Supplementary Information for

Spatial feasibility of large-scale farming for sustainable agriculture in China

This PDF file includes:

Supplementary Methods

Data sources

The data we used can be divided into two parts: spatial data and statistical data. The first part consists of two maps, one is for analysis of the current situation, the other for predicting and analyzing the effect of scale-farming.

(1) Lesiv et al used crowdsourcing to estimate the global distribution of field size [1]. They employed a large volume of volunteers to classify the area of sample sites. In China, there were 5,421 sample sites (Figure S3). We used these sites to calculate the average current field size for each county.

(2) Global land cover data are key sources of information for understanding the complex interactions between human activities and global change. FROM-GLC (Finer Resolution Observation and Monitoring of Global Land Cover) is the first 30 m × 30 m resolution global land cover map produced using Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) data. This series of data are produced by Gong et al and are available at http://data.ess.tsinghua.edu.cn/. We used FROM-GLC30 2017v1 to produce the scale farming map

and the share of area [2].

In addition, we also used data from the following sources:

(3) Statistical Yearbook in 2017. We collected local statistical yearbooks of all provinces and cities in China for 2017, and established a county-level database by compiling socio-economic data, agricultural data, and natural resource data of each county. All local statistical yearbooks are available at http://data.cnki.net/yearbook/. The database mainly includes population, sowing area, agricultural input and output, and fertilizer use. This data provides important support for our analysis of nitrogen input and yield differences at the county level.

(4) Nitrogen deposition map. Zhang et al provided a satellite-based, national assessment of wet and dry N_r deposition, constrained with national measurement [3]. We used it in the CHANS model for N deposition calculation.

(5) Consolidation cost. The cost of consolidation is from the website of China Land Consolidation and Rehabilitation (http://www.lcrc.org.cn/tdzzgz/zxgz/zdgcysfjs/). We collected data from 201 projects in 33 provinces and used this data to calculate average value for four different categories (divided by income and terrain). Details are showed in Table S2.

(6) China Rural Household Panel Survey (CRHPS). We also use China Rural Household Panel Survey (CRHPS). It was used to establish the relations between agricultural input and output and farm size. The CRHPS is a nationally representative survey covering all provinces except Xinjiang and Tibet. The original rural household data include 24,764 households that are registered as agricultural residents. These households consist of 77,132 individuals from 1,439 residential committees and villages, located in 363 selected counties in China. The CRHPS data are open to all researchers free of charge and full access to all data at http://ssec.zju.edu.cn/dataset/CRHPS/.

The survey collected information on household demographic features, agricultural and non-agricultural activities, and household income from these activities. It also collected information on the residential committees and villages where the households resided. A detailed run-down of all the variables used in our paper is provided in a later section.

(7) China Agricultural Yearbook 2017. China Agricultural Yearbook is reference book reflecting agriculture, forestry, animal husbandry and fishery in China. The data is provided by the National Bureau of Statistics, the Ministry of Agriculture and Rural Affairs, the National Forestry Bureau, the National Grassland Bureau and other relevant departments of China. We used the number of agricultural cost and output for every province.

(8) The third National Agricultural Census (NAC). The NAC was conducted by the National Bureau of Statistics of China. It is a decennial census that collects information of a nationally representative sample of rural households, as well as village and township governments, etc. We used the total number of agricultural labors.

Variables from the CRHPS 2015 to 2017

Farm size

Farm size refers to the operating farm land area including household's gross contracted land and transferred area subtracting idle land. Household's contractual area refers to the area of tenured arable land according to the rural Household Contract Responsibility System (HCRS). The area unit is hectare (ha).

Agricultural Labor

Agricultural labor per ha is the total number of farming labors divided by farm size. The labor includes family members, relatives and employees who work during busy and non-busy seasons. We weighted agricultural labors according to their working time. We first calculated labor hours according to labors' working during busy and non-busy seasons. We hypothesize that each labor works 8 hours a day, and long-term labor and family members engaged in agriculture work an average of 5 days a week. While assuming seven-fifths of the 365 days for long-term and family labors fully engaged in agriculture, we also weighted their working time with 80% considering non-busy seasons, part-time jobs and other situations. Therefore, total working hours of all labors

are calculated. We then deduct the number of agricultural labors using total working hours divided by the working hour of each long-term labor.

Agricultural Labor Cost

Agricultural labor cost ($) is the summarization of expenditures on employed labors and converted family labor cost. Expenditures on employed labors are directly reported in interviewees. And family labor cost is based on average county-level salaries of employed labors multiplying by family labor time inputs during busy and non-busy seasons, respectively.

Labor Productivity

Labor Productivity is the household agricultural gross income ($) divided by labor hours. The gross income is total market value of all crop yields directly reported by farmers.

Fertilizer

Fertilizer refers to chemical fertilizer purchases ($) divided by farm size.

Manure

Manure refers to expenditures on organic fertilizer and manure ($) divided by farm size.

Manure Ratio

Manure ratio is defined as the ratio between manure to the total chemical fertilizer input.

Cost

Cost is the total immediate input ($) per ha during farming. It includes all purchase of agricultural products such as seed and fertilizer, land transferred-in cost, machinery rental fee, depreciation of own machinery and labor input including both household labor and employment

labor.

Output

Output ($ per ha) is total market value of all crop yields directly reported by farmers including all grains and crash crops.

Profit

Agricultural profit ($ per ha) equals to the difference between total agricultural output and cost. In view of we considering converted cost of household labor and the depreciation cost of their own machinery, the net profit is mostly negative. In order to better observe the relationship between log-transformed profit and farm size, we add the absolute value of the minimum profit to all values to make them positive before log-transformed. The adjusted values are used for regression analysis with farm size to observe how much percentage changes of profit when the farm size changes by 1%.

