https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

Open Access Earth System Science Data Discussions

Long-term Land Cover Dataset of the Mongolian Plateau Based on Multi-source Data and Rich Sample Annotations

Juanle Wang[1,2,11*], Kai Li[1,2], Mengmeng Hong[1,3], Yating Shao[1,3], Zhichen Sun[1,3], Meng Liu[1,4], Fengjiao Li[1,5], Yuhui Su[1,6], Qilin Jia[1,7], Yaping Liu[1,3], Jiazhuo Liu[1,3], Jiawei Jiang[1,3], Altansukh Ochir[8], Davaadorj Davaasuren[9], Mengqiong Xu[1,4], Yamin Sun[1,10], Yifei Sun[1,2], Shaopu Huang[1,7], Weihao Zou[1,3], Tengfei Han[1,6], Feiran Sun[1,2]

[1]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
[2]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China;
[3]College of Geoscience and Surveying Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China;
[4]School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang, Jiangsu 222005, China;
[5]College of Geomatics, Xi'an University of Science and Technology, Xi'an 710054, China;
[6]School of Civil Engineering and Geomatics, Shandong University of Technology, Zibo, Shandong 255049, China;
[7]College of Ecology and Environment, Inner Mongolia University, Hohhot, Inner Mongolia 010022, China;
[8]Environmental Engineering Laboratory, Department of Environment and Forest Engineering, School of Engineering and Technology, National University of Mongolia, Ulaanbaatar 14201, Mongolia;
[9]Department of Geography, School of Art and Sciences, National University of Mongolia, Ulaanbaatar 210646, Mongolia;
[10]School of Resources and environment, Institute of Disaster Prevention Science and Technology, Sanhe 065201, China;
[11]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China;
Correspondence to: Juanle Wang (wangjl@igsnrr.ac.cn)

Abstract. The Mongolian Plateau (MP), with its unique geographical landscape and cultural features, plays an important role in the safety of the regional ecological environment and sustainable development in North Asia. Land-cover data form the basis of MP studies; however, global-scale data products often fail to reflect regional characteristics. Despite improvements in data resolution, the level of detail remains insufficient for MP studies. This study aimed to improve the land cover classification system for MP. Based on multi-source data and extensive sample annotations, we constructed a high-quality land cover dataset for the MP and analyzed regional land cover changes. In this study, we used Landsat 5 and 8 images as primary data, which were supplemented by NASADEM, nighttime light data, ESA WorldCover, and other multi-source data for auxiliary analysis. The MP was divided into 13 major land cover categories: forest, shrub, meadow steppe, real steppe, desert steppe, wetland, water, cropland, built areas, bare area, desert, sand, and ice. For these categories, we manually labelled the samples for seven periods: 1990, 1995, 2000, 2005, 2010, 2015, and 2020. We constructed a set of image features and used a random forest model to train and predict land-cover data from 1990 to 2020. The overall accuracy of the land-cover products was 83.9%, with a kappa coefficient of 0.817. Analysis of land cover changes over the past 30 years revealed that 65.04% of land cover types in the MP remained unchanged from 1990 to 2020. The sandy area decreased annually, with a significant trend of bare areas transitioning to vegetation. The land-cover dataset produced in this study includes land-cover data and labelled samples for seven periods, exhibiting high application potential. It can be used for

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024


Open Access Earth System Science Data Discussions

research on land-cover changes and related resource and environmental issues in the MP, providing data to support ecological protection and natural resource management in the region.

1 Introduction

The Mongolian Plateau (MP) is located at the heart of the Eurasian continent. It forms a relatively independent geographical unit with the Baikal Basin in the north, the Altai Mountains in the west, the edge of the monsoon region in the east, and the Gobi Desert in the south. The surface landscapes of MP are diverse and rich in mineral resources. It serves as a sanctuary for long-term records of Earth's life evolution, such as dinosaur fossils. It is a biodiversity hotspot ranging from the Siberian taiga to Asian desert steppes. In addition, the region is characterized by significant human cultural features. The long-standing blend of nomadic culture and modern agricultural civilization, rapid urbanization of central cities conflicting with environmental carrying capacity, close coupling of regional economic development with the Trans-Siberian Railway, and prominent sectors of animal husbandry and mineral resource development, alongside regional ecological protection, are all notable characteristics. Thus, the MP holds great significance for comprehensive natural and human studies and is a hotspot for research on global resources, the environment, ecology, disasters, and sustainable development.

Land-cover data are the most fundamental data resources for researching the resources and environments of MP. It provides information on the natural state of the surface and can be used for land degradation assessments (Pisano et al., 2022), ecological environment monitoring (Ribeiro et al., 2023; dos Santos et al., 2023) and has significant implications for Earth system science, global environmental change (de Lima et al., 2023; Gimbo et al., 2024), and sustainable development research (Rabehi et al., 2020; Mpakairi et al., 2022). Numerous studies have focused on land cover, which is primarily related to the ecological environment. Using ancient literature and land cover change data, Kempf (2022) reconstructed the historical land cover and environmental data of areas from Beijing to Inner Mongolia. This study concludes that the arid regions of northern China and Mongolia experienced severe desertification in the 20th century, exacerbated by overgrazing and resource extraction, which accelerated land degradation in the area. Zhang et al. (2022) evaluated the sustainability of MP development over the past 30 years using land cover data and related indicators of sustainable development goals. Luo et al. (2024) analyzed the concentration of microplastics in lake water within different environmental regions of the Inner Mongolia Plateau, including deserts, farmlands, grasslands, forests, and meadows. They found that microplastic distribution was positively correlated with farmland and artificial surface cover. This positive correlation indicates that human activity-related land cover types may exacerbate microplastic pollution in lakes. All of these studies relied on high-quality land-cover products. However, they face issues such as low temporal resolution and discontinuity in time scales, often limiting the research to a single year or widely spaced time periods.

Numerous land cover products have been developed globally, including the European Space Agency (ESA WorldCover) (Zanaga, 2021), GlobalLand30 by the National Geomatics Center of China (Chen et al., 2015), Copernicus Global Land Cover (GLC100, GLC200) by the European Commission (Buchhorn et al., 2020) (Buchhorn et al., 2020), Global Land

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

Cover Characterization (GLCC) by the US Geological Survey (Loveland et al., 2000), and Dynamic World by Google (Brown et al., 2022). However, these products often have limitations related to data inconsistency issues and limited applicability in the MP region. For example, the classification of bare areas and grasslands is often coarse and lacks a detailed secondary classification. Additionally, these products do not cover a sufficiently long time span to reflect the significant socioeconomic and political changes in Mongolia since 1990. In recent years, climate change has exacerbated the ecological vulnerability of MP. Targeted land-cover products can better analyze issues such as desertification and land degradation in the MP. Consequently, some researchers have begun to develop regional land-cover products specifically for MP. Wang et al. (2019) developed a land-cover mapping method based on object-oriented approaches to obtain land-cover data for Mongolia in 1990 and 2010. They classified Mongolia's land cover into categories, such as bare area, cropland, forest, meadow steppe, sand, built areas, desert steppe, ice, real steppe, and water bodies, enriching the classification types for grasslands. These data were updated in 2022 and 2023, expanding temporal coverage to 1990, 2000, 2005, 2010, 2015, and 2020 (Wang et al., 2022; Wang et al., 2023). However, these land-cover data still lack a more refined and comprehensive classification system, and the study area is limited to Mongolia, making it challenging to meet the need for detailed monitoring of the ecology, desertification, and land degradation of MPs. However, previous land-cover work based on visual interpretation is labor-intensive, lacks the accumulation of sample data, and thus cannot provide effective data for subsequent land-cover updates.

