Chin. Geogra. Sci. 2017 Vol. 27 No. 2 pp. 273–285
doi: 10.1007/s11769-017-0862-1

Springer Science Press
www.springerlink.com/content/1002-0063

Accuracy Comparison of Gridded Historical Cultivated Land Data in Jiangsu and Anhui Provinces

YUAN Cun[1, 2], YE Yu[1, 3], TANG Chanchan[1], FANG Xiuqi[1]

(1. *School of Geography*, *Beijing Normal University*, *Beijing* 100875, *China*; 2. *Encyclopedia of China Publishing House*, *Beijing* 100037, *China*; 3. *Key Laboratory of Environment Change and Natural Disaster*, *Ministry of Education*, *Beijing Normal University*, *Beijing* 100875, *China*)

Abstract: The spatial resolution of source data, the impact factor selection on the grid model and the size of the grid might be the main limitations of global land datasets applied on a regional scale. Quantitative studies of the impacts of rasterization on data accuracy can help improve data resolution and regional data accuracy. Through a case study of cropland data for Jiangsu and Anhui provinces in China, this research compared data accuracy with different data sources, rasterization methods, and grid sizes. First, we investigated the influence of different data sources on gridded data accuracy. The temporal trends of the History Database of the Global Environment (HYDE), Chinese Historical Cropland Data (CHCD), and Suwan Cropland Data (SWCD) datasets were more similar. However, different spatial resolutions of cropland source data in the CHCD and SWCD datasets revealed an average difference of 16.61% when provincial and county data were downscaled to a 10 × 10 km^2 grid for comparison. Second, the influence of selection of the potential arable land reclamation rate and temperature factors, as well as the different processing methods for water factors, on accuracy of gridded datasets was investigated. Applying the reclamation rate of potential cropland to grid-processing increased the diversity of spatial distribution but resulted in only a slightly greater standard deviation, which increased by 4.05. Temperature factors only produced relative disparities within 10% and absolute disparities within 2 km^2 over more than 90% of grid cells. For the different processing methods for water factors, the HYDE dataset distributed 70% more cropland in grid cells along riverbanks, at the abandoned Yellow River Estuary (located in Binhai County, Yancheng City, Jiangsu Province), and around Hongze Lake, than did the SWCD dataset. Finally, we explored the influence of different grid sizes. Absolute accuracy disparities by unit area for the year 2000 were within 0.1 km^2 at a 1 km^2 grid size, a 25% improvement over the 10 km^2 grid size. Compared to the outcomes of other similar studies, this demonstrates that some model hypotheses and grid-processing methods in international land datasets are truly incongruent with actual land reclamation processes, at least in China. Combining the model-based methods with historical empirical data may be a better way to improve the accuracy of regional scale datasets. Exploring methods for the above aspects improved the accuracy of historical cropland gridded datasets for finer regional scales.

Keywords: accuracy evaluation; spatial resolution; grid-processing method; grid size; historical period

Citation: Yuan Cun, Ye Yu, Tang Chanchan, Fang Xiuqi, 2017. Accuracy comparison of gridded historical cultivated land data in Jiangsu and Anhui provinces. *Chinese Geographical Science*, 27(2): 273–285. doi: 10.1007/s11769-017-0862-1

1 Introduction

Historical land use and land cover changes have attracted widespread research attention, and the reconstruction of high-resolution data has made great progress. In recent years, some research has focused on

Received date: 2016-07-06; accepted date: 2016-11-08

Foundation item: Under the auspices of National Natural Science Foundation of China (No. 41471156, 41501207), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA05080102), Special Fund of National Science and Technology of China (No. 2014FY130500)

Corresponding author: YE Yu. E-mail: yeyuleaffish@bnu.edu.cn

how to convert existing statistical data based on administrative units into a gridded format. The use of different source data, spatial resolutions, gridding processes, and grid sizes will affect the accuracy of reconstructed data. Thus, quantitative research on the effects of these factors is needed.

The international datasets of the Center for Sustainability and the Global Environment (SAGE) (Ramankutty and Foley, 2010) and the History Database of the Global Environment (HYDE) (Goldewijk *et al.*, 2011) have become representative examples of this research area. However, the accuracy of global datasets is still limited at the regional scale, and popular international global-scale or large-scale reconstruction methods may not be suitable for regional-scale land cover reconstruction (Kaplan *et al.*, 2011; Fuchs *et al.*, 2012; Li *et al.*, 2013). For instance in China, Li *et al.* (2010) and He *et al.* (2012) compared the SAGE and HYDE datasets with historical data on arable land in Northeast China and traditional Chinese agricultural areas, respectively, from the past 300 years. The trends of the SAGE dataset, particularly linear trends in land reclamation, contrasted sharply with those of the regional data. There was also a significant discrepancy in spatial distribution between the HYDE dataset and Chinese historical arable land records. Zhang *et al.* (2013) found that the average reclamation rate in the HYDE and PJ datasets was lower than that in Chinese regional data based on historical arable land records. Moreover, the HYDE dataset overestimated the reclamation rate in northern China and underestimated it in the Yangtze River Basin. Thus, approaches to regional-scale land cover reconstruction, particularly with regard to historical land cover changes in China, must be strengthened, and their applicability should be assessed.

Domestic scholars have conducted some research in this field (Lin *et al.*, 2009; Li *et al.*, 2016). The gridding process and the spatial resolution of source data affect the accuracy of reconstructed data. For example, the source of the gridded cultivated land data used by Ge *et al.* (Chinese Historical Cultivated Land Dataset, abbreviated as CHCD) (2003) was primarily provincial arable land data. However, Yuan *et al.* (2015) used source data of different spatial resolutions and found that county data improved accuracy by 16% over provincial data at a 10 km^2 grid size for Jiangsu and Anhui provinces (Suwan Cropland Dataset, abbreviated as SWCD).

In addition, the impact of the gridding method on the accuracy of regional-scale data must be evaluated. To construct both an international and domestic dataset, the appropriate limiting factors and gridding model need to be selected based on regional characteristics (Wei *et al.,* 2014). Zhang *et al.* (2014) distributed the arable land of Heilongjiang Province during the late 19th century in a grid with 1 km × 1 km pixels by building a comprehensive reclamation tendency model that included factors such as settlement, terrain, and water. Feng *et al.* (2014) proposed the idea of separate modeling and used different models for the gridding of traditional agricultural areas, Northeast China, Northwest China, and Qinghai-Tibet. Long *et al.* (2014) and Yang *et al.* (2015) used the CA model to reconstruct historical spatial patterns of cultivated land in Jiangsu and Shandong provinces. Luo *et al.* (2015) used different grid-processing methods for the Hehuang Valley in Tibet. Their results were fragmented because of factors such as urbanization, infrastructure, and forest husbandry that influence modern distribution patterns of arable land. Ignoring such factors and considering only natural and other human factors can polarize gridding results.

Overall, research pertaining to the impacts of gridding methods on data accuracy is lacking. At present, the main challenges and debates surrounding gridding model reconstruction methods include the validity of the hypothesis that historical land reclamation was within the scope of modern cultivated land and the impacts of the modern reclamation rate or range of potential reclamation rates as parameters of reconstruction models, temperature as a model factor on grid-processing, and different river and lake processing methods on gridded data. In addition, quantitative research on the effects of different grid sizes on data accuracy is worth conducting.

In summary, internationally popular global-scale reconstruction methods are not always suitable for regional-scale reconstruction of Chinese land cover changes; thus, regional-scale methods must be strengthened. Using cropland data from Jiangsu and Anhui provinces, this study quantitatively compared the accuracy of gridding results from different data sources, gridding methods, and grid sizes. It will provide a reference for the establishment of a regional dataset that accurately reflects China's historical regional land cover changes.

