirical growing degree-day and annual precipitation data. Previous yield gap analyses used a modeled water stress or aridity index (actual / potential evapotranspiration) as a moisture variable[9,10]. To remove model dependence, we examined several crop-relevant empirical moisture variables in preliminary analyses: an empirical aridity index (precipitation / Thornthwaite potential evapotranspiration), annual precipitation, or precipitation during the growing season. The different moisture variables had little effect on the magnitude or spatial patterns of yield gaps. Given the lack of sensitivity, we chose to utilize annual precipitation in our analysis in order to utilize the simplest possible metric.

In addition, we utilized a climate contour technique to identify climate outliers, equal-area binning, and a more restrictive definition of similar climates than climate bins used in previous analyses[9,10]. These methodological details prevent us from attempting to calculate AYs in anomalous climates and ensure an adequate sample size of grid cells within each climate bin. The more restrictive definition of similar climate generally results in slightly more conservative estimates of potential production, but allows us to capture a greater amount of yield variation due to climate for many crops (Table S3).

Limitations of the analyses

The fertilizer dataset, yield gap estimates, and yield models presented here are not without limitations, which we discuss below.

Our crop-specific, sub-national fertilizer application rates provide a more detailed and comprehensive picture of global nutrient use than previously assembled data[1,20,21], yet we still encountered considerable data limitations in many lower- and middle-income

countries. Furthermore, data from multiple sources did not always reconcile and we were forced to make a number of judgments on how best to integrate the data sources. In some case, our ability to reconcile the different datasets differed by nutrient, with more anomalies arising for phosphate and potash than nitrogen. For example, in the US, our initial estimates of nitrogen consumption (using the crop-specific application rates) aligned quite well with FAO nitrogen consumption, and we only needed to scale application rates for "untrusted" crops. For phosphate and potash in the US, our consumption estimates from the crop-specific data did not closely align with the FAO consumption data and we were forced to scale all the crop application rates downward a considerable amount. Improved underlying data is needed to reconcile these disagreements.

Attainable yield and yield gap estimates are quantified using census-derived observed yield data from Monfreda *et al*.[3], and AYs are identified from locations elsewhere on the globe with similar annual climate characteristics. As such, AYs are conservative estimates of landscape-scale achievable yields circa the year 2000 and not estimates of physiologically achievable potential yields (for a thorough discussion of yield potential quantification, see Lobell *et al*.)[22]. However, as noted in the main text, our AY calculations are likely more realistic than biophysical potential yields for defining intensification potential on regional and global scales. Yet a drawback of this technique is that we may underestimate AYs if no high-performing regions fall within a climate analog. For example, in certain climates we may not observe any truly high-performing areas, and thus we may be underestimating yields attainable under more intensive management regimes (e.g. we observe relatively small yield gaps in the Sahel for sorghum and millet due to a lack of high-performing grid cells elsewhere within those climate bins, although we speculate that it is unlikely these areas have truly small yield gaps). Yield gap and attainable yield quantification could be improved in further studies by attempting to include the effects of intra- and inter-annual climate variability. We also note that observed yields can be affected by farmer choice in multi-cropping systems. Further efforts should be made to analyze the magnitude of this effect on AY and yield gap calculations.

The yield models developed in this study describe observed patterns of agricultural productivity[3] for most major crops (Table S1), but are dependent upon input data of varying quality and scale. The models usefully characterize global patterns of input-use efficiency, but obscure fine-scale details about input-yield relationships and impacts of precision techniques. Yield models generally perform well for input-dependent crops grown in areas where data is likely more reliable, but do not perform as well for some tropical crops due to factors including a lack of high-performing climate analogs, poor data quality, or missing information about important management practices for these crops. Despite these limitations, we believe conclusions from the models can diagnose broad-scale trends across landscapes and regions. Moreover, the models provide a useful framework for further large-scale, quantitative analysis of agricultural management and crop production.

We lacked data to assess management practices other than fertilizers and irrigation in the analysis, although many of these practices may positively co-vary with our explanatory variables. Among the important practices to consider as high-quality data becomes available are the following: crop-specific manure application rates (particularly important in tropical settings)[1,23], distribution of advanced seed, drainage and water management, plant population, prevalence of agro-ecological techniques for improving soil health and nutrient recycling, common crop rotations and impacts on growing season length, advanced precision management techniques, and crop protection through chemical or agro-ecological means. Estimates of yield-limiting factors, nutrient imbalances, and inputs to close yield gaps could change as additional management practices, particularly organic nutrient inputs, are considered.