Rent

The rent is the average of transferred-in cost and transferred-out income. The value is all divided by transfer area to get the number for per hectare. And the former was 612.7 Yuan (RMB), the latter was 711.5 Yuan (RMB). Based on the exchange rate of 1:7, we can get that the average rent was 1418.8 $ ha^{-1}.

Supplementary Methods for Fig. 1

The current map (a) was directly extracted from map of dominant field sizes (dominant_field_size_categories.tif) by administrative boundary of China. We used the table of estimated dominant field sizes at each location to derive the average field size of every county in China. There were 5,421 sites in China (Fig. S3). We updated the data to the cropland area shown

in (b). Following the calculation, we made a point-grid with an interval of 1 km. At each point, the k nearest neighbors method was applied to the given point and we set k=5 as the recommendation to get the value for unknown points. The data analysis was done in the R environment. The following R packages were used: raster (https://CRAN. R-project.org/package=raster); RANN (https://CRAN.R-project. org/package=RANN); and sp (https://CRAN.R-project.org/package=sp). The area proportions were from their estimation (Details see Fig. S4).

The change is the value for large-scale farming minus current figure. Here, we only show the part where the result is positive. And the current field share in (d) is directly from the paper of Lesiv et al [1]. While the value for scale farming is calculated using 'Analysis Toolbox'. The legend is same as (a) and (b). Details see Fig. S4. We also calculated the change of average field size of each county for following calculations.

Supplementary Methods for Fig. 4

Current data is from China Agricultural Yearbook 2017. According to the define of every index, we collected total agricultural labor in China, agricultural cost and profit for every province. Then we assumed that these data only related to farm size and distributed these data to every county according to the average field size. We calculated agricultural labor, labor income, agricultural cost, agricultural output and agricultural profit for the average field size of each county.

The predicted calculation is based on current values and changes in the field size showed in Fig. 1d and according to relations between farm size and agricultural input and output in China (See Table 1). Here we use the change of labor productivity to represent the change of labor income. The changes are the predicted value minus the current one. We only show the figures of agricultural labor, labor productivity and agricultural cost in the body of the paper. The figures for labor output and labor profit see Fig. S10. We didn't show it because the change is little.

References:

1 Lesiv, M. et al., Estimating the global distribution of field size using crowdsourcing. GLOBAL CHANGE BIOL 25 174 (2019).

2 Yu, L. et al., Using a global reference sample set and a cropland map for area estimation in China. SCI CHINA EARTH SCI 60 277 (2017).

3 Liu, L., Zhang, X., Xu, W., Liu, X. & Wu, X., Fall of oxidized while rise of reduced reactive nitrogen deposition in China. J CLEAN PROD 272 122875 (2020).

4 Zhang, F., Chen, X. & Chen, Q., Guidelines for major crop fertilization in China. (China Agriculture Press, 2009).

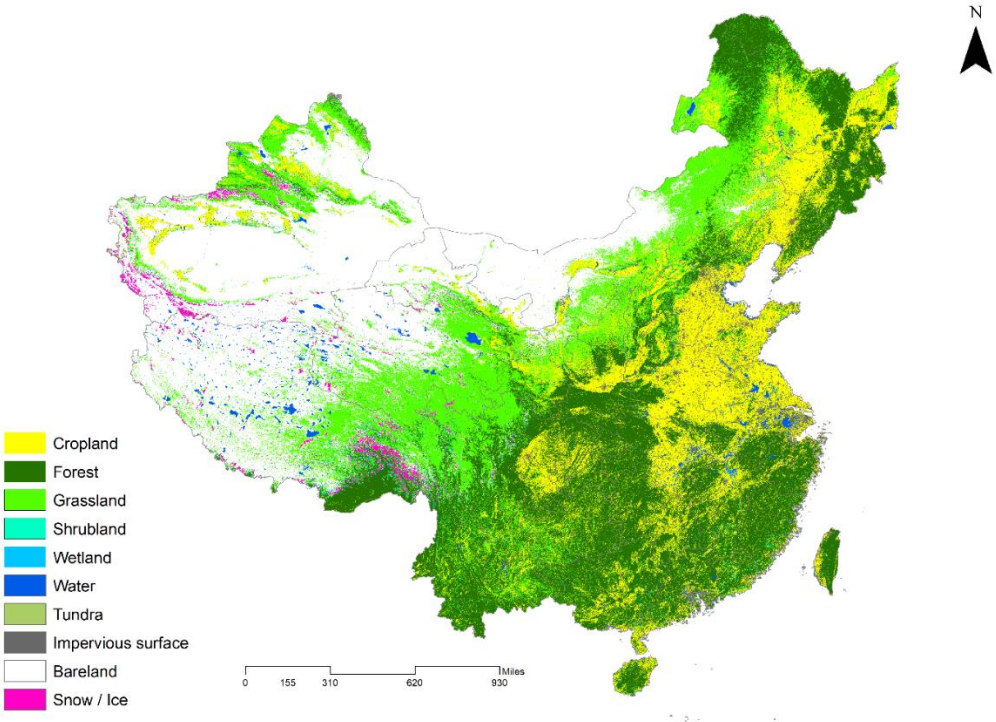

Fig. S1 China Land use (2017). This map is derived from FROM-GLC 2017v1 [2]. It shows the land use of China in 2017. There are 10 types of land, namely cropland, forest, grassland, shrubland, wetland, water, tundra, impervious surface, bareland and snow/ice. We extract cropland from this map for our analysis.