2 Materials and Methods

2.1 Study area

Jiangsu and Anhui provinces (collectively, the Suwan area) are traditional agricultural areas of China. They are considered active hot spots, influenced by both natural and anthropogenic factors and famously known as a land overflowing with fish and rice. The two provinces are located in the Yangtze River Delta, which comprises mainly hilly plains with low, flat terrain and a developed water system with many rivers and lakes (Fig. 1). Paddy fields account for 60.06% of Jiangsu Province's cultivated land, concentrated in the old rice areas of Taihu Lake, the Lixia River, and the Zhenjiang-Yangzhou hilly region. The majority of dry land in Jiangsu Province is located in Xuzhou, Huaiyin, and coastal and riverside areas (The agriculture of China, Jiangsu volume editor committee, 1998). Anhui Province spans the Huaihe River and the Yangtze rivers, which divide the province into three regions: Huaibei, Jianghuai, and Jiangnan. The paddy fields of Anhui Province are mainly located along the Yangtze River and in the Huaihe River hilly area. Dry land is mainly distributed on the Huaibei plain and in the Huaihe River hilly area. According to the administrative regional distribution, the arable land of Fuyang, Liuan, Suxian, and Chuzhou accounts for 55.26% of the province's total arable land (The agriculture of China, Jiangsu volume editor committee, 1998).

2.2 Data sources

For Jiangsu and Anhui, the frequently used gridded land data include the SAGE and HYDE datasets and the Chinese Historical Cultivated Land Dataset (CHCD) of traditional agricultural areas. The SAGE dataset was reconstructed by the Global Environment and Sustainable Development Center of the University of Wisconsin (http://daac.ornl.gov/ISLSCP_II/guides/historic_cropland_xdeg.html), covering the years of 800 to 1992 with a 0.5° × 0.5° spatial resolution and 10-year time resolution. The HYDE dataset was reconstructed by the Netherlands Environmental Assessment Agency (http://themasites.pbl.nl/en/themasites/hyde/download/index.html), with a reconstruction period of 10000 BC to 2000 AD, 5′ spatial resolution, and time resolution of 10 year after 1700 AD. The CHCD dataset, which is based on provincial data and has a spatial resolution of 60 km × 60 km, was obtained from Lin *et al.* (2009). The provincial data were mainly from the cultivated land data reconstructed by Ge *et al.* (2003), while data on years after 1949 were from the National Statistics Bureau (He *et al.*, 2012). The source dataset used in the gridding process of the arable land in Jiangsu and Anhui (Suwan Cropland Data, hereafter SWCD) included county-level data on the Qing Dynasty (1735) and Republic of China (1933), as provided by Zhao (2005); modern general survey data (1980s); and detailed survey data (1996). Modern land use data were obtained from the Earth System Science Data Sharing Platform (http://www.

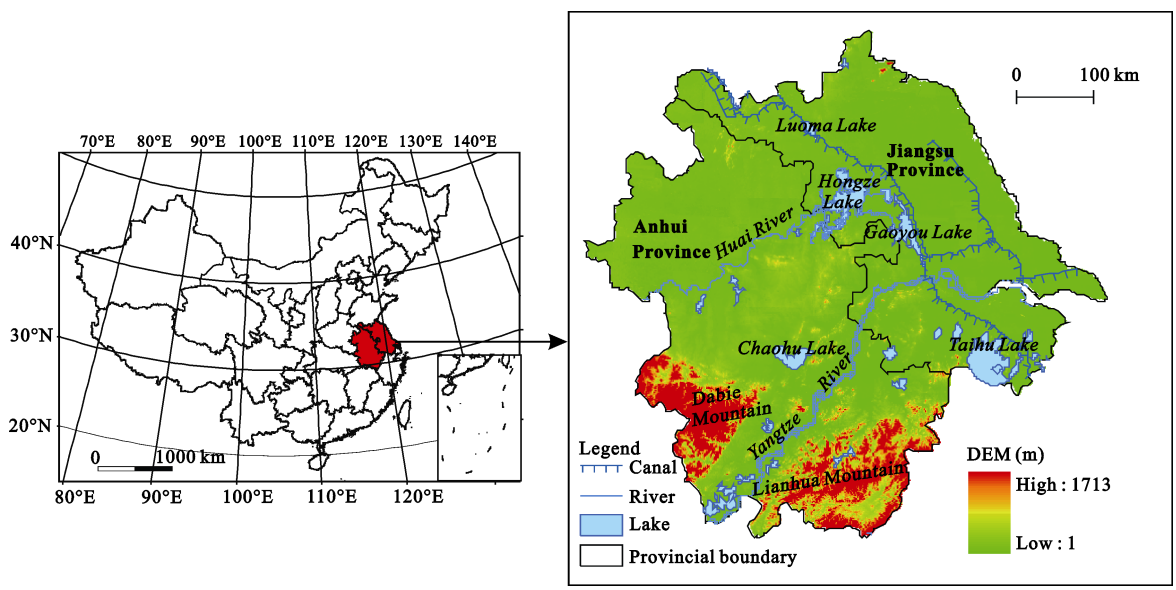

Fig. 1 Study area. DEM: Digital Elevation Model

geodata.cn/Portal/index.jsp).

We compared data in terms of time trends and spatial distribution. The SAGE, HYDE, CHCD, and SWCD datasets all have different time distributions. They were all transferred to corresponding time sections representing decadal-scale resolution, assuming that the arable land in a short time will not have a fundamental change limited by climate and topography. In this study, six time periods were selected based on the SWCD, including the mid-Qing Dynasty (1735), the late Qing Dynasty (1860), the Republic of China (1933), the early days of liberation (1950), the 1980s, and the late 20th century (1995). The SWCD data for 1735, 1933, and 1996 closely corresponded to the years 1730, 1930, and 1995; CHCD data for 1724, 1873, and 1933 data corresponded to the years 1730, 1860, and 1930; SAGE data for 1992 corresponded to 1995; and the average of the 1990 and 2000 data from the HYDE dataset was used in lieu of 1995 data.

2.3 Methodology

The arable land gridding process model was created by Yuan *et al.* (2015). The model was based on the potential arable land reclamation rate, area of rivers and lakes, slope, and average annual temperature, as expressed by Equation (1):

$$a(i) = N(i) \times \varepsilon(i) \times S'(i) \times T'(i) \tag{1}$$

where $\alpha(i)$ represents the extent of the priority arable land reclamation of grid i, $N(i)$ represents the potential arable land reclamation rate of a historical period (computed as the sum of the modern arable land reclamation rate and modern construction land reclamation rate), $\varepsilon(i)$ represents the standard value of land area within grid i, $S'(i)$ represents slope standardized value of grid i, and $T'(i)$ represents the annual temperature standard value of grid i. The cultivated land area of grid i of county k is determined by Equation (2):

$$\delta(i,k) = a(i,k) / \sum a(i,k) \times A(k) \tag{2}$$

where $\alpha(i,k)/\Sigma_i\alpha(i,k)$ is the proportion of cultivated land area of grid i of county k, and $A(k)$ is the total area of cultivated land in county k. The reclamation rate for grid i is determined using Equation (3):

$$K(i) = \delta / area(i) \tag{3}$$

where $area(i)$ is grid area, the value of which was 100 km^2 in this study. Using this model, cultivated land data from 1735, 1932, 1980, and 1996 were processed into a grid, and the gridded cropland dataset of Jiangsu and Anhui, or the SWCD, was constructed.

