ARTICLES
https://doi.org/10.1038/s43016-021-00300-1
nature food

Check for updates

Optimization of China's maize and soy production can ensure feed sufficiency at lower nitrogen and carbon footprints

Zitong Liu[1,10], Hao Ying[1,10], Mingyou Chen[1,10], Jie Bai[1], Yanfang Xue[2], Yulong Yin[1], William D. Batchelor[3], Yi Yang[4], Zhaohai Bai[5], Mingxi Du[6], Yixin Guo[7,8], Qingsong Zhang[1], Zhenling Cui[1]✉, Fusuo Zhang[1] and Zhengxia Dou[9]

China purchases around 66% of the soy that is traded internationally. This strains the global food supply and contributes to greenhouse gas emissions. Here we show that optimizing the maize and soy production of China can improve its self-sufficiency and also alleviate adverse environmental effects. Using data from more than 1,800 counties in China, we estimate the area-weighted yield potential (Y_{pot}) and yield gaps, setting the attainable yield (Y_{att}) as the yield achieved by the top 10% of producers per county. We also map out county-by-county acreage allocation and calculate the attainable production capacity according to a set of sustainability criteria. Under optimized conditions, China would be able to produce all the maize and 45% of the soy needed by 2035—while reducing nitrogen fertilizer use by 26%, reactive nitrogen loss by 28% and greenhouse gas emissions by 19%—with the same acreage as 2017, our reference year.

Soy and maize are feed grains that figure prominently in the global food security and sustainability agenda. Soy and maize account for 51% of all cereal grains produced globally. Since 2000, their trade on the international market has increased by 217% in quantity (from 135 to 293 million tons) and 424% in value (from US$21 to 89 billion), exceeding all other cereal grains[1]. Associated with soy and maize production are enormous environmental impacts[1–5]. For example, land expansion for soy production totalled 47.6 Mha since 2000, of which 44% occurred in the Brazilian Amazon and Cerrado regions, resulting in deforestation with substantial carbon losses and ecosystem damages[1,2,6,7]. As staple feed grains that are essential in modern livestock production, the heightened demands for soy and maize are closely linked with the growing appetite for animal-source foods (meats, milk, and eggs), particularly in developing economies[8–10].

China has an important role in this area. From 2000 to 2017, the per capita consumption of meat, milk and eggs increased by 75%, 150% and 38%, respectively[11]. Meanwhile, annual domestic consumption of maize and soy went up from 117.8 to 261.8 Mt and from 24.6 to 108.7 Mt, respectively[1]. The country managed to maintain its self-sufficiency for maize but became increasingly dependent on importation for soy. By 2017, China bought 66% of the soy traded on the international market to meet 90% of its domestic needs[1].

The consequences of the lopsided maize–soy production situation in China are multifaceted[2,3,6,7]. Domestically, China's maize production is widely known for high inputs (for example, fertilizers) and high levels of pollution (for example, nutrient losses, greenhouse gas (GHG) emissions)[3,12]. The diminishing share of soy production in the agrifood landscape has also been a topic of debate concerning national (food) security[13], in addition to issues pertaining to land use, biodiversity loss and implications for soil health[3]. Globally, China's increasing soy purchases have corresponded with hikes in the price of soy on the international market[14] (Supplementary Fig. 1). China could also be responsible for indirectly accelerating deforestation and GHG emissions in the Brazilian Amazon and Cerrado, where land has been cleared for expanded soy production[2,6,7]. Given the magnitude of the problem and its wide-ranging consequences—and anticipating even greater demands in the decades to come—we asked how China can optimize its resource allocation and enhance maize and soy production capacity so as to reduce its reliance on soy importation and advance both feed and food security sustainably.

In this study, we show that optimizing China's maize and soy production can help to address the issues raised above. Our conceptual framework is rooted in an analysis of both the current state and the potential of China's production systems: (i) Maize acreage is large (more than 30% of total cultivated land and the highest among all cereal grain crops in the country) but yield is low (roughly 50% of the yield potential)[15], indicating that there is a lot of room for improvement. (ii) Maize yield can increase substantially by implementing advanced management technologies that are already available[15–18]. Doing so would spare some maize acreage for soy while reducing pollution. (iii) Maize and soy have similar growth conditions—for example, planting regions and growing days—thus making the optimization strategies feasible. Through systems-based

[1]College of Resources and Environmental Sciences, National Academy of Agriculture Green Development, Key Laboratory of Plant-Soil Interactions, Ministry of Education, China Agricultural University, Beijing, China. [2]Maize Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China. [3]Biosystems Engineering Department, Auburn University, Auburn, AL, USA. [4]Key Laboratory of the Three Gorges Reservoir Region's Eco-Environment, Ministry of Education, Chongqing University, Chongqing, China. [5]Key Laboratory of Agricultural Water Resources, Center for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang, China. [6]Department of Geography, McGill University, Montreal, Quebec, Canada. [7]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [8]Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China. [9]Department of Clinical Studies - New Bolton Center, School of Veterinary Medicine, University of Pennsylvania, Kennett Square, PA, USA. [10]These authors contributed equally: Zitong Liu, Hao Ying, Mingyou Chen ✉e-mail: cuizl@cau.edu.cn

comprehensive analyses that combine field data with model simulation, together with a meta-analysis of the literature, we show that China would be able to produce 45% of the soy needed by 2035 while maintaining total maize self-sufficiency on the same acreage of our reference year (2017) for the two crops. Furthermore, implementing enhanced management practices proposed in the optimization strategies will have substantial co-benefits such as improving fertilizer efficiency, reducing reactive nitrogen losses and lowering GHG emissions.

Results

Yield benchmarks and attainable production capacity. Our first step was to quantify yield potential under irrigated (Y_p) and rain-fed conditions (Y_w) for maize and soy, which is a prerequisite for assessing maximal production capacity on existing cropland[19]. In China, maize and soy are grown in more than 1,800 counties that range from warm subtropics at 18° N to cool temperate climates at 53° N, with agroecosystems spanning arid to semi-arid to humid conditions. For each of those counties, we determined Y_p and Y_w using county-specific optimal agronomic parameters (for example, planting date, density and varieties; Supplementary Fig. 2 and Methods) as inputs to the Hybrid-Maize and SoySim models. To account for year-to-year variation, we used weather data for a 10-year period (2005–2014). The Y_p aggregated from all 1,839 counties averaged 15.6 Mg ha^{-1} for irrigated maize (range: 8.7–19.1 Mg ha^{-1}, median: 14.9 Mg ha^{-1}) and 4.3 Mg ha^{-1} for irrigated soy (range: 3.3–5.6 Mg ha^{-1}, median: 4.3 Mg ha^{-1}). In comparison, the Y_w was 11.7 Mg ha^{-1} for rain-fed maize (range: 5.0–16.6 Mg ha^{-1}, median: 11.6 Mg ha^{-1}) and 4.1 Mg ha^{-1} for rain-fed soy (range: 2.8–5.3 Mg ha^{-1}, median: 4.2 Mg ha^{-1}). Our simulated Y_p for maize is 9–10% greater than previous estimates[15,20]; the latter were based on less than optimum values for conditions such as planting date, density and varieties.

