J. Geogr. Sci. 2013, 23(6): 978-990
DOI: 10.1007/s11442-013-1057-z

Comparisons of cropland area from multiple datasets over the past 300 years in the traditional cultivated region of China

HE Fanneng[1], LI Shicheng[1,2], ZHANG Xuezhen[1], GE Quansheng[1], DAI Junhu[1]

1. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Land use/cover change is an important parameter in the climate and ecological simulations. Although they had been widely used in the community, SAGE dataset and HYDE dataset, the two representative global historical land use datasets, were little assessed about their accuracies in regional scale. Here, we carried out some assessments for the traditional cultivated region of China (TCRC) over last 300 years, by comparing SAGE2010 and HYDE (v3.1) with Chinese Historical Cropland Dataset (CHCD). The comparisons were performed at three spatial scales: entire study area, provincial area and 60 km by 60 km grid cell. The results show that (1) the cropland area from SAGE2010 was much more than that from CHCD; moreover, the growth at a rate of 0.51% from 1700 to 1950 and –0.34% after 1950 were also inconsistent with that from CHCD. (2) HYDE dataset (v3.1) was closer to CHCD dataset than SAGE dataset on entire study area. However, the large biases could be detected at provincial scale and 60 km by 60 km grid cell scale. The percent of grid cells having biases greater than 70% (<–70% or >70%) and 90% (<–90% or >90%) accounted for 56%–63% and 40%–45% of the total grid cells respectively while those having biases range from –10% to 10% and from –30% to 30% account for only 5%–6% and 17% of the total grid cells respectively. (3) Using local historical archives to reconstruct historical dataset with high accuracy would be a valuable way to improve the accuracy of climate and ecological simulation.

Keywords: cropland datasets; comparisons; past 300 years; traditional cultivated region; China

1 Introduction

On account of land use/cover changes, the land surface properties, including albedo, green area fraction, roughness, and plant phenology were modified. Through the biogeophysical (e.g., Findell *et al.*, 2009; Pitman *et al.*, 2011) and biogeochemical (e.g., Shevliakova *et al.*, 2009; Houghton *et al.*, 2012) effects on atmosphere, these conversions of the land cover have important implications for climate changes. For instance, human-induced land cover conversions from natural forest to cropland have potential to emit CO_2 to atmosphere and

Received: 2013-05-06 **Accepted:** 2013-06-02
Foundation: China Global Change Research Program, No.2010CB950901; National Natural Science Foundation of China, No.41271227; No.41001122
Author: He Fanneng (1963–), specialized in historical geography and environmental changes. E-mail: hefn@igsnrr.ac.cn

thus may lead to climate warming. Nowadays, numerical simulations are popular ways to research climatic effects of human activities. Therefore, as an important driving force of climate changes, human-induced land cover change is considered as an important boundary conditions of climate models and ecology system models (Thompson, 2000; Feddema *et al*., 2005; Foley *et al*., 2011). Historical land use/cover change (LUCC), especially in the past 300 years, is necessary for us to understand past climate change, diagnose climate formation mechanism, assess the sensitivity of the climate system to nature and human forcing and predict future climate change. As a result, studies on LUCC over the past 300 years received extensive attentions with the development of the global environmental change research (e.g., Shi *et al*., 2007; Voldoire *et al*., 2007).

To reveal historical LUCC in regional and global scales over the past 300 years, a large number of studies have been carried out (e.g., Bicik *et al*., 2001; Ge *et al*., 2004; Hamre *et al*., 2007; Ye *et al*., 2009; Zhang *et al*., 2011). Among these studies, the historical LUCC datasets created by the Center for Sustainability and the Global Environment (SAGE) (Ramankutty and Foley, 1999), University of Wisconsin (USA) and the History Database of the Global Environment (HYDE), Netherlands Environmental Assessment Agency (Klein Goldewijk, 2001) were most widely used in climate simulations (e.g., Matthews *et al*., 2003; Brovkin *et al*., 2004; Forster *et al*., 2007). For instance, using reconstructed croplands (Ramankutty and Foley, 1999) and pasturelands (Klein Goldewijk, 2001), Betts *et al*. (2007) simulated a global mean radiative forcing of –0.18 W m^{-2} relative to 1750 and –0.24 W m^{-2} relative to potential natural vegetation (PNV). Li *et al*. (2006) and Chen *et al*. (2006) studied the influence of land cover change from 1700 to 1990 on regional climate of China using HYDE2.0 dataset. However, as indicated by SAGE, the dataset captures the general patterns of cropland change over history, but not the fine details at local to regional scales and it should only be used for continental-to-global scale analysis and modeling (Ramankutty and Foley, 2010). Besides, by comparing these datasets with records from local historical documents, Li *et al*. (2010) found that the SAGE dataset overestimated cropland for Northeast China by 20.98 times in 1700 and 1.6 times in 1990 and the HYDE dataset did not reproduce the northward expansion of cropland across Northeast China over the last 300 years as well. Zhang *et al*. (2013) reported HYDE dataset could not exhibit spatial distribution of cropland in the mid-11th century as recorded by local historical documents. In a word, although much effect has been put into quantifying land cover change of the past 300 years, many uncertainties still exist, especially the uncertainties of global datasets in regional scales.