Using the cultivated land data from the mid-Qing Dynasty in 1735 as a representative example, we quantitatively analyzed the impact of the gridding method on the accuracy of gridding results. To investigate the effects of the potential arable land reclamation rate and annual temperature on accuracy, we individually removed each factor from the original model and compared the grid-processing results with those of Yuan *et al.* (2015). For example, when the potential arable land reclamation rate parameters are removed, the reclamation rate is calculated as Equation (4) and Equation (5):

$$a(i) = \varepsilon(i) \times S'(i) \times T'(i) \tag{4}$$

When the mean annual temperature is removed, the reclamation rate is calculated as Equation:

$$K(i) = a(i,k) / \sum a(i,k) \times A(k) / area(i) \tag{5}$$

The accuracy of the river and lake treatment method was then investigated, and the results from the HYDE dataset were compared with the grid-processing results of Yuan *et al.* (2015). Finally, the impact of gridding size on accuracy was evaluated. The grid size was set to

Table 1 Sources of cropland area data for Suwan

Dataset	Spatial resolution	Time slices (time resolution)	Time span	Data source
SAGE	0.5° × 0.5°	1–50 years	800–1992	Ramankutty and Foley (2010)
HYDE	5′ × 5′	10 years after 1700	10000 BC–2000 AD	Goldewijk *et al.* (2011)
CHCD	60 km × 60 km	1661, 1724, 1784, 1820, 1873, 1911	Since the Qing Dynasty	Lin *et al.* (2009) and He *et al*. (2012)
SWCD	10 km × 10 km	1735, 1933, 1980s, 1996	Since the Qing Dynasty	Yuan (2015)

Notes: SAGE is datasets of the Center for Sustainability and the Global Environment; HYDE is History Database of the Global Environment; CHCD is Chinese Historical Cultivated Land Dataset; SWCD is Suwan Cropland Dataset

1 km^2 and 10 km^2, and the above model was used to compare the results. Absolute and relative differences between the results of different treatments were calculated to analyze the impacts of grid-processing methods on accuracy.

3 Results

3.1 Comparison of data sources for grid-processing

In terms of total data, the SAGE dataset was larger than the SWCD dataset, while the HYDE and CHCD datasets were close in size to the SWCD dataset. Through comparative analysis, the HYDE and SWCD datasets were found to reflect Anhui Province accurately. There were only 2.15% more HYDE data from 1995 than SWCD data. The difference between the HYDE and SWCD datasets was greatest in 1930, when there were 13.33% more HYDE data than SWCD data (Fig. 2a). The difference in the amount of data in Jiangsu Province was greater than that in Anhui Province. In Jiangsu, the greatest difference appeared from 1930 to 1950, when there were 22.5% more HYDE data than SWCD data. The smallest difference (8.75%) occurred in 1980 for the two provinces. The difference between the CHCD and SWCD datasets was small. This difference was most noticeable in 1730 in Jiangsu Province (21.26%). The difference in the size of the datasets for the entire Suwan area was 11.66%. This was mainly because the administrative area of the CHCD data was based on the Qing Dynasty regionalization, in which Shanghai was returned to Jiangsu.

Overall, both for Anhui, Jiangsu and the complete Suwan area, the trends regarding cultivated land were basically the same between the two datasets. From 1730 to 1950, as a result of the Taiping Heavenly Kingdom War, all areas exhibited a relatively stable cultivated land area and weak growth (Fig. 2). Variations in cultivated land trends in each dataset were mainly due to the following factors: the gridding handle of the SAGE source data was a linear interpolation of cultivated land data; the HYDE source data were the cultivated land data of different time periods per capita arable land area; the CHCD dataset reflected trends using an alternative method of trend-data revision of historically arable land; and the SCWD data in 1735 were based on local chronicles that considered the discount in mu (1 ha = 0.0015 mu) and converted tax units into statistical data.

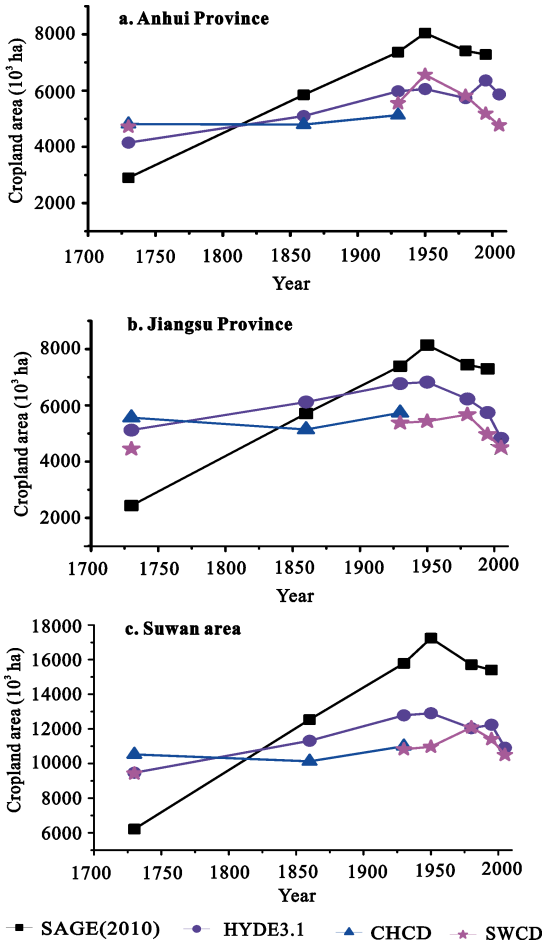

Fig. 2 Dataset of cropland area of Center for Sustainability and the Global Environment (SAGE), History Database of the Global Environment (HYDE), Chinese Historical Cultivated Land Dataset (CHCD), and Suwan Cropland Dataset (SWCD)

From the previous assessment and evaluation results of He *et al.* (2012), cropland data of the Jiangsu and Anhui provinces in the SAGE and CHCD datasets were found to differ significantly, while those of the CHCD and SWCD datasets differed only slightly. There are 154 counties in the Jiangsu and Anhui provinces, and approximately half (77 counties) of these counties' areas range from 1000 to 2000 km^2, while only five counties' areas exceed 3000 km^2. The spatial resolution of the CHCD dataset was 60 km × 60 km, and each grid was 3600 km^2. Therefore, this dataset could not accurately reflect the spatial distribution of land in the county and could not be used in our comparison.

The two provinces were found to differ in terms of spatial distribution between the source data of the HYDE3.1 and SWCD datasets. This difference was concentrated in areas around rivers, especially during

the Qing Dynasty and the Republic of China. The greatest difference occurred for the two provinces in 1930 (Fig. 3b), with the HYDE data 70% greater than the SWCD data in the majority of coastal areas and the Yangtze River canal area. The difference in spatial distribution in the Huaihe River coastal region also reached more than 40%. In the 1980s and 1995, the difference between the two datasets for the two provinces was less pronounced at less than 40%, but it was still relatively large along rivers and in lake areas (Fig. 3).

With different spatial resolutions of arable land data (comprising provincial and county domain data), the results were different, even with the same amount of data. These data were allocated into 10 km × 10 km grid cells through downscale processing. The gridded results differed between the simulated Jiangsu and Anhui provincial land data and county land data by an average rate of 16.61%. The relatively smaller difference rate (–10%–10%) occupied 24.55% of the grid, and the larger rate (>70% or ≤–70%) occupied 13.3% of the grid, primarily within the Hongze Lake Basin and northern Jiangsu along the Yangtze River Plain (Yuan, 2015).