A method based on sample point annotation has also been applied to global-scale datasets. The GLanCE land-cover sample set provided two million land-cover sample points worldwide from 1984 to 2020, covering 13 secondary categories (Stanimirova et al., 2023). Remelgado et al. (2020) collected over 8,000 cropland data samples from Central Asia between 2015 and 2018 and compiled samples of 40 crop types. Fritz et al. (2017) utilized crowdsourcing to obtain global land cover data from the Geo-Wiki platform for 10 land cover types from the Geo-Wiki platform. In the era of artificial intelligence and cloud computing, these sample annotation methods can continuously accumulate multi-type land-cover data, providing ample sample data support for model transfer and continuous updates of data products. However, owing to issues such as time, region, scale, and classification systems, the sample sets from these studies cannot fully meet the research needs of surface cover changes specific to MP.

To address the lack of long-term detailed land cover data for the MP, we constructed a sample database of the MP's current surface cover. Using the Google Earth Engine cloud platform, feature bands were constructed, and land cover data for the growing seasons from 1990 to 2020 in the MP were trained and generated. This dataset is expected to serve the purposes of resource investigation and utilization, land degradation and desertification monitoring, natural disaster prevention and control, ecological security assessment, and regional sustainable development in the MP.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

Open Access Earth System Science Data Discussions

2 Study area and data sources

2.1 Study area

The MP is located in the heart of the Eurasian continent and encompasses the entirety of Mongolia and China's Inner Mongolia Autonomous Region. The total area is 2.7 million km^2, with elevations ranging from approximately 80 to 4,200 m (Figure 1). The elevation of the MP generally decreases from northwest to southeast. The Altai, Khangai, and Khentii mountains in the northwest of the study area have higher terrain, with an average elevation of up to 3,000 m. The Greater Khingan Range in Inner Mongolia is divided into two parts with significant differences in elevation. The western part generally has an elevation above 1,000 m, whereas the eastern part has an elevation below 500 m. The surface cover types of MP are diverse but primarily consist of grasslands, bare areas, and forests.

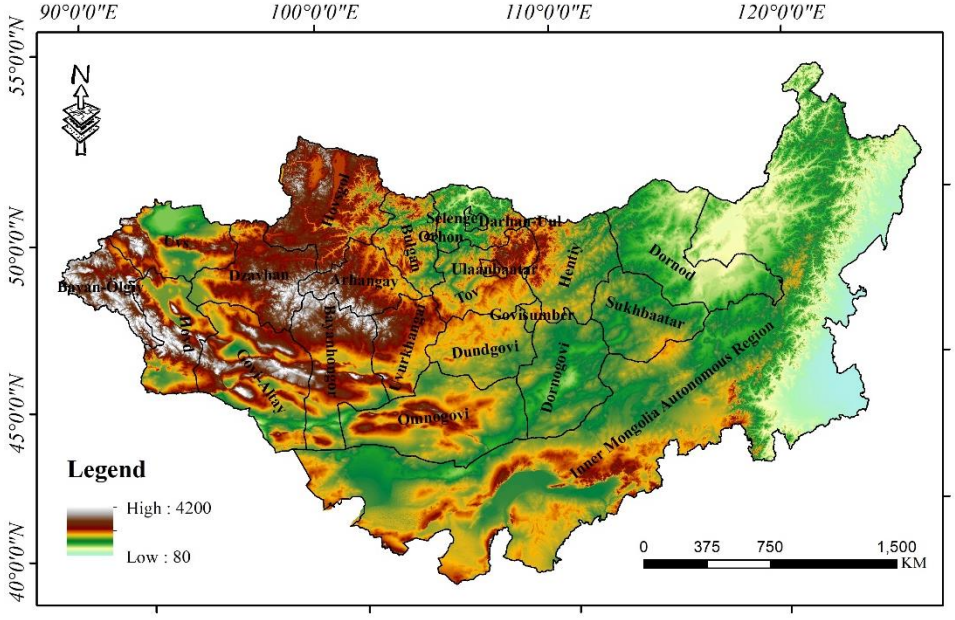

Figure 1: Study area.

2.2 Data Source

The data sources are listed in Table 1. Landsat 5 and Landsat 8 were used as base maps to classify the respective years. During the annotation of the sample points, NASDEM, ESA WorldCover, OpenLandMap Sand Content, and global wetland distribution data points were used to assist in identification. Landsat 5, Landsat 8, NASDEM, and nighttime light data were used to build feature bands, train the sample points and produce land cover data.

Table 1: **Data Sources**

Data Source	Band Name	Value	Time	Link in GEE
Landsat 5	B1, B2, B3, B4, B5, B7,	1-65,455	1984-2012	LANDSAT/LT05/C02/T1_L2

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

	QA_PIXEL			
Landsat 8	B2, B3, B4, B5, B6, B7, QA_PIXEL	1-65,455	2013-	LANDSAT/LC08/C02/T1_L2
NASADEM	elevation	80-4,200	2000	NASA/NASADEM_HGT/001
Night Light	median	/	2012-	NOAA/VIIRS/DNB/ANNUAL_V21
	avg_lights_x_pct	/	1992-2014	NOAA/DMSP-OLS/NIGHTTIME_LIGHTS
ESA WorldCover	/	10-100	2021	ESA/WorldCover/v200
OpenLandMap Sand Content	b0	1-100	1950-2018	OpenLandMap/SOL/SOL_SAND-WFRACTION_USDA-3A1A1A_M/v02
Global wetland distribution data points	/	/	/	https://rsis.ramsar.org/

3 Methods

This study aimed to further improve the land cover system of the MP based on initial study (Wang et al., 2019; Wang et al., 2022), expanding the spatial range to MP, enhancing temporal continuity of data, and supplementing it with wetland and shrub categories. The MP includes 13 major wetland areas, including Lake Uvs and its surrounding wetlands, Ayrag Nuur, and the Dalai Lake National Nature Reserve, which are crucial for regional water purification, flood control, and for providing diverse species habitats. The MP primarily has arid and semi-arid climates, where shrubs are a major land cover type. Therefore, the MP was divided into 13 categories: forest, shrub, meadow steppe, real steppe, desert steppe, wetland, water, cropland, built areas, bare area, desert, sand, and ice. The classification system is presented in Table 2.