This analysis uses a cross-sectional approach with spatial data circa the year 2000 to assess opportunities for intensification, but detailed analysis of temporal yield, attainable yield, and harvest efficiency data are also needed to understand intensification pathways. Beyond yields, the harvest efficiency of cropland has changed over time [4]; additional work must assess the opportunities and environmental tradeoffs for increasing cropping intensity and decreasing pre- and post-harvest crop losses. [24-43]

SUPPLEMENTARY FIGURES AND LEGENDS

a

global nitrogen consumption (kg N / grid cell ha)

0   50   100   150

b

global phosphate consumption (kg P_2O_5 / grid cell ha)

0   50   100   150

c

global potash consumption (kg K_2O / grid cell ha)
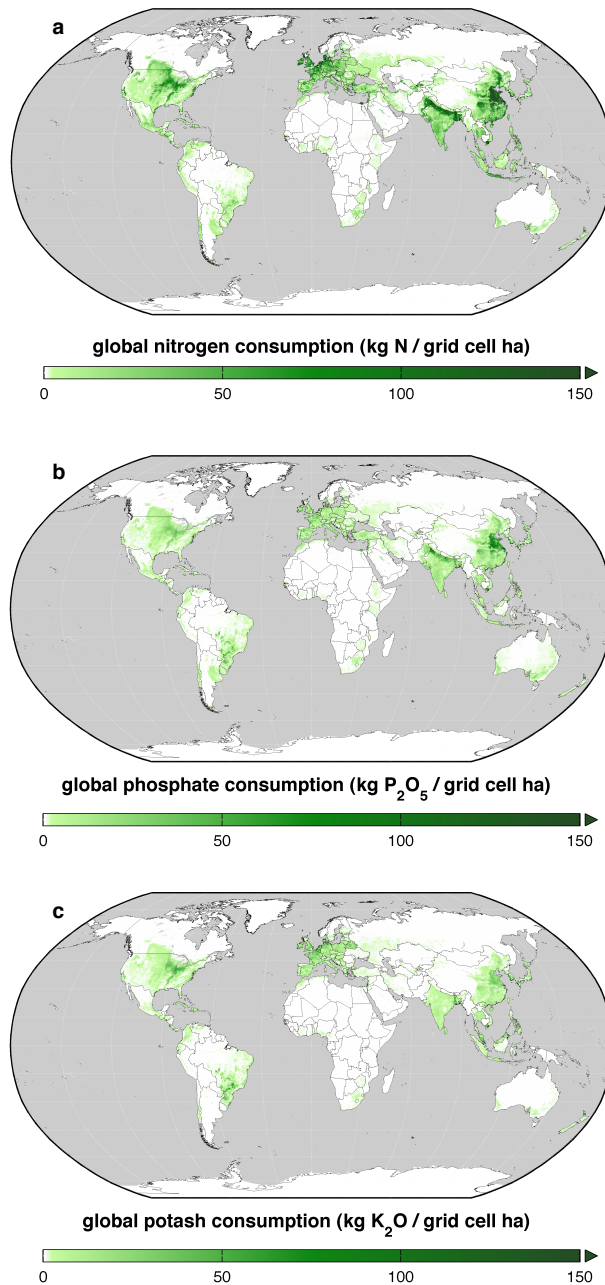
0   50   100   150