3.2 Influence of different factors selection and grid-processing methods on accuracy

3.2.1 Influence of potential land reclamation rate

As seen in Fig. 4 and Table 2, the spatial distribution of cultivated land was consistent overall, with only slight

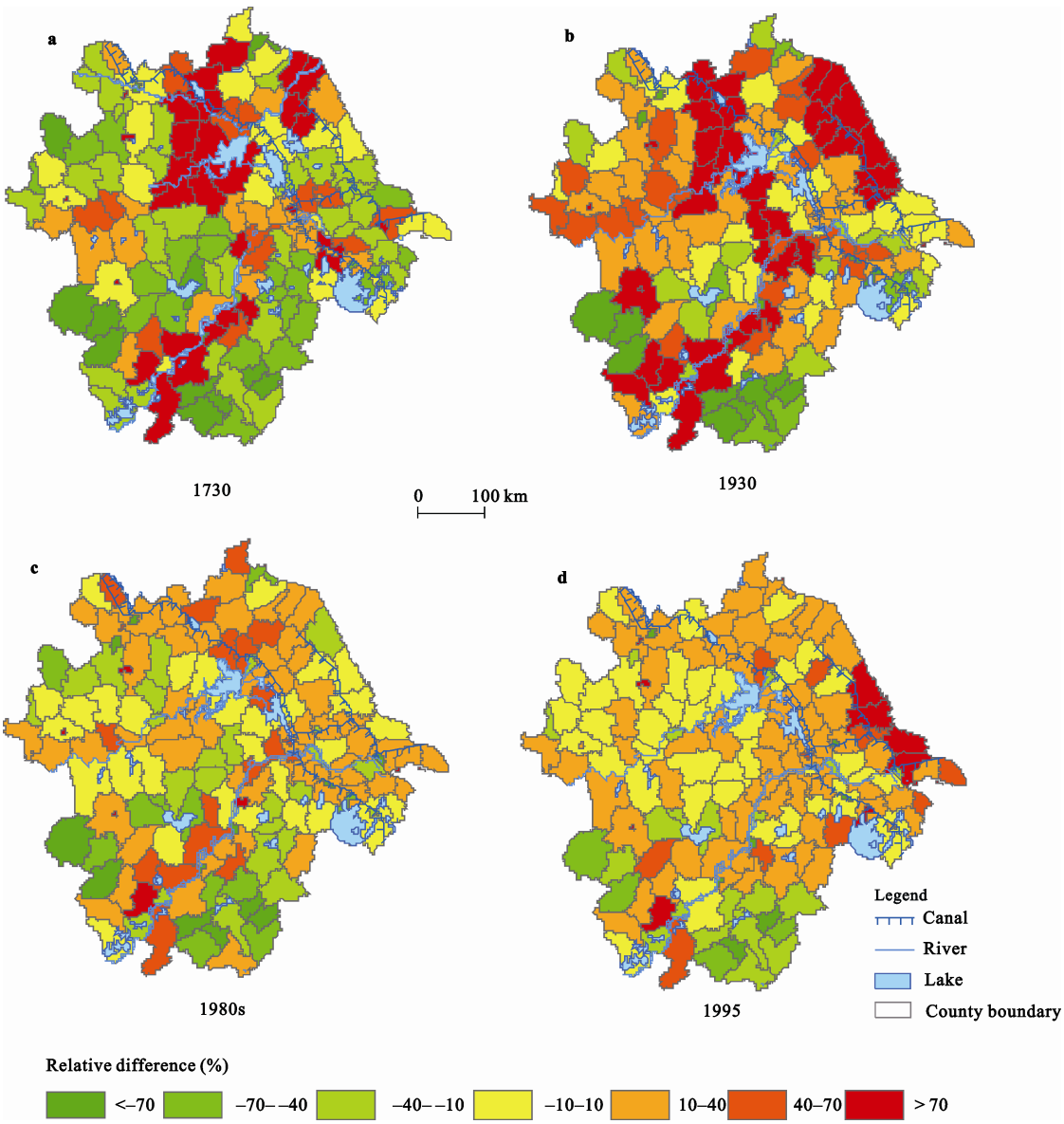

Fig. 3 Relative difference in county spatial distribution between HYDE and SWCD datasets

differences between the two models with and without the consideration of the potential arable land reclamation rate. Without the addition of the potential land reclamation rate, the cultivated land spatial distribution within the county area or adjacent to the county was more uniform, and spatial differences and standard deviations were small (Fig. 4a). With the added potential arable land reclamation rate, the results had the same characteristics as modern arable land distribution, with spatial differences to an extent (Fig. 4b). Because Jiangsu and Anhui provinces have a long history of traditional agriculture, their natural condition is more balanced. In the later development period, some locations became suitable for reclamation, but other places remained unsuitable. However, this difference was not very obvious, as 72.98% of grids exhibited an absolute difference in cultivated land area within 10 km^2, and 51.58% of grids had a relative difference rate of less than 10%. Only 10.92% of grids exhibited relatively large differences (>70% or <–70%). Approximately half of the grids with relatively large differences were distributed in mountainous areas where the reclamation rate mostly ranged from 10% to 20%, while others were scattered in the Suwan region's plains. When the potential cultivated land reclamation rate was added to the model, standard deviation increased by 4.05.

Because the simulation results of the two models in each grid were one-to-one, a paired-samples *t*-test was selected to analyze whether the difference in the results was significant. The premise of the test was that significant test data needed to be normally distributed. The data were found to be not normally distributed and needed to be converted to a normal distribution. After normal conversion, the *t*-statistic was –0.996, and the associated probability was 0.319, which was greater than the 0.05 significance level. Therefore, the difference was not statistically significant.

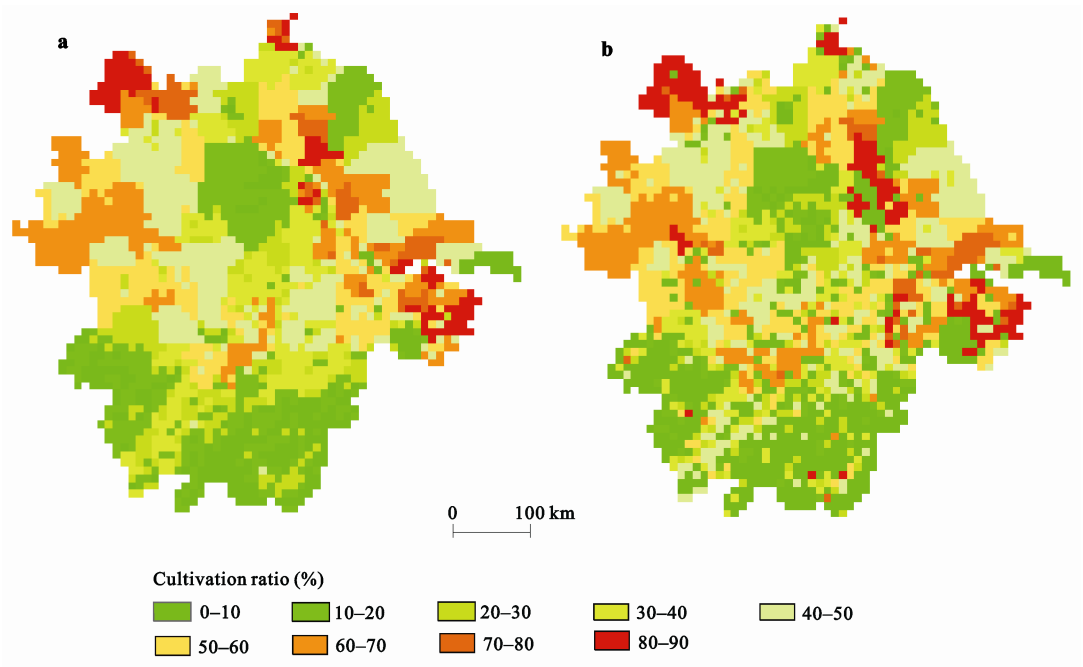

Fig. 4 Grid distribution results of models without (a) and with (b) the potential land reclamation rate parameter

Table 2 Comparison of reclamation rates between two models

Model	0–20%	20%–40%	40%–60%	60%–80%	80%–90%	Standard deviation
Model 1	10.8585	30.0689	48.6257	60.0391	84.2024	22.23
Model 2	5.3166	31.0482	49.3074	65.7471	85.3878	26.28

Notes: Model 1 refers to the gridding model without the potential cultivated land reclamation rate. Model 2 refers to the model with the potential cultivated land reclamation rate

3.2.2 Comparison on Influence of different temperature factors

In the original cultivated land allocation model, terrain is considered to be the main factor affecting cultivated land (Hall *et al.*, 1995), and climatic factors are not considered. Other model types are diverse and complex, with some using temperature as a climate indicator (e.g., HYDE dataset) and some using climatic potential productivity (Li *et al.*, 2012; 2014; He *et al.*, 2014; Luo *et al.*, 2014).