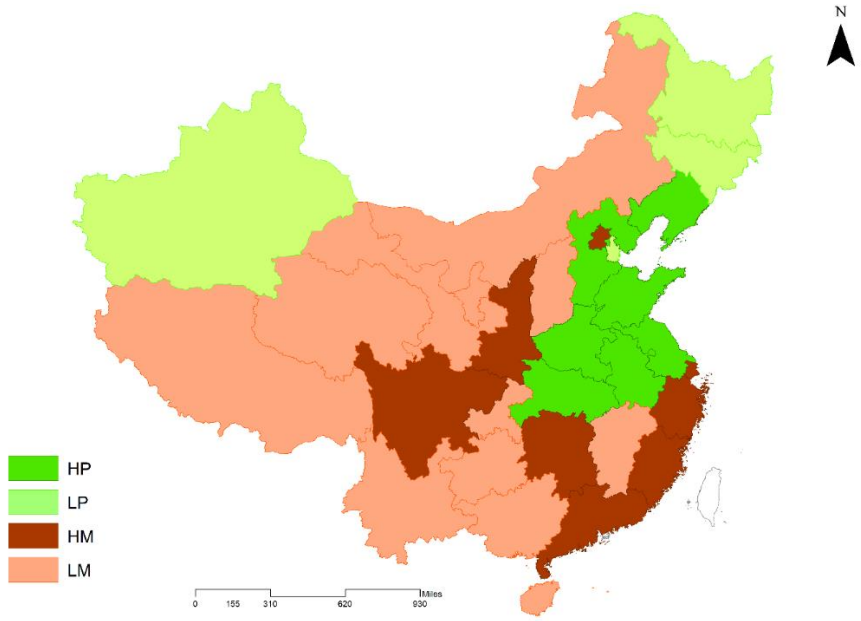

Fig. S2 Categories of regions. We divide the country's provinces into four categories according to terrain and local economic conditions. HP refers to high-income plain region. LP refers to low-income plain region. HM represents for high-income mountainous region. LM represents for low-income mountainous region.

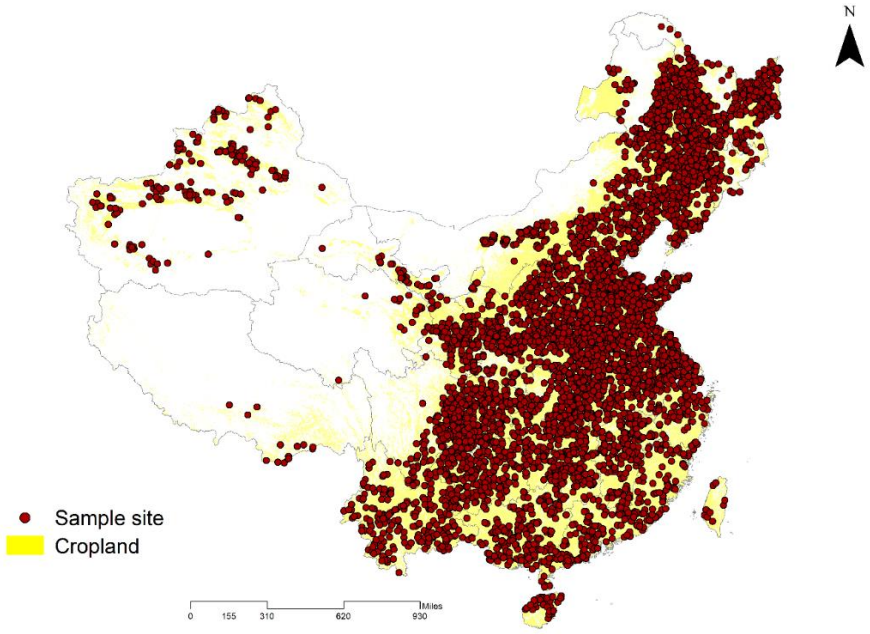

Fig. S3 Distribution of sample sites. The sample sites for field size are from the table of dominant field size provided by Lesiv et al [1]. There are 5421 sites, detailed data can be downloaded at http://pure.iiasa.ac.at/id/eprint/15526/. And the data was transferred to point shapefile by ArcGIS 10.2. Yellow area is cropland.

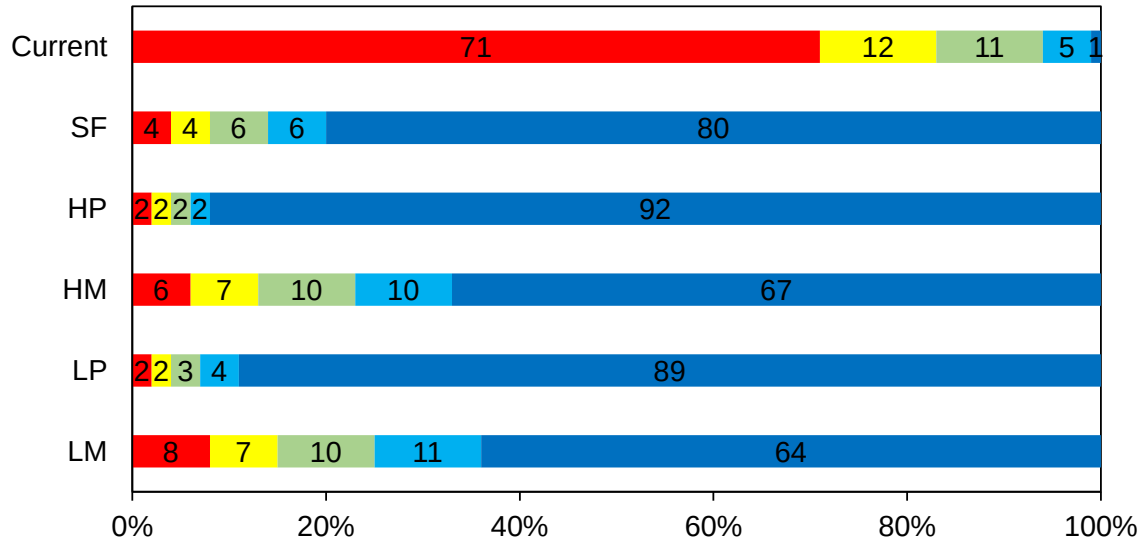

Fig. S4 Field size share in different regions. This figure shows the percentage of different field size in the four regions mentioned above. And SF refers to scale farming. The color is consistent with Fig. 1. The red color represents for field which is less than 0.6 hectare (ha), yellow for 0.6–2.6 ha, green for 2.6–16 ha, light blue for 16–100 ha and dark blue for field larger than 100 ha.

