Response to referee comments

We thank the two reviewers and the editor for the precious and constructive suggestions to improve our manuscript. We carefully revised our manuscript and addressed the comments of each of the two reviewers. Please find our point-by-point response below.

Referee #1

General comments. According to the World Meteorological Organization, 2023 is the hottest year on record. Therefore, it is of great significance to develop a long-term cropland dataset to explore the climatic effects of human land use. This study reconstructs millennial cropland for Northeast China. Topics fits the aims and scope of the ESSD. The following comments and suggestions should be considered for revisions.

Response: We sincerely thank the reviewers for valuing our work. Thank you for your insightful comments which have improved our work greatly. Please find our point-by-point response below.

1. --First, why only area estimation, and no spatial reconstruction? A 5' × 5' cropland dataset is developed for Northeast China from AD 1000 to 1200 by these authors (Gridded reconstruction of cropland cover changes in Northeast China from AD 1000 to 1200. https://link.springer.com/article/10.1007/s10113-023-02118-y). But in this study, only provincial-level or county-level cropland area is available. Why? Obviously, the datasets reconstructed in this study cannot be used by climate modelers. In addition, in terms of data from 1000 to 1200 years, is there any improvement in this paper compared to the paper mentioned above (Gridded reconstruction of cropland cover changes in Northeast China from AD 1000 to 1200)?

Response: Thank you for your helpful suggestion. In this study, 1000 to 1600 corresponds to historical provincial-level administrative districts, while 1700 to 2020 corresponds to modern county-level administrative districts.

The primary reasons are as follows: First, compared to the regional existing historical LUCC gridded reconstruction results, the cropland area data of administrative units is relatively more fundamental and reliable. Our research experiments indicate that there remains a certain degree of uncertainty in gridded reconstruction of cropland over long historical periods, even when using human factors supported by historical data in the allocation model (such as historical settlement points) (Jia et al., 2023). Long-term historical cropland gridded reconstructions need to consider using allocation methods that match the historical facts of different periods in local area. In theory, based on this dataset, researcher could even create datasets with a resolution of 1 meter or less. If climate modelers need to use gridded cropland datasets, they can easily convert our dataset into a customized grid spatial resolution dataset according to their required time range and main influencing factors (such as natural factors, human factors, etc.). For instance, some studies concluded that the HYDE dataset can be used as a map of agricultural potential and crop suitability, especially in periods before the advent of satellite imagery (Yu and Lu, 2018; Yu et al., 2021).

Second, for the 1700 to 2020 corresponds to modern county-level administrative districts, the average spatial scale of these counties in this dataset ranges between 0.5°×0.5° and 1°×1°. Theoretically, when readers use this dataset to convert it into the gridded dataset they need, the error can be controlled within between 0.5°×0.5° and 1°×1° even if they don't use natural or human factors to guide the allocation.

Third, this dataset can be applied in a wide range of scenarios (such as carbon emission and carbon neutrality, climate data construction, ecological footprint, and biological population assessment, etc.). The bilingual format and the administrative boundaries consistent with the current county-level administrative units in China also facilitate its use by scholars in the humanities and social sciences worldwide.

Realistically, compared to the primary data and reconstruction methods of Jia et al. (2023), this study directly used the results of the cropland area (1000-1200) of the above study. The main difference is that, this dataset provides provincial-level cropland area data for three time points (1000-1200) within the current administrative boundaries of Northeast China, consistent with the boundaries of the other 25 time points in this dataset. We are also very pleased to offer the reconstructed gridded cropland dataset in Northeast China from 1000 to 1200 as a reference solution for readers.

References:

Yu, Z. and Lu, C.: Historical cropland expansion and abandonment in the continental U.S. During 1850 to 2016, Glob. Ecol. Biogeogr., 27, 322-333, https://doi.org/10.1111/geb.12697, 2018.

Yu, Z., Jin, X., Miao, L., and Yang, X.: A historical reconstruction of cropland in china from 1900 to 2016, Earth Syst. Sci. Data, 13, 3203-3218, https://doi.org/10.5194/essd-13-3203-2021, 2021.

Jia, R., Fang, X., and Ye, Y.: Gridded reconstruction of cropland cover changes in Northeast China from ad 1000 to 1200, Reg. Envir. Chang., 23, 128, https://doi.org/10.1007/s10113-023-02118-y, 2023.

2. --Second, the applicability of the reconstruction method of estimating the cropland area for a small area by population. Generally speaking, estimating cropland by population is mostly applicable at continental to global scales. In the case of a small region, more other factors will affect the relationship between population and cropland.

Response: Thank you for your suggestion. The fundamental reason is that there are no direct historical records of cropland area in Northeast China from 1000 to 1600. When we selecting proxy indicators to reconstruct cropland area, the population data for this period is relatively complete and authoritative.

We acknowledge that at a regional scale, the factors affecting the relationship between population and cropland area are diverse and may change over time. Therefore, we combine historical facts (particularly from 1000 to 1600) and use the most authoritative historical population data in China: "*The History of Population of China*" (Wu and Ge, 2005a; Cao and Ge, 2005b). Based on the needs of this study, we extracted the population data (mainly divided into agricultural, non-agricultural, and military populations) that was consistent with the scope of this study area, and developed cropland calculation indicators for different historical periods corresponding to different population categories.

In several global LUCC datasets, such as HYDE, when estimating cropland based on population, either the population indicators (e.g.: per capita cropland) remain unchanged across different historical periods; or in the uncertainty estimates, homogeneously varied it with time across the globe; or to account for their uncertainties, different population databases were used and the upper and lower ends of an uncertainty range were assessed. However, these estimation methods can still lead to errors in cropland area on a global or regional scale. Therefore, when this study used historical population data to reconstruct cropland, special attention was paid to two key points: 1) authoritative and accurate population data, and 2) cropland areas corresponding to different population categories based on historical facts (Please see Line 632-654). Thus, we believe this method is relatively applicable for Northeast China from 1000 to 1600.

Line 632-654**:** (3) The two proxy indicators of 14 *Mu* (0.93 hm^2, the average annual potential cropland area per Man of the agricultural population) and 2 *Mu* (0.13 hm^2, the average cropland area per household in the nonagricultural population) from 1000 to 1600 may lead to inaccuracies in cropland estimation. The reasons for using population to reconstruct cropland during this period have been detailed in the previous section, necessitating further analysis and clarification of the corresponding cropland-related indicators.

