






ARTICLE

<https://doi.org/10.1038/s41467-019-09447-9>

OPEN

# Closing yield gaps for rice self-sufficiency in China

Nanyan Deng <sup>1</sup>, Patricio Grassini <sup>2</sup>, Haishun Yang<sup>2</sup>, Jianliang Huang <sup>1</sup>, Kenneth G. Cassman <sup>2</sup> & Shaobing Peng <sup>1</sup>

China produces 28% of global rice supply and is currently self-sufficient despite a massive rural-to-urban demographic transition that drives intense competition for land and water resources. At issue is whether it will remain self-sufficient, which depends on the potential to raise yields on existing rice land. Here we report a detailed spatial analysis of rice production potential in China and evaluate scenarios to 2030. We find that China is likely to remain self-sufficient in rice assuming current yield and consumption trajectories and no reduction in production area. A focus on increasing yields of double-rice systems on general, and in three single-rice provinces where yield gaps are relatively large, would provide greatest return on investments in research and development to remain self-sufficient. Discrepancies between results from our detailed bottom-up yield-gap analysis and those derived following a top-down methodology show that the two approaches would result in very different research and development priorities.

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement, MARA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070 Hubei, China. <sup>2</sup>Department of Agronomy and Horticulture, University of Nebraska-Lincoln, PO. Box. 830915, Lincoln, NE 68583-0915, USA. Correspondence and requests for materials should be addressed to S.P. (email: [speng@mail.hzau.edu.cn](mailto:speng@mail.hzau.edu.cn))

China must make strategic decisions about how to ensure food security for 1.4 billion people, and those decisions will have a large impact on global agriculture and land use. Currently self-sufficient in rice production, China's most important food crop, the annual production of about 206 million metric tons (MMT) represents 28% of global rice supply<sup>1,2</sup>. At issue is whether agricultural policy should target continued self-sufficiency in rice or accept becoming a major rice importer. Resolution of this issue will markedly influence global rice markets because reliance on imports for just 10% of China's rice consumption represents more than 35% of all internationally traded rice based on 2013–2015 global export statistics<sup>3</sup>.

Current self-sufficiency has been achieved by raising rice yields more than 50% since 1980 despite an 11% reduction in harvested rice area<sup>1</sup>. But the rate of rice yield growth has slowed markedly in recent years<sup>4</sup>, and prime farmland and water for irrigation are becoming increasingly scarce as the rural-to-urban demographic shift drives fierce competition for both land and water<sup>5,6</sup>. Given the prognosis for little increase, or even a reduction, in rice production area, making an informed decision about whether to pursue continued self-sufficiency depends on the potential for increasing yields on existing rice area. Yield-gap analysis provides the means for estimating this untapped production potential by estimating the difference (i.e., gap) between current farm yields and the potential yield that can be achieved when yield losses from nutrient deficiencies, pests, and diseases are minimized<sup>7</sup>.

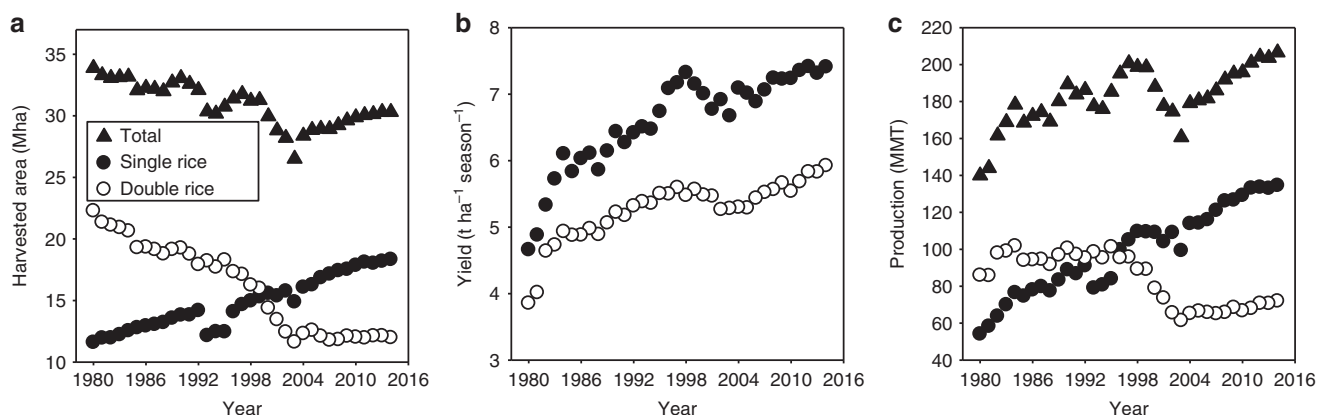
Robust estimation of rice yield gaps in China is complicated by the large variation in production systems and climates in which rice is grown, from warm sub-tropics at 18° N latitude to cool temperate climates at 50° N. Potential yield is a location-specific property because it depends on the local weather and crop growth duration. In the irrigated systems that dominate rice production in China, location-specific factors governing potential yield include the length of growing season, determined by temperature regime, and the amount of light intercepted by the leaf canopy during the crop growth period, determined by incident solar radiation, leaf area development, and persistence<sup>8</sup>. Hence, accurate yield gap estimation requires good quality, long-term weather records, and data on current crop yields and management practices with adequate spatial resolution to support simulation of potential yield across the large environmental variation that characterizes Chinese rice production<sup>9</sup>. To date, it has not been possible to perform such a detailed analysis due to lack of both a spatially explicit dataset on rice production systems across this wide range of environments, and a suitable upscaling technique for aggregating results to a national scale.

To fill this void, we report a spatially explicit yield-gap analysis of Chinese rice production using primary data and bottom-up scaling methods recently developed for the Global Yield Gap Atlas (GYGA) ([www.yieldgap.org](http://www.yieldgap.org))<sup>9,10</sup>. In this yield-gap analysis, we evaluate future scenario options based on estimated rice production capacity on current Chinese rice area, identifying specific regions and rice production systems that deserve highest priority for research and development investments to achieve greatest rice production on a limited supply of prime farmland. We find that a focus on increasing yields of double-rice systems in general (i.e., two rice crops per year planted and harvested in the same field), and on three provinces where yield gaps are relatively large and single-rice systems predominate (i.e., one rice crop per year on a given field), would provide greatest return on investments in research and development. Discrepancies between results from our detailed yield-gap analysis and those derived following a top-down methodology indicate that the two approaches would result in very different research and development priorities.

## Results

**Recent trends in rice cropping systems.** The geography of Chinese rice production has undergone enormous changes over the past 35 years. Whereas double-rice systems that dominate in the warm climates of central-south and south coastal regions accounted for 66% of national total harvested rice area in the 1980s, they currently represent less than 40% (Fig. 1). In contrast, area given to single-rice systems in cooler climates of central and northern regions has increased steadily although this expansion did not overcome the reduction in double-rice area. Hence, total harvested rice area has decreased by about 3.6 million hectares (Mha) since 1980. The main reasons for decrease in double-rice were (i) rapid urbanization in south coastal and central-south regions and associated conversion of rice land to housing, industry, and supporting infrastructure, and (ii) decreased rural labor availability and rising labor costs leading to lower net income than for single-rice systems, which require less labor<sup>11</sup>. Taken together, trends in production area and yields have resulted in the dominance of single-rice systems, which now account for more than 65% of national rice production (Fig. 1).

