SCIENCE CHINA
Earth Sciences

• RESEARCH PAPER •

November 2010 Vol.53 No.11: 1689–1699
doi: 10.1007/s11430-010-4053-5

Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data—A case study in Northeast China

LI BeiBei[1], FANG XiuQi[1,2*], YE Yu[1] & ZHANG XueZhen[2]

[1] School of Geography, Beijing Normal University, Beijing 100875, China;
[2] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Received February 7, 2009; accepted July 13, 2010; published online August 7, 2010

Historical cropland datasets are fundamental for quantifying the effects of human land use activities on climatic change and the carbon cycle. Two representative global land-use datasets, the Global Land Use Database (termed SAGE dataset) and the Historical Database of the Global Environment (termed HYDE dataset) have been established and used widely. Despite improvement of data quality and methodologies for extracting historical land use information, certain dataset limitations exist that need to be quantified and communicated to users so that they can make informed decisions on whether and how these land-use products should be used. The Cropland data of Northeast China (CNEC) is based on calibrated historical data and a multi-sourced data conversion model, and reconstructs cropland cover change in Northeast China over the last 300 years. Using the CNEC as a reference, we evaluated the accuracy of cropland cover for SAGE and HYDE in Northeast China at spatial scales ranging from the entire Northeast China to provinces and even individual raster grid cells. Neither SAGE nor HYDE reflects real historical land reclamation. Cropland areas in SAGE are overestimated by 20.98 times in 1700 to 1.6 times in 1990. Although HYDE is better, there are significant disagreements in cropland area and distribution between HYDE and CNEC, especially in the 18th and 19th centuries. The proportion of total grid cells whose relative error was greater than 100% was 63.55% in 1700 and 53.27% in 1780. Global cropland dataset errors over Northeast China originate mainly from both the reverse calculation method for historical cropland data based on modern spatial patterns, and modern land-use outputs from satellite data.

LUCC, dataset, accuracy assessment, Northeast China

Citation: Li B B, Fang X Q, Ye Y, et al. Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data—A case study in Northeast China. Sci China Earth Sci, 2010, 53: 1689–1699, doi: 10.1007/s11430-010-4053-5

Land use and land cover change (LUCC), one of the important human influences on the Earth system, has global environmental effects [1, 2]. LUCC can modify directly or indirectly the terrestrial ecosystem carbon cycle and energy exchange between land and atmosphere. Carbon budgets and climate change caused by LUCC are the focus of much research. When land use and land cover is changed from one type to another, it is usually accompanied by considerable carbon emissions [3]. LUCC also changes the physical properties of the Earth's surface, thereby affecting regional and global climate. Surface albedo is a significant component of the radiative forcing of global climate change influenced by human land use activities, and has been especially affected by agricultural exploitation since the industrial revolution [4–6].

The study of LUCC processes and their long-term environmental effects is a key to unlocking the history of human

*Corresponding author (email: xfang@bnu.edu.cn)

influence on the Earth system. Human land use activities released 157 Pg of carbon emissions to the atmosphere during 1850–2000 [7]. Land use change dominated by deforestation (and associated biomass burning) with contributions from changing agricultural practices were responsible for 25% increase in atmospheric CO_2 concentration since the industrial revolution, only second to the contribution of fossil fuel combustion [6]. HadAM3 AGCM model simulations show that global mean winter and spring temperature decreased 1–2°C as a result of land use change since 1700, with surface albedo change the primary factor in this cooling [8].

Accurate and complete historical land use data are fundamental for research on correlations between land use change, the carbon cycle and anthropogenic climate effects [9, 10]. Scientific programs under the framework of the International Geosphere-Biosphere Program (IGBP) such as BIOME300, LUCC, GCTE (Global Change and Terrestrial Ecosystem), GLP and iLEAPS (Integrated Land Ecosystem Atmosphere Processes Study) have made great progress in the reconstructing past environments, especially for global land cover over the last 300 years [11]. Many historical land use and land cover datasets have been established at global and regional scales [12]. Two representative global land use datasets are the Global Land Use Database (termed RF) and the Historical Database of the Global Environment (termed HYDE), which were established by Ramankutty and Foley from the Center for Sustainability and the Global Environment of Wisconsin-Madison University, and the Netherlands Environmental Assessment Agency, respectively [13, 14]. These two important datasets have been used widely for quantitative analysis of environmental effects of land use change [6, 9, 10]. Some research on carbon cycle history influenced by land use and land cover change has been conducted using SAGE and HYDE. Tian et al. [15] used cropland data from SAGE with a process-based Terrestrial Ecosystem Model (TEM) to simulate the combined effects of climate variability, increasing atmospheric CO_2 concentration and cropland establishment and abandonment on the exchange of CO_2 between the atmosphere and monsoon Asian ecosystems during 1860–1990. Brovkin et al. [16] applied both SAGE and HYDE to assess the role of changing anthropogenic (CO_2 emissions, land cover) forcing on the global climate system over the last 150 years using an earth system model of intermediate complexity. Oost et al. [17] used HYDE as one parameter in water and tillage erosion models to simulate the impact of agricultural soil erosion on the global carbon cycle. Other studies using these two datasets focused on the climate effects of LUCC globally and in China. Matthews et al. [18] investigated the radiative effect of changing human land use patterns on the climate of the past 300 years using the Uvic Earth System Climate Model, through analysis of surface albedo change induced by cropland extension reflected in RF. Wang et al. [19] investigated the impacts in China resulting from historical land cover modification extracted from HYDE using a regional climate model. Chen et al. [20] used an AGCM + SSIB model to simulate the sensitivity of climate changes because of land cover change across Eurasia, and the historical degradation of vegetation was estimated from HYDE. Using the improved regional climate model of the National Climate Center (NCC-RegCM), Li et al. [21] used HYDE to undertake a series of modeling experiments to investigate the impacts of historical land use change on regional climate in China.

SAGE and HYDE can be used to reconstruct spatially explicit global cropland maps that can be used in integrated models for identifying human environmental impacts. However, the accuracy of global datasets which were established quickly for modeling the Earth system processes still need to be assessed, especially at regional scales. If there are significant errors in the chronology and spatial distribution of the land use data, uncertainty in quantifying the environmental effects of LUCC based on simulations using these data would increase.