Firstly, the conclusion of 14 *Mu* per Man for agricultural population during the Liao and Jin Dynasties (1000~1200) is primarily derived from historical records in the Jin Dynasty (1200) and the relationship between population and cropland in the early Qing Dynasty (1661~1680) (Jia et al., 2023). There are two reasons why 14 *Mu* was used in the Yuan and Ming Dynasties (1300~1600): one reason is the agricultural household size and the ratios of Man in agricultural household in Northeast China during the Yuan and Ming Dynasties (1300~1600) are closer to those of the Liao and Jin Dynasties (1000~1200) (Table 1). And the per capita cropland area owned by agricultural population in the Liao-Jin-Yuan-Ming periods (1000~1600) consistently ranged between 4 and 5 *Mu* (0.27~0.33 hm^2), slightly higher than the subsistence level of 3 *Mu* per capita in previous studies for the same historical period in this region (Ye et al., 2009; Fang et al., 2006; Shi, 1990), which is relatively reasonable. The second reason is that there were no significant changes in agricultural production technology in Northeast China during the Liao-Jin-Yuan-Ming periods (1000~1600), and the population declined significantly compared with the Liao and Jin Dynasties (1000~1200) due to factors such as warfare. However, considering the social stability at standard time-points during the Yuan and Ming Dynasties (1300~1600), the strong willingness of the agricultural population towards cultivation, and the limitations of individual cultivation capabilities, the cropland from the Liao and Jin Dynasties could be relatively easily inherited and reclaimed by descendants.

Secondly, Similar to the agricultural population, considering the non-agricultural household size, stable agricultural production technology, the historical inheritance of most ethnic groups, this study continues to use 2 *Mu* as the calculation indicator of non-agricultural population in the Yuan and Ming Dynasties (1300~1600) (Cong, 1993a; Cong, 1993b; Wu and Ge, 2005a; Cao and Ge, 2005b; Liu et al., 2016).

References:

Wu, S. and Ge, J.: The History of Chinese Population, Volume 3, Fudan University Press, Shanghai, China, 2005a (in Chinese).

Cao, S. and Ge, J.: The History of Chinese Population, Volume 4, Fudan University Press, Shanghai, China, 2005b (in Chinese).

3. --Third, failure to evaluate the reliability or accuracy or uncertainties of the reconstructed dataset will affect the user's use of the dataset. The comparison with the global dataset does not indicate the reliability of the dataset developed in this paper, because the global dataset itself has a large degree of uncertainty. The fact that the reconstruction results in this paper are very different from the global dataset does not mean that the dataset developed in this paper is reliable.

Response: Thank you for your insightful suggestion. We acknowledge that the current paper's reliability, accuracy, or uncertainties assessments are not yet sufficiently comprehensive. We have made every effort to supplement the relevant assessments and uncertainty analysis as much as possible. Please see Line 428-495; 619-678, new Fig. 6, new Fig. 7 and Table S3.

***Line 428-495*: 4.1 Credibility assessment**

Based on the study of Fang et al. (2020), three methods including accuracy assessment, rationality assessment, and likelihood assessment, can be used to assess the credibility of historical LUCC dataset. Regarding the likelihood assessment, in reconstructing cropland area from 1985 to 2020, we selected eight RS products to assess the consistency. Based on the control of cropland survey data, this study identified high-consistency and high-priority pixels as cropland pixels for this dataset and evaluated and validated the accuracy of the integration results. Theoretically, compared with any single RS products used in this study during this period, the total amount of cropland area in this study is relatively more accurate and the spatial distribution is relatively more reasonable.

4.1.1 Accuracy assessment

The cropland data at lower spatial scales can be used to evaluate the accuracy of reconstructed cropland area. Due to the availability of county-level cropland survey data, we selected the county-level first general land investigation at 1985 and the county-level data from the third national land survey at 2020 for comparison. As shown in Fig. 6, the determination coefficients between the cropland area from this study and the cropland area from the survey data for 1985 and 2020 are 0.9582 and 0.9892 respectively. This indicated that the overall accuracy of the reconstructed cropland area at county-level was relatively high, and our constrained integration method that combines multisource cropland cover products with survey data can well match the spatial distribution of cropland cover in Northeast China.

In addition, from 1985 to 2020, the identified high-consistency and high-priority pixels as cropland pixels based on constrained integration method may lead to errors with survey data (Table S3). The relative errors between the cropland area of this study and the cropland survey data for the period 1985 to 2020 as -1.35%, 4.02%, 5.17%, 1.10%, 0.21%, -1.93%, 0.25% and 0.67%, respectively. The vast majority of errors are around 1%, with the larger errors are 4.02% and 5.17%, which indicates that the reconstructed cropland area in this study is relatively accurate from 1985 to 2020.

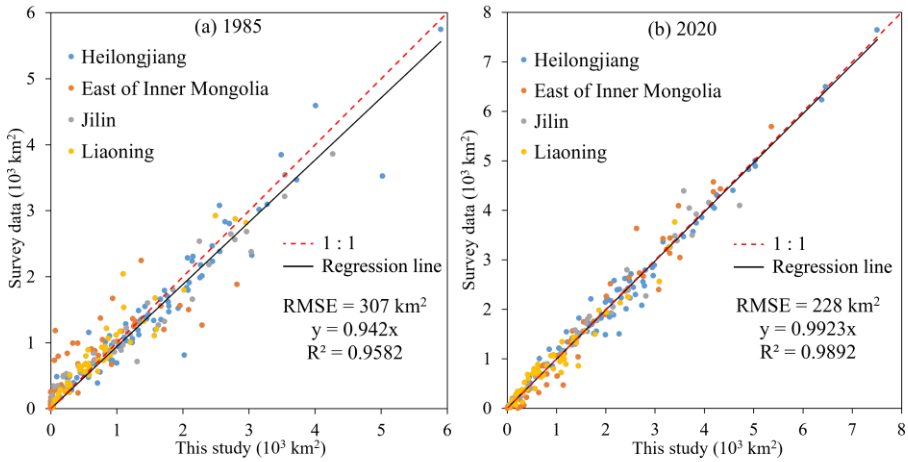

Figure 6: Correlation between the cropland data of this study and survey cropland data at county-level in the Northeast China in 1985 and 2020.

4.1.2 Rationality assessment

Due to the unavailability of actual historical land cover data, we used the actual historical agricultural development of Northeast China as a reference standard for rationality assessment. As one of the cases evaluating the distribution rationality of the HYDE3.2 cropland cover in Northeast China over the past millennium, Fang et al. (2020) analyzed changes in the northern boundary and spatial distribution of settlement relics in the Liao, Jin, Yuan, and Ming periods (916~1644), as well as changes in the cumulative number of towns and spatial distribution of towns in the three provinces of Northeast China during the Qing Dynasty (1644~1911). The unique development history of the Northeast China shown in this case is basically consistent with the process of increase or decrease and spatial distribution of the total cropland area during the same period reconstructed by this study, which reflects the rationality of this dataset.