**Current potential yield and yield gaps.** To achieve the required level of spatial resolution for robust estimation of irrigated rice yield gaps requires primary data for at least 10 years of daily weather records, digital maps of current rice production area and associated rice yields, and the dominant rice cropping systems



**Fig. 1** Rice production trends in China. Trends in harvested rice area (a), rice yield (b), and total production (c) for single- and double-rice cropping systems during the past 35 years (1980–2014) in China. Note that the yield for double-rice is on a per-harvested area basis so that total annual yield is twice the values shown. Data were obtained from ref. 1. Mha million hectares, MMT million metric tons. Source data are provided as a Source Data file

and management practices<sup>9,10</sup>. Fifty locations representing 16 climate zones were selected for the reference weather stations (RWS). Because both single- and double-rice systems are prevalent at some of these locations, a total of 847 simulations of rice potential yield were required using a well-validated ORYZA rice simulation model<sup>12</sup> (details on calibration and validation provided in Supplementary Information). These location-specific estimates were then scaled up, based on the proportion of national rice area represented by each location and rice cropping system, to give national estimates of potential yield and current farm yield of 9.8 and 6.8 metric tons per hectare ( $\text{t ha}^{-1}$ ) per crop in single- and double-rice systems, respectively, and a national average yield gap of  $3.0 \text{ t ha}^{-1}$ . Hence, current national average rice yield represents 69% of potential yield, which is approaching the 75–80% of the potential yield threshold at which farm yields typically stagnate at regional to national scales due to diminishing returns from further investment in yield-enhancing technologies and inputs<sup>13</sup>.

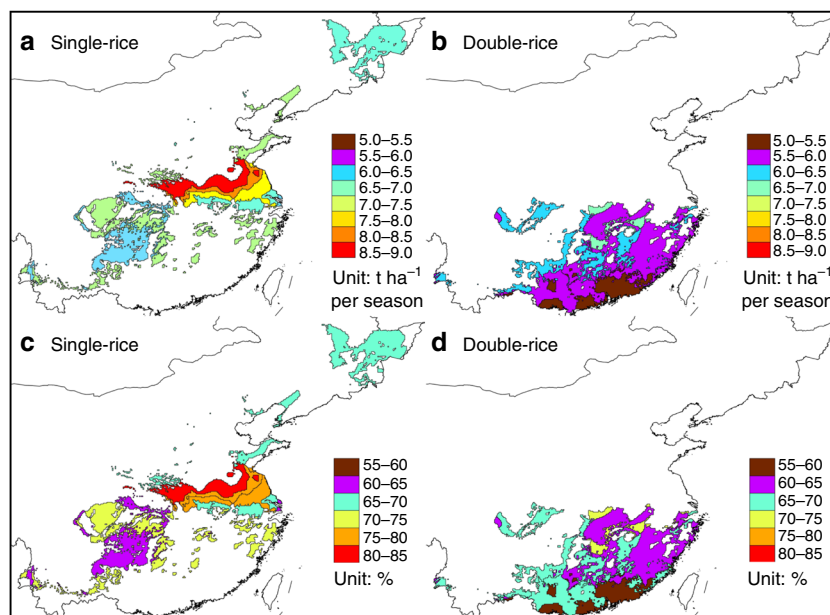
Evaluating yields by climate zone identifies large differences in potential yield among regions and rice cropping systems. For example, estimated potential yields ranged from 8.6 to  $10.8 \text{ t ha}^{-1}$  across climate zones and rice systems, while current farm yields varied from 5.2 to  $8.8 \text{ t ha}^{-1}$  (Fig. 2). Farm yields were highest in central regions for both single- and double-rice systems. Year-to-year variability in potential yield was small for both systems as indicated by a temporal coefficient of variation (CV) of 8%, which is typical of the high yield stability found in grain production systems with a reliable supply of irrigation water<sup>14</sup>. In double-rice systems, potential yields of early- and late-season rice crops were similar, with a national average of  $9.0 \text{ t ha}^{-1}$  for both crops. Although national potential yield of single-rice was 14% greater than that of double-rice per season ( $10.3$  versus  $9.0 \text{ t ha}^{-1}$ ), total annual potential yield from double-rice was  $18 \text{ t ha}^{-1}$  because two crops are produced each year from the same field.

Larger spatial variation in average current farm yields (CV = 14%) than in potential yields (CV = 8%) resulted in a wide range of yield gaps, from 18% to 41% of potential yield across the 16 climate zones evaluated. Current national farm yields for single-rice ( $7.4 \text{ t ha}^{-1}$ ) and double-rice per season ( $5.9 \text{ t ha}^{-1}$ ) were 72% and 66% of the potential yields estimated for each system, respectively. Hence, yield gaps of single-rice systems are very close to the 75–80% potential yield threshold at which farm yields tend to stagnate, whereas yield gaps in double-rice systems are considerably below this threshold.

Assuming the exploitable yield gap is estimated by the difference between current farm yield and 80% of potential yield (hereafter called the exploitable yield ceiling), exploitable yield gaps for each of the two crops in double-rice systems are 44% greater than for single-rice ( $1.3$  versus  $0.9 \text{ t ha}^{-1}$  per season, Table 1). Annual per hectare increase in rice production from closing exploitable yield gaps would be three-fold greater for double-rice than for single-rice ( $2.6$  versus  $0.9 \text{ t ha}^{-1}$ ). Correcting for the larger current production area of single-rice gives a total potential increase in rice production of about 16 MMT from double-rice and 15 MMT from single-rice with the closure of exploitable yield gaps on all current rice area (Fig. 3a). Taken together, a scenario of closing exploitable yield gaps on existing rice area would increase national rice production by 15% (+31 MMT, Table 1) compared to current rice production of 206 MMT (average of 2013–2015)<sup>1</sup>.

But achieving yields that are 80% of potential yield requires large inputs of fertilizer nutrients and aggressive use of pest control measures to minimize yield losses caused by diseases, insects, and weeds in these continuous rice systems. Precise timing of these inputs with regard to stage of crop development is also required, which means greater investment in labor and expertise to monitor crop status and make tactical modifications to field management during the growing season in response to weather and expected yield levels. Such intensive management may not be economically justifiable if marginal costs of the additional inputs, labor, equipment, and decision-support tools do not cover expected returns, which appears to be the case in California where irrigated rice yields have stagnated at 76% of potential yield<sup>15</sup>. Hence, if the exploitable yield ceiling for profitable rice production is only 75% of potential yield, increased production capacity would be 8% of current production (+16 MMT, Fig. 3a).

But achieving yields that are 80% of potential yield requires large inputs of fertilizer nutrients and aggressive use of pest control measures to minimize yield losses caused by diseases, insects, and weeds in these continuous rice systems. Precise timing of these inputs with regard to stage of crop development is also required, which means greater investment in labor and expertise to monitor crop status and make tactical modifications to field management during the growing season in response to weather and expected yield levels. Such intensive management may not be economically justifiable if marginal costs of the additional inputs, labor, equipment, and decision-support tools do not cover expected returns, which appears to be the case in California where irrigated rice yields have stagnated at 76% of potential yield<sup>15</sup>. Hence, if the exploitable yield ceiling for profitable rice production is only 75% of potential yield, increased production capacity would be 8% of current production (+16 MMT, Fig. 3a).