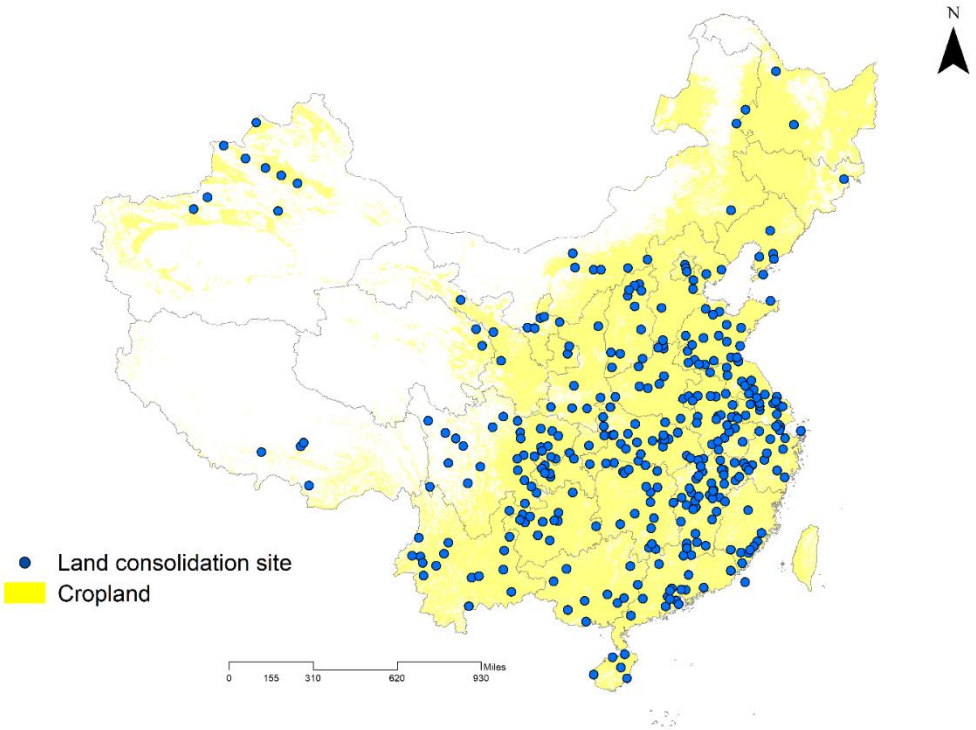

Fig. S5 Land consolidation sites. We collected land consolidation data from the website. It shows the distribution of land consolidation projects that almost cover all of China's provinces.

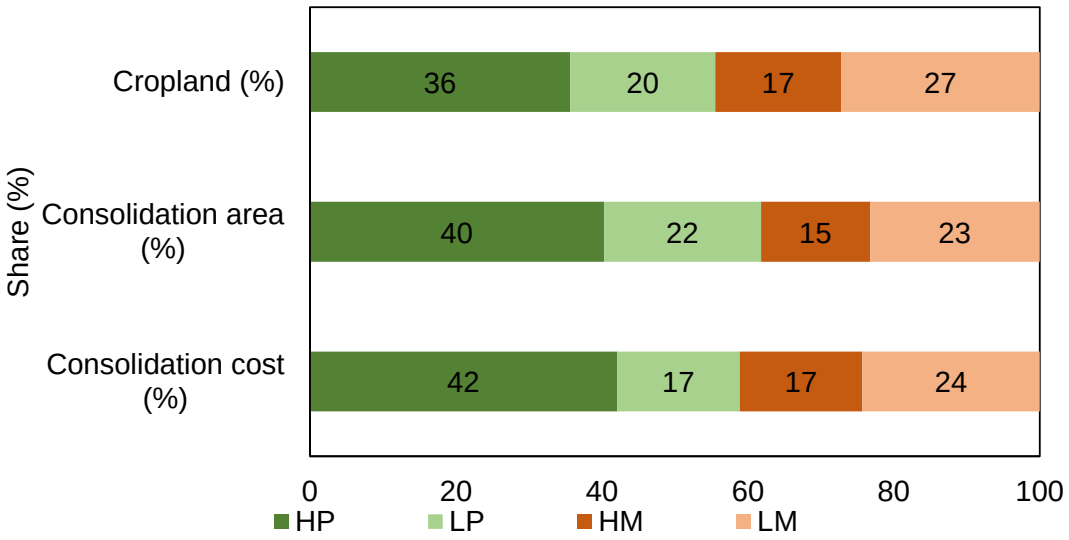

Fig. S6 Consolidation cost share. The cropland share of every region is calculated from cropland map. We calculated the proportion change based on Fig. S4 and the cropland area to get the consolidation area. Consolidation cost was calculated by cost per hectare and consolidation area. Here we only show the share of each region, details see Table S3.