In addition, this study attempts to briefly summarize the population changes, settlements changes (the settlement relics and the administrative division points derived from Jia et al. (2018) and the Historical Atlas of China (Tan, 1982a; Tan, 1982b)) (Fig. S2), warfare, and land policies that may have influenced land cultivation in Northeast China during the Liao, Jin, Yuan, and Ming periods (1000~1600). The population and settlements in Northeast China from 1000 to 1600 exhibited phase changes of expansion-reduction-expansion, with possible reasons including the Liao and Northern Song Dynasties signed the "*Chanyuan Alliance* (澶渊之盟)" in 1004 after war, the Jin and Southern Song Dynasties signed the "*Shaoxing Peace Treaty* (绍兴和议)" in 1141 after war, the Jin and Southern Song Dynasties signed the "*Longxing Peace Treaty* (隆兴和议)" in 1164 after war. During the three treaties and related wars, both the Liao and Jin dynasties in the north benefited significantly. They not only received reparations but also resettled large numbers of captives to the present-day Northeast China to engage in agricultural and other productive activities. Historical records also indicate that the rulers of the Liao and Jin dynasties during this period both attached much importance to agricultural production (Wu and Ge, 2022; Han, 1999; Toqto'A, 1974; Toqto'A, 1975).

From 1211, when Genghis Khan personally led the Mongol army to attack the Jin Dynasty, until 1233, the Mongols had essentially gained control over the entire Northeast China. Using this region as a base, they also conducted war against Goryeo (present-day Korean Peninsula), which lasted until 1259. From 1259 to 1287, the Mongols made several attempts to establish governing institutions in Northeast China, but faced continuous rebellions. It wasn't until the Yuan Dynasty subdued the rebellions and established the Liaoyang Province in 1287 that effective governance began in the Northeast China. However, during this period, the region suffered from continuous warfare, significant population loss, and severe disruptions to agricultural production (Xue, 2006, 2012). According to the Dynastic History of Yuan Dynasty, from 1294 to 1345, the Yuan government provided relief to Liaoyang Province 40 times. Additionally, rebellions in the Northeast China persisted from 1343 onwards, only being effectively subdued the rebellions by 1362, just six years before the collapse of the Yuan Dynasty in 1368 (Song, 1976; Xue, 2006, 2012).

In 1368, the Ming Dynasty was established, and remnants of the Yuan Dynasty retreated to the northern grassland, known as the Northern Yuan Dynasty (Tatar), which partly within our study area. It wasn't until 1389 that the Ming Dynasty established the "*Uriyangqa three Commanderies* (兀良哈三卫)" in the region from present-day Qiqihar city to Baicheng city, gaining certain practical control over the region. However, from 1399 to 1402, the Ming Dynasty faced the internal strife of the "*Jingnan Campaign* (靖难之役)" weakening its influence over the Northeast China, allowing some ethnic minorities to further occupy territories to the south. In 1409, the Ming Dynasty established the Dusi of Nuergan, reflecting their policy of appeasement and assimilation towards ethnic minorities in the Northeast China. In 1449, the Ming Dynasty experienced the "*Tumu Crisis* (土木之变)", prompting substantial efforts to fortify defensive structures. This also greatly strengthened the defensive capabilities of the Ming Great Wall in the Northeast China and confined the major agricultural population and agricultural areas of the Northeast China within the Dusi of Eastern Liao (south of the Ming Great Wall in the Northeast China). This situation persisted until the Ming Dynasty's collapse in 1644 (Cao and Ge, 2022; Fan, 2015; Cao and Ge, 2005; Zhang, 1974). All these pieces of evidence contribute to the validation of the rationality of our dataset to a certain degree.

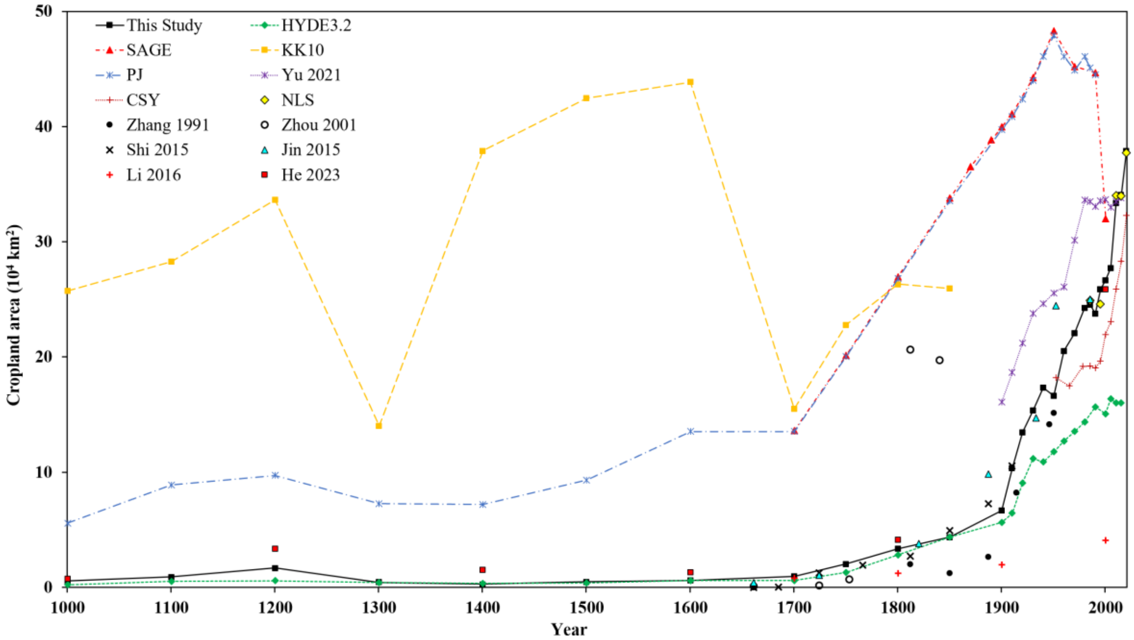

Figure 7: Comparison of total cropland area from global historical LUCC datasets, previous studies and this study in the Northeast China. The abbreviations used in the figure are as follows: HYDE3.2 refers to Goldewijk et al. (2017); SAGE refers to Ramankutty et al. (2008) and Ramankutty and Foley. (1999); KK10 refers to Kaplan et al. (2011); PJ refers to Pongratz et al. (2008); Yu 2021 refers to Yu et al. (2021); CSY denotes the Chinese Statistical Yearbook (refer to provincial and prefectural statistical yearbook); NLS denotes the National Land Survey (1985 refer to the first general land investigation; 1995 refers to the first national land survey; 2010 and 2015 refer to the second national land survey; 2020 refers to the third national land survey); Zhang 1991 refers to Zhang (1991); Zhou 2001 refers to Zhou (2001); Shi 2015 refers to Shi (2015); Jin 2015 refers to Jin et al. (2015); Li 2016 refers to Li et al. (2016); He 2023 refers to He et al. (2023).