Since 1700, land use and land cover over Northeast China have changed dramatically because of migration and reclamation. Based on information from historical documents, government files, Russian and Japanese investigations, official statistics and other scholars' publications, Ye et al. [22] reconstructed historical cropland cover change in Northeast China (Cropland data of Northeast China, CNEC data for short) over the last 300 years through unification processes including documentary data calibration and a multi-sourced data conversion model. These are high-resolution spatial, regional land use data verified by detailed materials at the same resolution. To assess the validity of applying global land use data at the regional scale, in this study we selected Northeast China as a case study area to evaluate the accuracy of SAGE and HYDE using regional CNEC data.

1 Data and methods

1.1 Data sources and reconstruction issues

We selected two global land use datasets. One is the SAGE dataset [13] from the Center for Sustainability and the Global Environment, University of Wisconsin-Madison (http://www.sage.wisc.edu/iamdata). The other is the HYDE dataset [14] (HYDE 3.0, published in August 2008) from The Netherlands Environmental Assessment Agency (http://www.pbl.nl/en/themasites/hyde/index.html). The spatial resolution of SAGE is 0.5° latitude by longitude grid cells, and the time resolution is 10 years from 1700 to 1992. HYDE has higher spatial resolution of 5 minute grid cells, and the time resolution is 10 years from 1700 to 2005. The reference regional historical cropland data used here is the CNEC covering the 300 years [22], the spatial resolution of which is 0.5°–1° latitude by longitude grid cells determined by the extent of administrative counties. The CNEC data cover the period 1683 to 2002. Because of constraints of historical documents and comparison between past and recent data, the time resolution is 50–100 years prior to the 20th century, and 5–10 years throughout the 20th century (Table 1).

Following general methods were used to reconstructing historical cropland cover of SAGE. First, global croplands in 1992 were characterized from remotely-sensed land cover classification data. An extensive database of historical croplands at the national and sub-national level was then compiled. The 1992 cropland area data were used as an initial condition for characterizing spatially historical croplands back in time for each political unit (Figure 1). One basic assumption is that within each political unit, the cropland pattern of 1992 represents historical spatial patterns. The reconstruction method was to simply adjust the 1992 crop cover pattern so that total cropland for that unit matched historical inventory data, i.e., the downscaling of historical cropland area was directed by contemporary cropland distribution. SAGE simulates crop cover backward in time, annually, through linear regression between inventory data.

The updated HYDE 3.0 implemented new allocation algorithms with time-dependent weighted cropland maps. (sub-)national crop area statistics were allocated to grid cells according to a combination of two weighted maps: a current one constructed from a 2000 cropland satellite map (Wcrop_satellite), and a historical one constructed according to human settlements and the nature of surrounding landscape (swamps, mountains, dense forests, poor soils, unfavorable climate). This is described by

$$W_{\text{crop},t} = \frac{\left[G_{\text{area}} - U_{\text{area},t}\right]}{G_{\text{area max}}} \times W_{\text{pop},t} \times W_{\text{suit}} \times W_{\text{river}} \times W_{\text{slope}} \times W_{\text{temp\_crop}},$$

where G_{area} is total land-area (excluding ice and snow cover), $U_{\text{area},t}$ is urban built-up area for year t, and $G_{\text{area max}}$ is maxi-

Table 1 Data format of SAGE dataset, HYDE dataset and CNEC data in last 300 years

Dataset	Interval	Time section	Spatial resolution
RF	10 years	1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 1992	0.5°
HYDE	10 years	1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2005	5′
CNEC	100–5 years	1683, 1735, 1780, 1908, 1914, 1931, 1940, 1950, 1955, 1960, 1970, 1980, 1990, 2002	County (~0.5°–1°)

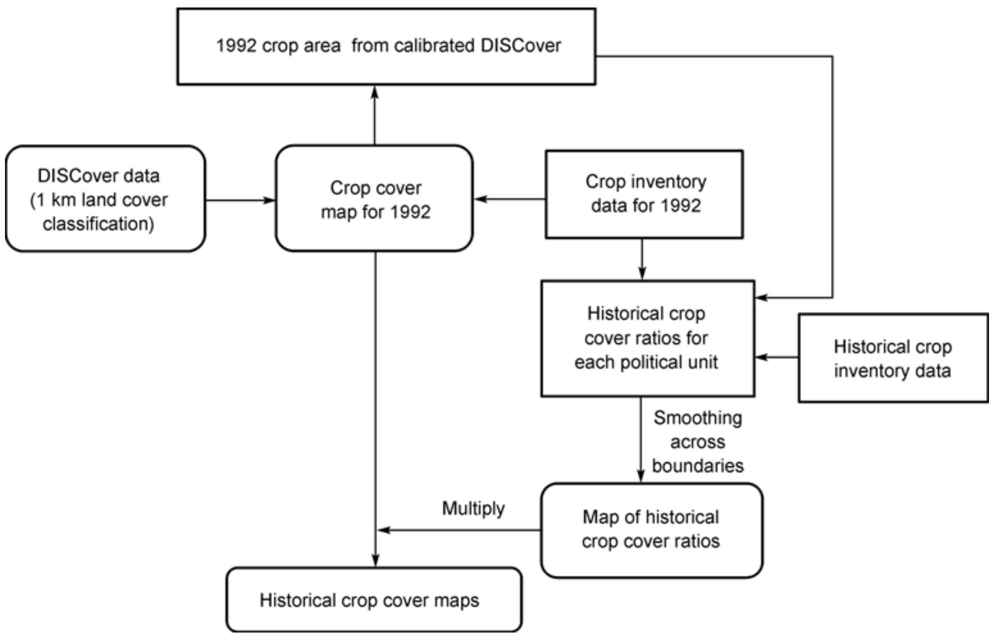

Figure 1 Approach used to reconstruct historical crop cover maps for SAGE [13]. Boxes with sharp corners indicate data at the level of political units, and boxes with rounded corners indicate spatially explicit maps. A crop cover map for 1992 was first derived by calibrating the DISCover data set against 1992 crop inventory data. The ratio of crop cover in the past to the crop cover in 1992 was then derived for each political unit and then further converted to a spatial map and smoothed across political unit boundaries. The resulting map was multiplied by the 1992 crop cover map to derive historical crop cover maps.

mum area of a 5′ grid cell. Allocation rules are: (i) in urban built-up areas (U_{area}) there was no allocation (no agricultural areas remaining); (ii) in areas with population density (W_{pop}) lower than 0.1 km^{-2} there was no allocation (no need for agriculture); (iii) land with highest soil suitability for crops was colonized first (W_{suit}); (iv) coastal areas and river plains were more favorable for early settlement, being easily accessible (W_{river}); (v) steep terrain with high slopes were less attractive for settlement and agriculture (W_{slope}); and (vi) below the threshold of mean annual temperature of 0°C no agricultural activity was assumed viable (W_{temp_crop}).