Table 2: Classification system of Mongolian Plateau

Class	Label	Description
forest	1	Densely vegetated areas covered with large trees and other vegetation. Mainly located in northern Mongolia and northeastern Inner Mongolia.
shrub	2	Areas dominated by low-growing woody plants.
meadow steppe	3	Areas with lush grassland cover, usually located near water bodies.
real steppe	4	Grasslands with high coverage, predominantly composed of xerophytic

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

		herbaceous plants.
desert steppe	5	Transitional areas between desert or bare area and real steppe.
wetland	6	Grassland areas rich in water sources, relatively fixed in location.
water	7	Surface water bodies such as lakes, rivers, and reservoirs.
cropland	8	Agricultural areas used for crop cultivation, mainly located around built areas.
built area	9	Urban and suburban areas dominated by man-made structures such as buildings, roads, and infrastructure.
bare area	10	Areas with sparse or no vegetation, usually consisting of exposed soil, rocks, or sand.
desert	11	Arid regions with very little precipitation, characterized by sand dunes and rocky terrain, mainly located in the southwestern corner of Inner Mongolia.
sand	12	Areas primarily covered by sand, mainly located in the Gobi region of the Mongolian Plateau.
ice	13	Areas covered with snow year-round, mainly located in the high-altitude and high-latitude regions of the Mongolian Plateau.

The technical approach was divided into three parts: image mosaicking, label annotation, and land-cover data production (Figure 2). Image mosaicking involves cloud removal and mosaicking to obtain cloud-free MP images, providing high-quality imagery for the subsequent steps. The second step is label annotation, which requires the establishment of a grid boundary with equal intervals for the MP. Target categories were annotated for each grid using base images (Landsat
130 imagery) and auxiliary reference feature images or products. The third step is the generation of land-cover products. This step involves constructing feature bands by combining multi-source data and using labelled data to train sample points on the Google Earth Engine cloud platform. Finally, the trained data were applied to Landsat imagery to generate land cover data for 1990-2020.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

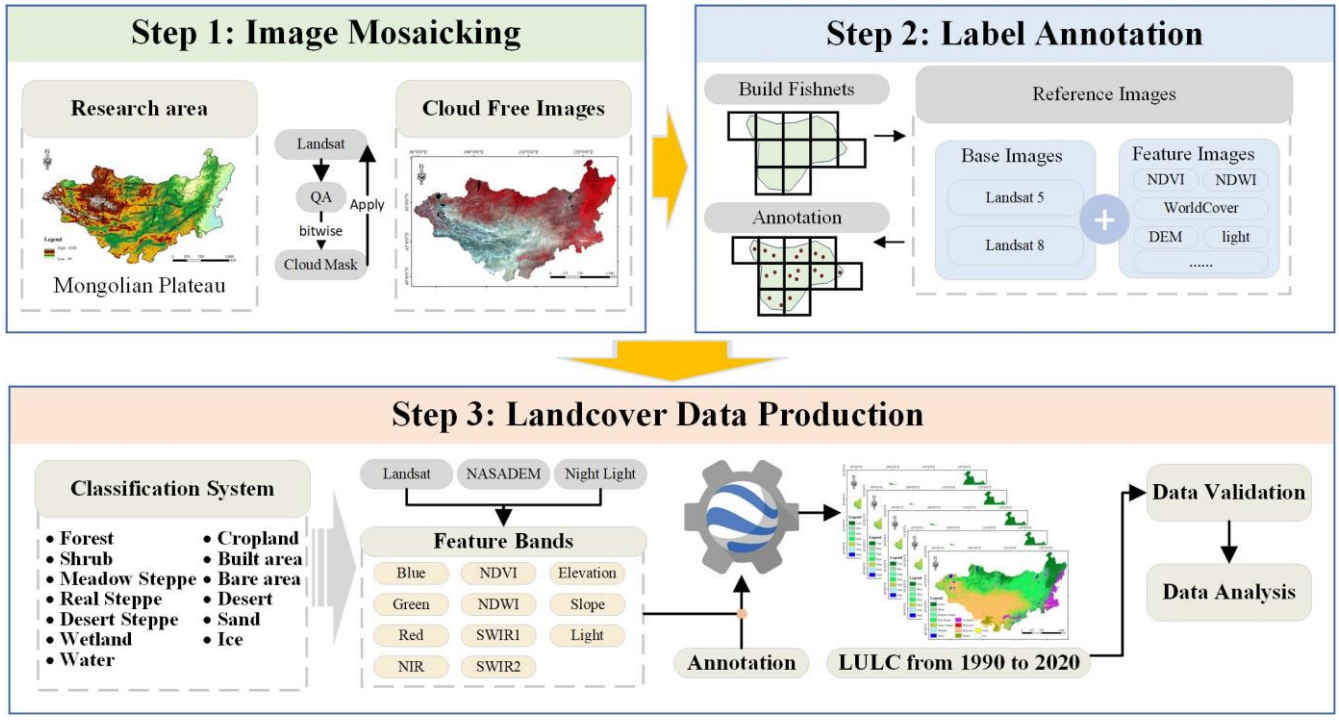

Figure 2: Overall workflow.

3.1 Image Mosaicking

Images from 1990 to 2020 were selected at five-year intervals as the base image data. Because Landsat 8 only started collecting data in 2013, data from Landsat 8 were used for 2015 and 2020, whereas data from Landsat 5 were used for 1990, 1995, 2000, 2005, and 2010. During the composting process, images from the growing season (June 1–August 31 of each year) were selected. The QA band was used to perform bitwise operations to create masks for clouds and cloud shadows, thereby removing the areas affected by interference. Each band was assigned a radiometric correction scale factor to convert the DN values into reflectance. The masked reflectance images for each year were mosaics. However, because of significant cloud interference, it is often difficult to synthesize a completely cloud-free image for the growing season each year, necessitating the use of other years for interpolation. Figure 3b shows the proportion of interpolation and image area for the current year. It can be seen that the proportion of interpolation images was the highest in 1990, but the interpolation area was also less than 6.5%. The interpolation areas for the other years were less than 2% of the overall area. Here, we used images from previous and subsequent years for interpolation, ultimately achieving a cloud-free image of the Mongolian Plateau for the growing season (Figure 3a). Figure 3a shows the false-color display results for the 2020 growing season images. Overall, the uniform color transition, absence of obvious banding, and clear visibility of vegetation cover, water bodies, and bare areas indicate that the image composition effect is good and can be used for the annotation and training of land cover data products.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