In this study, we aimed to evaluate the accuracy of SAGE and HYDE datasets in the traditional cultivated region of China (hereafter, TCRC for short) by comparing them with Chinese Historical Cropland Dataset (hereafter, CHCD for short) created using Chinese local historical documents. The evaluations were carried out at three spatial scales: entire study area, provincial scale and 60 km by 60 km grid cell scale. The study area of this paper, TCRC, is consisting of 18 modern provinces. Such evaluations could provide scientific insights on accuracy of SAGE and HYDE in TCRC and thus would be helpful for understanding uncertainties of simulations using the SAGE and HYDE datasets.

2 Data and methods

2.1 Data sources

There are three available cropland datasets, SAGE, HYDE and CHCD, covering TCRC over the past 300 years. The characteristics and reconstruction methods of these datasets are described as follows (Table 1).

Table 1 Information of SAGE2010, HYDE3.1 and CHCD datasets

Datasets	Reconstruction method	Land use type	Covering period	Temporal resolution	Spatial resolution
SAGE2010	Global map of cultivated land for 1992 and a 'hind-cast' modelling technique	Cropland	1700–1992	1–50 a	0.5°/1°
HYDE3.1	The mix of two weighting maps: a current map and a historical map	Cropland/ pasture	10000 BC–AD 2000	10–100 a	5′
CHCD	A cropland allocating model integrated slope and population density	Cropland	1661–1999	6–60 a	60 km

(1) SAGE dataset. This dataset describes the cropland area fraction of each grid cell from 1700 to 1992. It was first released in 1999 (named as SAGE1999 hereafter) (Ramankutty and Foley, 1999). In 2010, using more collections of ground-based data, the SAGE1999 was updated to SAGE2010 (Ramankutty and Foley, 2010). SAGE2010 has two spatial resolutions: one degree and half degree, and varied temporal intervals: 50-year for the period 1700–1850, 10-year for 1850–1980, and annual for 1986–1992. In this study, the SAGE2010 dataset with the half degree of resolution was assessed. The SAGE dataset was created by the following approach: using national-level cropland area from inventories in 1992 to calibrate the satellite-based DISCover dataset to obtain the spatially explicit and accurately quantitative cropland; then, keeping the relative spatial weights unchanged, to allocate the historical national-level cropland area into each grid cell; finally, spatially explicit historical cropland area dataset was achieved (see Ramankutty and Foley, 1999, for details).

(2) HYDE dataset. This dataset describes the cropland area (unit: km^2) of each grid cell. Since 1997 when HYDE dataset (v1.1) was firstly released (Klein Goldewijk and Battjes, 1997), it has been updated several times. The latest version is HYDE dataset (v3.1) (Klein Goldewijk *et al*., 2011). HYDE dataset (v3.1) has a spatial resolution of 5′ longitude/latitude and varied temporal intervals: 1000-year for the 10000BC–1BC, 100-year for the AD1–1700 and 10-year for AD 1700–2000. In this dataset, the original cropland area data was quantitatively estimated using the contemporary population and per capita cropland area for the period pre-1960 and was from FAO (2008) for the post-1961. The national/provincial-level cropland area was allocated spatially to grid cells using a mix of two weighting maps: a current map, which was constructed from a satellite map of AD 2000 (Klein Goldewijk and van Drecht 2006), and a historical map, which was constructed based on six rules by considering the effects of population density, temperature, land suitability for crops, distance to water, surface slope, and urban development on agricultural activities (for details, see Klein Goldewijk *et al*., 2011). In this paper, the HYDE3.1 was assessed.

(3) CHCD dataset. This dataset describes cropland area fraction for TCRC over the period from 1661 to 1999. It has a spatial resolution of 60 km by 60 km and varied temporal intervals. The original provincial cropland area data used to create the CHCD was collected and calibrated by Ge *et al*. (2004). The spatial allocation of historical cropland was carried out by Lin *et al*. (2009). The core idea of its allocating method is considering surface slope and

population density as main factors determining the spatial distribution of cropland (see Lin *et al*., 2009, for details).

2.2 Data processing

SAGE, HYDE, and CHCD datasets have different temporal intervals and spatial resolutions. In order to facilitate comparison, they were processed as follows.

(1) Updating of the CHCD dataset. For the post-1949, the original cropland area data used by CHCD was released by the National Bureau of Statistics (2000). This dataset was proofed to have a low level confidence (Crook, 1993; IIASA, 1999), especially for the period from 1960 to 1985. Therefore, in order to reflect the changes in cropland area in TCRC objectively, we updated the cropland area data post-1960 in CHCD using the results from Feng *et al*. (2005), the detailed national land use ground survey in 1996[1] and the cropland area change data released by the Ministry of Land and Resources of China (Liu, 2000).

(2) Unifying of the spatial resolutions. The spatial resolutions of SAGE2010 and HYDE3.1 are 0.5° and 5′ longitude/latitude, respectively, while CHCD has a spatial resolution of 60 km by 60 km. We firstly transformed the map coordinate of the SAGE2010 and HYDE3.1 from geographic coordinate (longitude/latitude) to Albers projection and then we resampled them to a cell size of 60 km by 60 km. Third, for SAGE2010 and CHCD, we calculated the cropland area of each grid cell via multiplying by the grid cell's area, 3600 km^2. Finally, for all the three datasets, we aggregated cropland area of all grid cells within a province to get the provincial totals and then get the total cropland area of the whole study area.