Figure S1: Global (a) nitrogen, (b) phosphate, and (c) potash consumption from fertilizer application as mapped using our crop-specific and, where available, sub-national dataset.
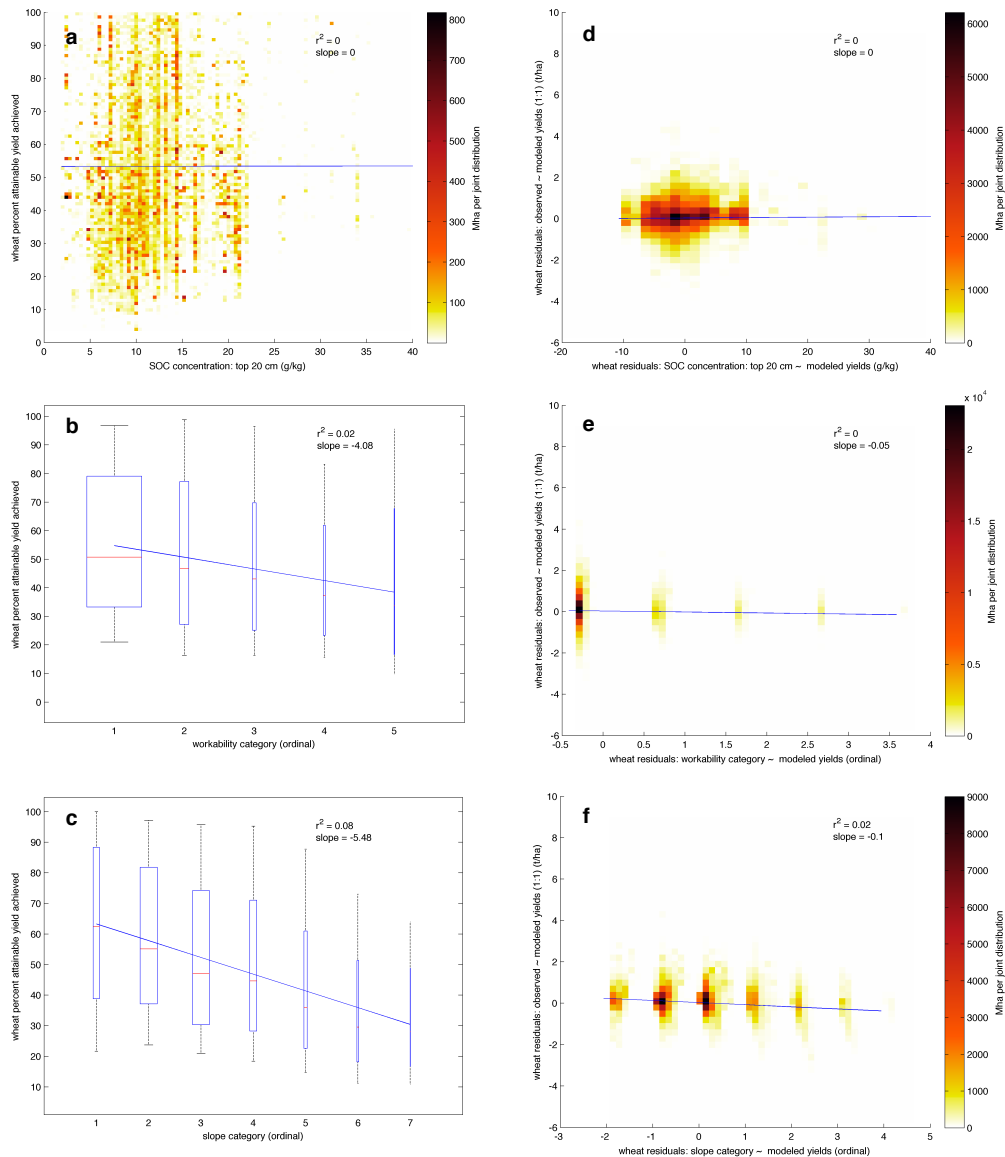
Figure S2. Wheat sensitivity analysis for (a,d) soil organic carbon, (b,e) soil workability category, and (c,f) slope category. Plots (a,b,c) show the raw correlation between yield gap (defined as the percent of attainable yield achieved) and the additional variables, while the added variable plots (d,e,f) show variation explained by these variables when controlling for management and climate. For (b,c), width of the boxplot is proportional to the area contained within each category, illustrating that most wheat area is grown in areas without substantial workability or slope constraints.