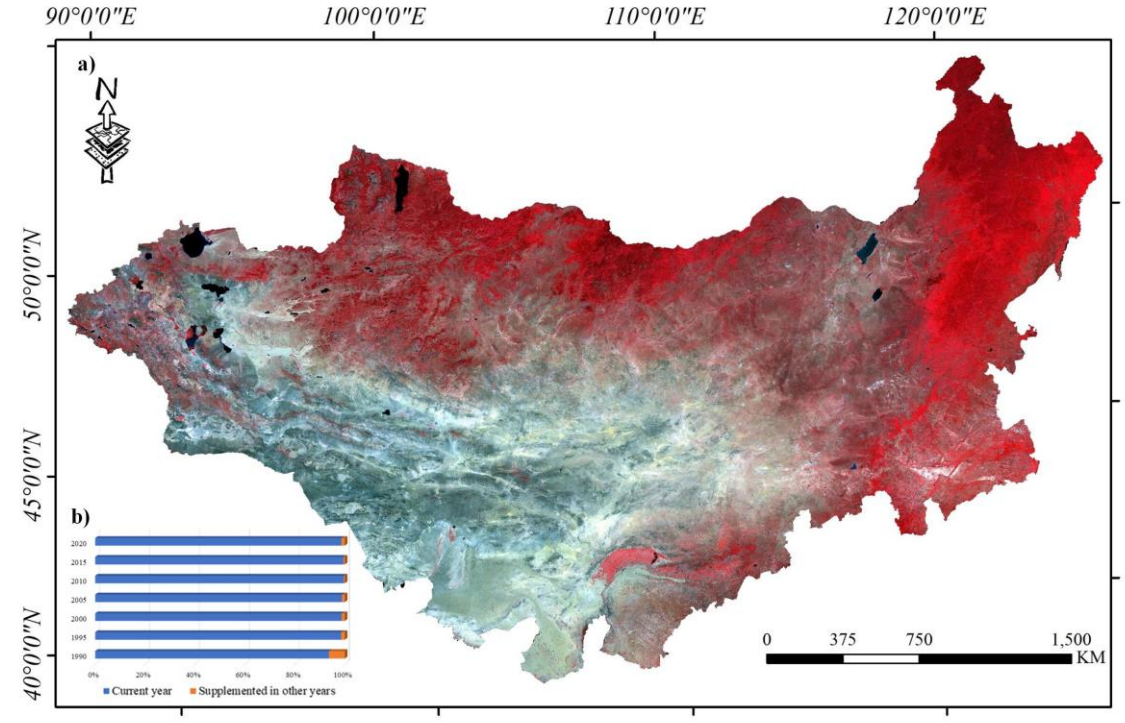

Figure 3: Mosaicked Image in 2020.

3.2 Label Annotation

A label collection team was established to collect land-cover label samples for Mongolia in 1995 and 2020, as well as land-cover label samples for Inner Mongolia at five-year intervals from 1990 to 2020. To ensure that the sample points were evenly distributed across the MP, the region was divided into 803 grids of 0.8°×0.5° (470 grids for Mongolia and 333 grids for Inner Mongolia). For each grid, visual interpretation was performed using Landsat remote sensing images from the corresponding year to annotate the land-cover type. During data collection, the screen displayed no more than two grid cells at a time, and at least 6-8 sample points were marked per grid and distributed relatively evenly. Based on these requirements, the sample collection personnel underwent uniform training. The regional samples for each period were independently collected by two personnel who served as mutual controls. A total of 43,223 samples were collected: 11,295 in 2020, 4,521 in 2015, 4,887 in 2010, 5,794 in 2005, 4,459 in 2000, 9,807 in 1995, and 2,460 in 1990 (Figure 4c). Ninety percent of the sample points were used for model training (Figure 4a), and ten percent were used for product accuracy validation (Figure 4b). The number of annotations for each category is shown in Figure 2c, with more sample points for forest, real steppe, desert steppe, cropland, and bare area and fewer sample points for snow, wetland, and shrub because of their more scattered distribution.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

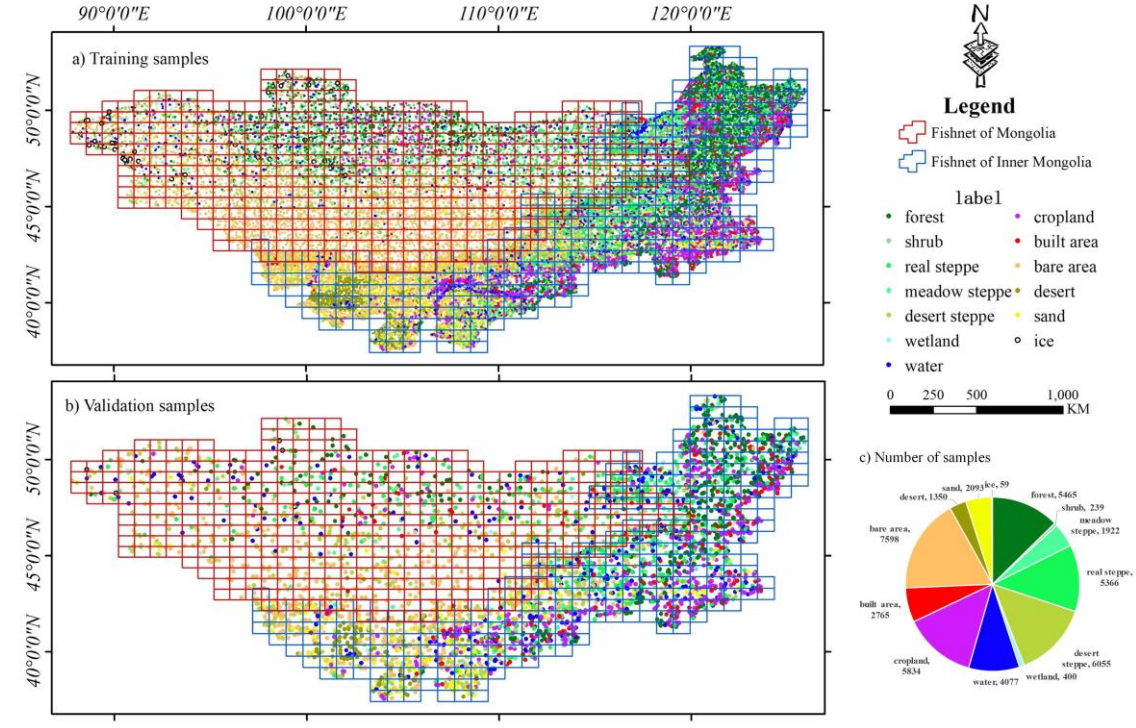

Figure 4: Distribution of labels in each category

3.3 Land cover data production

3.3.1 Construction of the Feature Dataset

To account for different land cover types, the feature dataset included blue, green, red, near-infrared, shortwave infrared 1, shortwave infrared 2, NDVI, NDWI, elevation, slope, and nighttime light data. The inclusion of blue, green, red, and near-infrared bands can effectively distinguish most land-cover types. NDVI can contribute to differentiating land cover types strongly associated with vegetation, such as forests, meadow steppes, real steppes, desert steppes, shrubs, croplands, and wetlands. The inclusion of the DEM and slope data can effectively reduce the misclassification of water bodies. Built areas often have significant spectral differences and some similarities with bare areas; therefore, the introduction of nighttime light data can aid in the classification of buildings, helping to mitigate the misclassification of built areas to some extent.

$$NDVI = \frac{NIR-Red}{NIR+Red}, \quad (1)$$

$$NDWI = \frac{Grenn-NIR}{Green+NIR} \quad (2)$$

where NIR, Red, and Green refer to the near-infrared, red, and green bands, respectively, in the remote sensing images. NDVI stands for Normalized Difference Vegetation Index, and NDWI stands for Normalized Difference Water Index.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

3.3.2 Model Training and Prediction

The Google Earth Engine (GEE) was used as the platform for data collection and model training. First, the historical sample point data were integrated, with the sample points for each year divided into training and validation points in a 9:1 ratio. The training and validation points were packaged and stored on the GEE platform. To address potential cognitive differences among the sample collection personnel and improve the issue of overfitting, images and data from the seven periods between 1990 and 2020 need to be standardized. According to Section 3.3.1, the feature dataset was constructed by extracting the 11 feature pixel values of the target year point-by-point and creating a training point set with a one-to-one correspondence between feature pixel values and label values. After comparison, the random forest algorithm that yielded the best classification results was used to train the training point set. The number of trees is set to 100. The features were combined to construct a predicted image set with 11 feature bands. A trained random forest classifier was applied to the prediction image set to produce land cover data products.