***Line 619-678*: 4.4 Uncertainty analysis**

In this study, the uncertainty mainly consisted in two aspects: the definition and selection of data, the application of methods. Regarding the data aspect: (1) In this study, the definition of cropland before 1950 is: the sum of arable land and land under permanent crops, and the temporary changes in land use and fallow land during historical periods were not considered. The cropland area for 1950 and later are basically consistent with the identification rules in the National Land Survey. Although the temporary changes in land use and fallow land during historical periods, this may still result in our reconstruction slightly less cropland than actual historical period.

(2) Due to the completeness of historical documents, the reconstruction results of cropland for seven time points from 1000 to 1600 in this study are at the provincial-level, which may not finely reflect the spatiotemporal characteristics of cropland. Especially between 1000 and 1300, the results may lead readers to mistakenly believe that cropland were evenly distributed across the entire Northeast China. However, based on the distribution of settlement relics during this period, cropland may mainly distribute on the Liaohe Plain and on the southern part of the Songnen Plain, then reduced southward into Liaoning Province.

(3) The two proxy indicators of 14 *Mu* (0.93 hm^2, the average annual potential cropland area per Man of the agricultural population) and 2 *Mu* (0.13 hm^2, the average cropland area per household in the nonagricultural population) from 1000 to 1600 may lead to inaccuracies in cropland estimation. The reasons for using population to reconstruct cropland during this period have been detailed in the previous section, necessitating further analysis and clarification of the corresponding cropland-related indicators.

Firstly, the conclusion of 14 *Mu* per Man for agricultural population during the Liao and Jin Dynasties (1000~1200) is primarily derived from historical records in the Jin Dynasty (1200) and the relationship between population and cropland in the early Qing Dynasty (1661~1680) (Jia et al., 2023). There are two reasons why 14 *Mu* was used in the Yuan and Ming Dynasties (1300~1600): one reason is the agricultural household size and the ratios of Man in agricultural household in Northeast China during the Yuan and Ming Dynasties (1300~1600) are closer to those of the Liao and Jin Dynasties (1000~1200) (Table 1). And the per capita cropland area owned by agricultural population in the Liao-Jin-Yuan-Ming periods (1000~1600) consistently ranged between 4 and 5 *Mu* (0.27~0.33 hm^2), slightly higher than the subsistence level of 3 *Mu* per capita in previous studies for the same historical period in this region (Ye et al., 2009; Fang et al., 2006; Shi, 1990), which is relatively reasonable. The second reason is that there were no significant changes in agricultural production technology in Northeast China during the Liao-Jin-Yuan-Ming periods (1000~1600), and the population declined significantly compared with the Liao and Jin Dynasties (1000~1200) due to factors such as warfare. However, considering the social stability at standard time-points during the Yuan and Ming Dynasties (1300~1600), the strong willingness of the agricultural population towards cultivation, and the limitations of individual cultivation capabilities, the cropland from the Liao and Jin Dynasties could be relatively easily inherited and reclaimed by descendants.

Secondly, Similar to the agricultural population, considering the non-agricultural household size, stable agricultural production technology, the historical inheritance of most ethnic groups, this study continues to use 2 *Mu* as the calculation indicator of non-agricultural population in the Yuan and Ming Dynasties (1300~1600) (Cong, 1993a; Cong, 1993b; Wu and Ge, 2005a; Cao and Ge, 2005b; Liu et al., 2016).

Regarding the method aspect: (1) From 1700 to 1980, cropland areas at multiple time points in this study were derived through linear interpolation and polynomial curve fitting. Although we have fully considered historical facts and other research conclusions (Fang et al., 2020; Ye et al., 2009; Fang et al., 2005) when selecting the interpolation time points, 1860 was chosen as the dividing point between slow growth and rapid growth. This method, compared to data recorded at each specific historical point, may affect the accuracy of the value at those standard time points.

(2) From 1700 to 1980, the county-level administrative boundaries in the published data used in this study differ from the modern county-level administrative boundaries used in this study. Especially in the CNEC data (Ye et al., 2009) in 1683, 1735 and 1780, there is county-level in Liaoning province, Assistant Governorate Jurisdiction (prefecture-level) in Heilongjiang and Jilin province. This would result in counties belonging to different Assistant Governorate Jurisdictions in present-day having the same cropland fraction. This problem is difficult to correct further because the lowest administrative level in Northeast China available in historical data during this period is Assistant Governorate Jurisdiction (prefecture-level).

(3) From 1985 to 2020, the land survey data utilized in this dataset might exhibit uncertainties in early cropland data due to backward technology and other factors. Additionally, the use of a uniform correlation coefficient to correct the cropland statistics data for the entire Northeast China may affect the accuracy of the cropland area in localized areas, this may lead to a lower cropland area at previous time-points. To mitigate the impact of these uncertainties on our dataset during this period, this study mainly adopts two methods: Firstly, this study mainly selects the standard time-points data after the nationwide surveys, avoiding the use of annual land change survey data. For instance, the cropland area in 1985 in this dataset is based on the first general land investigation around 1985; the cropland area in 1995 is based on the first NLS's standard time-point data on October 31, 1996; the cropland area in 2010 is based on the second NLS's standard time-point data on December 31, 2009; and the cropland area in 2020 is based on the third NLS's standard time-point data on December 31, 2019. Secondly, this study uses

correlation coefficients to correct the statistical data by category and time point. For instance, the average correlation coefficient of the second and third NLSs with corresponding statistical data is used to correct the statistical data for the 1990, 2000, and 2005; the correlation coefficient of the 1985 first general land investigation with corresponding statistical data is used to correct the statistical data for the 1950~1980.

Table S3. Supplementary information showing the total cropland area of global LUCC datasets, previous representative published studies and our dataset in Northeast China from 1000 to 2020 (10^4 km^2).