Closing the yield gap is key for pursuing sustainable food security[21,22]. To accurately assess county-level yield gaps for maize and soy, we first calculated the area-weighted yield potential (Y_{pot}) from Y_p (irrigated) and Y_w (rain-fed) at the county level, because our actual yield data do not distinguish irrigated from rain-fed production modes. Subsequently, county-level yield gap was computed as the difference between Y_{pot} and actual yield, the latter being the mean of farmers' yields recorded in Chinese yearbooks during 2005–2014. Together, the county-level yield gap varied from 1.35–10.93 Mg ha^{-1} for maize (5th–95th) and 0.84–3.56 Mg ha^{-1} for soy (Fig. 1c,d). Aggregated to the national level, the Y_{pot} is 13.1 Mg ha^{-1} for maize and 4.2 Mg ha^{-1} for soy (Fig. 1a,b); farmers' yields average 6.4 Mg ha^{-1} and 2.0 Mg ha^{-1}, respectively . Therefore, the national-level yield gap is 51% for maize and 52% for soy (Fig. 1c,d), suggesting that there is large room for improvement. For comparison, the yield gap was estimated at around 21% for maize in the USA[20] and around 36% for soy in Brazil[23].

Determining the attainable yield (Y_{att}) is critical in our analysis for optimizing China's maize and soy production. In previous studies, the attainable yield was assigned as 80% of the yield potential[21,22]. We examined the county-level yields that were actually achieved by high-producing farmers on the basis of a large-scale survey. The survey consisted of 4.85 million farmers in 1,243 counties, covering 78% of the total acreage of maize and soy production in China (Methods). For each county, we identified the top 10% producers and used the mean of their yields as Y_{att} (ref. [22]). County-by-county Y_{att} varied from 6.1 to 14.6 Mg ha^{-1} for maize ($n=997$, mean: 9.1 Mg ha^{-1}, median: 9.7 Mg ha^{-1}.) and 1.9 to 4.0 Mg ha^{-1} for soy ($n=246$, mean: 2.9 Mg ha^{-1}, median: 2.9 Mg ha^{-1}). The Y_{att} across all counties would be 40–98% of Y_{pot} for maize (mean: 68%, median: 65%) and 39–94% for soy (mean: 67%, median: 68%). Aggregated at the national level, Y_{att} would be 72% of Y_{pot} for maize and 71% for soy. We consider benchmarking county-specific Y_{att} on the basis of the mean actual yield of top 10% producers in each county to be a sound choice. It better reflects local socioeconomic and agronomic conditions, in contrast with the arbitrary assignment (80% of yield potential) that was used previously[21,22]. The Y_{att} values obtained in our analysis are probably conservative considering the highest records from field experiments in China (86–91% of Y_{pot} for maize; Supplementary Discussion)[16], but are realistic and feasible as the yields have been achieved by top producers in the counties.

Our next step was to map out acreage reallocation schemes for producing the maize and soy needed by 2035, projected to be 289 and 133 Mt, respectively (Methods). We assumed the same acreage as the reference year 2017 (50.6 Mha for the two crops[11]) and used the county-by-county Y_{att} as the production benchmark. We prioritized maize production by allocating enough acreage to produce the needed 289 Mt; the remaining acreage would then be used for soy. Without changing each county's total acreage for maize and soy, acreage reallocation followed three principles: (i) Cropping diversity. For each agroecological zone (see map in Supplementary Fig. 4a), we ranked all counties on the basis of their maize acreage as a percentage of total arable land in the county, then used the point of the 25th percentile as the cut-off for the maximal maize allowance. Counties with a maize acreage percentage greater than the cut-off point would have their 'excess' acreage reallocated to soy. For example, ranking in the northeast zone showed the 25th percentile to be 77%. Therefore, counties with a maize acreage greater than 77% of their total arable land would reduce their maize acreage to the cut-off and reallocate the remaining acreage to soy. (ii) Water availability. For counties with a persistent water deficit (see map in Supplementary Fig. 4b), a quarter of the current maize acreage would be reallocated to soy. This is because soy production consumes less water than maize production does on a per hectare basis (see Supplementary Discussion). (iii) Productivity constraints. Counties with a low Y_{att} (an indication of agro-biophysical and technological constraints) would convert 75% of the reference-year maize acreage to soy; this process started with counties with the lowest Y_{att} and moved up the ranking list until cumulative maize production output met the needed 289 Mt (see Methods).

The optimized maize and soy production scheme (Fig. 2) would have 30.7 Mha for maize (289 Mt) and the remaining 20.0 Mha for soy. This would mean a reallocation of 11.7 Mha from maize to soy compared to the reference year 2017. Production output for soy would total 60 Mt. China would be able to meet 100% of its needs for maize and 45% for soy by 2035. The required import of soy would be 73 Mt. In other words, under the optimized scheme, China would reduce its soy import reliance to 55%, compared with 90% in the reference year 2017. The relevant effects of this on international trade, in particular with regard to the two largest soy trade partners with China (Brazil and the USA), are presented in the Supplementary Discussion.

With the optimized scheme, yields at the national level for 2035 would average 9.4 Mg ha^{-1} for maize and 3.0 Mg ha^{-1} for soy (70% of the Y_{pot} for maize and 71% for soy). These would still be lower than the average yields achieved in the USA (11.9 and 3.3 Mg ha^{-1} in 2017) (FAO), and substantially lower than high-yield records in China (higher than 15 Mg ha^{-1} for maize)[16].

Nitrogen and carbon footprints and other effects. Advancing China's feed security with greater self-sufficiency must also address environmental and climate issues. We compared enhanced and conventional management scenarios to quantitatively assess a number of resource and environmental indices, including fertilizer uses, reactive nitrogen losses, GHG emissions and farmer cost–benefit analyses. The enhanced management scenario assumes the implementation of integrated soil-crop system management (ISSM), which is a comprehensive decision-support programme designed to provide agronomic and soil–water–nutrient management

