Assessment of land resource potential in Mongolia using remote sensing data and machine learning techniques

（リモートセンシングデータと機械学習を用いたモンゴルの土地資源の潜在力の評価）

OTGONBAYAR Munkhdulam

Doctor of Geography
Graduate School of Environmental Studies, Nagoya University

2023

Summary

This research addresses land resource potential assessment consisting of land suitability, land resources, and land potential. The main objective of this study was to assess and model land resource potential throughout Mongolia using state-of-the-art machine learning techniques and remote sensing data. The investigation is organized into seven chapters.

Chapter 1 provides the definitions, concepts, principles, historical development, and technological achievements of land assessment. Furthermore, the main issues of land assessment and the necessity for assessing land resource potential in Mongolia are described.

Chapter 2 investigates the land suitability assessment for the cropland in Mongolia. The main objective is to develop methods, tools, and criteria for detecting new crop areas with enough capacity for cultivation across the entirety of Mongolia. For the analysis, 9 constraints, and 17 multi-criteria factors, the Multi-Criteria Decision Making (MCDM) method and Geographic Information System (GIS)-based Analytical Hierarchy Processes (AHP) were used. The integrated assessment of constraint and multi-criteria factor analyses showed that 10.1% of the study area is highly suitable, 14.0% suitable, 15.5% moderately suitable, 16.3% unsuitable, and 12.9% highly unsuitable for cropland, with 31.2% as the constraint area. Within the framework of this research, evaluation methodology and criteria for assessing the suitability of agricultural cropland in Mongolia were developed.

Chapter 3 investigates pasture biomass, which is a component of land resources. The rational use of pastures as a source of feed is a vital issue for livestock pastoralism in Mongolia. The main objective of this study is to develop a robust methodology to estimate pasture biomass. Two regression models were compared and adopted for this study: Partial Least Squares (PLS) and Random Forest (RF). Both methods were trained to predict pasture biomass using a total of 17 spectral indices derived from Landsat 8 imagery as predictor variables. For training, reference biomass data from a field survey of 553 sites were available. This study confirms the high potential of a machine-learning regression model to predict pasture biomass. The developed model can be implemented easily, provided that sufficient reference data and cloud-free observations are available.

Chapter 4 investigates the climatologies of average monthly near-surface air temperature (Ta), which is the main indicator of nature-ecology that determines a nation's economic development, especially in drylands. Direct measurements of Ta at a height of 2 m above ground are only available from a limited number of meteorological stations. For

Mongolia, the spatial coverage of these measurements is inadequate. In addition, typical Ta time series comes with many missing values. On the contrary, satellite-derived land surface temperature (LST) data are continuous in both spatial-temporal coverages. The main objective of this study is to develop a robust statistical model to estimate climatologies of monthly average Ta over Mongolia using Moderate Resolution Imaging Spectroradiometer (MODIS) LST time series products and terrain parameters. The PLS and RF regression models were analyzed in this study linking data from 63 automatic weather stations (Ta) with Earth observation (EO) images. Both models were trained to predict Ta climatologies for each of the twelve months, using up to 17 variables as predictors. The four most predictive variables were day/nighttime LST, elevation, and latitude. Using the developed RF models, spatial maps of the monthly average Ta at a spatial resolution of 1 km were generated for Mongolia. This spatial dataset is used to estimate important bioclimatic and climatic variables in Mongolia. The method is transparent and relatively easy to implement.

Chapter 5 investigates bioclimatic and climatic variables (indices) assessment in Mongolia. The main objective is to explore alternative ways and to improve the temporal and spatial resolution of bioclimatic and climatic variables. Two-time series datasets monthly mean Ta, and monthly total precipitation (P) from Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data were used. Spatial maps of 19 bioclimatic variables and 6 climatic indices at a spatial resolution of 1 km were generated, representing the period 2002-2017. The success of the study was to the fact that climatologies of both Ta, as well as precipitation, can be retrieved from EO data over monthly intervals. In areas with sparse station density, EO data avoids otherwise necessary interpolation techniques. The main limitation of many EO products relates to the fact that data sets are still relatively short and that data from multiple satellites would have to be combined and normalized if longer time series are required. The advantage of the MODIS data set is, that it covers the most recent 15 years. In the future, spatial and temporal resolution and spatial coverage will favor EO data even more than other techniques as new satellites are launched at an unprecedented pace. For future research, recommend focusing on the improved quality, spatial, and temporal resolution of precipitation estimates.

Chapter 6 investigates the land potential assessment in Mongolia. The main objective is to assess land potential in Mongolia using a time series of environmental variables and four different regression models. The analysis used 25 environmental variables related to topography, climate, soil, and vegetation as explanatory variables. Reference biomass data from

a field survey of 12988 sites were used for training. The accuracy of the RF model (coefficient of determination (R^2)=0.73) model was much higher than PLS (R^2=0.46), Principle Component Regression (PCR) (R^2=0.55), and Classification and Regression Tree (CART) (R^2=0.60). Using the developed RF model, a spatial distribution map of land potential in Mongolia was generated at a resolution of 500 m. Compared with current pasture use, the land potential map showed that 52.3% of the territory has exceeded the land potential, of which 26.7% highly exceeded. This result showed that it is possible to assess land potential across the huge land surface of Mongolia using machine learning models, a time series of environmental datasets, and training data. In addition, the natural elements and processes are studied, and robust models are developed, while the potential of the newest generation of EO satellites was evaluated while leveraging modern machine learning techniques for information extraction.

Chapter 7 provides a summary conclusion based on the finding from this study. Developed models and spatial distribution maps within this study can contribute to reasonable decisions on sustainable use and land management. Attention needs to be paid to conveying the results that have been acquired to decision-makers and the public.

Acknowledgments

Looking back at my academic journey, I received scientifically and personally valuable knowledge and experience from many people who helped me make this study. The presented research in this thesis would not be possible without the contributions of my family, supervisors, advisors, colleagues, and friends.

First of all, I am enormously thankful to my loved family, husband B.Barkhasragchaa, daughter B.Nomin-Erdene, son B.Nomtoibayar, my mother-in-law B.Shiirev, brothers, sisters, and my siblings for their unlimited love, courage, and permanent support to make my study abroad. My loved dad, D.Otgonbayar, mom, Ts. Nyamaa, and my father-in-law S.Baldorj passed away however, I am always greatly grateful to them. I will phrase it Mongolian: "Юуны өмнө хайртай гэр бүлийнхэн, хань Б.Бархасрагчаа, охин Б.Номин-Эрдэнэ, хүү Б.Номтойбаяр, хадам ээж Б.Шийрэв, ах, эгч, дүү нарынхаа хязгааргүй их хайр, урам зориг, гадаадад суралцахыг минь байнга дэмжиж байсанд үргэлж талархаж явдагаа илэрхийлье. Хайрт аав Д.Отгонбаяр, ээж Ц.Нямаа, хадам аав С.Балдорж нар минь хорвоог орхисон хэдий ч тэдэндээ үргэлж талархаж явдаг."

I am deeply grateful to my supervisors Prof. Dr. Saito Histoshi, and Prof. Dr. Yokoyama Satoshi. My daily supervisor Saito Histoshi, I am deeply grateful to you for your patience and scientific expertise that helped me until finally submitted this thesis. Thank you very much for continuously being available to advise, comment, and correct me. I am deeply grateful to Prof. Dr. Yokoyama Satoshi for providing me with the great opportunity to undertake my Ph.D. research at the Graduate School of Environmental Studies (GSES), Nagoya University under your guidance and support. I consider myself very fortunate to have you as my supervisor.

I would like to thank my examiners of the Department of Environmental Earth Science, GSES Prof. Dr. Masato Shioda, and Department of Social and Environmental Studies/Collaborative Research Center for Disaster Mitigation, GSES Prof. Dr. Yasuhiro Suzuki. Thank you for agreeing to be committee members and for taking the time to review my PhD thesis and attend my defense. I am highly thankful to Jonathan Chambers for his efforts in rigorously English editing our five series scientific papers and the first and final chapters of my thesis.

I am deeply grateful to my advisors Prof. Dr. Clement Atzberger, and Prof. ScD. Amarsaikhan Damdinsuren. For the academic years of 2016/2017, and 2018/2019, I was a

visiting fellow at the Natural Resources and Life Science University (BOKU), Vienna, Austria under the advisor Prof. Dr. Clement Atzberger. I think this was a milestone in my professional life. I am deeply indebted to Prof. Dr. Clement Atzberger for his guidance, advice, valuable comments, and support who helped me to achieve my goals. I am also highly thankful to Prof. ScD. Amarsaikhan Damdinsuren for introducing me to the world of remote sensing and for all the guidance, and advice during my research. I am highly grateful to Anja Klisch, Matteo Mattiuzzi, Valentin Pesendorfer, Sebastian Böck, and Markus Immitzer who are researchers at the BOKU for their assistance in writing R code.

I express my sincere appreciation to the senior scientists ScD. B. Oyungerel, Dr. D.Enkhtaivan, Dr. E.Avirmed, the young scientist Dr. T.Renchinmyadag, and all my colleagues at the Division of Physical Geography and Environmental Study, of the Institute of Geography and Geoecology (IGG), Mongolian Academy of Sciences (MAS) for their cooperation, and continuous support. I would like to acknowledge the human resources A. Chimegsaikhan, the director Dr. A.Dashtseren, and all my colleagues at IGG, MAS for inspiring and support during my academic journey.

I would like to thank the examining committee members of the Department of Social and Human Environment, GSES Prof. Dr. Suzuki Yasuhiro, Prof. Dr. Takahashi Makoto, Prof. Dr. Imazato Satoshi, Prof. Dr. Kushima Momoyo, and Prof. Dr. Masaya IGA for their useful comments and advice, and for taking the time to attend my pre-defense. It was my pleasure to belong to this team of research students at the Laboratory of Geography. I am deeply grateful to all students for their fun chats that made my student life enjoyable. I will be missing the kind people and humid weather of Nagoya city.

Finally, I am greatly thankful to the staff of the Asian Satellite Campus Institute (ASCI) in Nagoya and Ulaanbaatar who are who are Dr. Prof. Akira Yamauch, Dr. Prof. Gangabaatar Dashbalbar, Ms. Odake Yuko, and Ms.Todrol Erdenesan, and the staff of the Student Affairs Office of GSES for their permanent support, useful instructions, and arrangement to provide all needs for the research.

CONTENTS

Summary i
Acknowledgments iv
List of Tables 1
List of Figures 3
List of Appendix Tables 6
List of Appendix Figures 7
Abbreviations 9
CHAPTER 1. Introduction 15
1.1 Definitions, concepts, and principles of land assessment 15
1.2 Historical development and technological achievements for land assessment classification systems, tools, and criteria 19
1.2.1 Land assessment classification systems 19
1.2.2 Land assessment tools 20
1.2.3 Land assessment criteria 22
1.3. The necessity for assessing land resource potential in Mongolia 25
1.3.1 The necessity for relating land resource potential concerns in Mongolia 25
1.3.2 Previous investigations and the necessity of a new technique in Mongolia 30
1.4 The objective and organization of this thesis 33
CHAPTER 2. Land suitability evaluation for cropland in Mongolia using the spatial MCDM method and GIS-based AHP 35
2.1 Introduction 35
2.2 Methodology and data 36
2.2.1 Methodology 36
2.2.2 Data 39
2.2.3 Developed criteria parameters for land suitability evaluation for agricultural cropland 40
2.3. Result 42
2.3.1 Result of constraint factor analysis based on the Boolean logic theory 42
2.3.2 Result of factor analysis based on the spatial MCDM method 42
2.3.3 Result of ranking and weights analysis of the criteria based on the AHP 44
2.3.4 Result of map layer overlay analysis based on suitability index 44
2.4 Conclusion 46

CHAPTER 3. Estimation of pasture biomass in Mongolia using Partial Least Squares (PLS) and Random Forest (RF) regression models and Landsat 8 imagery 47
3.1 Introduction 47
3.2 Methodology and data 50
3.2.1 Methodology 50
3.2.2 Data 52
3.2.3 Predictor variables and model development 55
3.3 Result 57
3.4 Discussion 59
3.5 Conclusion 61
CHAPTER 4. Estimation of climatologies of average monthly air temperature over Mongolia using MODIS Land Surface Temperature (LST) time series and machine learning techniques 62
4.1 Introduction 62
4.2 Methodology and data 64
4.2.1 Methodology 64
4.2.2 Data 67
4.2.3 Data analysis using RF and PLS models 70
4.3 Result 72
4.3.1 Comparison of RF and PLS models: variable importance and prediction accuracy 72
4.3.2 Maps of predicted air temperatures using RF models with the reduced feature set 74
4.4 Discussion 77
4.5 Conclusion 79
Appendix A 80
CHAPTER 5. Estimation of bioclimatic and climatic variables of Mongolia derived from a time series of remote sensing data 81
5.1 Introduction 81
5.2 Methodology and data 83
5.2.1 Methodology 83
5.2.2 Data 87
5.3 Result 89
5.4 Discussion 93
5.5 Conclusion 95
CHAPTER 6. Land potential assessment in Mongolia using linear and non-linear regression models and a time series of environmental variables 96

6.1 Introduction 96
6.2 Methodology and data 97
6.2.1 Methodology 97
6.2.2 Data 99
6.3 Results 106
6.3.1 The interrelationship between environmental variables 106
6.3.2 Importance variables and prediction accuracy using the RFR model 112
6.3.3 Comparison of linear and non-linear regression models 114
6.3.4 Map of predicted LP using RFR model 116
6.3.5 Interaction between LP and pasture use 116
6.4 Discussion 120
6.5 Conclusion 123
Appendix B. 124
Appendix C. 124
CHAPTER 7. Conclusion 125
7.1 Conclusion 125
7.2 Further studies 129
References 131
Appendix A 163
Appendix B 166
Appendix C 179

List of Tables

Table 1. 1 Definition of land suitability, land resources, and land potential 18
Table 1. 2 Nationwide research related to land suitability, land resources, and landscape potential 32
Table 2. 1 Details of used data 40
Table 2. 2 Evaluation of the multi-criteria parameters 41
Table 2. 3 Evaluation of the constraint criteria parameters 41
Table 2. 4 Defined ranking and weights of the criteria 44
Table 2. 5 Suitability classification results for cropland in Mongolia 45
Table 3. 1 Data characteristics and data sources. The upper section refers to the raster data, while the lower section lists the field data. Note that the mapping was done using Landsat 8 data. 53
Table 3. 2 Statistical descriptors (number, minimum, maximum, mean, and standard deviation) of the field-measured (reference) biomass samples as well as the seventeen analyzed spectral indicators and six spectral bands. For the abbreviations of spectral indices, the reader is referred to in Table 3.2. 54
Table 3. 3 Spectral indices used in this study. G-green wavelength, B-blue wavelength, R-red wavelength, NIR-Near-Infrared wavelength, SWIR- Short Wavelength Infrared, α a value of 0.3, a, b, c and d are coefficient where a = 1.7149, b = - 0.0157, c = 0.01281, d = - 0.0113. 56
Table 3. 4 Inter-correlations between spectral indices (n = 553). In the lower triangle, spectral indices with significant inter-correlation (e.g., $r \geq 0.90$) are highlighted (in gray). For the abbreviations, see Table 3.3. 56
Table 3. 5 Inter-correlation between Landsat 8 reflectances (n = 553). High inter-correlations ($r \geq 0.90$) are indicated in gray in the lower triangle 57
Table 3. 6 Summary statistics (R^2 and RMSE) for multi-variate biomass prediction models (n = 553) involving spectral reflectances (ρ), respectively, spectral VI. 57
Table 4. 1 Description of Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) products used in this study (source: Land Processes Distributed Active Archive Center (LP DAAC), 2019). 69
Table 4. 2 List of response/predictor variables and corresponding descriptive statistics (period 2002 to 2017). The list includes the measured air temperature (Ta) reference data at the weather station level (n = 712 for each of the twelve months) as well as the corresponding seventeen predictor variables extracted from satellite and other geo-data. For the acronyms of the variables, see Table 4.1. 71
Table 4. 3 Seven model subsets were studied. The seven groups were generated to study the relations between responses and up to 17 predictor variables. N_{var} indicates the number of predictor variables in each group. 71
Table 4. 4 Modeling results were obtained using the RF regression. Reported are the monthly summary statistics (coefficient of determination (R^2) and root-mean-square error (RMSE)) for Ta prediction models for each of the seven groups of variables. For details of groupings G1 to G7, see Table 4.3. 74
Table 5. 1 Formula and description of the bioclimatic variables (O'Donnell and Ignizio 2012). T_{avg}, T_{max}, and T_{min} are the monthly average, maximum, and minimum air temperature, and PPT is the monthly total precipitation 84

Table 5. 2 Performance measures used in this study: coefficient of determination (R^2), root mean squared error (RMSE), and normalized RMSE (nRMSE). The three statistics represent correlation (association), error (residual), and range normalized errors (Richter et al. 2012). 85
Table 5. 3 Two precipitation products with a high spatial resolution (Roca et al. 2019; Sun et al. 2018; Bai and Liu 2018; Beck et al. 2017). Only CHIRPS was used for this study. 88
Table 5. 4 Inter-correlation between estimated monthly air temperatures derived from MODIS LST, and World climatic datasets; precipitation CHRIPS and World climatic datasets (n=63). High inter-correlations ($r \geq 0.90$) are highlighted in blue. 89
Table 5. 5 Descriptor statistics of the estimated 19 bioclimatic variables (SatClim) for the years 2002-2017 (n=1 575 107 pixels) 90
Table 5. 6 Summary of statistics describing the correspondence between SatClim and WorldClim data over Mongolia (R^2, RMSE, and nRMSE). For the comparison, 19 variables were extracted from the full image (n=1 575 107 pixels). High correlations ($R^2 \geq 0.70$) are highlighted in light grey. 92
Table 6. 1 Data sources and their characteristics 100
Table 6. 2 The summary statistical description of all variables used in this study 101
Table 6. 3 Field-measured rangeland recovery data (RefData) used, being developed by the Ministry of Food and Agriculture (MoFA), ALAMGaC, IRIMHE, and Associations of Pasture Users Group (APUG) of Mongolia in 2015 (GGP-SDC 2015) 102
Table 6. 4 The five groupings generated from the 4 main, and 23 sub-explanatory variables 112
Table 6. 5 The summary statistics of the R^2 and RMSE for the LP prediction model, including five groups of variables using the RFR model 113
Table 6. 6 The interaction between historical pasture use and estimated LP in Mongolia 120

List of Figures

Figure 1. 1 Study area and its natural characteristics. (a) Digital Elevation Model (DEM) derived from the Shuttle Radar Topography Mission (SRTM) with automatic weather stations (n=63), (b) Köppen climate classification of the world (Kottek et al. 2006), (c) estimated annual average air temperature derived from MODIS MOD11A2 (v006) (Otgonbayar et al. 2019), (d) annual total precipitation (Fick and Hijmans 2017), (e) average annual NDVI derived from MODIS MOD13A2 (v006) for the period 2002-2017, (f) Terrestrial ecoregions of the world (Olson et al. 2001). 27

Figure 1. 2 Total livestock population for the period 1991–2021 (NSOM 2022b) 28

Figure 1. 3 Total population with urban and rural, and number of households with livestock for the period 1991–2021 (NSOM 2022c, 2022e) 28

Figure 1. 4 The ratio between the area of pastureland and the number of livestock (NSOM 2022a; NSOM 2022b). 29

Figure 1. 5 Annual mean air temperature, °C (top left) for the period 1961-1990, (top right) for the period 1991-2021. Yearly average temperature (bottom left); yearly average temperature anomalies (bottom right) for the period 1991-2021 29

Figure 1. 6 Annual total precipitation, mm: (top left) for the period 1961-1990, (top right) for the period 1991-2021. Yearly total precipitation (bottom left); the yearly total precipitation anomalies (bottom right) were derived from CRU-TS for the period 1991-2021. 30

Figure 2. 1 Land use constraint condition evaluation (Boolean map method). 42

Figure 2. 2 The main factors used in cropland suitability evaluation 43

Figure 2. 3 Suitable sites for cropland development (Multi-criteria factor analysis) 45

Figure 2. 4 Suitability classification map for cropland in Mongolia. 45

Figure 2. 5 Evaluation validation compared with the current extent of sown area 46

Figure 3. 1 General workflow for modeling pasture biomass using spectral predictors from Landsat 8 and for generating biomass maps 51

Figure 3. 2 (a) Base map Landsat 8 imagery from June to September 2016 with false color composite (Red: Band 5; Green: Band 4; Blue: Band 3), and Mongolia's 21 administrative units. Figure 3.2 (b) Sample points (n = 553) for which reference biomass information was available. Projection system world geographic system (WGS) 1984, central meridian 105 (zone 48), datum WGS1984 in Figure 3.2 (a) – (b). 52

Figure 3. 3 Frequency distribution of field-measured biomass samples available for the study and distributed over Mongolia (n = 553) 54

Figure 3. 4 Importance of spectral vegetation indices (VIs) for (a) PLS model, and (b) RF regression model. It is noted that the two model types use different methods to quantify the variable importance. 58

Figure 3. 5 Estimated biomass using Landsat 8 spectral vegetation indices (CL_{green}, SR, NDVI, EVI1, WDRVI, and $MSAVI_2$) and PLSR model. The resulting map was scaled to bit (0 - 255) 58

Figure 3. 6 Modeled versus measured biomass (n = 553) using the PLSR model. R^2 was 0.750 and RMSE = 101.10 kg ha^{-1}. 59

Figure 3. 7 Comparison estimated biomass values from the PLSR model and 17 years average MODIS vegetation product (MOD17) GPP (Kg C $(m^2)^{-1}$). R^2 was 0.817 and RMSE = 91.30 kg ha^{-1}. 59
Figure 4. 1 Frequency distribution of measured average 8-day air temperature reference data (n = 8544) from the 63 automatic weather stations for the years 2002–2017. Monthly statistics are depicted in Table 4.2. 70
Figure 4. 2 Correlation matrix between response and predictor variables (n = 8544). The saturation of the colors indicates the strength of the correlations. Positive correlations are shown in blue and negative correlations in red. In this graph, the air temperature data has been pooled across the twelve months. For the abbreviations, see Table 4.1. 70
Figure 4. 3 Random forest (RF) variable importance for each month. The importance is here given as the percentage increase in mean square error (%IncMSE). 73
Figure 4. 4 Comparison between measured and estimated monthly average Ta using LSTd, LSTn, and elevation for the RF regression model. 75
Figure 4. 5 Estimated monthly average Ta based on RF regression model using LSTd, LSTn, and elevation as predictor variables. (a) Spatial maps of estimated monthly average Ta over Mongolia at 1 km spatial resolution. (b) Monthly statistics of R^2 (blue) and RMSE (red) between observed and predicted air temperature. 76
Figure 4. 6 Estimated average Ta per season using LSTd, LSTn, and elevation as predictor variables. Spatial maps of seasonal-average Ta over Mongolia at 1 km spatial resolution using the RF (first column) and partial least squares (PLS) regression models (second column). In the last column, the difference between the two model outputs is shown. 77
Figure 5. 1 Inter-correlation between estimated monthly air temperatures derived from MODIS LST, and World climatic datasets; precipitation CHRIPS and World climatic datasets (n=63). High inter-correlations ($r \geq 0.90$) are highlighted in blue. 89
Figure 5. 2 (a) Modeled 19 SatClim bioclimatic variables using MODIS and CHIRPS data 2002–2017, (b) WorldClim variables with gridded data 1971–2000, (c) Frequency distributions of SatClim and WorldClim. 91
Figure 5. 3 Spatial distribution maps 6 different climatic indices such as hydrothermal, aridity, humidity, moisture, biological effectiveness, and bioclimatic potential over Mongolia. 92
Figure 6. 1 The distribution of the pasture monitoring sites across Mongolia from ALAMGaC in 2020. 103
Figure 6. 2 The results of the inter-analysis correlation between topography variables (n=19567425). The diagonal plots represent a histogram and distribution density, the top right plots represent a coefficient of determination (R^2) (Pearson), and the bottom left plots are represented scatter plots for a linear model distribution. 107
Figure 6. 3 The results of inter-correlation analysis between climatic variables (n=19567425). The diagonal plots represent a histogram and distribution density, the top right plots represent R^2 (Pearson), and the bottom left plots represent scatter plots in a linear model distribution. 108
Figure 6. 4 The results of inter-correlation analysis between soil variables (n=19567425). The diagonal plots represent histogram and distribution density, the top right plots represent R^2 (Pearson), and the bottom left plots represent scatter plots in a linear model distribution. 109
Figure 6. 5 The results of inter-correlation analysis between vegetation variables (n=19567425). The diagonal plots represent a histogram and distribution density, the top right plots represent R^2 (Pearson), and the bottom left plots represent scatter plots in a linear model distribution. 110

Figure 6. 6 The correlation matrix between dependent (n=12988) and 25 independent variables. The positive and negative correlations are represented in blue and orange, respectively. The acronyms of each variable are shown in Table 6.2. 111
Figure 6. 7 Evaluation of the important variables for the five groupings using the RFR model (n=12988). Importance is given to the increased node purity (IncNodePurity) 113
Figure 6. 8 Estimated LP based on the RFR model using field-measured recovery data of the 12988 sites and 25 environmental variables. A spatial distribution map of the assessment of the LP in Mongolia at a resolution of 500 m. Each class represents the rangeland recovery class: very high potential class 1, high potential class 2, moderate potential class 3, low potential class 4, and very low potential class 5, respectively. 116
Figure 6. 9 Spatial distribution maps of the livestock grazing capacity of Mongolia. (top left) mean of 1991-2000, (top right) mean of 2001-2010, (bottom left) mean of 2011-2020, and (bottom right) mean of 1991-2020, respectively. 118
Figure 6. 10 The interaction between historical pasture use (since the social transition), and estimated LP across Mongolia. (top left) mean of 1991-2000, (top right) mean of 2001-2010, (bottom left) mean of 2011-2020, and (bottom right) mean of 1991-2020 119
Figure 6. 11 The interaction between the status of pasture use in 2021 and the estimated results of LP in Mongolia. 119
Figure 7. 1 Results of land resources potential assessment in Mongolia. (a) the result of the land suitability analysis for cropland; (b) the result of pasture biomass analysis; (c) the result of the land potential analysis; (d) the ratio between the livestock density and estimated land potential. 127

List of Appendix Tables

Table B. 1 Climatic indices formulas and their sources that were used in this study 168
Table B. 2 The correlation matrix between topographic variables was derived from STRM DEM at a resolution of 90 m (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of the p-value (Pearson): "***" - <0.0001, "**" - <0.01, "." - <0.1. 169
Table B. 3 The correlation matrix between yearly mean climatic variables was extracted from CRU-TS and CRUNCEP datasets for the period 1991-2021 (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of each p-value (Pearson) was noted as <0.0001. 172
Table B. 4 The correlation matrix between soil variables (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of each p-value (Pearson) was noted as <0.0001. 173
Table B. 5 The correlation matrix between vegetation variables (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of each p-value (Pearson) was noted as <0.0001. 173
Table B. 6 Summary statistics of the PLSR model (n=12988) 173
Table B. 7 Summary statistics of the PCR model (n=12988) 174
Table B. 8 Summary statistics of the R^2 and RMSE for land potential prediction models, including 5 groups of variables using linear regressions (n=12988) 174

List of Appendix Figures

Figure A. 1 Variable importance in projections (VIPs), R^2, and RMSE for the 17 predictor variables in the twelve-monthly PLSR models. 163

Figure A. 2 Estimated monthly average Ta based on the PLS regression model and using LSTd, LSTn, and elevation as predictor variables. (a) Spatial maps of estimated monthly average Ta over Mongolia at 1 km spatial resolution. (b) Monthly statistics of R^2 (blue) and RMSE (red) between observed and predicted air temperature. 165

Figure B. 1 Variation of the (a) monthly mean air temperature, (b) monthly mean precipitation, (c) the monthly mean potential evapotranspiration, and (d) monthly mean vapor pressure derived from the CRU-TS v4.06 monthly climate dataset for the period 1991-2021 as an average of all weather stations (n=63) using a zonal statistics approach. 166

Figure B. 2 (left) Monthly mean temperature (Ta), and (right) monthly mean precipitation (Precip) for the period 1991-2021 as an average of all weather stations. Letters of the alphabet represent twelve months. 166

Figure B. 3 The correlation matrix between monthly mean Ta and Tmp (top), and monthly mean Precip and Prec (bottom). The monthly mean Ta and Precip data were obtained from weather stations for the period 1991-2021. Monthly mean Tmp and Prec data were extracted from CRU-TS for the same period. Extracted monthly mean Tmp and Prec data was an average of all weather stations. High correlations are highlighted in dark blue. 167

Figure B. 4 Variation of the annual mean soil moisture profile for the individual pixel was extracted from ERA datasets for the period 1991-2021. 168

Figure B. 5 Variation of the mean of (a) GPP, (b) LAI, (c) Fpar, (d) NDVI for individual pixels from MODIS and GIMMS during the vegetation growing season (April-August) for the years 2002-2021, and 1991-2015, respectively. The sample pixels were selected from the objective variable (n=12988). Each variable of the curves was counted by averaging all objective variables. 168

Figure B. 6 The spatial distribution maps of elevation (top left), slope (top center), aspect (top right), TPI (bottom left), and TWI (bottom center) derived from STRM DEM at a resolution of 90 m. 169

Figure B. 7 Spatial correlations between topographic variables. 170

Figure B. 8 Spatial distribution maps of yearly mean Srad, Tmp, Prec, PET, Vap, and Wnd were extracted from CRU-TS and CRUNCEP monthly climate-gridded datasets for the period 1991-2021. The climate indices of HCT, Iar, HFth, and MI were calculated from yearly mean datasets of Tmp, Prec, and PET for the period 1991-2021 using empirical equations (see, Table B.1). 171

Figure B. 9 Spatial distribution maps of yearly mean SOC, sHumus, and sMoisture. (left) The SOC variable was extracted from the Trends.Earth system at a resolution of 250 m (Conservation International, 2022). (center) The vector data of soil humus were obtained from the IGG, MAS (right). The soil moisture data were extracted from the ECMWF database at a resolution of 0.28º for the period 1991-2021. 172

Figure B. 10 Spatial distribution maps mean modisNDVI, gimmsNDVI, Fpar, GPP, and LAI during the vegetation growing season (April to August). modisNDVI, Fpar, GPP, and LAI derived from the MODIS Aqua and Terra satellites at spatial resolutions from 250 to 500 m for the period 2002-2021 (LP DAAC, 2022). The

mean gimmsNDVI derived from AVHRR/NOAA at a spatial resolution of 8 km for the period 1991-2015 during the vegetation growing season. 172
Figure B. 11 Estimated importance variables for the five groupings using the PLSR model (n=12988). 175
Figure B. 12 Regression tree of the CART was estimated from 23 explanatory variables with three steps of maximum tree depth (n=12988). The structure of each node is organized in a top-to-bottom direction. The right side of each level represents the important variables with a threshold value. In contrast, the left side of each level represents the less important variables with a threshold value. 176
Figure B. 13 Spatial distribution maps of livestock population density (sheep unit ha^{-1}) in Mongolia for the period 1991-2021. 177
Figure B. 14 Livestock population density (sheep unit) in Mongolia for each decade from 1991, and the 31-year average. 177
Figure B. 15 Spatial distribution maps of the total number of livestock population density (head ha^{-1}) in Mongolia for the period 1991-2021. 178
Figure C. 1 The trend of slope value of linear regression estimated from GIMMS and MODIS NDVI for the period 1990–2019 179
Figure C. 2 The trend of slope value of significant linear regression estimated from GIMMS and MODIS NDVI for the period 1990–2019 179
Figure C. 3 (a-f) Here, fitted components of trend, seasonal, and the remainder (evaluated noise) for the time series monthly NDVI of MODIS (left above) and GIMMS (left below). The abrupt change was observed in the trend component of the NDVI time series (right above and below). The grey, black, red, and blue lines represented a primary NDVI curve, fitted NDVI curve, direction and magnitude of abrupt change, and trend before and after the change, respectively. (a) Forest to barren land by mining activity (Latitude: 49.640, Longitude: 107.721). 180
Figure C. 4 Estimated strongly degraded area from MODIS and GIMMS NDVI time series for the period 1990–2019 using trend, BFAST, and RESTREND analysis methods 183
Figure C. 5 Generated strongly degraded area by human activity from Landsat OLI imagery, Google Earth map, and environmental geo-database of Mongolia 183

Abbreviations

ALAGG	Agency of Administration of Land Affair, Geodesy, and Cartography
AEZS	Agro-Ecological Zoning System
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
ALAS	Automated Land Assessment Systems
ALMGaC	Agency for Land Management, Geodesy, and Cartography
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AO	Additional Offset
ASCI	Asian Satellite Campus Institute
Aspct	Aspect
AVHRR	Advanced Very High-Resolution Radiometer
BCP	Bioclimatic potential
BEC	Climate biological efficiency indicators
BFAST	BFAST
BI	Brightness Index
BOKU	Natural Resources and Life Science University
CART	Classification and Regression Tree
CCD	Cold Cloud Duration
CHRIPS	Climate Hazards Group InfraRed Precipitation with Station
CI	Colorations Index
CL_{green}	Green Chlorophyll Index
CP	Compromise Programming
CR	Consistency Ratio
CRUNCEP	Climate Research Unit-National Centers for Environmental Prediction
CRU-TS	Climate Research Unit-Time Series
CSWR	Climate Space Weighted Regression
DEM	Digital Elevation Model
DN	Digital Number
DVI	Difference Vegetation Index
ELECTRE	Elimination and Choice Expressing Reality
Elv	Elevation
ENVI	Environment for Visualizing Images
EO	Earth Observation

EROS	Earth Resources Observation and Science
ESs	Ecosystem Services
ETM	Enhanced Thematic Mapper
ETM+	Enhanced Thematic Mapper Plus
EVI	Enhanced Vegetation Index
FAO	Food and Agricultural Organization
FEWS NET	Famine Early Warming Systems Network
Fpar	Fraction of Photosynthetically Active Radiation
FVC	Fraction of Vegetation Cover
GAEZS	Global Agro-Ecological Zones System
GBM	Generalized Boosted Model
GDP	Gross Domestic Product
GEE	Google Earth Engine
GIMMS	Global Inventory Monitoring and Modeling System
gimmsNDVI	GIMMS Normalized Difference Vegetation Index
GIS	Geographic Information Systems
GPP	Gross Primary Production
GSD	Ground Sampling Distances
GSES	Graduate School of Environmental Studies, Nagoya University
GSI	Top Grain Size Index
GWR	Geographically Weighted Regression
HDF	Hierarchical Data Format
HF_{th}	Humidity Index or Humidity Factor
HI	Hue Index
HTC	Hydrothermal Coefficient
I_{ar}	Aridity Index
IGG	Institute of Geography and Geoecology
IIASA	International Institute for Applied Systems Analysis
IRIMHE	Information and Research Institute of Meteorology, Hydrology, and Environment
IRP	International Resource Panel
LAI	Leaf Area Index
LASSO	Least Absolute Shrinkage and Selection Operator
Lat	Latitude
LCC	Land Capacity Classification
LCCS	Land Capacity Classification System

LDN	Land Degradation Neutrality
LGC	Livestock Grazing Capacity
LIDAR	Light Detection and Ranging
LMU	Land Mapping Units
Long	Longitude
LOO-CV	Leave One Out Cross Validation
LP DAAC	Land Processes Distributed Active Archive Center
LP	Land Potential
LST	Land Surface Temperature
LSTd	Daytime Land Surface Temperature
LSTn	Nighttime Land Surface Temperature
LUT	Land Utilization Types
Lvstk	Number of livestock
MAP	Map Analysis Package
MAS	Mongolian Academy of Sciences
MAUT	Multi-Attribute Utility Theory
MCDM	Multi-Criteria Decision Making
MCE	Multi-Criteria Evaluation
MERRA	Modern Era Retrospective-analysis for Research and Application
MI	Moisture Index
ML	Machine Learning
MLR	Multiple Linear Regression
MODIS	Moderate Resolution Imaging Spectroradiometer
modisNDVI	MODIS Normalized Difference Vegetation Index
MODM	Multi-Objective Decision-Making
MPRP	Mongolian People's Revolutionary Party
MSAVI	Modified Soil Adjusted Vegetation Index
MSI	Moisture Stress Index
NAMEM	National Agency for Meteorology, and Environment Monitoring
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
$NDVI_{green}$	Green Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near Infrared
NPP	Net Primary Production

nRMSE	Normalized Root Mean Squared Error
NSO	National Statistical Office
NSOM	National Statistical Office of Mongolia
ODW	One-Dimensional Weights
OLI	Operational Land Imager
OLS	Ordinary Least Squares
OOB	Out of Bag
OWA	Ordered Weighted Averaging
p	Spectral Reflectances
PCA	Principal Component Analysis
PCC	Pasture Carrying Capacity
PCR	Principal Component Regression
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Cloud Classification System
PET	Potential EvapoTranspiration
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
PNS	Photo-monitoring Network Sites
PPM	Photo Point Monitoring
P	Precipitation (automatic weather stations data)
Prec	Precipitation (reanalyzed data)
Precip	Precipitation (in situ metrological data)
PROMOTHEE	Preference Ranking Organization Method for Enrichment Evaluation
QA	Quality Assessment
QAs	Quality Assurances
QC	Quality Controls
QGIS	Quantum Geographic Information System
RADAR	Radio Detection and Ranging
RefData	Field-measured rangeland recovery data
RF	Random Forest
RFR	Random Forest Regression
RI	Redness Index
RiR	Ridge Regression
RMSE	Root Mean Square Error
RP5	Reliable Prognosis

RS	Remote Sensing
RST	Random Set Theory
RVI	Ratio Vegetation Index
RW	Random Weights
SAGA	System for Automated Geoscientific Analyses
SCP	Semi-automatic Classification Plugin
SDGs	Sustainable Development Goals
SDMs	Species Distribution Models
SDSs	Scientific Data Sets
SF	Scale Factor
sHumus	Soil Humus
Slp	Slope
SLR	Simple Linear Regression
sMoisture	Soil Moisture
SOC	Soil Organic Carbon
SPOT	Satellite Pour Observation Terra
SR	Simple Ratio
Srad	Solar radiation
SRTM	Shuttle Radar Topography Mission
STRK	Spatiotemporal Regression-Rriging
SU	Sheep Unit
SVM	Support Vector Machine
SW	Selected Weights
SWIR	Short Wavelength Infrared
Ta	Near-surface air temperature (automatic weather stations data)
TIR	Thermal Infrared
Tmp	Air temperature (reanalyzed data)
TOPSIS	Technique for Order by Similarly to Ideal Solution
TPI	Topographic Position Index
TVX	Temperature-Vegetation Index
TWI	Topographic Wetness Index
UN	United Nations
UNCED	United Nations Conference on Environment and Development
UNEP	United Nations Environment Program
USAID	United States Agency for International Development

USDA	United States Department of Agriculture
USGS	United States Geological Survey
Vap	Vapor pressure
VI	Vegetation Index
VIP	Variable Importance in Projection
VIs	Vegetation Indices
VR	Valid Range
WDRVI	Wide Dynamic Range Vegetation Index
$WDRVI_{green}$	Green Wide Dynamic Range Vegetation Index
Wnd	Wind
WSS	Water Storage in soil

CHAPTER 1. Introduction

1.1 Definitions, concepts, and principles of land assessment

Land assessment has multiple aspects (King 2019; FAO 2007; UN 1995; UNEP 1993). This thesis emphasizes land resource potential assessment consisting of land suitability, land resources, and land potential. The main objective of this study was to assess and model land resource potential throughout Mongolia using state-of-the-art machine learning techniques and remote sensing data. To achieve the objective, this study examines multiple perspectives of land resource potential such as land suitability, land resources, and land potential, focusing on developing robust models. The main reason for assessing land resource potential is to investigate what land resources are current, what land has high potential, how to link land use with land potential, and how to implement it. The first formal definition of land was defined by the Food and Agricultural Organization of the United Nations (FAO) in 1976 as "a delineable area of the earth's terrestrial surface, encompassing all attributes of the biosphere immediately above or below the surface", which was updated in 1995 (UN 1995; FAO 1976). According to Sombroek (1993; 1997), the land is physical material, which is a habitat for living organisms, and a source of materials necessary for their activities. Moreover, land assists in the supplying of fundamental ecosystem services such as managing biophysical and biochemical cycles or fiber and meal provision (Tilman et al. 2002). Furthermore, land includes all attributes of the biosphere immediately near the surface, the settlement pattern of humans, and the results of current and past human-induced activities (Rossiter 1996).

Land assessment has been critical for the FAO since it was established in 1945 (Sombroek 1997). The first land assessment framework was developed by FAO in 1972 and was published as "A framework for land evaluation" in 1976 in an FAO soil bulletin, which was revised in 2007 (FAO 1976; FAO 2007). As part of this work, the definitions, concepts, principles, and methods basis of the land assessment was finalized, in which land assessment is defined as "the evaluation of land performance when used for specified purposes" (FAO 1976; FAO 2007). Since the first framework, four guidelines on land assessment for extensive grazing, irrigated agriculture, rainfed agriculture, and forestry were processed by FAO (FAO 1991; FAO 1989; FAO 1984; FAO 1983). Before the FAO framework, in the early 1930s United States Department of Agriculture (USDA) defines land assessment as "an approach for the estimating the relative characteristics of land resources established upon measurable attributes of the land" (Klingebiel and Montgomery 1961; USDA 1983). Moreover, Nha (2017)

defines land assessment as the examination of data related to land such as topography, climate, soil, vegetation, and so forth, for detecting practical options for enhancing land usage. According to Sys, VanRanst, and Debaveye (1991), land assessment accords two essential aspects of the land such as physical resources (e.g., topography, geology, climate, soil, hydrology, and vegetation), and socioeconomic resources (e.g., farm size, management level, availability of manpower, market position and other human activities). The physical resources are accounted as relatively stable, while socio-economic resources are more dynamic depending on social and political conditions (FAO 1993; FAO 1991). Land assessment primarily deals with physical resources including topography, soil, and climate (FAO 1976; IRP 2016; Fischer et al. 2021), and these contain the possibilities for the hydrology, geology, and vegetation (Sys, VanRanst, and Debaveye 1991).

For the concepts and principles, there was little need to update these until the early 2000s, because they had widely been used in many countries around the world without requiring significant changes in the land assessment system (FAO 2007). During these years, land assessment investigations were performed to introduce more efficient and better-suited land to the agricultural industry for land use planning and land management. At that time, agricultural productivity (e.g., fiber, feed, food) was primarily considered in the land assessment. However, related to rapid land resource uses, environmental concerns such as land use change, soil fertile loss, erosion, and degradation have started a wider range. Environmental concerns were included in the first framework of the land assessment, however, this was only an outline.

The first UN Conference on the Human Environment was held in Stockholm in 1972. It was focused on environmental concerns, and it led to the establishment of the UNs Environment Program (UNEP) (Sohn 1973). Twenty years after, the Stockholm conference, the UN Conference on Environment and Development (UNCED) was held in Rio De Janeiro for the governments of the world. In this meeting, the sustainable development concept and issues of environmental resources protection were assumed to be at a wider range, which “marked a great step forward in bringing the role of land resources to wider attention” (UNCED 1993; Young 2000). At that time, it was believed that the main factor leading to the shortage of land resources is the production of food products caused by population growth. In 1993, the world's scientific academies convened and called for urgent measures to limit population growth (UN 1993). However, it recognized that the problem of population cannot be considered

in isolation, it is closely related to environmental resources and economic development. The third UN Conference on Population and Development in Cairo in 1994 highlighted the social status of women in society and education as one of the incentives to limit population growth. However, the idea that natural resources themselves are a limiting factor to population growth could not be put forward. From the second UN World Food Summit in Rome in 1996, and the UN Earth Summit in 1997, the main points were "responsible state, poverty, development of the rural sector, public participation and sustainability, disaster relief, and the role of women" and "on halting global warming by reducing greenhouse gas emissions", respectively (UN 1996; UN 1997). Related to these large number of environmental concerns, the scope and intention of the land assessment have required expanding and more systematically considering.

In this regard, the concepts and principles of land assessment were revised in the 2000s to more practically reflect requirements such as sustainable land use, sustainable livelihood, and environmental concerns (e.g., nature conservation, and avoiding degradation). Conceptually, land assessment is defined as “calls for matching of the land management and ecological necessaries of applicable kinds of land utilization with land quality, whilst considering local socio-economic conditions” (FAO 2007). For these principles, concerns related to environmental protection and land use sustainability led to the extending of the principles of land assessment with two new principles. Currently, eight principles are officially implemented in the land assessment system (details see FAO 2007, p29-30). These principles are there to ensure socio-economic equity by envisaging sustainable productivity of land, and environmental concerns.

On the other hand, land assessment comprises many different purposes. The definition of land assessment is distinct depending on the land assessment purposes. According to the literature review, most research on land assessment has been in the framework of land use suitability and assessment of physical resources of the land through both FAO and in many other countries (FAO 1976; FAO 1978; FAO 1984; FAO 1993; Young 2000; Fischer et al. 2021). Land suitability analysis was generally based on the detection of the advantages of various land use and, recently, nature conservation and sustainable management are being considered. Land resource assessment has been directed at determining land properties (e.g. topography, climate, soil, and vegetation) in an area, to understand its features, and resources (DES and Department of Resources 2021). Another new category that has entered the land assessment system in recent years is land potential. Land potential is the inherent potential of

the land, which contributes to finding out where production can be sustainably raised and determining where land can be restored (IRP 2016). Summarizing various concepts regarding land suitability, land resources, and land potential, concludes that it is characterized by physical resources and is purposed for sustainable land use, sustainable productivity, and reproduction. Therefore, this investigation considered multiple perspectives of land resource potential such as land suitability, land resources, and land potential (detailed definitions see in Table 1.1). This investigation will assist decision-makers in taking better-informed decisions about various land use alternatives, and it will contribute to achieving Land Degradation Neutrality (LDN) and the Sustainable Development Goals (SDGs).