Period	1000	1100	1200	1300	1400	1500	1600	1661	1685	1700	1724	1750	1753	1766	1800	1812	1820	1840	1850	1870	1887	1890	1900
This Study	0.55	0.91	1.69	0.44	0.28	0.49	0.59			0.96		2.04			3.33				4.36				6.66
HYDE3.2	0.23	0.53	0.57	0.42	0.34	0.37	0.59			0.59		1.29			2.81				4.36				5.63
SAGE										13.61		20.16			26.97				33.80	36.52		38.85	39.97
KK10	25.72	28.26	33.65	14.00	37.90	42.48	43.85			15.49		22.79			26.32				25.93				
PJ	5.58	8.88	9.72	7.27	7.19	9.30	13.53			13.53		20.04			26.80				33.57				39.74
Yu 2021																							16.07
CSY																							
NLS																							
Zhang 1991																2.02			1.21		2.66		
Zhou 2001								0.09			0.16		0.68			20.64		19.71					
Shi 2015								0.00	0.02		1.25			1.95		2.70			4.94		7.27		
Jin 2015								0.43			1.05						3.80				9.82		
Li 2016										0.78					1.23								1.98
He 2023	0.77		3.35		1.51		1.31								4.14								

Period	1910	1914	1920	1930	1933	1940	1946	1950	1952	1960	1965	1970	1978	1980	1985	1990	1995	2000	2005	2010	2015	2020
This Study	10.33		13.44	15.37		17.31		16.63		20.53		22.07		24.28	24.56	23.76	25.87	26.64	27.74	33.37	34.08	37.90
HYDE3.2	6.44		9.08	11.19		10.88		11.76		12.71		13.53		14.37		15.66		15.06	16.38	16.00	16.00	
SAGE	41.11			44.24				48.30				45.21				44.67		31.99				
KK10																						
PJ	40.87		42.37	44.02		46.08		47.92		46.09		44.90		46.08	45.12	44.49						
Yu 2021	18.64		21.21	23.78		24.63		25.56		26.09		30.13		33.61	33.49	33.05	33.55	33.68	32.99	33.50	33.79	
CSY									18.22		17.49		19.18		19.21	19.04	19.66	21.96	23.07	25.79	27.97	32.06
NLS															24.89		24.60			34.03	33.99	37.64
Zhang 1991		8.23					14.17	15.17														
Zhou 2001																						
Shi 2015	10.55																					
Jin 2015					14.71				24.46						25.01							
Li 2016																		4.09				
He 2023																		25.89				

The abbreviations used in the table are as follows: HYDE3.2 refers to Goldewijk et al. (2017); SAGE refers to Ramankutty et al. (2008) and Ramankutty and Foley. (1999); KK10 refers to Kaplan et al. (2011); PJ refers to Pongratz et al. (2008); Yu 2021 refers to Yu et al. (2021); CSY denotes the Chinese Statistical Yearbook (refer to provincial and prefectural statistical yearbook); NLS denotes the National Land Survey (1985 refer to the first general land investigation (Committee of Integrative Survey of Natural Resources and Committee of National Planning of Chinese Academy of Sciences, 1989); 1995 refers to the first national land survey (Li, 2000); 2010 and 2015 refer to the second national land survey; 2020 refers to the third national land survey (https://gtdc.mnr.gov.cn/Share#/)); Zhang 1991 refers to Zhang (1991); Zhou 2001 refers to Zhou (2001); Shi 2015 refers to Shi (2015); Jin 2015 refers to Jin et al. (2015); Li 2016 refers to Li et al. (2016); He 2023 refers to He et al. (2023).

4. --Fourth, writing is not done from the perspective of data development (Data description paper), it is more like a research paper. For example, the core content of the results should not be the analysis of the spatio-temporal characteristics of cropland changes, but the rationality, reliability, accuracy, and potential uses of the data products developed in this paper. More specific comments are as follows.

Response: Thank you for your insightful suggestion. We reorganized the paper to make it more like a data description paper. It mainly includes: 1. Deletion of the conclusion section (Please see "Conclusion section", Line 684-703); 2. Reorganized of introduction, results and discussion sections (Please see "Introduction section", Line 25-82; "Results section", Line 393-426; "Discussion section", Line 427-678); 3. Add the credibility assessment and uncertainty analysis (Please see Line 428-495; 619-678); 4. Potential uses are added to the introduction and abstract sections (Please see Line 22-24, 39-41).

***Line 22-24*:** Additionally, the dataset shows relatively better credibility assessment results, which can provide a refined data base for historical LUCC dataset reconstruction, carbon emission estimation, climate data construction, etc. The dataset can be downloaded from https://doi.org/10.6084/m9.figshare.25450468.v2 (Jia, 2024).

***Line 39-41*:** have been extensively employed in global change research. Such as carbon emission and carbon neutrality (Xu et al., 2024), climate data construction (Gortan et al., 2024), ecological footprint (Wang et al., 2024), and biological population assessment (Ye et al., 2024), etc.

5. --Title and Introduction. Why reconstruction for 1000 to 2020 in Northeast China? More explanations are necessary. Based on Figure 5, From 1000-1700, there was only a small area of cropland in Northeast China. Line 393, In 1200, cropland fraction of 1.17%; In 1400, line 395, cropland fraction of only 0.19%. The environmental impact

of such a small area of cropland is completely negligible. Based on figure 5, the topic for past 300 years (Ye, Y., Fang, X., Ren, Y., Zhang, X., and Chen, L.: Cropland cover change in northeast china during the past 300 years, Science China Earth Sciences, 52, 1172-1182, https://doi.org/10.1007/s11430-009-0118-8, 2009.) is good, but for 1000 to 2020 may be not a good research topic.

Response: Thank you for your insightful suggestion. We chose to reconstruct the cropland changes in Northeast China over the past millennium mainly for the following reasons: 1. From the LUCC projects carried out by IGBP and IHDP in the last century to the recent PAGES (The PAGES (Past Global Changes) project is an international effort to coordinate and promote past global change research. The primary objective is to improve our understanding of past changes in the Earth system in order to improve projections of future climate and environment, and inform strategies for sustainability.), LandCover6k (The goal of LandCover6k was to produce datasets on past land-cover and land-use on continental and global spatial scales that are useful for climate modeling studies on land-use as a climate forcing.) and other research projects, the importance of accurate long-term historical LUCC datasets has been emphasized.

2. The land reclamation in Northeast China exhibited a unique pattern: during the period between the two land reclamations (eleventh and twelfth centuries; from the nineteenth century to present), there was a prolonged period of nomadism in this area (Jia et al., 2023). In addition, the global historical LUCC datasets fail to demonstrate the historical fact of cropland cultivation in the study area from 1000 to 1200.

3. We want to better demonstrate the changes in human impact on terrestrial environments, ranging from near-natural original states (natural dominance) to significant alterations induced by extensive human intervention (anthropogenic dominance), particularly since the Industrial Revolution.