(3) Readjusting of the provincial administrative units. Over the past 300 years, along with national regimes changes from the Qing Dynasty to the Republic of China, and to the People's Republic of China, the provincial administrative units were changing. In order to use the data from the three periods, we readjusting the provincial administrative units using the current provincial administrative units as reference. 1) Merging Beijing, Tianjin and Hebei as Jing-Jin-Ji area; 2) merging Shanghai and Jiangsu as Hu-Ning area; 3) merging Chongqing and Sichuan as Chuan-Yu area; 4) merging Hainan and Guangdong as Yue-Qiong area; 5) merging Ningxia and Gansu as Gan-Ning area.

(4) Selecting of the hot temporal slices. As the temporal extent and interval of the three datasets are not uniform, we selected 14 time slices (years), which are 1700, 1720, 1780, 1820, 1870, 1890, 1910, 1930, 1950, 1960, 1970, 1980, 1990 and 2000, to evaluate HYDE3.1 and SAGE2010. HYDE3.1 is available for all the above time slices. And the data of 1720, 1780 and 1820 in SAGE (2010) were obtained by linear interpolation using the adjacent years' data. For CHCD, the data of 1685, 1724, 1784, 1873, 1893, 1913 and 1933 directly correspond to 1700, 1720, 1780, 1870, 1890, 1910 and 1930, respectively.

2.3 Analysis method

CHCD was created using local historical documents so the changes in cropland revealed by CHCD were most close to real history among the available three datasets. In this study, we used CHCD as real history to evaluate the SAGE and HYDE datasets. We exhibited their

[1] Office of the National Agricultural Zoning. China's agricultural land resources decadal analysis and evaluation(1986-1995), 1997.

bias through three aspects: absolute bias for each time slice (Eq. 1), relative bias (%) for each time slice (Eq. 2) and the growth rate of cropland area over the past 300 years (Eq. 3).

$$E_a = X_{It} - X_{Ct} \tag{1}$$

$$E_b = \left(X_{It} - X_{Ct}\right) / X_{Ct} \times 100\% \tag{2}$$

$$Q = \left(\sqrt[t_2 - t_1]{X_{t_2} / X_{t_1}} - 1\right) \times 100\% \tag{3}$$

where E_a and E_b are the absolute and relative biases of HYDE and SAGE dataset; X_{It} denotes cropland area at time point t from HYDE or SAGE and X_{Ct} for CHCD; Q is the annual growth rate of cropland area from t_1 to t_2; X_{t_1} and X_{t_2} are the cropland area at time points of t_1 and t_2, respectively.

3 Results

Corresponding to the above-mentioned comparison performances individually at entire study area scale, provincial scale and grid cell scale, we described the results separately as follows.

3.1 Total cropland of entire study area

3.1.1 CHCD dataset

As illustrated by Figure 1, the CHCD dataset reveals that cropland area of TCRC over the past 300 years had an increasing trend. The cropland increased by 10.64×10^6 hm^2 from 1724 to 1887 with the annual rate of ~0.1%. Due to natural disasters, wars and other reasons, the increasing trend was replaced by a decreasing trend from 1887 to 1949. It decreased by 6.03×10^6 hm^2 with the annual rate of –0.12%. Since 1949, the cropland area grew rapidly with fluctuations at an annual growth rate of 0.43% and reached a peak of 100.3×10^6 hm^2 in the 1980s. From 1980 to 2000, however, the cropland area reduced slowly and it was 94.7×10^6 hm^2 in 2000.

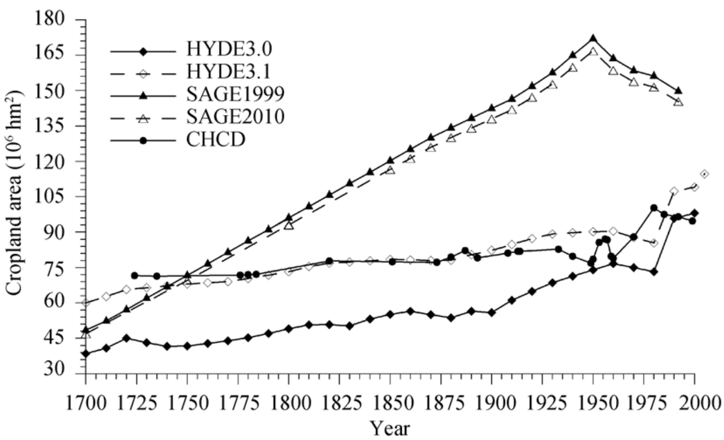

Figure 1 The cropland area of TCRC from 1700 to 2000 as revealed by SAGE, HYDE and CHCD

3.1.2 SAGE dataset

There are large discrepancies between the total cropland area of TCRC in the SAGE dataset and in CHCD dataset (Figure 1). Although the cropland area from SAGE2010 was reduced by ~3% in each time slice compared with that from SAGE1999, it still had great biases from the CHCD dataset. The SAGE2010 dataset showed that the cropland area of TCRC was ~47×10^6 hm^2 in 1700. It increased linearly at an annual rate of 0.51% since 1700 and reached the peak of 166.7×10^6 hm^2 in 1950 which was about 3.6 times more than that in 1700. After 1950, it declined at an annual rate of 0.34% (Figure 1). From the perspective of relative biases, the cropland area from the SAGE dataset was over 50% more than that from CHCD dataset (Table 2). Moreover, the relative biases increased as time went by and it reached up to 112.2% in 1950. Although the relative biases declined gradually since then, it was still as high as 51.1% even in 1980 when the cropland area of CHCD reached the historical peak. In one word, SAGE2010 overestimated the cropland area of TCRC and the variations in cropland area from SAGE couldn't reflect the real situations of land cultivation history as well.