Table 1. 1 Definition of land suitability, land resources, and land potential

Types	Definition of terms	Descriptions
Land suitability assessment	It is to detect the applicability of suitable places for specified purposes (Rossiter 1996; Verheye 1997; FAO 1976).	Land suitability identifies how well the Land Mapping Units (LMU) match the requirements of Land Utilization Types (LUT) (Rossiter 1996). LMU is a definitive area of land that is delineated on the map, and LUT is "a specific subdivision of a major kind of land use" (Beek 1975).
Land resources assessment	This is a process investigation of the land attributes (Vink 1975), which "encompasses the physical, biotic, environmental, infrastructural, and socioeconomic components of a natural land unit" (FAO 2023).	In particular, physical (natural) resources such as topography, near-surface climate, soil, hydrology, near-surface sedimentary layer, flora, and fauna (FAO 1976). These resources contain featuring the properties of land in a given area, more to appreciate its features, resources, and suitability for various purposes (DES and Department of Resources 2021).
Land potential assessment	It is "the capacity of the land to resist and recover from degradation" (IRP 2016).	It permits the generation of ecosystem services required to meet current needs without compromising our ability to meet future needs (Herrick et al. 2013; Herrick et al. 2019). Land potential defines potential resistance and resilience and represents the capacity to recover from degradation (Liebeg et al. 2017). Specifically, the land can maintain its capacity to provide ecosystem services through either resistance to change or changing rapidly to retain basic function, structure, identity, and response (Walker, Steffen, and Langridge 1999). Resistance is defined as the ability of a system to operate normally through maintenance in the event of a disturbance, while resilience is defined as the capacity to recover from disturbance (Haugum 2021).

1.2 Historical development and technological achievements for land assessment classification systems, tools, and criteria

Many experts from both the FAO and many other countries have been continuously contributing to the land assessment classification system, tools, and criteria. Many countries have developed their own methods and criteria for land assessment systems. However, the results have had different levels of success. This section presents an overview of historical development and technological achievements for land assessment classification systems, tools, and criteria.

1.2.1 Land assessment classification systems

Over the past seven decades, two types of classification systems have been widely used in land assessment from local to national levels all over the world: USDA Land Capacity Classification System (LCCS) and FAO Agro-Ecological Zoning System (AEZS).

The LCCS, initially developed by the Soil Conservation Service of the USDA in 1930, was the first widely used land assessment system, but it has now been updated in many countries. An Land Capacity Classification (LCC) "implies a characterization and a re-grouping of soil units in capacity classes for their present (actual) and future (potential) general use" (Sys, Van Ranst, and Debaveye 1991). The main content of the LCCS is to inform suitable land use to avert degradation (Herrick et al. 2016). In other words, LCC contributes to the identification of whether the land is appropriate for defined uses, and whether there are any degradation risks. Specifically, it is for classifying land by groups of soil focused on their potential for agricultural and other uses (Klingebiel and Montgomery 1961). The LCCS classifies the land into 8 classes based on a wide range of interpretations of topographic and soil factors that influence agricultural production and erosion of soil (IRP 2016). The first four (I-IV) classes are deliberated as acceptable for agronomic yields and cultivation. The next three (V-VII) classes are limited to agronomic yields and cultivation. The last (VIII) class is very restrictive, and usually cannot use any type of land resources. LCCS has been applied all over the world for land assessment systems to protect soil and natural resources for over 70 years. However, in recent years it has needed technological improvement. For instance, a digital platform for the input variables storage and management.

The AEZS for land assessment was initiated by FAO in 1978 for detecting the potential of agricultural production (FAO 1978-1981), and is based on the "Land evaluation

framework" from 1976 (FAO 1996). The AZES aimed to predict the potential yields for different crops. It is a resource of the land mapping unit, defined in terms of landform or topography, climate, soil, and other physical factors (e.g., land cover) (IRP 2016). The essential elements for classification are land resources inventory, inventory of LUT and crop, and land suitability assessment (e.g., potential maximum production, and corresponding constraints) (FAO 1978-1981). Related to the rapid development of geoinformatics technology, countries have been updating the classification system for land suitability. FAO collaborated with the International Institute for Applied Systems Analysis (IIASA) and has updated a classification of the Global Agro-Ecological Zones System (GAEZS). Since 2000, rapid advances in geo-information technology have been released three times in 2002, 2012, and 2021 (Fischer et al. 2021). The last version of GAEZS assessed sustainable potentials of the agricultural production (e.g., fodder, fiber, food, and pastureland grasses) for past, present, and future climatic conditions, obtained from big data at a spatial resolution of 5 arc minutes.

To summarize this section, both LCC and AEZ or GAEZ systems for land assessment have been broadly used at local and national scales all over the world. LCCS considers sustainable production, while AEZS is more focused on productivity. The disadvantage of both classification systems is that they do not explicitly consider sustainability. Moreover, the main limitation of AEZS (latest version) datasets still have a coarse spatial resolution of 0.5°. More recently, to develop new systems are testing that combines scientific knowledge with real data from monitoring sites. This led to reappraise system of land assessment.

1.2.2 Land assessment tools

This section presents the historical developments of tools (methods) of land assessment and achievements of the technology. Here four phases are highlighted, which are differentiated by approach and technology.

The earliest tool of land assessment is hand-overlay by hand-drawn paper maps. At that time, in order to overlay maps sun-light reflected on windows was generally used, along with transparent paper (McHarg 1996; Tyrwhitt 1950). In this phase, great progress was made with the introduction of printed photos and aerial imagery into land assessment systems. From these, aerial imagery was a powerful tool for improving land assessment at the local level (e.g., farm and watershed level) (Chrisman 1997).

The introduction of automatic tools in land assessment systems was the second phase of evolution. The first tool in computer-based Automated Land Assessment Systems (ALAS) was developed at Harvard (Hopkins 1977). This tool permitted land experts to develop their systems based on knowledge to adapt to the specifics of the locality (Collins, Steiner, and Rushman 2001). Moreover, ALAS and other land assessment systems and their user guidelines have, in the long-term served, land assessment, and are continuously developing Geographic Information Systems (GIS).

The introduction of mathematic analysis tools in the land assessment system was the third phase of evolution. In this phase, GIS-based mapping tools including "Map Algebra" and "Map Analysis Package (MAP)" were introduced in the land assessment system. However, this generated a practical need to reflect on the options of decision-makers in land assessment analysis (Goodchild et al. 1996). The main evolution of the third phase was the introduction of Boolean set theory and Fuzzy set logic theory in land assessment systems (Banai 1993). In particular, Fuzzy set theory (Zadeh 1965; Zadeh 1990; Zadeh 2007) as an extension of Boolean set theory was the main subject of the discussion of GIS-based land assessment systems such as land suitability analysis during the 1980s (Burrough et al. 1992). Various types of Multi-Criteria Evaluation (MCE) techniques based on Boolean and Fuzzy set theories have been developed for the decision-making of land assessment problems whose results depend on multiple criteria in this phase. From these, the most advanced tool was the integrated Multi-Criteria Decision Making (MCDM) method with GIS (Malczewski 1996). The main advantage of GIS-MCDM was structured for multiple criteria layers, which provides a spatially explicit assessment framework (Malczewski 2006). For estimating weights values for the multiple criteria, the analytic hierarchy process (AHP) was introduced, which is the MCDM method. The AHP is a mathematical approach for assessing values of the given functions focused on pairwise comparisons (Saaty 1980). The integration of the AHP, a Multi-Objective Decision-Making (MODM) method into GIS has further refined land assessment tools (Malczewski 1996; Malczewski and Rinner 2015).

The introduction of advanced artificial intelligence methods and data mining techniques in land assessment systems was the fourth phase of evolution. Since the early 1990s, AI techniques that can be used for some practical issues of geography have been introduced, and this was the first step of AI in land suitability analysis, known as a land assessment (Openshaw and Openshaw 1997). In the early stages of AI, the technologies were much more

experimental and costly. However, there has been growing interest in the introduction of AI devices in land assessment based-GIS. These technologies were introduced as ability advancement platforms for the process, which was a huge growth in the methodological processes. More recent improvements in land assessment systems are addressing the generation of robust models for real-time land potential using a time series of field survey reference data, cloud-based geospatial layers, and advanced machine learning methods.

To summarize this section, since the 1980s, advanced techniques of land resource assessment have come out, and they have begun to meet the new needs of digital information. In recent decades, many new powerful machine learning techniques, and cloud-based time series datasets are being developed and that is improving further land assessment tools. These tools are leading the efforts to reappraise land assessment techniques.

1.2.3 Land assessment criteria

The FAO has been using primarily physical resources including topography, climate, and soil variables for land assessment analysis (FAO 1976; FAO 1978-1981; FAO 1991; IRP 2016), and these contain possibilities for the hydrology, geology, and vegetation (Sys, VanRanst, and Debaveye 1991). USDA criteria have primarily used topography (e.g., location in the landscape, and slope) and soil (texture, depth, and other static properties of soil) variables that impact vegetation growth and erosion (USDA-NRCS 2013). Christian et al. (1953) noted land systems are defined as "repeated patterns of geology, topography, soil, and vegetation" (DES and Department of Resources 2021). Moreover, depending on the purpose of land assessment and the natural features and climate of the area, the criteria of topography, climate, soil, and vegetation were relatively different in some countries (FAO 2007). According to the literature review, the major criterion of land assessment is determined by topography, soil, and climate, and these have been widely used in land assessment from local to national levels all over the world (FAO 1976; USDA 1983; Sys, VanRanst, and Debaveye 1991; IRP 2016; USDA-NRCS 2023), whereas, the use of vegetation criteria for land assessment was generally different for all countries (IRP 2016; Dumanski 1997; FAO 1976). To better understand the contribution of physical resources to land assessment, this study focused on majority criteria such as topography, climate, soil, and vegetation variables and their interactions in the past.

Topography is the physical form and appearance of the specific land (e.g., plain, plateau, valley, depression, mountain, and hill). According to previous studies, the major

criteria of topography for land assessment were surface slope, and elevation (IRP 2016; Kumhalova et al. 2008). They allow us to determine the main attributes of the relief, which is an important physical resource. Specifically, slope and elevation have a strong effect on surface runoff through the land, which primarily affects land resilience as part of land potential (Herrick et al. 2016). For instance, water runoff through the steeper slope decreases water infiltration and transports soil nutrients, thereby limiting the amount of water used in vegetation, and slowing down the process of formation of organic material in the soil. In contrast, water runoff in flat surfaces is generally slow, and with more infiltration, and this allows enough moisture in vegetation during the long term.

Topography variables influence the microclimate in that they affect solar radiation, hydrothermal conditions (a combination of precipitation, and temperature), and wind. For instance, the steepness of the slope and the direction of the slope (aspect) affect the distribution of solar radiation. The northern side of the aspect is primarily shaded, and the southern side of the mountain receives more solar radiation, and this occurs in high-altitude regions. For this reason, mountain areas receive less radiant heat compared to flat areas, have a marked influence on vegetation growth, and affect potential land productivity. Mountains obstruct cold air from entering over the warm ground, which results in distinguished temperatures on both sides of the mountains. In other words, topography variables primarily affect water runoff and influence microclimate, which in turn influences vegetation, having an impact on soil formation. Less vegetation, on steeper slopes, affects erosion.

The main criterion controlling the potential productivity of land globally is climate. The primary climate criteria for land assessment were temperature, precipitation, and solar radiation (Sys, VanRanst, and Debaveye 1991; IRP 2016). In particular, temperature and precipitation and their distribution are the major criteria for differentiation in agro-climatic zones. Solar radiation is a natural resource that cannot be changed at any time, and it contributes to the intensity of photosynthesis. The cold temperatures of the tundra and alpine areas, and less precipitation of semi-arid regions limit vegetation growth, which in turn affects potential production of biomass, and influences soil formation and structure. Therefore, the climate is a key factor in controlling land potential productivity (Herrick et al. 2016).

The major soil criteria for land assessment were relatively static properties such as texture, and depth (IRP 2016; FAO 1976). For soil texture, clay-rich soil can store large amounts of nutrients, while loamy soils are better suited to retain available water for plants,

and sandy soils generally have higher water infiltration rates. The soil depth influences the land's direct and indirect potential productivity. Very deep soil is more resistant to soil erosion than shallow soil. The dynamic properties of soil including humus and soil moisture are significant criteria in land assessment. Soil humus is an organic matter layer that originated from the decomposition of living organisms by biological processes. It is rich in nutrients and its presence improves soil properties. The more humus in fertile soil, the better the nutrient-rich topsoil and the better the conditions for vegetation growth (Montgomery and Bikle 2021). The high content of soil moisture increases soil weathering. However, it increases nutrients in the soil and allows for greater biomass production, providing a positive feedback loop that increases land productivity.

Vegetation criteria are important for inland assessment because an area with natural vegetation stores rich nutrients. The amount of nutrients stored in vegetation directly depends on the vegetation types (Sys, VanRanst, and Debaveye 1991). However, vegetation is basically a dynamic feature, and vegetation and its productivity change as the climatic conditions change (Hengl et al. 2018). Thus, vegetation has, together with the climate, an important influence on the cycling of nutrients, affecting important soil properties that support the potential production of the land. Vegetation is useful in the reduction of land degradation (Weissmann and Shnerb 2014), and for assessing land potential (Herrick et al. 2013).

As shown, the interaction between physical resources of land has played a special role in converting world-sustaining light energy from the sun into life-sustaining nutrient energy through cycles of water, energy, and nutrients. Moreover, all topographic, climatic, and soil criteria influence water availability for the specific area.

To summarize this section, soil and climatic variables have been primarily used in the land assessment system over the seven decades. According to Young (2000) “recent years, justified criticism that soil surveys did not give the information needed for development planning at the international level” (Young 2000; McKenzie et al. 2008). As remote sensing sensors have evolved over the years, data have become available for highly informative remote measurements at different spectral, spatial, and temporal resolutions, which can be used to improve of the land assessment criteria. Therefore, achievements of the remote sensing data are leading efforts to reappraise criteria of land assessment.

1.3. The necessity for assessing land resource potential in Mongolia

From the overview of the historical development of land assessment the principles, tools, and criteria of land assessment are significantly changed at the international level in the last two decades. Recently, the integration of land productivity with environmental protection has been a major principle in the sustainable use of land resources. Therefore, globally a high priority is given to the optimum use of land resources (long-term sustainable use), and enhancing land management (conserving for the future generation). Because land resources are universal, they provide for the greater part of fiber, ~95% of human food, and all needs for fuel and building materials (Young 2000). Most of the developed countries in the world have already accountable and efficient policies, assessment systems, and management in place (McKenzie et al. 2008; Young 2000). For developing countries, there are many requirements to improve policy, assessment systems, and management (Young 2000; McKenzie et al. 2008). For instance, for Mongolia, there are several requirements to improve assessment of the potential of land resources: 1) paying attention to the long-term sustainable management, environmental conservation, and restoration of degraded land instead of increasing the use of land resources; 2) introduction of innovative techniques to accurately assess land resource potential; 3) aligning land use with land potential; 4) improving understanding by the public of the important role of land resources; 5) raising awareness of nature conservation among local populations and involving them in nature conservation activities; and 6) increasing the capacity to manage land resources (awareness of the need to improve land management).

1.3.1 The necessity for relating land resource potential concerns in Mongolia

Mongolia (1.56×10^6 km^2) is the 18^{th} largest, land-locked country in the world, located in the continental temperate zone of the northeastern hemisphere of the globe, and mainly located in the desert, with an extremely arid climate and variable topography (Figure 1.1). The climate is characterized by a long cold winter and a short dry-hot summer with low precipitation. Annual average precipitation is 20-220 mm, and it gradually increases from south to north (IRIMHE 2021). Approximately 85% of the total precipitation falls in the rainy season (June-August), and the remaining months are very dry. The annual average air temperature is around 0.00°C (coldest -36.58°C in winter, and warmest 30.62°C in summer) with strong temperature gradients (IRIMHE 2021). 65.1% of the territory of Mongolia covers drylands, of which 0.1% are hyper-arid, 31.3% arid, and 33.7% semi-arid, and it is ranked 4^{th} overall in Asia (Remus 2016). The drylands are very fragile and susceptible to global climate variability.

72.87% of the territory of Mongolia is used as agricultural land (arable land 0.45%, and pasture 72.42%) (NSOM 2022a). In other words, the majority of land resources are used for agriculture, including pasture. The main consumers of pasture are five types of livestock (sheep, goats, cattle, horses, and camels). One feature of Mongolian livestock husbandry is that they feed on rangeland plants all year round and use natural rangeland. Particularly, the major land use of drylands (related to water scarcity) is pasture with generally seasonal nomadic pastoralism.

In addition, half of the Mongolian population engages in the livestock husbandry sector. Livestock husbandry comprises ~85% of total agricultural products (MoFALI 2021). Livestock-derived product exports account for approximately 14.7% of Mongolia's foreign exchange earnings (MC 2022). The cornerstone of the Mongolian economy is land resource-based livestock pastoralism, which strongly influences the country's social and economic development. For Mongolia, livestock is renewable natural capital and is subject to state protection under the country's constitution (Densambuu et al. 2018). However, this renewable natural capital (livestock) is strongly dependent on all elements of land resources.

On the other hand, Mongolia transitioned from a socialist economy to a democratic society in 1990. Since the transition to a market economy, Mongolia has experienced significant social, economic, and environmental changes. For instance, between 1991 and 2021, the total livestock population rose by a factor of 2.63 (Figure 1.2) (NSOM 2022b), while the total population increased by 56.6% (Figure 1.3) (NSOM 2022c). Livestock population density, per hectare increased from 204 head in 1991 to 614 head in 2021, while the pasture area decreased by 11.4% (Figure 1.4). During the same period, the annual mean air temperature increased by 1.8°C (Figure 1.5), and the annual mean precipitation decreased by 74 mm (Figure 1.6). Mining as a component of the gross domestic product (GDP) rose from 9.0% in 2002 to 22.8% in 2022, while foreign exports rose from 53% to 85.0%, respectively (Byambatsogt 2017; NSOM 2022d).

Moreover, land degradation and drought frequency have increased. Specifically, according to the fifth nationwide report of land degradation and desertification assessment in Mongolia, 76.9% of the territory is affected by degradation, of which 18.6% are severely, and 4.6% are extremely degraded (IRIMHE 2020). Since 1991, in 11 years, being 1992, 1993, 1994, 1996, 1997, 2000, 2003, 2007, 2009, 2010, and 2017, partial drought during the growing season in Mongolia occurred (Munkhdulam et al. 2022).

To summarize this section, Mongolia is suffering the effects of land degradation, and desertification caused by overexploitation of pasture, mining, and climate change. The main challenge is increasing livestock pressures, which contribute significantly to land degradation under the global climate variability and dryness. The main reason for this is the mismanagement of agricultural land resources. Moreover, the previous solution relating to issues of food shortage and other needs of agriculture was through an increase of land use. Every year, there is a decrease in land, therefore, products to provide necessities for humans and livestock require increasing the potential production of the land. In other words, when the environment is deteriorating and natural resources are becoming scarcer, it is important to create sustainable land management that takes into account the interests of the present and the future and preserves the ecological balance, while adapting to climate change.

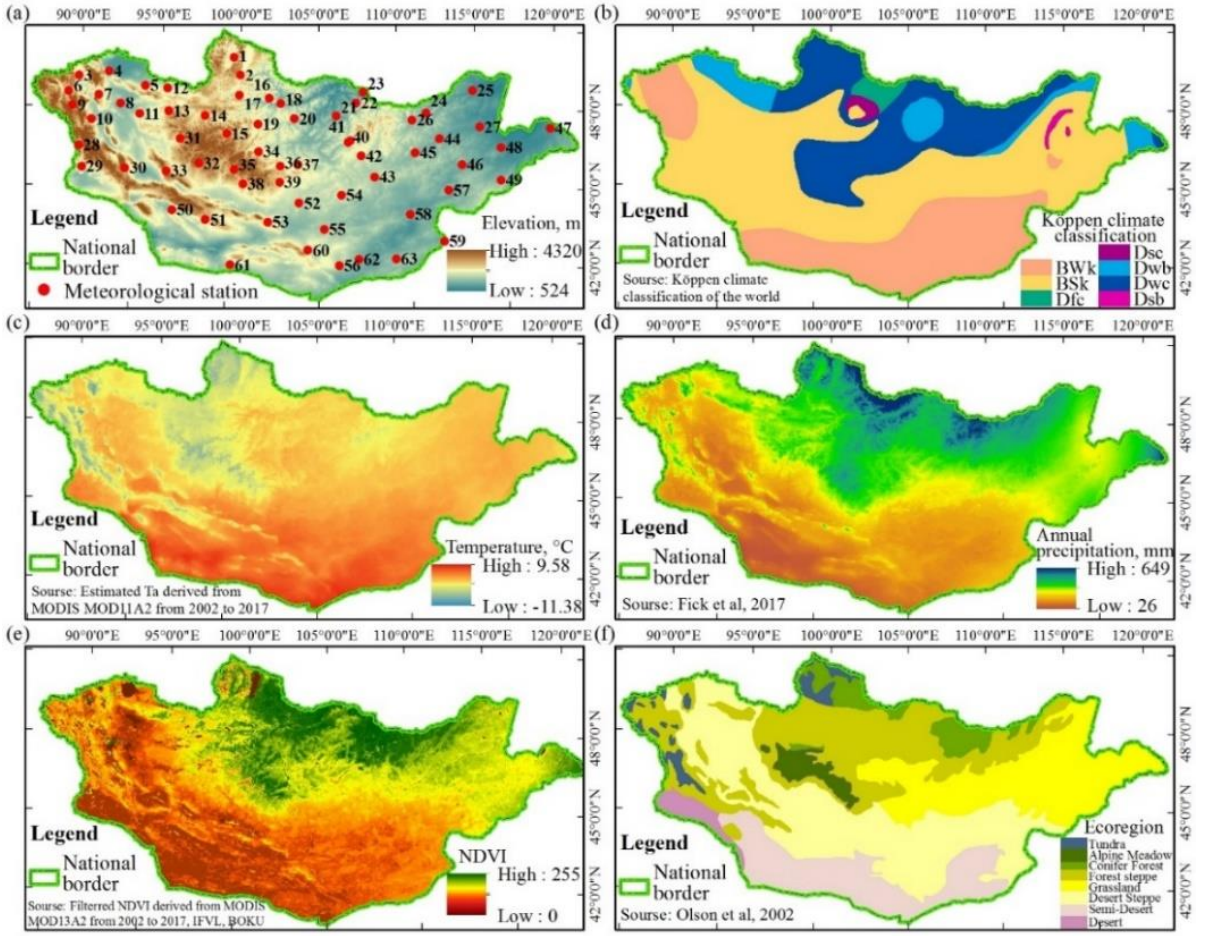


Figure 1. 1 Study area and its natural characteristics. (a) Digital Elevation Model (DEM) derived from the Shuttle Radar Topography Mission (SRTM) with automatic weather stations (n=63), (b) Köppen climate classification of the world (Kottek et al. 2006), (c) estimated annual average air temperature derived from MODIS MOD11A2 (v006) (Otgonbayar et al. 2019), (d) annual total precipitation (Fick and Hijmans 2017), (e) average annual NDVI derived from MODIS MOD13A2 (v006) for the period 2002-2017, (f) Terrestrial ecoregions of the world (Olson et al. 2001).

Therefore, robust information on land resources and their potential is needed for policy by state, regional, and local agencies, in the current situation, where the emergence of large-scale environmental concerns is possible. This robust information is a prerequisite for reasonable decisions on sustainable use and management, which will reduce the risk of mis-decision-making when the government is accountable and has efficient policies. In addition, the inherent land resource is the basis for improving the future productivity of the land. Therefore, it is timely to assess land resource potential (land resource, land suitability, and land potential) in Mongolia.

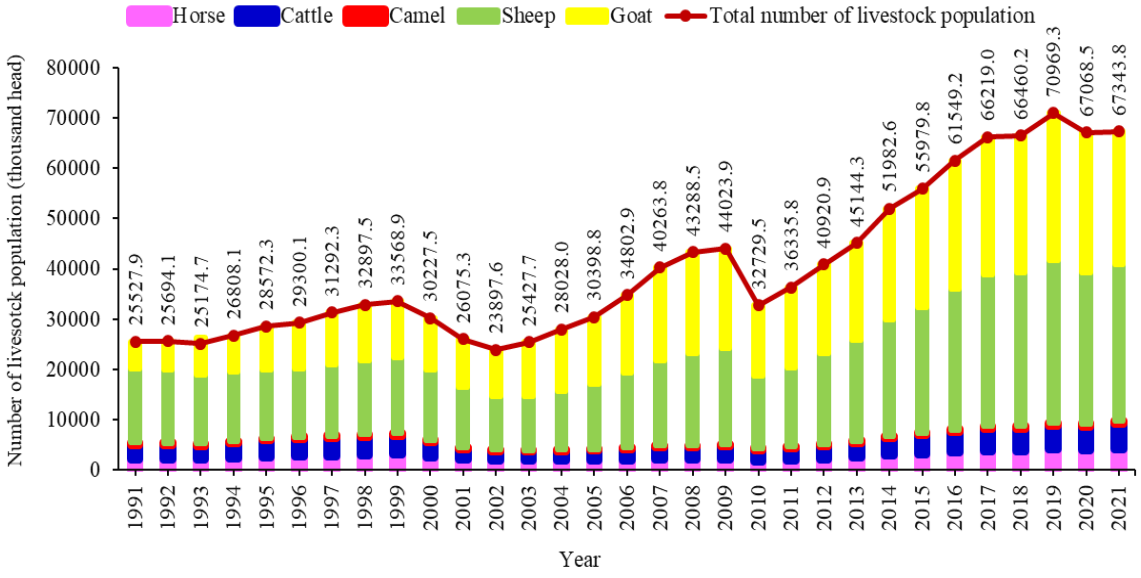


Figure 1. 2 Total livestock population for the period 1991–2021 (NSOM 2022b)

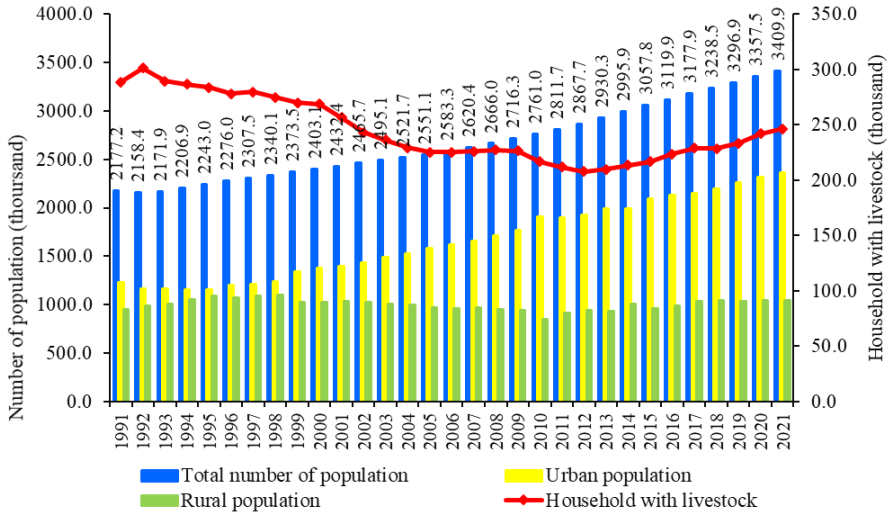


Figure 1. 3 Total population with urban and rural, and number of households with livestock for the period 1991–2021 (NSOM 2022c, 2022e).

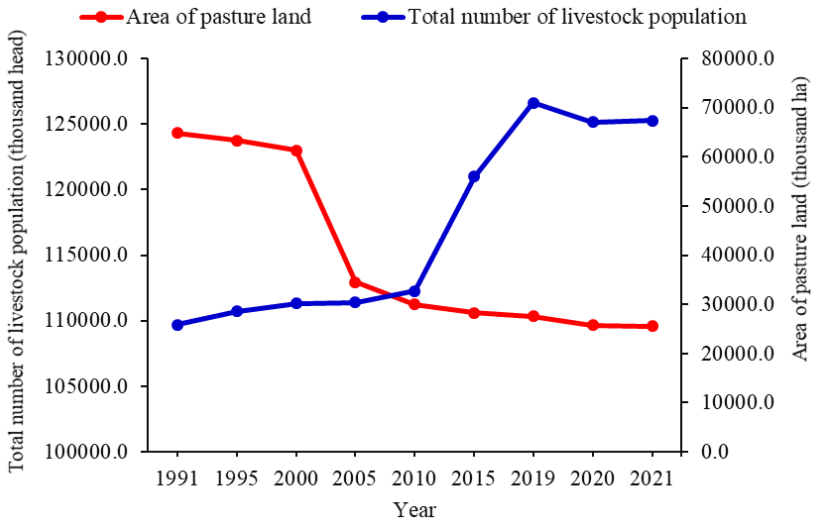


Figure 1. 4 The ratio between the area of pastureland and the number of livestock (NSOM 2022a; NSOM 2022b).

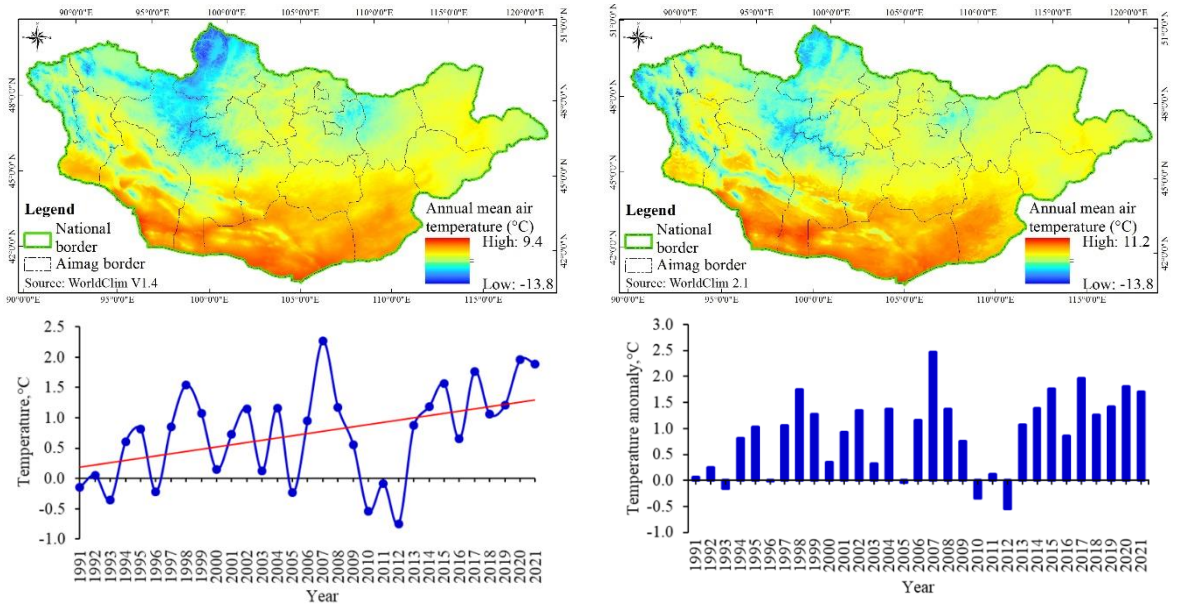


Figure 1. 5 Annual mean air temperature, ℃ (top left) for the period 1961-1990, (top right) for the period 1991-2021. Yearly average temperature (bottom left); yearly average temperature anomalies (bottom right) for the period 1991-2021.

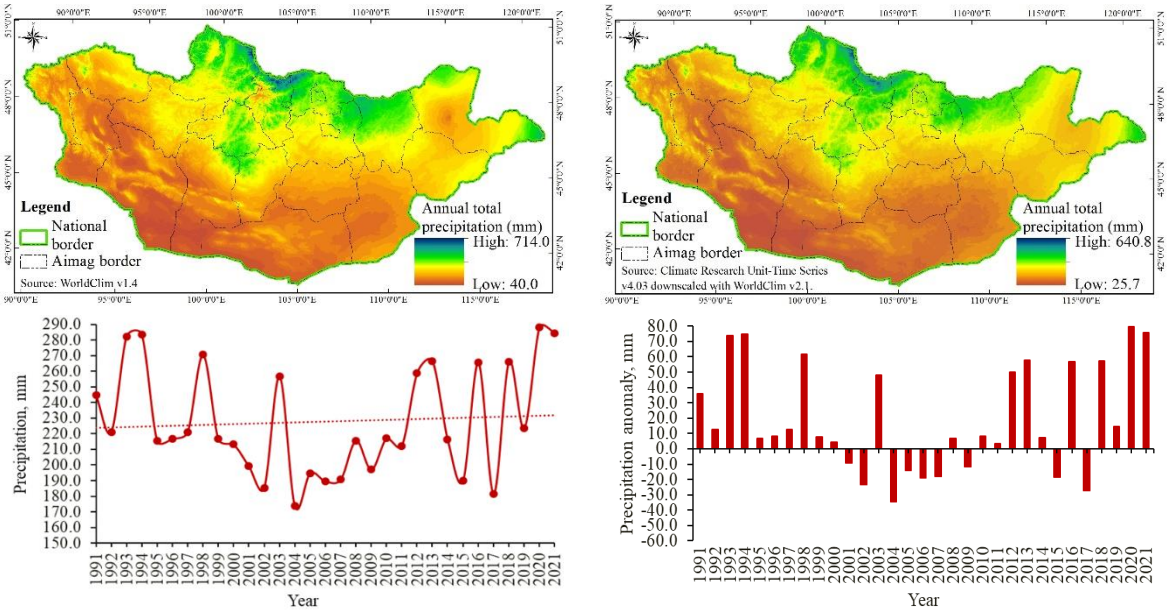


Figure 1. 6 Annual total precipitation, mm: (top left) for the period 1961-1990, (top right) for the period 1991-2021. Yearly total precipitation (bottom left); the yearly total precipitation anomalies (bottom right) were derived from CRU-TS for the period 1991-2021.

1.3.2 Previous investigations and the necessity of a new technique in Mongolia

In the 1950s, Mongolia was transformed from traditional pastoralism to a mix of industry, intensive livestock, and crop farming (MAS 1990). This was because the Mongolian People's Revolutionary Party (MPRP) approved the first "Five-Year Plan" in 1948, to develop the national economy based on central planning, with an emphasis on industrialization and collectivization of the agricultural sector (MAS 1990). Since the first plan, four additional plans were implemented in 1953-1957, 1958-1960, and between 1961 and 1985. Between 1950 and 1990, three main processes met, being industrialization, collectivization, and crop production. Specifically, the first three "Five-Year" plans were directed at developing large-scale industries to process agricultural products and natural resources for domestic demands. The 1960 constitution officially limited private ownership of livestock, abolished private holding, and transferred all production to the state. By 1963, the collectivization process was completed, and the entire agricultural sector was transferred to the state (MAS 1990).

Since the 1960s, science-based crop production was developed as an individual branch of agriculture (Gungaadash 2009). As a result of scientific and technological achievements, Mongolia overcame many technical difficulties related to geography, and climate as enormous mechanized farms were established, specializing in the production of some vegetables and grains (Neupert and Goldstein 1994). Moreover, four series of programs

entitled the “Atar campaign” were implemented for developing crop production in 1959, 1976, 2008, and 2020. At the same time, the state also paid great attention to the problem of reforming livestock production. The widespread development of veterinary services has almost completely eradicated previously widespread diseases of livestock (MAS 1990). Beginning in the 1970s, formal planning emphasized developing the mining industry. From the second half of the 1970s, the mining industry became an essential branch of the export-oriented economy (Neupert and Goldstein 1994).

In 1990, Mongolia completely transitioned to a market economy. The main reason was the rigidities imposed on the centrally planned system. After the transition, crop production decreased dramatically (the sown area decreased a factor of 3.96 from 846.1 thousand hectares in 1989 to 213.5 thousand hectares in 2000), investment in industrial almost stopped, and most of the large-scale industries collapsed (unable to obtain raw materials). At that time, the realization was that agriculture including nomadic livestock pastoralism, was the main cornerstone of Mongolia’s economy.

To summarize this section, since the 1960s, the utilization of land resources has intensively increased in Mongolia. Between 1960 and 1990, attention was paid to 2 components: the expansion of land for agriculture and the introduction of new technologies. A crucial activity of land assessment was the soil survey, which includes systematic investigation, description, and mapping. Studies covering large areas were primarily focused on landscape basis surveys, and then soil surveys for specific purposes. During this period, many advanced techniques, and technologies were introduced into the land assessment system, especially for agricultural land suitability assessment. Moreover, the knowledge of land assessment consistently rose, and there is a systematic understanding of the distribution of the topography, climate, soil, hydrology, and vegetation. Consequent to the land assessment, the thematic map of the landscape was processed in a way that emphasized the interactions of the topography, climate, soil, hydrology, and vegetation. However, information on the lack of land resource potential assessment for Mongolia, and the capacity of the institutions is weak. For instance, over the past 7 decades, five nationwide research works have been done related to the assessment of land suitability, land resources, and land potential (Table 1.2). In other words, there is a lack of research for assessing land resource potential in Mongolia. On the other hand, reappraising the suitability, resources, and potential of land originated from the “justified

criticism that soil surveys did not give the information needed for development planning at the international level" (IRP 2016; Young 2000; FAO 2007).

Table 1. 2 Nationwide research related to land suitability, land resources, and landscape potential

Period	Sources
1985	Mijiddorj, S. Comprehensive mapping of the natural resources of the Mongolian People's Republic based on satellite data. *Doctoral dissertation*
1990	Enkh-Amgalan, A. Economic assessment of the land resource of the Mongolian People's Republic to improve the territorial organization of agricultural production. *Doctoral dissertation*
1990	Jigj, S. Regional characteristics of the impact of agricultural activities on nature, the scientific basis of appropriate utilization of natural resources of the Mongolian People's Republic. *Doctoral dissertation*
2015	Munkhdulam, O., Jargaltulga, Ts., Mygmarjav, M. and Jigjidsuren, S. Assessment of land use suitability in Mongolia (cropland, pastureland, forest, urban, preserve and conserve, tourism and recreation, road and line network). *Scientific report sponsored by* Agency of Administration of Land Affair, Geodesy and Cartography of Mongolia
2020	Avirmed, E., Oyungerel, B., Renchinmyadag, T., Munkhdulam, O., Bayanjargal, B., Batnyam, Ts., Davaagatan, T., Purevsuren, M., Erdenesukh, S., Davaadorj, D., Sersmaa, J. and Dorjnamjaa, D. Landscape ecological potential of Mongolia. *Book financed by Mongolian Science and Technology Foundation*

Moreover, in the last two decades, robust information (data and knowledge) about land resource potential have been needed for policies, planning, and regulations related to sustainable land use, and management. Any decision made without reliable information is causing inefficient application of land resources, and land degradation (McKenzie et al. 2008). Therefore, to improve decision-making, management is needed to inform on the characteristics of land resources, and how they respond to disturbances. To obtain this necessary information, assessing, mapping, and modeling land resource potential is a necessity. Moreover, the assessment of land resource potential is not concerned with just physical resources (Young 2000). It represents the problem of sustainable use of land and resource management forming from the interaction between humans and the environment (IRP 2016; McKenzie et al. 2008; FAO 2007; FAO 1976).

Therefore, issues of land resource potential are sustainable use of resources and conservation for the next generation. Furthermore, the assessment of land resource potential

allows for characterizing the relatively stable properties of land (e.g., topography, climate, soil, and vegetation) in a given area, to figure out its resources, suitability, and potential. Therefore, this investigation is important to improve Mongolia's ability to conserve the environment, manage land resources, and support the implementation of priority environment programs. In the future, to avoid making costly mistakes (for instance, caused by mismanagement) and to increase the productivity of land, the assessment of land resource potential in Mongolia including land suitability, land resources, and land potential is necessary.

On the other hand, in Mongolia, two nationwide monitoring networks were established for the assessment, monitoring, and management of pasture land health: The National Agency for Meteorology, and Environment Monitoring (NAMEM) in 2011, and the Agency for Land Management, Geodesy and Cartography (ALMGaC) in 2017. 4200 of these 5716 monitoring sites are the responsibility of ALMGaC. It is possible to use these sites' data as reference data for training machine learning models. Moreover, most of the previous studies related to land assessment used small and medium-scale thematic maps. Remote sensing data provides a timely and accurate map of the land, as it allows for gathering information over large areas with regularity. Therefore, a recent investigation of land resource potential can evolve into integrated systems of multi-variable and multi-process that mix monitoring, assessment, mapping, and modeling. These changes are leading to reappraising techniques of land assessment. The assessment of land resource potential in Mongolia using new advanced techniques and time series remote sensing datasets will contribute to sustainably increasing agricultural productivity while adapting to climate variability. The results will be additionally used as baseline information for environmentally sustainable development in Mongolia.

1.4 The objective and organization of this thesis

The main objective of the research is to assess and model land resource potential throughout Mongolia using state-of-the-art machine learning techniques and remote sensing data. To achieve the objective, this study examines multiple perspectives of land resource potential such as land suitability, land resources, and land potential, focusing on developing robust models. Specifically, the investigation is to assess land resource potential in Mongolia for different purposes. Previously, only small-scale studies were done so far, with non-optimum use of EO and machine learning. This investigation seeks to find out if the land resource potential of a large country like Mongolia can be consistently assessed using machine learning techniques

and remote sensing data, which the generation of wall-to-wall maps of important indicators, such as land suitability, land resources, and land potential by the machine learning techniques. The potential of the newest generation of Earth Observation (EO) satellites is evaluated while leveraging modern machine learning techniques for information extraction. These validations were achieved in each stage. It is organized as follows.

Chapter 1 introduces the principles, historical development, and technological achievements of land assessment. Moreover, the main issues of land assessment and the necessity for assessing land resource potential in Mongolia are described.

Chapter 2 investigates land suitability for agriculture cropland in Mongolia. The main goal is to detect new crop areas with enough capacity for cultivation across the entirety of Mongolia. For the analysis, various sources of data (thematic maps and remote sensing data) and conventional research techniques were used.

Chapter 3 addressed the development of a robust methodology to assess pasture biomass across the huge land mass of Mongolia. Two models were compared: the non-linear machine learning regression, and the classical linear regression. For the analysis, single-source remote sensing data at a high resolution of 30 m, and reference data from a field survey of 553 sites were used.

Chapter 4 addressed the development of a robust statistical model to estimate climatologies of monthly average near-surface air temperature over Mongolia. The machine learning regression model was analyzed by linking a time series of remote sensing data, and reference data of 8544 points from 63 automatic weather station data.

Chapter 5 addressed the estimation of important bioclimatic, and climatic variables in Mongolia using a time series of two Earth Observation (EO) datasets. It concerns the use of the estimated bioclimatic and climatic variables in decision-making sustainable land management and environmental conservation while adapting to climate variability.

Chapter 6 addressed developing a robust methodology to assess land resource potential for agricultural land in Mongolia. Four different regression models were analyzed and linked with a time series of environmental variables, and reference data from a field survey of 12988 sites. Using the model a spatial distribution map of the land resource potential was developed, and consequently compared with current pasture use.

The final chapter provides a general conclusion based on findings from the investigation, and future research direction.

CHAPTER 2. Land suitability evaluation for cropland in Mongolia using the spatial MCDM method and GIS-based AHP

2.1 Introduction

Science-based agricultural production has been developing intensively in Mongolia since 1960 (Gungaadorj 2009). Between 1960 and 1989, the total sown area increased from 267.1 to 846.1 thousand hectares. From 1989 the total sown area fell, reaching 165.0 thousand hectares in 2006 (NSOM 2016). The sown areas rose steadily by 440.6 thousand hectares between the years 2006 and 2016. However, cropland remains 405.5 thousand hectares less than in 1989. In this same period, the total population increased 3.19 times while the amount of sown area declined by half as compared with the population growth.

Moreover, there is a significant difference in vegetable consumption between the urban and rural populations. Urban population vegetable consumption is double that of the rural population (NSC 2016). In 1960, 40.2% of the total population lived in settled areas. This increased to 66.4% by 2016. Population increase coupled with consumption increase resulted in an intensified demand for food. On the other hand, agricultural products, especially wheat and potato production, increased as a result of the national government crop development program ("ATAR III campaign") in 2008. Nowadays, potato and wheat consumption needs can be fulfilled by domestic production. However, of the total vegetable consumption (not including potatoes), 40-45% were imported (MC 2016).

The main vegetable imports (e.g., onion, garlic, cabbage, turnips, and other root seed vegetables) increased from 5438.4 tons in 1995 to 64107 tons in 2016, an increase of 11.7 times. Of these, 96-99% were imported from China. Mongolia remains strongly dependent on food security from neighboring countries. In addition, the soils of currently cultivated areas are degrading. The country is facing challenges (especially from local governments and community groups) to identify new crop areas with enough capacity for cultivation.

We previously studied this topic as "Land suitability evaluation for cropland based on GIS between 2014 and 2016", which was funded by the Mongolian Agency of Administration of Land Affairs, Geodesy and Cartography. In our preliminary study, used small and medium-scale digital thematic maps to analyze and assess land suitability for cropland. During the study, it was recognized that there was a need to improve the accuracy of input data using high-resolution satellite imagery for future research (Tsogtbayar and Otgonbayar 2016). Agricultural studies have broadly used Geographic Information System

(GIS) and Remote Sensing (RS) techniques. RS can provide a timely and accurate picture of the agricultural sector, as it is very suitable for gathering information over large areas with frequency and regularity (Atzberger 2013). The derived information is used for qualitative and quantitative analysis within near real-time production forecasts and for anticipating food security problems within the framework of monitoring agriculture.

The purpose of this study is to identify new crop areas with enough capacity for cultivation across the entirety of Mongolia. The specific objectives are as follows:

- Identify a methodology for land suitability evaluation for the cropland.
- Develop criteria parameters for land suitability evaluation for the cropland.
- Generate more accurate input data using high-resolution satellite imagery.
- Generate a map of land suitability for cropland in Mongolia

2.2 Methodology and data

2.2.1 Methodology

A combination of Boolean and Fuzzy logic theory, the spatial Multi-Criteria Decision Making (MCDM) method, Analytical Hierarchy Processes (AHP), and expert knowledge analysis were used. The general procedure for land suitability evaluation had several phases. The first phase was to define the objectives. The second phase was to select criteria, for which there are two kinds of factors and constraints (Eastman et al. 1995). The third phase was the standardization of the criteria; the fourth phase assessed the criteria ranking and weights; the fifth phase was to overlap the map layers; the sixth phase was accuracy assessment.

Creation of constraint map using Boolean logic theory

Constraints can be expressed in the form of Boolean logic (Eastman et al. 1995). Boolean logic can have only two outcomes, true (1) or false (0). A constraint factor is a discrete metric that can represent a true or false condition (Deng et al. 2014). Zero values are prohibited conditions, and 1 value is a permitted condition. Constraints in this particular study often include legal restrictions. These are current land-use policy restrictions. Condition assessments and prohibitions can be factors as well. The Boolean logic method must assume there is a definite cut-off point because there is no flexibility for assessing real uncertainty (Burrough 1992). Boolean logic cannot be used when environmental and socioeconomic factors are imprecise and incomplete. Under uncertain situations, Fuzzy (probabilistic) logic comes in handy (Prakash 2003).