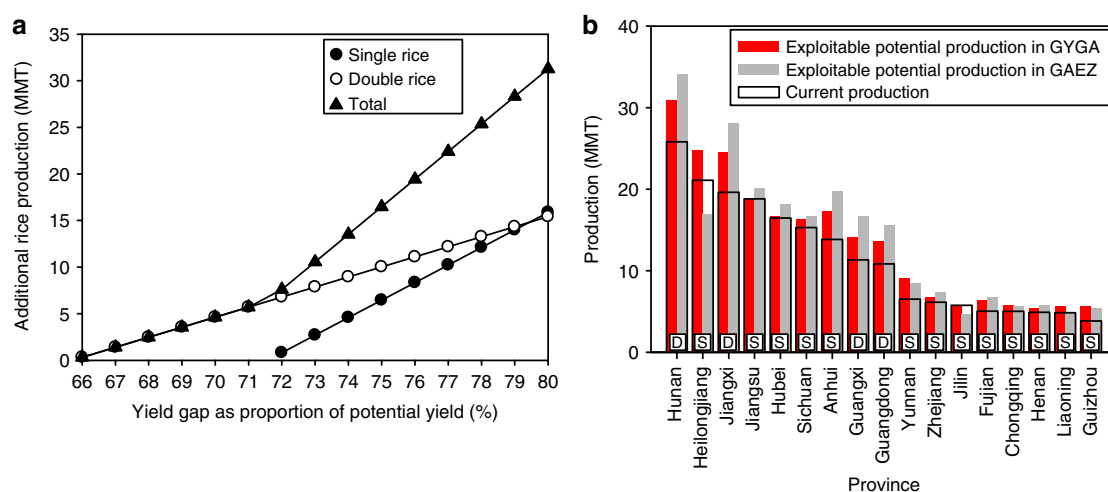


**Fig. 2** Current farm yields and yield gap in China. Current farm yields as absolute values (a, b) or as a percentage of potential yields (c, d). All values are reported on a per-harvest basis and are mapped at the climate zone spatial scale. Note that there may be several climate zones within areas showing the similar current farm yield, and thus having the same color in these figures. Source data are provided as a Source Data file

**Table 1** Yields and production of single- and double-rice systems in China and total national production under four scenarios projected to 2030

Scenarios	Rice system	Yield ( $\text{t ha}^{-1}$ season $^{-1}$ )	Production (MMT)	Total production (MMT)	Total production compared with demand of 217 MMT in 2030 (MMT)
S1	Single-rice	7.4	135.4	206	-11
	Double-rice	5.9	71.0		
S2, 80% $Y_p$ ceiling	Single-rice	8.3	150.5	224	7
	Double-rice	6.1	73.6		
S3, 75% $Y_p$ ceiling	Single-rice	7.8	141.2	215	-2
	Double-rice	6.1	73.6		
S4, 75% $Y_p$ ceiling	Single-rice	7.8	141.2	219	2
	Double-rice	6.5	77.9		

S1: Farm yields stagnate at current levels to 2030. S2 and S3: Rates of yield gain follow current trajectories based on regression of national rice yields versus year since 1985 to present for single- and double-rice to 2030 (Supplementary Fig. 6) and an exploitable yield ceiling that is 80% (S2) or 75% (S3) of potential yield ( $Y_p$ ). S4: Rates of yield gain in double-rice increase to the current yield growth rate of single-rice (an increase from 0.03 to 0.05  $\text{t ha}^{-1}$  per year per season) and an exploitable yield ceiling that is 75% of  $Y_p$ . In all four scenarios, there is no change in rice production area for each rice system, which is consistent with recent land use trends as explained in the text. Source data are provided as a Source Data file



**Fig. 3** Rice production in different cropping systems and provinces. **a** Additional rice production resulting from yield gap closure (expressed as % of potential yield) in single- and double-rice systems assuming no change in harvested area for each system and a maximum exploitable yield ceiling that is 80% of potential yield. **b** Current annual rice production and exploitable potential production for each major rice-growing province as estimated using protocols developed by the Global Yield Gap Atlas (GYGA) or by the Global Agro-Ecological Zones Model version 3.0 (GAEZ v3.0). Capital letters at bottom of each province production bar designate provinces in which double- (D) or single-rice (S) dominate. Exploitable potential production for each province is calculated as the product of provincial average rice planting area of 2013–2015 and the exploitable yield whereby all rice farmers achieve yields that are 80% of potential yield. The current production for each province is based on 2013–2015 average. MMT million metric tons. Source data are provided as a Source Data file

To evaluate plausible future scenarios of rice self-sufficiency, we use a projected total rice demand of 217 MMT in 2030 based on the average estimates of three studies<sup>16–18</sup>. In the four scenarios evaluated, we assume that total rice production area remains constant to 2030, which is consistent with recent trends for double-rice systems (Fig. 1), and also for central and northern provinces where single-rice systems prevail (justification for this assumption in Supplementary Materials and associated Supplementary Figs. 2, 7). Scenario 1 assumes farm yields remain at current levels to 2030. Scenarios 2 and 3 assume that current modest growth rates in rice yields (0.03 and 0.05  $\text{t ha}^{-1}$  per year in double- and single-rice, respectively, Supplementary Fig. 6) continue to 2030 and impose an exploitable yield ceiling of either 80% (scenario 2) or 75% (scenario 3) of potential yield (Table 1). By assuming the continuation of current rates of yield gain, we are assuming that improvements in rice cultivars, production practices, and changes in climate and atmospheric  $\text{CO}_2$  trajectories of the past 30 years will persist. Scenario 4 is the

same as scenario 3 but with higher growth rate in yield of double-rice to be equivalent to the yield growth of single-rice, which is modest compared to much faster growth rates experienced during the 1970s and 1980s. Although the higher rate of gain in yield of double-rice in scenario 4 represents a 60% increase compared to the rate of gain in recent decades, we believe such an increase is possible given the fact that double-rice systems have a larger exploitable yield gap than for single-rice systems and would benefit from greater research investment focused on closing this larger yield gap.

Import of 11 MMT would be required to meet projected 2030 rice demand under scenario 1, whereas scenario 2 with an exploitable yield ceiling 80% of potential yield results in a 7 MMT surplus (Table 1), which is equivalent to 11% of current global rice trade. But given concerns about the substantial environmental pollution associated with current intensive crop production practices<sup>19</sup>, a scenario with a 75% exploitable yield ceiling is perhaps more appropriate because it would require less fertilizer

input and less aggressive pest control measures. Under this less intensive production scenario, China would be nearly self-sufficient with a modest annual import requirement of 2 MMT. Strategic investment in provinces and rice systems with greatest yield gaps, however, could produce a small surplus even with the more conservative exploitable yield ceiling. For example, the greatest yield gaps are found in double-rice systems in general (Fig. 2), and in some provinces where single-rice dominates. Indeed, seven of the 17 provinces in which rice is a major crop account for 78% of the potential increase in rice supply if all rice farmers achieved yields that reach the exploitable yield ceiling (Fig. 3b). Disaggregation of exploitable production per region shows that the prioritization of crop management improvement interventions and sustainability measures in double-rice systems in general and single-rice systems in three provinces (Heilongjiang, Anhui, and Yunnan provinces) where yield gaps are relatively large, would provide greatest potential to ensure rice self-sufficiency in China.