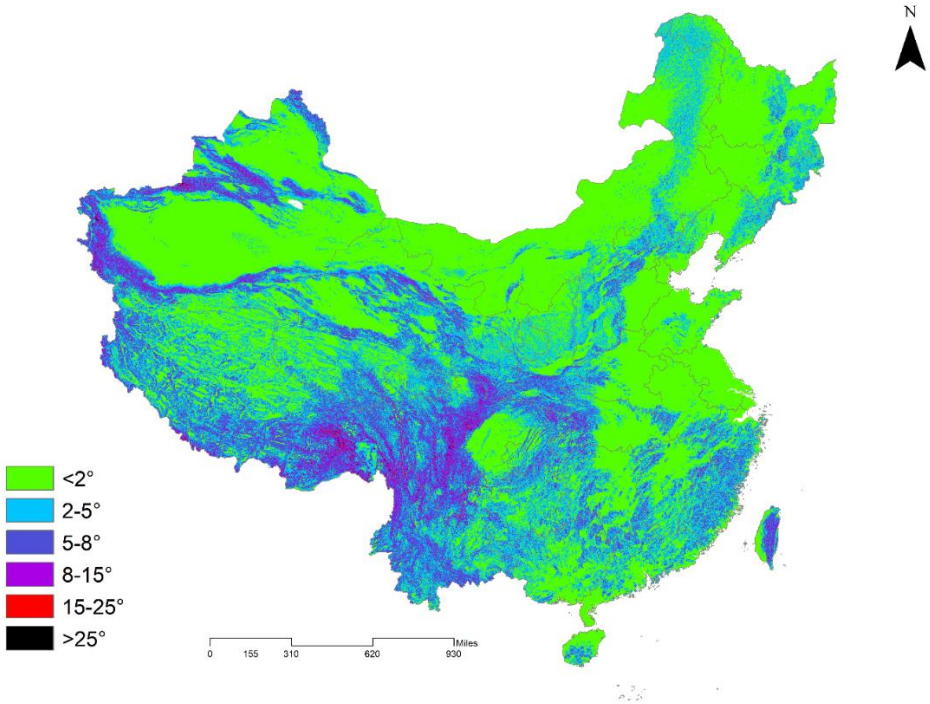

Fig. S7 Slope of China. The slope of China is range from 0 to 45 degrees. And we divided it into 6 levels, namely <2, 2–5, 5–8, 8-15, 15-25 and >25 degrees. It can be seen that most of the land is less than 8 degrees while great slopes located mainly in southwest region.

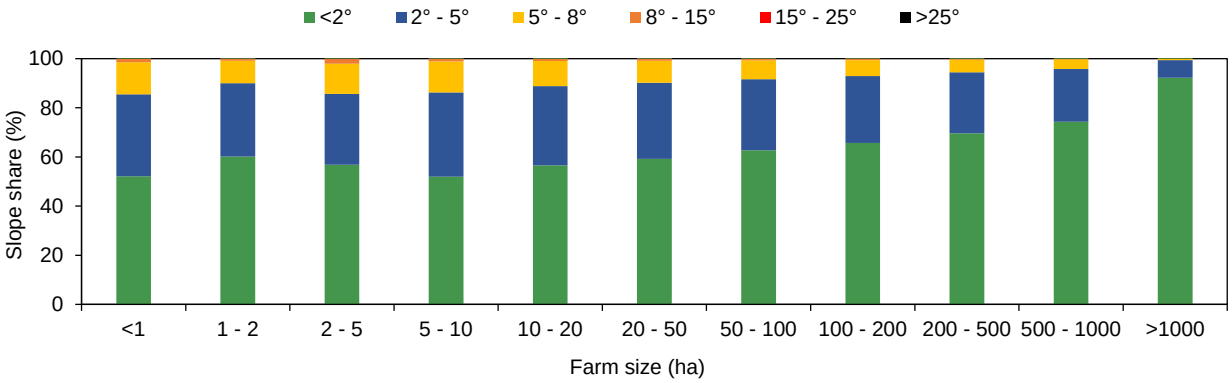

Fig. S8 Slope share of different field size. We choose slope to reflect the quality of land. And in this bar charts, we divided the arable land into 11 groups. The slope classification is according to "Regulation for gradation on agriculture land quality" of China. It is divided into 6 levels, namely <2, 2–5, 5–8, 8-15, 15-25 and >25 degrees, respectively. Here we didn't show the last class because it's little. As the increase of field size, the share of first slope class is increasing, too. It shows the rise in the quality of arable land.

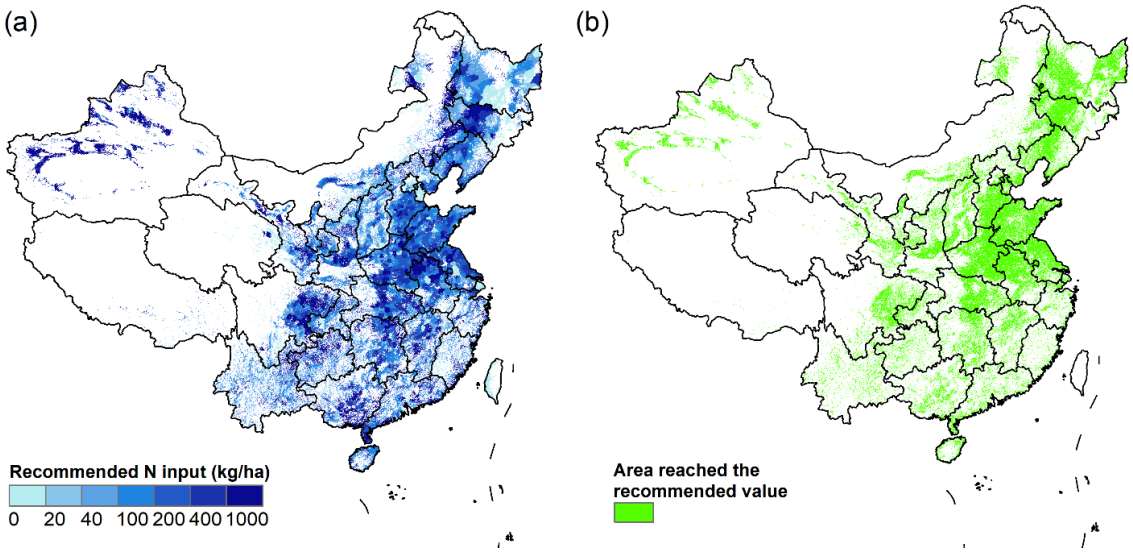

Fig. S9 Recommended N input. We use sowing area and recommended N fertilizer (Details see Table S8) for crops (rice, wheat, corn, millet, sorghum, barley, beans, potato, peanut, rapeseeds, cotton, hemp, tobacco, sugar beet, sugar cane, vegetable, fruits) to calculate the recommended N input for each county. And we compared this value with N fertilizer input for large-scale farming. The green area which occupied 74% cropland in (b) is the area where N input reached the recommended value.