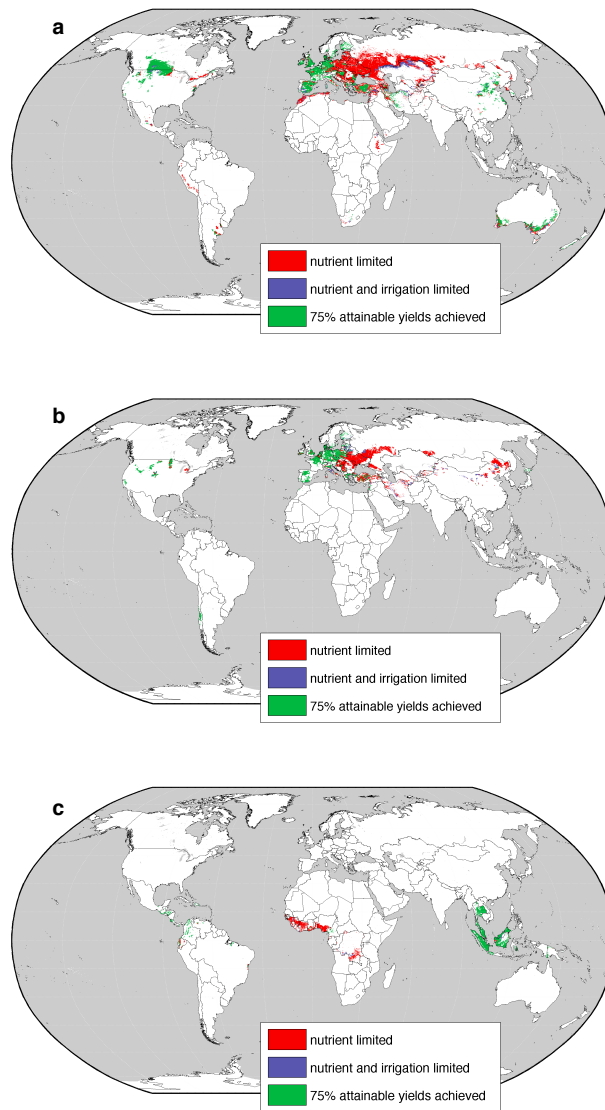
Figure S3: Management factors limiting yield gap closure to 75% of attainable yields for (a) barley, (b) sugar beet, and (c) oil palm. Median management models for these crops had the largest explanatory power (Table S1) across our 17 crops.
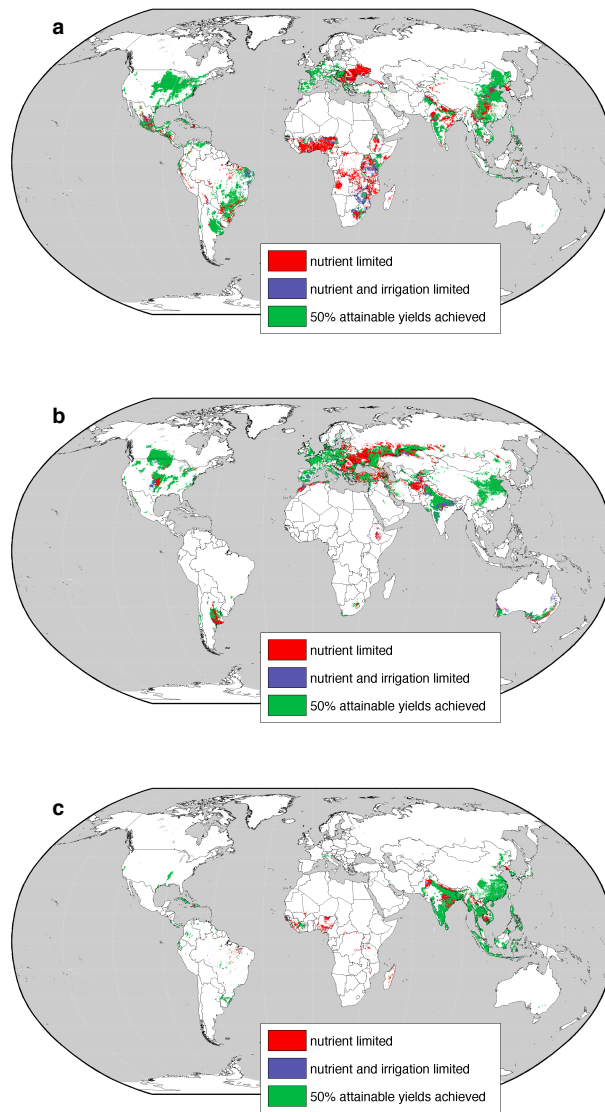
Figure S4: Management factors limiting yield gap closure to 50% of attainable yields for (a) maize, (b) wheat, and (c) rice.
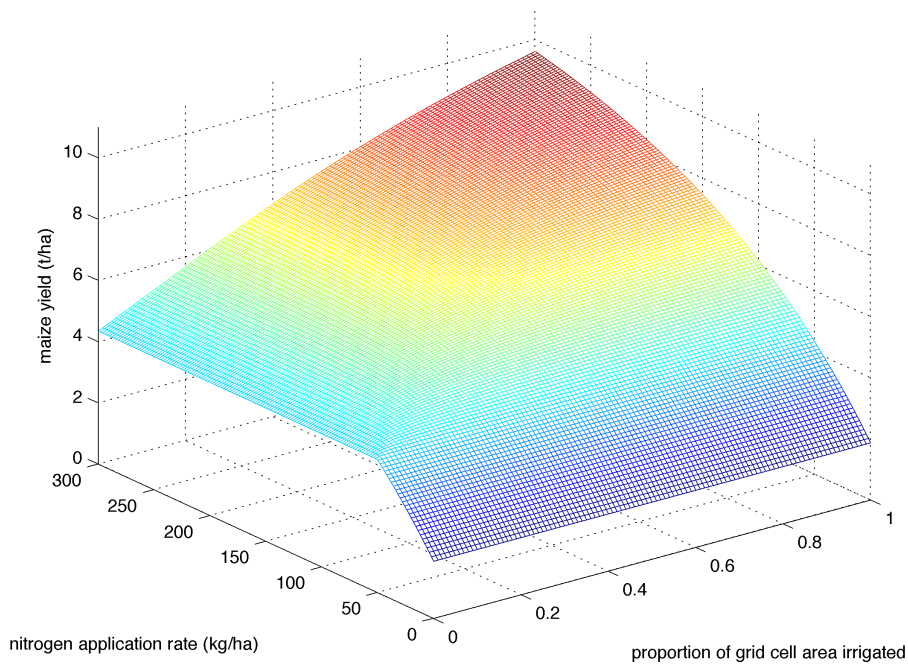
Figure S5: Example modeled yield surface response to increasing irrigated area proportion in a grid cell and nitrogen fertilizer application rate (kg/ha). Response is shown for maize climate bin 32 (GDD base 8°C = 1974 to 2321, precipitation = 573 to 648 mm/yr).