The major difference between HYDE and SAGE is that in the former allocation of crop and pasture is guided by the presence of historical population densities based on modern population densities. The historical population density map is a spatial allocation of national or sub-national level statistical data according to recent population distribution [23, 24], and this led to little change in spatial pattern during the last 300 years. Thus, historical and recent cropland data show similar spatial patterns.

CNEC data had county-level spatial resolution, and their temporal resolution varied from 50–100 years during the 17th–19th centuries to 5–10 years during the 20th century. Data from various sources and different periods were used to reconstruct cropland area in Northeast China during the last 300 years, including historical documents from the Qing Dynasty, statistical data from Chinese provincial administrations and Japanese and Russian field survey data from the first half of the 20th century, statistical data from the National Bureau of Statistics of China, and investigative data (from the 1980s and 1990s) from the Ministry of Land and Resources after 1949. To produce a unified data series, data need to be standardized to make them comparable and capable of being converted quantitatively to real land use conditions in a given period [22, 25]. Methods were adopted for calibrating and reconstructing cropland area at the county level in different historical periods. First, the Qing dynasty data were converted to one standard unit of cropland area, and gross cropland was calibrated by the hidden percentage of cropland area. Second, 20th century data were calibrated to the real cropland area by using a conversion model of multi-sourced cropland data, through which the investigative and statistical data could be converted. Finally, using the 1908 cropland data overlain by both data systems, cropland data from prior to the 20th century and during the 20th century were joined into one series.

1.2 Data processing

We evaluated cropland data accuracy of SAGE and HYDE by comparing cropland cover in Northeast China at spatial scales from the whole of Northeast China to each province and to smaller areas, using the CNEC data as a reference. Because of format differences among the three data sets, we standardized cropland data from RF, HYDE and CNEC before the accuracy assessment (Table 1).

The analysis years were chosen according those available in CNEC (Table 1). We therefore used SAGE and HYDE cropland data from 1700, 1740, 1780, 1910, 1920, 1930, 1940, 1950, 1980, 1990 and 2005. The 1683 CNEC data correspond to 1700 in SAGE and HYDE, the 1908 CNEC data correspond to 1910 in SAGE and HYDE, the 1931 CNEC data corresponds 1930 in SAGE and HYDE, and the 2002 CNEC data correspond to 2005 in HYDE.

We conducted a gross assessment of cropland area accuracy for the whole of Northeast China, and an individual provincial assessment for the same region. SAGE provided data in tabular format for each provincial administrative unit, and gross cropland area was obtained by summing the data from the three provinces. HYDE was in raster format, and was clipped using provincial and Northeast China boundaries to obtain province and gross cropland area. The CNEC data provided both provincial and gross cropland area [22].

We evaluated spatial differences between the global datasets and CNEC by grid accuracy assessment. Both SAGE and HYDE are comprised of grids, while the CNEC data presented according to county polygons. We standardized the data sets to a 1°×1° latitude by longitude grid resolution, which approximates the average county area in Northeast China, so that the conversion error from CNEC data was minimized. SAGE and HYDE were upscaled directly to a 1°×1° resolution. The CNEC polygon data were converted to a 1° gridded map on the assumption that the fractions in a grid derived from the county map of CNEC had the same cropland cover ratio in each county.

1.3 Methods

The gross and provincial assessments were conducted mainly to calculate the difference ratio in cropland area between the global dataset and CNEC data for each analysis period, i.e., SAGE or HYDE cropland area was divided by CNEC cropland area for the same time period.

After considering results of the gross and provincial assessments, we selected only HYDE for further analysis. The grid assessment included comparison of spatio-temporal changes in cropland area between HYDE and CNEC, the determination of absolute and relative error for each grid cell, and identifying loci of the cropland increase at century or 50-year timescale. Because of systematic and calculation error, only the absolute and relative error between CNEC and HYDE were characterized spatially, and not the absolute numerical difference. The formulae for absolute and relative error were:

$$\text{Absolute error} = X_{Hi} - X_{Ci},$$

$$\text{Relative error} = (X_{Hi} - X_{Ci})/X_{Ci} \times 100\%,$$

where X_{Hi} is the HYDE cropland area in '*i*' time period, and

X_{Ci} is the CNEC cropland area in 'i' time period.

2 Results and analysis

2.1 Gross assessment

Northeast China is an area where in the last 300 years the largest land cultivation activities by migrants have occurred in the country. Affected by prohibitive migration policy in Northeast China, cropland area increased slowly during the Qing Dynasty, and extensive land exploration and settlement occurred from the south to the north from the start of the 20th century. CNEC data showed that cropland area increased slowly initially, then at an exponential rate. The cropland area in the 17th–19th centuries was much less than at present and increased slowly. The gross amount of cropland was 5396 km^2 in 1683, 15266 km^2 in 1735, 21867 km^2 in 1780, and 81016 km^2 in 1908. From 1683 to 1840 the cropland area increased with an annual average rate of less than 1.0%. The annual growth rate reached 1.6% during 1840–1908. During the 20th century, the total cropland area increased more rapidly. In 1914, total area was 102245 km^2, and rose to 196519 km^2 by 1960, which was close to modern figures. Two phases of higher growth rates occurred in 1914–1931 and 1950–1960, when annual average growth rate of cropland area reached 2.3% and 3.1%, respectively. Between these two phases, total cropland area decreased during 1940–1950. In 1960–1980, it rose slowly with an annual average growth rate of 0.2%, and increased only 6791 km^2 in these 20 years. In the period 1980–2000, cropland area decreased with an annual average rate of –0.5%; in the period 1990–2000 it decreased rapidly at a rate of −1.2% [22].

SAGE showed a different time series for cropland area in Northeast China to that of CNEC. Total cropland area was already 113,210 km^2 in 1700, and increased with an annual average growth rate of 0.4% during 1700–1950. This can be simulated by a linear equation (y_i = 112.36i–179852.62, r^2 = 0.99, y_i is the cropland area in year i). The increasing trend in SAGE does not accord with CNEC during 1700–1950. After 1950, the cropland area decreased (Figure 2). In 1950–1970, 1970–1980 and 1980–1990, the cropland area decreased in 2.43×10^4, 4.81×10^4 and 0.75×10^4 km^2, respectively. This contrasts with CNEC, where cropland increased during 1950–1980, and only decreased 0.18×10^4 km^2 in 1980–1990, which was less than that of SAGE in the same period. The difference ratios between SAGE and CNEC were greater than 1 during 1700–1990, which meant that total cropland area over Northeast China for SAGE was greater than that of CNEC in each time period. In the 18th and 19th centuries, the difference ratio was larger than that of the 20th century: it was 20.98 in 1700, and decreased from 9.22 in 1780 to 1.6 in 1990. This means in recent years the cropland area in SAGE was 1.6 times of in CNEC figures (Table 2).