The Jiangsu and Anhui provinces have adequate water and light, and temperature exerts a relatively large impact on crop yields of all climate factors. Zhang (1982) reported that when temperature changed by 1℃, the corresponding crop yield changed by 10%. In our model, cultivated land area was considered with and without the annual average temperature parameter, and 96.07% of the grid exhibited a relative difference from –10%–10%, while 96.85% of the grid had an absolute difference ranging from –2–2 km^2. Thus, the effect of average annual temperature on the model was minimal.

The temperature factor selection index can vary and affect the distribution of cultivated land in terms of an average annual temperature or an accumulated temperature of greater than 0℃ or 10℃ (Fig. 5 and Table 3). Using the average annual temperature and an accumulated temperature greater than 0°C, we investigated the effects of temperature on cultivated land gridding results. We found that 85.86% of the grid exhibited relative differences from –10%–10%. In the Suwan region, temperature changes had little effect on farmland grid-processing results. Therefore, in some cases, the temperature factor can be omitted to simplify the model.

3.2.3 Influence of different land distribution methods near rivers and lakes

The SWCD and HYDE datasets differed in terms of spatial distribution (Fig. 6). SWCD data had lower reclamation rates in the three regions: mountainous areas of

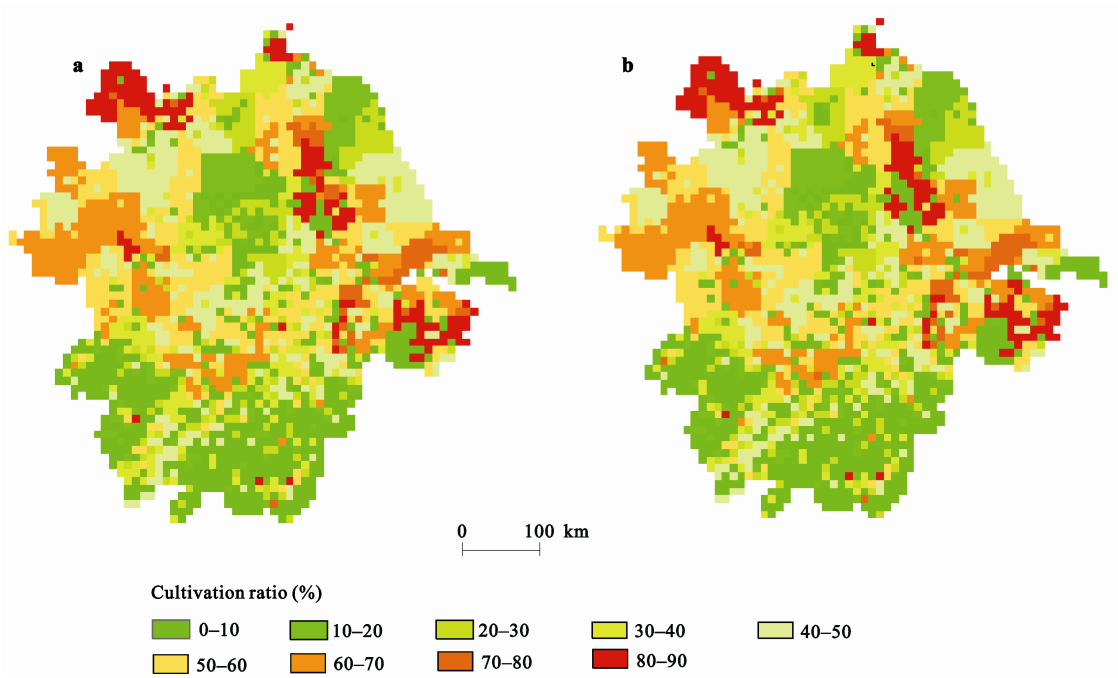

Fig. 5 Comparison on grid distribution of cultivated land using model parameters of 0℃ accumulated temperature (a) and annual average temperature (b)

Table 3 Comparison of reclamation rates between two models

Model	Average reclamation rate (%)					Standard deviation
	0– 20	20–40	40–60	60–80	80–90	
Model 1	5.4825	30.7298	49.0990	65.7169	87.8119	25.98
Model 2	5.3166	31.0482	49.3074	65.7471	85.3878	26.28

Note: Model 1 refers to the gridding model using 0℃ accumulated temperature. Model 2 refers to the model considering average annual temperature

the southern Suwan area; the abandoned Yellow River Estuary; and the region around Hongze Lake, including Sixian, Sihong, Xuyi, and Mingguang, but the reclamation rate of other plains was higher. The HYDE dataset's reclamation rates were higher on both sides of the ancient Yellow River and Yangtze River, while the reclamation rates of the mountains were low. In northwestern Anhui Province and the coastal areas of Jiangsu Province, the natural conditions were better, but the reclamation rates were low. This was mainly due to different treatment methods of cultivated land distribution near rivers and lakes; the HYDE dataset assumed that coastal and river plains were more conducive to farmland reclamation and therefore allocated more farmland area near water.

The relative difference in cultivated land area between the SWCD and HYDE datasets is shown in Fig. 6. In the HYDE dataset, 70% more cultivated land was distributed along river coastlines, in the Old Yellow River Estuary and in the areas surrounding Hongze Lake, than in the SWCD dataset. The plains far from rivers and lakes in the HYDE dataset were allocated at least 20% less arable land area.

3.3 Accuracy with different grid sizes

This study distributed cultivated land into 1 km × 1 km and 10 km × 10 km grids for the Jiangsu and Anhui provinces, respectively, according to the existing grid-processing model and then compared the results with remote-sensing data from 2000 as a representative example to determine accuracy based on grid size (Fig. 7).

To quantitatively evaluate the difference between the results with 10 km × 10 km and 1 km × 1 km grids, we calculated the absolute difference in area of the two units in the model distribution data and statistically analyzed the real spatial distribution of the remote-sensing data. Of the 1 km^2 grids, 91.42% exhibited an absolute difference in cultivated land area from –0.2–0.2 km^2, while 69.94% exhibited an absolute difference within 0.1 km^2 (Fig. 8a). Of the 10 km × 10 km grids, 70.46% exhibited an absolute difference per unit land area from –0.2–0.2 km^2, while 44.94% had an absolute difference concentrated within the –0.1–0.1 km^2 range (Fig. 8b). Hence, a 1 km^2 grid can improve data accuracy over a 10 km^2 grid, and the accuracy of grids with an absolute difference per unit area within 0.1 km^2 can be improved by 25%.

To evaluate the consistency of the effects of grid size on accuracy during historical and contemporary periods, the absolute differences in different grid sizes per land unit area (10 km × 10 km and 1 km × 1 km) in 2000 and 1735 were calculated (Fig. 9). In 1735, 70.55% of the grids exhibited an absolute difference between different grid sizes per unit land area within the range of –0.1–0.1 km^2, while in 2000, only 46.93% of the grids exhibited an absolute difference within this range. Therefore, the impact of grid size on accuracy is greater for modern versus historical data.