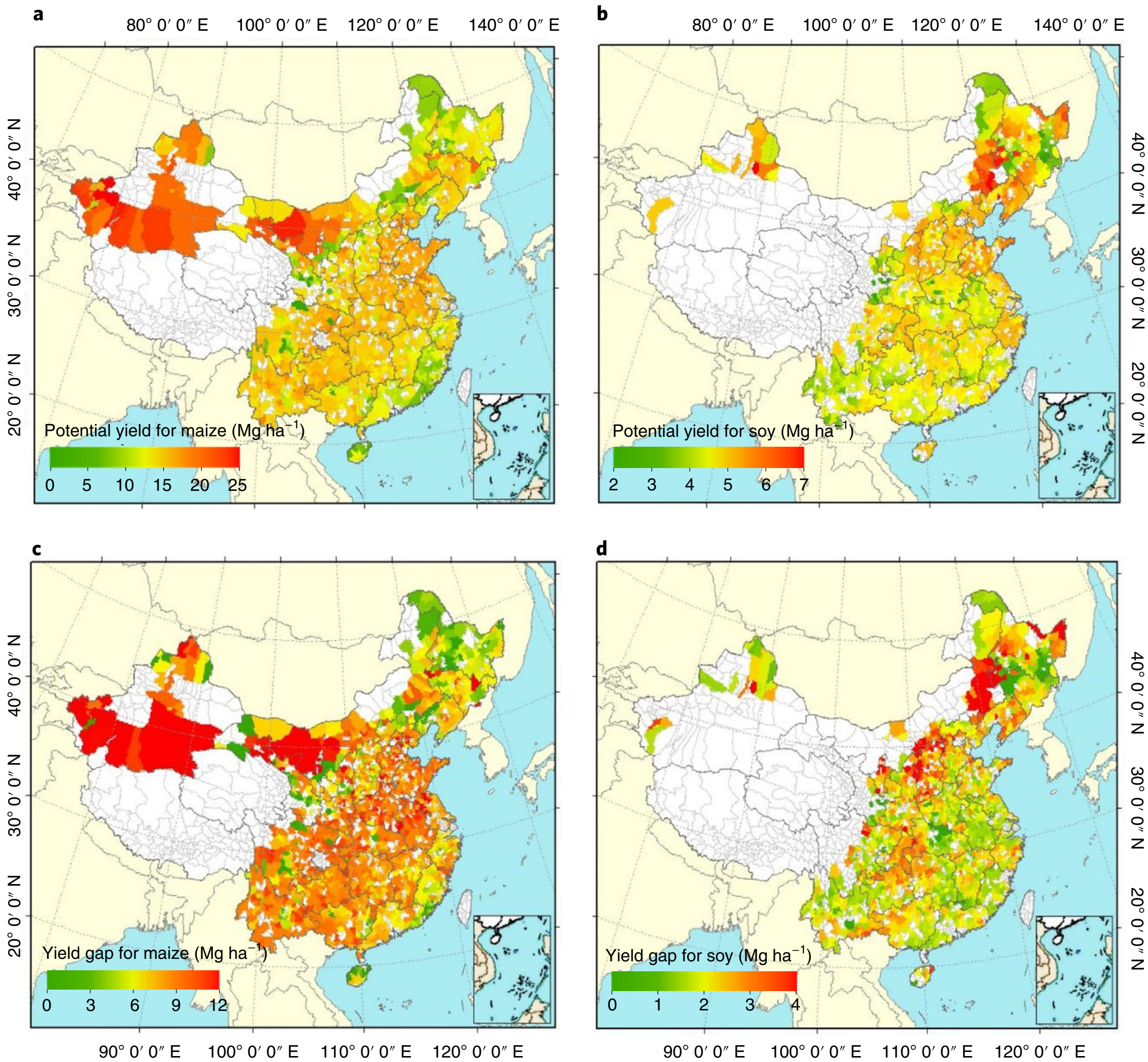

Fig. 1 | County yield potential and yield gaps for maize and soy in China. a,**b**, County-level area-weighted yield potential estimated from Y_p for irrigated and Y_w for rain-fed maize (**a**; $n=1{,}735$) and soy (**b**; $n=1{,}589$). **c**,**d**, County-level yield gap (that is, yield potential minus farmers' yield) for maize (**c**) and soy (**d**).

recommendations and has been widely tested in China[16–18]; the conventional scenario refers to the prevailing practices of farmers. First, we derived data on fertilizer and nutrient use efficiency for enhanced versus conventional scenarios from field trials of more than 8,000 site-years conducted during 2005–2014 through national collaboration networks[18] (Methods). Each trial featured side-by-side comparison of local farmers' practices (conventional) versus ISSM-based recommendations (the enhanced management scenario). Compared to local farmers' practices, the adoption of ISSM-based recommendations led to an increase in yield from 7.8 to 9.5 Mg ha^{-1} (12–40% for 6,089 site-years) for maize, and from 2.2 to 2.6 Mg ha^{-1} (6–40% for 2,072 site-years) for soy. Note that the yields obtained through ISSM are similar to the target yields (that is, Y_{att}) for assessing maize and soy production capacity by 2035 (9.4 and 3.0 Mg ha^{-1} for maize and soybean, respectively). Nitrogen use efficiencies (crop yields per unit of applied nitrogen fertilizer) increased from 40 to 51 kg kg^{-1} N for maize and from 51 to 62 kg kg^{-1} N for soy. Phosphorus use efficiencies went up as well, from 105 to 109 kg kg^{-1} P_2O_5 for maize and from 46 to 51 kg kg^{-1} P_2O_5 for soy. Under ISSM, potassium application rates increased, as recommended based on soil-test results, from 26 to 73 kg K_2O ha^{-1} for maize and from 15 to 20 kg K_2O ha^{-1} for soy (Supplementary Table 2).

From the large dataset established via those field trials, we calculated county-by-county fertilizer usage for the projected production of maize and soy by 2035 and their nutrient use efficiencies under enhanced (implementing ISSM-based recommendations) versus conventional scenarios. Aggregating county-by-county data to the national level, it would require 11.7 Mt N, 5.3 Mt P_2O_5 and 2.1 Mt K_2O for maize and soy production under the conventional scenario, but 6.7 Mt N, 3.8 Mt P_2O_5 and 2.9 Mt K_2O under the enhanced management scenario. In other words, adoption of ISSM could decrease total fertilizer use by 42% for N and 28% for P_2O_5, while increasing K_2O by 38% (Fig. 3a).

To assess reactive nitrogen losses under the enhanced versus the conventional scenarios, we created a random forest regression model that uses machine learning techniques to simulate reactive nitrogen loss through N_2O emissions, NH_3 volatilization and NO_3^- leaching (Supplementary Fig. 5). The model computes in situ reactive nitrogen losses by integrating eight parameters that are county specific, including precipitation, potential evapotranspiration, temperature and wind speed; soil texture, pH and organic matter; and nitrogen balance (see Methods). Loss factors are calculated as reactive nitrogen loss divided by the nitrogen applied. Our results indicated that loss factors would decrease from 23.4% under the conventional scenario to 21.9% under the enhanced scenario (Supplementary

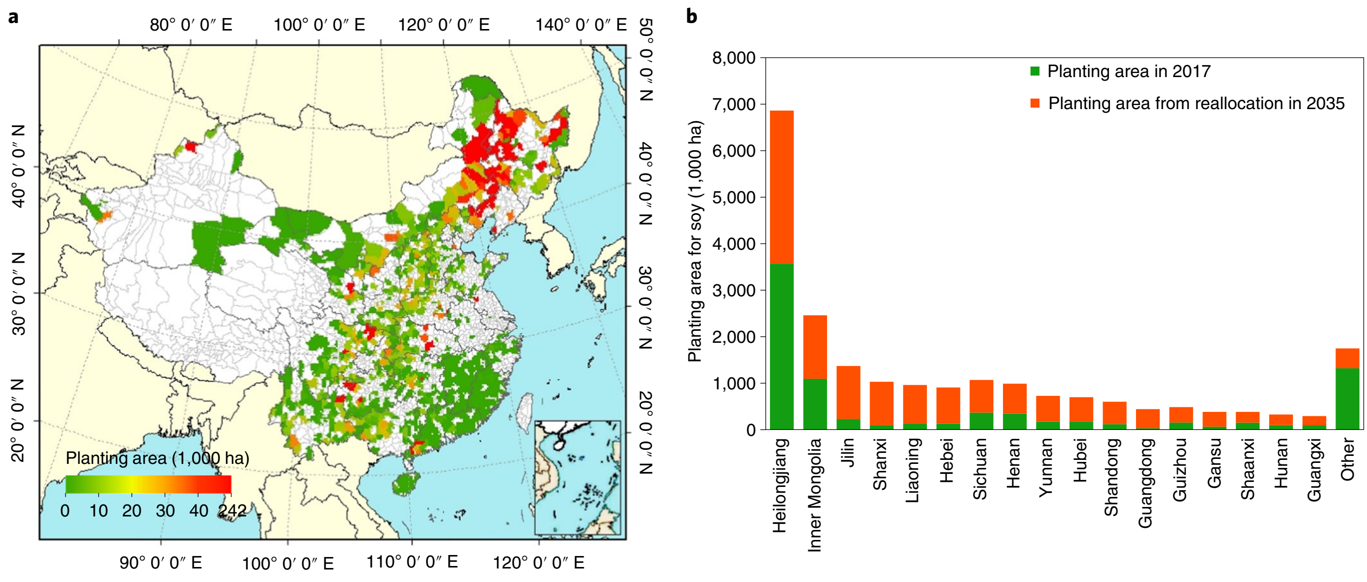

Fig. 2 | Acreage reallocation in the optimized production scheme. a, County level. **b**, Province level. Green pillars represent soy acreage in the reference year (2017); orange pillars represent acreage reallocated from maize to soy for the 2035 projection. Total acreage reallocation (11.7 Mha) considers crop diversity (28%), water availability (8%) and productivity constraints (64%).