4. Northeast China has now become one of the most important agricultural regions in China and the world. Focusing solely on the past 300 years of research is not conductive to the exploration of long-term LUCC effects in critical agricultural areas worldwide (He et al., 2023). If feasible, we are willing to expand the dataset regarded as "truth values" to encompass a broader historical span, thereby enhancing the

credibility of historical cropland area.

5. We also reorganized the introduction to describe the novelty of this manuscript (Please see Line 73-82).

***Line 73-82*:** The dataset in this study presents a critical update and extension of the former historical cropland cover change in the three provinces of Northeast China over the past 300 years (Ye et al., 2009). Throughout the prolonged agricultural development, the natural vegetation landscape in the Northeast China has undergone notable transformations. In this study, we used the improved historical cropland reconstruction methods to reconstruct 28 time-points cropland area by assimilating multiple data sources in Northeast China from 1000 to 2020. The mainly new features of this dataset include: (1) Extended the reconstruction period to 1000~2020, aligning with the standard time-points of internationally established global historical LUCC datasets; (2) the reconstruction included the entire East of Inner Mongolia, which area accounts for approximately 45% of the Northeast China. (3) the smallest administrative divisions for the reconstructed cropland are at the provincial-level from 1000 to 1600, and at the county-level from 1700 to 2020. Our main objective is to provide a long-term time series of cropland area change dataset in Northeast China that is close to the historical "truth value" under a unified standard.

References:

He, F., Yang, F., Zhao, C., Li, S., and Li, M.: Spatially explicit reconstruction of cropland cover for china over the past millennium., Science China Earth Sciences, 66, 111-128, https://doi.org/10.1007/s11430-021-9988-5, 2023.

Jia, R., Fang, X., and Ye, Y.: Gridded reconstruction of cropland cover changes in Northeast China from ad 1000 to 1200, Reg. Envir. Chang., 23, 128, https://doi.org/10.1007/s10113-023-02118-y, 2023.

6. --Data and Methods. Not clear enough. For example, Line 91-115, It only introduces population data, per household population data, and interpolates the population according to the population growth rate, and does not involve how to estimate the cropland area at all. Line 107-110 mentions how to estimate the area of cropland, but it is very simple and there is no specific method. As far as Northeast China is concerned, why such an estimate is reasonable is not explained at all. From the perspective of

historical land use reconstruction, estimating cropland area based on population as a proxy is only applicable to large-scale scales such as global and continental. For example, HYDE uses population to estimate the world's historical cropland. It also makes sense to reconstruct China's historical cropland in this way. But the Northeast is only a small region of China, so there is a lot of uncertainty in the results of this estimate.

Response: Thank you for your constructive suggestion, and we apologize if this was confusing. The main algorithm applied in the Liao, Jin, Yuan, and Ming Dynasties (1000-1600) can be found in the supplementary materials.

The reason and the credibility of using population to reconstruct cropland is similar to your second question, to which we have already responded and discussed more fully in the "Uncertainty analysis section". Please see Line 632-654. Regarding the importance of Northeast China, we added content in the "Introduction section". Please see Line 60-72.

***Line 632-654*: 4.4 Uncertainty analysis**

(3) The two proxy indicators of 14 *Mu* (0.93 hm^2, the average annual potential cropland area per Man of the agricultural population) and 2 *Mu* (0.13 hm^2, the average cropland area per household in the nonagricultural population) from 1000 to 1600 may lead to inaccuracies in cropland estimation. The reasons for using population to reconstruct cropland during this period have been detailed in the previous section, necessitating further analysis and clarification of the corresponding cropland-related indicators.

Firstly, the conclusion of 14 *Mu* per Man for agricultural population during the Liao and Jin Dynasties (1000~1200) is primarily derived from historical records in the Jin Dynasty (1200) and the relationship between population and cropland in the early Qing Dynasty (1661~1680) (Jia et al., 2023). There are two reasons why 14 *Mu* was used in the Yuan and Ming Dynasties (1300~1600): one reason is the agricultural household size and the ratios of Man in agricultural household in Northeast China during the Yuan and Ming Dynasties (1300~1600) are closer to those of the Liao and Jin Dynasties (1000~1200) (Table 1). And the per capita cropland area owned by agricultural population in the Liao-Jin-Yuan-Ming periods (1000~1600) consistently ranged between 4 and 5 *Mu* (0.27~0.33 hm^2), slightly higher than the subsistence level of 3 *Mu* per capita in previous studies for the same historical period in this region (Ye et al., 2009; Fang et al., 2006; Shi, 1990), which is relatively reasonable. The second reason is that there were no significant changes in agricultural production technology in Northeast China during the Liao-Jin-Yuan-Ming periods (1000~1600), and the population declined significantly compared with the Liao and Jin Dynasties (1000~1200) due to factors such as warfare. However, considering the social stability at standard time-points during the Yuan and Ming Dynasties (1300~1600), the strong willingness of the agricultural population towards cultivation, and the limitations of individual cultivation capabilities, the cropland from the Liao and Jin Dynasties could be relatively easily inherited and reclaimed by descendants.

Secondly, Similar to the agricultural population, considering the non-agricultural household size, stable agricultural production technology, the historical inheritance of most ethnic groups, this study continues to use 2 *Mu* as the calculation indicator of non-agricultural population in the Yuan and Ming Dynasties (1300~1600) (Cong, 1993a; Cong, 1993b; Wu and Ge, 2005a; Cao and Ge, 2005b; Liu et al., 2016).

***Line 60-72*:** Northeast China is one of the most important grain bases in China today. The grain output constitutes 25.18% of the national total, with corn and soybean contributing 41.64% and 56.20%, respectively (National Bureau Of Statistics, 2023). A study has indicated that the supply centers for China's three major grains (wheat, corn, rice) significantly moved to the Northeast from 2000 to 2020, while the demand centers did not move simultaneously. This shift underscores the rapidly increasing importance of the Northeast China in ensuring China's food security (Xuan et al., 2023). Furthermore, the majority of China's black soil is distributed in Northeast China, which provides an important foundation for the productivity of crops. A study has pointed out that compared to other global black soil regions, the Northeast black soil region's yields of eight major crops (excluding rice) remained in the top three among the world's main black soil distribution countries from 2000 to 2015, with Russia and Ukraine occupying the first two positions (Wang et al., 2024). Additionally, long-term precise cropland area change data reflects the significance for soil and water conservation research in Northeast China, thereby ensuring food security. A typical case study in the Northeast China examined the long-term effects of cultivation on soil carbon, nitrogen, and bacterial community in the Northeast black soil region. The results indicated that prolonged cultivation (e.g., 152 years) led to a negatively and exponentially decline in SOC and total nitrogen (Liu et al., 2024).