Taken together, the larger yield gap and slower yield growth rate suggests greater potential to accelerate yield growth of double-rice, assuming targeted investment in that goal. Raising the rate of yield growth of double-rice to that of single-rice would lead to a rice production surplus of 2 MMT by 2030 with a conservative exploitable yield ceiling of 75% of potential yield. Reducing labor requirements and associated costs in double-rice also would be an important objective of the research and development portfolio focused on these systems.

#### Bottom-up versus top-down approaches to estimate yield gaps.

Previous studies have used a top-down spatial framework based on meso-scale grids (roughly 100 km<sup>2</sup>) into which weather and current crop yield data, obtained from databases at much coarser spatial scales, are interpolated to fit the smaller grid size<sup>20–23</sup>. Likewise, these previous studies do not provide estimates of potential yield as distinguished by different rice cropping systems. In these top-down studies, potential yield is estimated by either the 90th percentile of current farm yields<sup>20</sup>, or by a generic crop model not specific for rice<sup>20–23</sup> and therefore not validated for the ability to estimate rice potential yield across the wide variation of rice-growing environments in China.

Comparison of estimates using our bottom-up approach with those using the top-down framework of Global Agro-Ecological Zone (GAEZ) protocols<sup>23</sup> shows a modest difference in national potential production of 17 MMT by 2030, which represents 8% of current national production, and a relatively large fraction (27%) of current global rice trade (documentation of comparison provided in Supplemental Materials and Supplementary Table 8). And while estimates of national potential yield by the two approaches are also in relatively close agreement (10.5 t ha<sup>-1</sup> for GAEZ versus 9.8 t ha<sup>-1</sup> as reported here), potential production estimates at the provincial level (Fig. 3b) and the estimates of potential yield and yield gap at climate zone scale (Supplementary Fig. 8) differ markedly. For example, in two of the provinces GAEZ potential production estimates are well below current rice production (2013–2015 average). Moreover, in five climate zones in central and southern China, representing 43% of total Chinese rice production, GAEZ estimates a potential yield 18% greater than reported here, and in the two northernmost rice-growing climate zones the GAEZ estimate is 22% less than our estimate (based on data shown in Supplementary Fig. 8). Hence, the relatively close agreement in potential yield estimates at national scale occurs by chance and masks large differences in estimates at finer spatial scales. Such large differences at provincial and climate zone scales between the two approaches would result in very different research and development priorities that seek to

focus investments on regions and cropping systems with greatest opportunities for increasing rice production.

#### Discussion

That small changes in projected Chinese rice production would have a large impact on global rice trade highlights the need for accurate estimates of potential production on existing rice area. Hence, the spatial resolution of our yield-gap analysis by climate zone, province, and cropping system is essential to adequately inform policies and strategic plans for food security and land use. For this reason our analysis (i) distinguishes between the two major rice cropping systems, (ii) utilizes a rice simulation model that has been rigorously validated for the ability to estimate potential yield across the major rice-growing regions in China (Supplementary Materials, Supplementary Figs. 3–5), (iii) relies on at least 10 years of measured weather data (Supplementary Materials, Supplementary Fig. 1), and (iv) employs a bottom-up scaling protocol validated for capacity to reproduce crop performance across large variation in climate<sup>9,10</sup>.

Our scenario analysis about rice self-sufficiency is based on the assumption that rice yield growth rate was the same across the whole country for each single- and double-rice systems. While our scenario assessment does not account for expected climate change, a number of studies have shown that the magnitude of climate change and associated impact on rice yields is highly uncertain<sup>24–28</sup>. Given the relatively small change in climate for the first half of the century<sup>27,28</sup>, together with the short timeframe of our assessment (15 years, from 2016 to 2030), as well as the associated uncertainty about the magnitude of change in climate factors influencing potential yield<sup>24–27</sup>, we believe the assessments of rice potential yield provided in our study are both robust and policy-relevant. Likewise, our assessment of future rice production is optimistic as it assumes that future conversion of prime rice land for residential, industrial, and recreational purposes will be compensated by the addition of new rice land area with a similar level of productivity<sup>5</sup>. If our assumptions on climate change and rice land fall short of reality, this will put additional pressure to accelerate the rate of rice yield gain on existing rice land.

Crop price supports and other types of subsidies to promote self-sufficiency in food production are not considered sound agricultural policies due to high costs, market distortions, and reduced incentives for innovation and efficiencies<sup>29</sup>. In contrast, Clapp<sup>30</sup> argues that populous countries like China may benefit from maintaining self-sufficiency or near self-sufficiency, in production of their primary staple food crops. Hence, for a country like China, reliance on imports for even a small portion of total rice demand represents such a large fraction of global rice trade that pressure from Chinese rice purchases can influence global rice prices in ways that might lead to higher food prices and reduced access, not only for China but also for other countries that import rice. Such pressures are of particular concern in years when global rice supply falls short due to major drought or flooding in rice exporting countries. The good news is that the scenarios evaluated in this study indicate that China has substantial flexibility to maintain self-sufficiency in rice without increasing cropland area devoted to rice production. Indeed, by strategic targeting of investments in research and development, it may be possible to maintain self-sufficiency with a net reduction in rice cropland area, especially if the rate of yield gain can be accelerated in double-rice systems.

#### Methods

**Protocols for yield-gap assessment and spatial upscaling.** For irrigated crops such as rice in China, the yield gap represents the difference between current farm yields and the potential yield when the crop is grown without limitations from

nutrient deficiencies, insect pests, or diseases<sup>7</sup>. To estimate yield gaps we followed protocols developed by the GYGA, which utilize primary, location-specific data to the extent possible<sup>9</sup> and a robust upscaling framework to estimate yield gap at larger levels of spatial aggregation such as climate zones (CZ), regions, and national scales<sup>10</sup>. These protocols have been rigorously evaluated for their ability to estimate yield gaps using Australian rainfed wheat as a test case<sup>31</sup>. All underpinning data and protocols are available on the GYGA website (<http://www.yieldgap.org/>). The GYGA protocols are based on the climate zonation scheme developed by Van Wart et al.<sup>32</sup> that is delineated by three variables: (i) growing degree days, which determine the potential length of the crop growing season, (ii) annual aridity index, which provides an estimate of water supply as the ratio between rainfall and potential evapotranspiration, and (iii) temperature seasonality, which distinguishes between temperate and tropical climates. One hundred and thirty-nine CZ are delineated in China using this climate zonation scheme.

Briefly, a digital map of rice crop area distribution (SPAM map, with 10 × 10 km grid-cell resolution)<sup>33</sup> was superimposed on the CZ map to identify weather stations located in areas with greatest density of rice production area (Supplementary Fig. 1). Buffer zones of 100-km radius surrounding each weather station, were clipped following van Bussel et al.<sup>10</sup>, so that their borders fall within the same CZ. Weather stations were selected in sequence starting with the station and associated buffer zone with largest rice area and continuing until ca. 50% of national rice area was covered by the selected weather station buffer zones. Earlier work has shown that the inclusion of additional weather stations to achieve greater coverage of rice area does not improve national estimates of yield gap<sup>34</sup>. Selected weather stations are hereafter referred to as RWS. Following this approach, 50 RWS were selected containing 48% of national harvested rice area within their associated buffer zones. The 50 RWS are located in 16 CZ, which together contain 85% of national rice area (Supplementary Fig. 1). Details about selected RWS locations and the dominant rice cropping system within each RWS buffer zone are provided in Supplementary Table 1.