Creation of factor map using spatial MCDM method

A factor is a criterion that can determine the suitability of specific outcomes for activities under consideration (Eastman et al. 1995). In this study, the spatial MCDM method was used in the creation of factor maps. Suitability levels for each of the factors were defined; these levels were used as a base to generate the factor maps (Ceballos-Silva and López-Blanco 2003). Land suitability evaluation is expressed by qualitative and quantitative parameters. This section used a combination of the spatial MCDM and the Fuzzy method. The main objective of land suitability analysis is to select the most optimal areas for a specific purpose. Land suitability analysis is a multi-criteria decision-making process (Parkash 2003). Land suitability analysis is an interdisciplinary approach that includes information from different factors such as environmental and socio-economic. A main advantage of the MCDM procedure is the decision rule relationship between the input and output map. The MCDM method is divided into 4 groups and 7 classes (Malczewski 1999): Multi-attribute and multi-objective decision-making methods based on an objective or attribute; individual and group decision-making methods based on the number of people involved in the decision-making process; decision-making under certainty and uncertainty methods based on the situation under which decision-making is being done and the nature of the criteria and spatial MCDM based on vector and raster data.

From these, multi-attribute, multi-objective, and spatial multi-criteria decision-making methods have been widely used in land-use suitability analysis. The multi-objective methods are based on mathematical programming models, and the multi-attribute methods are data-oriented (Malczewski 2004). Spatial MCDM is a process where geographical data can be combined and transformed into a decision (Parkash 2003). The main purpose of the spatial MCDM is to solve spatial decision-making problems originating from multiple criteria. The integration of spatial MCDM techniques with GIS has considerably advanced conventional map overlay approaches to land-use suitability analysis (Parkash 2003; Malczewski 2004; Carver 1991; Banai 1993; Eastman 1997; Thrill 2000). Land suitability analysis involves the incorporation of expert knowledge at various levels of decision-making. Experts, however, cannot be certain all the time, there is still uncertainty and imprecision. The MCDM method contains many different theories on how to improve the algorithm for processing imprecise or uncertain information, such as Fuzzy set theory, Elimination and Choice Expressing Reality III (ELECTRE III), Preference Ranking Organization Method for Enrichment Evaluation (PROMOTHEE), Multi-Attribute Utility Theory (MAUT), Technique for Order by Similarly

to Ideal Solution (TOPSIS) and Random Set Theory (RST). Many studies have recommended such as Eastman (1995), Sui (1994), Chen et al (2008), Dermirel, Demirel, and Kahraman (2009), Zhang and Achari (2010), Zarghami, Szidarovszky, and Ardakanian (2008), and Mosadeghi et al (2013). The fuzzy set theory technique is one of the most commonly used techniques for improving upon imprecise, incomplete, and vague information (McBratney and Odeh 1997). Fuzzy logic is like Boolean logic but fuzzier. Mathematician Lofti Zadeh presented a fuzzy set theory in 1965, illustrating a mathematically meaningful method to quantify the degree of uncertainty and imprecision of non-discrete data (Collins, Steiner, and Rushman 2001). The main point was that fuzzy data are obtained using an array of fuzzy membership functions with values that range from "0" to "1" (Zadeh 1965).

Standardization of criteria

All criteria used in the analysis were measured with different measurement values. Different values of criteria needed to be transformed into common values (Ligmann-Zielinska 2013). To implement this objective, used a criteria standardization procedure. Here used a simple linear scaling equation based on the fuzzy set method.

$$E_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where: E_i is the value of standardized in pixel i, X_{min} is the minimum value of the criteria, X_{max} is the maximum value.

Assessing ranking and weights of criteria

In land suitability analysis, there must be an evaluation that ranks the relative importance of the criteria. In this evaluation, many different factors such as geophysical, biophysical, climate, and socioeconomic were ranked. Ranked each criterion based on conclusions from the literature from professional experts. Next, came the important step of determining the weighting values for each criterion. There are many different approaches for assessing the weight of criteria based on MCDM techniques such as ELECTRE III (Joerin, Thériault, and Musy 2001), Ordered Weighted Averaging (OWA) (Malczewski 2006), Compromise Programming (CP) (Baja, Chapman, and Dragovich 2006), AHP (Saaty 1980; Wu 1998; Saaty 2008) and Fuzzy AHP (Prakash 2003; Mosadeghi et al. 2013), Sensitivity analysis (Chen, Yu, and Khan 2013) includes 3 different approaches: One-Dimensional Weights (ODW), Random Weights (RW), and Selected Weights (SW) (Pascual et al. 2010). From these, the most widely used method in spatial multi-criteria decision analysis for land suitability evaluation is the GIS-

based AHP because it calculates weight values associated with criteria maps through a pairwise comparison matrix. Moreover, the weighting values of each of the criteria can be compared against each other with an index consistency. AHP has been calculated by weighting values of the criteria, and it is expressed with the following equation.

$$W_{ij} = \frac{\sum X(ij)}{n} \quad (2)$$

Where: X_{ij}- the normalized value of a pairwise comparison matrix; n- the order of the matrix; W_{ij} – the weight of the criteria. The Consistency Ratio (CR) indicates the probability, and that the matrix ratings were randomly generated. The consistency of the pairwise comparison matrix is expressed by the consistency ratio index. When the CR exceeds 0.1 the weighting value is disagreeable, and when the index value is estimated below 0.1, the weighting value is agreeable.

$$CR = \frac{CI}{RI} \quad (3)$$

Where: CI- consistency index; RI-random index; CR- consistency ratio. Herein, calculating the consistency index was applied to the following common equation.

$$CI = \frac{\lambda_{max} - n}{n-1} \quad (4)$$

Where: CI- consistency index; λ_{max}- maximum eigenvalue, and n is the order of the matrix

Overlap of map layers

After describing the weights values of the criteria concerning their importance for land suitability analysis, all criteria maps have been overlaid using the suitability index. The formula used for calculating the suitability index of each layer was as follows:

$$S_i = \sum X_i * W_i \quad (5)$$

Where, X_i- values of each criterion; W_i - weight values of each criterion; S_i- suitability index

2.2.2 Data

In this research, various datasets were used such as satellite data of the Landsat 8, MODIS, Shuttle Radar Topographic Mission (SRTM), archive, and field survey data. Archived soil survey data obtained from the Agency of Administration of Land Affair, Geodesy and Cartography (AALAGG), biomass and climate data obtained from the Information and Research Institute of Meteorology, Hydrology, and Environment (IRIMHE), and thematic

maps of land use and soil obtained from Institute of Geography and Geoecology (IGG) of the Mongolian Academy of Sciences (MAS), and AALAGG (Table 2.1).

Table 2. 1 Details of used data

Type of data	Path/Row	Bands	Resolution, m	Date	Source
Raster data					
Landsat 8	123-143/24-31	2-7	30	Between on 1st June and 31st August, 2016	www.glovis.usgs.gov http//earthexplorer.gov
SRTM DEM	123-143/24-31	1	30, 90	Version 5.0	http//earthexplorer.gov
MODIS product	23-25/03-04 26/04	MOD13	250	Average 16 days, from 1st June to 31st August 2000-2016	www.ipdaac.usgs.gov
		MOD15	500		
		MOD17	1000		
Vector data					
Land use data					AALAGC
River Network					IGG, MAS
Soil humus, soil stone, soil pH					IGG, MAS
Distribution permafrost					IGG, MAS
Field-measured data					
Biomass data	969 sites	1 hectare	100centner/ha	1st August 2016	IRIMHE
Field survey soil data	137 sites	501 plots	1:1 cm	2013-2016	AALAGC
Meteorological data					
Mean the temperature, and total precipitation in the summer season					IRIMHE

2.2.3 Developed criteria parameters for land suitability evaluation for agricultural cropland

For the analysis, 6 main factors and 17 criteria for land suitability evaluation for agricultural cropland were developed. A criteria evaluation schema was then developed based on our own, and other countries' practices, literature, and expert knowledge (Table 2.2, and Table 2.3). The criteria evaluation was divided into two types such as multi-variables (factor) and constraint criteria parameters. A constraint is restraint criteria and it serves to limit the alternative. The constraint can also be often represented the legal restriction. That will be the decision based on the current land-use policy. It can apply for land use constraints condition assessment such as determined by the sum of factors prohibiting the use.

In this study, 9 constraints have been chosen and there are obtained range values 0 and 1. The land use constraints condition assessment is determined by the sum of factors prohibiting the use. The constraint factor assessment of land use is a true or false condition representation. Zero value is impossible, and 1 value is possible. In this study, a combination of constraint and factor analysis methods were used. There were nine constraint factors and 17 criteria factors. All constraints can be represented with values of 0 or 1. Suitability levels

between 0 and 5 were obtained for each of the factors. The levels were 5-highly suitable, 4 suitable, 3-moderately suitable, 2 unsuitable, and l-highly unsuitable (Table 2.2, and Table 2.3).

Table 2. 2 Evaluation of the multi-criteria parameters

Factor	Criteria	Unit	Standard scores for the corresponding class				
			Highly suitable (5 scores)	Suitable (4 scores)	Moderately suitable (3 scores)	Unsuitable (2 scores)	Highly unsuitable (1 score)
Topography	Slope	Degree	<3	3-6	6-9	9-12	>12
	Elevation	Meter	<1000	1000-1500	1500-2000	2000-3500	>3500
Soil	Soil humus	(%)	>4	3-3.9	2-2.9	1-1.9	<0.9
	Depth soil humus	(cm)	>20.1	15.1-20	10.1-15	5.1-10	<5
	Soil texture	-	Light clay (21-30), Sandy (10-20)	Mid-siltstone (31-45)	Sand (<10)	Heavy clay (45-60)	Clay (>60)
	pH	-	6.5-7.0	7.1-7.5 6.1-6.5	7.6-8 5.6-6.0	8.5-9 5.1-5.5	>9 <5
	Soil stone	%	<5.0	6-20	21-35	36-50	>50
	Estimated soil organic C	-	>0.50	0.35-0.50	0.25-0.35	0.15-0.25	<0.15
Vegetation	Estimated AGB	%	>75	55-75	30-55	10-30	<10
	NDVI	-	>0.50	0.35-0.50	0.25-0.35	0.15-0.25	<0.15
	LAI (MODIS)	m^2plant/ m^2ground	>60	40-60	20-40	10-20	<10
	GPP (MODIS)	Kg C/m^2	>0.035	0.02-0.035	0.01-0.02	0-0.01	3.0-3.27
Climate	Mean temperature summer season	(^{0}C)	19-22	15-19	>22	13-15	<13
	Sum of rainfall summer season	(mm)	>200.1	150.1-200	100.1-150	50.1-100	<50
	Estimated Hydro-Thermal coefficient	-	1.5-2.0	1.0-1.5	0.5-1.0	>2.0	<0.5
Hydrology	River density	km/km^2	>0.4	0.2-0.4	0.1-0.2	0.05-0.1	<0.05
	Permafrost distribution	-	Region of seasonal freezing	Sporadic& Prelatic	Island	Dis-continuous	Continuous
	Water index	-	>0.50	0.35-0.50	0.25-0.35	0.15-0.25	<0.15
Socio-economic	Distance from settlement area	Km	<100	100-200	200-400	400-800	>800
	Distance from infrastructure	Km	<50	50-100	100-300	300-500	>500
	Distance from surface water resources	Km	<3	3-6	6-12	12-15	>15

Table 2. 3 Evaluation of the constraint criteria parameters

Constraint criteria	Standard scores for the corresponding class		Requirement
	Prohibit land	Other land	
Forest land	0	1	Forest land for natural resources not be in used for cropland purposes
Urban land	0	1	Not be located in near settlement and in the settlement areas
Roads, high-voltage electricity transmission network areas	0	1	Avoid roads, high-voltage electricity transmission network areas
Cropland	0	1	Not be located cropland areas
Mining land	0	1	Not be located mining areas
Historical and cultural monuments areas	0	1	Not be located historical and cultural monuments areas (buffer zone with 500 m radius)
Archaeological sites	0	1	Not be located archaeological sites (buffer zone with 500 m radius)
River, lake, ponds	0	1	Be near to water reserve, but not in buffer zone
Springs, wells and water points	0	1	Be near to water reserve, but not in buffer zone (buffer zone with 500 m radius)

2.3. Result

2.3.1 Result of constraint factor analysis based on the Boolean logic theory

Assessment of the land use constraint conditions was determined by the sum of factors restricting land usage. The constraint factor assessment of land use is represented by a true or false condition. A “zero” value means impossible, and a “one” value means possible. We defined the forest, urban area, roads, high-voltage electricity transmission network areas, mining areas, historical and cultural monument areas, archaeological sites, rivers, lakes, springs, wells, and water points (near to water reserve, but not in the buffer zone) as completely unsuitable for cropland based on current land-use policy in Mongolia. Using the weighted linear combination method all constraint factors were combined. The analysis demonstrated a 31.2% constraint factor for the entirety of Mongolia (Figure 2.1).

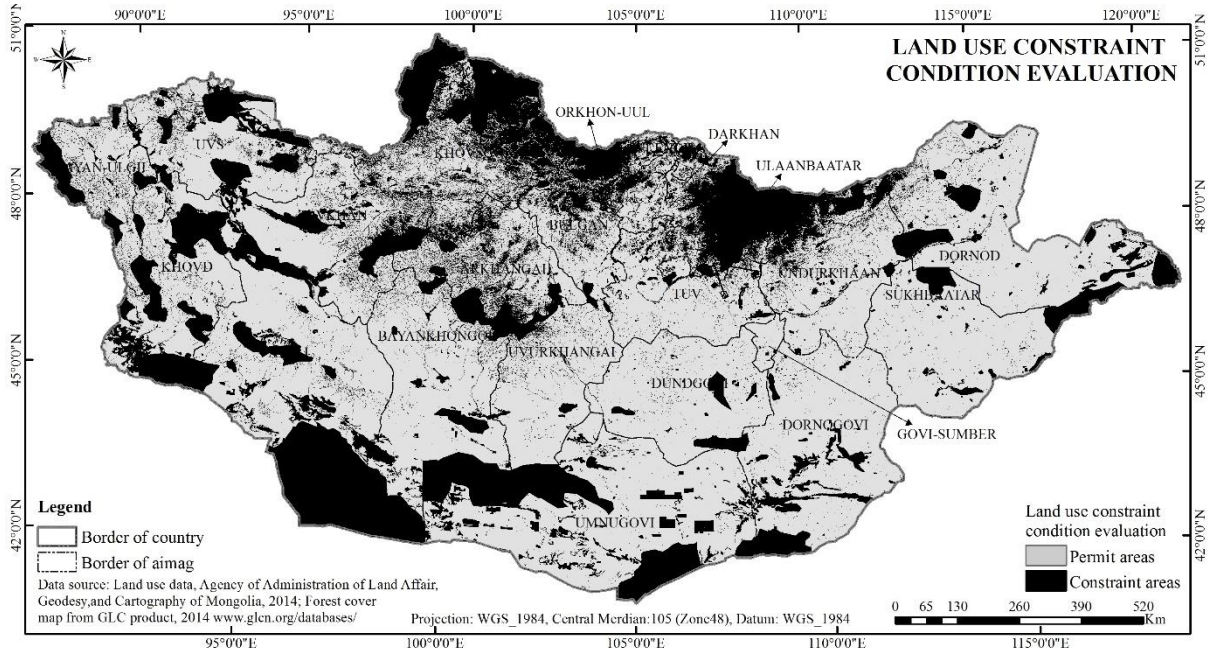


Figure 2. 1 Land use constraint condition evaluation (Boolean map method).

2.3.2 Result of factor analysis based on the spatial MCDM method

A comprehensive analysis of the study area used six major factors (topography, soil, vegetation, agro-climate, hydrology, and socio-economic) for land suitability evaluation at the primary level. There were a different number of criteria under each category totaling 22 at the secondary level (Table 2.2, column 2). In this analysis, 5 factors and 17 criteria were applied. The topography factor was important for maintaining slope stability and was critical to the distribution of other variables at a local scale (e.g., a steep terrain should not be tilled to prevent

soil erosion). Soil governed the type of vegetation that could grow most productively in a given area, and vegetation (e.g., its presence and health conditions) showed whether the land could be used productively. The agro-climatic factor was important because it affected the growth of vegetation and crops. The hydrology determined the amount of water available for plant growth. The role of these factors in the environment varied with land cover. Therefore, due to changing dominance in different areas, the same environmental factors could have dissimilar influences. Figure 2.2 shows the suitability value maps for 17 criteria, which represent the distribution of the suitability values within the study area using a continuous scale with values ranging from low to high.

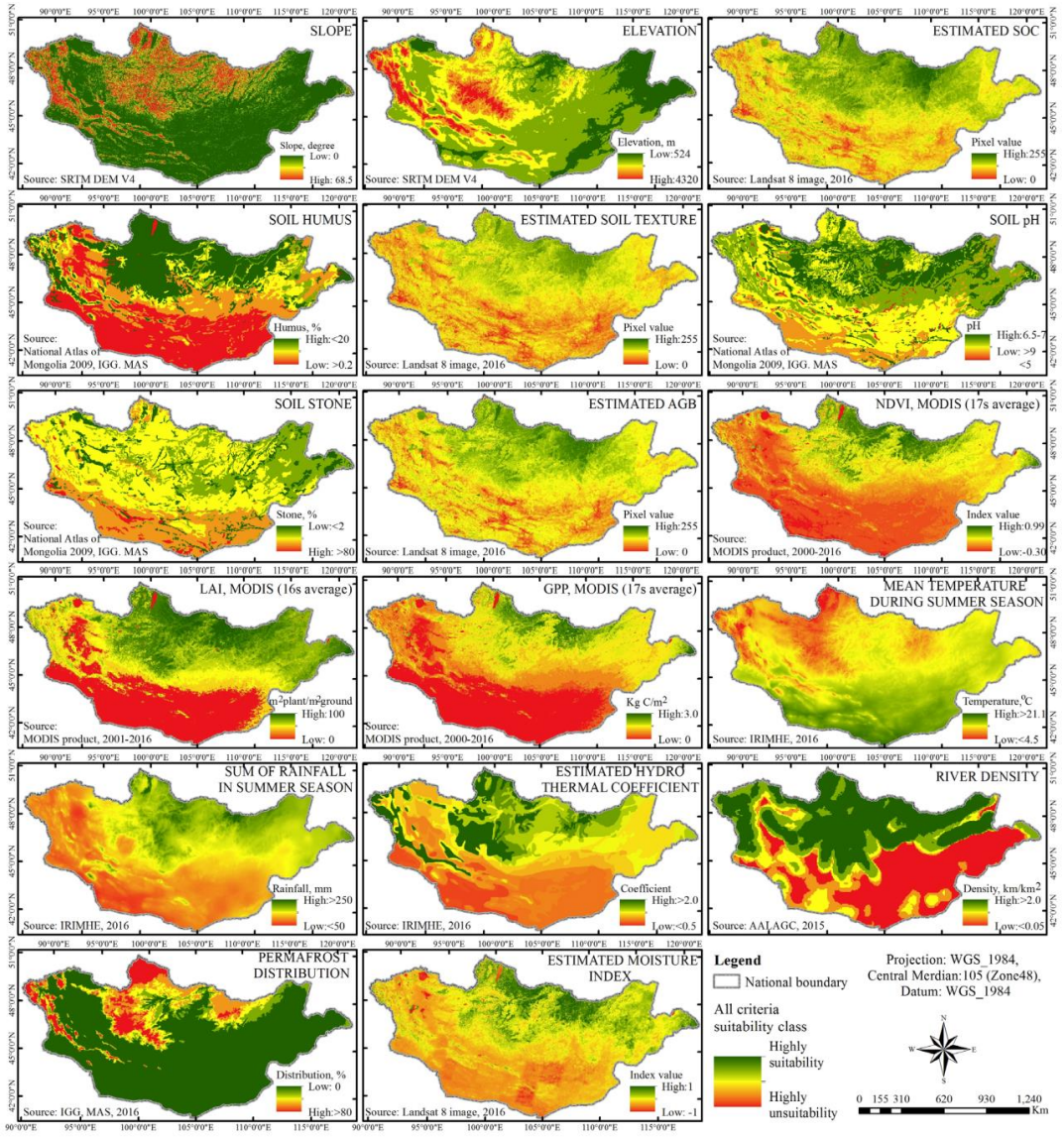


Figure 2. 2 The main factors used in cropland suitability evaluation

2.3.3 Result of ranking and weights analysis of the criteria based on the AHP

Table 2.4 shows the ranking of 17 factors based on a literature review and expert consultations, with the weights calculated using AHP based on GIS. The study has estimated a CR= 0.089, suggesting that there was a reasonable level of consistency in judgment.

Table 2. 4 Defined ranking and weights of the criteria

Factor	Weight	Criteria	Ranking	Weight	Function
Topography	0.172	Slope (S)	1	0.142	Linear
		Elevation (E)	5	0.030	Non linear
Soil	0.288	Humus (H)	1	0.142	Linear
		Estimated soil organic C (OC)	6	0.021	Linear
		Texture (T)	2	0.097	Linear
		pH (P)	7	0.014	Non linear
		Stone (SS)	7	0.014	Non linear
Vegetation	0.283	Estimated AGB (A)	1	0.142	Linear
		NDVI (N)	3	0.066	Linear
		LAI (L)	4	0.045	Linear
		GPP (G)	5	0.030	Linear
Agro-climatic	0.163	Estimated Hydro-Thermal coefficient (HT)	2	0.097	Linear
		Sum of rainfall in summer season (SR)	6	0.021	Linear
		Mean temperature in summer season (MT)	4	0.045	Linear
Hydrology	0.088	River density (R)	3	0.069	Linear
		Distribution permafrost (DP)	9	0.008	Non linear
		Estimated moisture index (M)	8	0.011	Linear
Consistency ratio (CR): 0.089					

2.3.4 Result of map layer overlay analysis based on suitability index

After weighing the importance of different criteria for land suitability analysis, seventeen criteria maps were overlaid using the suitability index.

$$S_i = 0.142*S+0.030*E+0.142*H+0.021*OC+0.097*T+0.014*P+0.014*SS+0.0142*A+0.066N+0.045 *L+0.030*G+0.097*HT+0.021*SR+0.045MT+0.069*R+0.008*DP+0.011*M$$

The results of the analysis show that 18.8% of the area studied was highly suitable, 20.2% was suitable, 19.0% was moderately suitable, 22.6% was unsuitable, and 19.3% was highly unsuitable (Figure 2.3). The results of the integrated assessment of constraint and factor analysis are shown in Figure 2.4, and Table 2.5. The integrated assessment shows that 10.1% of the area covered was highly suitable, 14.0% suitable, 15.5% moderately suitable, 16.3% unsuitable, 12.9% highly unsuitable and 31.2% was the constraint area. The results were then compared with the current extent of sown area, and the results are shown in Figure 2.5.

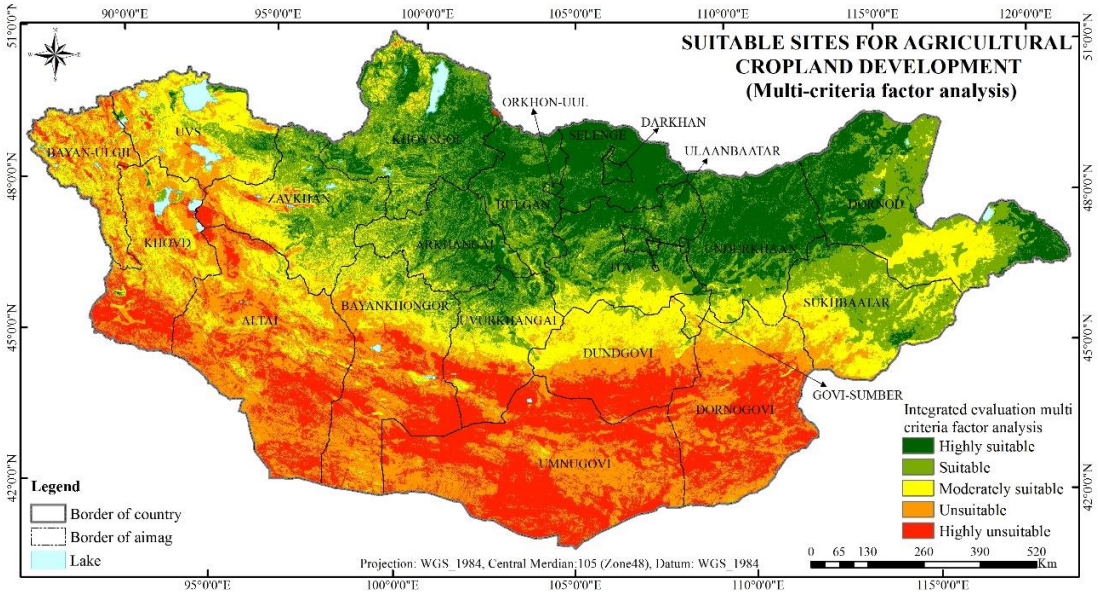


Figure 2. 3 Suitable sites for cropland development (Multi-criteria factor analysis)

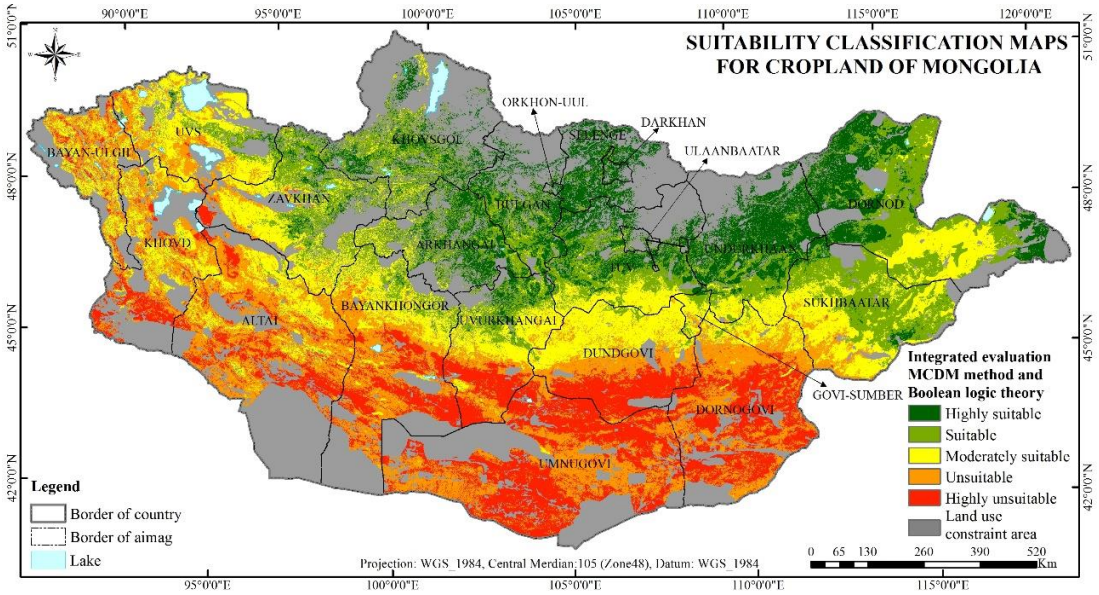


Figure 2. 4 Suitability classification map for cropland in Mongolia.

Table 2. 5 Suitability classification results for cropland in Mongolia

Suitability classification	Preliminary study result (used thematic map)		Current study result (used satellite data)	
	Area, km²	% of total area	Area, km²	% of total area
Highly suitable	83030	5.30	157707.5	10.1
Suitable	222457	14.2	219716.4	14.0
Moderately suitable	452747	28.9	243498.4	15.5
Unsuitable	249089	15.9	255180.5	16.3
Highly unsuitable	65797	4.20	201514.7	12.9
Constraint area*	482513	30.8	488982.5	31.2

Constraint area* is unsuitable based on current land-use policy.

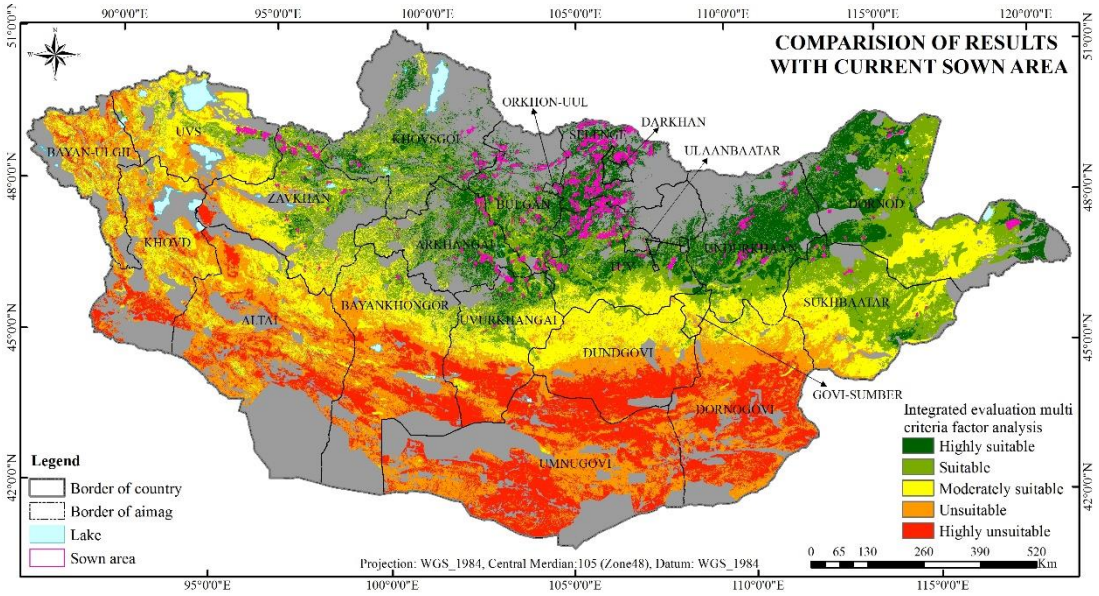


Figure 2. 5 Evaluation validation compared with the current extent of sown area

2.4 Conclusion

Since 1960, the method of wholesale selection was used for cropland areas. This was conducted based on a few parameters such as the general condition of the weather, the natural landscape, and the content of the soil. Our study, on the other hand, evaluated the extent of cropland in Mongolia, examining the results of a land suitability multi-criteria evaluation based on multiple factors such as topography, soil, vegetation, agro-climatic, hydrology, and constraints. Integrated assessment of constraint and factor analyses showed that 10.1% of the study area is highly suitable, 14.0% suitable, 15.5% moderately suitable, 16.3% unsuitable, and 12.9% highly unsuitable for cropland, with 31.2% as the constraint area. As shown in the results land suitability evaluation for cropland is possible using GIS and remote sensing technology based on a combination of multi-criteria decision output and matrix. The abovementioned method of land suitability evaluation for cropland can be used to save time for land management and it allows for the possibility of justifying policy decisions with science.

CHAPTER 3. Estimation of pasture biomass in Mongolia using Partial Least Squares (PLS) and Random Forest (RF) regression models and Landsat 8 imagery

3.1 Introduction

The rational use of pastures as a source of feed is a vital issue for the livestock industry in many countries around the world. For Mongolia, which is heavily reliant on natural pastures, it is important to use pastures sustainably, while minimizing adverse effects on predominantly fodder plants by the local carrying capacity. As a function of prevailing weather conditions, pasture biomass in Mongolia varies widely every year, complicating its management.

Pasture biomass-here defined as the amount of green mass produced in a year (Khosbayar and Narantuya 2014) - is one of the most essential environmental factors in animal husbandry, and additionally influences environmental processes such as the hydrological cycle, as well as soil erosion and degradation, especially in semi-arid areas. Consequently, biomass estimation in agroecosystems is important to understand its role in carbon exchange (Abdullah et al. 2011). Moreover, grasslands are an indispensable terrestrial ecosystem for maintaining the ecological balance of arid and semi-arid regions under global climate change (Wang et al. 2017). Considering the ecological importance of pastures, the generation of accurate pasture biomass maps is essential (Tserendash et al. 2000).

To determine livestock carrying capacity, planners must know the period for pasture usage, the area of pasture, its actual biomass, and the amount of grass consumed (Tserendash and Altanzul 2013). Of these information needs, pasture biomass is the most important. Only by accurately estimating pasture biomass and its spatial distribution is it possible to accurately identify the livestock carrying capacity in different natural zones. At the same time, such information allows decision-makers to set pasture use fees.

Given the importance of accurate biomass information, there is a great need for the development of accurate and transferable methods for biomass estimation (Eisfelder et al. 2014). Ideally, such methods should be reliable and applicable to large areas. Since the early 1980s, several biomass estimation methods have been developed based on Earth Observation (EO) data (Santos et al. 2002; Wylie et al. 2002; Ullah et al. 2012; Ramoelo et al. 2015; Liang et al. 2016; Wang et al. 2017). These various methods show benefits and drawbacks. Empirical regressions between biomass and spectral vegetation indices are the most popular and widely applied approaches (good examples are found in Cohen et al. 2003; Li et al. 2017). However, owing to their empirical nature, these regression models are site and sensor specific, and their

performance can be restricted by factors such as differences in surface properties and viewing geometry (Baret et al. 1987). Another major disadvantage of vegetation indices is that they generally use two or three spectral bands, ignoring potentially important information existing within other spectral bands (Atzberger 2013).

As remote sensing (RS) sensors have evolved over the years, data have become available from other sources, such as Light detection and ranging (LIDAR), microwave, and hyperspectral sensors. However, due to easy access and low cost, multi-spectral optical data are still the most widely used RS source to estimate pasture biomass in different parts of the world (Prince 1991; Bella et al. 2004; Numata et al. 2007). Various types of pasture biomass estimation techniques have been applied to optical RS data sets with different spatial and temporal resolutions (Lu 2006; Kumar et al. 2015; Wang et al. 2017). Dusseux et al (2015) for example, estimated grassland biomass in agricultural areas by applying the Normalized Difference Vegetation Index (NDVI) and two biophysical variables, the Leaf Area Index (LAI) and the Fraction of Vegetation Cover (FVC) on fine resolution Satellite Pour Observation Terra (SPOT) 5 images. Zandler, Brenning, and Samimi (2015) demonstrated the application of RapidEye data and Landsat 8 imagery for the investigation of total dwarf shrub biomass in arid environments. The study applied several modeling approaches such as stepwise, Least Absolute Shrinkage and Selection Operator (LASSO), Partial Least Squares (PLS), Ridge Regression (RiR), and Random Forest (RF). The best-performing method was the lasso regression model. Xie et al (2009) estimated grassland aboveground dry biomass in the Xilingol area of Inner Mongolia, China using Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), and Landsat Enhanced Thematic Mapper (ETM) imagery. Both models achieved reasonable results, but the ANN model provided a more accurate estimation than MLR. Mundava et al (2014) applied Landsat Enhanced Thematic Mapper Plus (ETM+) to test the relationship between rangeland aboveground biomass and remotely sensed indices by measuring dry and green biomass fractions, and they found that single vegetation indices were more accurate for green biomass than dry biomass. Ali et al (2017) used moderate-resolution RS images for biomass estimation by applying MLR, ANN, and an Adaptive Neuro-Fuzzy Inference System (ANFIS). The model evaluation showed that the ANFIS generated a better estimation of biomass compared to the ANN and MLR.

Machine learning methods, such as ANN and decision trees, use spectral information to minimize estimation errors through an adaptive learning process (Li et al. 2017). These

algorithms model complex non-linear relationships, which are then combined with ancillary information to find the best solutions (Verrelst et al. 2008). This has led to the widespread use of these approaches, in particular when combined with physically based canopy reflectance models (Jacquemoud and Baret 1990; Jacquemoud et al. 2000; Fan et al. 2014). However, complex parameterization and optimization procedures often prevent the application of such models (Atzberger 2004). Concerning physically-based models, the ill-posed nature of model inversion is the main limiting factor, often leading to unstable predictions as different canopy parameter combinations may yield similar spectral signatures (Combal et al. 2003; Baret and Buis 2008; Atzberger and Richter 2012; Verrelst et al. 2015; Verrelst et al. 2018; Verrelst et al. 2019).

More recently, the EO community has been exploiting RF models for biomass prediction (Wu et al. 2016; Wang et al. 2017). RF is a popular regression tree algorithm for multiple regression problems based on uncorrelated decision trees (Breiman 2001). For each decision tree, a new bootstrap sample is generated from the original data, and at each decision node, the algorithm randomly selects a subset of the predictors as candidates for splitting. To get the final regression model, the results of the individual trees are averaged.

As many spectral variables are often highly correlated with each other, methods from chemometrics such as Partial Least Square Regression (PLSR) have also been analyzed in EO (Rivera-Caicedo et al. 2017). Similar to classical Principal Component Analysis (PCA), these methods permit condensing information into a few uncorrelated latent variables. Compared to PCA, however, PLSR minimizes the co-variation to the variable(s) of interest (Maitra and Yan 2008). PLSR also reduces the negative impact of noise present in the dataset and can be further improved by removing highly correlated, as well as low-scoring, predictor variables (Li et al. 2017).

As in many other countries of the world, the amount of pasture biomass is very important for early drought detection and impact assessment (Kogan et al. 2004). Accurate biomass maps are also important for issues such as the hydrological cycle, soil erosion and degradation, the desertification process, pasture carrying capacity, pasture monitoring, and identification of pasture use fees. Pasture biomass can also be used as preliminary data for pasture management decision-making. As Mongolia has an extensive pastureland, carrying millions of livestock, the country needs an advanced technique to predict pasture biomass and must use its outcomes for its sustainable development. This research, (1) analysis PLSR and

RF models for predicting pasture biomass, to identify a suitable and robust mapping method, and (2) applies the best-performing model to generate a spatial distribution map of pasture biomass in Mongolia.

3.2 Methodology and data

3.2.1 Methodology

Partial least square (PLS) and random forest (RF) models were run in parallel on both, spectral reflectances (ρ) and spectral vegetation indices (VIs). This allows the identification of possible advantages and disadvantages of each method. It also permits confirmation expected advantages of spectral indices over simple reflectance values. The overall workflow is depicted in Figure 3.1. PLS regression (PLSR) was chosen because it can be used to predict or analyze a set of dependent variables from a set of independent variables or predictors (Abdi 2003). PLSR is a good alternative to the more classical MLR and principal component regression (PCR) methods because it is more robust and less susceptible to data redundancy and overfitting (Geladi and Kowalski 1986). PLSR was originally developed in chemometrics dealing with a high number of highly inter-correlating predictor variables (Wold et al. 1984; Wold et al. 2001). The method is now widely used by the remote sensing community, both for vegetation analysis (Atzberger et al. 2010; Inoue et al. 2012; Laurin et al. 2014) and soil-related studies (Rossel et al. 2006; Farifteh et al. 2007; Gomez et al. 2008). RF regression was chosen as it is capable of synthesizing regression or classification functions based on discrete or continuous data sets (Pal 2005; Mutanga, Adam, and Cho 2012). The main advantages of RF are (Immitzer et al. 2016; Hudak et al. 2008; Belgium and Dragut 2016):

- Variable distributions need not be unimodal or even normally distributed;
- high-dimensional and highly correlated data sets can be processed efficiently;
- over-fitting of the models is prevented;
- performance measures can be computed using only out-of-bag (OOB) data;
- information on the importance of each input variable for the model is provided.

The fact that the model performance can be computed along the way using OOB samples is a very appealing feature of RF (Genuer et al. 2010; Rodriguez-Galiano et al. 2012). This is possible, as not all observations are included in the respective bootstraps of the individual trees. The main disadvantage of RF models is that the averaging of the single trees tends to lead to an overestimation of small values and an underestimation of high values (Horning 2010;

Vanselow and Samimi 2014). Due to the underlying decision trees, it is also not possible for RF to predict beyond the range of the input data. Once trained, RF and PLS models were assessed against field survey biomass data. In a final step, calibrated models were applied to the entire study area to generate high-resolution biomass maps of Mongolia.

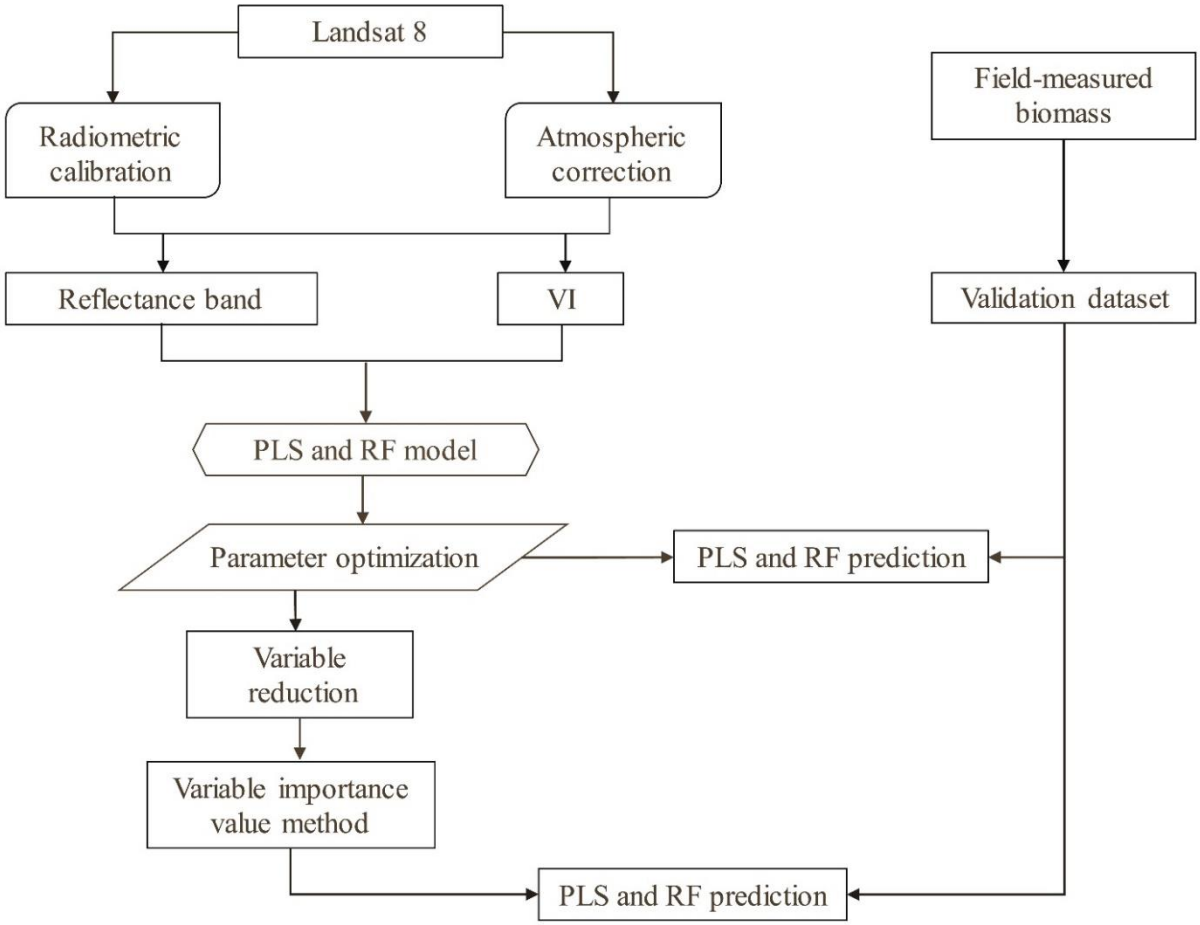


Figure 3. 1 General workflow for modeling pasture biomass using spectral predictors from Landsat 8 and for generating biomass maps

Statistics

The two widely used statistics, the coefficient of determination (R^2) in equation (1) and the root-mean-square error (RMSE) in equation (2), have been calculated to assess the quality of the models (Richter et al. 2012).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(V_{\text{est}}^{i} - \hat{V}_{\text{est}})^2}{\sum_{i=1}^{n}(V_{\text{est}}^{i} - \bar{V}_{\text{est}})^2} \quad (1)$$

Where V_{est}^{i} is the estimated variables, $\hat{V}_{\text{est}}$ is the average of the estimated variables, and $\bar{V}_{\text{est}}$ is the average of the predicted variables.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(V_{\mathrm{est}}^{i} - V_{\mathrm{obs}}^{i})^2} \qquad (2)$$

Where V_{obs}^{i} is the observed variables, and *n* is the number of observed variables.

3.2.2 Data

To develop a robust methodology to estimate pastureland biomass from Landsat 8 data, field-measured biomass samples were analyzed together with spectral information derived from Landsat 8 (Figure 3.2 (a)). The various datasets were acquired during the summer of 2016 (August) and are specified in Table 3.1. Sampling points are shown in Figure 3.2 (b). The mapping was done using Landsat 8 data. Moderate-resolution imaging spectroradiometer (MODIS) data was additionally used as a means to check and validate the resulting maps.

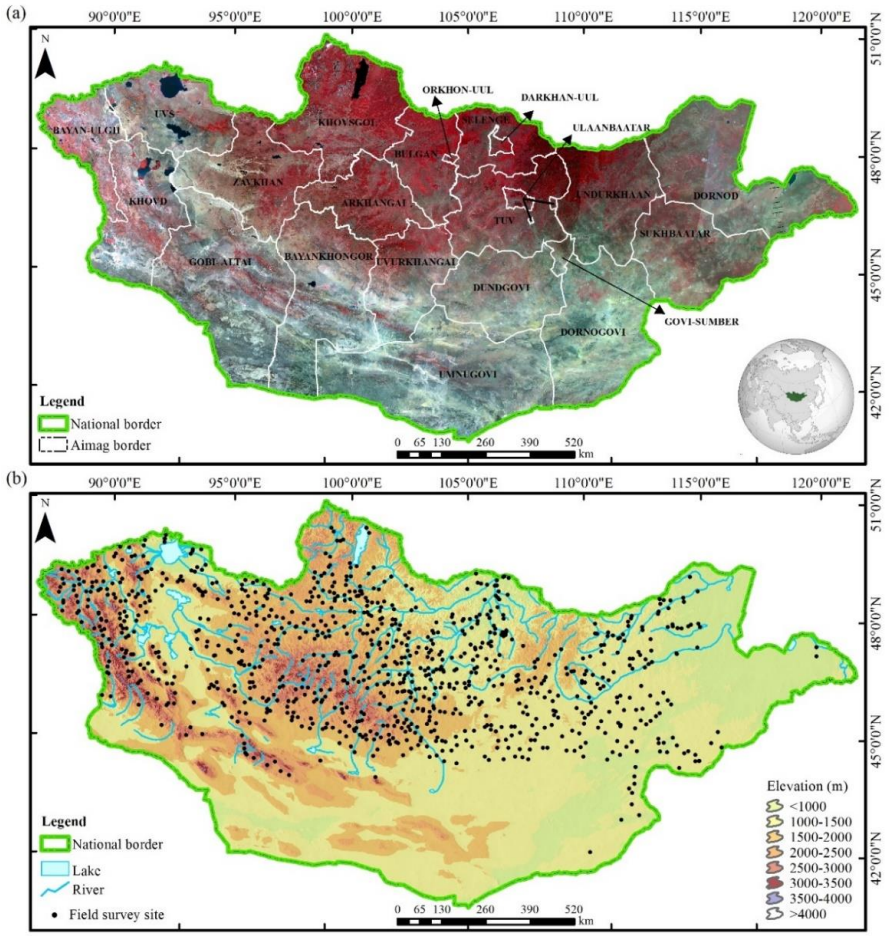


Figure 3. 2 (a) Base map Landsat 8 imagery from June to September 2016 with false color composite (Red: Band 5; Green: Band 4; Blue: Band 3), and Mongolia's 21 administrative units. Figure 3.2 (b) Sample points (n = 553) for which reference biomass information was available. Projection system world geographic system (WGS) 1984, central meridian 105 (zone 48), datum WGS1984 in Figure 3.2 (a) – (b).