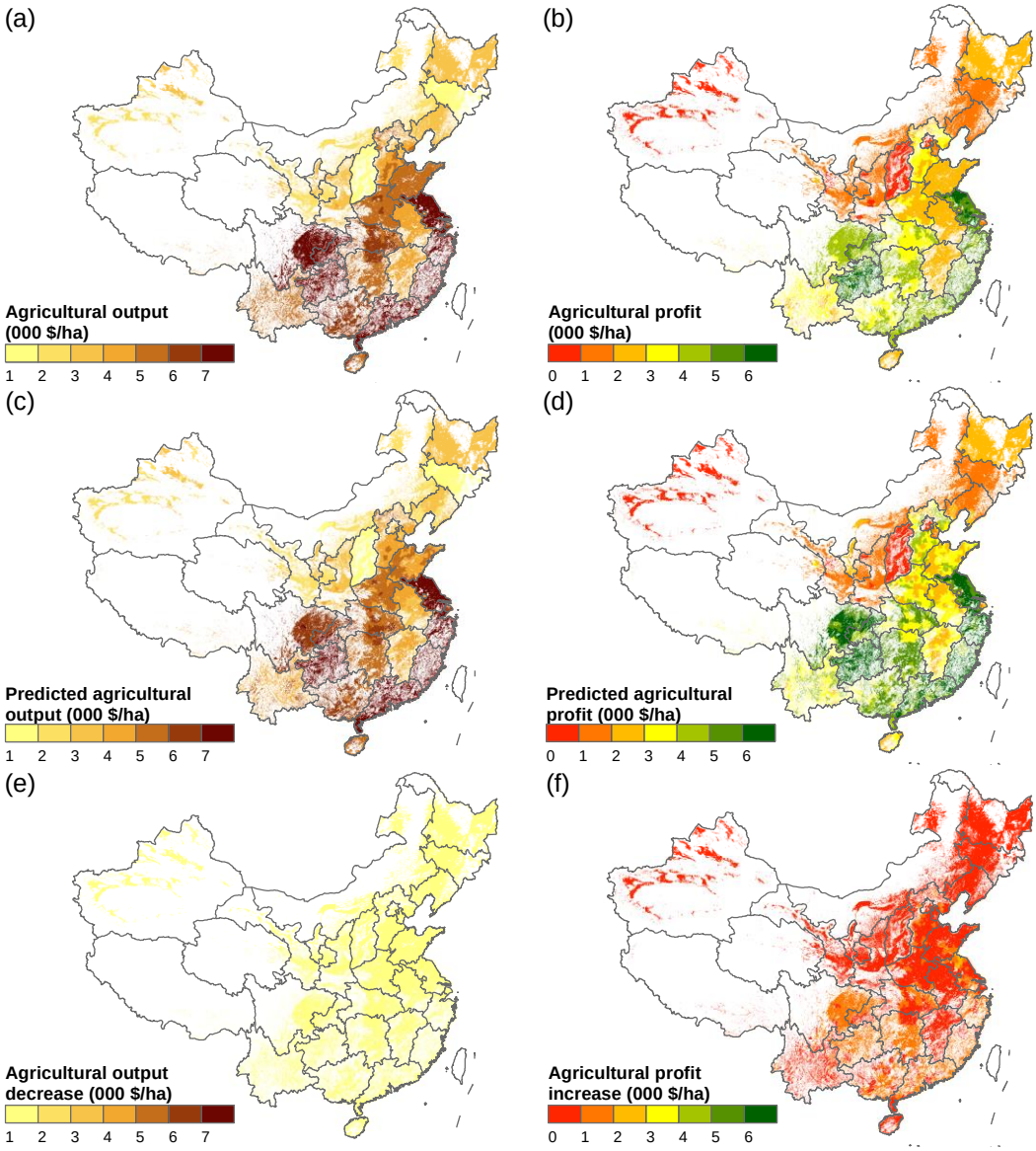

Fig. S10 Changes of agricultural output and profit. (a) Current agricultural output; (b) Current agricultural profit; (c) Predicted agricultural output of large-scale farming; (d) Predicted agricultural profit of large-scale farming; (e) Agricultural output decrease; (f) Agricultural profit increase. Agricultural output is total market value of all crop yields directly reported by farmers. It includes all grains and crash crops. Agricultural profit equals to the difference between total agricultural output and cost. Current data is from China Agricultural Yearbook 2017. The predicted calculation is based on current values and changes in the field size showed in Fig. 1d and according to relations between farm size and agricultural output and profit in China (See Table 1). The changes are the differences between predicted value and the current one.