7. --In addition, for 1000, 1100, and 1200, what’s the difference between this study and

the paper mentioned above (Gridded reconstruction of cropland cover changes in Northeast China from AD 1000 to 1200).

Response: Thank you for your comment. This study is a continuation of a previous study. Realistically, compared to the primary data and reconstruction methods of Jia et al. (2023), this study directly used the results of the cropland area (1000-1200) of the above study. The dataset in this study presents a critical update and extension of the former historical cropland cover change in the three provinces of Northeast China over the past 300 years (Ye et al., 2009). The main difference is that, this dataset provides provincial-level cropland area data for three time points (1000-1200) within the current administrative boundaries of Northeast China, consistent with the boundaries of the other 25 time points in this dataset. We are also very pleased to offer the reconstructed gridded cropland dataset in Northeast China from 1000 to 1200 as a reference solution for readers.

References:

Jia, R., Fang, X., and Ye, Y.: Gridded reconstruction of cropland cover changes in Northeast China from ad 1000 to 1200, Reg. Envir. Chang., 23, 128, https://doi.org/10.1007/s10113-023-02118-y, 2023.

Ye, Y., Fang, X., Ren, Y., Zhang, X., and Chen, L.: Cropland cover change in northeast china during the past 300 years, Science China Earth Sciences, 52, 1172-1182, https://doi.org/10.1007/s11430-009-0118-8, 2009.

8. --Line 170-197；Line 257-278. Introduce too much about the estimation methods in published papers (Ye et al, 2009; Tian et al., 2005). It needs to be drastically cut, and readers can read these papers at all. In short, the writing of the method section is too lengthy and will scare off the vast majority of readers.

Response: Thank you for your helpful suggestion, and we apologize if this was confusing. We have deleted the detailed description about the estimation methods in published papers in the main text to make the article more readable. Considering that the papers we cited in this section is not available in English, we have added these contents to the supplement material for readers who need it.

9. --Line 232. Correct negative or zero values of cropland. If the estimated results have a negative value, then there must be a problem with the previous interpolation and fitting methods, and we have reason to suspect that all the results obtained by the interpolation are problematic. Just correct negative or zero values of cropland isn't enough, what about the other results? From this point, it can be seen that this paper needs to have an uncertainty assessment of the estimation results, otherwise readers will not dare to use this data product to carry out downstream research.

Response: Thank you for your insightful suggestion, and we apologize if this was confusing. We have reorganized the structure of the original text, and added more specific and detailed description based on careful checking and correction of the existing errors in the original text. Please see Line 219-269 and the Table S2.

***Line 219-269*:** (3) Linear interpolation and polynomial curve fitting to obtain the cropland area: Previous studies have used the linear interpolation and polynomial curve fitting to reconstruct cropland areas (He et al., 2017; Jin et al., 2015; Ramankutty and Foley, 1999; Wei et al., 2016; Wei et al., 2021; Ye et al., 2015; Yu, 2019; Fang et al., 2021), and the interpolated data did not reduce the credibility of their datasets. In addition, previous studies have shown that in the process of reclamation in the Northeast China over the past 300 years, 1860 was a dividing point between slow growth and rapid growth, mainly due to the implementation of the immigration and reclamation policy by the Qing government (Fang et al., 2020; Ye et al., 2009; Fang et al., 2005; Kong and Feng, 1989). Therefore, this study selected the CNEC data (Ye et al., 2009) in 1683, 1735, 1780, 1908 and 1914 for linear interpolation and polynomial curve fitting of cropland area data for each county or district in the three provinces of the Northeast China, obtaining data for 1700, 1750, 1800, 1850 and 1900. In addition, this study selected the data from Tian (Tian, 2005) in 1724, 1782, 1868, and 1911; the CNEC data (Ye et al., 2009) in 1735; the data from Ye (Ye and Fang, 2012) in 1916 for linear interpolation and polynomial curve fitting to obtain cropland area data for 1700, 1750, 1800, 1850, and 1900 in the Eastern of Inner Mongolia. The problems that may be encountered during the operation and the corresponding solutions are as follows:

①Linear interpolation and determination of zero values. The time points involved in this issue include 1700 and 1750 for the three provinces of Northeast China; 1750, 1800, and 1850 for East of Inner Mongolia. For instance, in Northeast China, the cropland area in each county in 1700 is interpolated based on records from 1683 and 1735. At 1700, there are no negative values, but there may be zero values. Specifically, the cropland value in 1683 is 0, while there is definite value in 1735. Our solution involves consulted contemporary county gazetteers to verify the history of land reclamation in 1700. If so, a polynomial curve fitting trend extrapolation was applied to obtain the proportional relationship at the provincial level for adjacent points on the extrapolated trend. Then this proportion was multiplied by the cropland area of the county at the adjacent time-point to obtain the cropland area at that time-point. If the land was not reclaimed, the value at that time point was considered as zero. Similarly, other counties involved in interpolation adopt the same solution when encountering this situation.

②Polynomial curve fitting and correction of negative values. Besides the previously mentioned linear interpolation, polynomial curve fitting based on the least squares method may encounter problems with data points resulting in negative values. First of all, the main reason for this issue is our historical determination that 1860 was a dividing point between slow and rapid growth. Therefore, we use 1860 as a breakpoint and separate interpolated the data for Ye (Ye et al., 2009, Ye and Fang, 2012) and Tian (Tian, 2005) before and after this period. Second, for time points that cannot be directly obtained through linear interpolation, cropland need to be calculated by polynomial fitting backwards (1800 and 1850 in the three provinces of Northeast China; 1900 in East of Inner Mongolia). For instance, in Northeast China, cropland area in each county in 1800 and 1850 are derived from data in 1683, 1735, and 1780 using polynomial curve fitting method. Some counties may show a decline in cropland, potentially resulting in negative values in the extrapolation results. Our solution involves using the proportion of provincial administrative level to multiply by the cropland area in 1780 for correction in the counties' cropland area in 1800 and 1850. Third, for time points that cannot be directly obtained through linear interpolation, cropland need to be calculated by polynomial fitting forwards (1900 in the three provinces of Northeast China; 1700, 1910 in East of Inner Mongolia). For instance, in Northeast China, cropland area in each county in 1900 is derived from data in 1908 and 1914 using polynomial curve fitting method. Due to rapid growth of cropland in some counties from 1908 to 1914, the extrapolation for 1900 may result in negative values. Our solution involves using the proportion of provincial administrative level to multiply by the cropland area in 1908 for correction in the counties' cropland area in 1900.