**Simulating rice potential yield.** The ORYZA rice model has been used to simulate rice growth and development of different rice cultivars across a wide range of climatic conditions in Asia, Africa, and USA<sup>15,34–36</sup>. Briefly, ORYZA rice model is a daily-step model that simulates rice phenology, canopy leaf area growth, dry matter accumulation and partitioning, and grain yield formation. Simulation of crop phenology requires calibration of four parameters: development rates for juvenile (DVRJ), photoperiod-sensitive (DVRI), panicle development (DVPR), and reproductive phases (DVRR). Photoperiod sensitivity is assumed negligible, which is consistent with evidence for modern high-yield rice cultivars used in China. Daily rate of canopy CO<sub>2</sub> assimilation is calculated from daily incoming radiation, temperature, and leaf area index. Dry matter production is simulated based on the balance between daily CO<sub>2</sub> assimilation and respiration. Total biomass is partitioned into various organs depending on allocation rates calibrated for different developmental stages. Potential yield in this model is performed by setting the subroutine of PRODENV = “POTENTIAL” and NITROENV = “POTENTIAL”, which basically assumes that the crop grows without water and nutrient limitations and without incidence of pathogens, insect pests, and weeds. In this study, we used the most recent version of this model (ORYZA v3<sup>12</sup>). Although the model still has limitations in simulating rice growth under stress environment, e.g., low temperature, drought, and nitrogen-deficient, etc., it performs well in simulating potential yield under non-stress environment<sup>12,15</sup>. A detailed description of model structure and parameterization can be found in Bouman et al.<sup>36</sup> Because potential yield is location-specific, the model requires specification of long-term daily weather records and crop management practices for each RWS.

Simulation of annual potential yield and temporal variability in potential yield due to year-to-year variation in weather using the ORYZA rice model requires daily weather records, including maximum and minimum temperature, wind speed, relative humidity, precipitation, and solar radiation. For irrigated cropping systems, 10 years of weather data are sufficient for robust estimation of weather-related temporal variation in potential yield<sup>34</sup>. We obtained daily weather data for 11 years (2004–2014) for all weather variables (except solar radiation) from the National Meteorological Information Center of the China Meteorological Administration. We obtained solar radiation data from the NASA-POWER database (<https://power.larc.nasa.gov/>). Previous studies have shown that crop yield simulations based on NASA-POWER solar radiation are in close agreement with simulations based on ground-measured radiation across a wide range of environments and regions<sup>37</sup>. Erroneous and/or missing information in weather data were screened following the quality control methods described in van Wart et al.<sup>34</sup> and Grassini et al.<sup>9</sup>.

Crop management data required for simulation of potential yield include sowing date, plant density (i.e., number of plants per hill and number of hills per m<sup>2</sup>), and phenological duration of dominant cultivars. For each RWS and associated buffer zone, we obtained information about the predominant rice cultivars and current yield from local experimental data, and publications reporting agronomic field research conducted at those sites. The selected RWS were grouped into six regions according to CZ and dominant cultivar characteristics following Duan et al.<sup>38,39</sup>, and management practices within the same region were assumed to be similar (Supplementary Fig. 2). For example, single-rice systems dominate in the northeast, north, central and southwest regions, while double-rice systems

dominate in the central and south regions. In each region, one widely planted rice cultivar was used for single-crop systems, while two cultivars were used for double-rice systems, one for the early and the other for the late cropping season. A large number of new rice cultivars are released every year and immediately planted by farmers; parameterizing and simulating all of them would have been unfeasible. In this study, we selected the dominant high-yield modern cultivars with broad adaptation and good tolerance to pathogens and insect pests within each region. These cultivars were selected based on the rice variety database (<http://www.ricedata.cn/variety/>), published research papers, and recommendations from regional rice experts in China.

Since ORYZA rice model was developed for tropical areas, the model was calibrated in the present study to account for the climate conditions and dominant cultivars grown in China. The crop model parameters for an elite rice cultivar in China named Huanghuazhan was carefully calibrated as a standard crop file using experimental data from a 2-year (2012–2013) experiment performed in Huaqiao township (30°06' N, 115°45' E), Wuxue county, Hubei province, China. Huanghuazhan is the most common inbred rice cultivar planted in central and south China and is been widely grown across 7 major rice-producing provinces (<http://www.ricedata.cn/variety/>) with high and stable yield, good quality, and wide adaptability. Although Huanghuazhan is an inbred cultivar, the yield is comparable with hybrid cultivars<sup>40</sup>. Sowing date in the experiment was May 10th, which is typical in central China, and the seedlings were transplanted 26–31 days after sowing. Plant density was 28 hills per m<sup>2</sup> with 2–3 plants per hill. Fertilizers were applied at rates of 195 kg N ha<sup>-1</sup>, 33 kg P ha<sup>-1</sup>, and 125 kg K ha<sup>-1</sup>. Fertilizer-N was split into several applications: 36% as basal, 23% at tillering, 23% at panicle initiation, and 18% 1 week after panicle initiation. Fertilizer-P was applied in one single basal application while fertilizer-K was applied 30% before sowing, 40% at tillering, and 40% 1 week after panicle initiation. Insects, diseases, and weeds were periodically controlled using pesticides to avoid biomass and grain yield losses. Leaf area index, and dry matter of leaves, stems, and panicles were measured at different phenological stages.

For the other cultivars, model coefficients were calibrated using phenology, yield and biomass data from studies published in recent years (after 2005) in which crops received near-optimal management, that is, with non-limiting water and nutrient supplied and crops kept free of pathogens, insect pests, and weeds. In those studies involving multiple treatments, we used the highest yield measured across treatments or average yield among treatments if the treatment effect was not significant. Despite our efforts to derive the same set of genetic coefficients for the same rice cultivar irrespective of site or season, it was not possible to portray differences in yield and phenology between late and early season for one of them in one region (Teyou 582 in south region in double-rice systems). Hence, separate coefficients were derived for the early and late season for this cultivar. A similar approach has been followed in calibrating ORYZA in previous studies to portray differences between contrasting environments<sup>41</sup>.

Cultivar-specific parameters were derived following standard procedures for the ORYZA rice model, and related information is provided in Supplementary Table 2. Parameterization of crop characteristics was obtained by two utility programs in ORYZA rice model called *drate(v2).exe* and *param(v2).exe*. The *drate(v2).exe* is a program to determine the phenology development rate of a given variety, and the *param(v2).exe* is a program to estimate crop parameters such as assimilate partitioning, specific leaf area, and non-structure C&N translocation, etc.<sup>12,36</sup>. In our study, program *drate(v2).exe* was used to determine the phenology development rate by using the phenological stages and growth duration of local dominant cultivars. Program *param(v2).exe* was used to estimate the fraction of dry matter partitioned to shoot, the shoot dry matter partitioned to leaves, stems, and panicles at different phenological stages, and the fraction of carbohydrates allocated to stems that is stored as reserves by using the measured dry matter of organs at different phenological stages of local dominant cultivars. See details of calibrated parameters in Supplementary Tables 3–4.