3.3.3 Accuracy Validation

The remaining 10% of the validation points were used to extract the land-cover data values for the corresponding years. The accuracy was validated using a confusion matrix. The evaluation metrics for accuracy validation included the Overall Accuracy, Precision, Recall, F1 score, and kappa coefficient. These metrics range from 0 to 1, with values closer to 1 indicating a better classification performance. The precision, Recall, and F1 score were used to evaluate the classification performance of each land-cover type, whereas mPrecision, mRecall, and wF1 were used to assess the overall classification performance.

$$Accuracy = \frac{T}{T+F} \tag{3}$$

$$Recall_i = \frac{TP_i}{TP_i + \sum FN_i} \tag{4}$$

$$Precision_i = \frac{TP_i}{TP_i + \sum FP_i} \tag{5}$$

$$F1_i = \frac{2 \bullet Precision_i \bullet Recall_i}{Precision_i + Recall_i} \tag{6}$$

$$mRecall = \frac{1}{13}\sum_{i=1}^{13} Recall_i \tag{7}$$

$$mPrecision = \frac{1}{13}\sum_{i=1}^{13} Precision_i \tag{8}$$

$$wF1 = \sum_{i=1}^{13} w_i \bullet F1_i \tag{9}$$

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

In formulas (3-9), T represents the number of pixels where the labels and predictions are consistent, and F represents the number of pixels where the labels and predictions are inconsistent. TP_i is the number of pixels correctly predicted for category i, TN_i is the number of pixels correctly predicted for all categories other than category i, FN_i is the number of pixels incorrectly predicted for category i, and FP_i is the number of pixels that belong to category i but are predicted for other categories.

4 Results and Discussion

4.1 Results

Table 3 shows the confusion matrix for all validation samples across the seven periods in the MP dataset, with a total of 4,383 validation samples. Table 4 presents the overall statistical accuracy for each category. The overall validation accuracy was 83.9%, with a kappa coefficient of 0.817. The average precision, Recall is 81.4%, and weighted F1 scores were 86.1, 81.4, and 84.0%, respectively. Overall, the accuracy of the land cover data was relatively high. The Precision metrics for ice, wetlands, water, forests, meadow steppes, and built areas exceeded 90%. The Recall metrics for ice water, real steppe, and forest were above 90%. The combined F1 scores for Precision and Recall indicated that the classification performance for shrubs and sand was poorer than that for other land-cover types, with scores below 70%. The F1 scores for forests, desert steppes, wetlands, water, croplands, built areas, bare areas, and ice exceeded 80%. Overall, the highest classification accuracies were observed for ice, water, and forest, whereas the classification accuracy for shrubs and sand was relatively low. Looking at different years, the overall accuracies for 1990, 1995, 2000, 2005, 2010, 2015, and 2020 are 80.9%, 73.4%, 78.2%, 85.5%, 67.6%, 94.5%, and 97.7%, respectively. The Kappa coefficients were 0.78, 0.70, 0.75, 0.83, 0.62, 0.94, and 0.97, respectively. The highest classification accuracies were observed for 2015 and 2020.

Table 3 Confusion matrix

	forest	**shrub**	**meadow steppe**	**real steppe**	**desert steppe**	**wetland**	**water**	**cropland**	**built area**	**bare area**	**desert**	**sand**	**ice**
forest	475		1	44			1	6					
shrub		18		2	1					1		2	
meadow steppe	3	4	137	25	3			20	2	5			
real steppe	18		1	543	7			13	2	9			
desert steppe	7	1	1	82	464			11	1	52	6	5	
wetland			4	2		29	4	3		1			

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

water	1		1	1	1	380	5	2	7				
cropland	2	3	2	55	7	2	500	13	6				
built area	1	1		5	6		20	231	7				
bare area	1	1		11	11		1		636	2	67		
desert		1		5	3				32	96	5		
sand		1		3	8		1		50	3	164		
ice													6

Table 4 Accuracy indicators

	forest	**shrub**	**meadow steppe**	**real steppe**	**desert steppe**	**wetland**	**water**	**cropland**	**built area**	**bare area**	**desert**	**sand**	**ice**
Precision	0.935	0.600	0.932	0.698	0.908	1.000	0.982	0.862	0.920	0.789	0.897	0.675	1
Recall	0.901	0.750	0.688	0.916	0.737	0.674	0.955	0.847	0.852	0.871	0.676	0.713	1
F1 score	0.918	0.667	0.792	0.792	0.813	0.806	0.968	0.855	0.885	0.828	0.771	0.693	1

Overall Accuracy	83.9%	**Kappa**	0.817	**mPrecision**	86.1%	**mRecall**	81.4%	**wF1**	84.0%	