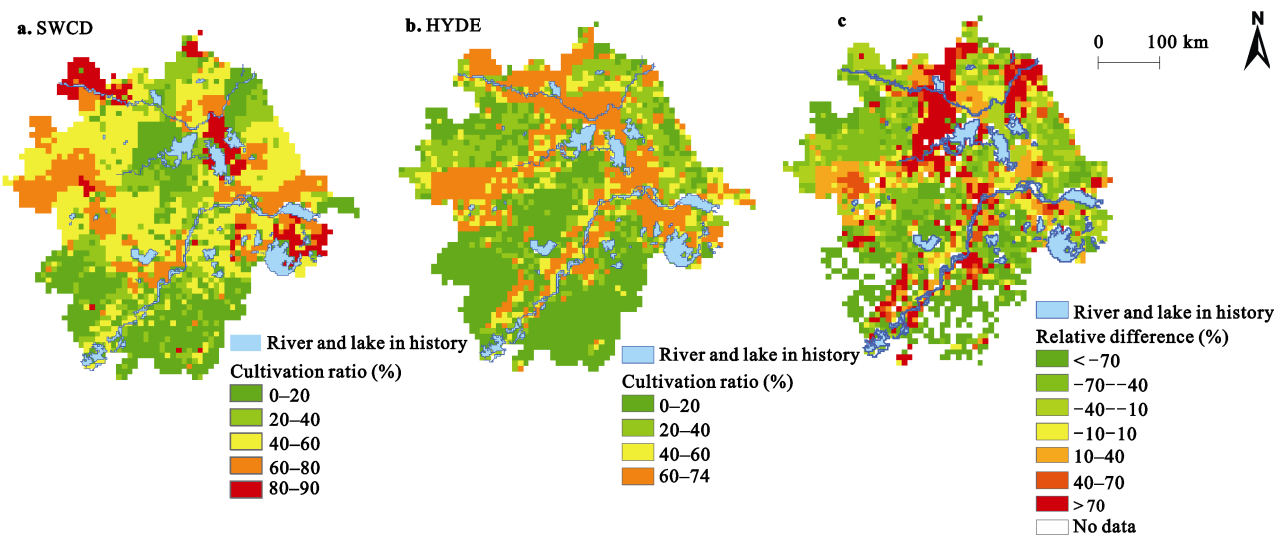

Fig. 6 Relative differences between SWCD and HYDE cultivated land data in a 10 km × 10 km grid for the Suwan region in 1735

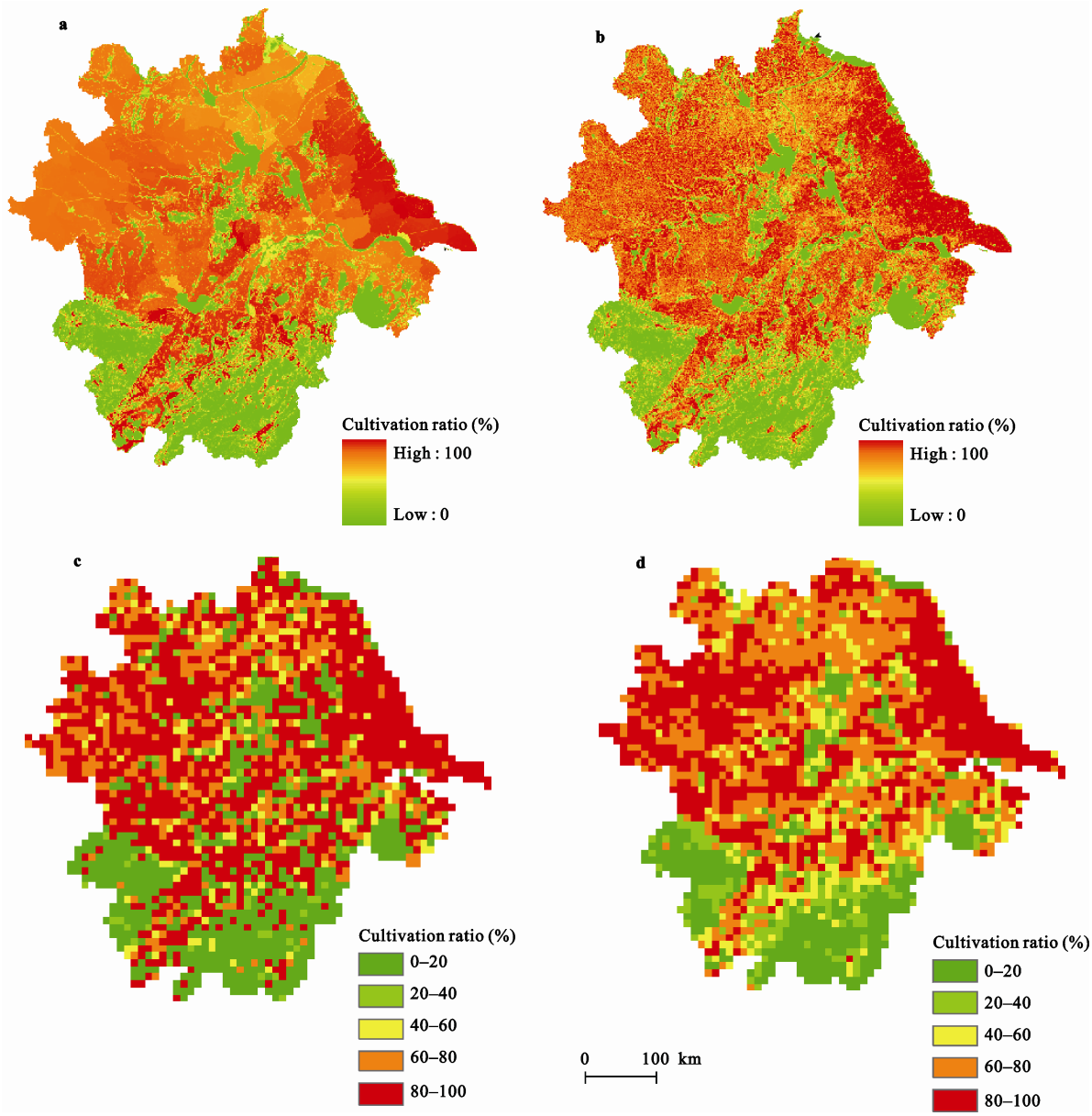

Fig. 7 Grids of 1 km × 1 km and 10 km × 10 km of arable land area in 2000 (a, c) and remote-sensing data (b, d)

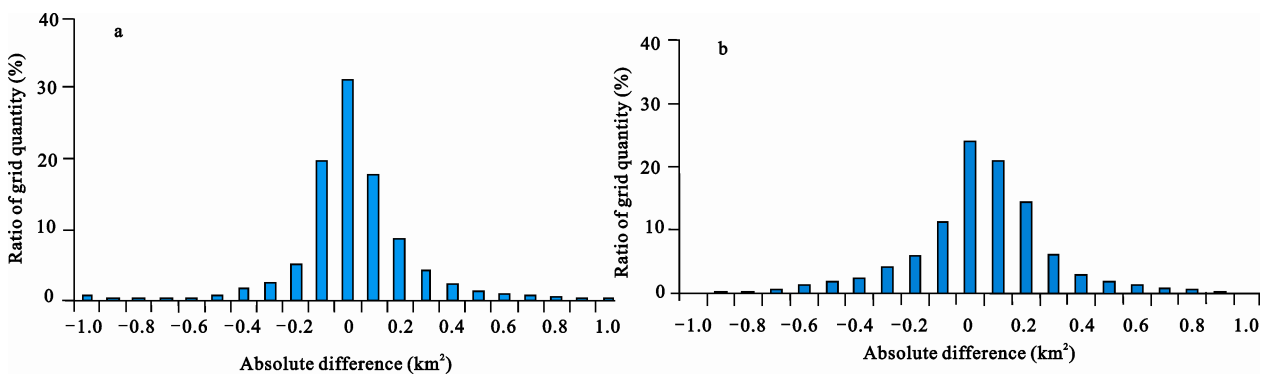

Fig. 8 Absolute difference histograms per unit land area of arable land gridding results with real distribution (a: grid size of 1 km × 1 km; b: grid size of 10 km × 10 km)