It should be noted that, considering the historical development process of Northeast China during the Qing Dynasty, war factors, and the encouraging land reclamation policies implemented by the Qing government after 1860, we determined that the cropland area in each county of Northeast China in 1900 would not significantly exceed that of 1908. During this period, in Northeast China, the total cropland area was gradually increasing and was not significantly affected by events such as the Second World War, which led to a notable decrease in cropland area in 1950 compared to 1930 and 1940. Therefore, when the extrapolated value for a county in 1900 exceeds that of 1908, the proportion of provincial administrative level is used to multiply by the cropland area in 1908 for correction in the county's cropland area in 1900.

③The determination of initial cultivation occurred between 1780 and 1908. Few counties in Northeast China where cropland was zero in 1683, 1735, and 1780, but had cropland in 1908. Our solution involves consulted contemporary county gazetteers to verify the history of land reclamation between 1800 and 1900. If local gazetteers indicate the initial cultivation occurred before 1860, this study applies the same method as described in "①Linear interpolation and determination of zero values". If the initial cultivation began after 1860, this study applies the same method as described in "②Polynomial curve fitting and correction of negative values". All the counties where this situation occurs can be found in Table S2.

Table S2. Supplementary information showing the county's determination of initial cultivation between 1780 and 1908

Counties where cropland need to be supplemented	Year	Adjacent counties for interpolation or based on the number of settlements
Nong'an County	1800, 1850	Revised according to local gazetteer of Nong'an County
Yongji County		Based on the historical origin of this region and Jilin City, revised according to the average proportional relationship between the two regions in 1908, 1914, 1931 and 1940
Huadian City		Dunhua City, Jingyu County, Fusong County
Shulan City		Jilin City, Changchun City, Jiaohe City, Wuchang City
Panshi City		Huadian City, Dongliao County, Dongfeng County, Meihekou City, Huinan County
Mulan County		Bayan County, Tonghe County, Bin County
Fuyu County		Yi'an County, Nehe City, Lindian County, Qiqihar City, Gannan County
Bei'an City		Kedong County, Wudalianchi City, Hailun City
Jiaohe City		Interpolated based on multiplying the number of

Changling County		settlements and the cropland area owned by unit settlement in this region in 1800 and 1850 respectively
Hunchun City		
Dunhua City		
Siping City		
Lishu County		
Yitong Manchu Autonomous County		
Gongzhuling City		
Shuangliao City		
Yanji City	1850	Revised according to local gazetteer of Yanji City
Tonghe County		Yilan County, Mulan County, Fangzheng County
Lanxi County		Suihua City, Zhaodong City, Harbin City, Qinggang County

10. --Results. ESSD readers are more concerned about the reliability, availability, and accuracy of data products. However, the spatiotemporal variation characteristics of cropland area are not the most important.

Response: Thank you for your helpful suggestion. We acknowledge that the current paper’s reliability, accuracy, or uncertainties assessments are not yet sufficiently comprehensive. We have made every effort to supplement the relevant assessments and uncertainty analysis as much as possible. And we deleted some description of the spatiotemporal variation characteristics of cropland area. Please see “Results section”, Line 393-426; “Discussion section”, Line 427-678.

11. --4.1 comparison. The comparison with the global dataset does not indicate the reliability of the dataset developed in this paper, because the global dataset itself has a large degree of uncertainty. The fact that the reconstruction results in this paper are very different from the global dataset does not mean that the dataset developed in this paper is reliable. Line 516-517, the following statement is not acceptable “Comparative analysis with global historical LUCC datasets indicates that the results of this study are relatively credible and more rational.”

Response: Thank you for your insightful suggestion. “Comparative analysis with global historical LUCC datasets indicates that the results of this study are relatively credible and more rational.” This statement is indeed inappropriate. We have deleted such statements in the original text and used more objective language to describe the differences between different datasets and analyze the possible reasons for the differences. Please see "Discussion section", Line 427-678.

12. -- technical corrections. Figure 5, no titles for x and y axes.

Response: Thank you for your suggestion. We revised the Fig. 4 and Fig. 7 in current manuscript.

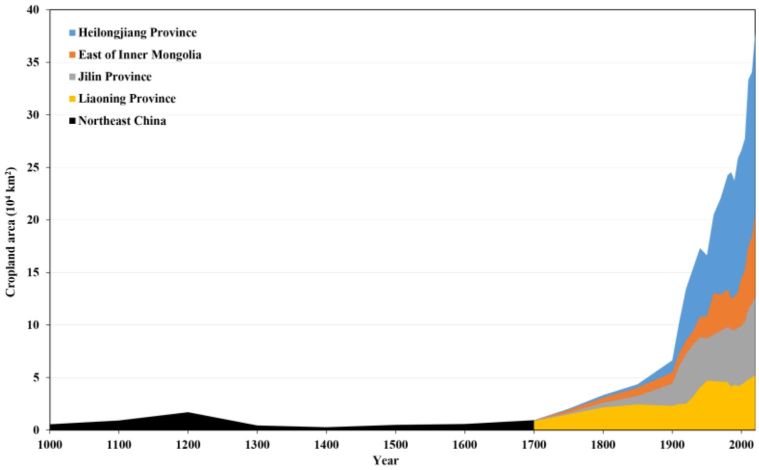

Figure 4: Changes in total cropland area in the Northeast China from 1000 to 2020.

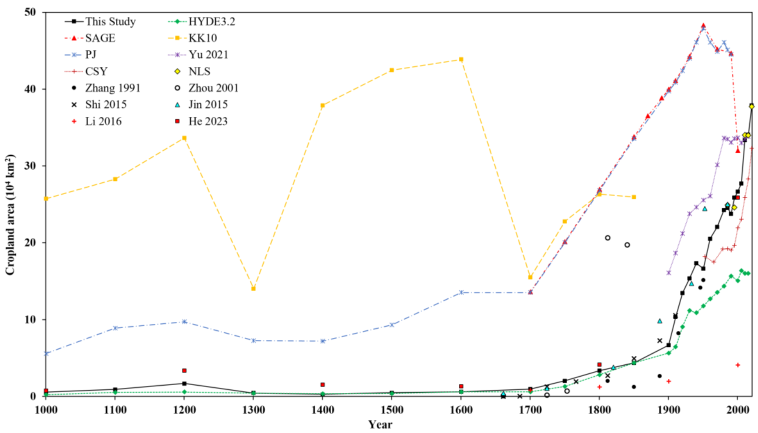

Figure 7: Comparison of total cropland area from global historical LUCC datasets, previous studies and this study in the Northeast China.

Once again, thank you very much for your constructive comments, which has greatly improved this manuscript.