Total cropland area in Northeast China as determined from HYDE increased from 1700, except for the period 1960–1980. The growth rate changed during the last 300 years: it rose slowly before 1900 and increased during the period 1900–1960. During the period 1980–2005, the cropland area increased most quickly. The HYDE cropland area was larger in CNEC from the 18th to early 20th centuries, but became smaller than CNEC after the 1930s. The difference ratio between HYDE and CNEC was largest in the 18th–19th centuries. In 1700 and 1780, it was 9.87 and 2.99, respectively. After the 1930s, the difference ratios were less than 1, and their average was 0.69. This means HYDE cropland area was ~30% lower than CNEC (Table 2). During 1700–1910, HYDE cropland area increased 3.6×10^4 km^2 at an annual average rate of 0.3%, but CNEC cropland area increased at a rate of 6.2%. HYDE reflected the accelerated growth rate trend in the 20th century. However, the annual average rate was lower than for CNEC at a century timescale, and decadal variability in cropland area differed from CNEC. During 1910–1960, HYDE cropland area increased

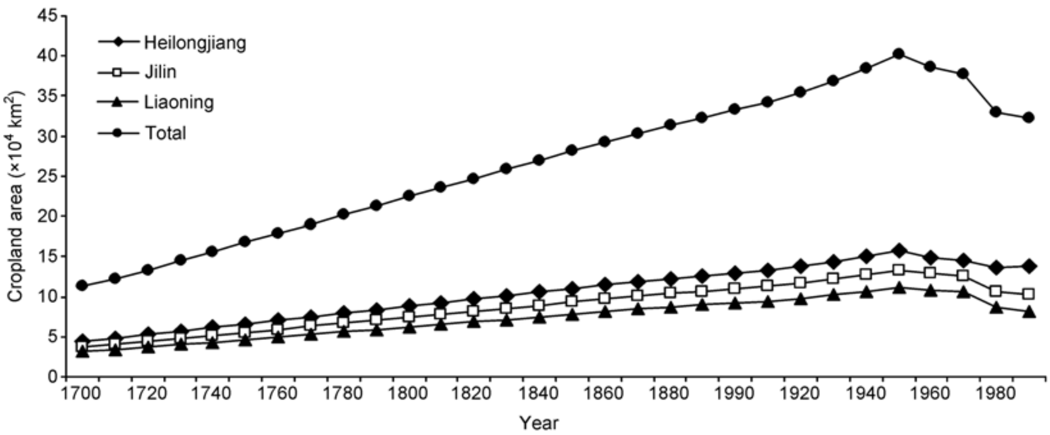

Figure 2 Changes in cropland area in Northeast China from the SAGE dataset during 1700–1990.

Table 2 Cropland areas (km^2) of from RF, HYDE and CNEC datasets and difference ratio (in bracket, %)

Region	Dataset	1700	1780	1910	1930	1950	1990
Northeast China	RF	113210 (20.98)	201650 (9.22)	341800 (4.22)	367800 (2.45)	401600 (2.77)	321810 (1.60)
	HYDE	532D44 (9.87)	65417 (2.99)	89672 (1.11)	101219 (0.68)	109405 (0.75)	140252 (0.70)
	CNEC	5396 (1.00)	21867 (1.00)	81016 (1.00)	150128 (1.00)	145058 (1.00)	201516 (1.00)
Liaoning Province	RF	31370 (5.81)	55900 (3.15)	94780 (4.18)	101990 (2.54)	111340 (2.40)	80630 (1.86)
	HYDE	11938 (2.21)	14792 (0.83)	20445 (0.90)	23206 (0.58)	25173 (0.54)	32673 (0.76)
	CNEC	5390 (1.00)	17749 (1.00)	22683 (1.00)	40212 (1.00)	46312 (1.00)	42992 (1.00)
Jilin Province	RF	37510 (–)	66820 (26.21)	113290 (3.32)	121940 (2.22)	133160 (3.14)	102990 (2.09)
	HYDE	17276 (–)	21111 (8.28)	28796 (0.84)	32578 (0.59)	35270 (0.83)	44980 (0.91)
	CNEC	0 (–)	2549 (1.00)	34172 (1.00)	54813 (1.00)	42342 (1.00)	49168 (1.00)
Heilongjiang Province	RF	44330 (–)	78930 (50.31)	133730 (5.53)	143870 (2.61)	157100 (2.79)	138190 (1.26)
	HYDE	24030 (–)	29514 (18.81)	40431 (1.67)	45435 (0.82)	48962 (0.87)	62599 (0.57)
	CNEC	0 (–)	1569 (1.00)	24161 (1.00)	55103 (1.00)	56404 (1.00)	109356 (1.00)

linearly with an annual average rate of 0.5%, but that of CNEC had a growth rate of 2.3% during 1914–1931 which increased to 3.1% during 1950–1960. HYDE cropland area decreased 0.55×10^4 km^2 during 1960–1980, which was the inverse of CNEC whose cropland area increased by 0.68×10^4 km^2. Changes in HYDE cropland area were also the opposite CNEC after 1980: HYDE cropland area increased at an annual average rate of 2.0%, while CNEC cropland area reduced at an annual average rate of –0.5%.