Table 3 and Supplementary Discussion). Per hectare reactive nitrogen loss would be fertilizer application rate multiplied by the loss factor. Aggregated to the national level, reactive nitrogen loss would be 37.9 kg ha^{-1} (ranging from 24.3 to 56.1 kg ha^{-1} for 1,735 counties) for maize and 9.2 kg ha^{-1} (ranging from 5.1 to 17.6 kg ha^{-1} for 1,763 counties) for soy production under the enhanced management scenario, compared to 71.3 kg ha^{-1} (maize) and 13.6 kg ha^{-1} (soy) under the conventional scenario. The total reactive nitrogen loss would be 45% lower under the enhanced management scenario (1.3 Mt) compared to the conventional scenario (2.5 Mt) (Fig. 3b and Supplementary Fig. 6).

We also performed a life-cycle analysis to compute the carbon footprints (as carbon dioxide equivalent; CO_2-eq) for the two management scenarios, accounting for three contributing factors: field N_2O emissions; GHG emissions associated with the production and usage of fertilizers; and GHG emissions related to crop production operations (see Methods). All calculations were made county-by-county, and the results were expressed in CO_2-eq ha^{-1}. Aggregated to the national level, carbon footprints would be 3,074 and 1,192 CO_2-eq ha^{-1} for maize and soy under the enhanced management scenario, compared to 5,175 and 1,551 CO_2-eq ha^{-1} under the conventional scenario. Total carbon footprints for producing maize and soy projected for 2035 would be 38% lower under the enhanced scenario (118 Mt CO_2-eq), compared to the conventional scenario (190 Mt CO_2-eq) (Fig. 3c).

Our cost–benefit analysis (Methods) indicated that farmers who implemented the enhanced management scenario would reduce cost by US$52 ha^{-1} compared to the conventional scenario (Fig. 3d and Supplementary Table 4). This would add up to US$2.7 billion nationwide (2035 projection). It is also interesting to note that by optimizing China's maize and soy production, total fertilizer use for 2035 would be similar to the amounts consumed in the reference year 2017 (more phosphorus and potassium but less nitrogen), whereas the total reactive nitrogen loss would decrease by 27% and GHG emissions by 19% from the reference levels (Supplementary Table 3). In addition, there can be a number of tangible benefits besides the reductions in resource and environmental-climate burdens described above. For example, residual soil nitrogen from the soy crop residues can lower fertilizer nitrogen needs for the subsequent maize crop in a maize–soy rotation, thereby reducing nitrogen-induced pollution and GHG emissions[22]. Also, an improved maize–soy rotation or intercropping could help increase soil organic carbon, improve soil water retention and promote soil health by inhibiting pathogens and soil-borne diseases[24,25].

Discussion

The findings we present in this study will be valuable for informing policies and guiding decision-making processes going forward. Operationally, there may be both barriers and opportunities; a few examples are briefly discussed here. First and foremost, addressing the bottom line for farmers—that is, the economic return—is fundamental. The current market situation would make the conversion of maize acreage to soy an unattractive option (net benefits averaging US$1,485 ha^{-1} for maize versus US$1,086 ha^{-1} for soy; Supplementary Table 5). The provision of financial incentives could have a key role in motivating smallholder farmers towards more soy production. In 2017 and 2018, Chinese soy producers were provided US$384 ha^{-1} in subsidy from the central government; a hefty US$706 ha^{-1} was provided in Heilongjiang province, where growth conditions particularly favour soy production[26]. Such incentive programmes would help promote a transition towards more soy planting. In fact, China's soy acreage increased from 6.8 to 8.2 Mha during 2015–2017[11]. For the three provinces where the highest increases in soy acreage are projected in our simulation for 2035 (Heilongjiang, Inner Mongolia and Jilin), there is a trend with soy acreage increasing 0.64 Mha per year during 2015–2017; continuing this trend would bring their combined soy acreage to 13.3 Mha by 2035, exceeding our projected 11.0 Mha (Fig. 2).

That farming in China is performed by millions of smallholders who are resource poor and knowledge limited could be a major challenge for broadly implementing enhanced management technologies[27]. But recent success in engaging and empowering smallholders with markedly improved production and environmental performance is encouraging[18,28]. Still, transferring enhanced management technologies into the hands of millions of smallholders will require institutional, infrastructural and capital commitments[28]. Furthermore, steady and forward-looking policies on land tenure and land stewardship that emphasize long-term soil

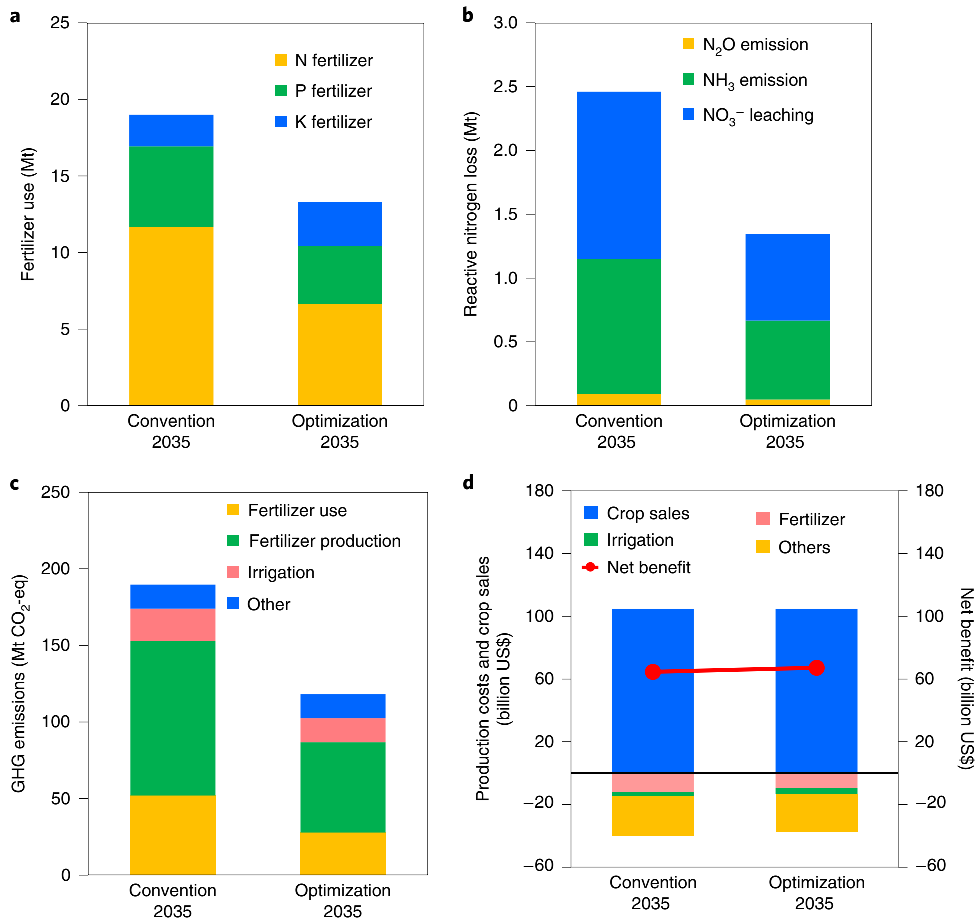

Fig. 3 | Fertilizer input, reactive nitrogen losses, GHG emissions and cost–benefit analysis for projected maize and soy production in 2035 under conventional versus enhanced management scenarios. a, Estimated nitrogen (N), phosphorus (P) and potassium (K) fertilizer use. **b**, Reactive nitrogen loss. **c**, GHG emissions. **d**, Production costs and crop sales. Production costs include costs for fertilizer, irrigation and other (for example, machinery, labour, pesticides, seed and so on). Net benefit (red dots and line in **d**) is the difference between crop sales and production costs.

health and productivity are essential. In Heilongjiang and Xinjiang, for example, large-scale farms have shown an increased yield growth for maize and soy production, compared to smallholders (Supplementary Fig. 8).