a

major cereals: ΔN application rate (kg/ha) possible

-200 -180 -160 -140 -120 -100 -80 -60 -40 -20 0

b

major cereals: ΔP$_2$O$_5$ application rate (kg/ha) possible

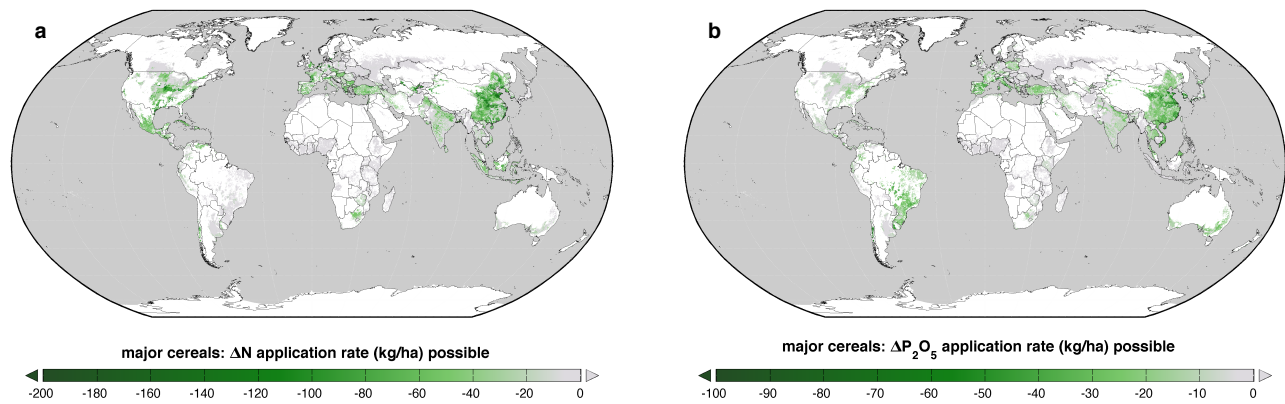-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0

Figure S6: Model results indicate decreases in (a) nitrogen and (b) phosphate application rates are possible without affecting yields for maize, wheat, and rice.

a

major cereals: ΔP_2O_5 application rate (kg/ha)

0   10   20   30   40   50   60

b

major cereals: ΔK_2O application rate (kg/ha)

0   10   20   30   40   50   60

c

major cereals: ΔP_2O_5 application rate (kg/ha)

-60   -40   -20   0   20   40   60

d

major cereals: ΔK_2O application rate (kg/ha)
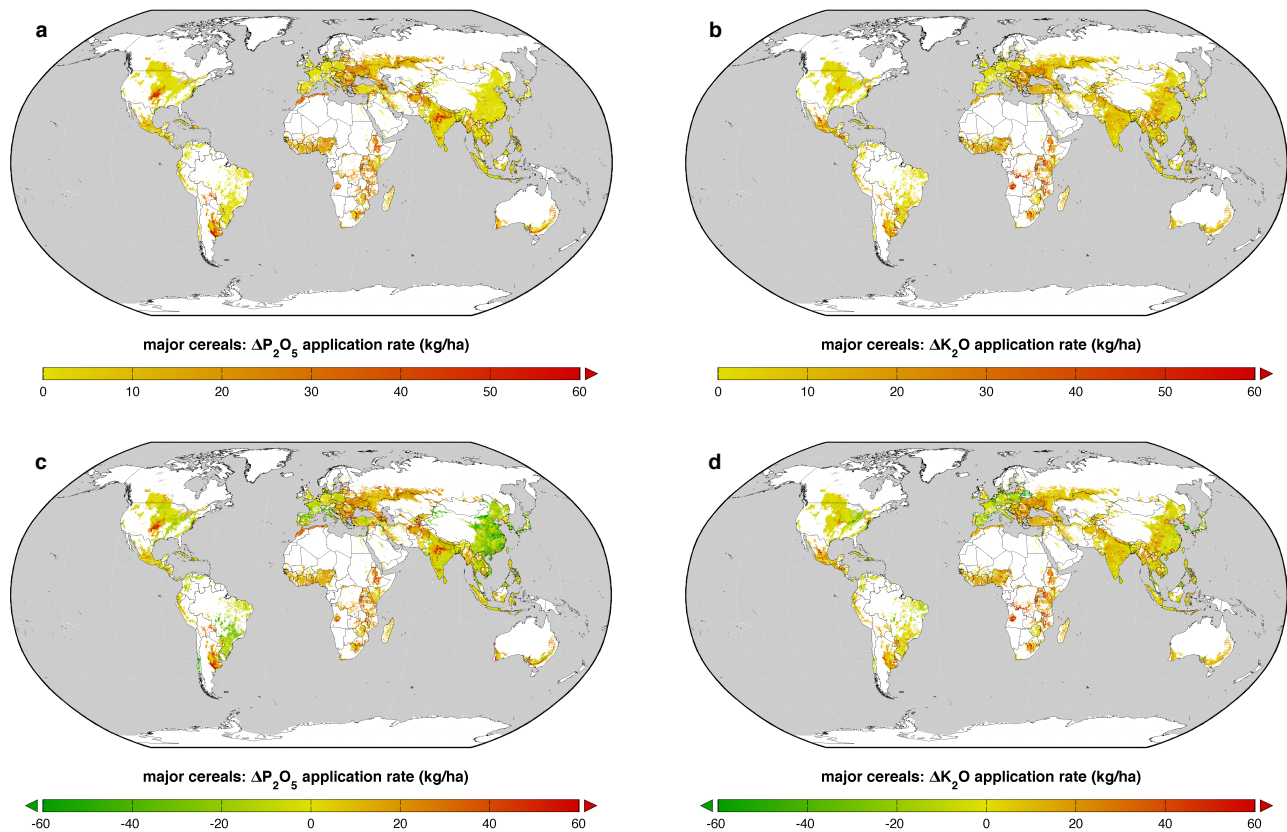
-60   -40   -20   0   20   40   60