2.2 Provincial assessment

CNEC data indicated that the three provinces of Northeast China had different reclamation processes, and locus of cropland area increase moved from the south to the north. At the beginning of the Qing Dynasty, cropland was only distributed in Liaoning Province. In 1683, there were 5396 km^2 of cropland in Liaoning Province, and none in Jilin and Heilongjiang provinces. The increase of cropland area in Liaoning Province was greatest during the 17th and 18th centuries, and its cropland area increased 1.2×10^4 km^2 from 1683 to 1780. Up to 1780, the cropland areas of Liaoning, Jilin and Heilongjiang provinces were 17749, 2549 and 1569 km^2, respectively. During the 19th century, cropland increased mainly in Jilin and Heilongjiang provinces, and cropland areas of these two provinces were both larger than that of Liaoning Province at the beginning of the 20th century. In 1908, Jilin Province had the largest cropland area of the three provinces, ~34172 km^2, and it increased by 3.16×10^4 km^2 from 1780. Over the same period, cropland area increased by 2.26×10^4 km^2 in Heilongjiang Province and 0.49×10^4 km^2 in Liaoning Province. During the 20th century, the largest increase in cropland was in Heilongjiang Province, where cropland area became largest of the three provinces by the 1930s, and exceeded the sum of Jilin and Liaoning provinces at the end of the 20th century. The growth of cropland in the three provinces experienced different decadal changes in the 20th century. The periods of cropland increase were longest in Heilongjiang Province, which had three growth periods in 1914–1940, 1950–1960 and 1960–1980 with annual average rates of 2.0%, 4.4% and 1.1%, respectively. In 1940–1950 and 1980–2000, the cropland areas decreased. There were two cropland growth periods in Jilin Province in 1914–1931 and 1950–1960 with annual average rates of 1.9% and 2.7%, respectively; otherwise, cropland area decreased. The two growth periods in Liaoning Province were 1914–1931 and 1940–1960 with annual average rates of 23.1% and 1.6%, respectively.

SAGE cropland area changes in the provinces showed similar trends, and all were simulated by a linear equation which explained 99% of variance during 1700–1950 (Figure 3). After 1950, cropland decreased in the three provinces, except for a limited increase in Heilongjiang Province during 1980–1990. SAGE provincial data did not reflect periodic growth at the century timescale or fluctuations at the decadal timescale. SAGE cropland areas in all three provinces dataset were larger than that in the CNEC for each time period. The difference ratios between SAGE and CNEC in Jilin and Heilongjiang provinces were much larger than that in Liaoning Province. The difference ratios of Jilin and Heilongjiang provinces for 1700 were unable to be computed because no cropland existed at this time. In 1780, the difference ratios for Jilin and Heilongjiang provinces were 26.21 and 50.31, respectively, and in Liaoning Province was only 3.15. This indicates the overestimation of cropland area in RF, mainly in Jilin and Heilongjiang provinces during the 18th–19th centuries.

HYDE cropland changes at the provincial scale had the same characteristics as for gross cropland across Northeast China. The cropland area in the three provinces slowly increased during 1700–1910, growth accelerated during 1910–1960, declined during 1960–1980, and then grew rapidly after 1980. The differences in cropland growth feature between Liaoning, Jilin and Heilongjiang provinces were not reflected by HYDE, especially at the decadal timescale. The cropland area ratio between the three provinces was maintained mostly at 1∶1.40∶1.95, indicating that HYDE was unable to record the changing locus of cropland increase after 1700. The difference ratio between HYDE

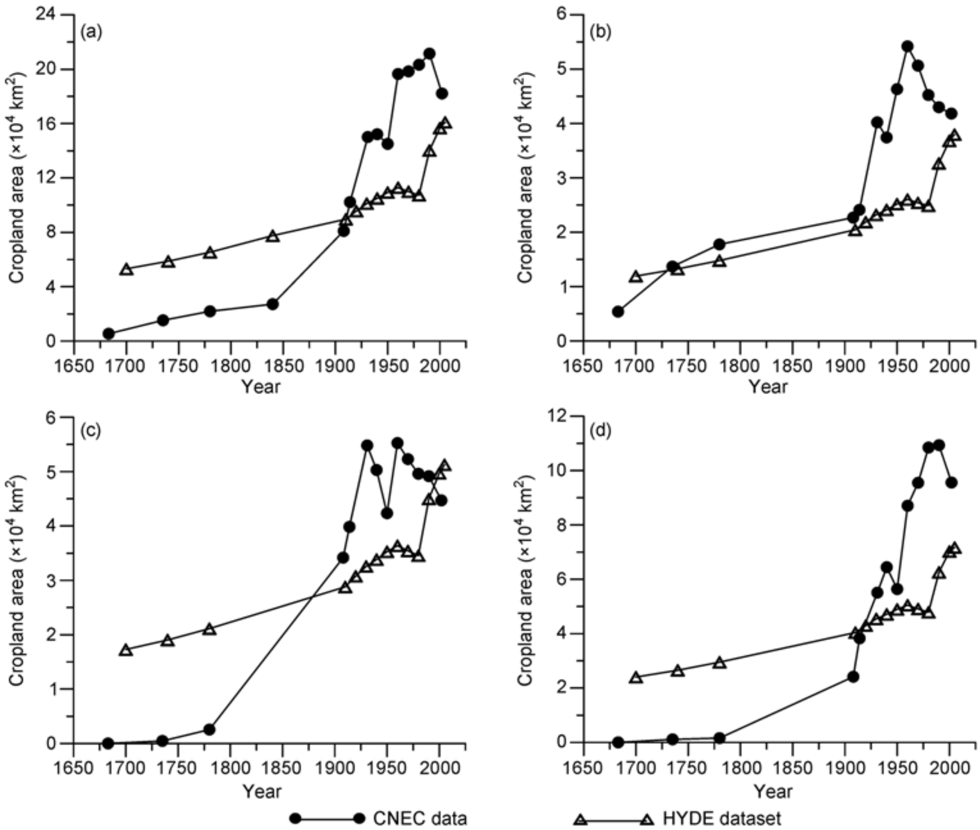

Figure 3 Comparison of cropland area changes in Northeast China between HYDE and CNEC datasets. (a) Gross cropland area in Northeast China; (b) cropland area in Liaoning Province; (c) cropland area Jilin Province; (d) cropland area in Heilongjiang Province.

and CNEC showed independent changes in different province. For Liaoning Province, the difference ratio was 2.21 in 1700, and was below 1 after 1740 with phases close to 1 in 1740–1920 (0.90) and 1990–2000 (0.82). For Jilin Province, the difference ratio was much larger than 1 in the 18th–19th centuries, and was below 1 after 1910 with phases close to 1 in 1910 (0.84), 1950 (0.83) and 1990 (0.91). For Heilongjiang Province, the difference ratio was also greater than 1 in the first 200 years of the 300 year analysis period, and was below 1 after 1930 with phases close to 1 in 1930 (0.82) and 1950 (0.87). Similar to RF, this suggests that HYDE reconstruction errors came mainly from Jilin and Heilongjiang provinces during the 18th–19th centuries.

2.3 Grid assessment

Results from the gross and provincial assessments showed that SAGE differed considerable from CNEC and HYDE, and showed cropland overestimations for all the time periods. Because of their comparability, HYDE and CNEC data only were used for grid cell assessment for cropland spatial distribution error analysis.