More broadly, large volumes of food-processing residues and waste are generated routinely[1]; China is not an exception[29]. Leveraging livestock to upcycle those biomass materials for meat, milk and egg production is a viable option, as animals are natural bioprocessors capable of digesting the biomass and capturing some of the nutrients (calories, proteins, minerals and so on) that are contained in the otherwise wasted materials. Such livestock-enabled upcycling would, in turn, reduce the needs for maize and soy as feed grains[30]. In addition, policies and actions aimed at changing consumer food-wasting behaviours, as well as promoting healthy eating, are important[29].

The comprehensive analyses presented above are subject to limitations. First, with counties being the base unit, intra-county variations in soil and other conditions are not considered because of data gaps at the sub-county level. Second, although we included 10-year weather variability (2005–2014) in our analysis, the potential impact of climate change on crop production for the coming decade is not factored in. Previous studies have reported yield increases in some places but decreases in others according to climate variability[31]. Given the broad range of biophysical and climatic conditions across China, we assume that the yield-boosting and yield-inhibiting effects of climate change might cancel each other out to a certain extent and thus the net effects may be small. Third, our analyses use existing data generated in the recent past; for example, the attainable yield was set as the actual yield achieved by top-producing farmers. Meanwhile, continued agricultural innovations such as new and improved seeds (including genetically modified seeds) or more advanced management technologies are likely to further boost yield[10]. These potential improvements may compensate for the negative effects of climate change but were not taken into account in our analysis.

Our study goes beyond previous work on quantitatively defining the yield gap to assess the production improvement capacity for a given crop, as we optimize maize and soy production for improved productivity and environmental outcomes. The refined county-level resolution at which our comprehensive analyses are performed and the subsequent results reflect diverse production conditions across the country. China can address its maize and soy needs more sustainably and productively while lowering resource–environment–climate burdens domestically and elsewhere. To attain these goals, China needs to design innovative and effective policies, establish implementation strategies with measurable milestone targets and take concrete and decisive actions. Our conceptual framework is transferable to other regions or countries.

Methods

Maize and soy yield potential. The Hybrid-Maize (2017) and SoySim (1.0) models were used to simulate the yield potential of maize and soy. A detailed description of model structure and parameterization has been published previously[32–34]. These models are process-based and have been used to simulate crop development, dry matter production and yield potential in the USA and China[15,16]. We have tested the models in previous studies[16,35] under irrigated as well as rain-fed conditions. We ran the models using county-level weather data from 2005–2014 for numerous combinations of varieties, sowing dates and plant densities. A combination of maize growing degree days and soy maturity group, sowing date and plant densities that maximized yield potential over the 10 seasons was selected as the optimal management inputs (Supplementary Fig. 2 and Supplementary Discussion). We simulated Y_p and Y_w for each county using optimal management, respectively. An area-weighted yield potential, Y_{pot}, was obtained per county based on the ratio of irrigated and rain-fed areas and relevant Y_p and Y_w values (equation (1)). Irrigated and rain-fed land was identified in each county using SPAM 2005 (http://mapspam.info; Supplementary Fig. 3a,b).

Model simulation of annual Y_p and Y_w requires daily weather records and basic soil information. Daily weather records from 2005–2014 were obtained for a total of 1,839 counties (Supplementary Fig. 7). The daily maximum and minimum temperatures, relative humidity, precipitation and reference evapotranspiration data were obtained from interpolated weather data with a spatial resolution of 0.25° × 0.25° and designed for high-resolution climate modelling (CN05.1_A_1961_2015_B_025x025.ctl) from 2,416 station observations. Solar radiation data were obtained from the gridded China Meteorological Forcing Dataset (CFMD), developed by the Institute of Tibetan Plateau Research, Chinese Academy of Sciences[36,37]. The gridded meteorological data were subsequently resampled for each county through bilinear interpolation. This process resulted in the creation of one weather dataset per county per day.

Soil properties as model inputs were obtained from the Harmonized World Soil Database, v.1.1 (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). Soil-type data were obtained from the Resource and Environment Data Cloud Platform (http://www.resdc.cn/data.aspx?DATAID=202).

Area-weighted yield potential yield for each county was determined using the following equation:

$$Y_{\mathrm{pot}} = \frac{Y_{\mathrm{p}} \times \mathrm{SA_i} + Y_{\mathrm{w}} \times \mathrm{SA_r}}{\mathrm{SA}} \quad (1)$$

where Y_{pot} stands for area-weighted yield potential, Y_p for irrigation yield potential, Y_w for rain-fed yield potential, SA_i for irrigated area in hectare, SA_r for rain-fed area in hectare and SA for both production modes (that is, SA_i plus SA_r). SA_i and SA_r are determined by multiplying the total planted area for the crop per county (data from China Municipal Statistical Yearbook) by the irrigated versus the rain-fed ratio (obtained using SPAM 2005).

Current farmers' yields and yield gaps. Data on maize and soy yields and hectares cultivated during 2005–2014 were extracted from the China Municipal Statistical Yearbook that included records from 1,735 counties for maize and 1,589 counties for soy in China. The average current yield was calculated as the average yield over the 10-year (2005–2014) time period to represent current farmers' yield. The yield gap was the difference between the farmers' yield and area-weighted Y_{pot}.

Estimation of Y_{att}. The Y_{att} was set at the per-country average yield achieved by the top 10% producers based on a large-scale survey. The survey was conducted during 2005–2014 in 997 counties for maize and 246 counties for soy identified from the yield database as described above, involving 4.58 million maize growers and 0.27 million soy producers. These producers were interviewed face-to-face by county extension agents using a questionnaire designed to obtain information on yield, crop varieties and fertilizer use (see a previous report[18] for more detailed survey description). We calculated the average Y_{att} in each province to be used as the attainable yield for counties without survey data (Supplementary Table 1).

Projection for 2035. China's future demands for maize and soy were projected for the year 2035. The demands were assumed to be primarily for animal feeds, which were projected based on human consumption of animal products. The human population was projected to increase from 1.407 billion in 2015 to 1.461 billion by 2035 with 70% urban and 30% rural distributions differing in dietary characteristics[5,38]. The projected animal-source food consumption by 2035 is 14 kg of animal protein per capita per year, reflecting a 30% increase from that consumed in 2015 in China (11 kg of protein per capita per year), but still lower than that of the USA and Europe. We used the nutrient flows in food chains, environment and resource (NUFERNUFER) model[38] to calculate annual nitrogen and phosphorus inputs and outputs in crop and animal production and food processing, retail and consumption at the regional scale, and to project the feed grain demands for 2035, resulting in 289 Mt for maize and 133 Mt for soy (10% and 22% higher than the reference year of 2017). Much of these increases are associated with a rapid increase in monogastric animals and intensive livestock systems, which require soy as the source of high-quality protein for the animals[9].