Table 2 Cropland area of TCRC and its relative biases among HYDE3.1, SAGE 2010 and CHCD datasets

Years	CHCD	SAGE(2010)		HYDE3.1	
	Cropland area (10^6 hm^2)	Cropland area (10^6 hm^2)	Relative biases (%)	Cropland area (10^6 hm^2)	Relative biases (%)
1700	59.29	46.97	–20.8	59.98	1.2
1720	71.56	55.99	–21.8	65.65	–8.3
1780	72.15	83.63	15.9	70.35	–2.5
1820	77.75	102.43	31.7	76.82	–1.2
1870	77.17	125.94	63.2	77.95	1.0
1890	79.08	134.04	69.5	80.35	1.6
1910	81.84	141.93	73.4	84.79	3.6
1930	82.68	152.74	84.7	89.20	7.9
1950	78.57	166.71	112.2	90.21	14.8
1960	78.85	158.58	101.1	90.43	14.7
1970	87.92	153.74	74.9	87.72	–0.2
1980	100.26	151.47	51.1	85.35	–14.9
1990	96.51	145.50	50.8	107.29	11.2
2000	94.70	–	–	109.00	15.1

3.1.3 HYDE dataset

The cropland area of TCRC part in HYDE3.1 was much closer to CHCD than HYDE3.0 as the cropland area of HYDE3.1 increased by 41.6% from that of HYDE3.0. And the cropland area from HYDE3.1 was much closer to that from CHCD than the SAGE dataset too. The relative biases of HYDE3.1 were in an interval of –15% to 15% (Table 2). Among the fourteen time slices in comparison, nine time slices had relative biases less than 10% and the largest value of relative biases was merely 15.1%.

However, when it comes to the trends, the HYDE3.1 and CHCD are not close. Since the early 20th century, especially since 1960, due to the different data sources they used, the overall trends differed from each other. For example, in 1980, it was the relatively low value and the cropland area was 85.4×10^6 hm^2 in HYDE3.1, but it was high value and the cropland area was 100.3×10^6 hm^2 in CHCD. Since then, the cropland area of HYDE3.1 showed a rapid growth tendency while CHCD decreased slowly (Figure 1).

3.2 Provincial cropland area

3.2.1 Comparisons between SAGE2010 and CHCD

As illustrated by Figure 2, the variations in provincial cropland area were similar with total cropland area of entire study area. A linear increasing trend (~0.51% per year) from 1700 to 1950 and a linear decreasing trend (0.14%–0.44% per year) from 1950 to 1990 could be found for each province. Such patterns of variations were obviously different from local historical documents-based CHCD dataset. As shown by CHCD, in the Qing Dynasty and Republic of China, there was rapid cropland increase in Southwest China (Yunnan, Guizhou and Chuan-Yu). In Chuan-Yu area, the cropland increased at an annual rate of as high as 1.22% which might be the largest rate during the past 300 years. Such rapid increase in cropland area mainly resulted from a policy which changed territorial tribal chiefs into officials and huge immigration from other provinces. However, the cropland area of Shanxi, Gan-Ning area, and Henan grew slowly with fluctuations and it even decreased in Hu-Ning area. These spatial variabilities of cropland changes could not be represented by SAGE2010 dataset.