Field data was made available by the Information and Research Institute of Meteorology, Hydrology, and Environment (IRIMHE). In Mongolia, IRIMHE is responsible for nationwide rangeland monitoring covering 1450 monitoring plots representing all baghs in Mongolia. Meteorology technicians in 320 soums collect the primary data yearly at 1450 plots using a new (standardized) methodology since 2011. The newly standardized methodology includes foliar canopy cover, core species composition, basal gaps of perennial plants, plant height, and biomass. Measurement methods include line-point intercept, gap intercept, air-dry biomass at 1 cm clipping height, and photo points (GGP-SDC 2015).

Table 3. 1 Data characteristics and data sources. The upper section refers to the raster data, while the lower section lists the field data. Note that the mapping was done using Landsat 8 data.

Raster data							
Type of data	Band	Resolution (m)	Path/Row	Acquisition date	Path/Row	Acquisition date	Source
Landsat 8	2 - 7	30	123/027	23 July 2016	134/025-030	24 June 2016	www.glovis.usgs.gov; http//earthexplorer.gov
			124/027-028	25 June 2016	134/025-026	19 Aug 2015	
			124/027-028	12 July 2015	135/025	28 Aug 2016	
			125/027-028	21 July 2016	135/026-029	12 Aug 2016	
			126/025-028	13 Aug 2016	135/030	09 June 2016	
			127/025-030	08 Aug 2016	136/024-030	02 July 2016	
			128/025-030	27 Aug 2016	136/024-030	30 June 2015	
			129/026-030	02 Aug 2016	136/029-030	01 Aug 2015	
			129/030	02 Aug 2016	137/024-030	10 Aug 2016	
			129/031	15 June 2016	137/024-025	08 Aug 2015	
			130/026-030	25Aug 2016	137/026-027	23 Aug 2015	
			130/031	09 Aug 2016	138/024-029	18 Sept 2016	
			131/025-031	31 July 2016	138/024-029	14 June 2016	
			131/025	15 July 2016	138/029	02 Sept 2016	
			131/025-027	14 Aug 2015	139/025-029	02 Sept 2016	
			132/025-031	07 Aug 2016	139/025-029	23 July 2016	
			132/025-031	20 June 2016	139/027-029	24 Aug 2016	
			132/025-028	11 Aug 2015	140/025-029	31 Aug 2016	
			132/031	23 Aug 2016	140/025-029	28 July 2015	
			133/025-031	08 Aug 2016	141/026-028	05 June 2016	
			133/025-026	11 July 2015	141/026-028	04 Aug 2015	
			133/025-031	07 Aug 2016	142/025-027	13 Aug 2015	
			133/025-028	11 July 2015	143/026	01 Aug 2016	

Type of data	Product	Resolution (m)	Horizontal and vertical title number	Acquisition date	Source
MODIS	MOD17	1000	h23v03, h23v04 h24v03, h24v04 h25v03, h25v04 h26v04	16-day composites, from 1st June to 31st August 2000-2016	www.lpdaac.usgs.gov

Field data					
Type of data	Number of sites	Sampling size (ha)	Unit	Date	Source
Biomass data	553	1	kg ha^{-1}	August 2016	IRIMHE

Field survey

From the IRIMHE field survey, a total of 553 biomass samples were available. The frequency distribution of the field-measured biomass for the 553 sites is shown in Figure 3.3. Some descriptive statistics for the reference data set are reported in Table 3.2. Biomass from each plot at each field was collected, sealed in plastic bags, sent to a meteorological station, and

plotted for analysis. In the laboratory, each field-measured biomass was dried, and the dry weight was calculated. The dry weight was divided by the surface area of the plot, and then the weight was converted to kg ha^{-1}. At the beginning of August 2016, when the data were collected, the value of field-measured biomass ranged from 20 to 1000 kg per hectare with a mean biomass value of 257 kg ha^{-1} and a standard deviation of 208 kg ha^{-1}. The left skew of the distribution is noted.

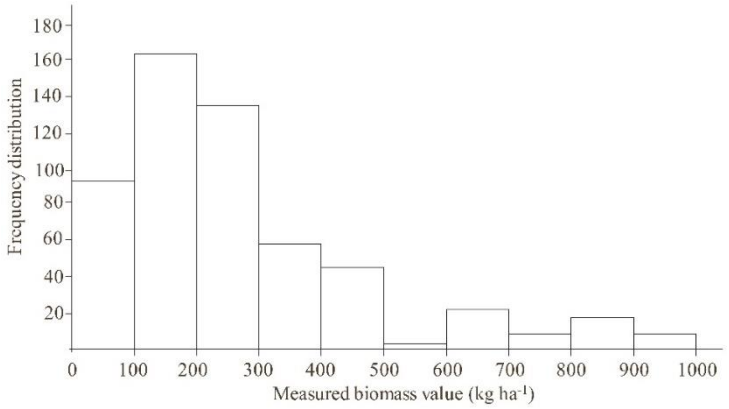


Figure 3. 3 Frequency distribution of field-measured biomass samples available for the study and distributed over Mongolia (n = 553)

Table 3. 2 Statistical descriptors (number, minimum, maximum, mean, and standard deviation) of the field-measured (reference) biomass samples as well as the seventeen analyzed spectral indicators and six spectral bands. For the abbreviations of spectral indices, the reader is referred to in Table 3.2.

Variables	No. samples	Min	Max	Mean	Std. deviation
CL_{green}	553	0.20	3.57	1.08	0.74
SR	553	1.20	4.57	2.08	0.74
WDRVI	553	-0.46	0.15	-0.25	0.15
NDVI	553	0.09	0.64	0.32	0.13
EVI_1	553	0.02	0.27	0.11	0.05
EVI_2	553	0.04	0.38	0.18	0.07
$MSAVI_2$	553	0.03	0.36	0.17	0.06
SOC	553	1.72	1.73	1.74	0.00
$NDVI_{green}$	553	0.26	0.67	0.47	0.10
$WDRVI_{green}$	553	-0.09	0.03	-0.02	0.03
NDWI	553	-0.33	0.11	-0.14	0.07
CI	553	-0.04	0.31	0.18	0.06
GSI	553	0.05	0.31	0.22	0.05
HI	553	0.03	1.10	0.69	0.23
RI	553	1.08	4.56	3.02	0.80
MSI	553	0.79	1.99	1.34	0.20
BI	553	0.05	0.14	0.09	0.01
B	553	0.034	0.120	0.060	0.017
G	553	0.041	0.167	0.085	0.026
R	553	0.053	0.229	0.124	0.042
NIR	553	0.115	0.324	0.233	0.037
$SWIR_1$	553	0.183	0.417	0.310	0.048
$SWIR_2$	553	0.111	0.359	0.245	0.053
Biomass	553	0.20	10.00	2.57	0.20

Remote Sensing data

In total, 104 scenes from Landsat 8 satellite were analyzed and processed. Data were downloaded from (USGS Earth Explorer 2018), focussing on the period from June to August 2016 (see Table 3.1). Data were available at a spatial resolution of 30 m and covering six spectral bands in the visible, near, and shortwave infrared. Before analysis, the Landsat 8 imagery was radiometrically and atmospherically corrected. A radiometric correction was implemented in ENVI (Environment for Visualizing Images) 5.0 software with the radiometric calibration module. Atmospheric correction was implemented with the quantum geographic information system (QGIS) 2.18.2 Semi-automatic Classification Plugin (SCP), parameterized with a mid-latitude summer, a rural aerosol model, no aerosol retrieval, and 40 km visibility.

Due to different orbits, sensor overpass times, and locally changing atmospheric conditions, the primary difficulty was to harmonize the dataset into a homogeneous image mosaic. The mosaic preprocesses (MOSPREP) algorithm with bundle color balancing method was used for this purpose (PCI Geomatica 2015a). The bundle color balancing method applies a global adjustment of the mean and standard deviation (sigma) of each image using a "block-bundle" method between it and each of its overlapping images, and then using “dodging points” to make smaller local adjustments between pairs of images once they have been mosaicked (PCI Geomatica 2015b; 2017). MODIS data (downloaded from Land processes distributed active archive center 2018) were used to check the plausibility of the final results. Details about this dataset are reported in Table 3.1.

3.2.3 Predictor variables and model development

A large number of well-known vegetation indices were used to estimate vegetation biomass using Landsat 8 imagery. The indices and their formula are listed in Table 3.3. Both model types were also run using spectral reflectances (ρ) as inputs. Table 3.4 and Table 3.5 highlight the high inter-correlation between the various VIs, respectively, the spectral reflectances. For both models (PLSR and RF regression), all 17 indices and all 6 spectral bands were used simultaneously to predict field-measured biomass from Landsat 8 data.

Table 3. 3 Spectral indices used in this study. G-green wavelength, B-blue wavelength, R-red wavelength, NIR-Near-Infrared wavelength, SWIR- Short Wavelength Infrared, α a value of 0.3, a, b, c and d are coefficient where a = 1.7149, b = - 0.0157, c = 0.01281, d = - 0.0113.

Spectral index	Abbr	Formula	Reference
Green Normalized Difference Vegetation Index	$NDVI_{green}$	$\frac{(NIR - G)}{(NIR + G)}$	(Gitelson et al. 1996)
Simple Ratio	SR	$\frac{NIR}{R}$	(Jordan 1969)
Green Chlorophyll Index	CL_{green}	$\frac{NIR}{R} - 1$	(Gitelson et al. 2005)
Normalized Difference Vegetation Index	NDVI	$\frac{(NIR - R)}{(NIR + R)}$	(Rouse et al. 1974)
Enhanced Vegetation Index 1	EVI_1	$2.5 \times \frac{(NIR - R)}{(1 + NIR + 6 \times R - 7.5 \times B)}$	(Liu and Huete 1995)
Enhanced Vegetation Index 2	EVI_2	$2.5 \times \frac{(NIR - R)}{(1 + NIR + 2.4 \times R)}$	(Jiang et al. 2008)
Wide Dynamic Range Vegetation Index	WDRVI	$\frac{(\alpha \times NIR - R)}{(\alpha \times NIR + R)}$	(Gitelson 2004)
Green Wide Dynamic Range Vegetation Index	$WDRVI_{green}$	$\frac{(\alpha \times NIR - G)}{(\alpha \times NIR + G) + \frac{(1 - \alpha)}{(1 + \alpha)}}$	(Gitelson et al. 2012)
Modified Soil Adjusted Vegetation Index 2	$MSAVI_2$	$\frac{NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - R)}}{2}$	(Qi et al. 1994)
Colorations Index	CI	$\frac{R - G}{R + G}$	(Gallagher et al. 2004)
Hue Index	HI	$\frac{2 \times R - G - B}{G + B}$	(Tse-Wei et al. 2008)
Brightness Index	BI	$\frac{\sqrt{G^2 + R^2 + NIR^2}}{3}$	(Escadafal and Bacha 1996)
Redness Index	RI	$\frac{R^2}{B + G}$	(Huete and Escadafal 1991)
Top Grain Size Index	GSI	$\frac{NIR - B}{NIR + B + G}$	(Xiao et al. 2006)
Normalized Difference Water Index	NDWI	$\frac{NIR - SWIR}{NIR + SWIR}$	(Gao 1996)
Moisture Stress Index	MSI	$\frac{SWIR}{NIR}$	(Datt and Ravallion. 1990)
Soil Organic Carbon	SOC	$EXP\ a + b \times R + c \times G + d \times B$	(Chen et al. 2000)

Table 3. 4 Inter-correlations between spectral indices (n = 553). In the lower triangle, spectral indices with significant inter-correlation (e.g., r ≥ 0.90) are highlighted (in gray). For the abbreviations, see Table 3.3.

Indices	CL_{green}	SR	WDRVI	NDVI	EV_1	EV_2	$MSAVI_2$	SOC	$NDVI_{green}$	$WDRVI_{green}$	NDWI	CI	GSI	HI	RI	MSI	BI
CL_{green}	/	1.00	0.99	0.97	0.97	0.94	0.94	0.81	0.84	0.83	0.65	-0.65	-0.69	-0.66	-0.64	-0.61	-0.35
SR	1.00	/	0.99	0.97	0.97	0.94	0.94	0.81	0.84	0.83	0.65	-0.65	-0.69	-0.66	-0.64	-0.61	-0.35
WDRVI	0.99	0.99	/	1.00	0.99	0.97	0.96	0.83	0.89	0.88	0.63	-0.60	-0.66	-0.63	-0.61	-0.60	-0.38
NDVI	0.97	0.97	1.00	/	0.99	0.98	0.97	0.84	0.92	0.90	0.61	-0.56	-0.64	-0.60	-0.59	-0.59	-0.38
EV_1	0.97	0.97	0.99	0.99	/	0.99	0.99	0.78	0.91	0.89	0.67	-0.54	-0.58	-0.56	-0.54	-0.65	-0.28
EV_2	0.94	0.94	0.97	0.98	0.99	/	1.00	0.72	0.90	0.86	0.70	-0.54	-0.58	-0.56	-0.54	-0.68	-0.19
$MSAVI_2$	0.94	0.94	0.96	0.97	0.99	1.00	/	0.70	0.89	0.85	0.71	-0.54	-0.57	-0.55	-0.53	-0.69	-0.17
SOC	0.81	0.81	0.83	0.84	0.78	0.72	0.70	/	0.75	0.83	0.30	-0.50	-0.59	-0.55	-0.55	-0.25	-0.80
$NDVI_{green}$	0.84	0.84	0.89	0.92	0.91	0.90	0.89	0.75	/	0.98	0.50	-0.19	-0.32	-0.25	-0.24	-0.49	-0.38
$WDRVI_{green}$	0.83	0.83	0.88	0.90	0.89	0.86	0.85	0.83	0.98	/	0.44	-0.19	-0.33	-0.25	-0.25	-0.41	-0.52
NDWI	0.65	0.65	0.63	0.61	0.67	0.70	0.71	0.30	0.50	0.44	/	-0.50	-0.36	-0.39	-0.35	-0.99	0.21
CI	-0.65	-0.65	-0.60	-0.56	-0.54	-0.54	-0.54	-0.50	-0.19	-0.19	-0.50	/	0.89	0.96	0.92	0.45	0.13
GSI	-0.69	-0.69	-0.66	-0.64	-0.58	-0.58	-0.57	-0.59	-0.32	-0.33	-0.36	0.89	/	0.98	0.98	0.31	0.32
HI	-0.66	-0.66	-0.63	-0.60	-0.56	-0.56	-0.55	-0.55	-0.25	-0.25	-0.39	0.96	0.98	/	0.99	0.35	0.25
RI	-0.64	-0.64	-0.61	-0.59	-0.54	-0.54	-0.53	-0.55	-0.24	-0.25	-0.35	0.92	0.98	0.99	/	0.31	0.27
MSI	-0.61	-0.61	-0.60	-0.59	-0.65	-0.68	-0.69	-0.25	-0.49	-0.41	-0.99	0.45	0.31	0.35	0.31	/	-0.26
BI	-0.35	-0.35	-0.38	-0.38	-0.28	-0.19	-0.17	-0.80	-0.38	-0.52	0.21	0.13	0.32	0.25	0.27	-0.26	/

Table 3. 5 Inter-correlation between Landsat 8 reflectances (n = 553). High inter-correlations (r ≥ 0.90) are indicated in gray in the lower triangle

Variables	B	G	R	NIR	$SWIR_1$	$SWIR_2$
B	/	0.967	0.905	0.596	0.707	0.784
G	0.967	/	0.973	0.641	0.754	0.842
R	0.905	0.973	/	0.581	0.770	0.883
NIR	0.596	0.641	0.581	/	0.731	0.585
$SWIR_1$	0.707	0.754	0.770	0.731	/	0.937
$SWIR_2$	0.784	0.842	0.883	0.585	0.937	/

3.3 Result

The coefficient of determination (R^2) of the four models combining two model types (PLSR and RF) and two spectral inputs (VI and ρ) ranged between 0.689 and 0.767 (Table 3.6). The corresponding RMSE was in the range of 102.70 kg ha^{-1} - 103.40 kg ha^{-1}. Overall, the RF model combined with spectral indices gave the highest accuracies. With the RF biomass model, R^2 of 0.764 and RMSE of 98.00 kg ha^{-1} could be obtained. Results for PLSR were similar but with slightly lower R^2. In general, models involving spectral reflectances as predictor variables scored lower than models using spectral indices. While the difference between VIs and ρ was relatively small for RF, it was slightly higher for PLSR (Table 3.6).

Table 3. 6 Summary statistics (R^2 and RMSE) for multi-variate biomass prediction models (n = 553) involving spectral reflectances (ρ), respectively, spectral VI.

	PLSR		RF	
	VI	ρ	VI	ρ
R^2	0.749	0.689	0.764	0.740
RMSE	1.011	1.034	0.980	1.027
n	553	553	553	553

Concerning the variable importance, the top eight spectral indices (VI) were identical for PLSR and RF regression models (Figure 3.4), with only slight variations in the respective rankings. The eight most important VIs for both model types were green chlorophyll index (CL_{green}), simple ratio (SR), wide dynamic range vegetation index (WDRVI), normalized difference vegetation index (NDVI), soil organic carbon concentration (SOC), enhanced vegetation index (EVI_1), modified soil adjusted vegetation index 2 ($MSAVI_2$) and enhanced vegetation index 2 (EVI_2). As the differences between PLSR and RF were only marginal (Table 3.6), used the computationally simpler PLS model to produce a biomass map of Mongolia (Figure 3.5).

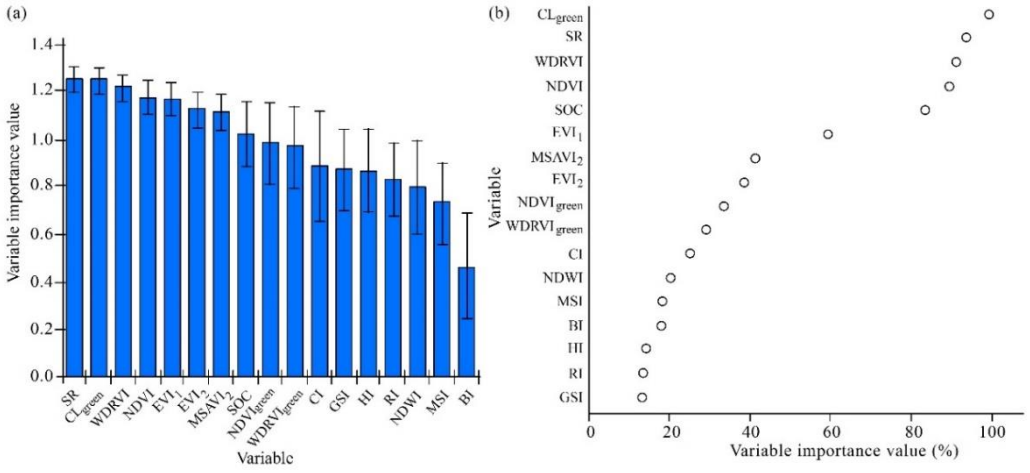


Figure 3. 4 Importance of spectral vegetation indices (VIs) for (a) PLS model, and (b) RF regression model. It is noted that the two model types use different methods to quantify the variable importance.

The model is summarized in equation (3), and the scatterplot between the reference and modeled biomass is shown in Figure 3.6.

$$\mathrm{BM} = -0.331 + 0.415\,(\mathrm{CL})_{\mathrm{green}} + 2.125\,(\mathrm{NDVI}) + 0.415\,(\mathrm{SR}) + 3.860\,(\mathrm{EVI})_1 + 1.987\,(\mathrm{WDRVI}) + 4.082\,(\mathrm{MSAVI})_2 \quad (3)$$

Where BM is biomass. PLSR is easy to apply to large datasets as it involves only one multi-variate (linear) model, whereas RF would require the application of 500 decision trees, for each 30 m pixel of the region of interest. To further check the plausibility of the generated biomass map, the modeled biomass values were also compared against the 17-year average gross primary production (GPP) product from MODIS (MOD17) at a 1km spatial resolution (Figure 3.7).

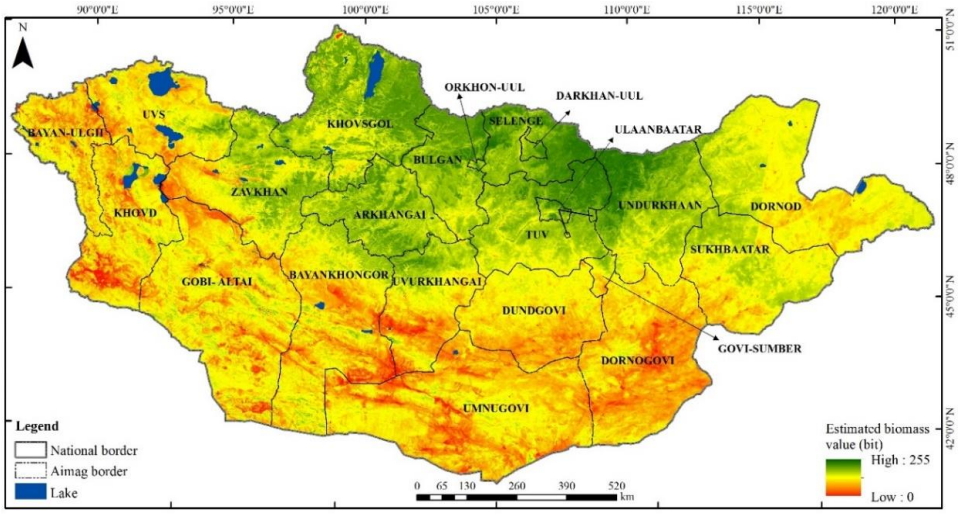


Figure 3. 5 Estimated biomass using Landsat 8 spectral vegetation indices (CL_{green}, SR, NDVI, EVI1, WDRVI, and $MSAVI_2$) and PLSR model. The resulting map was scaled to bit (0 - 255)

The close relation confirms that the modeled spatial pattern in pasture biomass is closely related to the GPP product (averaged from June to August), albeit generated at a much higher spatial resolution (30 m) compared to the former 1 km.

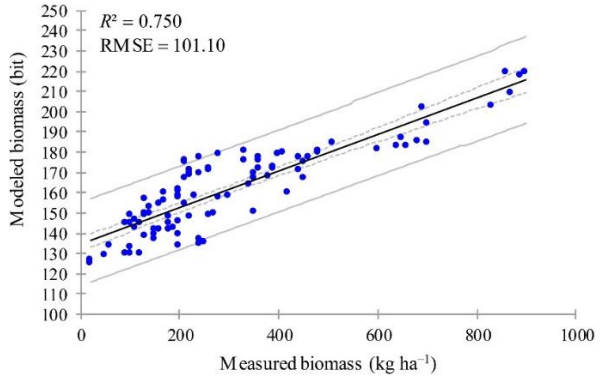


Figure 3. 6 Modeled versus measured biomass (n = 553) using the PLSR model. R^2 was 0.750 and RMSE = 101.10 kg ha^{-1}.

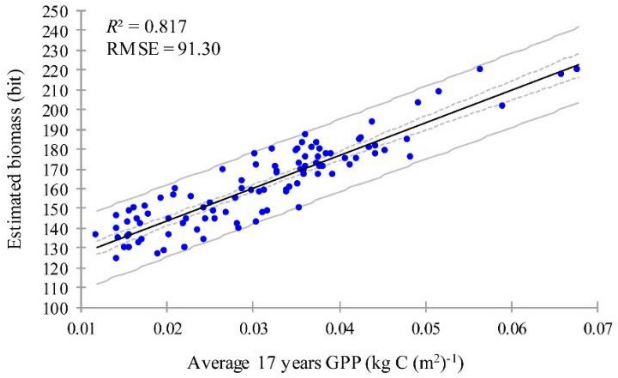


Figure 3. 7 Comparison estimated biomass values from the PLSR model and 17 years average MODIS vegetation product (MOD17) GPP (Kg C $(m^2)^{-1}$). R^2 was 0.817 and RMSE = 91.30 kg ha^{-1}.

3.4 Discussion

This study showed pasture biomass values modeled with RF (R^2 = 0.764, RMSE = 98.00 kg ha^{-1}) and by PLSR (R^2 = 0.750, RMSE = 101.10 kg ha^{-1}). Satellite sensors such as multi-spectral imagers, Radio Detection and Ranging (RADAR), and LIDAR provide highly informative remote measurements at different spectral, spatial, and temporal resolutions, which can be used to estimate biomass in pastures, as well as forests and cropland. Compared to traditional field measurement methods, only remote sensing techniques permit the estimation of aboveground biomass at coarse to fine scales and over extended spatial extents (Lu 2005). This explains the intensive use of remote sensing data for the estimation of pastureland biomass (Prince and Tucker 1986; Anderson et al. 1993; Todd et al. 1998; Kogan et al. 2004; Clevers et al. 2007; Xie et al. 2009; Ren et al. 2011; Edirisinghe et al. 2011; Xiaoping et al. 2011; Laliberte et al.

2011; Eisfelder et al. 2012; Li et al. 2013a; Dusseux et al. 2015; Ali et al. 2017). In these studies, various methods such as linear, power, and logistic regression, Multiple Linear Regression (MLR), Artificial Neural Network (ANN), Partial Least Square (PLS), Random Forest (RF) regression, Support Vector Machine (SVM), and Adaptive Neuro-Fuzzy Inference System (ANFIS) were used. These studies found correlations between field-measured biomass and different reflectance at various wavelengths, as well as the vegetation indices derived from remotely sensed data. For instance, Anderson et al (1993) calculated the relationship between the difference (DVI), ratio (RVI), NDVI vegetation indices, and dried green vegetation biomass for estimating biomass in semi-arid grassland areas. VIs were compared with ground sample estimates using sample point, spectral class, and the greenness strata approach. No strong relationship between VIs and dried green biomass was found with the sample point approach (NDVI $R^2 = 0.019$, RVI $R^2 = 0.018$, and DVI $R^2 = 0.018$). A positive correlation between VIs and dried green biomass was found with the spectral class approach (NDVI $R^2 = 0.930$, RVI $R^2 = 0.960$, and DVI $R^2 = 0.960$). A strong correlation between dried green biomass and NDVI was observed when the data were combined into greenness strata. Kogan et al (2004) estimated pasture biomass in Mongolia using vegetation health indices derived from AVHRR (biomass anomaly $R^2 = 0.658$). Clevers et al (2007) calculated grassland biomass values with SVM ($R^2 = 0.590$, RMSE = 2.10), PLS ($R^2 = 0.550$, RMSE = 2.23), and SLR ($R^2 = 0.480$, RMSE = 2.37), respectively. The results indicate that one band in the near-infrared (NIR) region and a spectral band in the red-edge region is important for predicting fresh and dry biomass when using a vegetation index. Indices based on short-wave infrared bands performed worse. Xie et al (2009) Pasture biomass values were estimated using ANN ($R^2 = 0.817$, RMSE = 0.39) and MLR ($R^2 = 0.591$, RMSE = 0.50). Li et al (2013a) five vegetation indices, plus red (R) and near-infrared (NIR) spectral bands derived from an 8-day MODIS product (MOD09Q1) were analyzed as an input for three models- MLR, ANN, and ANFIS (Ali et al. 2017). The model evaluation showed that the ANFIS ($R^2 = 0.850$, RMSE = 11070.00 kg ha^{-1}) improved the estimation of biomass as compared with the ANN and MLR. The results suggested that NDVI, RVI, and MSAVI have the best potential to estimate aboveground grassland biomass. Dusseux et al (2015) assessed the relationship between 3 variables derived from SPOT imagery and biomass measurement for estimating grassland biomass. The results showed an R^2 value of 0.680 against 0.300 and 0.500 for NDVI and FVC respectively. The squared Pearson correlation coefficient between observed and estimated biomass using LAI derived from SPOT imagery was 0.730. Li et al

(2017) estimated the grassland LAI using the RF approach and Landsat imagery in the meadow steppes of Hulunbuir, China. The results showed that the RF parameters have a small effect on the performance of RF and the RMSE was 196.00 kg ha^{-1}.

3.5 Conclusion

In this study, a methodology for estimating pasture biomass using Landsat 8 satellite data and field measurements of biomass was developed and evaluated. The analysis focussed on RF and PLS regressions. A large number of spectral predictor variables were assessed, including fourteen vegetation indices, two moisture indices, and one soil index. We also checked the predictive power of the six original spectral reflectances. Reference biomass samples from 553 sites of a field survey were available for modeling.

PLS regression analysis showed that a high correlation between biomass and Landsat indices can be obtained. Statistics between measured and modeled biomass indicated a good accuracy ($R^2 = 0.750$ and RMSE = 101.10 kg ha^{-1}). RF regression gave similar results, with a slightly higher correlation between biomass and spectral vegetation indices, yielding R^2 = 0.764 and RMSE = 98.00 kg ha^{-1}. The four most important spectral indices for both models were CL_{green}, SR, WDRVI, and EVI_1.

Compared to the spectral indices, Landsat's spectral reflectances were much less successful when used in the PLS modeling, and slightly less predictive in the RF model. This confirms that spectral indices provide a simple, yet powerful, way to reduce perturbing effects due to, for example, scene illumination and soil brightness. By combining several spectral indices in one model, their specific contribution and sensitivity can be leveraged, provided that the models used can cope efficiently with inevitable highly inter-correlated predictor variables. Both, PLSR and RF are robust against well-known correlations between various spectral indices.

Based on the optimum PLSR model, a spatial distribution map of pasture biomass in Mongolia was developed. The final map was produced at a spatial resolution of 30 m and depicts the biomass distribution around August 2016. The PLSR model was preferred for this task, as it is computationally much simpler compared to the application of a large number of decision trees involved in RF modeling. A comparison of the obtained biomass map against a 17-year average MODIS GPP products at 1km spatial resolution showed a good correlation, indicating that detailed biomass products can be generated even for very large areas.

CHAPTER 4. Estimation of climatologies of average monthly air temperature over Mongolia using MODIS Land Surface Temperature (LST) time series and machine learning techniques

4.1 Introduction

Near-surface air temperature (Ta) is a key descriptor of the climate (Nieto et al. 2011). Ta is a critical variable to the effective understanding of the many physical and biological processes between the atmosphere and land systems (Shamir and Georgakakos 2014; Benali et al. 2012; Stisen et al. 2007) because it regulates many land surface processes such as photosynthesis, respiration, and evaporation (Prihodko and Goward 1997). As air temperature influences nearly all biotic processes (Hooker, Duveiller, and Cescatti 2018), the climatologies of Ta also permit a good characterization of terrestrial environmental conditions (Prihodko and Goward 1997; Peón, Recondo, and Calleja 2014). As this variable can change quickly over space, cost-efficient mapping procedures are needed that can depict Ta using high spatial resolution.

Since the early 1980s, various interpolation methods have been used to estimate Ta has given adequate sample points (Lam 1983; Thiébaux 1991). The literature shows that the most common interpolation techniques are global interpolators, thin plate smoothing splines and different forms of kriging (Ishida and Kawashima 1993; Hutchinson 1983), inverse distance weighting (Willmott et al. 1985), and climatologically aided interpolation (Willmott et al. 1995). In the comparative study of Burrough, McDonnell, and Lloyd (2015), most interpolation methods gave similar results. However, interpolation errors typically range between 1 and 3°C (Mostovoy et al. 2006; Vogt, Viau, and Paquet 1997) depending on the spatial and temporal resolution of recorded Ta data and the density of the station network (Vancutsem et al. 2010).

Direct measurements of Ta at a height of 2 m above ground are only available from a limited number of meteorological stations. In many cases such as Mongolia, the spatial coverage of these measurements is inadequate; in addition, the typical Ta time series comes with many missing values (Hooker, Duveiller, and Cescatti 2018). On the contrary, satellite-derived Land Surface Temperature (LST) data are continuous in both spatial–temporal coverages and are relatively inexpensive. However, the satellite does not directly measure Ta but only the LST. Based on the physical linkage between LST and Ta, several authors have offered methods to estimate Ta using remote sensing satellite data (Prihodko and Goward 1997; Dash et al. 2002; Oke 1988).

During past decades, a large body of research has been collected regarding the retrieval of LST from satellite-based thermal infrared (TIR) data (Chatterjee et al. 2017), in particular, that related to a better understanding of emissivity and atmospheric effects (Dash et al. 2002; Li et al. 2013b; Prata et al. 1995). As a result, LST can be retrieved nowadays relatively accurately from remotely sensed TIR data (Cresswell et al. 1999). Several studies have demonstrated that Ta and LST data are highly correlated (Stoll and Brazel 1992). However, as expected, large differences have been noticed (Garand, Buehner, and Wagneur 2004) which are related, for example, to physical properties and atmospheric conditions (Dickinson 1983; Jin and Dickinson 2010). Three major approaches have been used to estimate Ta from LST data (Zakšek and Schroedter-Homscheidt 2009):

- Energy-balance parameterization based on thermodynamic approaches (Mostovoy et al. 2006; Oke 1988; Sun et al. 2005),
- contextual approaches based on temperature–vegetation index relations (TVX) (Prihodko and Goward 1997; Zhu, Lu, and Jia 2013; Czajkowski et al. 2000), and
- Statistical approaches using various forms of regression techniques (Mostovoy et al. 2006; Vogt, Viau, and Paquet 1997; Cresswell et al. 1999; Kilibarda et al. 2014; Chen et al. 2015; Xu, Qin, and Shen 2012; Yan et al. 2009).

Good exemplary studies estimating air temperatures with MODIS LST products using the aforementioned methods can be found for example in Bartkowiak et al (2019), Lu et al (2018), Zhou et al (2017), Janatian et al (2017), Ho et al (2014), Duan et al (2014), and Benali et al (2012).

Within the last two decades, statistical approaches, including simple and advanced regression, (e.g., linear and multiple regression, and machine learning techniques) have been developed to estimate Ta from Moderate Resolution Imaging Spectroradiometer (MODIS) LST products with varying levels of success. More recently, several studies have investigated more complex and advanced approaches to estimate Ta from MODIS LST products, such as Geographically Weighted Regression (GWR) and Climate Space Weighted Regression (CSWR) (Hooker, Duveiller, and Cescatti 2018), Spatiotemporal Regression-Rriging (STRK) (Kilibarda et al. 2014), stepwise (Janatian et al. 2017; Noi, Kappas, and Degener 2016), Random Forest (RF) (Ho et al. 2014; Noi, Kappas, and Degener 2016; Li and Zha 2019; Yoo et al. 2018; Yang, Cai, and Yang 2017; Meyer et al. 2016), Generalized Boosted Model (GBM) (Meyer et al. 2016), cubist (Noi, Kappas, and Degener 2016; Meyer et al. 2016), Support

Vector Machine (SVM) (Ho et al. 2014), Ordinary Least Squares (OLS) (Ho et al. 2014) and M5 model tree (Emamifar, Rahimikhoob, and Noroozi 2013). To ensure high modeling accuracy, several papers have highlighted the usefulness of multivariate and non-parametric algorithms such as RF and STRK. For instance, Kilibarda et al. (2014) estimated mean, maximum, and minimum daily Ta with a spatial resolution of 1 km at a global scale using STRK with MODIS 8-day time-series LST products along with elevation, wetness index, and geographical location. The performance of STRK to predict Ta from MODIS LST products was compared with the performance of the linear regression model. The results indicated that the root-mean-square errors (RMSEs) for predicting mean, maximum, and minimum daily Ta are ±2°C for areas with a high density of stations and from ±2°C to ±4°C for areas with a coarse station density. The lowest accuracy was 6°C in Antarctica and at locations with high altitudes. Yoo et al. (2018) estimated maximum and minimum daily Ta in two megacities using LST data from MODIS Terra/Aqua and seven auxiliary variables based on the RF machine learning method, resulting in an RMSE of 1.1°C and 1.2°C for maximum and minimum Ta, respectively, in Seoul, and an RMSE of 1.7°C and 1.2°C for maximum and minimum Ta, respectively, in Los Angeles. Several authors have concluded that machine learning techniques perform better than more conventional methods which provide multi-variables and nonlinear and nonparametric regression and classification (Janatian et al. 2017; Ho et al. 2014; Yoo et al. 2018; Meyer et al. 2016; Zhang et al. 2016; Keramitsoglou et al. 2016). Machine learning algorithms are particularly useful for cases where no deterministic model is available to solve the problem. Our research objective was to develop a robust empirical model to estimate climatologies of average monthly Ta across Mongolia at 1 km spatial resolution using time series of MODIS Terra LST products, terrain parameters (elevation, slope, and aspect), and other ancillary information.

4.2 Methodology and data

4.2.1 Methodology

RF regression

The well-known RF regression (RFR) method (Li and Zha 2019; Yang, Cai, and Yang 2017) was chosen as the main approach to model the relation between our response variable (Ta) and the predictor variables listed in Tables 4.1 and 4.2 (LST MODIS products plus elevation, slope, aspect, latitude, and longitude). RF is a non-linear statistical ensemble method that leverages

uncorrelated decision trees for regression. Developed by Breiman (2001), it is capable of modeling discrete and/or continuous data sets (Muntanga, Adam, and Cho 2012; Pal 2005). RF predictions are obtained by aggregating a large number of individual regression decision trees where each decision tree is built from bootstrapped training samples (as in bagging) and variables are randomly selected at each decision node. The algorithm then randomly selects a subset of the predictors as candidates for splitting (Breiman 2001; Liaw and Wiener 2002). To obtain the final regression model, the results of all the individual trees are averaged. Good examples of the benefits and drawbacks of RF are given by (Immitzer et al. 2016; Belgiu and Drăgut 2016; Hudak et al. 2008).

The RF algorithm provides Out-of-bag error (OOB) estimates and variable importance rankings (Rodriguez-Galiano et al. 2012; Genuer, Poggi, and Tuleau-Malo 2010), as not all observations are included in the respective bootstraps of the individual trees. In each tree at each split, the enhancement in the split-criterion importance measure is characterized by the splitting variable and aggregates individually all the trees in the forest for each variable (Trevor, Tibshirani, and Friedman 2006). Variable importance is measured by computing the increase in mean square error (MSE) when the OOB data for each variable are again computed but without the left-out variable (Breiman 2001; Prasad et al. 2006). The variable importance measures can assist in defining which variables are most important in the reduction of prediction error (Belgiu and Drăgut 2016). Two kinds of variable importance measure widely use the "randomForest" package in R (Liaw and Wiener 2002; Ishwaran 2007): (1) a percent increase in the mean square error (%IncMSE) and (2) an increase in node purity (IncNodePurity). From these, our analysis computed and analyzed %IncMSE. However, we also checked the IncNodePurity indicator but found similar results (not shown). In our research, the basic algorithm shown in equation (1) was used to build the RF predictor for regression (Trevor, Tibshirani, and Friedman 2006):

$$\hat{f}\,(x) = \frac{1}{B}\sum_{b=1}^{B} T_b\,(x) \qquad (1)$$

A new bootstrap sample for each decision tree T_b that includes X = x_1…… x_i with responses Y = y_1…… y_i bagging repeatedly (B times) selects a random sample from training data and each unpruned decision tree is increased in the sample. To increase each tree T_b, the following steps are repeated at each terminal node of the tree:

- Randomly select *m* variables from *p* variables

- Pick the variable that best splits and the corresponding split point
- Split the node into two nodes.

As mentioned above, to implement the RFR model, two parameters must be set: the number of decision trees (n_{tree}) and the number of variables to select for the best split (m_{try}). For both hyperparameters, standard settings have been chosen. Each decision tree is independently increased to its maximum size, focusing on a new bootstrap sample from the training data (2/3 of samples). The remaining 1/3 of the samples, not used to fit the given decision tree, are referred to as the out-of-bag sample. The OOB sample is used to calculate the OOB error rate and variable importance. For quantifying the OOB error (prediction error) for each RF decision tree, using equation (2), i.e.,

$$Error_{OOB} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

where $\hat{y}_i$ is the estimated output of OOB samples, y_i is the actual output, and *n* is the total number of OOB samples. RF regression is flexible and easy to use in comparison to other machine learning algorithms, even without hyperparameter tuning.

PLS regression

For comparison- and to assess the differences between linear and non-linear models-a prominent linear modeling technique was used: partial least square regression. PLS regression (PLSR) is widely used by the remote sensing community for vegetation analysis (Atzberger et al. 2010; Laurin et al. 2014; Inoue et al. 2012), soil-related studies (Gomez et al. 2008; Farifteh et al. 2007; Rossel et al. 2006), and climate and ecological studies (Fernandes et al. 2018; Liu, Peng, and Wang 2018; Ceglar et al. 2016; Carrascal, Galván, and Gordo 2009) amongst others.

PLSR is a multivariate linear regression method used to predict a set of dependent variables from a set of independent variables or predictors (Abdi 2003). PLSR was originally developed for econometrics and chemometrics (Wold 1982), where commonly a large number of strongly correlated predictor variables exist (Wold 2001). PLSR reduces the variables to a smaller set of uncorrelated components and performs least squares regression on these components instead of on the original data. Compared to other techniques, PLSR is more robust and less susceptible to data redundancy and over-fitting (Geladi and Kowalski 1986).

PLSR extracts a set of latent variables that explain the correlation between dependent and independent variables. The optimum number of latent variables for each generated model is implemented using the minimum value of residual mean squared error and the Leave-One-

Out-Cross-Validation (LOO-CV) methods, e.g., jackknife and bootstrap (Mevik and Cederkvist 2004). To assess which variables are most contributing to the PLSR model, used the Variable Importance in the Projection (VIP) method (Wold 2001), as seen in equation (3), i.e.,

$$VIP_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 \, SSY_f \, J}{SSY_{total} \, F}} \tag{3}$$

where VIP_j is a measure of the contribution of the j variable in the PLSR model, W_{jf} is the weight value for the j variable and f latent variables (components), SSY_f is the sum of squares of explained variance for the f latent variable and J number of the predictor (independent) variables, SSY_{total} is the total sum of squares explained as the response (dependent) variables, and F is the total number of latent variables. The VIP values determine the contribution of the predictor variables to the PLSR latent variables. A VIP value greater than 0.80 ensures that only relevant variables are considered (Wold 2001). In Mkhabela, Bullock, and Sapirstein (2018), the VIP threshold of predictor variables that were identified as the most relevant variables ranged between 0.83 and 1.21. Predictor variables with ≤0.80 VIP values were classified as less important while variables with VIP values ≥1.20 were considered the most influential.

Model evaluation and statistics

Two widely used statistics were calculated to assess the accuracy of the models (Richter et al. 2012), including the R^2 and the RMSE. The R^2 describes the percentage of explained variance whereas the RMSE summarizes the deviations of predictions from the one-to-one line. As both models provide quantitative information about the importance of different variables, also report these findings. For the RF regression model, assessed the importance of the individual predictors in Ta estimates focused on the %IncMSE (Ho et al. 2014). For the PLS regression model, used the VIP method (Farres et al. 2015; Wold 2001; Mkhabela, Bullock, and Sapirstein 2004).

4.2.2 Data

Remote sensing data

MODIS LST products are distributed by the Land Processes Distributed Active Archive Center (LP DAAC) in a hierarchical data format or HDF file. We used observations from MOD11

from the Terra satellite. MODIS generates two daily observations, one for daytime (LSTd) and one for nighttime (LSTn) at approximately 10:30 and 22:30 local time, respectively. The new collection 6 (c006) of MODIS LST products has been used to estimate Ta (Duan et al. 2018). This dataset was made available in 2016. It covers the entire period (2002-2017) and data are of higher quality compared to the earlier collection(s), which had been used for previous studies such as Benali et al (2012), Vancutsem et al (2010), Emamifar, Rahimikhoob, and Noroozi (2013), Zhang et al (2016), Oyler et al (2016), and Xu, Knudby, and Ho (2014). The LST accuracy of the c006 products is reported as being approximately twice as good as collection 5 (c005) due to the incorporation of the emissivity adjustment model in the MODIS split-window algorithm (Duan et al. 2018). For instance, the c006 LST product reduced the RMSE of bare soil sites of the c005 LST product by 1.24°C during the day and 0.58°C at night (Duan et al. 2018, p. 88).

To cover the land surface of Mongolia, seven tiles of granules with horizontal (h) and vertical (v) title numbers h23v03, h23v04, h24v03, h24v04, h25v03, h25v04, and h26v04 had been used. The MODIS MOD11A2 c006 data were obtained through the online Data Pool at the National Aeronautics and Space Administration (NASA), the LP DAAC, and the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. The retrieved MODIS LST used the generalized split-window algorithm (Wan and Dozier 1996) to derive surface temperature from the recorded at-satellite radiances.

This study, used the MODIS Terra 8-day LST product (MOD11A2) at a spatial resolution of 1 km, gridded in the Sinusoidal projection intervals, and covering the period 2002–2017. The HDF file for the MOD11A2 product includes 12 different scientific data sets (SDSs), as shown in Table 4.1. A detailed description of SDSs is given by (Wan 2008; Wan 2014). This MOD11A2 product includes daytime and nighttime LST data (LSTd and LSTn), quality information (QCd and QCn), observation information (DvA, NvA, DvT, and NvT), emissivity data (Em31 and Em32), and clear sky coverage (CsD and CsN). The HDF file for this product also contains associated quality science dataset layers which provide users with information regarding the usability and usefulness of the data products. The MODIS LST quality science dataset layers are binary encoded and bit packed. The quality assurances (QAs) layer containing integer values had been converted to a bit binary value for interpretation (Wan

et al. 2002; Wan 2007). The quality controls (QC) are defined by bit flags such as mandatory quality assessment (QA) flags, data quality flags, emissivity quality flags, and cloud error flags.