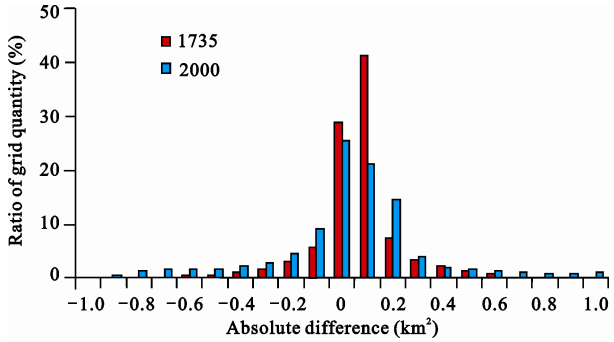

Fig. 9 Absolute difference in grid-processing results for different grid sizes per unit land area

4 Discussion

Based on the above results, we conclude that the spatial resolution of source data, impact factor selection on grid model and the size of grids all influence the accuracy of gridded historical cultivated land datasets in the Jiangsu and Anhui provinces. Furthermore, these factors might represent the primary limitation of global land datasets applied at the regional scale. Present popular international global or large-scale reconstruction methods should be modified when used for reconstruction of Chinese historical regional-scale cultivated land cover. Comparing our results with those from other similar studies, we further consider whether these influences also existed in the reconstructions of other regions in China or in China as a whole. Maybe such considerations should inspire researchers to develop better modified methods for regional scale reconstruction, especially for China.

First, other similar studies also presented the influence of spatial resolution of source data on the accuracy of historical cultivated land datasets similar to this study. He *et al.* (2012) found that the trends in land reclamation in China in the SAGE dataset sharply contrasted with those pertaining to regional data in CHCD. There was also a significant discrepancy in spatial distribution between the HYDE dataset and Chinese historical arable land records. Similar questions also exist in datasets in Northeast China. There are marked disagreements on cropland areas and distribution patterns between the HYDE dataset and CNEC data, especially in the 18th and 19th centuries. The HYDE dataset cannot reflect the actual historical reclamation progress in Northeast China (Li *et al.*, 2010). This research further analyzes the influence of different spatial resolutions of provincial and county-level data in Jiangsu and Anhui provinces. When those are all downscaled to the 10 km × 10 km grid for comparison, they have an average difference rate of 16.61%, and the larger rate (>70% or ≤–70%) occupies 13.3% and is located primarily within the Hongze Lake Basin and northern Jiangsu along the Yangtze River Plain. This proves that the model hypothesis of linear trend in international land datasets is incongruent with the actual land reclamation process, at least in China and especially at the finer regional scale. Applying county-level data sourced from historical records that relatively reflect authentic cultivated land cover changes for gridded datasets would be more dependable compared to the use of coarser-resolution data. Combining the model-based methods with historical empirical data may be a better way to improve the accuracy of regional scale datasets.

Second, previous similar studies have selected different impact factors in the grid model. Initially, researchers thought that the topography was the most important factor for spatial allocation of cropland at the regional scale. Then, the climate factor, modern arable land reclamation rate and water factor were added to make the model more complex and accurate (i.e., closer to the actual allocation). However, this study evaluates the influence of factor selection on the grid model, which most likely creates a regional discrepancy. Lin *et al*. (2009) and He *et al*. (2011) selected topographic and demographic factors, while Li *et al*. (2012) and Luo *et al*. (2014) selected the potential arable land reclamation rate, topography factors, climate production potential, and water factors. In the land-grid model of Jiangsu and Anhui in this study, natural factors mainly comprise the potential arable land reclamation rate, terrain factors (elevation and slope), climatic factors, and water factors. We found that use of the potential arable land reclamation rate of the Suwan region can increase the spatial difference in cultivated land distribution, and the standard deviation of the model increased by 4.05. In this area, temperature had little influence on the model, and more than 90% of the grid had a relative difference within 10% with the consideration of annual mean temperature. In addition, the difference in spatial distribution of cropland between HYDE and SWCD data was concentrated in areas along the Yangtze River canal area and coastal areas, especially during the Qing Dynasty and the Republic of China. This mainly resulted from

the model assumption in HYDE that the river is primarily used for agricultural irrigation, such that more cropland was allocated there, while in fact it was mainly utilized for transporting grain at that time. Another cause is that the water system has developed over many hundreds of years. In this study, we apply period-specific water system routes but not modern water system maps, as was used in HYDE, in the gridding process, which undoubtedly improved the accuracy of grid datasets.

Third, past research has used different grid sizes, such as Li *et al*. (2010) using land data in Northeast China with a grid size of 0.5° × 0.5°, Lin *et al*. (2009) using traditional Chinese agricultural land data with a grid size of 60 km × 60 km, Li *et al.* (2016) using a grid size of 10 km × 10 km, and Luo *et al*. (2014) using a grid size of 2 km × 2 km. This study quantitatively evaluated the difference between the results with the 10 km × 10 km and 1 km × 1 km grids. The results show that a 1 km^2 grid can improve the accuracy per unit area by 25%. Grid size selection should be fit for the size of geomorphic units. Due to different intensities of land reclamation at different times, the effects of grid size on data may vary. Therefore, in the choice of grid size, the impact of landform units, land use patch size, and other factors should be considered.

5 Conclusions

This study quantitatively analyzed the influence of different data sources, grid-processing methods, and grid sizes on accuracy of gridded cropland data in the Jiangsu and Anhui provinces. The applicability of approaches to historical regional-scale cropland cover reconstruction was assessed and some related suggestions were given. The main findings are as follows:

(1) The temporal trends of the HYDE, CHCD, and SWCD datasets with various data sources were more similar. However, different spatial resolutions of cropland source data in the CHCD and SWCD datasets revealed an average difference of 16.61% when provincial and county data were downscaled to the 10 km × 10 km grid for comparison.

(2) The inclusion of potential arable land reclamation rate can increase the difference in the spatial distribution of cropland area allocation; however, the standard deviation only increased by 4.05 in this study. Considering temperature factors, more than 90% of the grid exhibited relative differences within 10%, and over 90% of the grid had absolute differences within 2 km^2. Due to the assumption that there was more arable land close to rivers and lakes, the HYDE dataset allocated 70% more of the grid near river coastlines, the abandoned Yellow River Estuary, and the areas surrounding Hongze Lake than did the SWCD dataset.

(3) Grid processing with the grid size of 1 km^2 improved the accuracy of gridded results over that of 10 km^2 grid size for modern cropland data in 2000, with the absolute difference in the unit land area within 0.1 km^2 improved by 25%, which was much more pronounced in modern times than in historical periods.

(4) In conjunction with the outcomes of other similar studies, this paper proves that some model hypotheses and grid-processing methods in international land datasets truly do not fit with the actual land reclamation process, at least in China and especially at the finer regional scale. Combining the model-based methods with historical empirical data may be a better way to improve the accuracy of regional scale datasets.

Of course, continuing study will present new challenges on availability of accurate historical data on arable land with multiple time sections and high resolution, consideration of the influence of manmade factors such as changes in science, technology, and policy and so on.