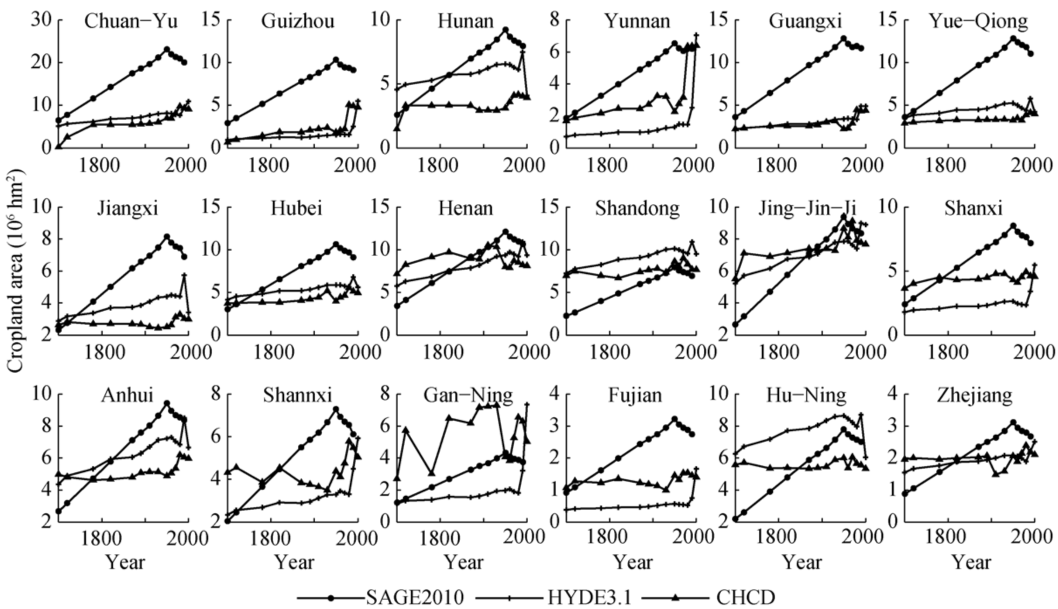

Figure 2 Provincial cropland area from 1700 to 2000 as revealed by HYDE3.1, SAGE2010 and CHCD

Additionally, the provincial cropland area amounts from SAGE2010 were mostly larger than that from CHCD. Among the 18 provinces, there are 12 provinces whose cropland area from SAGE2010 was much more than that from CHCD at each time slice over the past 300 years. The greatest discrepancies occurred in 1950 when the SAGE2010 reached its peak. Taking Chuan-Yu area as an example, the SAGE2010 had a cropland area of 23.12×10^6 hm^2 in 1950 which was more than 3 times of cropland area of 7.02×10^6 hm^2 in CHCD. In

Guangxi, SAGE2010 reached the peak in 1950, but it was a lower phase in CHCD. The SAGE2010 was higher by 478% than the CHCD. In addition, there are four provinces (i.e. Henan, Jing-Jin-Ji, Hu-Ning and Zhejiang) whose cropland area from SAGE2010 was more than that from CHCD before the mid-19th century and was less than that from CHCD after the mid-19th century. For instance, the cropland area from SAGE2010 was smaller than that in CHCD before 1870. The largest relative bias was about –52% which appeared in 1720. However, after 1870, it was almost greater than that in CHCD and the largest relative bias was ~52% which appeared in 1950. Besides, in Shandong and Gan-Ning area, the cropland area from SAGE2010 was lower than those from CHCD during the whole study period. For example, the cropland area in Shandong from SAGE2010 was 2.23×10^6 hm^2 in 1700 while it was 6.97×10^6 hm^2 in CHCD and the relative bias was –68%.

3.2.2 Comparisons between HYDE3.1 and CHCD

As shown by Figure 2, the HYDE3.1 exhibited that in each province the cropland area had an increasing trend over the last 300 years. Before the 1950s, the provincial annual growth rates ranged from 0.12% to 0.24%, which indicate a small spatial variability. Such extensive increases were consistent with CHCD. However, the HYDE3.1 didn't capture the large spatial variability of increasing trend as revealed by CHCD. In detail, the annual growth rates of cropland area in Chuan-Yu and Yunnan were 1.5% and 0.3%, respectively, but the cropland area in Fujian and Zhejiang reduced slightly at an annual rate of –0.03% and –0.1% respectively. After the 1950s, the discrepancy between the HYDE3.1 and CHCD enhanced. The different trends between the HYDE3.1 and CHCD could be found extensively. For instance, HYDE3.1 showed an increasing trend in Shanxi, Shaanxi, Gan-Ning and Fujian and a decreasing trend in Hunan, Hu-Ning and Jiangxi, all of which were conflicted with trends revealed by CHCD.

In terms of absolute cropland area, among the 18 provinces, there are five provinces (i.e. Chuan-Yu, Guangxi, Jing-Jin-Ji, Zhejiang and Henan) in which the cropland area from HYDE3.1 were close to those from CHCD. The relative biases ranged from –25% to 35%. For instance, in Guangxi and Jing-Jin-Ji the relative biases were –5% and 10% respectively in 1700, and they are –20% and 16% respectively in 1930, which is relatively low. There are seven provinces (i.e. Hunan, Hu-Ning, Guangdong, Jiangxi, Hubei, Shandong, and Anhui) in which cropland area from HYDE3.1 were generally larger than those from CHCD. Taking Hunan Province as an example, all of the relative biases in the 11 time slices were over 50% and the largest one occurring in 1700 was up to 211%. There are six provinces (i.e. Yunnan, Guizhou, Shaanxi, Shanxi, Gan-Ning, and Fujian) in which cropland area from HYDE3.1 was generally less than those from CHCD. For example, the cropland area of 13 time slices in Shanxi in HYDE3.1 was only about 50% of those in CHCD.

3.3 Comparisons and analysis on grid cell scale

The above-mentioned findings demonstrate that SAGE2010 very largely overestimates the cropland area of TCRC. It is invaluable to perform grid cell-based comparison between SAGE2010 and CHCD. Therefore, the following grid cell-based comparison was carried out only for HYDE3.1. Considering the availability of original data, we selected 1724, 1784, 1820, 1873 and 1911 to carry out the grid cell-based comparison. Since the available time slices of HYDE3.1 were not completely consistent with those of CHCD, the adjacent time slices, i.e. 1720, 1780, 1870 and 1910 of HYDE3.1 were selected.