To retrieve and pre-process the products, the *MODIS* R-package (MODIS acquisition and processing package v1.1.4) was used (Mattiuzzi et al. 2018). The package is run in the R software system and environment for statistical computing and graphics (R Core Team 2019). The *MODIS* R-package allows automatic downloading of data and processing such as changing file format, mosaicking, subsetting, and time-series filtering (Mattiuzzi et al. 2019). Using the package, digital numbers (DN) of MODIS Terra LST products were converted into LST (Table 4.1). Additionally, three terrain parameters (elevation, slope, and aspect) originating from SRTM DEM (Rabus et al. 2003) were retrieved. All raster data were re-projected to MODIS sinusoidal projection.

Table 4. 1 Description of Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) products used in this study (source: Land Processes Distributed Active Archive Center (LP DAAC), 2019).

Variable Type	Acronym	Units	Data Type	Fill Value	Valid Range (VR)	Scale Factor (SF)	Additional Offset (AO)
Daytime LST	LSTd	Kelvin	16 bit	0	7500 to 65,535	0.02	N/A
Nighttime LST	LSTn	Kelvin	16 bit	0	7500 to 65,535	0.02	N/A
Day clear sky coverage	CsD	N/A	16 bit	0	1 to 65,535	0.0005	N/A
Night clear sky coverage	CsN	N/A	16 bit	0	1 to 65,535	0.0005	N/A
View zenith angle of daytime	DvA	Degree	8 bit	255	0 to 130	1.0	−65
View zenith angle of nighttime	NvA	Degree	8 bit	255	0 to 130	1.0	−65
Time of daytime (local solar)	DvT	Hours	8 bit	255	0 to 240	0.1	N/A
Time of nighttime (local solar)	NvT	Hours	8 bit	255	0 to 240	0.1	N/A
Emissivity band 31	Em31	None	8 bit	0	1 to 255	0.002	0.49
Emissivity band 32	Em32	None	8 bit	0	1 to 255	0.002	0.49
Quality controls of day LST	QCd	Bit	8 bit	N/A	0 to 255	N/A	N/A
Quality controls of night LST	QCn	Bit	8 bit	N/A	0 to 255	N/A	N/A

Note: Digital number (DN) = VR × SF − AO.

In situ meteorological data

Sixty-three synoptic weather stations are present in Mongolia. Their geographical locations are indicated. The weather stations provide Ta every three h, i.e., eight times a day. Ta data between 2002 and 2004 were obtained from the Mongolian Information Research Institute of Meteorology, Hydrology, and the Environment (IRIMHE). Data from 2004 to 2017 was downloaded from the "Reliable Prognosis (RP5)". From the three-hourly meteorological data, the average air temperature was calculated for every 8 days of MODIS LST, taking into account the eight daily observations. This led to a total of 8544 meteorological data points from 63 automatic weather stations covering the same period as the MODIS LST products, allowing for the development of prediction models between the remotely sensed data and Ta. The frequency distribution of the measured Ta reference data from 63 weather stations for the

period 2002–2017 is shown in Figure 4.1 (n = 8544). The value of measured Ta ranged from –36.6°C to 27.2°C with a mean value of 0.7°C and a standard deviation of 14.6°C.

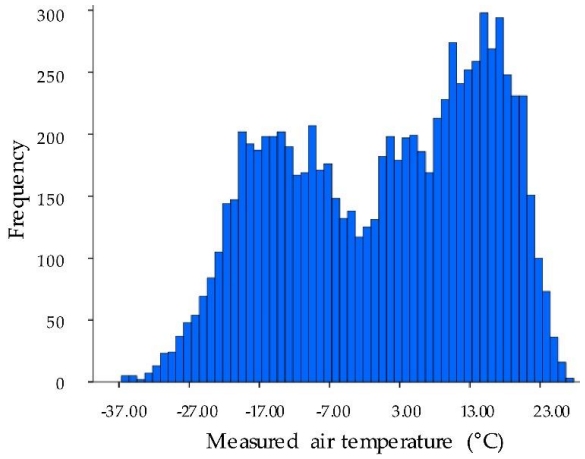


Figure 4. 1 Frequency distribution of measured average 8-day air temperature reference data (n = 8544) from the 63 automatic weather stations for the years 2002–2017. Monthly statistics are depicted in Table 4.2.

4.2.3 Data analysis using RF and PLS models

RFR and PLSR models were trained to predict Ta using up to 17 predictor variables. The use of two competing approaches permits the evaluation of the benefits of using non-linear machine learning approaches (e.g., RF) compared to classical linear regression models (e.g., PLS). Twelve of the seventeen variables were derived from LST time-series products of the Terra MODIS for the period 2002-2017 (Table 4.1). The five remaining variables were elevation, slope, and aspect (extracted from SRTM DEM), and geographical location (latitude and longitude) of weather stations (extracted from vector data). Summary descriptive statistics of the response and the 17 predictor variables are reported in Table 4.2.

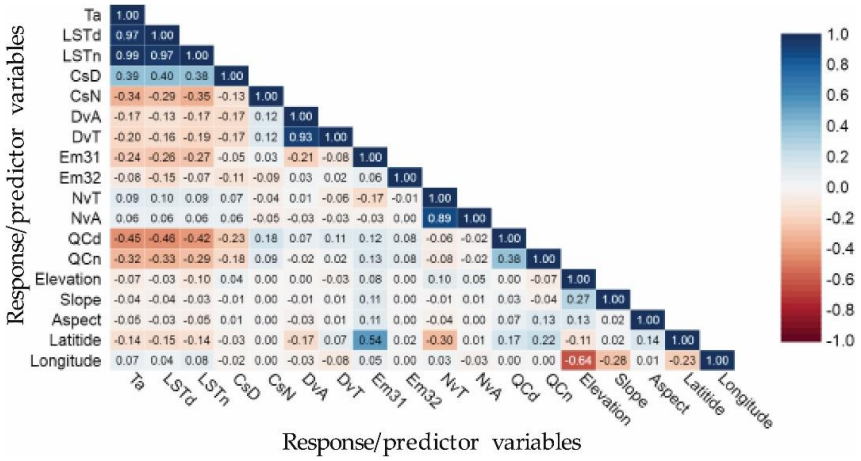


Figure 4. 2 Correlation matrix between response and predictor variables (n = 8544). The saturation of the colors indicates the strength of the correlations. Positive correlations are shown in blue and negative correlations in red. In this graph, the air temperature data has been pooled across the twelve months. For the abbreviations, see Table 4.1.

Longitude was included as this indirectly depicts (for Mongolia) the distance to the sea (Tsogtbaatar and Khudulmur 2014). The correlation matrix (Figure 4.2) reveals a strong correlation between Ta and daytime/nighttime LST of MODIS, as well as several other correlations and redundancies. Based on these intercorrelations and taking into account that the number of variable sets should be relatively small, the predictor variables were grouped into seven different groupings (Table 4.3).

Table 4. 2 List of response/predictor variables and corresponding descriptive statistics (period 2002 to 2017). The list includes the measured air temperature (Ta) reference data at the weather station level (n = 712 for each of the twelve months) as well as the corresponding seventeen predictor variables extracted from satellite and other geo-data. For the acronyms of the variables, see Table 4.1.

Variable	No. of Samples (*n*)	Minimum	Maximum	Mean	Standard Deviation
Ta_{01}	712	−36.60	−6.60	−20.83	5.36
Ta_{02}	712	−35.10	−0.60	−16.60	5.38
Ta_{03}	712	−20.50	4.80	−6.73	4.22
Ta_{04}	712	−7.60	12.40	3.92	3.32
Ta_{05}	712	2.80	19.30	10.75	3.05
Ta_{06}	712	9.20	24.90	16.98	3.06
Ta_{07}	712	11.60	27.20	19.49	3.21
Ta_{08}	712	8.70	25.60	17.16	3.34
Ta_{09}	712	2.60	19.70	10.48	3.07
Ta_{10}	712	−8.20	11.70	1.49	3.08
Ta_{11}	712	−22.70	0.30	−9.62	3.76
Ta_{12}	712	−31.50	−6.10	−17.85	4.40
LSTd	8544	−36.90	48.60	13.43	20.08
LSTn	8544	−42.50	24.40	−5.75	14.47
CsD	8544	0.00	0.13	0.06	0.02
CsN	8544	0.00	0.13	0.07	0.02
DvA	8544	−55.00	62.00	5.04	−52.35
DvT	8544	10.40	12.10	11.82	0.71
Em31	8544	0.96	0.99	0.98	0.50
Em32	8544	0.97	0.99	0.98	0.49
NvA	8544	−65.00	56.00	−0.48	−56.47
NvT	8544	20.80	22.70	21.90	1.29
QCd	8544	2.00	133.00	61.98	16.58
QCn	8544	2.00	145.00	55.14	19.70
Elevation	63	667.00	2255.00	1369.10	411.70
Slope	63	0.08	19.60	N/A	N/A
Aspect	63	6.34	358.10	N/A	N/A
Latitude	63	42.97	51.11	N/A	N/A
Longitude	63	89.93	118.67	N/A	N/A

Table 4. 3 Seven model subsets were studied. The seven groups were generated to study the relations between responses and up to 17 predictor variables. N_{var} indicates the number of predictor variables in each group.

	Acronym	Variables	N_{var}
Group 1	G1	LSTd and LSTn	2
Group 2	G2	LSTd, LSTn, and elevation	3
Group 3	G3	Elevation, slope, aspect, latitude, and longitude	5
Group 4	G4	Combined G1 and G3	7
Group 5	G5	CsD, CsN, DvA, DvT, Em31, Em32, NvA, NvT, QCd, and QCn	10
Group 6	G6	Combined G1 and G5	12
Group 7	G7	Combined G1, G3, and G5	17

4.3 Result

4.3.1 Comparison of RF and PLS models: variable importance and prediction accuracy

The estimated importance of the 17 predictor variables in the RF regression model is shown in Figure 4.3 for each of the twelve months. Under the top three in each month, LSTn appeared 12 times, LSTd 11 times, elevation 7 times, latitude 5 times, and aspect once. The remaining variables were never found under the top three in those rankings. Strong seasonality in the ranking can also be observed. For example, the warm season Ta (April to October) was strongly dependent on elevation. Conversely, the cold season Ta was more heavily affected by latitude. Using all 17 predictor variables led to R^2 values in the range from 0.83 (April) to 0.96 (August), while RMSE were between 0.91°C (September) to 1.92°C (February) (see Table 4.4, column G7).

Similar results were obtained for the PLS models (Appendix A Figure A.1). Using PLS regression, the variables most often listed under the top three were LSTn (12 times), LSTd (12 times), elevation (4 times), latitude (4 times), and emissivity (4 times). Again, the ranking was season-dependent. The variables LSTn, LSTd, and latitude were the most important variables for estimating Ta in autumn and winter (September to February). The Ta for spring (March–May) was strongly dependent on LSTn, LSTd, and emissivity. For the summer months (June–August), LSTn, LSTd, and elevation were strongly influenced by the estimation of Ta for summer.

Using the entire set of 17 predictor variables for estimating the monthly average air temperatures, the accuracies of PLS models were constantly lower compared to the RF models (Appendix A Table A.1). The PLS models gave R^2 a measured and estimated monthly Ta between 0.74 and 0.86 and RMSE from 1.20°C to 2.19°C (Table A.2, column G7). Concerning variable importance, the three variables LSTn, LSTd, and elevation were identical for PLS and RF regression models. This shows that LSTn, LSTd, and elevation play a key role in modeling Ta, with all other variables having a significantly smaller impact.

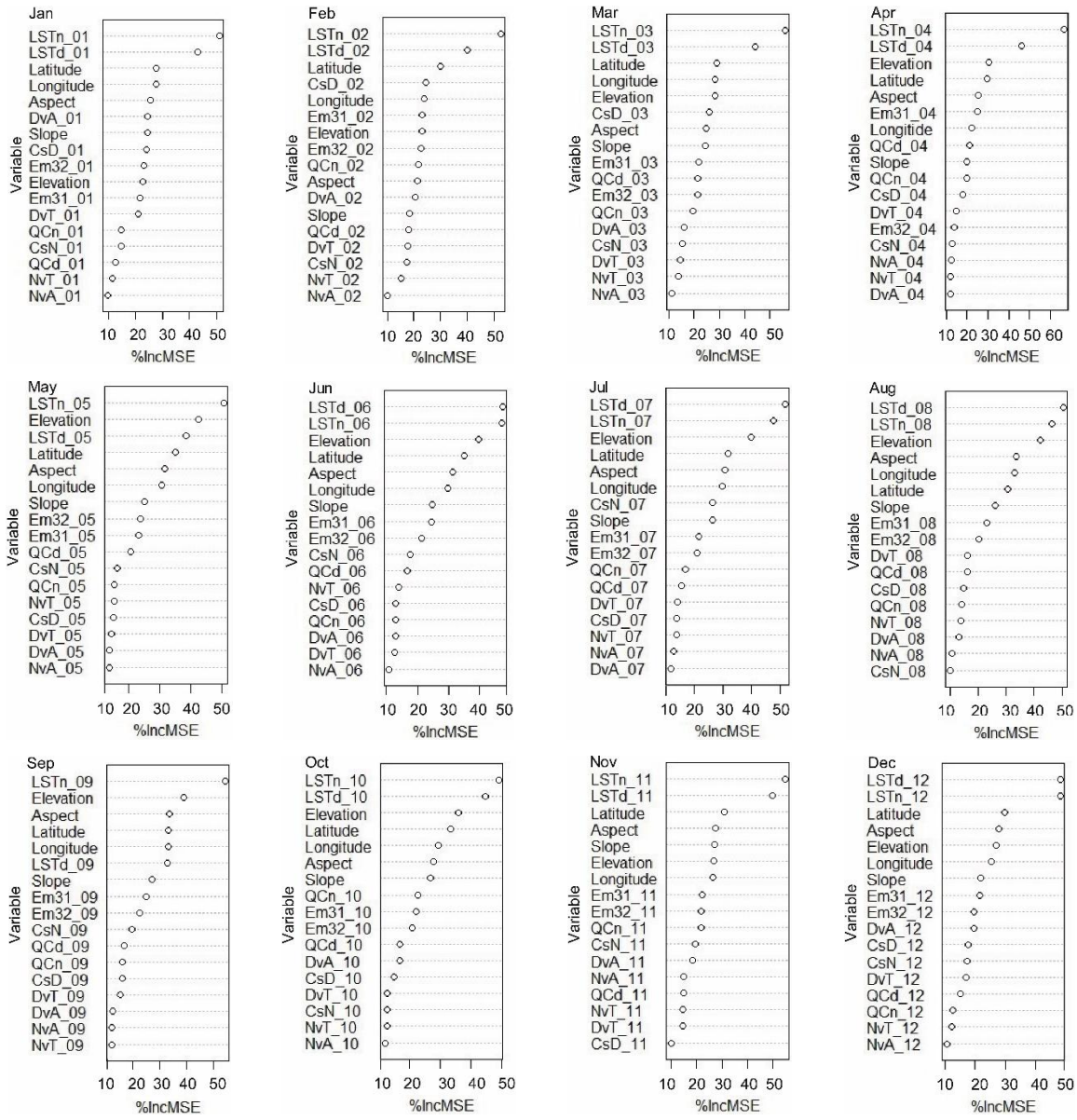


Figure 4. 3 Random forest (RF) variable importance for each month. The importance is here given as the percentage increase in mean square error (%IncMSE).

To further study the impact of the different predictor variables, the seven variable groupings highlighted in Table 4.3 were analyzed in more detail. Results for each month and the annual average air temperature are shown in Table 4.4 for the RF models. Compared to the full set of 17 variables (column G7), the reduced set with only three predictor variables LSTn, LSTd, and elevation (G2) achieved comparable results, again highlighting and confirming the importance of these three predictor variables. None of the other five groupings (G1 and G3 to G5) were able to yield similar model performances. The same findings also hold for the PLS models (Appendix A Table A.2 for details) but with constantly lower accuracies compared to

the RF models. PLS models and groupings G1 to G7 were, therefore, skipped for the remainder of the study.

Table 4. 4 Modeling results were obtained using the RF regression. Reported are the monthly summary statistics (coefficient of determination (R^2) and root-mean-square error (RMSE)) for Ta prediction models for each of the seven groups of variables. For details of groupings G1 to G7, see Table 4.3.

	n	G1		G2		G3		G4		G5		G6		G7	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
January	712	0.85	2.07	0.91	1.59	0.52	3.71	0.89	1.77	0.40	4.14	0.88	1.86	0.90	1.71
February	712	0.83	2.24	0.87	1.93	0.49	3.83	0.86	2.05	0.46	3.95	0.85	2.06	0.88	1.92
March	712	0.76	2.09	0.85	1.64	0.55	2.84	0.83	1.77	0.46	3.12	0.81	1.83	0.85	1.67
April	712	0.77	1.59	0.82	1.41	0.47	2.42	0.80	1.47	0.38	2.62	0.79	1.51	0.83	1.40
May	712	0.76	1.48	0.88	1.05	0.75	1.52	0.84	1.07	0.39	2.37	0.81	1.34	0.88	1.08
June	712	0.77	1.45	0.88	1.04	0.76	1.50	0.88	1.07	0.31	2.53	0.81	1.33	0.88	1.07
July	712	0.84	1.29	0.93	0.84	0.81	1.41	0.91	0.97	0.32	2.65	0.87	1.17	0.87	0.93
August	712	0.83	1.38	0.95	0.92	0.83	1.37	0.90	0.97	0.35	2.68	0.86	1.27	0.96	0.92
September	712	0.81	1.32	0.91	0.91	0.84	1.22	0.88	1.05	0.40	2.37	0.85	1.19	0.91	0.91
October	712	0.83	1.28	0.90	0.99	0.70	1.68	0.88	1.08	0.47	2.23	0.85	1.22	0.89	1.03
November	712	0.81	1.65	0.87	1.34	0.54	2.55	0.85	1.47	0.42	2.86	0.84	1.52	0.86	1.39
December	712	0.84	1.76	0.89	1.44	0.60	2.79	0.88	1.55	0.40	3.43	0.86	1.64	0.89	1.49

4.3.2 Maps of predicted air temperatures using RF models with the reduced feature set

Both the results of the variable importance rankings (Figure 4.3) and the grouping of variables (Table 4.4) indicate that relatively simple RF prediction models can be built to estimate Ta using only daytime/nighttime LST and elevation information. Scatterplots between measured and estimated monthly average air temperatures using only these three predictor variables are shown in Figure 4.4. Corresponding maps of modeled air temperatures at 1 km spatial resolution and covering the entire land mass of Mongolia are shown in Figure 4.5a (see Figure A2 and Table A3 in Appendix A for corresponding scatterplots and maps generated using PLS models). The scatterplots in Figure 4.4 reveal that the RF-predicted Ta is well distributed around the 1-to-1 line, with no apparent systematic deviations. In particular, we do not see any autocorrelation in the errors, or saturation effects. The errors are generally low and the explained variance (R^2) is mostly above 0.85. Generally, however, the RMSE increases slightly during the colder months (Figure 4.4).

The maps in Figure 4.5(a) depict in high spatial detail the model predictions. As expected, the predicted Ta decrease with elevation (Figure 1.1(a)) but reveals additional detail and information. Monthly analyses of the coefficient of determination (R^2; in blue) and root mean square error (RMSE; in red) for the period 2002–2017 are shown in Figure 4.5(b). Overall, a good agreement between observed and estimated Ta values was found but reflected again in

the aforementioned seasonal pattern. Large discrepancies were found to occur in transition months, such as the start or end of seasons.

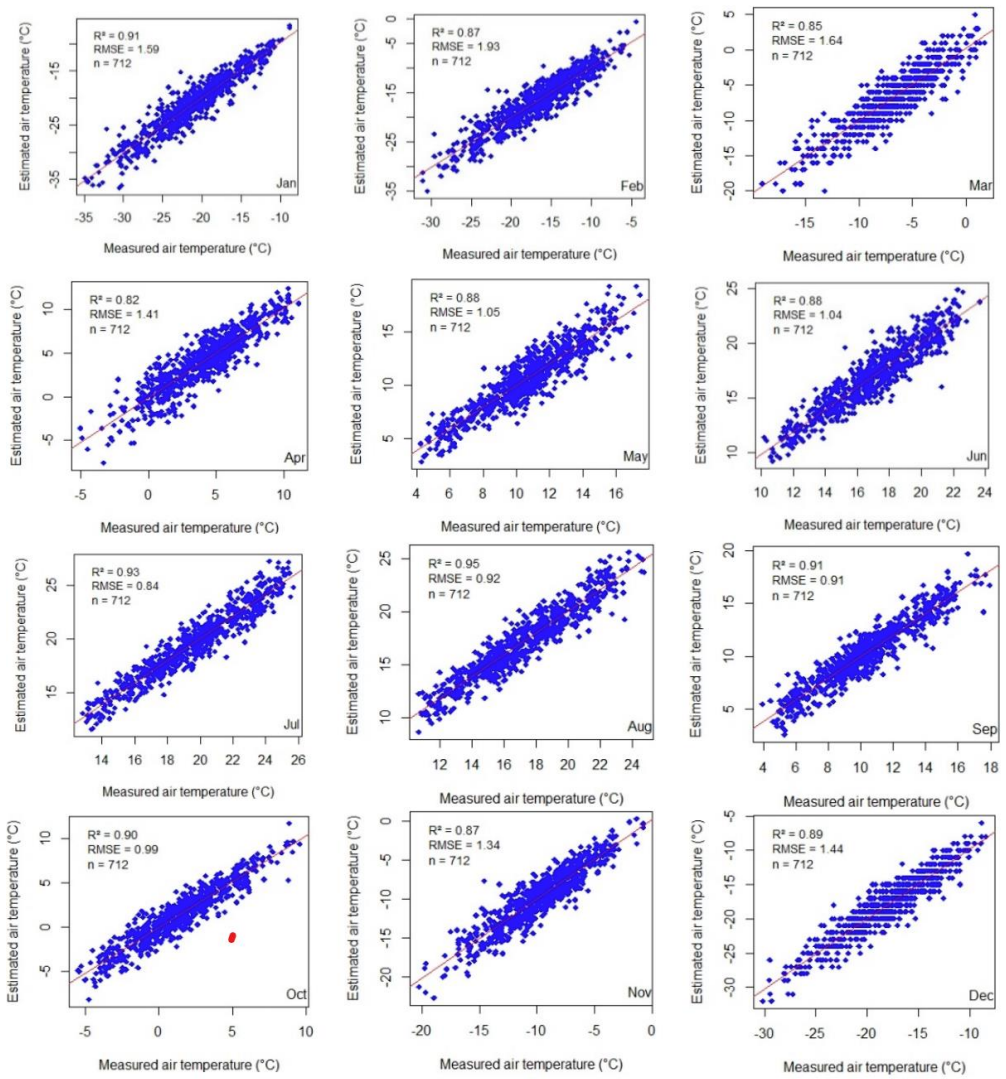


Figure 4. 4 Comparison between measured and estimated monthly average Ta using LSTd, LSTn, and elevation for the RF regression model.

The PLS-generated maps of monthly average air temperatures are generally similar to the maps derived from RF models (see Appendix A Figure A.2). However, a more detailed analysis reveals sometimes larger differences, even if modeled air temperatures are averaged by season. For example, Figure 4.6 clearly shows that large method-specific differences occur (maps in the third column). The differences show large seasonal fluctuations. The deviations moreover show a clear north-south gradient with generally lower Ta estimated using PLS compared to RF (reddish colors). The deviations are usually strongest during the warmer months. As the RF model outperformed the PLS model when evaluated against the observed

Ta (Table 4.4 for RF and Appendix A Table A.3 for PLS), we interpret these findings as mainly being the result of a systematic underestimation of Ta by the PLS model.

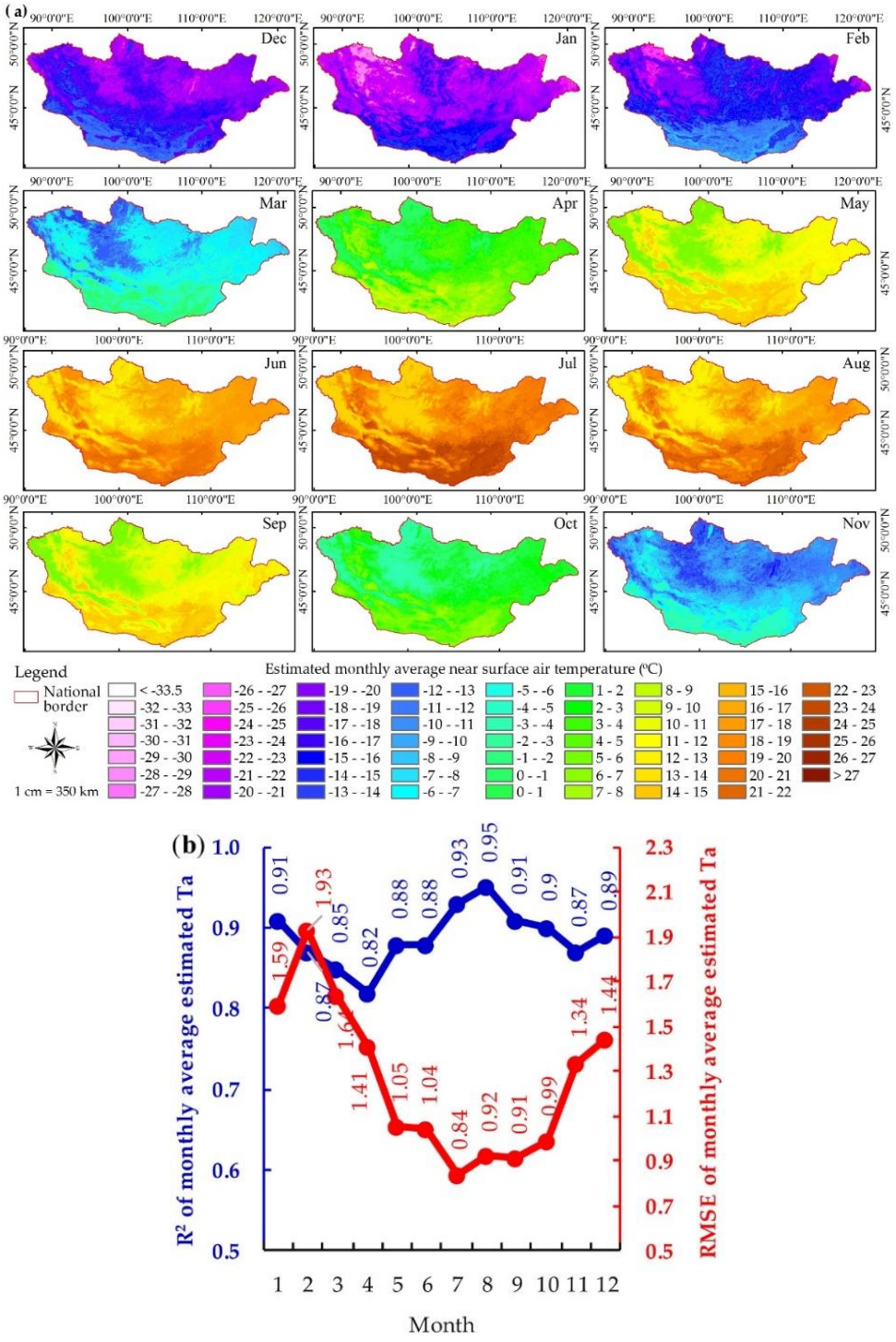


Figure 4. 5 Estimated monthly average Ta based on RF regression model using LSTd, LSTn, and elevation as predictor variables. (a) Spatial maps of estimated monthly average Ta over Mongolia at 1 km spatial resolution. (b) Monthly statistics of R^2 (blue) and RMSE (red) between observed and predicted air temperature.

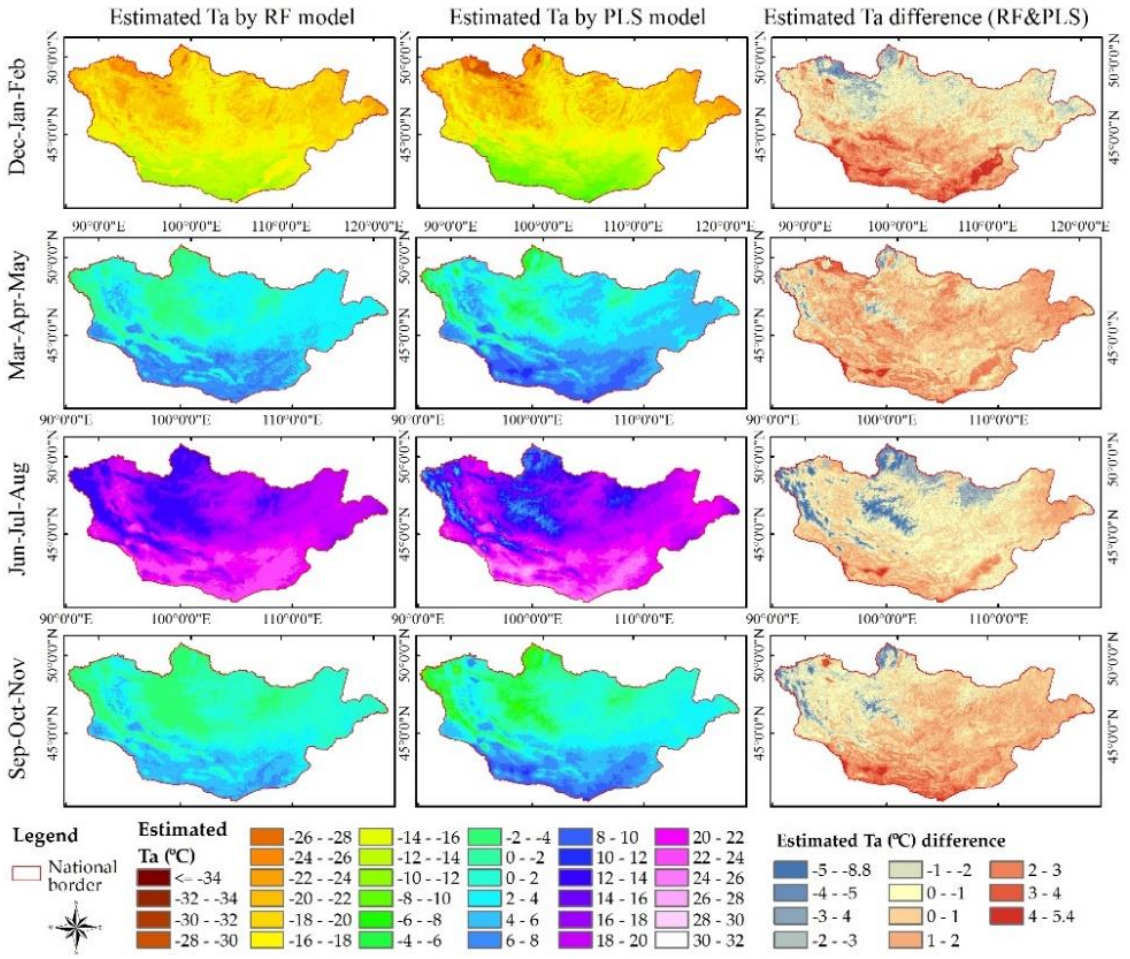


Figure 4. 6 Estimated average Ta per season using LSTd, LSTn, and elevation as predictor variables. Spatial maps of seasonal-average Ta over Mongolia at 1 km spatial resolution using the RF (first column) and partial least squares (PLS) regression models (second column). In the last column, the difference between the two model outputs is shown.

4.4 Discussion

Using relatively simple RF models driven by a few predictor variables, climatologies of monthly air temperatures in Mongolia could be obtained in this study with high accuracy (RMSE of about 0.84-1.93°C). Without any hyperparameter tuning, the non-linear RF models outperformed linear PLS models by other studies (Yoo et al. 2018; Xu, Knudby, and Ho 2014; Otgonabayar et al. 2019; Sun, Passi, and Jain 2016). Amongst the variables studied, the MODIS-derived land surface temperatures (day and night) together with elevation were the three most important predictors. The studies of Noi, Kappas, and Degener (2016), and Kilibarda et al (2014) have also reported the high importance of day- and nighttime LST observations as well as elevation.

As LST (both day and night) and elevation can be readily produced at 1 km spatial resolution, the models calibrated against weather station data permitted the creation of maps of

average air temperature for each of the twelve months in unprecedented detail and accuracy. Although the RF-generated maps often follow elevation, the inclusion of remotely sensed land surface temperature from MODIS improved the accuracy and spatial detail.

The results of the importance analysis indicated that nighttime LST was slightly more important compared to LSTd. The same result has been noted in China (Li and Zha 2019) and in Portugal (Benali et al. 2012). By these studies, we argue that nighttime observations are probably more predictive because LSTn is not affected by reflected solar radiation when using TIR sensors (Vancutsem et al. 2010). The daytime land surface temperature was nevertheless found to be important, as it reveals the strength of the latent heat flux and the energy available for generating sensible heat (Li and Zha 2019).

Other studies have confirmed that Ta predictions are possible using satellite observations and that there is a strong relationship between Ta, LSTn, and LSTd (Chen et al. 2015). Several studies have produced Ta estimations using MODIS LST data using multivariate linear and non-parametric regression methods (Ho et al. 2014; Noi, Kappas, and Degener 2016; Li and Zha 2019; Yoo et al. 2018; Yang, Cai, and Yang 2017; Mkhabela, Bullock, and Sapirstein 2018). These already published studies have shown different levels of success. The performance of multivariate and non-parametric regression models has been strongly dependent on environmental parameters such as vegetation cover, slope, aspect, elevation, quality of MODIS LST products, and applied filter techniques. For instance, the accuracy of the MODIS LST has been found to depend on the employed split-window algorithm, cloud cover, and terrain parameters (Chen et al. 2015). Nonetheless, the estimation of Ta derived from MODIS LST studies using multivariate and non-parametric algorithms is suitable for generating results at high accuracy. For instance, Li and Zha (2019) estimated the monthly average Ta for the territory of China at a spatial resolution of 1 km using RF regression with MODIS LST, Normalized Difference Vegetation Index (NDVI), nighttime light, and elevation. Using this dataset, the RMSE of the monthly average Ta ranged between 1.57°C to 1.99°C. Our study has shown that monthly average Ta can be accurately estimated using LSTn, LSTd, and elevation with similar RMSE ranging from 0.91°C to 1.93°C. The method is relatively easy to implement provided that there is a sufficient amount of training data with corresponding EO time-series observations.

4.5 Conclusion

In this study, PLS and RF regression models were applied to estimate the monthly average Ta in Mongolia for the period 2002-2017 using MODIS LST time-series products and terrain parameters. Meteorological data from 63 automatic weather stations were used to calibrate and validate the PLSR and RF models. Both models were trained to predict Ta using up to 17 variables as predictor variables. Twelve variables were derived from LST time-series products of Terra MODIS and three variables were extracted from an SRTM DEM (elevation, slope, and aspect). The geographical location (longitude and latitude) was used as an additional variable. For training, a total of 8544 meteorological data points from 63 automatic weather stations and corresponding MODIS LST were used. Both datasets covered the period 2002–2017. Using only day/nighttime LST and elevation as predictor variables, the correlation between measured and estimated monthly average Ta RMSE ranged from 1.20°C to 2.19°C for the PLSR and 0.84°C to 1.93°C for the RF. The significantly lower errors of the RF models confirm the benefits of this machine-learning approach compared to traditional (linear) modeling techniques (e.g., PLSR). We, therefore, recommend the use of RF models for similar studies.

Concerning the MODIS land surface temperature data, we found that this information contributed significantly to the modeling of air temperature. For example, it was not possible to obtain similarly low errors in the modeled air temperature using only terrain parameters as predictors. It is recommended that day- and nighttime LST be used simultaneously as both variables scored high in the feature importance metric.

Both machine learning models (RF and PLSR) represented well seasonal and spatial variations in Ta when time-series of LST were included as predictor variables. Using the models, maps of the monthly average Ta of Mongolia were developed at a spatial resolution of 1 km which was representative of the period 2002–2017. Although errors in the predicted Ta were generally low, the residual errors showed a significant seasonality; the warmer months were generally better modeled compared to the extremely cold winter months. Probably, the increased errors during the winter months reflect a lower accuracy in the input (LST) data. Further research is warranted to better understand the seasonality of the model quality.

Despite these trends, we firmly believe that this spatial dataset may be useful for various environmental applications; for instance, it may be useful for better assessing bioclimatic variations within the huge land mass of Mongolia. The developed methodology is

relatively easy and transparent and can be applied in different geographic regions, provided that enough weather stations are available to permit a model calibration. The spatial resolution of the final map product mainly depends on the ground sampling distance of the employed satellite sensors. As sensor technology advances at a rapid pace, the current 1 km spatial resolution can be further improved shortly.

Appendix A.

Supplementary data related to PLS regression results can be found in Appendix A.

CHAPTER 5. Estimation of bioclimatic and climatic variables of Mongolia derived from a time series of remote sensing data

5.1 Introduction

Climate resource is the main indicator of nature-ecology that determines a nation's economic development (Mijiddorj 2016). Various climatic variables (indices) are important to perceive climate mechanisms, and assess climatic conditions (Deniz, Toros, and Incecik 2011). Moreover, a large number of ecological studies have used climate-based models: two prominent examples are ecological niche models (Feilhauer et al. 2012; Waltari et al. 2007) and Species Distribution Models (SDMs) (Anderson 2012). The United States Geological Survey (USGS) developed for such purposes climate indices, which can be referred to as bioclimatic variables (O'Donnell and Ignizio 2012). Bioclimatic variables are widely used in species distribution modeling (Attorre 2007; Waltari et al. 2014; Salas et al. 2017). SDMs integrate information on species' appearance with environmental features to estimate their distributional range (Vega et al. 2018). SDMs are moreover valuable for other applications across evolutionary ecology and biology (Title and Bemmels 2018).

Besides those applications, bioclimatic variables also capture features of climate (Mesquita and Sousa 2009) that are directly related to plant physiological processes determining primary productivity (Leathwick et al. 2003). The bioclimatic variables represent the types of seasonal trends relevant to the physiological constraints of different species (O'Donnell and Ignizio 2012). Bioclimatic variables also include information on annual conditions, as well as seasonal mean climate conditions, and intra-year seasonality (O'Donnell and Ignizio 2012; Fick and Hijmans 2017). These variables represent annual trends, seasonality, and extreme or limiting environmental factors (Hijmans et al. 2005). Because of these characteristics, bioclimatic variables are widely used for vegetation mapping (Fraklin 1995; Hengl et al. 2018), and to study the effects of climate change on species distribution for past, current, and future scenarios (O'Donnell and Ignizio 2012; Walther et al. 2005; Peng 2000; Sykes et al. 1996), to monitor exotic and invasive species (Arriaga et al. 2004), for regional planning (Bryan and Crossman 2008), ecosystem distribution (Thompson et al. 2004), and to assess drought risk (Incerti et al. 2007). At the global level, a set of 19 gridded datasets was developed within WorldClim based on weather stations, involving data from the Global Climate Network Dataset (GHCN) (Lawrimore et al. 2011), the World Meteorological Organization climatological database, and additional minor database-specific weather stations

(WMO 2014). In WorldClim, the bioclimatic variables were derived from two climatic data sources to generate more biologically meaningful variables (O'Donnell and Ignizio 2012), which are monthly mean, minimum, and maximum temperature, and monthly total precipitation.

There are two versions of WorldClim bioclimatic variables. WorldClim version 1.4 is a global climate-gridded data set for the years 1961-1990 (excluding Antarctica) at 3 resolutions (2.5 minutes, 5 minutes, 10 minutes) (Marchi et al. 2019; Hijmans et al. 2005). WorldClim version 2.0 is a new dataset containing grids with interpolated data from between 9000 and 60000 weather stations for 4 different spatial resolutions from 30 seconds (~1 km) to 10 minutes (~340 km) for the years 1971-2000 (Fick and Hijmans 2017). In addition, Michael et al (O'Donnell and Ignizio 2012) developed a set of 20 bioclimatic variables as continuous raster surfaces between 1985 and 2009. Moreover, Vega et al (2018) reproduced interpolation methods from WorldClim to create MERRAclim, a global set of 19 bioclimatic variables that includes Antarctica. MERRA (Modern Era Retrospective-analysis for Research and Application) is a NASA (National Aeronautics and Space Administration) atmospheric data reanalysis of satellite information. MERRAclim contains three datasets of 19 bioclimatic variables for the years 1980, 1990, and 2000, using hourly temperature and humidity data from 1980 to 2000 at three different resolutions (2.5 minutes, 5 minutes, 10 minutes) (Vega et al. 2018).

In parallel, various large-scale gridded interpolated temperature and precipitation datasets at different spatiotemporal resolutions have been developed from in-situ measurements to estimate bioclimatic variables (Marchi et al. 2019; Vega et al. 2018; Fick and Hijmans 2017; Hijmans et al. 2005). Unfortunately, in-situ measured temperature and precipitation data with long temporal coverage are only available from a limited number of meteorological stations with inadequate spatial coverage (Otgonbayar et al. 2019). These data sets, therefore, suffer from uneven geographic coverage, with many areas of the Earth poorly represented (Hijmans et al. 2005).

On the contrary, EO satellites capture the entire Earth's surface at much denser Ground Sampling Distances (GSD) and with high temporal revisit frequency (usually 1 day). This data permits estimating the monthly mean, minimum, and maximum surface temperature (Benali et al. 2012), as well as monthly total precipitation (Sun et al. 2018). As sensor technology advances at a rapid pace, advanced geo-informatics techniques offer an opportunity

to estimate monthly temperature more accurately, and collect precipitation data derived from remote sensing sensors such as multispectral imagery, Radio Detection Ranging (RADAR), and Light Detection and Ranging (LIDAR) at different spectral, spatial, and temporal resolutions. For instance, Fick and Hijmans (2017) determined, satellite data enhanced by 5-15% prediction quality of temperature variables, especially low spatial density area. And improving the accuracy of precipitation data, they suggested using satellite-based precipitation data as covariates. Amiri et al (2020) estimated 19 bioclimatic variables from temperature and precipitation instrumental records (Model 1), and remote sensing data (Model 2) at a resolution of 1 km during 2001-2017 in Isfahan province of Iran together with three topographic variables using five different regression models. Accuracy statistics in Model 2 were higher than in Model 1. This study proved that bioclimatic variables derived from the satellite were more effective.

Our main goal is to explore alternative ways to improve the temporal and spatial resolution of bioclimatic variables derived from remotely sensed data. The specific aim of this study is to estimate bioclimatic, and climatic variables using time series of land surface temperature (LST) from Moderate Resolution Imaging Spectroradiometer (MODIS), and precipitation (P) from Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data and to apply the model to the entire land surface of Mongolia. For this analysis, we estimated monthly maximum, mean, and minimum air temperature from Terra MODIS satellite LST (MOD11A2) for the period 2002-2017 using the Random Forest (RF) regression model and three predictors (Otgonbayar et al. 2019; Otgonbayar and Sumya 2020).

5.2 Methodology and data

5.2.1 Methodology

To calculate 19 bioclimatic variables at 1 km spatial resolution we used the functions listed in Table 5.1. The functions use as inputs satellite-derived air temperature (monthly maximum, monthly average, monthly minimum), and monthly total precipitation. All calculations were done in R for statistical computing and graphics (Ripley 2001), and System for Automated Geoscientific Analyses (SAGA GIS) for analysis of spatial data (SAGA 2013). To estimate six climate indices used. The analysis used the ‘biovars’ function of the ‘dismo’ package in R. To test the differences between our set of bioclimatic variables (“SatClim”) and WorldClim

bioclimatic variables, we used coefficient of determination (R^2), root mean squared error (RMSE), and normalized root mean squared error (nRMSE) as described in Table 5.2.

Table 5. 1 Formula and description of the bioclimatic variables (O'Donnell and Ignizio 2012). T_{avg}, T_{max}, and T_{min} are the monthly average, maximum, and minimum air temperature, and PPT is the monthly total precipitation

Variable name	Unit	Formula	Description
Annual mean temperature	°C	$Bio\ 1 = \frac{\sum_{i=1}^{i=12} Tavg_i}{12}$	The annual mean temperature approximates the total energy for an ecosystem.
Annual mean diurnal range	°C	$Bio\ 2 = \frac{\sum_{i=1}^{i=12}(Tmax_i - Tmin_i)}{12}$	Mean of the monthly temperature range. This variable can help provide information relating to the relevance of temperature variation for different species
Isothermality	%	$Bio\ 3 = \frac{Bio\ 2}{Bio\ 7} \times 100$	Isothermality quantifies how large the day to night temperatures fluctuate relative to the summer to winter (annual) fluctuations. A species distribution may be influenced by larger or smaller temperature oscillation within a month relative to the year and this variable is useful for confirming such information.
Temperature seasonality (Standard deviation)	°C	$Bio\ 4 = SD\{Tavg_1, \dots, Tavg_{12}\}$	Temperature seasonality is a measure of temperature change over the year. A large Standard Deviation (SD) is the larger variability of temperature.
Maximum temperature of the warmest month	°C	$Bio\ 5 = \max(\{Tmax_1, \dots, Tmax_{12}\})$	Monthly maximum temperature incidence over a given year (time series) or averaged span of years. This variable is useful to test that species distribution is affected and influenced by warm temperature anomalies over the year.
Minimum temperature of the coldest month	°C	$Bio\ 6 = \min(\{Tmin_1, \dots, Tmin_{12}\})$	Monthly minimum temperature incidence over a given year (time series) or averaged span of years. This variable is useful to test that species distribution is affected and influenced by cold temperature anomalies over the year.
Annual temperature range	°C	$Bio\ 7 = Bio\ 5 - Bio\ 6$	The measure of temperature fluctuation over a given year. This variable is useful to investigate whether species distribution is affected by the range of extreme temperature conditions.
Mean temperature of wettest quarter	°C	$Bio\ 8 = \frac{\sum_{i=1}^{i=3} Tavg_i}{3}$ or $Q_{PPT_{max}}$	The quarterly variable is based on 3 months interval that is a mean temperature that prevails during the wettest season. The variable is useful for analyzing how such environmental factors can influence species season distribution.
Mean temperature of driest quarter	°C	$Bio\ 9 = \frac{\sum_{i=1}^{i=3} Tavg_i}{3}$ or $Q_{PPT_{min}}$	The variable provides mean temperature during the driest 3 months of the year which is useful for analyzing how such environmental factors can influence species season distribution.
Mean temperature of warmest quarter	°C	$Bio\ 10 = \frac{\sum_{i=1}^{i=3} Tavg_i}{3}$ or $Q_{T_{max}}$	The variable provides mean temperature during the warmest 3 months of the year which is useful for analyzing how such environmental factors can influence species season distribution.
Mean temperature of coldest quarter	°C	$Bio\ 11 = \frac{\sum_{i=1}^{i=3} Tavg_i}{3}$ or $Q_{T_{min}}$	The variable provides mean temperature during the coldest 3 months of the year which is useful for analyzing how such environmental factors can influence species season distribution.
Annual precipitation	mm	$Bio\ 12 = \sum_{i=1}^{i=12} PPT_i$	The variable is recognized by the sum of 12 monthly precipitation values which is useful in ascertaining the significance of water availability to the species distributions.
Precipitation of wettest month	mm	$Bio\ 13 = \max(\{PPT_1, \dots, PPT_{12}\})$	The variable is recognized by total precipitation values that prevail during the wettest month. The wettest month is useful if extreme precipitation conditions during the year affect a species potential range.
Precipitation of driest month	mm	$Bio\ 14 = \min(\{PPT_1, \dots, PPT_{12}\})$	The variable is recognized by total precipitation values that prevail during the driest month. The driest month is useful if extreme precipitation conditions during the year affect a species potential range.
Precipitation seasonality	mm	$Bio\ 15 = \frac{\text{SD}\{PPT_1, ., PPT_{12}\}}{1 + (Bio\ 12/12)}$	The variable is a measure of the variation in monthly precipitation totals over the year. This variable is expressed by percentage where larger percentages represent greater variability of precipitation.
Precipitation of wettest quarter	mm	$Bio\ 16 = max \sum_{i=1}^{i=3} PPT_i$	This quarterly provides total precipitation during the wettest 3 months of the year which can be useful for testing how such environmental factors can influence species season distribution.
Precipitation of driest quarter	mm	$Bio\ 17 = min \sum_{i=1}^{i=3} PPT_i$	This quarterly provides total precipitation during the driest 3 months of the year which can be useful for testing how such environmental factors can influence species season distribution.
Precipitation of warmest quarter	mm	$Bio\ 18 = \sum_{i=1}^{i=3} PPT_i$ or $Q_{T_{max}}$	This quarterly provides total precipitation during the warmest 3 months of the year which can be useful for testing how such environmental factors can influence species season distribution.
Precipitation of coldest quarter	mm	$Bio\ 19 = \sum_{i=1}^{i=3} PPT_i$ or Q_{min}	This quarterly provides total precipitation during the coldest 3 months of the year which can be useful for testing how such environmental factors can influence species season distribution.