Table S1 Agricultural fertilizer use (Top 10% countries)

Rank	World Bank (2016)		Food and Agriculture Organization (2017)	
	Area	Fertilizer use (thousand kg ha^{-1})	Area	Fertilizer use (million t)
1	Singapore	30.2	China	29.8
2	Qatar	6.8	Americas	24.5
3	Hong Kong	2.7	India	17
4	New Zealand	1.8	Brazil	5.2
5	Malaysia	1.7	Pakistan	3.4
6	Bahrain	1.3	Indonesia	3
7	Ireland	1.2	Canada	2.5
8	Kuwait	0.8	France	2.2
9	Colombia	0.7	Turkey	1.8
10	Egypt	0.6	Viet Nam	1.5
11	Costa Rica	0.6	Russia	1.5
12	Seychelles	0.5	Germany	1.5
13	China	0.5	Thailand	1.5
14	Oman	0.5	Mexico	1.5
15	Belize	0.5	Egypt	1.4
16	Viet Nam	0.4	Ukraine	1.4
17	South Korea	0.4	Bangladesh	1.2
18	Trinidad and Tobago	0.4	Poland	1.2
19	Ecuador	0.3	Australia	1.1
20	Lebanon	0.3	Spain	1.1

Note: The value in the World Bank database is fertilizer use per hectare in 2016, while the value in Food and Agriculture Organization of the United Nations (FAO) is total fertilizer use in 2017. We can see there are some difference between the ranking in both lists. China takes the leading position in total fertilizer use with the number of 29.8 million ton. And China ranks 13th for fertilizer use per hectare in 500 kg ha^{-1}.

Table S2 Consolidation cost samples

Province ID	Province Name	Region Name	Number of sample site	Average Consolidation cost (thousand Yuan ha^{-1})
11	Beijing	HM	1	18.2
12	Tianjin	LP	2	18.4
13	Hebei	HP	2	17.9
14	Shanxi	LM	2	17.6
15	Inner Mongolia	LM	7	17.7
21	Liaoning	HP	2	23.9
22	Jilin	LP	1	20.6
23	Heilongjiang	LP	1	16.5
31	Shanghai	HM	-	-
32	Jiangsu	HP	17	36.7
33	Zhejiang	HM	7	36
34	Anhui	HP	12	19.1
35	Fujian	HM	2	34.1
36	Jiangxi	LM	21	34.7
37	Shandong	HP	4	21.7
41	Henan	HP	9	21.9
42	Hubei	HP	14	31.7
43	Hunan	HM	13	22.7
44	Guangdong	HM	12	31.2
45	Guangxi	LM	7	23.4
46	Hainan	LM	6	60.8
50	Chongqing	LM	5	22.3
51	Sichuan	HM	8	23.6
52	Guzhou	LM	12	30.4
53	Yunnan	LM	12	27.3
54	Tibet	LM	1	30.7
61	Shannxi	HM	4	19.8
62	Gansu	LM	5	22
63	Qinghai	LM	3	23.6
64	Ningxia	LM	2	22.6
65	Xinjiang	LP	7	18.2

Note: Province ID is the administrative code of each province in China. Province name is full name. The order of regions is related to terrain and local economic conditions (GDP rank). HP refers to high-income plain region. LP refers to low-income plain region. HM represents for high-income mountainous region. LM represents for low-income mountainous region. And average consolidation cost is the cost for a project divided by the total area.

Table S3 Consolidation cost

Region name	Cropland (million ha)	Percentage of consolidation (%)	Consolidation area (million ha)	Average consolidation cost ($ ha^{-1})	Consolidation cost (billion $)
HP	48.4	91	44.0	3,530	155.4
LP	26.7	88	23.5	2,634	61.9
HM	23.0	71	16.3	3,787	61.9
LM	36.7	69	25.4	3,535	89.8
SUM	134.8	-	109.2	-	368.9

Note: HP refers to high-income plain region. LP refers to low-income plain region. HM represents for high-income mountainous region. LM represents for low-income mountainous region. The cropland is the sum of cropland of the whole region. Area needed consolidation is calculated by the change of field size in different regions (Fig. S4). And consolidation area is total cropland plus percentage. Average consolidation cost is from Table. S2 and the numbers have been converted to US dollars at an exchange rate of 1:7. Then we got the total cost.

Table S4 Number of plots and sum area for each field size for predicted scale-farming

Field size (ha)	<0.64	0.64-2.56	2.56-16	16-100	>100
Number of plots (million)	30.12	4.64	1.36	0.24	0.05
Sum area (million ha)	5.64	5.28	7.66	8.39	134.88

Note: Field size classification is from Lesiv et al [1]. And here we show the number of plots and sum area for each group. It can be seen that although number of plots doesn't change a lot, but the area share of large field has increased a lot.

Table S5 Regression results of farm size to agricultural input and out changes

	Labor			Chemical use			Cost and profit		
	Ln Person ha^{-1}	Ln Labor cost ($ ha^{-1})	Ln LP ($ hr^{-1})	Ln Fer ($ ha^{-1})	Manure ($ ha^{-1})	MF ratio	Ln Cost ($ ha^{-1})	Ln Profit ($ ha^{-1})	Ln output ($ ha^{-1})
Ln Farm size (ha)	-0.73***	-0.73***	0.33***	-0.26***	21.64***	0.58**	-0.62***	0.08***	-0.03
Dy/Dx					4.19***	0.09**			
Region	County	County	County	County			County	County	Province
Model	OLS	OLS	OLS	OLS	Tobit	Tobit	OLS	OLS	OLS
N	16717	11489	16499	16277	12424	11815	12124	12025	8249
F	387	192	174	135	8	0.09	208	40	103
Adjust R2	0.53	0.47	0.33	0.35			0.49	0.26	0.38

*** $p < 0.001$; ** $p < 0.05$; * $p < 0.01$. LP, Labor productivity; Fer, Chemical fertilizer; MF ratio, Manure fertilizer ratio. And labor person input has been weighted according to with working time of different labors. OLS refers to Ordinary Least Squares regression analysis. We use Tobit model rather than OLS for Manure and MF ratio considering there are too many zeros of the two variables. The effect of multiple crop index, plant type, plot numbers, year and region effect have been controlled in all OLS regressions. Province rather than County was controlled in Output regression and region was not controlled in the Tobit model both due to data limitations. Furthermore, we regressed output with farm size while additionally controlling fertilizer, machine, seed, pesticides and labor input.