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

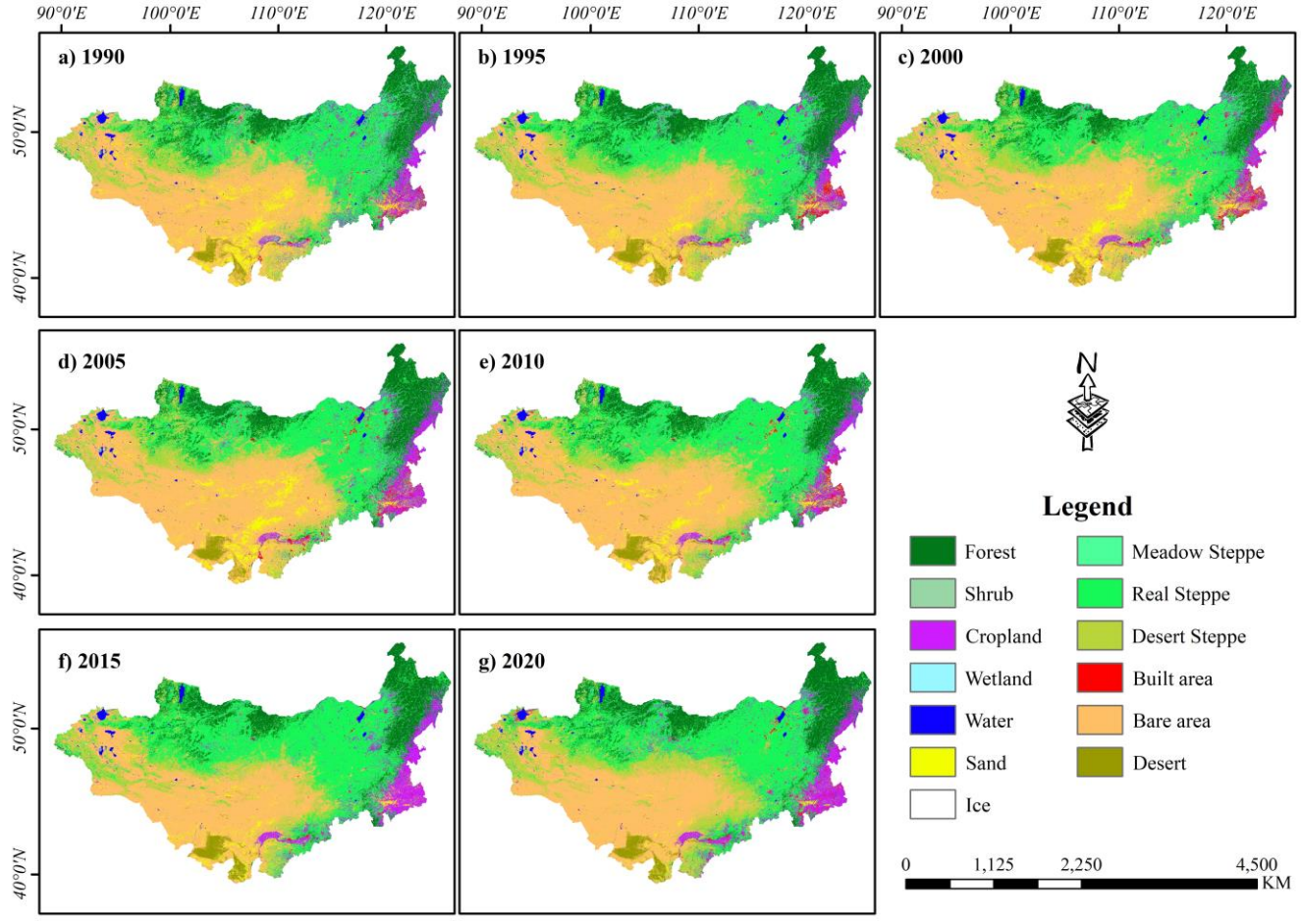

Figure 5: Land cover map of the Mongolian Plateau in 1990, 1995, 2000,2005,2010,2015, and 2020.

4.2 Analysis of Overall Trends in Land Cover Change on the Mongolian Plateau

The area statistics for each land cover type were calculated for each year (Figure 5), and the land cover transitions every five years were analyzed (Figure 6). The bare area was the most dominant land cover type in the MP, accounting for over 30%, followed by the real steppe, desert steppe, and forest (accounting for 29%, 17%, and 11%, respectively, in 2020). From 1990 to 2020, 65.04% of the land cover remained unchanged, with bare areas and real steppes constituting the majority at 37% and 30% of this stable portion, respectively. Forest and desert steppe follow-ups accounted for 13% and 10%, respectively. Over the past 30 years, real steppes, sand, wetlands, and ice areas have significantly decreased. The reduction in sand was particularly notable, with much of the sand transitioning to bare areas. Calculations show that the sand area decreased from 145,000 km^2 in 1990 to 24,000 km^2 in 2020, an 83% reduction compared with 1990. The wetland, ice, and desert steppe areas decreased by 78%, 65%, and 60%, respectively, compared with 1990. Although the bare area decreased by 13%, it

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

Open Access Earth System Science Data Discussions

experienced the largest absolute reduction of 125,000 km^2. The areas that increased the most included the desert steppe by 153,000 km^2, real steppe by 106,000 km^2, and cropland by 48,000 km^2.

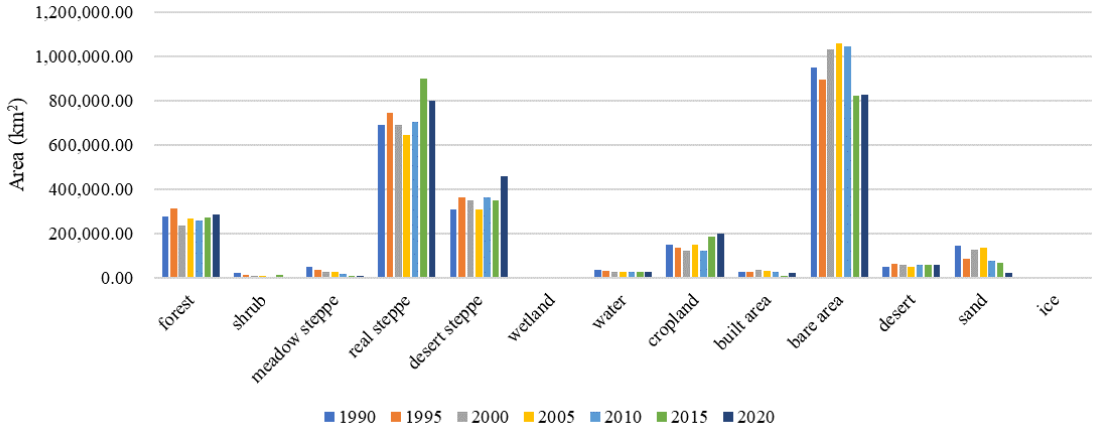

Figure 6: Annual area changes of land cover.

From the land cover transitions every five years (Figure 6), it can be seen that forested areas had the highest transition between 1995 and 2000, mostly converting to real steppe. The real steppe consistently transitions to forest each year and alternates between conversion to and from croplands. The trend in the bare areas was opposite to that of vegetation (forest, real steppe, cropland), with a significant downward trend in the past ten years due to the gradual transition of bare areas to vegetation. The total area of the desert steppe has not changed significantly, but in recent years, most desert steppes have transitioned from a bare area. Between 2010 and 2015, a large portion of the desert steppe transitioned to real steppe, reflecting an increasing vegetation cover on the MP.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

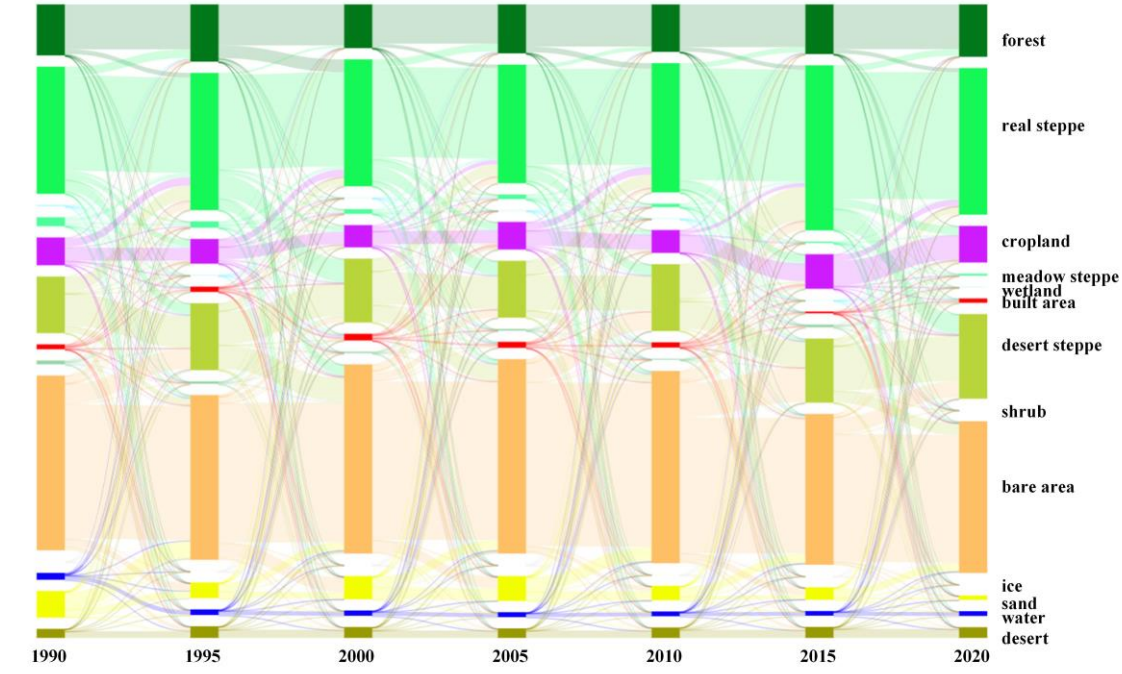

Figure 7: Land cover transfer every 5 years.

4.3 Data Description and Application Potential

This dataset is a land cover data product for the MP, spanning 1990–2020, and includes seven periods: 1990, 1995, 2000, 2005, 2010, 2015, and 2020. This describes the changes in surface cover during the growing season of the MP. The values ranged from 1 to 13, representing forest, shrub, meadow steppe, real steppe, desert steppe, wetland, water, cropland, built area, bare area, desert, sand, and ice, respectively. Additionally, the dataset provides label samples for MP for 1990, 1995, 2000, 2005, 2010, 2015, and 2020. This dataset can be used for research on land cover changes, related resources, and environmental issues in the MP. Considering the characteristics of MP, a classification system suitable for this region was developed. Based on the grassland classification, secondary classifications such as meadow steppe, real steppe, and desert steppe have been derived. This can better serve to monitor grass yield changes, regulate animal husbandry development, and evaluate ecological service functions in the MP.

5 Conclusion

Through the comprehensive use of multi-source remote sensing data in this study, we obtained 43,223 labelled samples of 13 land-cover types for seven years from 1990 to 2020 in the MP using manual annotation. Land cover data products for the MP from 1990 to 2020 were generated and analyzed by integrating Landsat 5, Landsat 8, NASADEM, nighttime light data, and other sources. The main conclusions of this study are as follows.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

This study enriches the land-cover classification system for MP based on previous studies. Considering the characteristics of the study area, a classification system for 13 main land-cover types was established: forest, shrub, meadow steppe, real steppe, desert steppe, wetland, water, cropland, built area, bare area, desert, sand, and ice. The overall classification accuracy was 83.9%, with a kappa coefficient of 0.817. The classification results were satisfactory, especially for categories such as forests, grasslands, wetlands, and water bodies. From 1990 to 2020, the main land-cover types were bare areas, real steppes, desert steppes, and forests. In this study, we found that the area of bare land decreased annually, whereas the areas of real steppe, desert steppe, and cropland increased. In particular, the sand area decreased significantly from 145,000 km^2 in 1990 to 24,000 km^2 in 2020, representing a reduction of 83%. In terms of land cover transitions, in recent years there has been a notable trend of bare areas transitioning to vegetation (forest, grassland, and cropland), reflecting a gradual improvement in vegetation cover in the MP. These changes are important for improving the regional ecological environment. The land-cover dataset generated in this study provides valuable foundational data for researching land-cover changes and related resources and environmental issues in the MP. Secondary grassland classifications in the dataset can better serve the needs of natural resource investigation and utilization, land degradation and desertification monitoring, natural disaster prevention and control, ecological security assessment, and regional sustainable development research and applications.

6 Data Accessibility

The MP_LULC land cover product, sample set, and corresponding documentation generated in this study are accessible via the Science Data Bank at 10.57760/sciencedb.07381 or https://www.scidb.cn/en/s/3AfYjm (Wang, et al., 2024). The product is stored in GeoTIFF format and includes ". ovr" pyramid files necessary for opening in ArcGIS," dbf: Statistical files for each category, among other relevant files. The data set was packaged as a. zip file. In the raster data, labels 1-13 correspond to forest, shrub, meadow steppe, real steppe, desert steppe, wetland, water, cropland, built area, bare area, desert, sand, and ice, respectively.

To facilitate user access and computation, this dataset is also available for online access via the Google Earth Engine (projects/ee-carylee/assets/MP_ dataset/land cover). The data were stored in the ImageCollection format, containing images for seven periods. Users can filter each year to obtain the dataset for each specific year.

Acknowledgment

This research was funded by the National Key R&D Program of China (grant number 2022YFE0119200), Science & Technology Fundamental Resources Investigation Program of China (grant number 2022FY101902), Key Project of Innovation LREIS (grant number KPI006), Key R&D and achievement transformation program of Inner Mongolia Autonomous Region (grant number 2023KJHZ0027), Construction Project of the China Knowledge Center for Engineering

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

Sciences and Technology (grant number CKCEST-2023-1-5), and Mongolian Foundation for Science and Technology (grant number NSFC_2022/01, CHN2022/276).

Competing interests

The authors declare that they have no conflict of interest.

Author contribution

Juanle Wang designed the experiments and prepared the manuscript. Kai Li developed the model code, performed the land cover production, and prepared the manuscript. Kai Li, Mengmeng Hong, Yating Shao, Zhichen Sun, Meng Liu, Fengjiao Li, Yuhui Su, Qilin Jia, Yaping Liu, Jiazhuo Liu, Jiawei Jiang, Altansukh Ochir, Davaadorj Davaasuren, Mengqiong Xu, Yamin Sun, Yifei Sun, Shaopu Huang, Weihao Zou, Tengfei Han, and Feiran Sun labelled the samples of land cover type in the Mongolian Plateau.

References

BROWN, C. F., BRUMBY, S. P., GUZDER-WILLIAMS, B., BIRCH, T., HYDE, S. B., MAZZARIELLO, J., CZERWINSKI, W., PASQUARELLA, V. J., HAERTEL, R., ILYUSHCHENKO, S., SCHWEHR, K., WEISSE, M., STOLLE, F., HANSON, C., GUINAN, O., MOORE, R. & TAIT, A. M. 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Scientific Data, 9.

BUCHHORN, M., LESIV, M., TSENDBAZAR, N.-E., HEROLD, M., BERTELS, L. & SMETS, B. 2020. Copernicus Global Land Cover Layers—Collection 2. Remote Sensing, 12, 1044.

CHEN, J., CHEN, J., LIAO, A. P., CAO, X., CHEN, L. J., CHEN, X. H., HE, C. Y., HAN, G., PENG, S., LU, M., ZHANG, W. W., TONG, X. H. & MILLS, J. 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. Isprs Journal of Photogrammetry and Remote Sensing, 103, 7-27.

DE LIMA, G. N., FONSECA-SALAZAR, M. A. & CAMPO, J. 2023. Urban growth and loss of green spaces in the metropolitan areas of Sao Paulo and Mexico City: effects of land-cover changes on climate and water flow regulation. Urban Ecosystems, 26, 1739-1752.

DOS SANTOS, M. P., NETO, B. D. S., CARDOSO, A. C. P., DOS SANTOS, I., COELHO, B. S., LEITE, S. A., FERNANDES, D. R. R., CARVALHO, G. A. & CASTELLANI, M. A. 2023. Natural parasitism of the coffee leaf miner: climate factors, insecticide, and landscape affecting parasitoid diversity and their ecosystem services in coffee agroecosystems. Bulletin of Entomological Research.

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

FRITZ, S., SEE, L., PERGER, C., MCCALLUM, I., SCHILL, C., SCHEPASCHENKO, D., DUERAUER, M., KARNER, M., DRESEL, C., LASO-BAYAS, J. C., LESIV, M., MOORTHY, I., SALK, C. F., DANYLO, O., STURN, T., ALBRECHT, F., YOU, L., KRAXNER, F. & OBERSTEINER, M. 2017. A global dataset of crowdsourced land cover and land use reference data. Scientific Data, 4.

GIMBO, R., MUKUNDI, J. B., MBURU, D. M. & BALIRWA, J. S. 2024. Dynamics of land use land cover change in an East African inland valley wetland. African Journal of Ecology, 62.

KEMPF, M. 2022. Documentary Evidence of 17th Century Landcover and Climate Change in Northern China and Mongolia Compared to Modern Spectral Greening Trends. Land, 11.

LOVELAND, T. R., REED, B. C., BROWN, J. F., OHLEN, D. O., ZHU, Z., YANG, L. & MERCHANT, J. W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. International Journal of Remote Sensing, 21, 1303-1330.

LUO, S., WU, H. N., XU, J. F., WANG, X. J., HE, X. D. & LI, T. 2024. Effects of lakeshore landcover types and environmental factors on microplastic distribution in lakes on the Inner Mongolia Plateau, China. Journal of Hazardous Materials, 465.

MPAKAIRI, K. S., DUBE, T., DONDOFEMA, F. & DALU, T. 2022. Spatio-temporal variation of vegetation heterogeneity in groundwater dependent ecosystems within arid environments. Ecological Informatics, 69.

PISANO, L., ZUMPANO, V., PEPE, M., LISO, I. S. & PARISE, M. 2022. Assessing Karst Landscape Degradation: A Case Study in Southern Italy. Land, 11.

RABEHI, W., BENTEKHICI, N., BOUHLALA, M. A., BENHARRATS, F., ZEGRAR, A., RAHLI, H. S., KAROUI, M. S. & BENHAMOUDA, F. 2020. Monitoring and Estimation of the Sustainable Development Goal -Fifteen- by Remote Sensing Tools, Assessment of Change in Land Cover (Sub-Indicator 15.3.1), Case of Algeria. 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2garss), 239-241.

REMELGADO, R., ZAITOV, S., KENJABAEV, S., STULINA, G., SULTANOV, M., IBRAKHIMOV, M., AKHMEDOV, M., DUKHOVNY, V. & CONRAD, C. 2020. A crop type dataset for consistent land cover classification in Central Asia. Scientific Data, 7.

RIBEIRO, I., DOMINGOS, T., MCCRACKEN, D. & PROENçA, V. 2023. The use of domestic herbivores for ecosystem management in Mediterranean landscapes. Global Ecology and Conservation, 46.

STANIMIROVA, R., TARRIO, K., TURLEJ, K., MCAVOY, K., STONEBROOK, S., HU, K. T., ARéVALO, P., BULLOCK, E. L., ZHANG, Y. T., WOODCOCK, C. E., OLOFSSON, P., ZHU, Z., BARBER, C. P., SOUZA, C. J. R., CHEN, S. J., WANG, J. A., MENSAH, F., CALDERóN-LOOR, M., HADJIKAKOU, M., BRYAN, B. A., GRAESSER, J., BEYENE, D. L., MUTASHA, B., SIAME, S., SIAMPALE, A. & FRIEDL, M. A. 2023. A global land cover training dataset from 1984 to 2020. Scientific Data, 10.

WANG, J. L., LI, K., HONG, M. M., SHAO, Y. T., SUN, Z. C., LIU, M., L, F. J., SU, Y. H., JIA, Q. L., LIU, Y. P., LIU, J. Z., JIANG, J. W., OCHIR, A., DAVAASUREN, D., XU, M. Q., SUN, Y. M., SUN, Y. F., HUANG, S. P., ZOU, W. H.,

https://doi.org/10.5194/essd-2024-237
Preprint. Discussion started: 10 July 2024

HAN, T. F., SUN, F. R. 2024. Long-term Land Cover Dataset of the Mongolian Plateau Based on Multi-source Data and Rich Sample Annotations (V2). Science Data Bank. https://doi.org/10.57760/sciencedb.07381

WANG, J., WEI, H., CHENG, K. & AL., E. 2022. Updatable dataset revealing decade changes in land cover types in Mongolia. Geoscience Data Journal, 9, 341-354.

WANG, J. L., CHENG, K., LIU, Q., ZHU, J. X., OCHIR, A., DAVAASUREN, D., LI, G., WEI, H. S., CHONOKHUU, S., NAMSRAI, O. & BAT-ERDENE, A. 2019. Land cover patterns in Mongolia and their spatiotemporal changes from 1990 to 2010. Arabian Journal of Geosciences, 12.

WANG, J. L., XU, S. X., YANG, F., LI, K. & SHAO, Y. T. 2023. A dataset of land cover classifications with a spatial resolution of 30m in Mongolia in 2005 and 2015. China Scientific Data, 8.

ZANAGA, D. V. D. K., R.; DE KEERSMAECKER, W.; SOUVERIJNS, N.; BROCKMANN, C.; QUAST, R.; WEVERS, J.; GROSU, A.; PACCINI, A.; VERGNAUD, S.; CARTUS, O.; SANTORO, M.; FRITZ, S.; GEORGIEVA, I.; LESIV, M.; CARTER, S.; HEROLD, M.; LI, LINLIN; TSENDBAZAR, N.E.; RAMOINO, F.; ARINO, O. 2021. ESA WorldCover 10 m 2020 v100.

ZHANG, Y., WANG, J. & AL., E. 2022. Land Cover Change Analysis to Assess Sustainability of Development in the Mongolian Plateau over 30 Years. Sustainability, 14, 6129.