Table 5. 2 Performance measures used in this study: coefficient of determination (R^2), root mean squared error (RMSE), and normalized RMSE (nRMSE). The three statistics represent correlation (association), error (residual), and range normalized errors (Richter et al. 2012).

Formula	Description	Range	Reference
$R^2 = 1 - \frac{\sum_{i=1}^{n}(V_{est}^{i} - \hat{V}_{est})^2}{\sum_{i=1}^{n}(V_{est}^{i} - \bar{V}_{est})^2}$	The R^2 measures the correlation between the predicted and observed value (fraction of explained variance)	0 to 1	(Richter et al. 2012)
$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(V_{est}^{i} - V_{obs}^{i})^2}$	The RMSE is a measure of the average magnitude of errors along the 1-to-1 line	Data unit	
$nRMSE = \frac{RMSE}{Range_{(obs)}}$	Normalizing the RMSE facilitates the comparison between datasets or models with different scales. NRMSE is the ratio of the RMSE to the variance of the observed variable.	0 to ∞	(Barzegar et al. 2016)

Where V_{est}^{i} – estimated variables, $\hat{V}_{est}$ – average of the estimated variables, $\bar{V}_{est}$ – average of the predicted variables, V_{obs}^{i} – observed variables, $Range_{(obs)}$ – range of the observed variables, and n- number of observed variables

For the climatic indices, the widely used six climatic indices were applied to generate spatial distribution mapping of the hydrothermal, aridity, and moisture condition of Mongolia equation (1-13).

1. *Selyaninov (1966) hydrothermal coefficient*

$$HTC = \frac{\sum P}{[0.1 * \sum T_{>X^oC}]} \quad (1)$$

HTC- Hydrothermal coefficient

ΣP- Annual total precipitation in a warm period, mm (daily mean temperature >10^0C)

$\sum T_{>X^oC}$-amount of air temperature >10^0C

$> X^OC$- temperature threshold value (x≥10^0C)

2. *De Martonne (1925) aridity index*

$$I_{ar} = \frac{P}{T + 10} \quad (2)$$

I_{ar}- Aridity index (мм/^{o}C)

P- Monthly precipitation sum (mm)

T- Monthly mean air temperature (^{o}C)

3. *Thornthwaite (1948) humidity factor*

$$HF_{th} = \frac{P_{I-XII}}{E_o} \quad (3)$$

P_{I-XII}- Annual precipitation sum (mm)

E_o- Annual evaporating capacity (mm)

$$E_o = 1.6 * \left(\frac{10T_i}{J}\right)^a \quad (4)$$

T_i-Mean monthly temperature (^{o}C)

i- Monthly heat index (^{o}C)

$$i = \left(\frac{T_i}{5}\right)^{1.514} \quad (5)$$

a- A coefficient that varies with the heat index and is given by

$$a = 6.75 * 10^{-7}J^3 - 7.71 * 10^{-5}J^2 + 1.79 * 10^{-2}J + 0.49 \quad (6)$$

4. Mezentsev (1969) moisture index

$$MI = \frac{P}{[0.2 * \sum T_{>10^oC} + 306]} \quad (7)$$

P- Annual precipitation sum (mm)

$T_{>10^oC}$- an amount of air temperature >10^{0}C

5. Ivanov, climate biological effectiveness indicator

$$BEC = (0.01\sum T_{>10}) * КУ \quad (8)$$

BEC- Climate biological efficiency indicators

$\sum T_{>10}$- Amount of air temperature >10^{o}C

КУ- Moisture content coefficient

$$E = 0.0018(25 + t)^2 * (100 - f) \quad (9)$$

t- Average monthly air temperature

f- Average monthly air relative humidity

$$КУ = \frac{P}{E_o} \quad (10)$$

P- Annual precipitation sum, мм

E_0- (9) annual evaporation capacity

6. Shasho (1985) bioclimatic potential (Kobysheva 2005)

$$BCP = Kp(ky)\ \frac{\sum t^0_{>10}}{\sum t_{ak(баз)}} \quad (11)$$

BCP- Bioclimatic potential

$Kp(ky)$- Indices of heat and water availability

$\sum t^0_{>10}$- Amount of air temperature >+10^{o}C

$\sum t_{ak(\text{баз})}$- 1000 °C (mean value for a cultivated area in the territory of Mongolia)

$$Kp(ky) = \lg(20K_{\text{увл}}) \quad (12)$$

$K_{\text{увл}}$- annual air humidity coefficient

$$\text{K}_{\text{увл}} = \frac{\text{P}}{\sum \text{d}} \quad (13)$$

P- Annual precipitation (mm)

∑d-Monthly amount lacks moisture

5.2.2 Data

Temperature data

Air temperature measurements from station data can in principle be interpolated to derive spatial maps (Robeson 1994). However, interpolation errors are often significant, depending on local conditions and the spatial and temporal resolution of measured air temperature data and station density (Dodson and Marks 1997). Similar to many other countries, Mongolia's weather station network for air temperature observations has insufficient spatial coverage. Satellite-derived LST data provide continuous spatial and temporal coverage and might, therefore, be used to model the temperature fields. However, satellites only measure land LST, and hence air temperature has to be estimated (Hooker et al. 2018). A suitable approach was presented by (Otgonbayar et al. 2019) where monthly maximum, minimum, and average air temperature over Mongolia was estimated using MODIS LST (MOD11A2, v006) time series and the Random Forest (RF) regression model. MODIS LST was obtained through an online data pool at the National Aeronautics and Space Administration (NASA) Land Processes Distributed Active Archive Centre (LP DAAC). Using the approach presented, we created spatial maps of monthly maximum temperature, minimum temperature (Otgonbayar and Sumya 2020), and average temperature (Otgonbayar et al. 2019) for the period of 2002-2017, at a spatial resolution of 1 km.

Precipitation data

Over the past two decades, numerous precipitation products have been generated from gauge-radar and gauge-satellite harmonized precipitation analysis at regional to global levels (Bai and Liu 2018; Li et al. 2013b; Price et al. 2014). Detailed information on the precipitation database combining rain gauge, satellite, and reanalysis products can be found for instance in Roca et al

(2019) and Beck et al (2017). Here, used two combined gauge-satellite datasets with a fine spatial resolution: CHIRPS and PERSIANN-CCS (Table 5.3). To select the most appropriate precipitation data, compared CHIRPS and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System (PERSIANN-CCS) (details see, Otgonbayar et al. 2021a). According to our findings, the CHIRPS data was overall far more accurate than PERSIANN-CCS. The main limitation of the CHIRPS data was its limited spatial coverage. Indeed, CHIRPS only covers the area 50S°-50N° whereas the northern part of Mongolia goes up to 52N°. To generate a gap-free wall-to-wall map for the entire territory of Mongolia, we filled the part of Mongolia between 50°N and 52°N with data from the Climate Hazards Center's Precipitation Climatology data version 1.0. CHIRPS monthly total precipitation datasets were obtained from the Climate Hazard Center website with a spatial resolution of 0.05°, spatial coverage of 87°E-120E°, 41°N-50°N, and covering the period between 2002 and 2017. These datasets were developed in collaboration with scientists at the USGS Earth Resources Observation and Science (EROS), supported by the United States Agency for International Development's (USAID) Famine Early Warming Systems Network (FEWS NET). The datasets are built on 'smart' interpolation techniques, estimates focused on infrared Cold Cloud Duration (CCD) observations that are available in GeoTIFF, NetCDF, and BIL formats. The unit is mm per period, including mm per day, pentad, and month (Funk et al. 2015).

Table 5. 3 Two precipitation products with a high spatial resolution (Roca et al. 2019; Sun et al. 2018; Bai and Liu 2018; Beck et al. 2017). Only CHIRPS was used for this study.

Product name	Acronym	Data used	Spatial Coverage	Spatial resolution	Temporal coverage	Temporal resolution	Reference
Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)- Cloud Classification System	PERSIANN-CCS	Gauge-satellite	60°N-60°S	0.04°	2003-present	Hourly, Daily	(Nguyen et al. **2018**)
Climate Hazards Group Infrared Precipitation with Stations	CHIRPS v2.0	Gauge-satellite	50°N-50°S	0.05°	1981-present	Daily	(Funk et al. 2015; Chris Funk et al. 2015)

Here, we compared between derived from satellites (MODIS and CHRIPS) for the period 2002-2017 and the weather station-based WorldClim datasets. WorldClim datasets include grids interpolated in situ station data for the 1970-2002 time period (Figure 5.1). Estimated monthly maximum, mean, and minimum temperatures derived from MODIS LST are highly

correlated with World Climatic temperature datasets compare to precipitation datasets (Table 5.4).

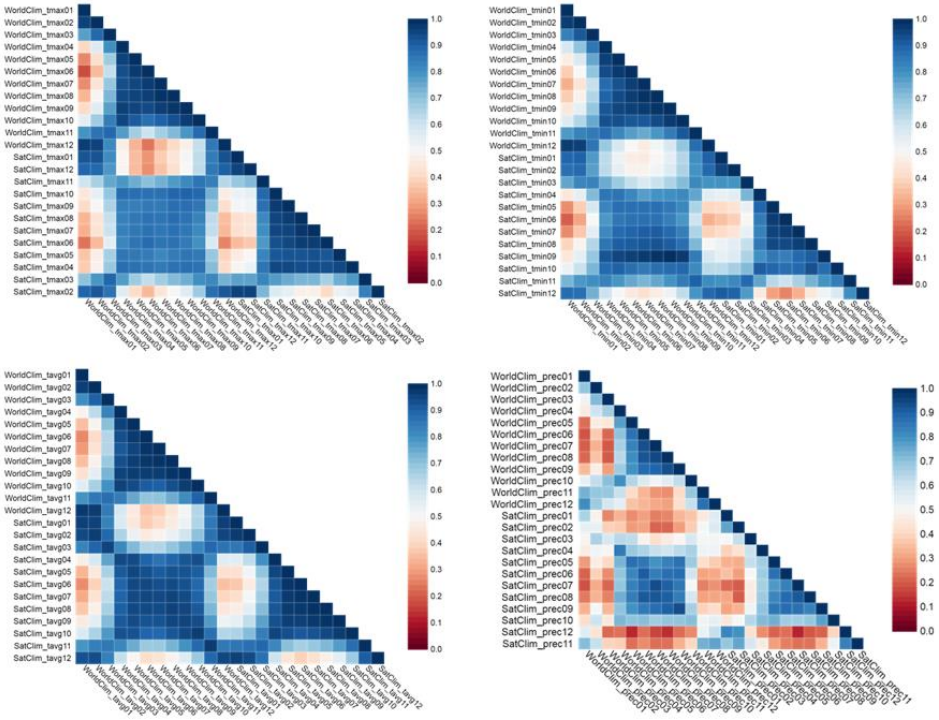


Figure 5. 1 Inter-correlation between estimated monthly air temperatures derived from MODIS LST, and World climatic datasets; precipitation CHRIPS and World climatic datasets (n=63). High inter-correlations (r≥0.90) are highlighted in blue.

Table 5. 4 Inter-correlation between estimated monthly air temperatures derived from MODIS LST, and World climatic datasets; precipitation CHRIPS and World climatic datasets (n=63). High inter-correlations (r≥0.90) are highlighted in blue.

Month	Maximum temperature	Mean temperature	Minimum temperature	Total precipitation
01	0.93	0.97	0.90	0.66
02	0.92	0.97	0.89	0.54
03	0.90	0.96	0.83	0.58
04	0.88	0.95	0.86	0.85
05	0.86	0.96	0.88	0.88
06	0.89	0.95	0.87	0.91
07	0.86	0.97	0.91	0.95
08	0.88	0.97	0.93	0.91
09	0.88	0.96	0.99	0.94
10	0.89	0.97	0.89	0.71
11	0.84	0.98	0.91	0.74
12	0.91	0.96	0.89	0.62

5.3 Result

In Table 5.5, descriptive statistics of the estimated 19 bioclimatic variables are reported. Spatial maps of 19 bioclimatic variables at a spatial resolution of 1 km for the period 2002-2017 are

shown in Figure 5.2. In this figure, we also provide a comparison between the estimated bioclimatic variables (SatClim) and WorldClim version 2. The results of the statistical analysis are reported in Table 5.6.

For almost all of the 19 SatClim and WorldClim bioclimatic variables, high correlations ($R^2 \geq 0.70$) were revealed in the linear regression between the > 1.5 x 10^6 pairs of values. Only the annual mean diurnal range (02) and Isothermality (03) found lower – but still moderate- correlations ($R^2 = 0.40$-0.46). Only for these two variables did the normalized RMSE (nRMSE) slightly exceed 10%. The nRMSE of the remaining 17 bioclimatic variables were all below 8% (with six variables nRMSE< 4%). The seven precipitation-related bioclimatic variables were generally closer correlated with WorldClim compared to the eleven temperature-related bioclimatic variables (Table 5.6). Examining the consistency of retrieved frequency distribution (WorldClim versus SatClim), we found a generally very similar pattern, often characterized by multi-modal distributions (Figure 5.2, last column).

Table 5. 5 Descriptor statistics of the estimated 19 bioclimatic variables (SatClim) for the years 2002-2017 (n=1 575 107 pixels)

Variables	Maximum	Mean	Minimum	Standard deviation
Annual mean temperature (01)	11.2	1.5	-13.4	3.6
Annual mean diurnal range (02)	12.6	7.3	4.4	1.0
Isothermality (03)	24.6	15.4	9.4	1.9
Temperature seasonality (04)	1960.0	1332.5	686.3	162.8
Maximum temperature of the warmest month (05)	32.2	22.3	-1.1	4.4
Minimum temperature of the coldest month (06)	-11.8	-25.4	-39.8	4.2
Annual temperature range (07)	64.4	47.7	28.3	4.9
Mean temperature of wettest quarter (08)	29.1	18.1	-3.2	4.4
Mean temperature of driest quarter (09)	2.9	-16.6	-31.0	4.9
Mean temperature of warmest quarter (10)	29.1	18.5	-3.2	4.5
Mean temperature of coldest quarter (11)	-7.6	-18.6	-32.0	4.2
Annual precipitation (12)	53.8	15.2	1.7	7.5
Precipitation of wettest month (13)	166.0	51.3	7.2	26.1
Precipitation of driest month (14)	10.0	1.5	0.0	0.8
Precipitation seasonality (15)	192.9	109.3	47.7	10.7
Precipitation of wettest quarter (16)	436.0	139.1	18.4	69.9
Precipitation of driest quarter (17)	30.0	5.3	0.0	2.6
Precipitation of warmest quarter (18)	436.0	133.7	12.3	70.7
Precipitation of coldest quarter (19)	53.6	6.1	0.0	3.1

Together, our results demonstrate that the spatial pattern, value ranges, and frequency distributions of WorldClim were generally well retrieved using the satellite-derived inputs of SatClim. For the two variables annual mean diurnal range and isothermality, the lower correlations can be attributed to the fact that temperature extremes enter the calculations; variables that are generally less well retrieved using satellite-based modeling techniques.

The spatial distribution maps of climatic indices such as Selyaninov's (1966) hydrothermal coefficient, De Martonne's (1925) aridity index, Thornthwaite's (1948) humidity factor, Mezentsev's (1969) moisture index, Ivanov's (1962), climate biological effectiveness indicator, Shasho's (1985; 2005) bioclimatic potential using remote sensing data of air temperature and precipitation over Mongolia has been shown in Figure 5.3 (Details see, Otgonbayar and Sumya 2020).

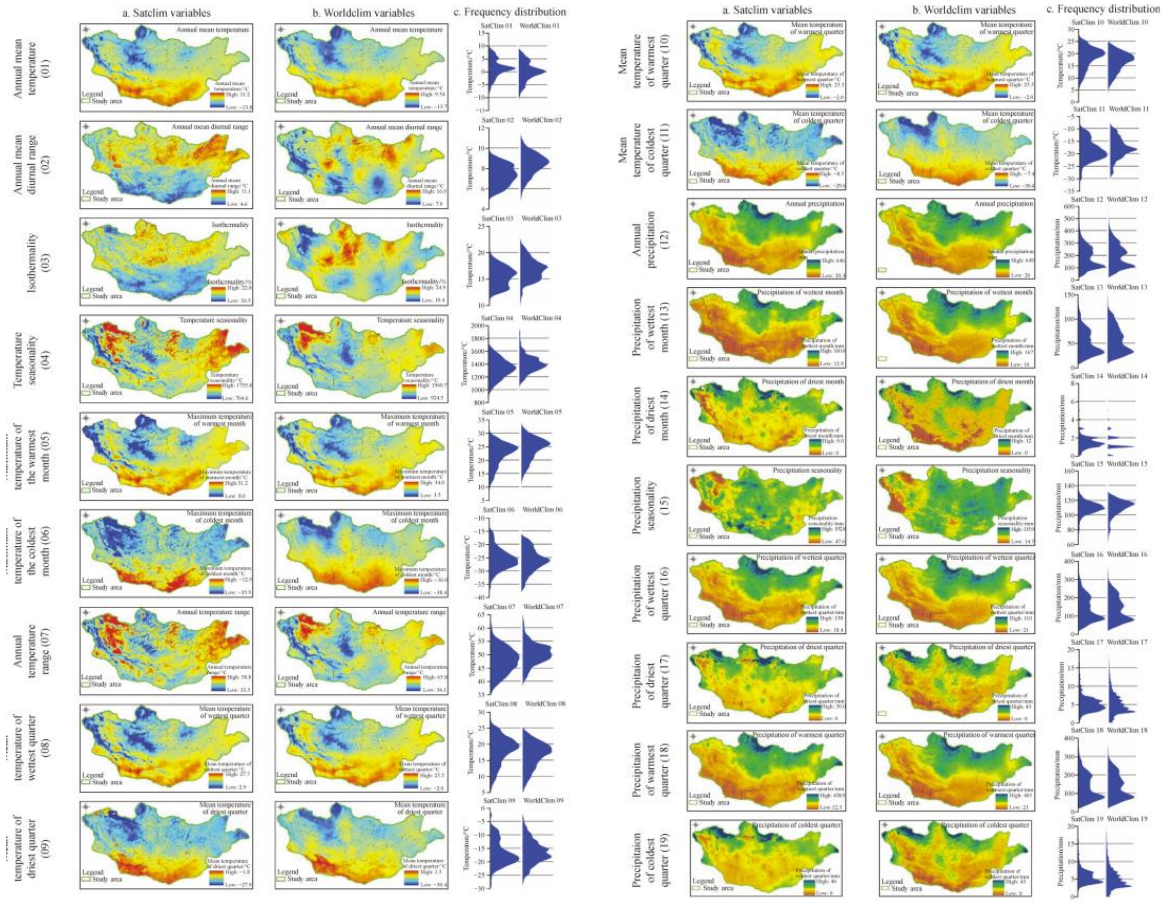


Figure 5. 2 (a) Modeled 19 SatClim bioclimatic variables using MODIS and CHIRPS data 2002–2017, (b) WorldClim variables with gridded data 1971–2000, (c) Frequency distributions of SatClim and WorldClim.

Table 5. 6 Summary of statistics describing the correspondence between SatClim and WorldClim data over Mongolia (R^2, RMSE, and nRMSE). For the comparison, 19 variables were extracted from the full image (n=1 575 107 pixels). High correlations ($R^2 \geq 0.70$) are highlighted in light grey.

Variable	R^2	RMSE	nRMSE (%)
Annual mean temperature (01)	0.97	0.61°C	2.48
Annual mean diurnal range (02)	0.46	0.95°C	11.57
Isothermality (03)	0.40	1.76%	11.59
Temperature seasonality (04)	0.86	61.14°C	4.80
Maximum temperature of the warmest month (05)	0.91	1.29°C	3.88
Minimum temperature of the coldest month (06)	0.76	2.05°C	7.31
Annual temperature range (07)	0.72	2.61°C	7.21
Mean temperature of wettest quarter (08)	0.95	1.00°C	3.10
Mean temperature of driest quarter (09)	0.70	2.69°C	7.95
Mean temperature of warmest quarter (10)	0.94	1.07°C	3.32
Mean temperature of coldest quarter (11)	0.93	1.32°C	5.43
Annual precipitation (12)	0.94	2.70mm	5.19
Precipitation of wettest month (13)	0.90	8.13mm	5.12
Precipitation of driest month (14)	0.73	0.54mm	5.42
Precipitation seasonality (15)	0.76	8.50mm	5.85
Precipitation of wettest quarter (16)	0.92	19.80mm	4.74
Precipitation of driest quarter (17)	0.78	1.65mm	5.50
Precipitation of warmest quarter (18)	0.87	2.20mm	0.52
Precipitation of coldest quarter (19)	0.70	0.04mm	3.64

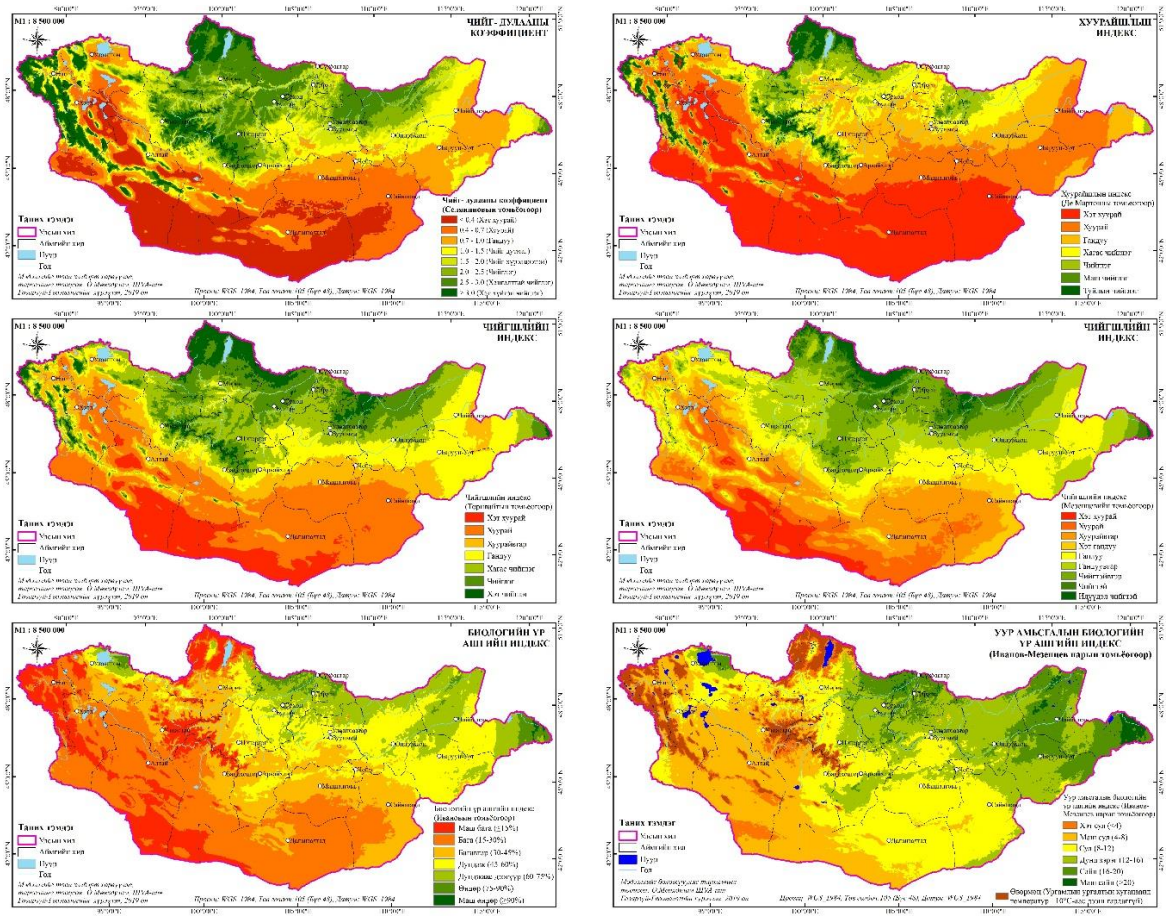


Figure 5. 3 Spatial distribution maps 6 different climatic indices such as hydrothermal, aridity, humidity, moisture, biological effectiveness, and bioclimatic potential over Mongolia.

5.4 Discussion

Bioclimatic variables show information about annual conditions (e.g., 01, 02, 07 and 12), seasonal variations (e.g., 05, 06, 13, 14), and intra-year seasonality (e.g., 08-11, 16-19). These variables are represented as indicators relevant to the physiological restrictions of species and are valuable for several applications (O'Donnell and Ignizio 2012).

In recent years, at the global level, bioclimatic variables mostly have been estimated from two commonly used types of datasets, namely WorldClim datasets (Fick and Hijmans 2017; Marchi et al. 2019), and MERRAclim datasets (Vega et al. 2018). WorldClim version 1 and version 2 are global gridded datasets at a spatial resolution of ~1 km^2. WorldClim datasets are representative of the period 1961-1990, and 1970-2000, respectively (Fick and Hijmans 2017). WorldClim climate datasets, and bioclimatic variables produced by geo-statistical interpolation methods (i.e., kriging and spline).

MERRAclim bioclimatic variables estimated from MERRAclim datasets, which are produced station-based hourly data of air temperature, and specific humidity gridded data (instead of precipitation) from the Modern-Era Retrospective Analysis for Research and Applications Reanalysis (MERRA) using a spline interpolation method for the years 1980, 1990, and 2000 (Gerta et al. 2017). Therefore, the MERRA dataset is a climate reanalysis dataset focused on weather stations and modern remote sensing data. The disadvantage of MERRAclim bioclimatic variables with a coarse spatial resolution (10 arc-minutes, 5 arc-minutes, and 2.5 arc-minutes).

Moreover, several studies (Waltari et al. 2014; Brown and Comrie 2002; Kurtzman and Kadmon 1999; Nikolova and Vassilev 2006) using various interpolation methods including kriging (co, simple, and ordinary), thin plate smoothing splines and inverse distance weighting (IDW) to simultaneously on precipitation and temperature datasets had a different level of success, and generally revealed larger errors for precipitation as compared to temperature (Mesquita and Sousa 2009). Interpolation errors commonly depend on the spatial and temporal resolution of recorded data (Otgonbayar et al. 2019). Moreover, the uncertainty of the interpolation-based method was rising the time and asymmetry difference between future studies and interpolation-based climate datasets (e.g., WorldClim) might lead to unsuitable predictions (Amiri et al. 2020). On the contrary, satellite-derived data are continuous in spatial and temporal coverage. Moreover, real-time access to satellite data has led to the creation of more up-to-date climate data (Amiri et al. 2020).

To remedy those limitations, estimated bioclimatic variables (SatClim) using MODIS LST and CHIRPS data for the years 2002-2017. For this analysis, estimated monthly maximum, mean, and minimum air temperature from Terra MODIS satellite LST (MOD11A2) for the period 2002-2017 using the Random Forest (RF) regression model and three predictors (Otgonbayar et al. 2019; Otgonbayar and Sumya 2020). We examined the relationship between SatClim and WorldClim bioclimatic variables version 2.0 for the entire territory of Mongolia using coefficient of determination (R^2), root mean squared error (RMSE), and normalized root mean squared error (nRMSE), which represent spatial correlation (association) and error (residual) (Richter et al.2012).

In general, and considering WorldClim as a "reference", the spatial pattern of all 19 bioclimatic variables were well retrieved from MODIS and CHIRPS data and had moderate to highly positive correlations, with similar (often multi-modal) frequency distributions. The lower performance of the two variables annual mean diurnal range (02) and Isothermality (03) can be attributed to the fact that temperature extremes enter into their calculation. These temperature extremes are often underestimated using satellite-derived input data (Janatian et al. 2017; Duan et al. 2018; Hooker et al. 2018). Other 17 variables were estimated with normalized RMSE of < 8% with six of the 17 variables nRMSE <4% (Table 5.6).

Amiri et al (2020) estimated 19 bioclimatic variables from temperature and precipitation instrumental records (Model 1), and remote sensing data (Model 2) at a resolution of 1 km during 2001-2017 in Isfahan province of Iran together with three topographic variables using five different regression models. Accuracy statistics in Model 2 were higher than in Model 1. This study proved that bioclimatic variables derived from the satellite were more effective.

The success of our satellite-derived method can be attributed to the fact that precipitation and temperature can be relatively well retrieved remotely (Li et al. 2013b; Paredes-Trejo et al. 2017; Kidd et al. 2010; Funk et al. 2015; Beck et al. 2017; Sun et al. 2018), and especially in highly elevated or mountainous areas (Fick and Hijmans 2017). In those areas, spatially and temporally continuous grids of land surface temperature (LST) are valuable inputs for accurate and robust air temperature retrievals with monthly resolution (Otgonbayar et al. 2019). Similarly, by observing cloud top temperatures, it is possible to model monthly precipitation fields with relatively high accuracy (Bai and Liu 2018). Without Earth Observation (EO) data, these primary variables have to be modeled and/or interpolated from

sparse station data, often not capturing well local peculiarities (Atzberger and Rembold 2013; Vancutsem et al. 2010; Benali et al. 2012).

5.5 Conclusion

Spatial maps of 19 bioclimatic variables at a spatial resolution of 1 km were generated for the entire territory of Mongolia, representing the period 2002-2017. The analysis used two different satellite time series data: MODIS LST, and CHIRPS. To estimate monthly maximum, mean, and minimum air temperature the RF regression model was used with time series of LST as a predictor variable. Monthly total precipitation data were obtained from CHIRPS version 2.0. Seventeen bioclimatic variables derived from MODIS and CHIRPS data had a strong positive correlation with the WorldClim bioclimatic variables, and their frequency distributions were close. Two variables were the lower performance as an annual mean diurnal range (02) and Isothermality (03) can be attributed to the fact that temperature extremes enter into their calculation. These temperature extremes are underestimated by applying satellite-derived input data (Janatian et al. 2017; Duan et al. 2018; Hooker et al. 2018). As a consequence of the successful retrieval of the bioclimatic variables, we are confident that the estimated 19 bioclimatic variables will be very useful for a range of applications, in particular, if a higher spatial resolution is required such as for species distribution modeling.

The success of the modeling can be attributed to the fact that climatologies of both air temperature, as well as precipitation, can be well retrieved from EO data, in particular, if aggregated over monthly intervals and for regions such as Mongolia. In areas with sparse station density, EO data avoids otherwise necessary interpolation techniques.

The main limitation of many EO products relates to the fact that data sets are still relatively short (e.g., MODIS LST starting only in 2002) and that data from multiple satellites would have to be combined and normalized if longer time series are required. The advantage of the MODIS data set is however that it covers the most recent 15 years. In the future, spatial and temporal resolution and spatial coverage will favor EO data even more than other techniques as new satellites are launched at an unprecedented pace. For future research, we recommend to focusing on the improved quality, spatial, and temporal resolution of precipitation estimates. For the climatic indices, spatial distribution maps of the aforementioned six indices in Mongolia were generated. For Mongolia, generating 6 climatic indices was significantly important to understand an aridity status.

CHAPTER 6. Land potential assessment in Mongolia using linear and non-linear regression models and a time series of environmental variables

6.1 Introduction

Linking land use with land potential (LP) is one of the most efficient strategies to limit land degradation and is vitally important in nature conservation, human development, and economic growth (Herrick et al. 2013; IRP 2016). From our findings, outstanding publications on LP were written by Herrick et al (2013), Herrick et al (2016), and Herrick et al (2019). Additionally, the scientific report "Unlocking the sustainable potential of land resources systems strategies and tools", developed by the United Nations Environmental Program (UNEP), furnished a very beneficial history fabric for LP assessment (IRP 2016). The term LP has been applied in the field of land degradation since 2013. The International Resource Panel (IRP) described LP as the " … inherent, long-term potential of the land to sustainably generate ecosystem services" (IPR 2016). It is thus generally agreed that the LP contains three elements: 1) potential production, 2) degradation resistance, and 3) resilience which is the capacity to recover from degradation (Herrick et al. 2013).

LP is significantly linked to the three aforementioned variables (Liebig et al. 2017; Herrick et al. 2019). How these variables influence each other and how to assess LP using these variables is still a challenge despite the important research done since the early days of the land evaluation systems (Trudgill and Briggs 1979). Notwithstanding this, as a result of the contemporary development of land evaluation systems, Earth Observation (EO), and geospatial technologies (Dong et al. 2019; Mariathasan, Bezuidenhoudt, and Olympio 2019), it is nowadays possible to address these challenges using advanced technologies, such as the cloud-based (Google Earth Engine-GEE) analysis of time-series of spectral and environmental variables and machine learning algorithms (Gonzalez-Roglich et al. 2019; Teich et al. 2019). To assess the LP, we can now investigate natural elements and processes, develop complex system simulation models, and analyze the coupling mechanism of natural and human factors and the dynamic change of the land surface system. At the same time, the potential of the newest generation of EO satellites is leveraged together with modern machine learning techniques for information extraction.

Increasing awareness and understanding of LP is valuable in describing where production can be sustainably raised and determining land that could/should be restored. Assessment of LP can also assist decision-makers in making knowledge-based decisions about

various land use alternatives. Any progress in this area will also contribute to achieving Land Degradation Neutrality (LDN) (Kust, Andreeva, and Lobkovskiy 2020; Liniger et al. 2019), and Sustainable Development Goal (SDG) 15.3 (Giuliani et al. 2020; Sims et al. 2019). LDN is "a condition where further land degradation is prevented and already degraded land can be restored" (UNCCD 2016; Cowei et al. 2018).

Assessment of LP is a complicated process, and this kind of study was lacking in Mongolia. Previous studies by Otgonbayar et al (2018), and Avirmed et al (2020) were focused to estimate landscape ecological potential. These studies used a combination of the Multi-Criteria Decision-Making (MCDM) method with Analytical Hierarchy Processes (AHP) to assess landscape-ecological potential. The current study aim is to assess LP in Mongolia using a time series of environmental variables, and linear-nonlinear regression models. The specific objectives are as follows:

- To analyze the relationship between environmental variables (topography, climate, soil, and vegetation) for a better understanding of their interactions
- To assess the importance of environmental variables in LP assessment using explanatory and objective variables
- To develop a prediction model to assess LP and to generate a spatial map of LP in Mongolia
- To identify the interaction between LP and pasture use to better understand the current situation of LP

6.2 Methodology and data

6.2.1 Methodology

Random Forest Regression (RFR) model

To develop an LP assessment model and to examine the importance of the 25 explanatory variables, the RFR model was applied. We used the IncNodePurity method from the package "randomForest" (Breiman 2001; Liaw and Wiener 2002) in RStudio (R core team 2022). The Random Forest (RF) is a non-linear, machine learning algorithm based on an ensemble method developed by Breiman (2001), which is "capable of modeling discrete or continuous datasets" (Pal 2005). The regression and classification models of the RF include the top-down approach to find optimal binary node splits by locally lowering variants at the terminal. The main benefits of the RFR model are 1) the distribution of the variable need not be normally distributed, 2) a

big, multi-dimensional, and strongly inter-correlated dataset can be modeled efficiently, 3) it prevents overfitting of the model, 4) performance measures can only be calculated using OOB error data and 5) it detects the significance of each explanatory variable in the model (Immitzer et al. 2016; Belgium and Dragut 2016; Hudak et al. 2008). In other words, it is possible to know which input variables are more effective (Otgonbayar et al. 2019). The RF algorithm measures two parameters: the importance of the explanatory variables and the internal structure of the variables (Breiman 2001; Liaw and Wiener 2002). To perform the RFR model sets two parameters some decision trees "ntree", and the number of variables to select at each split of tree "mtry" are used. For detecting the importance of the variables "Percent Increase in the Mean Square Error (%IncMSE)" and "Increase in Node Purity (IncNodePurity)" are calculated.

Comparison of linear and nonlinear regression models

To validate estimated importance variables in the RFR model and find the advantages and differences of the RFR model, three different regression models were compared: Partial Least Square Regression (PLSR), Principal Component Regression (PCR), and Classification and Regression Tree (CART). Two of the three regressions were classical linear regression models (PLSR, PCR), and the remaining one was non-linear machine learning (ML) regression model (CART). In other words, this comparison was done to detect the differences between classical linear regression and the ML regression models. Moreover, to find out the inter-differences between linear regression, the PLSR and PCR models were compared. In contrast, to find out inter-differences between non-linear ML regressions the CART regression model was used. The details of the PLSR, PCR, and CART can be found in Abdi (2003), Lasaponara (2006), and Breiman et al (1984), respectively.

Method for calculation of livestock grazing

To link pasture use with LP, we estimated livestock grazing capacity (LGC). LGC expresses the number of livestock that can be grazed on one hectare of pasture for a given period without adversely affecting the growth, development, and regeneration of pasture plants (IMALGC 2019). LGC- livestock grazing capacity, N- total number of livestock population in sheep unit (SU) per hectare, and PCC- pasture carrying capacity (Eq.2). Standard units were used to convert the number of five types of livestock to sheep unit: 1 goat = 0.9 SU, 1 sheep = 1 SU, 1 camel = 5 SU, 1 cattle = 6 SU, and 1 horse = 7 SU (IMALGC 2019).

$$LGC = \left(\frac{N}{PCC}\right) \times 100\% \quad (1)$$

$$PCC = \frac{BM_1}{BM_2 D} \quad (2)$$

PCC- pasture carrying capacity in SU, BM_1- the actual biomass of pasture (kg ha^{-1}), BM_2- livestock daily consumption grass (kg day^{-1}), which is an average of 1.64 kg day^{-1} SU^{-1} during the summer-fall, and 1.32 kg day^{-1} SU^{-1} during winter-spring (IMALGC 2019). D-grazing time (106 days during summer, 56 days during fall, and 150 days during winter and spring).

Model evaluation

Widely applied 4 statistics including coefficients of correlation (r) and determination (R^2), root mean square error (RMSE), and normalized root mean square error (nRMSE) were used. To detect the relationship between objective and explanatory variables used r. And the R^2 and RMSE were used to assess the accuracy of the RFR, PLSR, PCR, and CART regression models (Richter et al. 2012; Barzeger et al. 2016). The nRMSE was used to detect a normalized error of RFR. The r is a measure of the relationship between variables, R^2 determines the proportion of the explained variance, RMSE represents the residual error, and nRMSE expresses range normalized errors.

6.2.2 Data

This study, used 29 environmental variables of which one objective variable, 25 explanatory variables, and 3 validate variables derived from satellites, reanalyzed datasets, field surveys, and a statistical database. The objective variable was rangeland recovery data from 12,988 sites obtained from the Agency for Land Administration and Management, Geodesy, and Cartography (ALAMGaC). 25 explanatory variables related to topography, climate, soil, and vegetation were derived from the Shuttle Radar Topographic Mission (SRTM), Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Very High-Resolution Radiometer (AVHRR), Climate Research Unit-Time Series (CRU-TS), Climate Research Unit-National Centers for Environmental Prediction (CRUNCEP) and the Trends. Earth system. 16 of the 25 spatial explanatory variables were time-series data for the period 1991-2021, 1991-2015, and 2002-2021. The remaining 3 validated variables were the number of livestock and in situ meteorological datasets such as temperature and precipitation from the National Statistical Office (NSO) of Mongolia, and the Information Research Institute of Meteorology, Hydrology, and the Environment (IRIMHE) of Mongolia, respectively. 23 explanatory variables were re-projected (WGS84) and resampled (n=19567425) at a spatial

resolution of 500 m (excluding latitude and longitude). Each variable's characteristics and source are shown in Table 6.1. The descriptor statistics of all variables are reported in Table 6.2.

Table 6. 1 Data sources and their characteristics

Data name	Data type	Path/row or tile name	Spatial resolution	Temporal coverage	Source
DEM	SRTM satellite	Extracted from global data	90 m	-	https://worldclim.org
Temperature, Precipitation, Potential evapotranspiration , Vapor pressure	Reanalysis data from CRU-TS v4.06	Extracted from global data	0.5×0.5°	1991-2021	https://crudata.uea.ac.uk
Radiation, Wind	Reanalysis data from CRUNCEP v07	Extracted from global data	0.5×0.5°	1991-2021	https://downloads.psl.noaa.gov
Climatic indices (Hydrothermal, Aridity, Humidity, Moisture)	Estimated from CRU-TS v4.06	-	0.5×0.5°	1991-2021	-
Temperature, precipitation	In situ-metrological data	63 weather stations	-	1991-2021	IRIMHE 2022
Soil Organic Carbon	From Trends.Earth	Extracted from global data	250 m	-	https://docs.trends.earth
Soil moisture	Reanalysis of data from ERA	Extracted from global data	0.28×0.28°	1991-2021	https://esa-soilmoisture-cci.org
Soil humus	Vector data	-	-	-	IGG, MAS
NDVI (MCD13) Fpar (MCD15) LAI (MCD15) GPP (MOD17)	MODIS satellite	23-25/03-04, 26/04	250 m 500 m 500 m 500 m	2002-2021	https://ladsweb.modaps.eosdis.nasa.gov
NDVI (GIMMS3g)	AVHRR satellite	Extracted from global data	8 km	1991-2015	https://earthexplorer.usgs.gov
Field-measured objective data	Archived data from field measurements	308 sites 2115 sites 2625 sites 3200 sites 4740 sites	- - - - -	2016 2017 2018 2019 2020	ALAMGaC
Number of livestock	Statistical data	332 soums (sub-province)	-	1991-2021	NSOM

Table 6. 2 The summary statistical description of all variables used in this study

Variable name	Variable acronym	Units	Number of pixels	Minimum	Maximum	Mean	Standard deviation
Elevation	Elv	m	19567425	524.00	4320.00	1488.51	539.27
Slope	Slp	Degree	19567425	0.00	62.18	3.84	5.07
Aspect	Aspct	Degree	19567425	-1.00	359.94	172.34	106.18
Topographic Position Index	TPI	-	19567425	-543.00	647.73	0.02	42.83
Topographic Wetness Index	TWI	-	19567425	-7.47	17.46	-1.21	3.38
Latitude	Lat	Degree	19567425	41.58	52.14	n.a	n.a
Longitude	Long	Degree	19567425	87.74	119.93	n.a	n.a
Solar radiation	Srad	kJ m^{-2}	19567425	2467.56	3332.14	2930.24	158.12
Air temperature	Tmp	°C	19567425	-18.26	8.86	-0.61	3.83
Precipitation	Prec	mm	19567425	14.81	559.93	204.87	95.76
Potential evapotranspiration	PET	mm	19567425	8.76	40.75	28.29	5.94
Vapor pressure	Vap	hPa	19567425	0.10	7.31	5.05	0.84
Wind	Wnd	ms^{-1}	19567425	1.01	5.04	2.92	0.83
Hydro-Thermal coefficient	HTC	-	19567425	0.07	14.00	1.59	1.23
Aridity index	I_{ar}	-	19567425	1.00	56.00	3.55	6.36
Humidity index	HF_{th}	-	19567425	2.00	487.99	26.28	17.13
Moisture index	MI	-	19567425	0.08	2.12	0.63	0.31
Soil Organic Carbon	SOC	tons C ha^{-1}	19567425	0.00	608.00	106.84	48.76
Soil humus	sHumus	%	19567425	0.00	20.00	3.62	4.54
Soil moisture	sMoisture	m^3m^{-3}	19567425	672.21	4696.93	1795.74	557.72
MODIS NDVI	modisNDVI	-	19567425	-0.07	0.87	0.33	0.21
GIMMS NDVI	gimmsNDVI	-	19567425	0.00	0.69	0.18	0.11
Fraction of Photosynthetically Active Radiation	Fpar	%	19567425	0.00	0.92	0.27	0.23
Gross Primary Productive	GPP	Kg C m^{-2}	19567425	0.00	0.07	0.01	0.01
Leaf Area Index	LAI	m^2plant m^{-2} ground	19567425	0.00	5.59	0.66	0.72
Field-measured objective data	RefData	-	12988 sites	1.00	5.00	n.a	n.a
Air temperature	Ta	°C	63 weather stations	-30.58	21.62	-	14.05
Precipitation	Precip	mm		0.00	145.50	22.39	29.42
Number of livestock	Lvstk	Million heads	332 soums	23.63	71.10	39.96	15.60

Field-measured rangeland recovery data

This study used field-measured rangeland recovery data as an objective variable to develop an LP assessment model. The rangeland recovery data (RefData) were observed in 12,988 photo-monitoring sites and obtained from the ALAMGaC for the period 2016-2020. These monitoring sites aim to assess rangeland health at the national level (GGP-SDC 2015). The RefData were based on reference data of pasture state/ecological potential (the health state of

plant community in the given area), the process of recovery which is divided into 5 classes (Table 6.3). Each rangeland recovery class includes “leaf canopy cover, species composition, and basal gaps of perennial vegetation, vegetation height, and vegetation biomass" information (Densambuu et al. 2018). “Class I expresses that natural condition is very good in that the plant community is not changed or near reference conditions (not degraded); Class II expresses that the plant community is very low changed but can quickly be recovered within a short period with favorable climate conditions; Class III expresses that the plant community is changed but can be recovered within 5-10 years; Class IV expresses that plant community is strongly changed and usually costly to recover; and Class V expresses that with poor plant community and soil degradation. It is impractical to recover” (GGP-SDC 2015). Rangeland recovery data is represented by the LP because it is the capacity to resist and recover from land degradation. The field monitoring sites are shown in Figure 6.1.