Optimal acreage reallocation. For attainable production capacity, increases in the maize and soy yields were assumed to reach the Y_{att} in each county. When maize demand was fully met, the remaining acreage would be allocated to soy (the combined maize and soy acreage in each county would remain the same as in 2017). The optimization process consisted of three steps. First, for each of the four regions (northeast, northwest, north central, and south China; Supplementary Fig. 4a), we ranked the counties based on their ratio of maize acreage as a percentage of arable land. We then identified the top 25th percentile and used that as the cut-off point for maximal maize allowance, which turned out to be 77%, 43%, 75% and 28% for the four regions, respectively. The 'excess' acreage for a given county (that is, above the cut-off) would be allocated to soy. Through this step, 3.33 Mha of maize acreage was reallocated to soy. Second, the reference-year maize acreage was reduced by 25% in the water deficit regions, which consist of 186 counties (Supplementary Fig. 4b), based on current policy because of the fragile ecological environment. This led to the reallocation of 0.99 Mha from maize to soy. Finally, all counties were ranked based on Y_{att}. Starting from the the lowest Y_{att}, counties would be taken one-by-one with 75% of reference-year maize acreage reallocated to soy, until national cumulative maize production reached 289 Mt. In the end, a total of 30.7 Mha would be used to produce 289 Mt maize, and 20.0 Mha for producing 60 Mt soy (Fig. 2).

Estimation of fertilizer use. County-level nitrogen, phosphorus and potassium fertilizer rates were calculated by dividing the crop yield by the fertilizer use efficiency (kg grain kg^{-1} nutrient). The latter, including nitrogen, phosphorus and potassium fertilizer efficiencies for enhanced versus conventional management scenarios, were estimated using the results from large-scale on-farm experiments (Supplementary Table 2). In total, 6,089 (maize) and 2,073 (soy) site-years of field trials were conducted to increase grain yield and optimize fertilizer inputs using integrated management, with sites spread across 654 counties for maize and 141 counties for soy from 2005 to 2015. Each field trial included farmers' practices and ISSM-based recommendations. Results for maize have been reported[18]. For soy, the results, not reported previously, are used in the current study.

Estimation of reactive nitrogen loss. The Web of Science and China National Knowledge Infrastructure databases were searched to identify articles published between January 1995 and August 2018 for N_2O, NH_3 or NO_3^- loss during maize and soy production. The screen criteria were: (i) measurement of reactive nitrogen losses in fields throughout the growing season; (ii) inclusion of a control with zero nitrogen input in the measurement of reactive nitrogen loss; (iii) nitrogen application in the form of urea or ammonium, excluding slow-release or controlled-release fertilizers or organic materials, such as manure and compost; and (iv) determination of N_2O emissions using closed static chambers, NH_3 emissions using continuous air flow chambers, venting, a Drager-Tube or a microclimate with wind tunnels, and NO_3^- leaching using a suction cup and lysimeters, soil analysis or hydrological modelling. In total, 139 peer-reviewed studies contributing 634 observations (282 from 89 studies on N_2O, 246 from 49 studies on NH_3 and 106 from 18 studies on NO_3^-; see Supplementary Discussion) were identified. The loss factor for N_2O, NO_3^- and NH_3 was calculated from net losses obtained from these studies, and expressed as a percentage of nitrogen applied.

The loss factors during maize and soy production were determined at the county level. A random forest regression model[39] was developed using the parameter measurements synthesized from the 139 studies mentioned above. The model used ensemble machine learning techniques and presented a set of binary decision rules based on input variables. Data were processed in three steps. First, the model randomly selected a bootstrap sample of observations ('in-bag' data), which were the equivalent of the entire set of observations, along with replacements. About 37% of the initial observations were unselected, referred to as 'out-of-bag' (OOB) data[40]. The model selected a regression tree for each sample and randomly selected explanatory variables with m_{try} features (m_{try}: number of variables randomly sampled as candidates at each split. Note that the default values are different for classification (sqrt(p) where p is number of variables in x) and regression (p/3))[39]. A regression tree was built based on the in-bag data and m_{try} variables selected. All splits of the tree were examined with predictor variables, and the best split at each step was determined to classify and build the regression tree. The regression and MSE_{OOB} were developed as follows:

$$\mathrm{MSE_{OOB}} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}^{\mathrm{OOB}}\right)^2 \quad (2)$$

where MSE_{OOB} is the mean square error of the prediction, y_i is observation i and $\hat{y}^{OOB}$ is the predicted value of OOB. The OOB observation i in the regression tree denotes the mean reactive nitrogen loss for the in-bag observations at the same terminal node.

Nitrogen loss factors served as the dependent variable in the random forest regression model (Supplementary Fig. 5). Major factors affecting reactive nitrogen losses were climate (precipitation, potential evapotranspiration, mean temperature and mean wind speed), soil (sand content, pH and organic matter) and nitrogen balance (nitrogen applied minus above-ground nitrogen uptake), as determined using random forest modelling with 10-fold cross validation. The dataset was evaluated by dividing it into ten subsets of equal size; seven subsets were used for modelling and three were used for testing with the random forest model. The predictions were used to calculate the root mean square error of reactive nitrogen losses. Input variables for the regression tree were ranked according to their contributions to the mean square error.

Annual loss of nitrogen was computed for the counties in which production remained at the current level (2017), conventional (2035) and enhanced management scenarios (2035). The projection was conducted by random forest modelling. The climate and soil characteristics parameters were entered into the model to obtain the best-fitting simulation based on the regression coefficient of determination and variation. The amounts of N_2O, NH_3 and NO_3^- losses were calculated by multiplying loss factors by the rate of nitrogen application (kg N ha^{-1}) for each county.

Estimation of GHG emissions. Total GHG emissions from the 2017 level and in conventional and enhanced management scenarios for 2035 were calculated using a life-cycle analysis approach[41]. The occurrence of emissions included: (1) nitrogen fertilizer application, directly and indirectly from N_2O release, calculated through a meta-analysis using the random forest model (as above); (2) fertilizer (nitrogen, phosphorus and potassium) production; (3) use of electricity for irrigation and other factors, including the use of pesticides and diesel fuel for tilling and harvesting. The following calculations were made:

$$\begin{aligned}\text{Total GHG} &= 298 \times (\mathrm{N_2O_{indirect}} + \mathrm{N_2O_{direct}}) \times 44/28\\ &+\mathrm{N_{input}} \times \mathrm{EF_N} + \mathrm{P_{input}} \times \mathrm{EF_P} + \mathrm{K_{input}} \times \mathrm{EF_K} + 9.2\\ &\times \mathrm{Irr} \times \mathrm{EF_{irr}} + \mathrm{Pest_{input}} \times \mathrm{EF_{Pest.}} + \mathrm{Fuel_{input}} \times \mathrm{EF_{Fuel}},\end{aligned} \tag{3}$$

where N_2O_{direct} and $N_2O_{indirect}$ are the N_2O (kg ha^{-1}) emitted directly (from the soil during nitrogen application) and indirectly through volatilization of NH_3, respectively, accounting for soil redeposition and NO_3^- leaching[42]; 298 is the CO_2 equivalent of N_2O in terms of global warming; and 44/28 is the conversion rate of N to N_2O. N_{input}, P_{input} and K_{input} are the rates of N, P_2O_5 and K_2O application (kg ha^{-1}), respectively; EF_N, EF_P and EF_K are GHG emissions in CO_2-eq per kilogram of N, P_2O_5 and K_2O fertilizer production, respectively; 9.2 is electricity consumption by irrigation; Irr is irrigation amount (mm); EF_{irr} is the GHG emissions from the consumption of electricity; Pest $_{input}$ and $Fuel_{input}$ are pesticide and diesel fuel inputs (in kg), respectively; and EF_{Pest} and EF_{Fuel} are GHG emissions from pesticide and diesel fuel use, respectively.