Figure S7. Projected changes in (a) phosphate application rates and (b) potash application rates necessary to close maize, wheat, and rice yield gaps to 75% of attainable yields, and (c,d) projected net changes when eliminating input imbalances and inefficiencies.
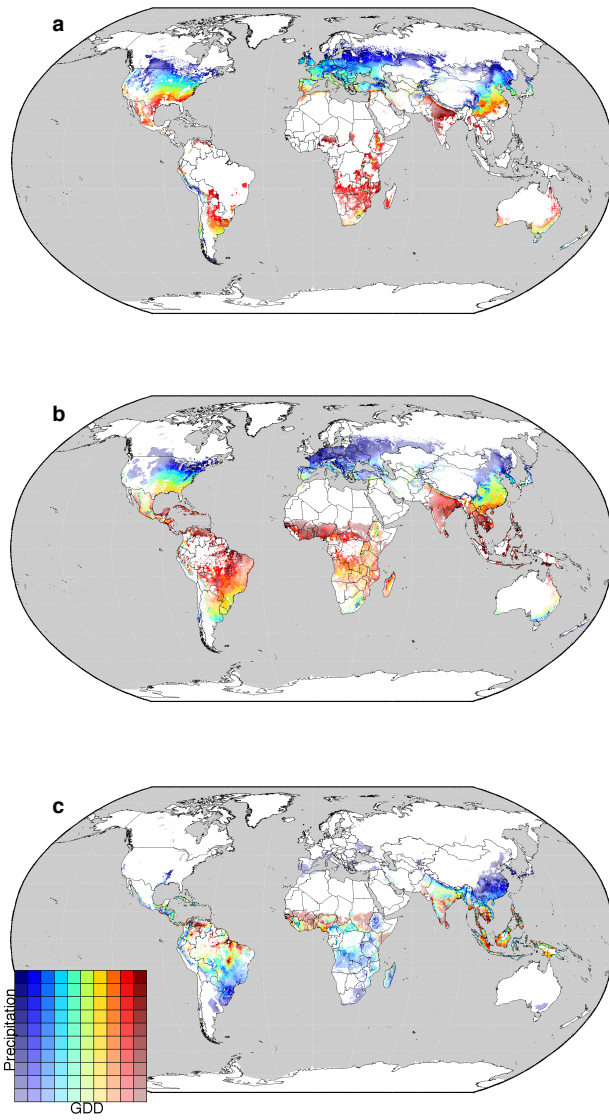
Figure S8: Crop-specific climate bin maps for (a) maize, (b) wheat, and (c) rice.