Table 6. 3 Field-measured rangeland recovery data (RefData) used, being developed by the Ministry of Food and Agriculture (MoFA), ALAMGaC, IRIMHE, and Associations of Pasture Users Group (APUG) of Mongolia in 2015 (GGP-SDC 2015)

Class type	Description
Class I	*“The plant network is at or close to reference conditions (non-degraded) or calls for 1-three developing seasons for recuperation from minor changes (slightly degraded); match-stocking rate to forage supply and use temporary seasonal deferment as needed” (GGP-SDC 2015)*
Class II	*“The plant community is altered and may be rapidly recovered (3-5 growing seasons) with favorable climatic conditions or a change in management (e.g., stocking rate reduction, seasonal deferment, rotation). The nature of alteration is not regarded as a significant long-term threat to the provision of forage and other ecosystem services” (GGP-SDC 2015).*
Class III	*“The plant network is altered and can take 5-10 developing seasons to get better with modified management (stocking price reduction, seasonal deferment, and long-time period rest). Alteration represents a significant loss of important ecosystem services (and are related to anthropogenic drivers), but recovery is possible in time” (GGP-SDC 2015).*
Class IV	*“The plant community is altered due to the local loss of key plant species, invasion of noxious plant species, or alteration of hydrology that is unlikely to be recovered for over a decade to many years without extensive interventions inclusive of species removal, seeding, or manipulations to get better historic hydrological function an ecological threshold was crossed). Previous ecosystem services have been lost and are usually costly to recover” (GGP-SDC 2015).*
Class V	*“The plant community is altered due to extensive soil loss, accelerated erosion rates, or salinization. Altered plant-soil feedbacks or permanent changes in the soil profile maintain the degraded state. Previous surroundings offerings had been misplaced and it also includes impractical to get better them (frequently seemed like proper desertification). The recovery class hypothesizes timelines for the recovery of the reference state based on vegetation cover and composition data interpreted according to expert knowledge and existing studies when available. For example, the presence of remnant perennial grasses suggests that recovery can occur within several years. The recovery classes allow standardized interpretations across multiple state and transition models to allow for reporting and visualization of rangeland restoration needs” (GGP-SDC 2015).*

The ALAMGaC is responsible for the photo-monitoring network sites (PNS) including rangeland recovery data to control the impact of pasture utilization. PNS is based on the photo point monitoring (PPM) method which indicates whether the vegetation cover of the majoring species is at a feasible level and obtains information on how the structure of these

majoring species is changing (Booth and Cox 2008). The land managers in 330 soums (sub-provinces) have been collecting yearly primary data on key vegetation species, vegetation biomass, vegetation cover, and vegetation recovery class since 2016 using the standardized PPM method. These field measurements are primary datasets examined by the land manager of ALAMGaC and were then entered into the national photo-monitoring database. Currently, the PNS of the ALAMGaC covers 4740 monitoring sites (Figure 6.1). All sites represent natural zones, administrative units, and seasonal pasture regions of Mongolia (ALAMGaC 2021).

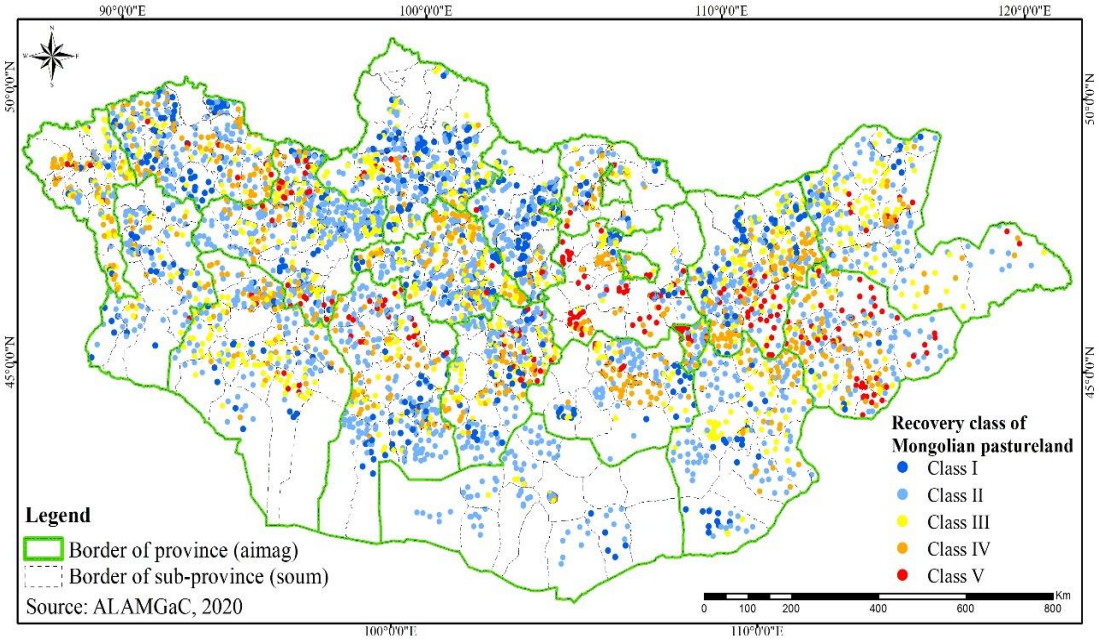


Figure 6. 1 The distribution of the pasture monitoring sites across Mongolia from ALAMGaC in 2020.

Remote sensing and reanalysis of data

Topographic data: To show the contribution of topographic variables in the assessment of the LP 7 variables were chosen of which 5 were widely used terrain parameters such as elevation (Elv), slope (Slp), aspect (Aspct), topographic position index (TPI), and topographic wetness index (TWI). The remaining two were the location of latitude (Lat), and longitude (Long). The Elv, Slp, Aspct, TPI, and TWI were calculated from STRM digital elevation model (DEM) at a resolution of 90 m. The Lat and Long were extracted from the central point of the gridded vector data.

Climatic data: To investigate the contribution of climate to the LP 10 variables were applied. These variables were obtained from CRU-TS, CRUNCEP, and in situ-weather stations (Table

6.2). Specifically, the time series of the monthly mean air temperature (Tmp), monthly total precipitation (Prec), monthly mean capability evapotranspiration (PET), and month-to-month suggest vapor pressure (Vap) derived from the CRU-TS version (v) 4.06 monthly climate dataset for the period 1991-2021 (Harris et al. 2020) (Appendix B, Figure B.1). Annual mean solar radiation (Srad), and wind (Wnd) were derived from Climate Research Unit-National Centers for Environmental Prediction (CRUNCEP) v07 (Viovy 2018). Both CRU-TS and CRUNCEP were reanalysis climate datasets at a resolution of 0.5º for the period of 1991-2021 (Harris et al. 2020). The climate index efficiently expresses the interaction of hydro-thermal conditions over large areas (Croitoru et al. 2013; Voropay and Ryazanova 2018). Therefore, we used 4 different climatic indices in this study, Selyaninov's hydro-thermal coefficient (HCT), De Martonne's aridity index (Iar), Thornwaite's humidity index (HFth), and Mezentsev's moisture index (MI) (Table B.1) based on the literature review (Nyamtseren, Feng, and Deo 2018; Natsagdorj, Munkhbat, and Gomboluudev 2019; Otgonbayar, Sumiya, and Tovuudorj 2021a; Otgonbayar, Erdenedalai, and Dalantai 2017). In order to examine the relationship between reanalysis and in situ meteorological datasets, we applied monthly mean near-surface air temperature (Ta) (Figure B.2 (left)) and monthly mean precipitation (Precip) data (Figure B.2 (right)) obtained from 63 weather stations (Figure 1.1(a)) for the period 1991-2021. Monthly mean Ta and Tmp had strong correlations with the coefficient of correlation (r) being higher than 0.92 (Figure B.3 (top)). Corresponding months of Precip and Prec had a good correlation with r=0.58-0.84 (Figure B.3 (bottom)).

Soil data: To assess the contribution of soil to LP 3 variables were used soil organic carbon (SOC), soil moisture (sMoisture), and soil humus (sHumus). The SOC variable was extracted from the Trends.Earth system (Conservation International 2022). Trends.Earth is an open-access system for estimating land degradation indicators (Gonzalez-Roglich et al. 2019). The SOC facts have been produced through the International Soil Reference and Information Centre's (ISRIC) SoilGrids at a spatial resolution of 250 m and with a depth layer of 0-30 cm (ISRIC 2022). For soil moisture data, five different global reanalysis datasets (Li et al. 2021) have been widely used in climate assessment and terrestrial environmental studies (Gallego-Elvira et al. 2016). From these, the fifth generation of ERA's sMoisture has a relatively higher resolution compared to others therefore, these data were used. The soil moisture data were extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) database at a resolution of 0.28º for the period 1991-2021 (Figure B.4). The sHumus data over Mongolia

were obtained from the Institute of Geography and Geoecology, Mongolian Academy of Sciences (IGG, MAS).

Vegetation data: To assess the contribution of vegetation to PL, 5 variables Normalized Difference Vegetation Index (modisNDVI), Fraction of Photosynthetically Active Radiation (Fpar), Leaf Area Index (LAI), and Gross Primary Productive (GPP) of MODIS and the third generation (3g) of Global Inventory Modeling and Mapping Studies (GIMMS) NDVI (gimmsNDVI) were used (Figure B.5). More specifically, gimmsNDVI data were derived from the AVHRR sensor data of the National Oceanic and Atmospheric Administration (NOAA) at a spatial and temporal resolution of 8 km and a bi-monthly composite, covering the period from 1982 to 2015 (Pinzon and Tucker 2014). The original gimmsNDVI3g datasets were developed by Tucker (2005). For the data analysis, bi-monthly gimmsNDVI3g datasets were used covering a 25-year (1991-2015) corresponding with the vegetation growing season (April-August). For the processing of the datasets applied the "GIMMS" R package (Detsch 2016). The remaining four vegetation variables were derived from MODIS collection 006 (c006) datasets for the period of 2002-2021 during the vegetation-growing season (April-August) in a sinusoidal projection acquired from the Land Processes Distributed Active Archive Center (LP DAAC 2022). Specifically, MCD13A2 (combined MYD13A2 and MOD13A2) NDVI datasets were extracted from the MODIS Aqua and Terra satellites at a spatial resolution of ~250 m. MCD15A2 (combined MYD15A2 and MOD15A2) Fpar and LAI datasets were extracted from the MODIS Aqua and Terra satellites at a spatial resolution of ~500 m. MOD17A2 GPP datasets were extracted from the MODIS Terra satellites at a spatial resolution of ~500 m. For the data processing download, mosaic, re-sample, reprojection, image mask, and crop were used with the "MODIS" R package (Matteo et al. 2016).

Livestock data

To compare the current state of land use with LP, used the number of livestock at the soum (administrative unit) level of Mongolia from 1991 to 2021 was obtained from the NSO of Mongolia, as counted yearly in December (NSOM 2022b). Every December, the officers of each soum take a census of five types of livestock, which are camels, horses, cattle, sheep, and goats. The yearly census of livestock data is entered into the database of NSO of Mongolia.

6.3 Results

6.3.1 The interrelationship between environmental variables

The relationship between topographic variables

We selected Elv, Slp, Aspct, TPI, TWI (Figure B.6), Lat, and Long to analyze the contribution of the topographic variables to LP. These variables have been widely applied to investigate spatial influences on the hydrological process and to determine the hydrological flow directions and paths, as well as describe biological and physical processes such as growing native vegetation processes and net primary production (NPP) (Makhaya et al. 2022; Blackburn et al. 2022). Primarily, examined whether there is any correlation between the topographic variables (n=19567425). The correlation analysis showed (Figure 6.2) that the highest positive correlation (r=0.55) was detected between elevation and slope. In contrast, the highest negative correlation (r=-0.63) was detected between elevation and longitude. The lowest correlations were detected between aspct and the remaining 6 variables. The significance levels (Table B.2) of Elv, Slp, TPI, and TWI noted $p<0.0001$ for each pair of variables. The Elv, Slp, Lat, and Long noted $p<0.0001$ for each pair of variables. The between Aspct and Elv, Aspct and TPI, Aspct and Lat, TPI and Lat, TPI, and Long ranged $p<0.01$, and $p<0.1$, respectively (Table B.2). The corresponding frequency distributions are characterized by two left skew, a bimodal, and two normal distributions, respectively (Figure 6.2). The spatial relationship between three topographic variables (Elv, Slp, and Aspct) showed that a high value of Slp was associated with a high value of Elv (Figure B.7).

From these analyses, the main features of the topographic variables were detected in the western, central, and northern parts of Mongolia which are primarily mountainous (Figure B.7). In the east and southeastern areas, which are dominated by flat plains, undulating plains, and hollows, no significant relationship was detected (Figure B.7).

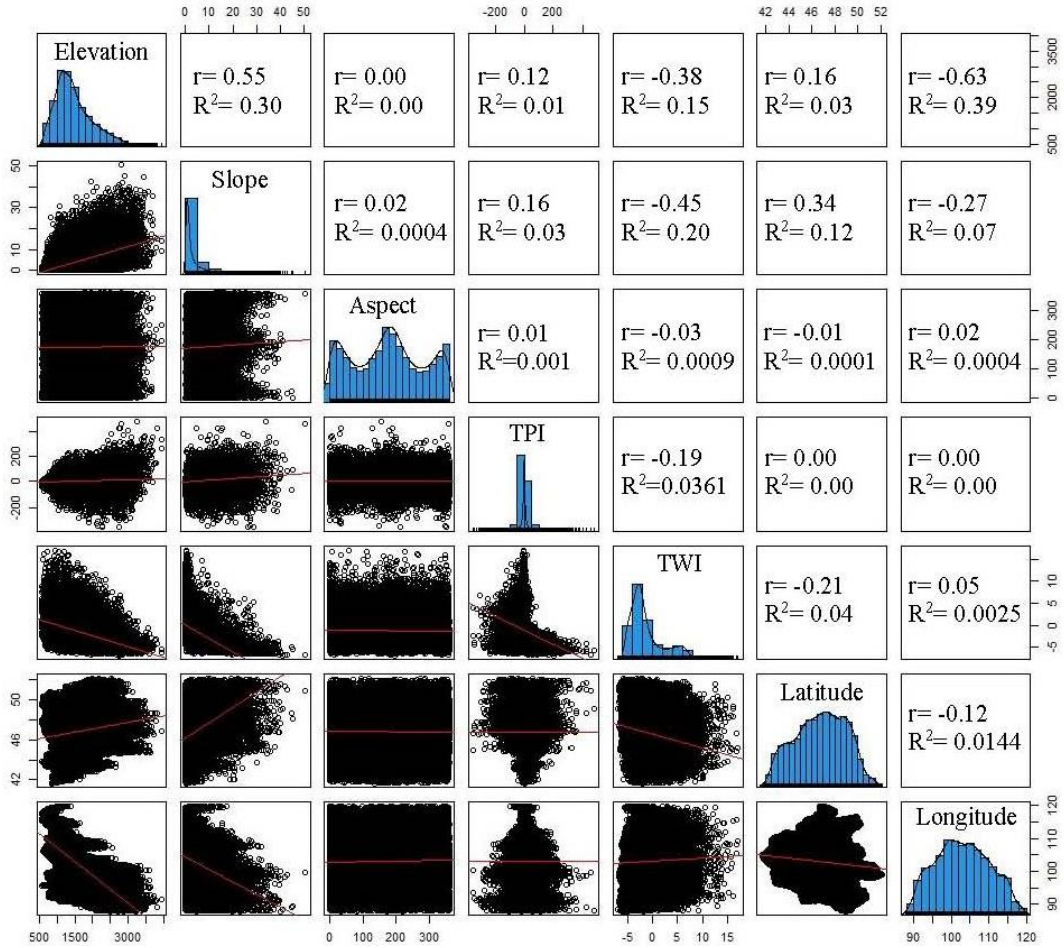


Figure 6. 2 The results of the inter-analysis correlation between topography variables (n=19567425). The diagonal plots represent a histogram and distribution density, the top right plots represent a coefficient of determination (R^2) (Pearson), and the bottom left plots are represented scatter plots for a linear model distribution.

The relationship between climatic variables

We selected 10 variables to analyze the contribution of climate variables to LP (Figure B.8). These variables play an important role in the effective explanation of the interaction between the atmosphere and the land system because many land processes are regulated by climatic parameters (Robert et al. 2020). In other words, climatic variables have a strong effect on LP (IRP 2016). More specifically, the heat and areas that lack moisture limit vegetation growth, while low temperatures also limit vegetation growth (Ólafur 2019). These limitations occur in drylands, cold tundra, and the high mountain regions in Mongolia. We examined whether there is any correlation between the climatic variables (n=19567425). The coefficients of r and R^2 between climatic variables ranged from -0.80 to 0.96 and from 0.14 to 0.92 (Figure 6.3), respectively.

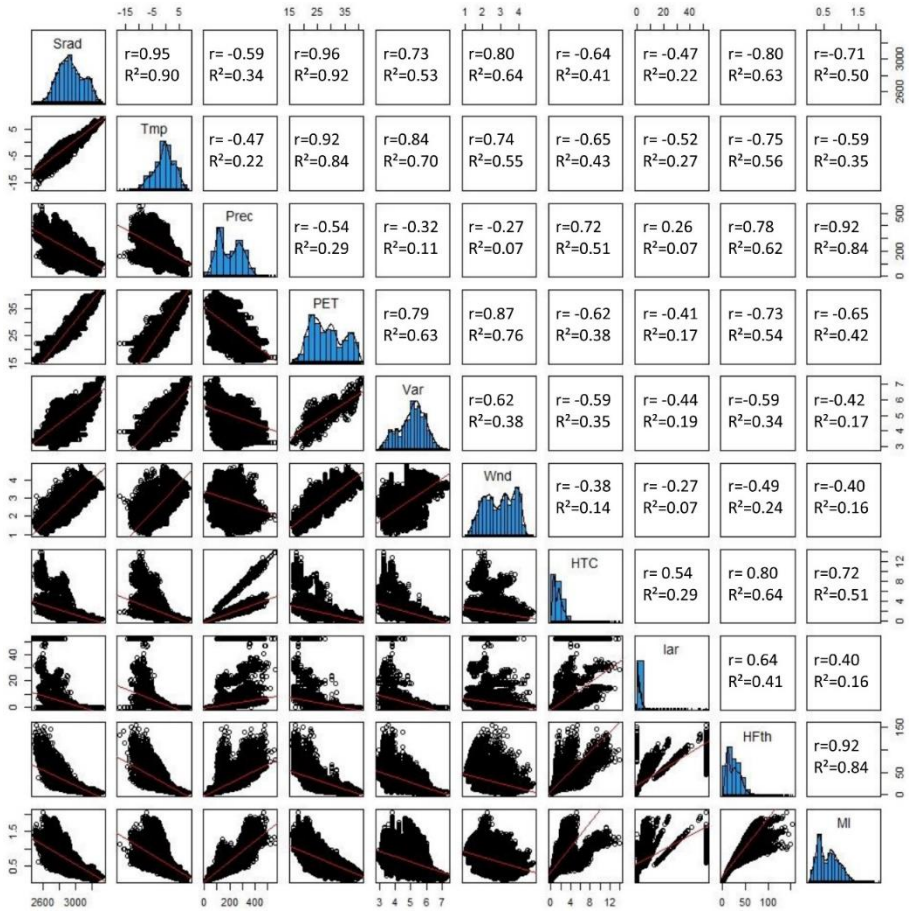


Figure 6. 3 The results of inter-correlation analysis between climatic variables (n=19567425). The diagonal plots represent a histogram and distribution density, the top right plots represent R^2 (Pearson), and the bottom left plots represent scatter plots in a linear model distribution.

The significance levels noted that $p<0.0001$ for all pair variables. Specifically, strong correlations ($R^2=\geq 0.60$) were detected between Srad and Tmp ($R^2=0.90$); Srad and PET ($R^2=0.92$); Srad and Wnd ($R^2=0.64$); Srad and HFth ($R^2=0.63$); Tmp and PET ($R^2=0.84$); Tmp and Var ($R^2=0.70$); Prec and MI ($R^2=0.84$); Prec and HFth ($R^2=0.62$); PET and Vap ($R^2=0.63$); PET and Wnd ($R^2=0.76$); HTC and HFth ($R^2=0.64$); HFth and MI ($R^2=0.84$). The abovementioned variables were statistically significant (Table B.3). The frequency distributions of yearly mean Srad, Tmp, Prec, PET, Vap, and Wnd were nearly designed, which was mainly characterized by bi-model and multi-model. The frequency distribution of yearly mean climatic indices such as HCT, Iar, HFth, and MI, were characterized by left skew (Figure 6.3). The correlation analysis detected that 24 of the 45 pair variables had a significant relationship ($R^2=\geq 0.60$): Srad (4 times), PET (4 times), HFth (4 times), Tmp (3 times), Prec (2 times), MI (2 times), Vap (2 times), Wnd (2 times) and HCT (1 time). According to this, the

influence of solar radiation, evaporation, and dryness are significant in the interaction of climate variables in Mongolia.

The relationship between soil variables

We selected 3 variables (SOC, sHumus, and sMoisture) to analyze the contribution of the soil variables to LP (Figure B.9). These are critical variables for determining LP. Specifically, high concentrations of SOC and sHumus have many positive influences on a net primary product, food production, soil aggregate stability, and pesticide sorption potential, to reduce erosion and pollution of soil (ISRIC 2002). Moreover, the water storage capacity of soil is a key indicator for determining LP in drylands (Trudgill and Briggs 1979). Water Storage in Soil (WSS) provides water to plants during dry periods (Seybold, Herrick, and Brejda 1999). The potential of the land is limited in arid climates where WSS is required. We examined whether there is any correlation between the soil variables (n=19567425).

The correlation analysis showed that the r between soil variables ranged from 0.58 to 0.69 (Figure 6.4). The highest correlations (r=0.69) were detected between SOC and sMoisture. The 3 soil variables had a positive relationship with each other. The significance level of p-values was noted as <0.0001 for all pair variables (Table B.4). The frequency distributions of SOC, sHumus, and sMoisture were characterized by left skew, multi-model, and bi-model, respectively (Figure 6.4).

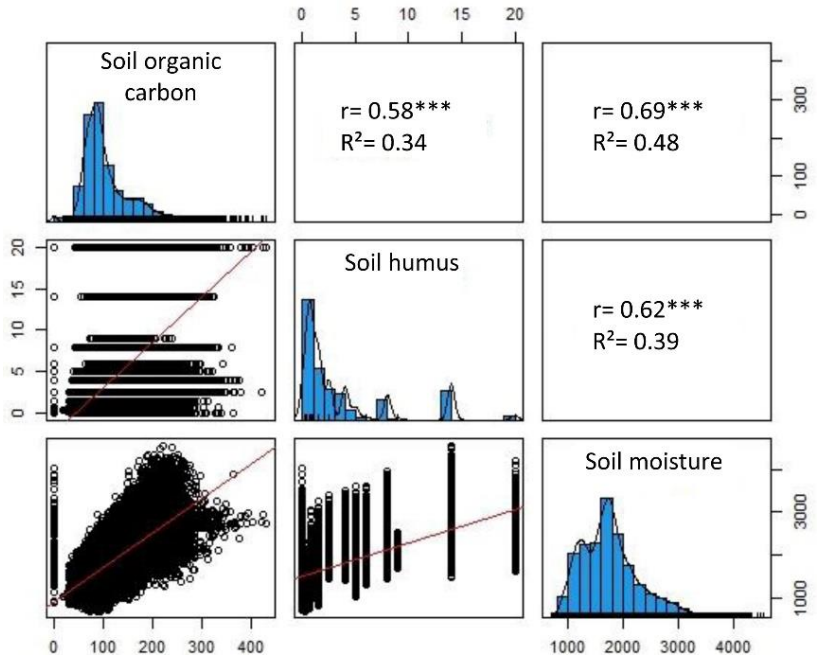


Figure 6. 4 The results of inter-correlation analysis between soil variables (n=19567425). The diagonal plots represent histogram and distribution density, the top right plots represent R^2 (Pearson), and the bottom left plots represent scatter plots in a linear model distribution.

The relationship between vegetation variables

We selected 5 vegetation variables modisNDVI, gimmsNDVI, Fpar, GPP, and LAI for the period 1991-2021 and 2002-2021 during the vegetation growing season (April-August) to analyze the contribution of vegetation to LP (Figure B.10). Vegetation is a vital variable in the interaction between climate and land processes. This includes biogeochemical processes and the water cycle, which converts solar energy into energy for forage (biomass), with this being the base of food chains. Furthermore, high-density vegetation reduces land surface runoff, soil erosion, and degradation. The relationship (r) between vegetation variables ranged from 0.88 to 0.97 (Table B.5).

The strongest correlation (r=0.97) was detected between GPP and Fpar. The lowest correlation (r=0.88) was detected between gimmsNDVI and LAI. The significance level of the p-values was <0.0001 for each pair of variables. The frequency distribution of modisNDVI, gimmesNDVI, Fpar, and GPP was characterized by multi-model, and a left skew of the LAI distribution was reported (Figure 6.5). From these analyses, the native vegetation with dense, stable, and continuous growth in a given area can indicate a high potential for the land. Therefore, vegetation is one of the key indicators for assessing LP (Hengl et al. 2018).

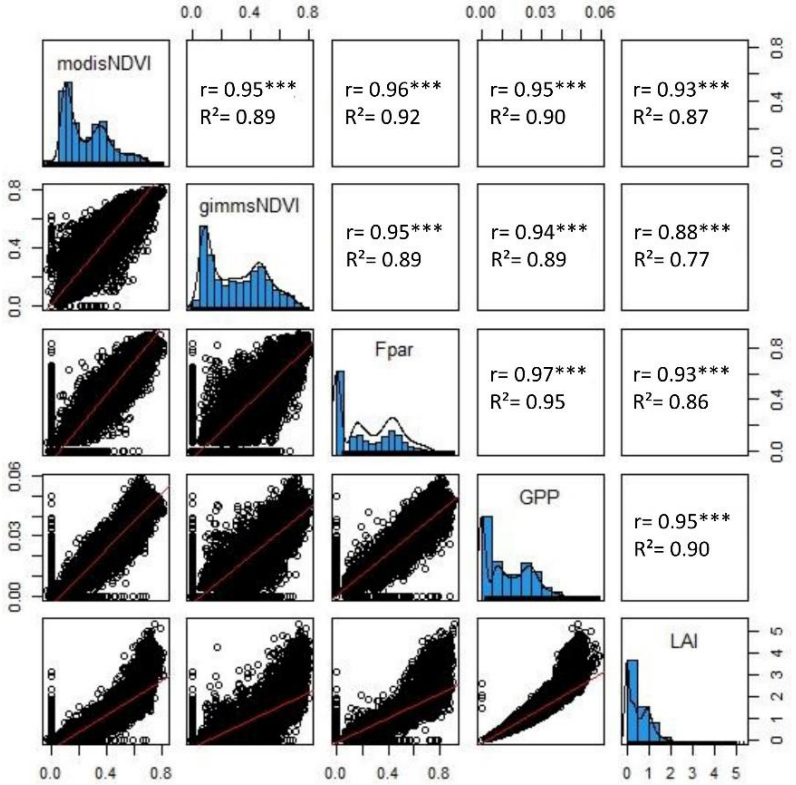


Figure 6. 5 The results of inter-correlation analysis between vegetation variables (n=19567425). The diagonal plots represent a histogram and distribution density, the top right plots represent R^2 (Pearson), and the bottom left plots represent scatter plots in a linear model distribution.

The relationship between objective and explanatory variables

The inter-correlation between topography, climate, soil, and vegetation is shown in Figure 6.2-Figure 6.5. In this part, analyzed the relationship between the objective (dependent) variable (n=12988) and the 25 explanatory (independent) variables. The correlation results revealed that 25 explanatory variables were sufficiently correlated with each other (excluding aspect) (Figure 6.6).

More specifically, all vegetation variables, all soil variables, and 5 climatic variables had a positive correlation with other variables, while the climatic variables of the PET, Srad, Tmp, Vap, and Wnd had a negative correlation. For the relationship between the response variable and 25 explanatory variables, high positive correlations (r=>0.65) were detected between the response variable and the 7 environmental variables modisNDVI (r=0.69), Fpar (r=0.68), GPP (r=0.68), MI (r=0.67, gimmsNDVI (r=0.67), Prec (r=0.66), and LAI (r=0.66). The lowest correlation was between the response variable and the aspect. Based on these correlation results and considering the number of variables should be a reduction in the explanatory variables. Therefore, we generated 5 groupings from the selected 4 main, and 25 sub-variables (Table 6.4).

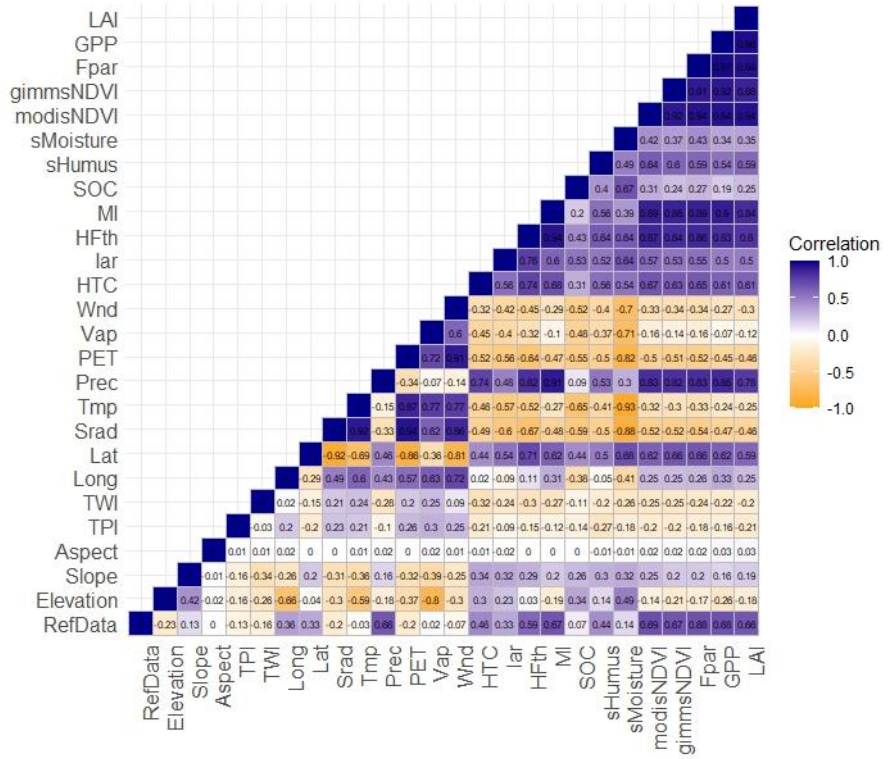


Figure 6. 6 The correlation matrix between dependent (n=12988) and 25 independent variables. The positive and negative correlations are represented in blue and orange, respectively. The acronyms of each variable are shown in Table 6.2.

Table 6. 4 The five groupings generated from the 4 main, and 23 sub-explanatory variables

Group name	Variable category	Variable acronym
Group 1 (G1)	Topography	Elv, Slp, Aspct, TPI, TWI, Lat, Long
Group 2 (G2)	Climate	Srad, Tmp, Prec, PET, Vap, Wnd, HTC, Iar, HFth, MI
Group 3 (G3)	Soil	SOC, sMoisture, sHumus
Group 4 (G4)	Vegetation	modisNDVI, gimmsNDVI, Fpar, LAI, GPP
Group 5 (G5)	All environmental	Integrated G1, G2, G3 and G4

6.3.2 Importance variables and prediction accuracy using the RFR model

The 3 highest ranked variables in the RFR model were Lat, Long, Elv for G1, Prec, MI, HFth for G2, sHumus, sMoisture, SOC for G3 and modisNDVI, gimmsNDVI, GPP for G4 (Figure 6.7, G1-G4). For the 5th grouping (G5), all vegetation variables were ranked in the top five, while two climatic, and two topographic variables were ranked between 6 and 9 (Figure 6.7, G5).

The RFR results showed that the LP can be determined by climate, soil, topography, and vegetation variables. However, vegetation variables were most important and served as indicators of the LP (Parfenava et al. 2019). For reason, vegetation variables are institutes such as climate, soil, and topography is a complex indicator that includes many environmental factors and in turn affects processes such as snow accumulation, soil moisture, surface runoff, infiltration, and erosion (Hengl et al. 2018). Hengl et al. (2018) study revealed that natural vegetation is useful for estimating LP. Specifically, Fpar is a "fraction of photosynthetically active radiation absorbed by the green parts of the canopy", and strongly depends on leaves, canopy structure, an optical property of soil, and irradiance. Moreover, the United Nations (UN) Global Climate Observing System defined "Fpar is one of the key 50 variables accepted as necessary to define global climate" (Qin et al. 2018). Several studies confirmed that Fpar is one of the fundamental variables to define the terrestrial state (Qin et al. 2018; Liu, Shao, and Liu 2015; Myneni, Knyazikhin, and Park 2015). GPP is a collection of complex processes that convert light energy and water into chemical energy, absorbing carbon and releasing oxygen. Especially, for the dry and temperate regions' photosynthesis, vegetation transmits large amounts of water therefore, GPP strongly impacts the interaction between land and atmosphere processes. Therefore, the main drivers of GPP are including climate, soil richness, and species composition of vegetation (Roy et al. 2001). LAI is a key variable of vegetation structure that strongly responds to vegetation in the climate (Fang et al. 2019). Moreover, the strong correlation of NDVI with GPP and Fpar makes NDVI a useful indicator of land resilience

(Yengoh et al. 2014). For the 5 groupings, accuracies of the RFR models explained R^2 values in the ranges from 0.60 (G3) to 0.73 (G5), while RMSE was explained from 0.55 (G5) to 0.67 (G3) (Table 6.5).

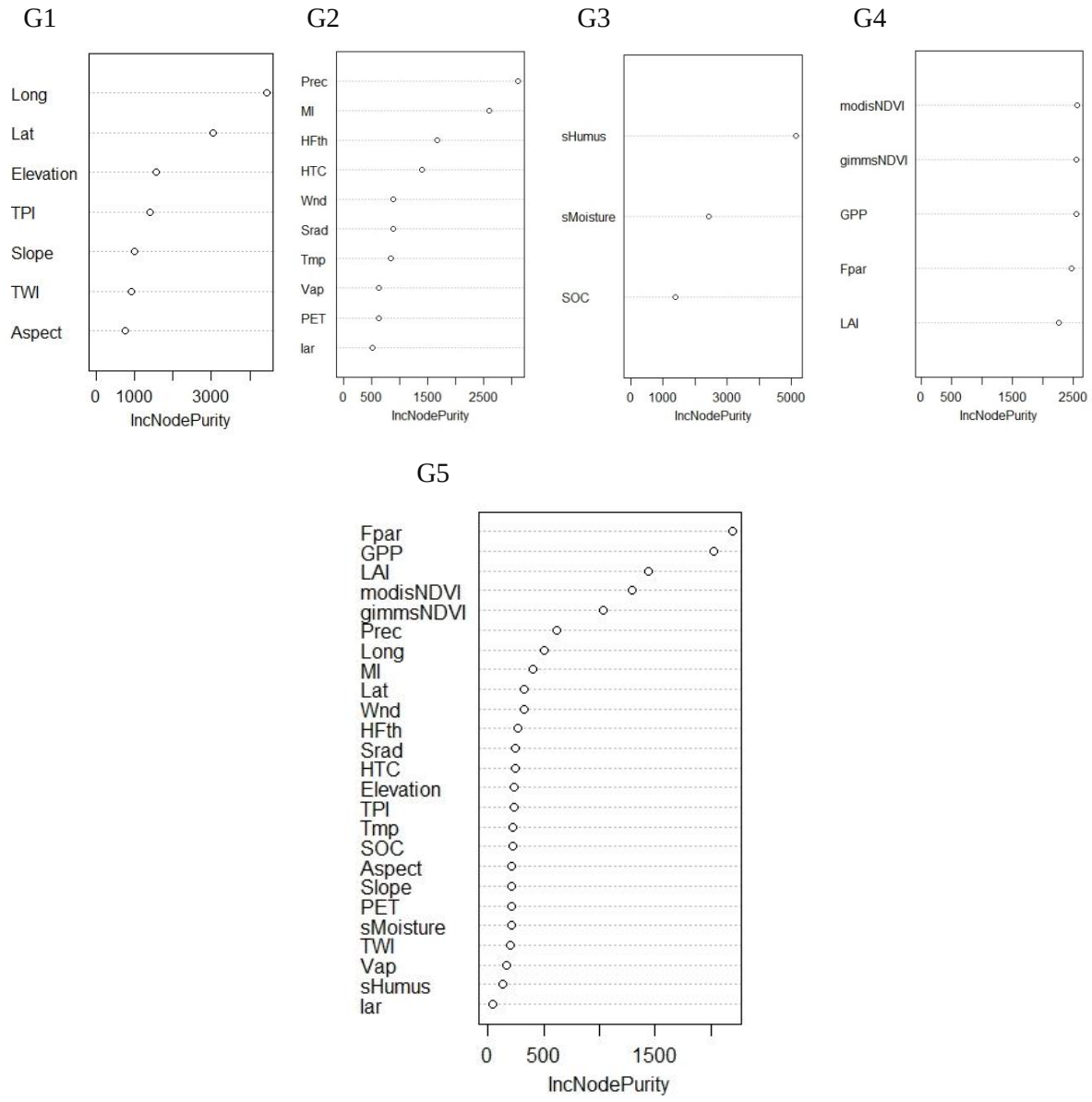


Figure 6. 7 Evaluation of the important variables for the five groupings using the RFR model (n=12988). Importance is given to the increased node purity (IncNodePurity)

Table 6. 5 The summary statistics of the R^2 and RMSE for the LP prediction model, including five groups of variables using the RFR model

n	G1		G2		G3		G4		G5	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
12988	**0.71**	0.57	**0.72**	0.56	**0.60**	0.67	**0.70**	0.58	**0.73**	0.55

6.3.3 Comparison of linear and non-linear regression models

To validate estimated importance variables in the RFR model and find the advantages and differences of the RFR model were compared to linear (e.g., PLSR, PCR) and non-linear (e.g., CART) regression models.

Comparison of linear regression models

To find out the differences between linear regressions PLSR was compared with PCR. We examined the following six conditions in both PLSR and PCR models, which were used to determine the unbiased of the estimators (β parameters): 1) β parameters must have linearity; 2) dependent and independent variables collected from a random sample, which can represent the function of a set; 3) independent variables must not have collinearity and multi-collinearity (must not be high inter-correlations (r=1.0) among them); 4) the conditional expectation of error given by the values of explanatory variables when the expected value of response variable is zero; 5) the conditional variance of the error of explanatory variables must be constant or all the same. It is important to estimate the standard error of β parameters, and 6) the error does not correlate with the independent variables and must have a normal distribution. If the developed models satisfy these six conditions, the model is called a classical linear regression model. Results of PLSR and PCR analysis showed that the first five conditions were satisfied in both models (Table B.6, Table B.7). 23 of the 25 variables had a non-normal distribution (see, Figure 6.2-Figure 6.5). Using the 25 explanatory variables with objective data of 12988 sites to estimate LP, the accuracy of the PCR model (R^2=0.55, RMSE=0.71, F(25, 12964)=624.24, p<0.0001) was slightly higher than the PLSR model (R^2=0.46, RMSE=0.78). Details of both linear regression results are shown in Table B.6 and Table B.7. Moreover, the accuracies of the PCR model for all 5 groupings were much higher than PLSR (Table B.8).

Comparison of linear and nonlinear regression models

To the examination of the estimated variable importance in the RFR, we estimated the Variable Importance in Projection (VIP) in the PLSR model. The VIP defines the contribution of each dependent variable in the PLSR model. If the value of VIP is less than 0.80 that indicates the explanatory variable is less important. The estimated variable importance from the PLSR model was similar to the results of the RFR model for each corresponding group (Figure B.11). Specifically, the highest three ranked variables for the first three groupings in the PLSR model (Figure B.11, G1-G3) were the same as the RFR results. For G4, nearly similar results were indicated (Figure B.11, G4). For G5, modisNDVI, GPP, Fpar, gimmsNDVI, and MI were

ranked in the top five, while Prec, LAI, HFth, HTC, sHumus, and Long were ranked between 6 and 11 in the PLSR model (Figure B.11, G5). These eleven explanatory variables with >0.80 VIP values were categorized as being more important. Estimated importance variables in both RFR and PLSR models were similar. However, the accuracy of the predicted model in the PLSR was lower than the RFR model.

Comparison of nonlinear regression models

In this part, for identifying the threshold value of explanatory variables in the assessment of LP the CART method was used, and the result to compare with RFR, PLSR, and PCR. The CART is a classification and regression technique based on the iterative method to find the best split and was developed by Breiman et al (1984). The splitting structures between root and terminal (leaf) nodes in the CART are considered criteria for the different expected classifications. Hence, in the regression tree of CART, the nodes between dependent and independent variables are explained efficiently (Nandintsetseg, Shinoda, and Erdenetsetseg 2018). In other words, the regression (RT) of the CART is recursively constructed to determine all potential splits of independent variables and define the threshold value of explanatory variables to construct the largest potential tree. For the CART analysis, used 25 environmental variables with 12,988 sites. To avoid overfitting, limited the tree size and set it to three steps (maximum depth) for each grouping. The results of the CART analysis are shown in Figure B.12. For topographic (G1), the most important variables in the assessment of LP or land recovery were the geographical location of Long and Lat with threshold values of ≥97.7º, and ≥47.2º, respectively. In addition, the terrain parameters of Elv and Slp were ranked between 3 and 4 with threshold values of ≥946 m, and ≥2.93º, respectively. For climate (G2), the most important variables were Prec with a threshold value of ≥258 mm, and MI with a threshold value of ≥0.96. For soil variables (G3), the most important variables were sHumus with a threshold value of ≥2.0, and sMoisture with a threshold value of ≥2324 m^3 m^{-3}. For vegetation (G4), the most important variables were Fpar with a threshold value of ≥0.38, and gimmsNDVI with a threshold value of ≥0.39. For the 23 environmental variables (G5), the 7 most important variables (maximum depth 5) were Fpar with a threshold value of ≥0.38, GPP with a threshold value of ≥0.02, LAI with a threshold value of ≥0.34, Long with a threshold value of ≥95.7º, gimmsNDVI with a threshold value of ≥0.39, Wnd with a threshold value of ≥3.1m s^{-2} and Prec with a threshold value of ≥329.5 mm. Accuracies of the RFR model for all 5 groupings were much higher than PLSR, PCR, and CART regressions. Specifically, the RFR model explained 73.0%

(RMSE=0.55) of the total objective data (n=12988) and 25 environmental variables in the assessment of LP.

6.3.4 Map of predicted LP using RFR model

The results indicated that the RFR can develop a prediction model from large amounts of data and long-term optimization. Moreover, non-linear ML regressions do not require the explanatory variables to be normally distributed. In addition, a key benefit of the RFR model was more efficiency for a large number of variables with heterogeneity and un-normal distribution. We generated a spatial distribution map of LP in Mongolia at a resolution of 500 m using the developed RFR model from grouping G5 (Figure 6.8). The LP classes were from very high to very low which is represented by the recovery classes. The classes were: class 1- very high potential, class 2- high potential, class 3- moderate potential, class 4- low potential, and class 5- very low potential, respectively (Table 6.3). The results revealed that 13.32% of the territory of Mongolia has a very high potential, 20.52% high, 30.84% moderate, 22.65 low, and 12.66% very low LP.

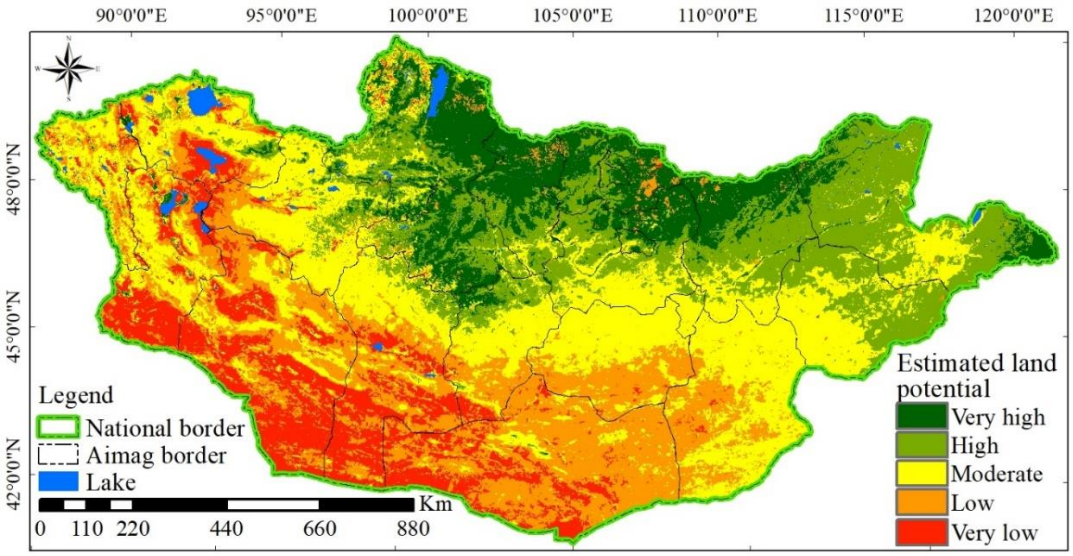


Figure 6. 8 Estimated LP based on the RFR model using field-measured recovery data of the 12988 sites and 25 environmental variables. A spatial distribution map of the assessment of the LP in Mongolia at a resolution of 500 m. Each class represents the rangeland recovery class: very high potential class 1, high potential class 2, moderate potential class 3, low potential class 4, and very low potential class 5, respectively.