We calibrated and validated the model using independent datasets. We used experimental data from one site-year in each experiment for calibration, and used the other site-year(s) for validation (hereafter referred to as validation 1). We performed an additional validation (hereafter referred to as validation 2) using published yield, biomass and growth duration data for the cultivars included in the calibration to supplement validation 1. These data were not used for model calibration because of the lack of detailed information required for model calibration or because there was not enough information to ensure that the crops were managed to reach potential yield; hence, it was expected that a number of the observed yields reported in these studies would fall below the simulated potential yield. Details about experimental data for model calibration and validation can be found in Supplementary Tables 2–4 and Supplementary Fig. 3.

Degree of association and agreement between simulated and observed variables was assessed by the coefficient of determination ( $r^2$ ), root mean square error (RMSE), and RMSE expressed as percentage of the observed mean (RMSE<sub>n</sub>), which were calculated as follows:

$$r^2 = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]} \quad (1)$$

$$RMSE = \left[ \left( \sum (x - y)^2 / n \right)^{0.5} \right] \quad (2)$$

$$RMSE_n = \left[ \left( \sum (x - y)^2 / n \right)^{0.5} / M_{mean} \times 100\% \right] \quad (3)$$

where  $x$  and  $y$  represent the simulated and observed values, and  $n$  represents the number of paired values.  $r^2$  close to 1 and RMSE and  $RMSE_n$  close to 0 indicate a good agreement between simulated and observed values.

Observed grain yields across experiments ranged from 8.4 to 14.4 t ha<sup>-1</sup> and were in close agreement with simulated values after model calibration as indicated by relatively low RMSE 0.89 t ha<sup>-1</sup> (validation 1), which represented 9% of the mean observed yield (Supplementary Fig. 5). Likewise, validation 1 indicated close agreement between simulated and observed aboveground dry matter and growth duration, with low RMSE for shoot biomass (0.80 t ha<sup>-1</sup>) and growth duration (5 days). In the validation 2 database, yield, total biomass, and growth duration ranged from 6.5–11.6 t ha<sup>-1</sup>, 10.8–20.7 t ha<sup>-1</sup>, and 106–185 days, respectively. Results in validation 2 also showed reasonable agreement between simulated yield and observed yield, with RMSE representing 15% and 16% of the mean observed yield and biomass, respectively. As expected, a number of observed yields were above the 1 to 1 line (Supplementary Fig. 5c). Growth duration was in close agreement as indicated by the low RMSE of 5%. To summarize, reasonable agreement between simulated and observed yields using two different datasets gives confidence that the calibrated ORYZA rice model is robust at reproducing potential yield across the wide range of climates and rice cropping systems in China.

We performed a review of previous efforts in calibrating and/or validating rice simulation models in studies aiming to estimate rice potential yield at regional and national levels (Supplementary Table 5). In most of those studies, model coefficients were neither calibrated nor validated using local experimental data; instead, potential yield was simulated using default coefficients derived somewhere else or published in the literature. In some studies, the model was calibrated and/or evaluated only for one or few varieties and/or using data from few site-years. An exception was two studies using the database from the Agrometeorological Experimental Stations of the Chinese Meteorological Administration (CMA)<sup>26,42</sup>. The CMA database has advantages given its large number of sites and information on rice crop phenology and yield. However, the data are collected from farmer fields that do not receive management conducive to expression of potential yield. Likewise, the database only has data on rice phenology and yield at harvest time, which are insufficient for a robust model calibration and/or evaluation. Hence, the detailed model calibration and evaluation performed for a large number of cultivars across a wide range of environments in China is a clear strength of our study compared with previous efforts aiming to estimate potential yield at regional and national levels.

All data sources and uncertainties, and associated quality control measures to mitigate possible biases, are listed in Supplementary Table 6. In all cases, data sources fall in the so-called tier 1 of data availability/quality for crop yield gap analysis described in Grassini et al.<sup>9</sup>.

**Current farm yield and yield gaps.** Available yield data from the 5 most recent years (2010–2014) were retrieved for each of the counties that overlap with the selected RWS from national and provincial statistical bureaus. However, county-level farm yield data are not accurate for a number of reasons, and the magnitude of inaccuracy varies across counties and is difficult to predict<sup>43,44</sup>. In contrast, provincial level farm yield data are more reliable and accurate because a combination of different methods are used, including remote sensing and ground truthing<sup>45</sup>. Hence, using county-level data without adjustment to be consistent with the provincial-level yield data gives inaccurate yield gap estimates when results are aggregated to larger spatial scales. To adjust county-level yield data in a province, we increased or decreased farm yields by an equivalent percentage so that the weighted average farm yields across all RWS within that province equaled the provincial official average yield of 2010–2014. However, we did not adjust farm yield of counties in provinces for which there was little difference (within ±5%) between the official provincial farm yield and the upscaled provincial farm yields following GYGA protocols. Average current yield was calculated as the average year over the 5-year (2010–2014) time period to represent yield in the current season.

The dominant cropping system (single- or double-rice) was identified for each RWS buffer zone and used as the basis for simulating potential yields and for estimating yield gaps. In some parts of central China and in northern provinces, only a single-rice crop is grown each year because the growing season is too short for double-rice. Hence, the yield gap for each RWS where a single-rice crop is grown was calculated as the difference between the single-rice potential yield and the current farm yield. In south and some parts of central China, farmers practice a double-rice cropping system. However, county-level data provide only the average yield for the two crops. But the area of early- and late-season rice crops is almost identical at 19% and 20% of total rice area, respectively, based on the estimates from ref. 1. We therefore simulated yields of early- and late-rice crops separately, and used the average potential yield of the two crops to calculate yield gaps for each RWS where double-rice systems dominate.

To upscale potential yield, current farm yield, and yield gap estimates from RWS to larger spatial scales, weighted averages for each variable were calculated by

the proportional contribution of rice area within each spatial unit contributing to the spatially aggregated value at the CZ or national scale<sup>10</sup>.

**Estimating rice production potential.** Current farm yields tend to stagnate when they reach 75–80% of potential yield (called the exploitable yield ceiling) due to diminishing returns from investment in additional production inputs and effort as yields approach the potential yield ceiling<sup>13,15,34</sup>. Hence, prospects for increasing rice production at any spatial scale are best indicated by the exploitable portion of the yield gap, which is the difference between current farm yields and 75% or 80% of potential yields estimated at the CZ and national scales. All else equal, achieving yields that are 80% of potential yield requires greater input of nutrients and more aggressive pest control measures than production at 75% of potential yield. Given concerns about the substantial environmental pollution associated with current intensive crop production practices<sup>19</sup>, scenarios with either an 80% or 75% exploitable yield ceiling were evaluated. Additional exploitable production potential of single- versus double-rice systems was estimated by the difference between 75% or 80% of potential yield and current farm yields for each rice cropping system over the total national production area.