The gridded CNEC reconstructions indicated the spatial expansion of cropland because of land exploitation in Northeast China during the last 300 years. In 1683, cropland existed only on the plain of the lower Liaohe River in Liaoning Province. Cropland spatial distribution did not change much during the 17th–19th centuries. Until the middle of the 18th century, agricultural areas in Northeast China with extensive cropland cover were still mainly restricted to Liaoning Province. Only small cropland areas distributed in central Jilin Province. Because of migration policy, dramatic change occurred from the late 19th to early 20th centuries, resulting in the northern reclamation boundary extending to central Heilongjiang Province. There were two accelerated cultivation periods during the 20th century, and three agricultural regions with extensive cropland cover had formed in Northeast China. During 1914–1931 cropland cover extended from central Jilin and Heilongjiang provinces to their eastern and western borders of the present day (particularly to the west), and the percentage of cropland cover greater than 50% in each province increased remarkably. During 1950–1980, the newly-increased cropland was concentrated on the Sanjiang Plain and south of the Xiao Hinggan Mountains in Heilongjiang Province. The ratio of cropland cover also increased on old reclaimed land.

There were significant differences between HYDE and CNEC in cropland expansion and ratio of cropland cover increase (Figure 4). The historical cropland distribution of HYDE had a similar spatial pattern to more recent cropland

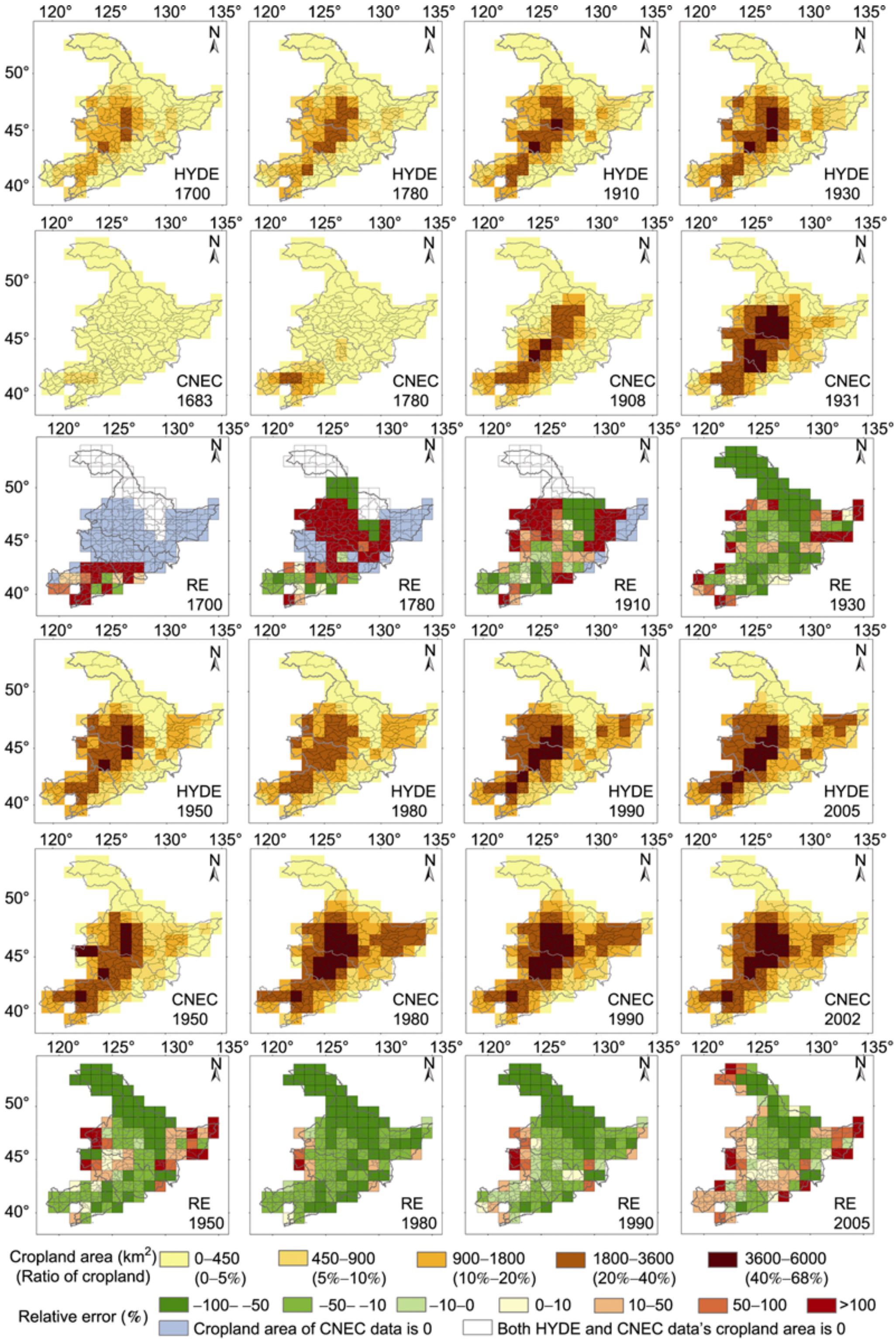

Figure 4 Changes of cropland area in Northeast China of HYDE dataset and CNEC data and relative error between two datasets. The statistic unit is 1° in latitude by longitude.

distribution, with the Songnen Plain and Liaohe Plain agricultural regions having high ratios of cropland cover. Cropland growth was expressed by increase in cropland area in the Songnen Plain and Liaohe Plain, and the expansion from the central plain to the eastern and western margins. HYDE is therefore contrary to the history record of reclamation that expanded from the south to the north along the central plain of Northeast China. HYDE recorded some cropland in

Heilongjiang Province in 1700, and cropland cover reached 5%–10% on the Sanjiang Plain in the early 20th century. Large-scale cultivation in Heilongjiang Province which showed by HYDE was much earlier than historical records, because the first northward progression of agricultural regions was during the period 1796–1901, and the main exploitation of the Sanjiang Plain was after 1949.