Table S6 Summary Statistics for variables used in regression analysis

	Obs	Mean	SD	Min	Max
Ln Farm size (ha)	20,766	-0.99	1.26	-22.33	6.20
Ln Person ha^{-1}	19,935	1.21	1.38	-6.44	4.36
Ln Labor Cost (\$ ha^{-1})	19,252	8.23	2.20	-1.48	12.79
Ln Labor Productivity (\$ hr^{-1})	19,482	-1.29	1.57	-11.38	6.39
Ln Fertilizer (\$ ha^{-1})	19,145	5.87	1.12	-6.50	8.87
Manure (kg ha^{-1})	14,971	50.94	175.68	0	2142.86
MF ratio	13,162	4.27	437.09	0	50000
Ln Cost (\$ ha^{-1})	14,588	9.34	1.47	1.35	15.15
Ln Profit (\$ ha^{-1})	14,380	12.50	0.25	-0.06	12.76
Ln Output (\$ ha^{-1})	20,155	7.64	1.22	-2.42	30.37
Ln Multiple Crop Index	17,545	-0.03	0.76	-3.30	2.30
Plot numbers (Categorical variable)	20,420	3.07	1.97	1	6
Plant type (Categorical variable)	23,480	5.30	3.81	1	10

Table S7 Summary Statistics for changes

	Current	Large-scale farming	Changes (%)
Average field size (ha)	2.77	12.24	441.9
N fertilizer (kg ha^{-1})	214.26	124.37	58.0
Manure (kg ha^{-1})	50.86	59.24	116.5
Cropland input (kg ha^{-1})	356.38	272.63	76.5
Cropland yield (kg ha^{-1})	147.75	139.76	94.6
NUE (%)	44.23	52.36	118.4
N surplus (kg ha^{-1})	207.02	118.32	57.2
Agricultural labor (person ha^{-1})	2.80	1.70	60.8
Labor income (\$ $person^{-1}$)	2,540.03	6,214.29	244.7
Agricultural cost (\$ ha^{-1})	2,185.36	1,093.68	50.0
Agricultural profit (\$ ha^{-1})	2,679.25	3,212.71	119.9
Agricultural output (\$ ha^{-1})	4,864.61	4,582.63	94.2

Table S8 Recommended N application rate for different crops (kg N ha^{-1})

	rice	wheat	corn	millet	sorghum	barley	beans	potato	peanut	rapeseeds	cotton	hemp	tobacco	sugar beet	sugar cane	vegetable	fruits
Beijin	135	120	135	120	180	135	24	60	60	90	165	135	84	157.5	285	270	240
Tianji	135	120	135	120	180	135	24	60	60	90	165	135	84	157.5	285	300	240
Hebei	135	150	135	120	180	135	24	60	60	90	165	135	84	157.5	285	345	255
Shanx	135	105	135	120	180	135	24	60	60	90	105	135	84	157.5	285	300	285
Inner	135	90	135	120	180	135	24	60	75	90	105	135	84	157.5	285	315	345
Liaoni	90	120	135	120	180	135	45	90	60	90	105	135	84	157.5	285	345	240
Jilin	90	105	150	120	180	135	24	105	75	0	120	135	84	157.5	285	240	405
Heilo	90	105	135	120	180	135	24	75	75	165	0	135	84	157.5	285	225	450
Shang	180	120	135	120	180	135	45	90	67.5	120	240	135	84	157.5	285	225	330
Jiangs	150	150	135	120	180	135	45	90	82.5	120	165	135	84	157.5	285	270	375
Zhejia	150	120	135	120	180	135	42	90	75	120	240	135	84	157.5	285	270	270
Anhui	150	150	135	120	180	135	24	90	90	120	165	135	84	157.5	285	270	480
Fujian	150	90	120	120	180	135	42	90	75	75	90	135	84	157.5	285	255	225
Jiangx	150	90	120	120	180	135	42	105	75	75	240	135	84	157.5	285	255	225
Shand	180	150	150	120	180	135	42	105	90	120	165	135	84	157.5	285	315	405
Henan	180	150	135	120	180	135	24	60	105	120	165	135	84	157.5	285	255	435
Hubei	120	120	120	120	180	135	42	90	90	120	240	135	84	157.5	285	225	300
Hunan	120	120	120	120	180	135	45	90	75	120	240	135	84	157.5	285	225	270
Guang	135	120	120	120	180	135	52.5	105	75	75	0	135	84	157.5	285	210	225
Guang	135	90	120	120	180	135	24	90	75	75	90	135	84	157.5	285	195	225
Haina	135	0	120	120	180	135	42	90	75	0	0	135	84	157.5	285	195	300
Chong	120	90	120	120	180	135	24	90	75	120	90	135	84	157.5	285	210	225
Sichu	120	120	120	120	180	135	42	90	75	120	90	135	84	157.5	285	240	225
Guizh	120	90	120	120	180	135	24	90	75	120	90	135	84	157.5	285	195	225
Yunna	120	90	135	120	180	135	42	90	60	120	270	135	84	157.5	285	195	225
Tibet	120	150	120	120	180	135	40.5	105	75	120	0	135	84	157.5	285	240	225
Shaan	150	120	120	120	180	135	24	60	75	120	270	135	84	157.5	285	270	225
Gansu	180	90	120	120	180	135	24	90	90	120	270	135	84	157.5	285	270	225
Qingh	0	120	150	120	180	135	0	90	90	120	0	135	84	157.5	285	270	225
Ningx	150	90	165	120	180	135	24	60	90	120	300	135	84	157.5	285	300	225
Xinjia	180	120	150	120	180	135	34.5	120	90	120	210	135	84	157.5	285	330	225

Note: The value is from guidance on scientific fertilization of major crops in 2010 by the Ministry of Agriculture of the People's Republic of China and Zhang et al [4]