**Future scenarios.** Rice production scenarios to 2030 were evaluated based on the following assumptions:

- (i) Yield would stagnate at current levels to 2030 or follow current trajectories based on regression of national rice yields versus year for single- and double-rice (Supplementary Fig. 6).
- (ii) Amount of cropland devoted to rice production remains unchanged at the level of 2011–2014 average, which is consistent with current trajectories based on the fitted trends of observed harvested area (Supplementary Fig. 6). We plotted harvested area and yield against year (1985–2014) for different rice cropping systems by linear function or by a linear-segment piecewise function using SigmaPlot 10.0 (Systat. Software, Inc., San Jose, CA, USA), as shown in Supplementary Fig. 6.

Linear function:

$$y = at + b \quad (4)$$

where  $t$  is year,  $y$  is single-rice harvested area or single-, double-rice yield.

Two-linear-segments piecewise function:

$$t_1 = \min(t), \text{ which is the year 1985} \quad (5)$$

$$t_2 = \max(t), \text{ which is the year 2014} \quad (6)$$

$$y = \begin{cases} \frac{a(t_1 - t) + b(t - t_1)}{t_1 - t_1}, & t_1 \leq t \leq T_1 \\ \frac{b(t_2 - t) + c(t - T_1)}{t_2 - T_1}, & T_1 \leq t \leq t_2 \end{cases} \quad (7)$$

where  $t$  is year,  $y$  is total- or double-rice harvested area, and  $T_1$  is the breakpoint year. The  $r^2$  for the linear regressions of total-, single-, and double-rice harvested area are 0.73, 0.88, and 0.93, and the  $r^2$  for the linear regressions of single- and double-rice yield are 0.78 and 0.71. All estimated parameters were statistically significant (Student's  $t$ -test,  $P < 0.0001$ ).

The linear regression for single-rice yield:

$$y = 0.0488t - 90.788 \quad (8)$$

The linear regression for double-rice yield:

$$y = 0.0263t - 47.255 \quad (9)$$

- (iii) While a similar regression for single-rice area shows a continued linear increase through 2015, disaggregation of these data by province shows that most of the recent increase occurred in Heilongjiang province, and to a lesser extent in Jilin province while rice production area in remaining single-rice provinces has remained constant or even decreased in recent years (Supplementary Fig. 7). Moreover, recent government policies have advocated for reduced rice area in Heilongjiang due to limited water resources.
- (iv) Total rice demand in 2030 is 217 MMT from the average of three recent studies<sup>16–18</sup>, which is 5.3% higher compared with current rice production of 206 MMT (average of 2013–2015 from NBSC<sup>1</sup> to represent current rice production).
- (v) An exploitable yield ceiling of either 80% or 75% of potential rice yield.

**Comparison of studies assessing rice production scenarios.** We cross-validated our scenario assessment results by comparing them with results from recent studies assessing future rice for China at country level (Supplementary Table 7). The goal of some of the studies was not to estimate rice self-sufficiency per se but they reported data that allowed us to assess it. These studies followed diverse approaches as basis for their assessments. We found that the results from our scenario assessment focusing on rice systems and regions fall within the range of projected self-sufficiency ratios with scenarios considering other factors e.g., socio economic,

technology development, and climate change etc., which adds confidence on the results reported here.

**Comparison of yield differences by GAEZ and GYGA methods.** Global Agro-Ecological Zones Model version 3.0 (GAEZ v3.0)<sup>23</sup> was used for comparison of yield-gap analyses. Compared to the GYGA up-scaling method, a down-scaling method is used in GAEZ v3.0 such that both weather and agricultural crop production data (e.g., current yields and crop production area) from much coarser spatial scales are interpolated into 5 arc-min grid cells (roughly 100 km<sup>2</sup>). National estimates are then estimated by aggregation of data from all grids in which there is rice production area. In contrast to our GYGA approach, which utilized provincial-level data on proportion of rice area under either single- or double-rice systems to estimate area of each system within RWS station buffer zones, GAEZ assigned one rice cropping system to each 5 arc-min grid cell by matching growth cycle and temperature requirements of rice with length of time available for crop growth<sup>23</sup>.

In GAEZ the “Agro-climatic exploitable yields with high input level” data layer is defined as potential climatic yield with optimal management practices ( $Y_{p,GAEZ}$ ), which is a proxy for potential yield as estimated using GYGA methodology ( $Y_{p,GYGA}$ ). The map of Agro-climatic exploitable yields with high input level for irrigated rice (map) was downloaded on May 26, 2018 from GAEZ website: <http://gaez.fao.org/>.

We calculated  $Y_{p,GAEZ}$  at CZ, province and country levels and compared them with  $Y_{p,GYGA}$  at those same levels of spatial upscaling using the following approach. First, each GAEZ grid cell (5 arc-min) in  $Y_{p,GYGA}$  map was superimposed with the irrigated rice harvested area from the SPAM map<sup>33</sup> to assign a harvested rice area to each grid cell. Second, each grid cell was assigned to a GYGA CZ based on the CZ map. The two processes were performed in ArcGIS 10.2. Weighted  $Y_p$  values based on harvested area were then estimated at the CZ, province and country level. Potential yield in GAEZ was calculated:

$$Y_p = \frac{\sum_{i=1}^n Y_{pi} \times A_i}{\sum_{i=1}^n A_i} \quad (10)$$

where  $Y_p$  is potential yield, and  $A$  is harvested area. For  $Y_p$  at CZ level,  $i$  is a grid cell and  $n$  is the number of grid cells within a CZ. For  $Y_p$  at province level,  $i$  is a grid cell and  $n$  is the number of grid cells within a province. For single-rice  $Y_p$ ,  $i$  is a CZ and  $n$  is the number of CZ within single-rice system. For double-rice  $Y_p$ ,  $i$  is a CZ and  $n$  is the number of CZ within double-rice system. For  $Y_p$  at country level,  $i$  is a cropping system and  $n$  is the number of cropping systems within a country.  $Y_g$  in GAEZ ( $Y_a$  as a percent of  $Y_p$ ) was calculated:

$$Y_{g,GAEZ}(\%) = \frac{Y_{a,GYGA}}{Y_{p,GAEZ}} \times 100\% \quad (11)$$

where  $Y_a$  in GYGA was used to calculate  $Y_g$  for GAEZ because the  $Y_a$  in GAEZ is outdated (from the year 2000).

## Data availability

All data generated or analyzed during this study are available within this paper and its Supplementary Information files. The source data underlying Table 1, Figs. 1–3, Supplementary Figs. 1–8, and Supplementary Table 8 are provided as a Source Data file. A reporting summary for this article is available as a Supplementary Information file.