References

Feng Yongheng, Zhang Shihuang, He Fanneng *et al.*, 2014. Separate reconstruction of Chinese cropland grid data in the 20th century. *Progress in Geography*, 33(11): 1546–1555. (in Chinese)

Fuchs R, Herold M, Verburg P H *et al.*, 2012. A high-resolution and harmonized model approach forreconstructing and analyzing historic land changes in Europe. *Biogeosciences Discussions*, 9(10): 14823–14866. doi: 10.5194/bgd-9-14823-2012

Ge Quansheng, Dai Junhu, He Fanneng *et al.*, 2003. Analysis of the quantitative change of cultivated land resources in China in the past 300 years and the driving factors analysis. *Progress in Natural Science*, 13(8): 825–832. (in Chinese)

Goldewijk K K, Beusen A, Van Drecht G *et al.*, 2011. The HYDE 3.1 spatially explicit database of human induced global land-use change over the past 12000 years. *Global Ecology and Biogeography*, 20(1): 73–86. doi: 10.1111/j.1466-8238.2010.00587.x

Hall C A S, Tian H, Qi Y *et al.*, 1995. Modelling spatial and temporal patterns of tropical land use change. *Biogeography*

Journal, 22(4/5): 753–757. doi: 10.2307/2845977

He Fanneng, Li Shicheng, Zhang Xuezhen, 2011. The reconstruction of cropland area and its spatial distribution pattern in the Mid-northern Song Dynasty. *Acta Geographica Sinica*, 66(11): 1531–1539. (in Chinese)

He Fanneng, Li Shicheng, Zhang Xuezhen, 2012. Comparisons of reconstructed cropland area from multiple datasets for the traditional cultivated region of China in the last 300 years. *Journal of Geographical Sciences*, 67(9): 1190–1200. (in Chinese)

He Fanneng, Li Shicheng, Zhang Xuezhen, 2014. Spatially explicit reconstruction of forest cover of Southwest China in the Qing Dynasty. *Acta Geographica Sinica*, 33(2): 260–269. (in Chinese)

Kaplan J O, Krumhardt K M, Ellis E C *et al*., 2011. Holocene carbon emissions as a result of anthropogenic land cover change. *The Holocene*, 21(5): 775–791. doi: 10.1177/0959683610386983

Li B B, Jansson U, Ye Y *et al.*, 2013. The spatial and temporal change of cropland in the Scandinavian Peninsula during 1875–1999. *Regional Environmental Change*, 13(6): 1325–1336. doi: 10.1007/s10113-013-0457-z

Li Beibei, Fang Xiuqi, Ye Yu *et al.*, 2010. The global land use data set the accuracy of regional assessments, in the northeast of China as an example. *Science China Earth Sciences*, 40(8): 1048–1059. (in Chinese)

Li Shicheng, He Fanneng, Chen Yisong, 2012. Gridding reconstruction of cropland spatial patterns in Southwest China in the Qing Dynasty. *Progress in Geography*, 31(9): 1196–1203. (in Chinese)

Li Shicheng, He Fanneng, Zhang Xuezhen, 2014. An approach of spatially-explicit reconstruction of historical forest in China: a case study in Northeast China. *Acta Geographical Sinica*, 69(3): 312–322. (in Chinese)

Li Shicheng, He Fanneng, Zhang Xuezhen, 2016. A spatially explicit reconstruction of cropland cover in China from 1661 to 1996. *Regional Environmental Change*, 16(2): 417–428. doi: 10.1007/s10113-014-0751-4

Lin Shanshan, 2007. *A Study on Cropland Gridding Data Recon-Struction over Chinese Traditional Agricultural Area in Qing Dynasty*. Beijing: University of Chinese Academy of Sciences. (in Chinese)

Lin Shanshan, Zheng Jingyun, He Fanneng, 2009. Gridding cropland data reconstruction over the agricultural region of China in 1820. *Journal of Geographical Sciences*, 19: 36–48. doi: 10.1007/s11442-009-0036-x

Long Ying, Jin Xiaobin, Li Miaoyi *et al.*, 2014. A constrained cellular automata model for reconstructing historical arable land in Jiangsu province. *Geographical Research*, 33(12): 2239–2250. (in Chinese)

Luo Jing, Chen Qiong, Liu Fenggui *et al.*, 2015. Methods for reconstructing historical cropland spatial distribution of the Yellow River-Huangshui River valley in Tibetan Plateau. *Progress in Geography*, 34(2): 207–216. (in Chinese)

Luo Jing, Zhang Yili, Liu Fenggui *et al.*, 2014. Reconstruction of cropland spatial patterns for 1726 on Yellow River-Huangshui River Valley in northeast Qinghai-Tibet Plateau. *Geographical Research*, 33(7): 1285–1296. (in Chinese)

Ramankutty N, Foley J A, 2010. ISLSCP II historical croplands cover, 1700–1992. In: Hall F G *et al*. (eds.). *ISLSCP Initiative II Collection.* Tennessee: Oak Ridge National Laboratory Distributed Active Archive Center, 1–20. doi: 10.3334/ORNLDAAC/966

The agriculture of China, Anhui volume editor committee, 1998. *The Agriculture of China*, *Anhui Volume*. Beijing: China Agriculture Press. (in Chinese)

The agriculture of China, Jiangsu volume editor committee, 1998. *The Agriculture of China*, *Jiangsu Volume*. Beijing: China Agriculture Press. (in Chinese)

Wei Xueqiong, Ye Yu, Cui Yujuan *et al.*, 2014. Review of China's historical land cover change reconstructions. *Advances in Earth science*, 29(9): 1037–1045. (in Chinese)

Yang Xuhong, Guo Beibei, Jin Xiaobin *et al.*, 2015. Reconstructing spatial distribution of historical cropland in China's traditional cultivated region: methods and case study. *Chinese Geographical Science*, 25(5): 629–643. doi: 10.1007/s11769-015-0753-2

Ye Y, Fang X Q, Ren Y Y *et al.* 2009. Cropland cover change in Northeast China during the past 300 years. *Science in China Series D*: *Earth Sciences*, 52(8): 1172–1182. doi: 10.1007/s11430-009-0118-8

Yuan Cun, 2015. *Gridding and Accuracy Comparison of Cropland Data in Jiangsu and Anhui Provinces During the Past 300 years.* Beijing: Beijing Normal University. (in Chinese)

Yuan Cun, Ye Yu, Fang Xiuqi, 2015. Rasterizing cropland data and accuracy comparison in Jiangsu and Anhui ProVinces in the Mid-Qing Dynasty. *Progress in Geography*, 34(1): 83–91. (in Chinese)

Zhang Jiacheng, 1982. Possible impact of climate variation on agricultural in China. *Geography Research*, 1(2): 8–15. (in Chinese)

Zhang Lijuan, Jiang Lanqi, Zhang Xuezhen *et al.*, 2014. Reconstruction of cropland over Heilongjiang Province in the late 19th century. *Acta Geographica Sinica*, 69(4): 448–458. (in Chinese)

Zhang Xuezhen, He fanneng, Li Shicheng, 2013. Reconstructed cropland in the mid-eleventh century in the traditional agricultural area of China: implications of comparisons among datasets. *Regional Environmental Change*, 13(5): 969–977. doi: 10.1007/s10113-012-0390-6.

Zhao Yun, 2005. *Jiangsu and Anhui Area Land Use and its Driving Mechanism.* Shanghai: Fudan University. (in Chinese)

Zhu Feng, Cui Xuefeng, Miu Lijuan, 2012. China's spatially-explicit historical land-use data and its reconstruction methodology. *Progress in Geography*, 31(12): 1563–1573. (in Chinese)