The absolute error of gridded HYDE distribution changes over time, and can be divided into two stages. First, during the 17th–18th centuries, the absolute error of most grids was positive, and high-error grids were concentrated in western Jilin Province and south-central and southwest Heilongjiang Province. Second, during the 20th century the absolute error of more grid cells became negative. In 1910, cells with positive absolute error were concentrated in southwest Heilongjiang Province and western Sanjiang Plain, and the negative ones were distributed in central Jilin and Heilongjiang provinces. After 1930, the absolute error of most grids was negative, and theses were located on Songnen Plain. Spatial changes in absolute error illustrate that HYDE underestimated cropland area in the Songnen Plain after 1900, and had significant spatial error in 1900 even when gross and provincial cropland area in HYDE and CNEC were similar.

The relative grid cell errors of HYDE were caculated for each time period. The relative error of most grid cells was greater than 100% in the 18th–19th centuries, mainly in the range of −10% to 100% and >100% in the early 20th century, and mainly in the range of −10%–100% after the 1930s (Table 3). The proportion of grids with greater than 100% relative error in 1700 and 1780 was 63.55% and 53.7%, respectively. The grids were located mainly in the modern agricultural regions of Jilin and Heilongjiang provinces. The lower relative error (0–10%) was in 1700 and 1780 in the Xiao Hinggan Mountains, where no large-scale exploitation took place in the 18th century. During the 20th century, the range, mean and variance of HYDE relative error were all lower than prior to the 20th century. In 1910, 31.78% of grid cells had greater than 100% relative error, where were distributed mainly on western Songnen and plains, and 20.56% of grid cells with 0–10% relative error were distributed mainly in uncultivated areas in the northern Xiao Hinggan Mountains. These above results indicated that cropland cover distribution in HYDE for 1910 still had obvious errors. After the 1930s, the relative error decreased, and grid cells with greater than 100% relative error comprised less than 15% of the region, where were dispersed on the eastern Sanjiang Plain and western Liaohe Plain. The relative error of most grid cells was −10%–100%, which were distributed mainly on the Liaohe and Songnen plains (−10% to −100%) and Xiao Hinggan and Changbai Mountains (−50% to −100%). Even in 2005, the proportion of grid cells with −10% to 10% relative error comprised only 15.89% of the region. This means modern cropland cover error in HYDE cannot be dismissed. In conclusion, HYDE had a lower accuracy of cropland distribution during the 18th–19th centuries, with the error originating mainly from overestimations in Jilin and Heilongjiang provinces. In the 20th century, HYDE overestimated cropland area in newly-cultivated areas, while underestimating cropland on old reclaimed land.

3 Discussion and conclusions

The gross, provincial and grid accuracy of global historical crop dataset in Northeast China can be summarized as follows:

(1) Results of gross and provincial assessments indicated that both SAGE and HYDE had obvious errors in the historical reconstruction of cropland area in Northeast China. The data could not reflect the real cropland changes over time, and therefore would increase uncertainties in model simulation for Northeast China if used directly.

(2) SAGE overestimated cropland area in Northeast China in all time periods, especially in the 18th and 19th centuries. Cropland area growth displayed a linear increase as recorded SAGE during 1700–1950, which was in total disagreement with the verified historical reclamation history of Northeast China. Cropland area changes at the provincial

Table 3 Relative error statistics of HYDE

Year	The proportion of total grid cells (%)							Min (%)	Max (%)	Mean (%)	Standard deviation
	−100– −50	−50– −10	−10–0	0–10a)	10–50	50–100	>100b)				
1700	0.00	2.80	0.00	27.10	3.74	2.80	63.55	41.55	1417.43	245.31	3.43
1780	13.08	6.54	0.93	20.56	0.93	4.67	53.27	−100	21449.86	2142.75	53.11
1910	14.95	13.08	4.67	20.56	9.35	5.61	31.78	−100	3405.62	288.63	6.97
1930	42.06	26.17	0.93	4.67	9.35	3.74	13.08	−100	1125.08	8.3	1.94
1950	25.23	27.10	2.80	18.69	12.15	4.67	9.35	−100	1671.72	9.37	1.96
1980	50.47	37.38	2.80	0.93	6.54	0.93	0.93	−100	283.74	−51.2	0.50
1990	36.45	33.64	11.21	4.67	7.48	4.67	1.87	−100	305.04	−34.37	0.59
2005	19.63	25.23	6.54	9.35	21.50	8.41	9.35	−97	899.73	16.39	1.26

a) Includes grid cells where both HYDE and CNEC cropland area is 0; b) includes grid cells unable to be computed because CNEC cropland is 0.

level did not show the shift in locus of farmland increase from the south to the north during the last 300 years.

(3) HYDE overestimated cropland area in the 18th and 19th centuries and underestimated it in the 20th century. Its reconstruction errors came mainly from Jilin and Heilongjiang provinces during the 18th century. HYDE cropland increase rate also differs from CNEC at both century and decadal timescales. It also could not represent the shifting locus of cropland increase since 1700.

(4) The grid assessment indicated significant disagreements in spatial distribution between HYDE and CNEC. HYDE had the highest error in the 18th century mainly from overestimation in Jilin and Heilongjiang provinces. In the 20th century, HYDE overestimated cropland in newly-cultivated areas and underestimated cropland on old reclaimed land.

The global datasets errors in cropland data across Northeast China originate mainly from two aspects. One is the reverse calculation method for historical cropland data based on modern spatial patterns. The other is the modern land-use map based on satellite data. The basic approach of the reconstruction methods using by the two global datasets is to reconstruct historical cropland distribution in uniform time intervals using modern spatial patterns as a reference and for linear interpolation. The establishment of both SAGE and HYDE included producing weighted maps based on modern data (land-cover map for RF, and population density map for HYDE) and allocating historical cropland according to map weighting. This idealized reverse calculation method ignores influences on cropland expansion and growth rate of reclamation by coupled natural and social factors such as migration, war and government policies. This reverse calculation method does not reflect human reclamation proceeding in Northeast China because the agricultural regions here were formed by large-scale migration and cultivation. This results in deviation of SAGE and HYDE from the known history of land use change. Modern maps of cropland cover are the foundation of the two global datasets, but their high-level accuracy cannot be guaranteed. The modern 5′-resolution cropland map of SAGE was derived from the DISCover land cover dataset [26], the remote sensing interpretation accuracy of which is only 66.9% at a global scale [27]. Although the DISCover data were combined with a variety of national and sub-national agricultural inventory data (primarily from the Food and Agriculture Organization, 1995), Ramankutty and Foley [26] recognized that their global cropland area is much larger than that of FAO. Especially in former Soviet Union, China, and some other countries, it is 1.5 time or more than the FAOSTAT data. HYDE reconstructed cropland maps with a 5′ resolution based on satellite data and agricultural statistics from FAO for the period 1990–2000. Two satellite datasets used by HYDE were IGBP DISCover data and GLC2000 data [28], which had similar accuracy of remote sensing interpretation (68.6%) to the DISCover data [29]. The combination of DISCover and GLC maps resulted in a satisfactory match with FAO data for all countries in HYDE. However, when compared with CNEC, the HYDE cropland area for Northeast China still has considerable error, especially in Heilongjiang Province where HYDE cropland is 42.7% and 26.51% less than CNEC for 1990 and 2000, respectively.