SUPPLEMENTARY TABLES

crop	median within-bin rmse (t/ha)	median within-bin r^2	global rmse (t/ha)	global r^2
wheat	0.80	0.51	0.83	0.74
rice	0.93	0.33	0.94	0.74
maize	1.27	0.47	1.38	0.79
soybean	0.39	0.29	0.38	0.69
barley	0.56	0.66	0.59	0.83
sorghum	0.49	0.08	0.56	0.81
millet	0.37	0.07	0.30	0.56
cotton	0.45	0.25	0.48	0.80
rapeseed	0.33	0.14	0.34	0.79
groundnut	0.43	0.16	0.45	0.82
sunflower	0.34	0.33	0.32	0.66
sugarcane	14.04	0.33	16.47	0.43
potato	4.88	0.49	5.57	0.67
cassava	4.53	0.21	3.58	0.39
oil palm	2.09	0.85	2.21	0.91
rye	0.80	0.22	0.90	0.57
sugar beet	7.43	0.68	8.05	0.81

Table S1. Root mean squared error (RMSE) and r-squared statistics for the median (defined by r-squared) within-climate bin management model and the combined suite of 100 climate bin-specific models. R-squared statistics for within-bin models measure within-bin yield variance explained by the management model, whereas the global r-squared statistics measure the global yield variance explained by the suite of models.

data source	spatial coverage	data type	years
FAO[4]	194 countries	national-level consumption	1997-2003
IFA/IFDC/IPI/PPI/FAO[2]	88 countries and Taiwan	national-level application rates by crop	1994-2001
Australian Bureau of Statistics[24, 25]	Australia	sub-national consumption	2000-2002
FAO[26]	Brazil	sub-national application rates by crop; average sub-national application rates	2002
FAO[27]	Bulgaria	national-level application rates by crop	1998
Statistics Canada[28]	Canada	sub-national consumption	2000
National Bureau of Statistics of China[29]	China	sub-national consumption	1997-2002
FAO[30]	Cuba	national-level application rates by crop	1997-2002
UNIFA[31]	France	average sub-national application rates	1997-2003
Statistisches Bundesamt, BMELV[32]	Germany	sub-national consumption	2005-2006
FAO[33]	India	sub-national application rates by crop	2003-2004
India Department of Agriculture and Cooperation, Agricultural Census Division[34]	India	sub-national consumption	2001-2002
FAO[35]	Iran	sub-national consumption	2005
ISTAT[36]	Italy	sub-national consumption	2002
FAO[37]	Korea, DPR	national-level application rates by crop	1998-2000
Statistics New Zealand[38]	New Zealand	sub-national consumption	2007
National Fertilizer Development Centre[39]	Pakistan	sub-national consumption	1997-2003
FAO[40]	Poland	sub-national application rates by crop; average sub-national application rates	2003
Ministerio de Agricultura, Pesca y Alimantacion[41]	Spain	sub-national consumption	2005
Turkish Statistical Institute[42]	Turkey	sub-national consumption	1999-2001
AAPFCO[5]	USA	sub-national consumption for select states	2001-2002
USDA ERS[43]	USA	sub-national application rates by crop for select states	1997-2003

Table S2. Fertilizer data sources, data type, and spatial and temporal coverage.

crop	Δ global potential production	r^2 - Licker et al. & Johnston et al. climate bins	r^2 - Mueller et al. climate bins
wheat	-8%	0.36	0.41
rice	-4%	0.58	0.59
maize	-3%	0.58	0.60
soybean	-2%	0.54	0.56
barley	-11%	0.37	0.49
sorghum	1%	0.67	0.67
millet	-5%	0.45	0.43
cotton	-6%	0.57	0.60
rapeseed	-5%	0.58	0.64
groundnut	2%	0.70	0.69
sunflower	-7%	0.39	0.41
sugarcane	-4%	0.20	0.20
potato	-15%	0.17	0.25
cassava	-1%	0.19	0.18
oil palm	0%	0.57	0.48
rye	-8%	0.24	0.33
sugar beet	-9%	0.26	0.42

Table S3. Due to methodological differences, estimates of potential production increases are slightly more conservative than those in previous analyses. Differences between the climate bins used in this analysis ("Mueller") and the climate bins used by Licker et al. [9] and Johnston et al. [10] ("Licker") are analyzed by comparing the differences in global potential production with AYs (using 95th percentile yields within a climate bin, [Mueller potential production - Licker potential production] / Licker potential production) and the amount of global yield variation explained by the climate bins (calculated using an r-squared where the "modeled" yields are the area-weighted average yields within a climate bin). These calculations only include grid cells shared between both analyses, meaning that the grid cells cannot be defined as climate outliers according to the Mueller climate contour method and they must be in a Licker climate bin containing at least five grid cells.