3.3.1 Comparison of overall spatial pattern

Figure 3(a-j) illustrates the overall spatial pattern of cropland in the HYDE3.1 and CHCD datasets. We can find that the HYDE3.1 dataset generally has consistent spatial pattern with CHCD dataset. Both of them showed the North China Plain, Kuan-Chung Plain, the middle-lower Yangtze River and Sichuan basins were heavily developed while the southwest, the southeast coast and the Loess Plateau were slightly developed. However, there are differences in the quantitative development intensity and domain of development. Firstly, the HYDE3.1 revealed that the growth of cropland area over the past 300 years was mainly represented by increase in development intensity in the existing agriculutral areas. However, as shown by CHCD, along with the increase in deveopment intensity in the existing agriculutral area, the new agriculutral areas were gradually devloped. Secondly, HYDE3.1 has larger heavy development areas (reclamation rate>40%), such as Huang-Huai-Hai Plain and the middle-lower Yangtze River Basin, than CHCD and smaller slight development areas (reclamation rate<40%), such as Southwest China and southeast coastal area, than CHCD. Thirdly, CHCD illustrate a spatial differeces in variations in development intensity. An increaing development intensity was found in Southwest China while a decreasing development intensity occurred in Hu-Ning, Shandong and Anhui. However, no such spatial differences could be found in the HYDE3.1 dataset.

3.3.2 Analysis of the relative biases on grid cell scale

Table 3 shows the grid numbers of each grade of relative difference. We found that the grid cells having relative biases over 70% (<–70% or >70%) accounted for 56%–63% of the total grid cells. Among these large relative biases grids, the grid cells having relative biases over 90% (<–90% or >90%) accounted for 40%–45% of the total grid cells. The grid cells having negative biases (<–70%) were two to three times more than those having the positive biases (>70%). The grid cells having relative biases no more than 10% only accounted for 5%–6% and those having relative biases no more than 30% accounted for ~17%. Although the relative biases decreased gradually from 1720 to 1910, the decreased amplitudes were too small. The proportions of grid cells having relative bias more than 90% (⩽–90% or >90%) decreased by only 5% (from 45% to 40%). The proportions of grid cells having relative bias no more than 10% increased by only 1.5% (from 4.5% to 6%). Such findings demonstrate that the spatial distribution pattern of the cropland on grid cell scale have great discrepancies between the two datasets partly owing to using different allocating methods.

From the perspective of spatial distribution of the relative biases (Figure 3), the grid cells having positive relative biases more than 90% mainly distributed in northern Jiangsu, the east-central parts of Hubei and Hunan, the southern Sichuan Basin and southern part of China. The grid cells having negative relative differences more than 90% mainly distributed in Southwest China, the Loess Plateau and the southeast coastal area. And the grid cells having relative biases no more than 10% scattered across the study area and, in comparing, many grid cells could be found in the northern Sichuan Basin and the southern Huang-Huai-Hai Plain.