Received: 19 August 2018 Accepted: 11 March 2019

Published online: 12 April 2019

## References

- NBSC. *National Bureau of Statistics of China, China Statistical Yearbook*. (China Statistics Press, Beijing, China, 1980–2016).
- FAOSTAT. *Statistical Yearbook. Asia and Pacific Food and Agriculture*. (FAO, Bangkok, Thailand, 2015).
- USDA Foreign Agriculture Service. Grain: World Markets and Trades. <https://www.fas.usda.gov/data/grain-world-markets-and-trade> (2017).
- Grassini, P., Eskridge, K. M. & Cassman, K. G. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* **4**, 2918 (2013).
- Wang, L. et al. Spatial and temporal changes of arable land driven by urbanization and ecological restoration in China. *Chin. Geogr. Sci.* <https://doi.org/10.1007/s11769-018-0983-1> (2018).
- Yan, T., Wang, J. & Huang, J. Urbanization, agricultural water use, and regional and national crop production in China. *Ecol. Model.* **318**, 226–235 (2015).
- van Ittersum, M. K. et al. Yield gap analysis with local to global relevance—a review. *Field Crops Res.* **143**, 4–17 (2013).
- Evans, L. T. *Crop Evolution, Adaptation and Yield* (Cambridge University Press, Cambridge, United Kingdom and New York, USA, 1996).
- Grassini, P. et al. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Res.* **177**, 49–63 (2015).
- van Bussel, L. G. J. et al. From field to atlas: upscaling of location-specific yield gap estimates. *Field Crops Res.* **177**, 98–108 (2015).
- Chen, F., Sushil, P. & Ding, S. Changing rice cropping patterns: evidence from the Yangtze River Valley, China. *Outlook Agric.* **42**, 109–115 (2013).
- Li, T. et al. From ORYZA2000 to ORYZA (v3): an improved simulation model for rice in drought and nitrogen-deficient environments. *Agric. For. Meteorol.* **237**, 246–256 (2017).
- Cassman, K. G., Dobermann, A., Daniel, T. W. & Yang, H. Meeting cereal demand while protecting natural resources and improving environmental quality. *Annu. Rev. Environ. Resour.* **28**, 315–358 (2003).
- Grassini, P., Specht, J. E., Tollenaar, M., Ciampitti, I. & Cassman, K. G. *Crop Physiology (2nd Ed.) Chapter 2. High-yield Maize-Soybean Cropping Systems in the US Corn Belt* (Elsevier Inc., 2015).
- Espe, M. B. et al. Yield gap analysis of US rice production systems shows opportunities for improvement. *Field Crops Res.* **196**, 276–283 (2016).
- Chen, X. et al. Producing more grain with lower environmental costs. *Nature* **514**, 486 (2014).
- Alexandratos, N. & Bruinsma, J. World agriculture towards 2030/2050: the 2012 revision. In *ESA Working Paper No. 12-03* (FAO, Rome, 2012).
- Robinson, S. et al. *The International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT): Model Description for Version 3* (2015).
- Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671 (2002).
- Licker, R. et al. Mind the gap: how do climate and agricultural management explain the ‘yield gap’ of croplands around the world? *Glob. Ecol. Biogeogr.* **19**, 769–782 (2010).
- Mausser, W. et al. Global biomass production potentials exceed expected future demand without the need for cropland expansion. *Nat. Commun.* **6**, 8946 (2015).
- Costinot, A., Donaldson, D. & Smith, C. Evolving comparative advantage and the impact of climate change in agricultural markets: evidence from 1.7 million fields around the world. *J. Political Econ.* **124**, 205–248 (2016).
- IIASA/FAO. *Global Agro-ecological Zones (GAEZ v3.0)* (IIASA, Luxembourg, Austria and FAO, Rome, Italy, 2012).
- Li, Y. et al. Integrated assessment of China’s agricultural vulnerability to climate change: a multi-indicator approach. *Clim. Change* **128**, 355–366 (2015).
- Zhang, H. et al. Contributions of climate, varieties, and agronomic management to rice yield change in the past three decades in China. *Front. Earth Sci.* **10**, 315–327 (2016).
- Lv, Z. et al. Climate change impacts on regional rice production in China. *Clim. Change* **147**, 523–537 (2018).
- Challinor, A. J. et al. A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Change* **4**, 287–291 (2014).
- IPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2014).
- Anderson, K. *Agricultural Trade, Policy Reforms, and Global Food Security* (Springer, New York, USA, 2016).
- Clapp, J. Food self-sufficiency: making sense of it, and when it makes sense. *Food Policy* **66**, 88–96 (2017).
- Hochman, Z., Gobbett, D., Horan, H. & Garcia, J. N. Data rich yield gap analysis of wheat in Australia. *Field Crops Res.* **197**, 97–106 (2016).
- Van Wart, J. et al. Use of agro-climatic zones to upscale simulated crop yield potential. *Field Crops Res.* **143**, 44–55 (2013).
- You, L. et al. Spatial Production Allocation Model (SPAM) 2005 v2.0. Available from <http://mapspam.info> (2014).
- van Wart, J., Kersebaum, K. C., Peng, S., Milner, M. & Cassman, K. G. Estimating crop yield potential at regional to national scales. *Field Crops Res.* **143**, 34–43 (2013).
- van Oort, P. A. J. et al. Assessment of rice self-sufficiency in 2025 in eight African countries. *Glob. Food Sec.* **5**, 39–49 (2015).
- Bouman, B. A. M. et al. *ORYZA2000: Modeling Lowland Rice*. (International Rice Research Institute and Wageningen University and Research Centre, Los Baños, Philippines, 2001).
- van Wart, J., Grassini, P. & Cassman, K. G. Impact of derived global weather data on simulated crop yields. *Glob. Change Biol.* **19**, 3822–3834 (2013).
- Duan, J. & Zhou, G. Climatic suitability of single cropping rice planting region in China. *Chin. J. Appl. Ecol.* **23**, 426–432 (2012).
- Duan, J. & Zhou, G. Dynamics of decadal changes in the distribution of double-cropping rice cultivation. *Chin. Sci. Bull.* **58**, 1955–1963 (2013).
- Zhou, S. C. et al. Breeding and application of Huanghuazhan, a new early, middle or late rice variety with good quality and wide adaptability. *J. Agric. Sci. Technol.* **10**, 77–83 (2008).



41. Cao, B. et al. Evaluation of ORYZA2000 for simulating rice growth of different genotypes at two latitudes. *Agron. J.* **109**, 2613–2629 (2017).
42. Zhang, T., Yang, X., Wang, H., Li, Y. & Ye, Q. Climatic and technological ceilings for Chinese rice stagnation based on yield gaps and yield trend pattern analysis. *Glob. Change Biol.* **20**, 1289–1298 (2014).
43. Okamoto, K. & Kawashima, H. Estimating total area of paddy fields in Heilongjiang, China around 2000 using Landsat Thematic Mapper/Enhanced Thematic Mapper Plus data. *Remote Sens. Lett.* **7**, 533–540 (2016).
44. Jin, C., Xiao, X., Dong, J., Qin, Y. & Wang, Z. Mapping paddy rice distribution using multi-temporal Landsat imagery in the Sanjiang Plain, northeast China. *Front. Earth Sci.* **10**, 49–62 (2016).
45. Liu, J. et al. Spatial and temporal patterns of China's cropland during 1990–2000: an analysis based on Landsat TM data. *Remote Sens. Environ.* **98**, 442–456 (2005).

## Acknowledgements

The authors would like to thank Dr. David Lobell for commenting on an earlier version of the paper and suggesting the comparison of bottom-up versus top-down yield-gap analysis approaches. This research was supported by the earmarked fund for the National Key Research and Development Program of China (No. 2016YFD0300210), China Agriculture Research System (CARS-01-20), the Program of Introducing Talents of Discipline to Universities in China (the 111 Project no. B14032), the Program for Changjiang Scholars and Innovative Research Team in University (IRT1247), and the Daugherty Water for Food Global Institute at University of Nebraska-Lincoln.

## Author contributions

All authors conceived the study and wrote the paper. N.D., H.Y. and P.G. performed the statistical analysis and simulations. All authors contributed to editing the paper.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-09447-9>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks Ando Radanielson and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019