Irrigation is an important factor for calculating GHG emissions and cost–benefit analysis (discussed later). We estimated irrigation water use by dividing the crop yield by the rate of irrigation water productivity (IWP). A meta-analysis was performed to quantify the effects of the enhanced management strategies on IWP (Supplementary Discussion). A review of peer-reviewed scientific journal articles on IWP for conventional and enhanced management practices published after January 2000 was conducted based on the following screening criteria: (i) all data were from field experiments; (ii) data on yield and irrigation volumes were provided; and (iii) traditional management (local water volume with furrow or flooding irrigation) and at least one optimal management strategy was applied. Optimal water management strategies included optimized irrigation volume or improved technologies, such as drip or sprinkler irrigation. A total of 83 peer-reviewed studies describing 287 paired observations of optimal and traditional water management was identified.

Cost–benefit analysis. A cost–benefit analysis was performed based on input costs and crop sales. The benefit was the gross sales of crop products (US\$ ha^{-1}), and the net benefit was the benefit minus the input costs. Input costs included purchases of nitrogen, phosphorus and potassium fertilizer, irrigation (drip and sprinkling) equipment and other operational needs (pesticides, seeds, machinery and labour). The prices of nitrogen, phosphorus and potassium fertilizer, irrigation and other costs were obtained from the National Agricultural Products Cost–Benefit Compilation of Information (2018) and the National Agricultural Products Wholesale Market Price Information System (pfsc.agri.cn) (Supplementary Table 5). Seed expenditures would increase according to (i) farmers' shifts from older varieties to newer varieties (based on https://www.cnhnb.com/); and (ii) increased seed for high planting density. The increase in machinery cost was calculated based on the increased use of fuel (diesel) for fertilization and water for irrigation, seeding and harvesting. The cost of irrigation equipment[43] in the enhanced scenario included all expenses related to drip irrigation and sprinkling irrigation equipment, which were annualized assuming a lifespan of 5 years and a discount rate of 10%.

Data management. The databases were developed and maintained using Microsoft Excel 2013. Weather records were analysed using MATLAB R2017a. Data obtained from national surveys were stored and analysed using the SQL Server 2008. Graphics were drawn using SigmaPlot v.12.5 and Microsoft Excel 2013. A publicly released map of China was obtained from the Resource and Environmental Data Cloud Platform (http://www.resdc.cn), and all map-related operations were performed using ArcGIS 10.2 software (www.esri.com/en-us/arcgis). Computations and predictions were performed using the Random Forest Environmental Model, R 3.5.1.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data supporting the findings of this study are available within the paper and its Supplementary Information and Supplementary Data files. Source data are provided with this paper.

Code availability

The custom code generated for this study is available in the Supplementary Data file.

Received: 3 September 2020; Accepted: 11 May 2021;
Published online: 14 June 2021

References

1. *FAOSTAT: Statistics Division of the Food and Agriculture Organization of the United Nations* (FAO, 2019); http://www.fao.org/faostat/en/#home
2. Taherzadeh, O. & Caro, D. Drivers of water and land use embodied in international soybean trade. *J. Clean. Prod.* **223**, 83–93 (2019).
3. Sun, J. et al. Importing food damages domestic environment: evidence from global soybean trade. *Proc. Natl Acad. Sci. USA* **115**, 5415–5419 (2018).
4. Ying, H. et al. Safeguarding food supply and groundwater safety for maize production in China. *Environ. Sci. Technol.* **54**, 9939–9948 (2020).
5. Hill, J. et al. Air-quality-related health damages of maize. *Nat. Sustain.* **2**, 397–403 (2019).
6. Fuchs, R. et al. US–China trade war imperils Amazon rainforest. *Nature* **567**, 451–454 (2019).
7. Rajao, R. et al. The rotten apples of Brazil's agribusiness. *Science* **369**, 246–248 (2020).
8. Foley, J. A. et al. Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
9. Bai, Z. H. et al. China's livestock transition: driving forces, impacts, and consequences. *Sci. Adv.* **4**, 8534 (2018).
10. Cassman, K. G. & Grassini, P. A global perspective on sustainable intensification research. *Nat. Sustain.* **3**, 262–268 (2020).
11. National Bureau of Statistics of China. *China Statistical Yearbook* (in Chinese) (China Statistics Press, 2001, 2018).
12. Cui, Z. L., Vitousek, P. M., Zhang, F. S. & Chen, X. P. Strengthening agronomy research for food security and environmental quality. *Environ. Sci. Technol.* **50**, 1639–1641 (2016).
13. Liu, W. F. et al. China's food supply sources under trade conflict states and limited domestic land and water resources. *Earths Future* **8**, e2020EF001482 (2020).
14. *Commodity Markets* (World Bank, 2019). https://www.worldbank.org/en/research/commodity-markets
15. Liu, B. H. et al. Estimating maize yield potential and yield gap with agro-climatic zones in China—distinguish irrigated and rainfed conditions. *Agr. Forest Meteorol.* **239**, 108–117 (2017).
16. Chen, X. P. et al. Integrated soil–crop system management for food security. *Proc. Natl Acad. Sci. USA* **108**, 6399–6404 (2011).
17. Chen, X. P. et al. Producing more grain with lower environmental costs. *Nature* **514**, 486–489 (2014).
18. Cui, Z. L. et al. Pursuing sustainable productivity with millions of smallholder farmers. *Nature* **555**, 363–366 (2018).
19. Van Ittersum, M. K. et al. Yield gap analysis with local to global relevance—a review. *Field Crops Res.* **143**, 4–17 (2013).
20. *Global Yield Gap Data* (Global Yield Gap Atlas, 2020). http://www.yieldgap.org
21. Cassman, K. G., Dobermann, A., Walters, D. T. & Yang, H. Meeting cereal demand while protecting natural resources and improving environmental quality. *Annu. Rev. Environ. Resour.* **28**, 315–358 (2003).
22. Lobell, D. B., Cassman, K. G. & Field, C. B. Crop yield gaps: their importance, magnitudes, and causes. *Annu. Rev. Environ. Resour.* **34**, 179–204 (2009).
23. Sentelhas, P. C. et al. The soybean yield gap in Brazil—magnitude, causes and possible solutions for sustainable production. *J. Agr. Sci.* **153**, 1394–1411 (2015).
24. Drinkwater, L. E., Wagoner, P. & Sarrantonio, M. Legume-based cropping systems have reduced carbon and nitrogen losses. *Nature* **396**, 262–265 (1998).
25. Snapp, S. S. et al. Biodiversity can support a greener revolution in Africa. *Proc. Natl Acad. Sci. USA* **107**, 20840–20845 (2010).
26. Price Bureau of the National Development and Reform Commission of China. *China Agricultural Products Cost-Benefit Compilation of Information 2017* (in Chinese) (China Statistics Press, 2017).
27. Spielman, D. J., Byerlee, D., Alemu, D. & Kelemework, D. Policies to promote cereal intensification in Ethiopia: the search for appropriate public and private roles. *Food Policy* **35**, 185–194 (2010).
28. Zhang, W. F. et al. Closing yield gaps in China by empowering smallholder farmers. *Nature* **537**, 671–674 (2016).