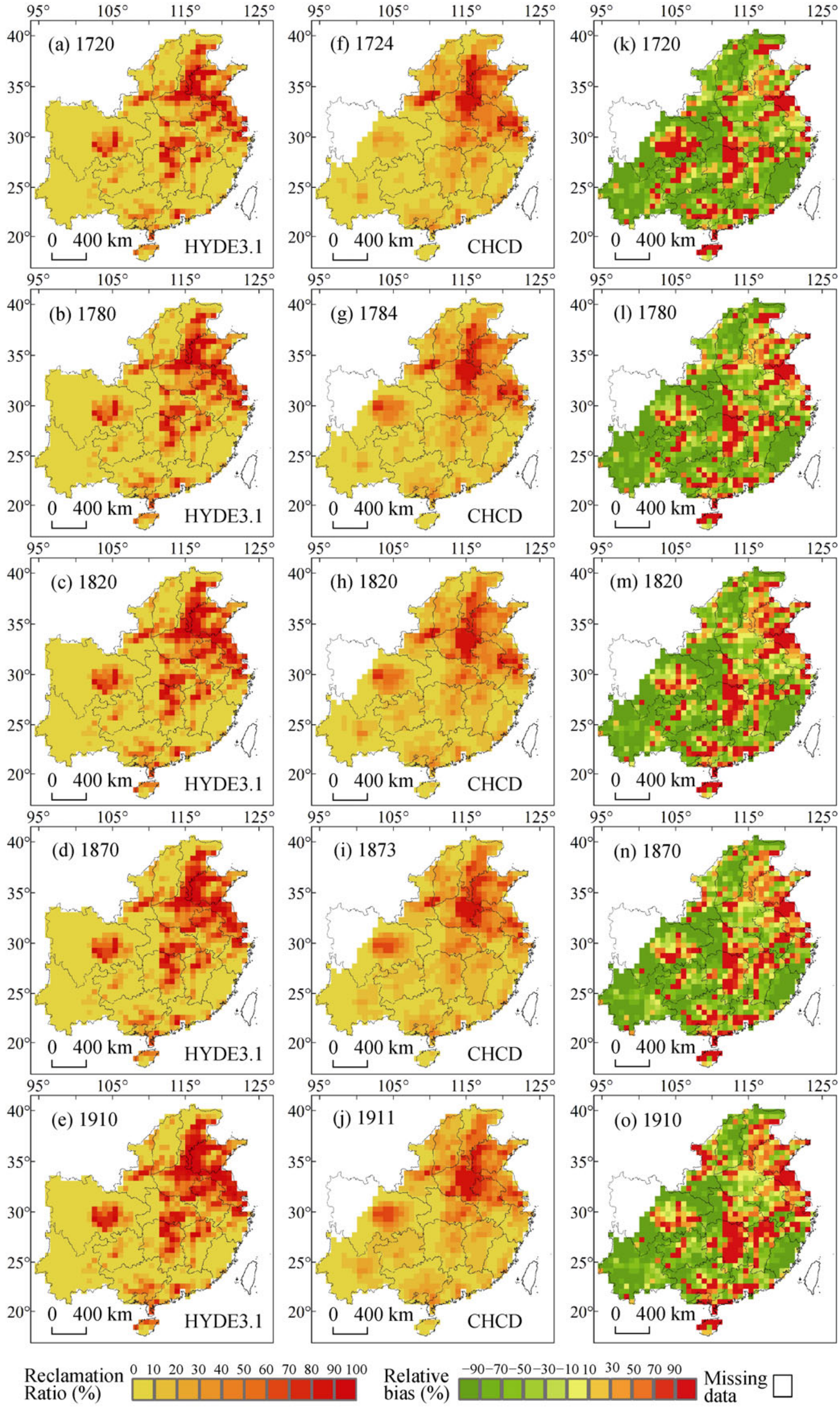

Figure 3 The spatial patterns of cropland distribution revealed by HYDE3.1 (left panel) and CHCD (middle panel) and the grid cell-based biases of HYDE3.1 (right panel) (60 km by 60 km grid cell)

Table 3 The percentage of grid cells of different relative biases between the HYDE3.1 and CHCD datasets

Relative biases (%)	1720	1780	1820	1870	1910
< –90	31.23	31.05	28.81	27.87	26.74
–90 to –70	14.30	12.53	11.78	12.62	11.58
–70 to –50	8.52	8.95	9.67	8.20	6.95
–50 to –30	7.78	7.37	6.94	7.05	8.00
–30 to –10	6.31	7.05	5.99	7.26	7.68
–10 to 10	4.52	5.16	5.99	5.57	6.42
10 to 30	3.26	4.95	5.57	5.26	5.47
30 to 50	5.36	4.11	4.52	4.84	5.47
50 to 70	1.89	3.26	3.26	5.05	4.21
70 to 90	2.94	3.16	4.31	2.63	4.21
> 90	13.88	12.42	13.14	13.67	13.26
Standard deviation (%)	5.02	4.42	4.21	3.65	3.59

4 Conclusions

Based on the above-mentioned findings, we drawn the following conclusions on the reliability of the cropland area from SAGE2010 and HYDE3.1 for TCRC over the past 300 years.

(1) SAGE2010 dataset estimated the cropland area of TCRC through linear interpolation from 1700 to 1950. The estimated cropland increased linearly at a rate of 0.51% from 1700 to 1950 and decreased linearly at a rate of –0.34% after 1950. Such linearly variations with a turn point in 1950 were inconsistent with real history. Moreover, the SAGE2010 largely overestimated the cropland in the TCRC. The largest overestimation occurred in 1950. The estimated cropland area from SAGE2010 dataset was more by 112% than that from CHCD dataset. On provincial scale, there are also large discrepancies.

(2) The HYDE3.1 dataset has similar total cropland area of entire study area with CHCD. However, large discrepancy on the spatial distribution of cropland could be found. The proportion of the grid cells having relative biases more than 70% (<–70% or >70%) and 90%(<–90% or >90%)account for 56%–63% and 40%–45% respectively while the percent of the grid cells having relative biases no more than 10% and 30% account for only 5%–6% and 17% respectively. It is implicated that HYDE3.1 has less ability to reproduce the spatial pattern of cropland at a scale of 60 km by 60 km in TCRC. It is needed to develop new methodology to allocate the cropland spatially.

(3) SAGE and HYDE are two global scale historical land use datasets. Their reconstruction principles and methods are reasonable on global scale. They can capture the large scale patterns of historical cropland changes, but not the fine details at local to regional scales. Therefore, it should be cautious about applying such global dataset at regional/local research. And it will be an important research direction in the future to develop new gridding methods and improve the understanding of the spatial distribution of cropland in the grid cell scale. In terms of China, making full use of local historical archives to create historical accurate land use dataset is a valuable way to improve the quality of regional climate and ecological simulation.

References

Betts R A, Falloon P D, Klein Goldewijk K *et al*., 2007. Biogeophysical effects of land use on climate: Model simulations of radiative forcing and large-scale temperature change. *Agricultural and Forest Meteorology*, 142(2–4): 216–233.

Bicik I, Jelecek L, Stepanek V, 2001. Land-use changes and their social driving forces in Czechia in the 19th and 20th centuries. *Land Use Policy*, 18(1): 65–73.

Brovkin V, Sitch S, Von Bloh W *et al*., 2004. Role of land cover changes for atmospheric CO_2 increase and climate change during the last 150 years. *Global Change Biology*, 10(8): 1253–1266.

Chen X, Lei M, Tang J P, 2006. Simulating the effect of changed vegetation on the climate change in Eurasia. *Advances in Earth Science*, 21(10): 1075–1082. (in Chinese)

Crook Frederick W, 1993. Underreporting of China's cultivated land area: Implications for world agricultural trade. *China International Agricultural and Trade Report*: 33–39.