Accuracy assessment of global land use datasets for regional analysis needs to be studied further. Because of limitations in historical data and the reconstruction method, accuracy of reconstructed historical data is difficult to ensure at resolutions higher than the administrative unit level for historical statistics. Therefore, historical land-use data at regional scales are important for verifying and improving global land use data. This is necessary to support regional studies and integration between regional and global modeling. Two initiatives can be taken. The first is to correct the linear growth trend of global datasets using reconstructed regional series. The second is to substitute higher spatial resolution reconstructed regional data with that used in global datasets.

We would like to thank Professor Changming Sun at CSIRO for his kind help. This work was supported by National Natural Science Foundation of China (Grant Nos. 40571165, 40901099), Beijing Normal University Independent Research Fund (Grant No. 2009SAP-2) and National Key Technology R & D Program (Grant No. 2007BAC03A11).

1 Vitousek P M, Mooney H A, Lubchenco J, et al. Human domination of the Earth's ecosystems. Science, 1997, 277: 494–499
2 Lambin E F, Turner B L, Geist H J, et al. The causes of land-use and land-cover change: Moving beyond the myths. Glob Environ Change, 2001, 11: 261–269
3 Bolin B, Sukumar R. Global perspective. In: Watson R T, Noble I R, Bolin B, et al., eds. Land Use, Land-use Change, and Forestry. Cambridge: Cambridge University Press, 2000. 25–45
4 Hansen J E, Sato M, Lacis A, et al. Climate forcings in the Industrial era. Proc Natl Acad Sci USA, 1998, 95: 12753–12758
5 Feddema J, Oleson K, Bonan G, et al. A comparison of a GCM response to historical anthropogenic land cover change and model sensitivity to uncertainty in present-day land cover representations. Climate Dyn, 2005, 25: 581–609
6 IPCC. 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 and New York: Cambridge University Press, 2007. 180–185, 512
7 Houghton R A. Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. Tellus Ser B-Chem Phys Meteorol, 2003, 55: 378–390
8 Betts R A, Falloon P D, Goldewijk K K, et al. Biogeophysical effects of land use on climate: Model simulations of radiative forcing and large-scale temperature change. Agric Forest Meteorol, 2007, 142: 216–233
9 Zhen J Y, Lin S S, He F N. Recent progress in studies on land cover change and its regional climatic effects over China during historical times. Adv Atmos Sci, 2009, 26: 793–802
10 Ge Q S, Dai J H, He F N, et al. Land use changes and their relations with carbon cycles over the past 300 a in China. Sci China Ser D-Earth Sci, 2008, 51: 871–884
11 LUCC Scientific Steering Committee. Key findings of LUCC on its research questions. Glob Change Newsl, 2004, 63: 12–14
12 Goldewijk K K, Ramankutty N. Land cover change over the last three centuries due to human activities: The availability of new global

data sets. Geo J, 2004, 61: 335–344
13 Ramankutty N, Foley J A. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Glob Biogeochem Cycle, 1999, 13: 997–1027
14 Goldewijk K K, Drecht G V. HYDE 3: Current and historical population and land cover. In: Bouwman A F, Kram T, Goldewijk K K, eds. Integrated Modelling of Global Environmental Change: An Overview of IMAGE 2.4. Bilthoven: Netherlands Environmental Assessment Agency (MNP), 2006
15 Tian H, Melillo J M, Kicklighterc D W, et al. Regional carbon dynamics in monsoon Asia and its implications for the global carbon cycle. Glob Planet Change, 2003, 37: 201–217
16 Brovkin V, Sitch S, Bloh W V, et al. Role of land cover changes for atmospheric CO_2 increase and climate change during the last 150 years. Glob Change Biol, 2004, 10: 1253–1266
17 Oost K V, Quine T A, Govers G, et al. The impact of agricultural soil erosion on the global carbon cycle. Science, 2007, 318: 626–629
18 Matthews H D, Weaver A J, Eby M, et al. Radiative forcing of climate by historical land cover change. Geophys Res Lett, 2003, 30: 1055–1059
19 Wang H, Pitman A J, Zhao M, et al. The impact of land-cover modification on the June meteorology of China since 1700, simulated using a regional climate model. Int J Climatol, 2003, 23: 511–527
20 Chen X, Lei M, Tang J P. Simulating the effect of changed vegetation on the climate change in Eurasia (in Chinese). Adv Earth Sci, 2006, 21: 1075–1082
21 Li Q P, Ding Y H, Dong W J. A numerical simulation study of impacts of historical land-use changes on the regional climate in China since 1700. Acta Meteorol Sin, 2007, 21: 9–23
22 Ye Y, Fang X Q, Ren Y Y, et al. Cropland cover change in Northeast China during the past 300 years. Sci China Ser D-Earth Sci, 2009, 52: 1172–1182
23 Goldewijk K K. Estimating global land use change over the past 300 years: The HYDE database. Glob Biogeochem Cycle, 2001, 15: 417–433
24 Goldewijk K K. Three centuries of global population growth: A spatial referenced population density database for 1700–2000. Popul Environ, 2005, 26: 343–367
25 Ye Y, Fang X Q, Dai Y J, et al. Calibration of cropland data and reconstruction of rate of reclamation in Northeast China during the period of Republic of China (in Chinese). Prog Nat Sci, 2006, 16: 1419–1427
26 Ramankutty N, Foley J A. Characterizing patterns of global land use: An analysis of global croplands data. Glob Biogeochem Cycles, 1998, 12: 667–685
27 Scepan J. Thematic validation of high-resolution global land-cover data sets. Photogramm Eng Rem Sens, 1999, 65: 1051–1060
28 Goldewijk K K, Drecht G V, Bouwman A F. Mapping contemporary global cropland and grassland distributions on a 5×5 minute resolution. J Land Use Sci, 2007, 2: 167–190
29 Mayaux P, Eva H, Gallego J, et al. Validation of the global land cover 2000 map. IEEE Trans Geosci Remote Sens, 2006, 44: 1728–1739