REFERENCES

1. Potter, P. A., Ramankutty, N., Bennett, E. M. & Donner, S. D. Characterizing the Spatial Patterns of Global Fertilizer Application and Manure Production. *Earth Interactions* **14**, 1–22 (2010).
2. IFA IFDC IPI PPI FAO *Fertilizer use by crop*. 1–67 (FAO: Rome, 2002).
3. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles* **22**, GB1022 (2008).
4. *FAOSTAT*. (Food and Agriculture Organization of the United Nations:).
5. Association of American Plant Food Control Officials *Commercial Fertilizers 2002*. (2002).
6. Tilman, G. D. Forecasting Agriculturally Driven Global Environmental Change. *Science* **292**, 281–284 (2001).
7. World Bank *Country and lending groups: World Bank list of economies (July 2010)*. (2010).
8. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles* **22**, GB1003 (2008).
9. Licker, R. *et al.* Mind the gap: how do climate and agricultural management explain the "yield gap" of croplands around the world? *Global Ecology and Biogeography* **19**, 769–782 (2010).
10. Johnston, M. *et al.* Closing the gap: global potential for increasing biofuel production through agricultural intensification. *Environmental Research Letters* **6**, 1–11 (2011).
11. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978 (2005).
12. Portmann, F. T., Siebert, S. & Doell, P. MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles* **24**, GB1011 (2010).
13. Paris, Q. The von Liebig hypothesis. *Am J Agr Econ* **74**, 1019–1028 (1992).
14. Frank, M., Beattie, B. & Embleton, M. A Comparison of Alternative Crop Response Models. *Am J Agr Econ* **72**, 597–603 (1990).
15. Lal, R. Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands. *Land Degrad Dev* **17**, 197–209 (2006).
16. Batjes, N. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (version 1.0). *Wageningen: ISRIC—World Soil Information* (2006).
17. Sanchez, P. A. *et al.* Digital Soil Map of the World. *Science* **325**, 680–681 (2009).
18. Nachtergaele, F. *et al. Harmonized World Soil Database*. 1–50 (FAO, IIASA, ISRIC, ISSCAS, JRC: 2012).
19. van Velthuizen, H. *et al. Mapping biophysical factors that influence agricultural*

production and rural vulnerability. 1–93 (FAO: Rome, 2007).

20. MacDonald, G. K., Bennett, E. M., Potter, P. A. & Ramankutty, N. Agronomic phosphorus imbalances across the world's croplands. **108**, 3086–3091 (2011).

21. Liu, J. *et al.* A high-resolution assessment on global nitrogen flows in cropland. *Proceedings of the National Academy of Sciences USA* **107**, 8035–8040 (2010).

22. Lobell, D. B., Cassman, K. G. & Field, C. B. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annual Review of Environment and Resources* **34**, 179–204 (2009).

23. Smil, V. Phosphorus in the environment: Natural flows and human interferences. *Annu Rev Energ Env* **25**, 53–88 (2000).

24. Australian Bureau of Statistics *Agricultural Commodities 2000-01*. 1–40 (Canberra, 2002).

25. Australian Bureau of Statistics *Agricultural Commodities 2001-02*. 1–48 (Canberra, 2003).

26. Food and Agriculture Organization of the United Nations *Fertilizer use by crop in Brazil*. 1–64 (Rome, 2004).

27. Food and Agriculture Organization of the United Nations *FertiStat*. **2011**,

28. Statistics Canada *Census of Agriculture: Table 8.1*. 1–26 (2001).

29. National Bureau of Statistics of China *China Statistical Yearbooks: 1997-2002*.

30. Food and Agriculture Organization of the United Nations *Fertilizer use by crop in Cuba*. 1–40 (Rome, 2003).

31. Union des industries de la fertilisation UNIFA *Evolution des doses moyennes de N, P2O5, K2O en kg/ha fertilisable depuis plus de 30 ans dans votre région*. (2009).

32. Statistisches Bundesamt Bundesministeriums für Ernährung Landwirtschaften und Verbraucherschutz Statistisches Jahrbuch über Ernährung, Landwirtschaft und Forsten. (2007).

33. Food and Agriculture Organization of the United Nations *Fertilizer use by crop in India*. 1–59 (Rome, 2005).

34. India Department of Agriculture and Cooperation Agricultural Census Division *Input Survey 2001-02*. (2002).

35. Food and Agriculture Organization of the United Nations *Fertilizer use by crop in the Islamic Republic of Iran*. 1–78 (Rome, 2005).

36. Istat *Annuario statistico italiano 2004*. (2004).

37. Food and Agriculture Organization of the United Nations *Fertilizer use by crop in the Democratic People's Republic of Korea*. 1–46 (Rome, 2003).

38. Statistics New Zealand *Agricultural census tables: Fertiliser and lime applied by region and type*. (2007).

39. National Fertilizer Development Centre *Agricultural statistics of Pakistan 2004-2005*. (Islamabad, 2005).

40. Food and Agriculture Organization of the United Nations *Fertilizer use by crop in Poland*. 1–49 (Rome, 2003).

41. Ministerio de Agricultura Pesca y Alimantacion *Estadistica mensual de consumo de fertilizantes en la agricultura*. (2006).

42. Turkish Statistical Institute, *Bazi Tarimsal Gostergeler: İllere göre gübre kullanımı*.

(2009).

43.   United States Department of Agriculture Economic Research Service *Fertilizer Use and Price*. (2010).