FAO, 2008. Food and Agriculture Organization of the United Nations (FAO), FAOSTAT, Rome, Italy Available at: http://www.fao.org (accessed October 2008).

Feddema J J, Oleson K W, Bonan G B *et al*., 2005. The importance of land-cover change in simulating future climates. *Science*, 310(5754): 1674–1678.

Feng Z M, Liu B Q, Yang Y Z, 2005. A study of the changing trend of Chinese cultivated land amount and data reconstructioning: 1949–2003. *Journal of Natural Resources*, 20(1): 35–43. (in Chinese)

Findell K L, Pitman A J, England M H *et al*., 2009. Regional and global impacts of land cover change and sea surface temperature anomalies. *Journal of Climate*, 22(12): 3248–3269.

Foley J A, Ramankutty N, Brauman K A *et al*., 2011. Solutions for a cultivated planet. *Nature*, 478(7369): 337–342.

Forster P, Ramaswamy V, Artaxo P *et al*., 2007. Changes in Atmospheric Constituents and in Radiative forcing. In: Solomon Susan, Qin D H, and Manning M *et al*.(eds), 2007. Climate Change 2007: The Physical Science Basis Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Ge Q S, Dai J H, He F N *et al*., 2004. Spatiotemporal dynamics of reclamation and cultivation and its driving factors in parts of China during the last three centuries. *Progress in Natural Science*, 14(7): 605–613.

Hamre L N, Domaas S T, Austad I *et al*., 2007. Land-cover and structural changes in a western Norwegian cultural landscape since 1865, based on an old cadastral map and a field survey. *Landscape Ecology*, 22(10): 1563–1574.

Houghton Ra, Van Der Werf Gr, Defries Rs *et al*., 2012. Chapter G2 Carbon emissions from land use and land-cover change. *Biogeosciences Discuss*, 9(1): 835–878.

IIASA, 1999. What is the most likely area of China's cultivated land. http://www.iiasa.ac.at.

Klein Goldewijk K, 2001. Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles*, 15(2): 417–433.

Klein Goldewijk K, Battjes J, 1997. A hundred year (1890–1990) database for integrated environmental assessments (HYDE, version 1.1). National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

Klein Goldewijk 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 12,000 years. *Global Ecology and Biogeography*, 20(1): 73–86.

Li B B, Fang X Q, Ye Y *et al*., 2010. Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data: A case study in Northeast China. *Science in China* (*Series D*), 53(11): 1689–1699.

Li Q P, Ding Y H, Dong W J, 2006. A numerical simulation on impact of historical land-use changes on regional climate in China since 1700. *Acta Meteorologica Sinica*, 64(3): 257–270. (in Chinese)

Lin S S, Zheng J Y, He F N, 2009. Gridding cropland data reconstruction over the agricultural region of China in 1820. *Journal of Geographical Sciences*, 19(1): 36–48.

Liu Y C, 2000. Investigation Datasets of China Land Resources. Beijing: China Land Press, 22–23.(in Chinese)

Matthews H D, Weaver A J, Eby M *et al*., 2003. Radiative forcing of climate by historical land cover change.

Geophysical Research Letters, 30(2): 1055–1059.

National Bureau of Statistics, 2000. China Agricultural Statistics Yearbook of 2000. Beijing: China Statistics Press, 169–181. (in Chinese)

Pitman A J, Avila F B, Abramowitz G *et al*., 2011. Importance of background climate in determining impact of land-cover change on regional climate. *Nature Climate Change*, 1(9): 472–475.

Ramankutty N, Foley J A, 1998. Characterizing patterns of global land use: An analysis of global croplands data. *Global Biogeochemical Cycles*, 12(4): 667–685.

Ramankutty N, Foley J A, 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles*, 13(4): 997–1027.

Ramankutty N, Foley J, 2010. ISLSCP II Historical Croplands Cover 1700–1992. Dataset. http: //daac.ornl.gov/.

Shevliakova E, Pacala S W, Malyshev S *et al*., 2009. Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. *Global Biogeochemical Cycles*, 23(GB2022), doi: 10.1029/2007GB003176. 7

Shi Z G, Yan X D, Yin C H *et al*., 2007. Effects of historical land cover changes on climate. *Chinese Science Bulletin*, 52(18): 2575–2583.

Thompson Robert S, 2000. BIOME 300: Understanding the impacts of human activities on land cover over the past 300 years. *IGBP Newsletter*, (43): 2–3.

Voldoire A, Eickhout B, Schaeffer M *et al*., 2007. Climate simulation of the twenty-first century with interactive land-use changes. *Climate Dynamics*, 29(2): 177–193.

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*), 52(8): 1172–1182.

Zhang X Z, He F N, Li S C, 2013. Reconstructed cropland in the mid-11th century in the traditional agricultural area of China: Implications of comparisons among datasets. *Regional Environmental Change*, doi: 10.1007/s10113-012-0390-6.

Zhang X Z, Wang W C, Fang X Q *et al*., 2011. Vegetation of Northeast China during the late seventeenth to early twentieth century as revealed by historical documents. *Regional Environmental Change*, 11(4): 869–882.