29. Liu, J. G., Lundqvist, J., Weinberg, J. & Gustafsson, J. Food losses and waste in China and their implication for water and land. *Environ. Sci. Technol.* **47**, 10137–10144 (2013).
30. Dou, Z. X., Toth, J. D. & Westendorf, M. L. Food waste for livestock feeding: feasibility, safety, and sustainability implications 2017. *Glob. Food. Secur.* **17**, 154–161 (2018).
31. Challinor, A. J. et al. A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Change* **4**, 287–291 (2014).
32. Yang, H. S. et al. Hybrid–Maize—a maize simulation model that combines two crop modeling approaches. *Field Crops Res.* **87**, 131–154 (2004).
33. Yang, H. S., Dobermann, A., Cassman, K. G. & Walters, D. T. Features, applications, and limitations of the Hybrid–Maize simulation model. *Agron. J.* **98**, 737–748 (2006).
34. Setiyono, T. D. et al. Simulation of soybean growth and yield in near optimal growth conditions. *Field Crops Res.* **119**, 161–174 (2010).
35. Guo, J. M. et al. Designing corn management strategies for high yield and high nitrogen use efficiency. *Agron. J.* **108**, 922–929 (2016).
36. Wu, J., Gao, X. J., Giorgi, F. & Chen, D. L. Changes of effective temperature and cold/hot days in late decades over China based on a high resolution gridded observation dataset. *Int. J. Climatol.* **37**, 788–800 (2017).
37. Yang, K. & He, J. *China Meteorological Forcing Dataset (1979–2018)* (National Tibetan Plateau Data Center, 2019); https://doi.org/10.11888/AtmosphericPhysics.tpe.249369.file
38. Ma, L. et al. Exploring future food provision scenarios for China. *Environ. Sci. Technol.* **53**, 1385–1393 (2019).
39. Prasad, A. M., Iverson, L. R. & Liaw, A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**, 181–199 (2006).
40. Audsley, E. et al. *Harmonisation of Environmental Life Cycle assessment for Agriculture: Final Report Concerted Action MR3-CT94-2028* (Silsoe Research Institute, 1997).
41. Liu, Q. et al. Biochar application as a tool to decrease soil nitrogen losses (NH_3 volatilization, N_2O emissions, and N leaching) from croplands: options and mitigation strength in a global perspective. *Global Change Biol.* **25**, 2077–2093 (2019).
42. Eggelston, S. et al. *2006 IPCC Guidelines for National Greenhouse Gas Inventories* (IGES, IPCC, 2006).
43. Lin, H. F. et al. Effect of irrigation method on farmers' planting decision and the economy: a case in Zhangbei County, Hebei Province. *Chin. J. Eco Agric.* **27**, 1293–1300 (2019).

Acknowledgements

We acknowledge all those who provided local assistance or technical services involving the farmer survey. We also thank Z. Wu for editing the manuscript. This work was financially supported by the National Key Research and Development Program of China (2016YFD0200105), the Taishan Scholarship Project of Shandong Province (no. TS201712082) and the Science and Technology Plan Project of Qinghai Province (2019-NK-A11-02).

Author contributions

Z.C. designed the research and supervised the project. Z.L., M.C., H.Y. and F.Z. performed research. Z.L., Z.B., Y. Yin, M.C., J.B., Y.X., Q.Z., Y. Yang, H.Y. and M.D. collected and analysed the data. Z.C., Z.D., Z.L., W.D.B. and Y.G. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43016-021-00300-1.

Correspondence and requests for materials should be addressed to Z.C.

Peer review information *Nature Food* thanks Martin van Ittersum and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Zhenling Cui

Last updated by author(s): Apr. 5, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Hybrid-Maize model 2017 and SoySim model 1.0 were used to simulated the yield potential of maize and soy respectively.
Data analysis	Weather records were analyzed using MATLAB R2017a. Data obtained from national surveys were stored and analyzed using SQL Server 2008. Computations and predictions were performed using Random Forest Environment Model, R 3.5.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data presented in our study as Figures and Tables (including Extended Data Figures and Tables) are all available in the Text and Supporting Information or otherwise can be obtained from corresponding authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This work is an integrated study which contains: (i) crop models, Hybrid-Maize (2017) and SoySim (1.0), for accurately predicting yield potential; (ii) large scale statistical data summary (county-level yield and sowing area of maize, and soy); (iii) large farm survey data summary (the yield and fertilizer N application of top 10% and average producers of each county);(iv) the NUFERNUFER model for predicting maize and soy demands by 2035;(v)a random forest (RF) regression model for estimating Nr losses, and to assess Nr losses, GHG emission and Cost-benefit of conventional and optimized scenarios in 2035;(iv) the impact of reduction of China's soy importation on global trade and economy by GTAP model.
Research sample	For the yield potential and improvement, the county-level yield potential of maize and soy was simulated using Hybrid-Maize and SoySim respectively and current yield were extracted from China Municipal Statistical Yearbook. For the impact on N losses and greenhouse gas emissions, we conducted a literature search using Web of Science and the China National Knowledge Infrastructure (CNKI) for relevant articles published between January 1995 and August 2018. Selection criteria were as follows: (i) Nr losses must have been measured during field operations and throughout the entire growing season; (ii) Nr losses must have contained at least two N input levels, including a zero-N control; and (iii) N application was in the form of urea and ammonium. For outlook of yield improvement, the yield and fertilizer N application of top 10% and average producers of each county obtained from a face-to-face survey by county extension agents.
Sampling strategy	For the yield potential and improvement, we estimated the yield potential and current yield in 1,735 counties for maize and 1,589 counties for soy from 2005 to 2014, together accounting for 97% of the total hectares for these crops in China. We aggregated the 141 GTAP regions into 19 region-blocks based on their proportions in the soy market. The 5 primary factors were also aggregated into 3 factors (including land, capital and labor).For the impact on N losses and greenhouse gas emissions, we collected data from 158 peer-reviewed studies, include 612 observations (90 studies with 274 observations for N2O emission, 50 studies with 234 observations for NH3 volatilization, and 18 studies with 104 observations for NO3− leaching). For outlook of yield improvement, we summarized a survey 2005-2014 in 997 counties for maize and 246 counties for soybeans, involving 4.58 million producers for maize and 0.27 million producers for soy.
Data collection	We used Hybrid-Maize and SoySim model to estimate the yield potential of maize and soy based on optimized managements in planting date, plant density, and maturity condition. We obtained county-level current yield from China Municipal Statistical Yearbook. For the impact on N losses and greenhouse gas emissions, we conducted a literature search using Web of Science and the China National Knowledge Infrastructure (CNKI) for relevant articles published between January 1995 and August 2018. We calculated the yield and fertilizer N application of top 10% and average producers of each county obtained from a face-to-face survey by county extension agents.
Timing and spatial scale	All the statistical data were summarized from 1961 to 2017; and model simulation data were simulated from 2005 to 2014; Nr loss data of maize and soybean production from publications between January 1995 and August 2018.
Data exclusions	None.
Reproducibility	Our study is an integrated study mainly based on model simulation ,statistical data and reference data. Our results can be reproduced when following the described methods and data.
Randomization	This is not relevant to our study because our work is not an "Experimental" study but an integrated data analysis. We used national statistical data and published data to do the integrated analysis.
Blinding	Blinding is not possible in our study. Because we have no choice to collect the national statistical data or reference data (just according to key words) considering the blinding principle.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging