

Dataset/atlas name (equivalent to resource name in metadata)**Data Documentation**

I.	Dataset/atlas content features.....	2
i.	Abstract.....	2
ii.	Elements (content fields).....	2
iii.	Temporal cover	2
iv.	Spatial cover	2
II.	Subject/industry scope of dataset/atlas.....	3
i.	Subject scope.....	3
ii.	Industry scope	3
iii.	Other classifications (optional)	3
III.	Accuracy of dataset/atlas	3
i.	Time frequency	3
ii.	Spatial reference, accuracy, and granularity.....	3
IV.	Dataset/atlas storage management.....	4
i.	Data quantity	4
ii.	Type format	4
iii.	Update management	4
V.	Quality control of the dataset/atlas	4
i.	Production mode.....	4
ii.	Data sources (condition selection).....	4
iii.	Methods of the data acquisition and processing (condition selection).....	5
VI.	Sharing and usage method of the dataset/atlas	5
i.	Sharing methods and restrictions	5
ii.	Contact information of the sharing service (condition selection).....	5
iii.	Conditions and methods of usage	6
VII.	Intellectual property rights of the dataset/atlas	6
i.	Property rights (optional)	6
ii.	Reference method of the dataset/atlas	6
iii.	Usage contacts of the datasets/atlas	6
VIII.	Others (optional).....	7

I. Dataset/atlas content features

i. Abstract

This data set chooses 200km buffer areas along both sides of the China-Mongolia Railway (Mongolia section) as the study area, and introduces the P-BSHADE method, based on the normalized differential vegetation index NDVI and a small amount of ground-measured sample data, to simulate evenly distributed samples. Conduct an interpolation experiment. Based on the measured sample points and simulated sample points, combined with the three vegetation indices of NDVI, enhanced vegetation index, and net photosynthesis, the optimal model for estimating grass yield was established, and the annual grass yield data of the study area from 2000 to 2019 was retrieved and temporal and spatial pattern change characteristics was analyzed.

ii. Elements (content fields)

(The list describes the feature items (content fields) contained in the dataset/atlas. The code dictionary table should be given if the element is a code, and the meaning of the code should be indicated. Each data file (layer) should be described accordingly if the datasets/atlas have multiple data files (layers).)

Table 1 Description of data element content

Data name	Item (field)	Field name in Chinese	Field measure unit	Field code description	Remarks
NDVI-INV	Value	Annual grass production (per yield)	Kg/hm ²		

iii. Temporal cover

(Temporal cover means the time information of the dataset/atlas content that can be recorded as a time point or a time period (start time and end time).)

2000-2019

iv. Spatial cover

(The spatial information of the dataset/atlas includes the location description information or the latitude and longitude coordinates (latitude and longitude data should follow the format of degree, minute and second). It can be a single location or a spatial range (latitude and longitude coordinates in the upper left and lower right corners).

The 200km buffer zone along the east and west sides of the China-Mongolia Railway (Mongolia

section) is selected as the study area, located between 101°26'~113°51'E, 42°38' ~50°30'N, and Sukh Bator Province, Eastern Gobi Province, Central Gobi Province, Gobi Sumbel Province, Kent Province, Central Province, Ulaanbaatar City, Orkhon Province, Burgan Province, Houhangai Province, Selenga Province, Thirteen provinces and cities including Dar Khan Ula and Kusugul provinces.

II. Subject/industry scope of dataset/atlas

i.Subject scope

(The subject classification of the dataset/atlas is based on the “Subject Classification and Code (GB/T 13745-2008)”. The general requirements of a primary classification or a number of auxiliary classifications should reach the second level of classification, and may also reach the third level of classification.)

Earth Science

ii.Industry scope

(The industry classification of the dataset/atlas is based on the “Industry Classification of National Economy and Code Table (GB/T 4754-2002)”. The general requirements of a primary classification or a number of auxiliary classifications should reach the second level of classification, and may also reach the third level of classification.)

Disaster prevention and mitigation, ecological environment, land resources, regional economy, social development and other industries

iii.Other classifications (optional)

(Other categories can be applied, but should reflect the dataset/atlas characteristics.)

III. Accuracy of dataset/atlas

i.Time frequency

(Time frequency is the representation content of datasets/atlas' time frequency, such as multi-year average, average, monthly, daily, yearly, month by month, day or hour.)

1 year

ii.Spatial reference, accuracy, and granularity

(This part is the spatial reference, accuracy, and granularity of datasets/atlas. The spatial reference includes coordinate system, projection mode, elevation system, etc. Spatial accuracy means the vector data scale or raster data resolution, etc. Spatial granularity is in accordance with the continent, the state, province, county, and other divisions.)

Spatial reference: GCS_WGS_1984

Accuracy: 1 time

Spatial resolution: 250m×250m

Granularity: Site

IV. Dataset/atlas storage management

i.Data quantity

(Data quantity means dataset/atlas' data volume (in MB) or record number (for monitoring data or table data).)

6225.92MB

ii.Type format

(Type format means dataset/atlas storage medium (paper and electronic media, including CD, hard disk, server, etc.), structure type (text, table, vector, raster, etc.), and its specific format and version.)

The data set is stored on the hard disk, and the data structure type is raster data.

iii.Update management

(Dataset/atlas update plan: the data provider should indicate whether there is a data update plan, and if necessary, specify the update frequency and the person responsible for it. The information of the person responsible for updating management includes institute name, name of the person responsible, contact information (mailing address, telephone, and e-mail), and other relevant information.))

Update from time to time

V. Quality control of the dataset/atlas

i.Production mode

(The production mode of the dataset/atlas mainly includes raw data (field monitoring, investigation survey, test analysis, experimental test, acquisition, and similar data) and processing data (electronic, digital, integrated conversion, standardized processing, computational simulation, and similar data).)

Firstly, based on the interpolation of the long-term NDVI and the ground-measured sample point data, uniformly distributed simulated samples are obtained, and then the data set is obtained by combining the measured samples, the simulated samples and the vegetation index.

ii.Data sources (condition selection)

(The source, accuracy of the raw data, and scope of application for the raw data collected, purchased, exchanged, and shared and the data generated by processing must be explained. You may leave this blank and indicate "self-built" when the dataset is the first-hand observation data.)

Remote sensing data: MODIS satellite remote sensing data from the official website of NASA.

Measured data: The author's research team and Mongolian National University scholars jointly conducted field surveys to obtain sample data.

Auxiliary data: including Mongolian administrative boundary vector data, Mongolian traffic data and grassland type data, among which Mongolian administrative boundary vector data and traffic data are provided by the Chinese Academy of Sciences Data Cloud-Resources Innovation Platform, and the grassland type data is provided by Mongolian National University Provide assistance.

iii. Methods of the data acquisition and processing (condition selection)

(For the data of collection, monitoring, testing, processing, must explain the collection and the processing methods. The specifications include the equipment, standards, and methods for data acquisition, monitoring, testing, and usage and methods and models (the model origin and parameters) or software of processing, calculation method, and the methods (standards, etc.) of error control and processing. The data provider may leave this blank and just indicate “raw data, without processing” when the dataset is the original acquisition without any processing.)

The measured data on the ground is 10m×10m in size for each plot. Three plots of 0.5m×0.5m are randomly selected in the plot, and special scissors are used to harvest the green plants above the ground. Weigh the three plots. The average value of the fresh weight is taken as the grass yield of the sample point. A camera was used to take pictures to record the land cover types around the sample plots, and a handheld GPS was used to determine the latitude and longitude coordinates and elevation of the sample sites. Electronically record the measured data on the ground, and use Microsoft Excel software to organize the recorded data in a unified format, including the longitude and latitude coordinates of the sample point, the fresh weight of the grass produced, and the elevation. The standardized measured data is imported into the ArcGIS software platform, and after a series of processing such as coordinate and format conversion, it is exported as a Shapefile format data file to generate vector sample point data.

The MODIS data is formatted, projected and spliced through the MRT tool developed by the National Aeronautics and Space Administration. The ArcGIS software platform is used for image cropping and maximum synthesis. The vegetation index images from 2000 to 2019 are extracted year by year based on the latitude and longitude coordinates of the measured sample points. Meta value.

Introduce the P-BSHADE method, perform spatial interpolation based on the normalized differential vegetation index NDVI and the data of the ground measured sample points, obtain uniformly distributed simulated sample points, and establish the optimal model for grass yield estimation, and obtain the research area by inversion 2000-2019 Annual grass production data.

VI. Sharing and usage method of the dataset/atlas**i. Sharing methods and restrictions**

(The sharing method can be one of three possible types: full and open sharing, agreement sharing, and temporarily unshared. Full and open sharing is when the user can freely access the dataset/atlas. Agreement sharing means obtaining a dataset/atlas after signing a shared agreement with an application for the responsible party or a delegated regulatory agency. Temporarily unshared means that the data cannot be accessed at present or only be shared after the approval of the management department. Agreement sharing needs to explain the specific requirements, while temporarily unshared needs to explain the reasons and the limit of the open time.)

Fully open sharing

ii. Contact information of the sharing service (condition selection)

(Full and open sharing need to provide an online link address or a service contact. Protocol sharing needs

to provide the time and contact information of the service. The data provider may leave this blank and just indicate “service provided by the collection management institute” if the dataset/atlas is only served by the collection management institution.)

Online link address:

The service contact information is as follows:

Name: Yanjie Wang

Mailing address: No. 11, Datun Road, Chaoyang District, Beijing

Zip code: 100101

E-mail: wangyanjie@lreis.ac.cn

iii. Conditions and methods of usage

(The environmental conditions when to use the datasets/atlas should be provided, including the necessary software tools, hardware requirements, and operation of the steps, methods, or precautions.)

Use ArcGIS, ENVI and other software to open.

VII. Intellectual property rights of the dataset/atlas

i. Property rights (optional)

(The dataset/atlas property ownership information includes the property owner contact information.)

The property rights of the "Dataset of Grass Production Volume of China-Mongolia Railway (Mongolia Section) from 2000 to 2019" are owned by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

ii. Reference method of the dataset/atlas

(Dataset/atlas usage statement and citation: (i) usage statement refers to identify the dataset/atlas used in the user's research results text, while (ii) citation refers to identify the dataset/atlas used in the references of the research results, or list must be referenced published papers or monographs closely related to the dataset/atlas. It can be in either Chinese or English form or only in Chinese.)

(The usage statement can follow the following method: <Dataset/atlas name> from Disaster Risk Reduction Knowledge Service System of International Knowledge Center for Engineering Sciences and Technology (IKCEST) “<Project name (number)>”; reference method is recommended to take the following method: <Dataset/atlas property owner name> (can be multiple, sorted by contribution size, separated by “;”). <Organization name> (can be multiple, sorted by contribution size, separated by “;”). <Dataset/atlas name>. <Generation time>. When actually written, “< >” is replaced by a specific value.)

iii. Usage contacts of the datasets/atlas

(The contact persons or agencies, who curate the data and provide a data sharing service, should be listed, including their name, address, postcode, telephone, and e-mail).

Contact person

Name: Yanjie Wang

Mailing address: No. 11, Datun Road, Chaoyang District, Beijing

Zip code: 100101

E-mail: wangyanjie@lreis.ac.cn

VIII. Others (optional)

In addition to the above, other information must also be explained.

Data documentation author information			
Data documentation author		Update time	
Organization			
Contact information			
Address		Postcode	
Telephone		E-mail	