6.3.5 Interaction between LP and pasture use

In this part, we considered the interaction between LP and pasture use. For reason, the largest share of land use in Mongolia is agricultural, which is generally pasture land. In 2021, 72.87%

of Mongolian land surface was used as agricultural land (cropland 0.45%, and pastureland 72.42%) (NSOM 2022a). The main consumers of pasture are five types of livestock (sheep, goats, cattle, horses, and camels) and wild animals. One feature of Mongolian livestock husbandry is that they feed on rangeland plants all year round and use natural rangeland. For Mongolia, livestock is renewable natural capital and is subject to state protection under the country's constitution (Undarmaa et al. 2018). The cornerstone of the Mongolian economy is pasture-based livestock husbandry, which strongly influences the country's economic development. Half of the Mongolian population engages in the livestock husbandry sector. Specifically, the total number of households in Mongolia in 2021 was 920165, and of these, the number of households with livestock (246302), and herder households (188610) was 47.0% (NSOM 2022e). Livestock husbandry comprises ~85% of total agricultural products, accounting for 11.2% of gross domestic product (GDP) (MoFALI 2021). In addition, livestock and livestock-derived raw product exports account for approximately 29.7% of Mongolia's foreign exchange earnings (MC 2022). Therefore, livestock husbandry is an integral part of Mongolia's social and economic development and is directly dependent on natural pastures.

Since the 1990 transition to a market economy, the utilization of land resources has increased greatly. For instance, the total livestock population rose from 25.5 million in 1990 to 67.3 million in 2021 by a factor of 2.63 (NSOM 2022b) (Figure A.2). To match pasture use with the LP needs considerable attention, as the number of livestock continues to increase under climate change. If do not consider matching pasture utilization with LP, the future could be faced land degradation and desertification challenges.

To compare the current state of pasture use with LP, firstly, generated spatial distribution maps of livestock density from the number of livestock at the Soum (administrative unit) level of Mongolia for the period 1991-2021 (Figure B.13), and for each decade from 1991 to 2021 (Figure B.14). And then, generated spatial distribution maps of the livestock grazing capacity in Mongolia (Figure 6.9) using the density of livestock population, Eq (2) for each decade from 1991 to 2021, and the 31-year average. When evaluating livestock grazing capacity, the number of livestock in 1 ha of pasture is expressed as a percentage compared to the pasture carrying capacity. The value of LGC is $\leq$ 25.0% over resources pasture, 25.1-50.0% is sufficient pasture, 50.1-75.0% slightly overgrazed, 75.1-100.0% moderately overgrazed and $\geq$100.1% strongly overgrazed (IMALGC 2019). Overgrazing refers to unbalance between the number of livestock and pasture carrying capacity, which declined the usefulness, productivity,

and biodiversity of the land and is affected by land degradation and desertification (Mysterud 2006; Lee 2008).

The results showed (Figure 1.1, Figure B.15 that the total number of livestock gradually increased from 1991 up until 1999, and then decreased from 2000 to 2002 from the impacts of the *dzud* and drought. "*Dzud* is a Mongolian term for severe winter condition" (Chadraabal et al. 2020; Sternberg and Batbuyan 2013; Sternberg et al. 2010; Batjargal et al. 2002). A good definition of *Dzud* is given in Natsagdorj et al (2003), Suttie (2005), Roa et al (2015), and Shinado (2017).

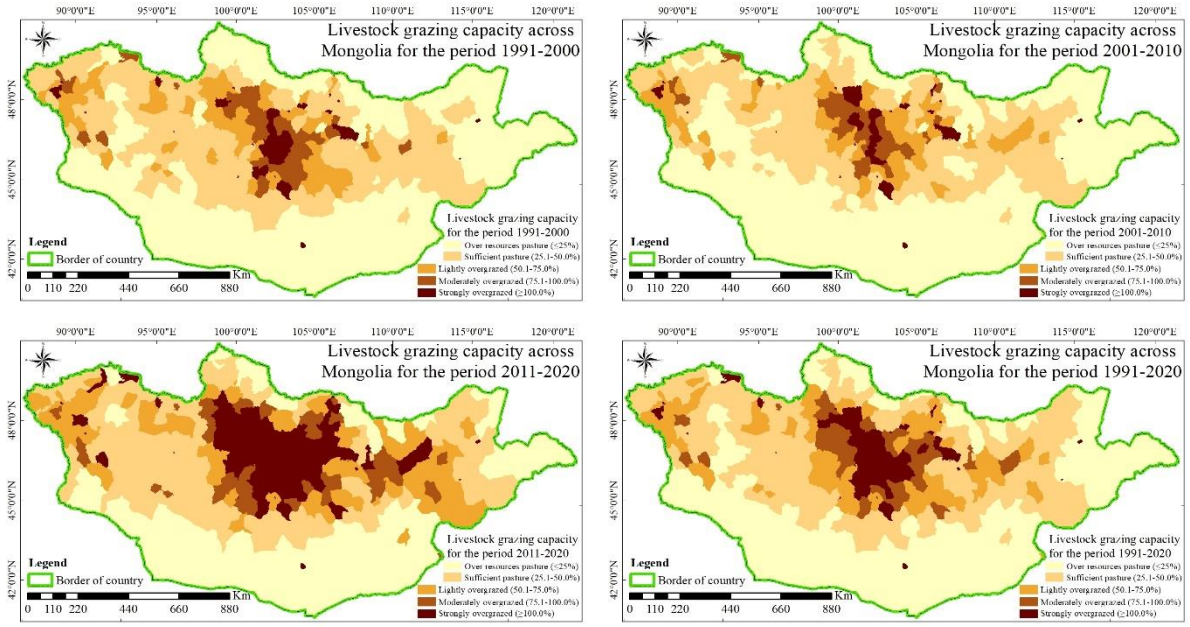


Figure 6. 9 Spatial distribution maps of the livestock grazing capacity of Mongolia. (top left) mean of 1991-2000, (top right) mean of 2001-2010, (bottom left) mean of 2011-2020, and (bottom right) mean of 1991-2020, respectively.

From 2003 until 2009, the livestock population gradually increased, and in 2010 it dramatically decreased. Finally, from 2011 to 2021 it dramatically increased. Most growth has been observed in the last 10 years. Between 1991 and 2000 total livestock density population per 100 hectares ranged from 10 to 205 head; from 2001-2010 it ranged from 10 to 256 head, from 2011 to 2020 it ranged from 10 to 616 head. Since 2011, climate conditions have been relatively stable with normal humidity. The weather conditions have affected livestock growth. On the other hand, the government and herders have been gaining considerable experience in preventing and responding to the risks of *dzud* and drought. However, the pastureland is being overgrazed at an alarming pace (Figure B.15). To detect the interaction between LP and pasture

use, compared the results of LP with livestock grazing capacity. The three decades of the analysis show that in 1991-2010 19.23%, in 2001-2010 20.33%, in 2011-2020 46.6% of the territory have exceeded its LP, of which 0.80%, 5.73%, and 22.02% have highly exceeded the LP (Figure 6.10, Table 6.6). The results of the status of pasture use and the estimated LP revealed that in 2021, 52.34% of the territory of the country has exceeded the LP, of which 26.73% highly exceeded (Figure 6.11). The exceed LP refers to a mismatch between LP and livestock grazing capacity.

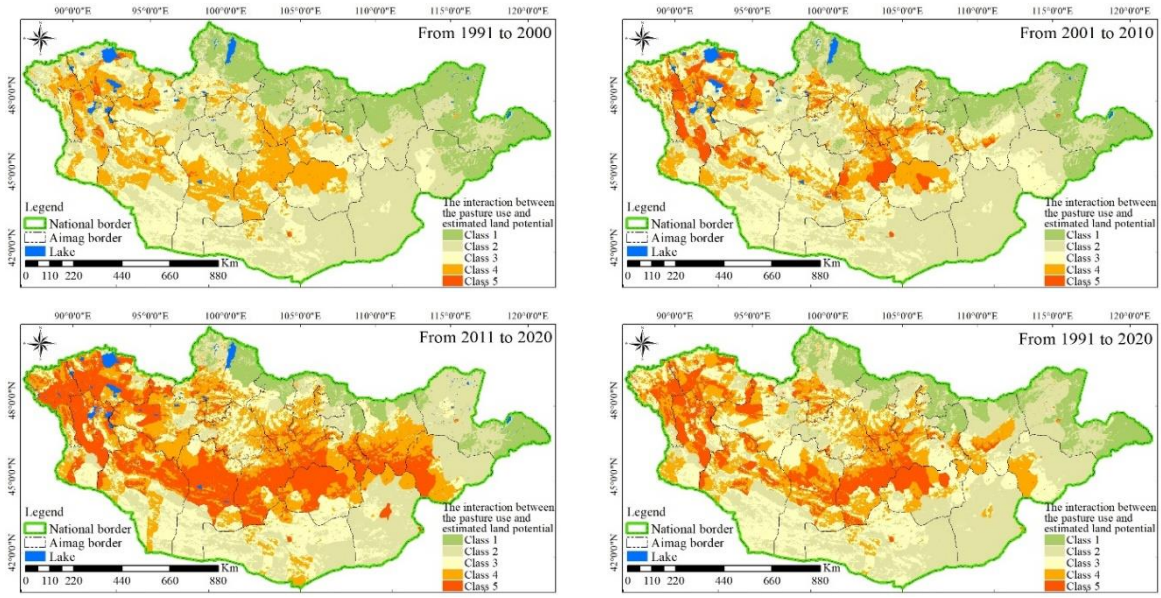


Figure 6. 10 The interaction between historical pasture use (since the social transition), and estimated LP across Mongolia. (top left) mean of 1991-2000, (top right) mean of 2001-2010, (bottom left) mean of 2011-2020, and (bottom right) mean of 1991-2020.

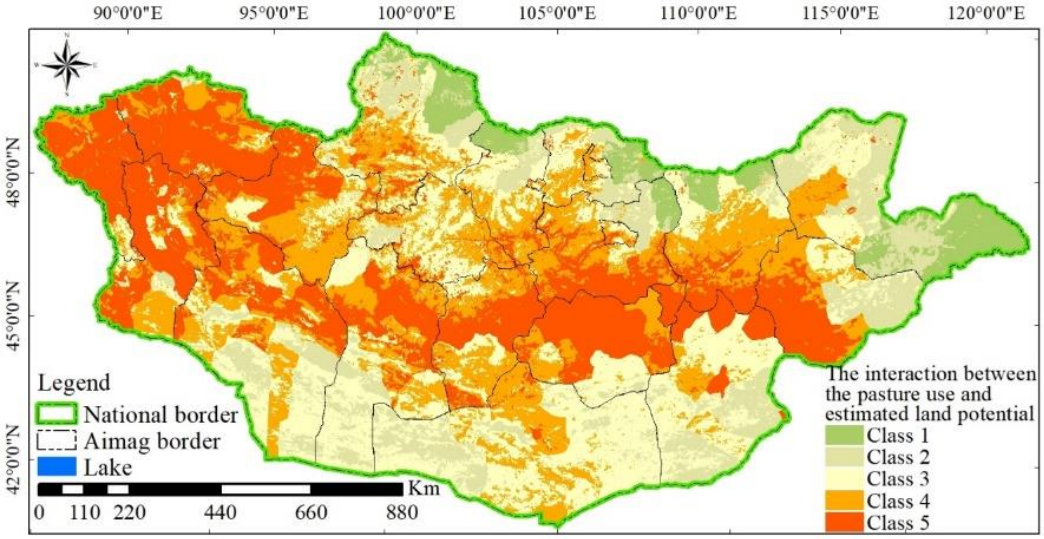


Figure 6. 11 The interaction between the status of pasture use in 2021 and the estimated results of LP in Mongolia.

Table 6. 6 The interaction between historical pasture use and estimated LP in Mongolia

Class	1991-2000		2001-2010		2011-2020		1991-2021	
	km^2	%	km^2	%	km^2	%	km^2	%
Class 1 (very high potential or over-sufficient pasture)	267971.81	17.11	208152.46	13.29	108720.18	6.94	159057.29	10.16
Class 2 (high potential or sufficient pasture)	606641.23	38.74	593304.12	37.89	363725.26	23.23	484635.11	30.95
Class 3 (moderate potential or slightly overgrazed)	390305.86	24.92	446265.61	28.50	363822.72	23.23	430886.05	27.52
Class 4 (low potential or moderately overgrazed)	288547.14	18.43	228582.85	14.60	384970.45	24.58	314993.70	20.11
Class 5 (very low potential or strongly overgrazed)	12533.96	0.80	89694.96	5.73	344761.39	22.02	176427.85	11.27

6.4 Discussion

Indeed, unbalance between LP and land use results in land degradation. The LP contains three elements, namely (i) potential productivity, (ii) degradation resistance, and (iii) resilience (Herrick et al. 2013). LP assessment to support the creation of Ecosystem Services (ESs) demands understanding both current potential and capacity to resist and recover from degradation, which is generally termed resilience (O'Connell et al. 2016). Resilience is "the capacity to recover from disturbance", and resistance is "the capacity of a system to maintain function through a disturbance" (IPR 2016). Degradation appears when disturbances cause an adverse change in the ability to provide ESs (Herrick et al. 2016), which depends on both the disturbance and the system being disrupted. Moreover, the effect of the disturbance on land can be controlled by bio-geo-physical processes. The effect of a new disturbance on land is possible to estimate from its cumulative impacts on this process with past disturbances (IPR 2016). On the other hand, the level of land degradation varies broadly depending on the regime of disturbance and the capacity of recovery (Cowie et al. 2018). Herrick et al (2019) reported that it is a challenge to precisely predict land response to disturbance. However, topography, climate, and soil processes can be applied to enhance the prediction of LP. Herrick et al (2013) noted that the potential resilience of land also depends on the three mentioned factors that represent potential productivity. The responses of land depend on the basic set of bio-geophysical processes for that reason, LP assessment is required by some bio-geophysical processes combined at various spatial and temporal levels (IPR 2016). Therefore, our study focused on assessing LP as a basic foundation for sustainable land use, applying a time series of environmental variables. This study used field-measured data from 12988 sites, and 25 environmental variables such as surface, climate, soil, and vegetation factors.

First, detected the relationship between each environmental variable, namely topography, climate, soil, and vegetation using the correlation matrix (n=19567425). The results showed that the strongest correlations were detected between vegetation variables (r=0.88-0.96), then climatic variables (r=-0.80-0.96), and soil variables (r=0.58-0.69), respectively. Weak correlations were detected between topographic variables (-0.63-0.55). Second, assess the importance of the 25 explanatory variables in the RFR model. Fpar, GPP, LAI, modisNDVI, gimmsNDVI, Prec, Long, MI, and Lat played an important role in modeling LP. The study results indicate that vegetation and climate variables were the main factors controlling LP and productivity at the national level. Third, generated a spatial distribution map of LP at a resolution of 500 m using the developed model in RFR. Finally, compared the prediction model result of LP with current pasture use. The results of the four analyses showed that interactions of topography, climate, soil, and vegetation varied in impact on LP.

For instance, topography variables such as elevation, slope, aspect (slope direction), and topographic position index have a higher influence on surface runoff rate. The runoff rate is higher on steeper slopes, which decreases the infiltration of water (IPR 2016). Lower water infiltration affects plant water availability and thus slows down the productivity of soil organic matter. Moreover, slope, aspect, and TWI also strongly influence LP by regulating water shortages due to evapotranspiration, especially in high latitudes. Therefore, all topographic variables influence the availability of vegetation water.

For the climatic variables, Prec and MI were more important in the LP assessment. However, interactions of climatic variables and the other three environmental variables had varied impacts on LP. For instance, the location of the aspect with high solar radiation had more capacity for productivity but was restricted by low temperature. Moreover, the reserve of soil organic matter was low in dryland regions with low annual mean precipitation. For Mongolia, where a dry-cold climate predominates, the natural nitrogen recovery process takes a long time because of the impact on the nutrient cycle. Therefore, all climatic variables are primary factors for controlling potential productivity, which is the main contributing factor to the potential of land recovery.

For the soil variables, the static soil property (e.g., sHumus) contributed more to LP assessment compared to the dynamic soil property (e.g., sMoisture and sSOC). Soils with a high humus content had more capacity for biomass production and water infiltration. Therefore, it had a high availability of water for vegetation. The high capacity of soil water keeping and

soil humus affect the vegetation nutrients. Therefore, water and nutrient availability to vegetation affects soil structure and nutrient cycling, which are responses to LP.

For the vegetation variables, all variables were important in the LP assessment. Because the contribution of vegetation to LP depends on properties of soil, climate, and topography, the effect is mainly on vegetation rate and species composition. The potential of the land is determined by the ability to be resilient if the area is affected by exposure, and the ability to continue normal functioning (resistance) even though it has undergone negative changes and degradation. Therefore, in the present when human activities influence natural changes, it may be more rational to follow the natural order rather than oppose it.

We investigated to detection of land degradation in Mongolia using a time series of Normalized Difference Vegetation Index (NDVI) of Moderate Resolution Imaging Spectroradiometer (MODIS), and Global Inventory Modeling and Mapping Studies (GIMMS). For detecting land degradation, used Sen's slope, RESTREND analysis, and Breaks For Additive Season and Trend (BFAST) algorithm. For the analysis, monthly mean NDVI derived from MODIS and GIMMS for the period 2002-2019, and 1982-2013 were used (Appendix C).

Estimated trend significance values derived from GIMMS and MODIS NDVI from 1990 to 2019 are shown in Figure C.1. The trend analysis result showed that the positive trends observed in central, north, and northeastern Mongolia increased from the north to the northeast. On the other hand, negative trends are observed in southern Mongolia and all areas of the west, some forested areas in the north and northeast, as well as grassland areas in the east and around Ulaanbaatar. Especially in the northwest and southern parts of Mongolia, the land is also sparsely vegetated. In order to illustrate the significance of the trend, we used Sen's slope method. Figure C.2 shows, high-density vegetated areas were not significant, for instance, in the northern and northeastern parts of Mongolia. However, contrasting trends could balance out, so it was significant to ensure that hypothesis for the determining the linear trend is met for each analysis. BFAST algorithm iteratively calculates the number change and time, and features of change by their amplitude and direction. Therefore, for detection of the trend, seasonal, and remainder components, both gradual and abrupt changes of time series monthly NDVI (MODIS of 2002–2019 and GIMMS of 1981–2013), used the BFAST algorithm. As could be seen from the BFAST analysis (Figure C.3 (a-f)), the areas of positive and negative NDVI trends are strongly related to land cover change. Validation of smaller, more localized NDVI trends can be done with fine-resolution satellite imagery. By the use of the NDVI time

series analysis and BFAST algorithm, we detected negative trends in the areas with forest fire, deforestation, mining activities, and urban expansion (Figure C.4). To illustrate a human-induced degraded area, generated from the Landsat Operational Land Imager (OLI)imagery, Google Earth map, and vector data (Figure C.5). The vector data was obtained from the environmental geo-database of Mongolia. This result revealed that one of the most popular practices in the assessment and monitoring of land degradation could be the analysis of time-series remotely sensed NDVI. The positive trend of NDVI represents the areas of vegetation recovery, and the negative trend represents the human-induced degradation of vegetation cover.

6.5 Conclusion

The objective of this study was to assess LP in Mongolia using a time series of environmental variables and an RFR model. For the analysis, used 25 explanatory variables derived from three different satellites and reanalyzed datasets. 16 of the 25 explanatory variables were time-series data for the period 1991-2021, 1991-2015, and 2002-2021. For training, one objective variable (rangeland recovery data) of 12988 sites from the ALAMGaC for the period 2016-2020 was used. Using the developed RFR model, the spatial map of LP of Mongolia, at a spatial resolution of 500 m, was generated with R^2=0.73, RMSE=0.55, and nRMSE=11.00%. The research results revealed that 13.32% of the territory of Mongolia is a very high potential, 20.52% high, 30.84% moderate, 22.65% low, and 12.66% very low LP. Time series of environmental datasets derived from EO satellites proved to be highly informative and can be used to estimate LP. Moreover, RFR provided an efficient model for a large number of variables with heterogeneity compared to a classical linear model (e.g., PLSR, PCR) and non-linear (e.g., CART) regression models. The statistical model emphasizes that the environmental variables derived from EO data clearly define the potential of the land. The main advantages of this study are the use of more informative field-measured recovery data (including canopy cover, species composition, and basal gaps of perennial vegetation, vegetation height, and vegetation biomass) and a much more powerful machine-learning algorithm. Finally, compared the results of the LP assessment with current pasture utilization. The resulting map of LP compared with pasture use in 2021 showed that 52.34% of the territory has exceeded the LP, of which 26.73% has highly exceeded. These are alarming numbers that should be taken into account in future policy decisions.

Appendix B.

All data that provide the findings of this chapter is included in Appendix B.

Appendix C.

Supplementary data related to land degradation analysis can be found in Appendix C.

CHAPTER 7. Conclusion

7.1 Conclusion

The main objective of this study was to assess and model land resource potential throughout Mongolia using state-of-the-art machine learning techniques and remote sensing data. To achieve the objective, this study examines multiple perspectives of land resource potential such as land suitability, land resources, and land potential, focusing on developing robust models. The main contribution of this assessment is to determine what land resources are current, what land has high potential, how to link land use with land potential, and how to implement it. In addition, the investigation seeks to find out if the land resource potential of a large country like Mongolia can be consistently assessed using machine learning techniques and remote sensing data. The first step of the land assessment potential was related to investigating, assessing, and modeling as the main part of the land assessment. The second stage was explaining the results of these investigations in the link to land management, and development policy. This chapter addresses the findings related to land assessment (land suitability, land resources, and land potential), management, and policy.

The first conclusion is explaining the result of the land suitability assessment for cropland in terms relevant to sustainable management and policy. The result of the cropland suitability analysis showed that 10.1% of the total territory of Mongolia is highly suitable, 14.0% suitable, 15.5% moderately suitable, and 29.1% unsuitable for farming, with 31.2% as the constraint areas (Figure 7.1(a)). The spatial distribution revealed that suitable to highly suitable areas for cropland are observed in the north, central, and northeastern parts of Mongolia. The limited suitability areas occurred in the northwestern, the southern edge of Mongolia's central and eastern parts increased from the northwest to the east. As shown in the results land suitability evaluation for cropland is possible using GIS and remote sensing technology based on a combination of multi-criteria decision output and matrix. The abovementioned method of land suitability evaluation for cropland can be used to save time for land management and it allows for the possibility of justifying policy decisions with science.

The results of this basic research need to be integrated with applied and adaptive research areas to link to sustainability and development policy. Specifically, as of 2021, 3.54% (sown area 1126.2 thousand hectares, and fallow land 211.6 thousand hectares) (ALAMGaC 2021) of the total suitable area (~37695.1 thousand hectares or 24.1%) is used for agricultural cropland. 15.8% of the current used area for agriculture cropland is abandoned (ALAMGaC

2021). Although there is a need to increase agricultural production (fodder and food) due to livestock pastoralism and population growth, a goal must be not to degrade land resources, and possibly improve land productivity by combining sustainable production with protection of nature, which is the basic principle of land management. On the other hand, the areas identified as suitable for cultivation can be used for pastures and other purposes. Therefore, the area available for food and feed production may be several times smaller than estimated, so there is a risk of land shortage if there are no certain actions. Furthermore, it is necessary to take care of the currently used land to preserve the newly identified land for the benefit of future generations. To implement it the following is considered (IRP 2016; McKenzie et al. 2008; Young 2000):

Prioritize soil protection activities (e.g., prohibit farming on steep slopes, increase soil nutrients, and use the protection approach of biology for caring for the soil);

Improvement of the irrigation system (e.g., improving the efficiency of water use, construction of ponds for water storage, development of water transfer dams and canals system);

Develop more productive cultivation varieties (e.g., appropriate in soil constraints, and drought resistant);

Raising awareness of nature conservation of the farmers, (e.g., involving them in nature conservation activities, and creating awareness of the need to improve land management).

The second conclusion is explanation of the results of the land resources assessment including pasture biomass in terms relevant to sustainable management and policy. For Mongolia, which is heavily reliant on natural pastures, the rational use of pastures as a source of feed is a vital issue for the livestock industry. The result of the pasture biomass assessment showed that detailed biomass products can be generated even for very large areas using machine learning techniques, spectral indices derived from remote sensing data, and reference data from field surveys. The result of the pasture biomass assessment showed that 14.9% of the total territory of Mongolia is very high-density biomass (200-255 bit), 25.4% high density (100-200 bit), 29.4% moderately density (50-100 bit), 22.8% low density (5-50 bit), and 7.6% very low density (<5 bit) (Figure 7.1(b)). The spatial distribution map revealed that pasture biomass increases gradually from south to north. Pasture biomass was highly correlated with temperature and precipitation. The desert, semi-desert, and dry steppe in the southern and

northwestern parts of the country have low-density biomass. High-density biomass is observed in the forest, forest-steppe, and grassland zones. Biomass is important to use pastures sustainably, minimizing adverse effects on predominantly fodder plants by the local carrying capacity (Amarsaikhan et al. 2023; Otgonbayar et al. 2021b). However, due to weather conditions, pasture biomass in Mongolia varies yearly, which complicates management. Particularly, the main land use system of arid and semi-arid regions, which occupies 65.0% of the territory of Mongolia is seasonal nomadic pastoralism, and the land resource potential including pasture biomass in this region strongly depends on rainfall (related to water scarcity with seasonality). In addition to unstable pasture biomass, the dramatic growth of livestock and natural disasters (drought, *dzud*, heavy snow, and dust storms) complicates pasture management. Issues of pasture management are more complex compared to crop management. The nomadic pastoralism of drylands (supply fodder) is especially complex to adapt to global change. The following activities might be good practices in order not to irreversibly degrade the vegetation of pastures.

Increase farming with forage legumes in order to supplement feeding in the dry season to crop perennial vegetation;

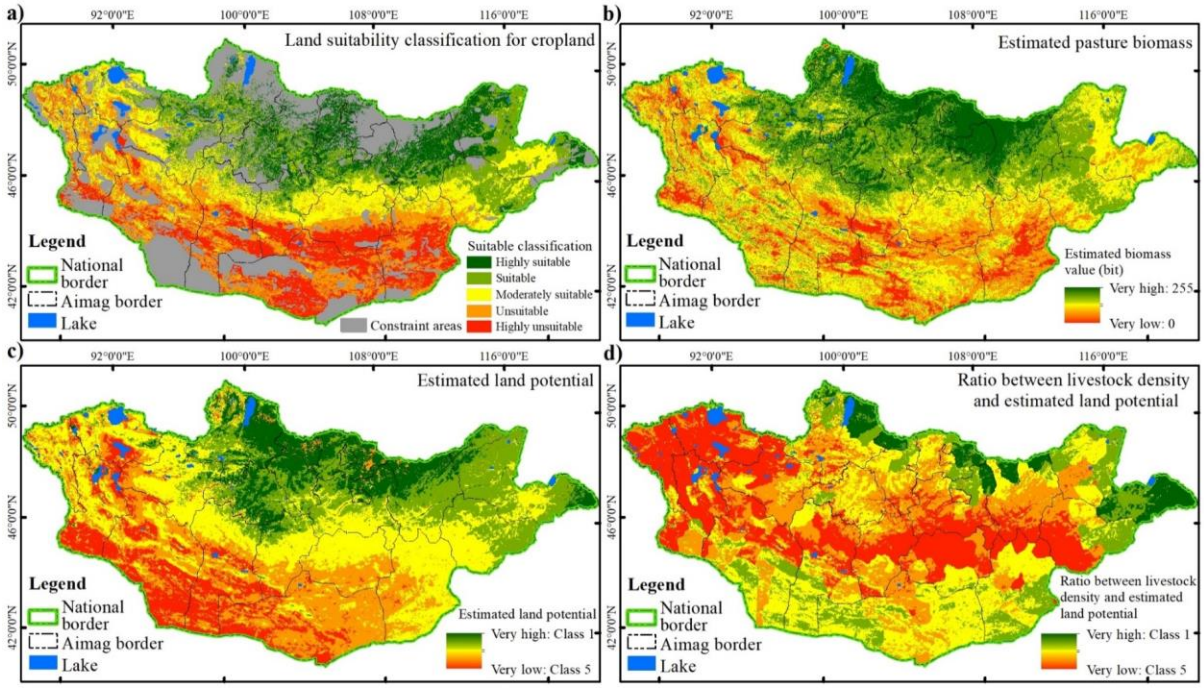


Figure 7. 1 Results of land resources potential assessment in Mongolia. (a) the result of the land suitability analysis for cropland; (b) the result of pasture biomass analysis; (c) the result of the land potential analysis; (d) the ratio between the livestock density and estimated land potential.

The third conclusion is explanation of the results of the land potential assessment in terms relevant to sustainable management and policy. The research results showed that 13.32% of the territory of Mongolia has a very high potential, 20.52% high, 30.84% moderate, 22.65% low, and 12.66% very low land potential (Figure 7.1(c)). The spatial distribution revealed that very high potential occurs in forests regions in the north (around the Khovsgol mountains), northeast (Khentii mountains), and east (Khyngan mountains) of Mongolia, as well as forest-steppe regions in the central part (around the north side of Knagay mountains). High potential is observed in some forested regions in forest-steppe and grassland regions in central, northern, and eastern Mongolia. Moderate potential occurs in all regions of the west (Altai mountains), and steppe regions (semi-arid regions) from west to east. Low to very low potential occurred in desert steppe, and desert (arid regions) in south and southwestern Mongolia. The comparison of the results of the land potential with pasture use showed that 52.34% of the territory has exceeded the land potential, of which 26.73% highly exceeded. This study showed that in all regions of western Mongolia, the central parts (steppe and semi-arid regions), and southeastern parts (steppe or grassland regions) are very risky (Figure 7.1(d)). Particularly in the western part of Mongolia land potential productivity decrease is caused by a mismatch between land use and land potential. The results of our research on land degradation have confirmed this. Although herders in the arid and semi-arid regions (drylands) have been using the pastures seasonally, degradation is continuously increasing. Despite the fact these drylands is well suited to nomadic livestock pastoralism (for rotational grazing) with seasonality, the lack of fodder caused by weather condition (especially, meteorological drought) makes pasture management difficult. At the same time, when the number of livestock population rapidly increases and pastures are overgrazed, the drought frequency increases (being an integral part of arid and semi-arid regions), affecting natural vegetation growth, reducing biomass, limiting seed dispersal, and slowing down the natural regeneration process of grasslands. This is a very serious issue related to land productivity. With the growth in livestock placing rising pressures on drylands, the Mongolian government should provide high priority to optimal land utilization, enhanced land management, and avoiding land degradation. It is necessary to take the following measures for long-term sustainable use of land resources and improve land potential. These include:

Creating opportunities for the natural regeneration of overgrazed pastures by temporarily releasing them from use.

Ensuring the implementation of laws and regulations aimed at reducing overexploitation of pasture.

Increasing vegetation cover by planting perennials.

Implementation of measures to protect the soil of agricultural fields from erosion and damage, and to support farming with forest strips.

Taking measures to improve the physical characteristics of the soil by revegetation and use of green manures in excessively degraded areas.

Local governments should be taking initiative to reduce land degradation

Sorting out the problem of multi-branched dirt roads and not creating new ones;

"Government Environmental Policy", "Government Food and Agriculture Policy", "Green Development Policy", "National Biodiversity Program", "Climate Change Adaptation Program", "Land Degradation Zero Program", and "Billion Tree National Program" to ensure coordination and improve interdisciplinary policy, planning, and coordination.

To summarize this section, this research revealed that the utilization of land resources in arid and semi-arid regions exceeds the actual potential of the land. To avoid land degradation, the users (herders, farmers) of land resources realize the need to control the land resources of the area and put them into practice. Because herders are well known to increase livestock mortality when drought and *dzud* years come, despite this, herders prefer to raise livestock to take more of the benefits of this. Therefore, the priority needs to be to raise awareness of nature conservation among the public to involve them in nature conservation activities.

7.2 Further studies

In order not to make costly mistakes in the future, it is necessary to carry out the following activities.

To develop unified guidelines for land assessment, mapping, and modeling to support the proper use, protection, and improvement of land resources. This would contribute to evidence-based decision-making related to sustainable long-term management of land resources (policy, regulation, and planning).

There is a need to develop a mobile application that integrates scientific and local community (herders and farmers) knowledge to take real-time information on a given area's potential productivity (current land status, capacity of the land resilience and resistance, and

recovery from degradation). It would allow the opportunity to increase fundamental information, knowledge of land resources, and cloud-based geospatial databases.

There is a need of rigorous requirements for the goals of proper use of land resources, to avoid land degradation, and make decisions based on scientific research. When there is a better understanding of the contribution of land resources to human well-being, and when the right awareness and attitudes are created, the measures implemented in the field of land management improvement will bring more results. Better information, more research, more training, and promotion will be important parts of the effort to improve management. Therefore, the priority needs to be raising awareness of the public related to matching land use with land potential and protecting the land.

References

Avirmed, E., Oyungerel, B., Renchinmyadag, T., Munkhdulam, O., Bayanjargal, B., Batnyam, Ts., Davaagatan, T., Purevsuren, M., Erdenesukh, S., Davaadorj, D., Sersmaa, J. & Dorjnamjaa, D. (2020). *Landscape ecological potential of Mongolia [Монгол орны ландшафтын экологийн чадавх](pp. 404)*. Ulaanbaatar, Mongolia: Namnan Design Press. ISBN978-9919-23-153-8.

Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social Sciences* 6 (4): 792-795. Available online: http://www.utd.edu/~Herve/Abdi-PLSR2007-pretty.pdf.

Abdullah, H. M., Akiyama, T., Shibayama, M., & Awaya, Y. (2011). Estimation and validation of biomass of a mountainous agroecosystem using sampling, spectral data, and QuickBird satellite image. *International Journal of Sustainable Development & World Ecology*, *18*(5), 384-392.

ALAMGaC (Administration and Management, Geodesy and Cartography). (2021). state consolidated report of the unified land fund. Accessed on 09 June 2023, Available online at https://en.gazar.gov.mn/report/gnst

Ali, I., Cawkwell, F., Dwyer, E., & Green, S. (2017). Modeling managed grassland biomass estimation by using multitemporal remote sensing data—A machine learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), 3254-3264.

Amarsaikhan, E., Erdenebaatar, N., Amarsaikhan, D., Otgonbayar, M., & Bayaraa, B. (2023). Estimation and mapping of pasture biomass in Mongolia using machine learning methods. *Geocarto International*, *38*(1), 2195824.

Amiri, M., Tarkesh, M., Jafari, R., & Jetschke, G. (2020). Bioclimatic variables from precipitation and temperature records vs. remote sensing-based bioclimatic variables: Which side can perform better in species distribution modeling? *Ecological informatics*, *57*, 101060.

Anderson, R. P. (2012). Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, *1260*(1), 66-80.

Anderson, G. L., Hanson, J. D., & Haas, R. H. (1993). Evaluating Landsat Thematic Mapper derived vegetation indices for estimating above-ground biomass on semiarid rangelands. *Remote sensing of environment*, *45*(2), 165-175.

Arriaga, L., Castellanos, A. E., Moreno, E., & Alarcón, J. (2004). Potential ecological distribution of alien invasive species and risk assessment: a case study of buffel grass in arid regions of Mexico. Conservation biology, 18(6), 1504-1514.

Attorre, F., Alfo', M., De Sanctis, M., Francesconi, F., & Bruno, F. (2007). Comparison of interpolation methods for mapping climatic and bioclimatic variables at the regional scale. International Journal of Climatology: A Journal of the Royal Meteorological Society, 27(13), 1825-1843.

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote sensing of environment*, *93*(1-2), 53-67.

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems, and major information needs. *Remote sensing*, *5*(2), 949-981.

Atzberger, C., & Richter, K. (2012). Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. Remote Sensing of Environment, 120, 208-218.

Atzberger, C., & Rembold, F. (2013). Mapping the spatial distribution of winter crops at a sub-pixel level using AVHRR NDVI time series and neural nets. *Remote Sensing*, *5*(3), 1335-1354.

Atzberger, C., Guérif, M., Baret, F., & Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Computers and electronics in agriculture, 73(2), 165-173.

Bai, P., & Liu, X. (2018). Evaluation of five satellite-based precipitation products in two gauge-scarce basins on the Tibetan Plateau. *Remote Sensing*, *10*(8), 1316.

Banai, R. (1993). Fuzziness in geographical information systems: contributions from the analytic hierarchy process. *International Journal of Geographical Information Science*, *7*(4), 315-329.

Baret, F., Champion, I., Guyot, G., & Podaire, A. (1987). Monitoring wheat canopies with a high spectral resolution radiometer. *Remote Sensing of Environment*, *22*(3), 367-378.

Baret, F., & Buis, S. (2008). Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. *Advances in land remote sensing: System, modeling, inversion and application*, 173-201.

Bartkowiak, P., Castelli, M., & Notarnicola, C. (2019). Downscaling land surface temperature from MODIS dataset with random forest approach over alpine vegetated areas. *Remote Sensing*, *11*(11), 1319.

Barzegar, R., Adamowski, J., & Moghaddam, A. A. (2016). Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. *Stochastic environmental research and risk assessment*, *30*, 1797-1819.

Batjargal, Z., Oyun, R., Togtokh, N., & Sangidansranjav, S. (2002). Lessons learned from the dzud 1999–2000. *Fundamental issues affecting the sustainability of the Mongolian Steppe*, 73-98.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & De Roo, A. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589-615.

Beek, K. (1975). Land utilization types in land evaluation.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, *114*, 24-31.

Bella, D., Faivre, R., Ruget, F., Seguin, B., Guerif, M., Combal, B., ... & Rebella, C. (2004). Remote sensing capabilities to estimate pasture production in France. *International Journal of Remote Sensing*, *25*(23), 5359-5372.

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, *124*, 108-121.

Le Bissonnais, Y. L. (1996). Aggregate stability and assessment of soil crushability and erodibility: I. Theory and methodology. *European Journal of soil science*, *47*(4), 425-437.

Blackburn, K. W., Libohova, Z., Adhikari, K., Kome, C., Maness, X., & Silman, M. R. (2022). Influence of land use and topographic factors on soil organic carbon stocks and their Spatial and vertical distribution. *Remote Sensing*, *14*(12), 2846.

Booth, D. T., & Cox, S. E. (2008). Image-based monitoring to measure the ecological change in rangeland. *Frontiers in Ecology and the Environment*, *6*(4), 185-190.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Cart. *Classification and regression trees*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Brown, D. P., & Comrie, A. C. (2002). Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA. *Climate Research*, *22*(2), 115-128.

Bryan, B. A., & Crossman, N. D. (2008). Systematic regional planning for multiple objective natural resource management. *Journal of environmental management*, *88*(4), 1175-1189.

Burrough, P. A., MacMillan, R. A., & Van Deursen, W. (1992). Fuzzy classification methods for determining land suitability from soil profile observations and topography. *Journal of Soil Science*, *43*(2), 193-210.

Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (2015). Principles of geographical information systems. Oxford University Press, USA.

Byambatsogt, Ts. (2017). Economic diversification and focus problems. *Research and statistics department of the Bank of Mongolia*. Accessed on 21 April 2023, Available online at https://www.mongolbank.mn/file.

Carrascal, L. M., Galván, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, *118*(5), 681-690.

Ceglar, A., Toreti, A., Lecerf, R., Van der Velde, M., & Dentener, F. (2016). Impact of meteorological drivers on regional inter-annual crop yield variability in France. *Agricultural and forest meteorology*, *216*, 58-67.

Chadraabal, A., Shinoda, M., Suzuki, Y., & Komiyama, H. (2020). Mitigation of severe wintertime disasters in northern Mongolia through the early implementation of local action. *International Journal of Disaster Risk Reduction*, *50*, 101739.

Chatterjee, R. S., Singh, N., Thapa, S., Sharma, D., & Kumar, D. (2017). Retrieval of land surface temperature (LST) from Landsat TM6 and TIRS data by single channel radiative transfer algorithm using satellite and ground-based inputs. *International journal of applied earth observation and geoinformation*, *58*, 264-277.

Chen, F., Kissel, D. E., West, L. T., & Adkins, W. (2000). Field-scale mapping of surface soil organic carbon using remotely sensed imagery. *Soil Science Society of America Journal*, *64*(2), 746-753.

Chen, Y., Yu, J., & Khan, S. (2013). The spatial framework for weight sensitivity analysis in AHP-based multi-criteria decision making. *Environmental modeling & software*, *48*, 129-140.

Chen, F., Liu, Y., Liu, Q., & Qin, F. (2015). A statistical method based on remote sensing for the estimation of air temperature in China. *International Journal of Climatology*, *35*(8), 2131-2143.

Chrisman, N. (1997). Exploring Geographic Information System. John Willey & Sons. *Inc., New York*.

Christian, C. S., Blake, S. T., Nokes, L. C., & Stewart, G. A. (1953). No. 1 General Report on Survey of Katherine–Darwin Region, 1946. *CSIRO Land Research Surveys*, *2010*(1), 1-177.

Clevers, J. G. P. W., Van der Heijden, G. W. A. M., Verzakov, S., & Schaepman, M. E. (2007). Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering & Remote Sensing*, *73*(10), 1141-1148.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, *84*(4), 561-571.

Collins, M. G., Steiner, F. R., & Rushman, M. J. (2001). Land-use suitability analysis in the United States: historical development and promising technological achievements. *Environmental management*, *28*, 611-621.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnere, A., ... & Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. Remote sensing of environment, 84(1), 1-15. doi:10.1016/S0034-4257(02)00035-4.

Cowie, A. L., Orr, B. J., Sanchez, V. M. C., Chasek, P., Crossman, N. D., Erlewein, A., ... & Welton, S. (2018). Land in balance: The scientific conceptual framework for Land Degradation Neutrality. *Environmental Science & Policy*, *79*, 25-35. doi.org/10.1016/j.envsci.2017.10.011.

Cresswell, M. P., Morse, A. P., Thomson, M. C., & Connor, S. J. (1999). Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model. *International Journal of Remote Sensing*, *20*(6), 1125-1132.

Czajkowski, K. P., Goward, S. N., Stadler, S. J., & Walz, A. (2000). Thermal remote sensing of near-surface environmental variables: application over the Oklahoma Mesonet. *The Professional Geographer*, *52*(2), 345-357.

Dash, P., Göttsche, F. M., Olesen, F. S., & Fischer, H. (2002). Land surface temperature and emissivity estimation from passive sensor data: Theory and practice-current trends. *International Journal of remote sensing*, *23*(13), 2563-2594.

Datt, G., & Ravallion, M. (1990). Regional Disparities, Targeting, and Poverty in India. *Agricultural Policies WPS*, *375*, 1–47.

Deniz, A., Toros, H., & Incecik, S. (2011). Spatial variations of climate indices in Turkey. *International Journal of Climatology*, *31*(3), 394-403.

Densambuu, B., Sainnemekh, S., Bestelmeyer, B., & Budbaatar, U. (2018). National report on the rangeland health of Mongolia: Second Assessment. *Green Gold-Animal health project, SDC*, *62*.

DES (Department of Environment and Science) and Department of Resources (2021). Queensland Land Resource Assessment Guidelines-Volume 1: *Soil and land resource assessment.* Department of Environment and Science and Department of Resources, Brisbane, Queensland.

Detsch F. (2016). Package "GIMMS". Accessed on 25 June 2022, Available online at https://cran.r-project.org/web/packages/gimms/gimms.pdf.

Dickinson, R. E. (1983). Land surface processes and climate—Surface albedos and energy balance. In *Advances in geophysics* (Vol. 25, pp. 305-353). Elsevier.

Dodson, R., & Marks, D. (1997). Daily air temperature interpolated at a high spatial resolution over a large mountainous region. *Climate Research*, *8*(1), 1-20.

Dong, J., Metternicht, G., Hostert, P., Fensholt, R., & Chowdhury, R. R. (2019). Remote sensing and geospatial technologies in support of a normative land system science: Status and prospects. *Current Opinion in Environmental Sustainability*, *38*, 44-52. doi.org/10.1016/j.cosust.2019.05.003.

Duan, S. B., Li, Z. L., Wu, H., Leng, P., Gao, M., & Wang, C. (2018). Radiance-based validation of land surface temperature products derived from Collection 6 MODIS thermal infrared data. International journal of applied earth observation and geoinformation, 70, 84-92.

Duan, S. B., Li, Z. L., Tang, B. H., Wu, H., Tang, R., Bi, Y., & Zhou, G. (2014). Estimation of the diurnal cycle of land surface temperature at high temporal and spatial resolution from clear-sky MODIS data. *Remote Sensing*, *6*(4), 3247-3262.

Dumanski, J. (1997). Criteria and indicators for land quality and sustainable land management. *ITC Journal*, *3*(4), 216-222.

Dusseux, P., Hubert-Moy, L., Corpetti, T., & Vertès, F. (2015). Evaluation of SPOT imagery for the estimation of grassland biomass. *International Journal of Applied Earth Observation and Geoinformation*, *38*, 72-77. doi:10.1016/j.jag.2014.12.003.

Edirisinghe, A., Hill, M. J., Donald, G. E., & Hyder, M. (2011). Quantitative mapping of pasture biomass using satellite imagery. *International Journal of Remote Sensing*, *32*(10), 2699-2724. doi:10.1080/01431161003743181.

Eisfelder, C., Kuenzer, C., & Dech, S. (2012). Derivation of biomass information for semi-arid areas using remote-sensing data. *International Journal of Remote Sensing*, *33*(9), 2937-2984. doi:10.1080/01431161.2011.620034.

Eisfelder, C., Klein, I., Niklaus, M., & Kuenzer, C. (2014). Net primary productivity in Kazakhstan, it is spatiotemporal patterns and relation to meteorological variables. *Journal of Arid Environments*, *103*, 17-30.

Emamifar, S., Rahimikhoob, A., & Noroozi, A. A. (2013). Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree. *International Journal of Climatology*, *33*(15), 3174-3181.

Enkh-Amgalan, A. (1990). Economic assessment of the land resource of the Mongolian People's Republic to improve the territorial organization of agricultural production. *Doctoral dissertation.*

Escadafal, R., & Bacha, S. (1996). Strategy for the dynamic study of desertification. In *Proceedings of the ISSS International Symposium Ouagadougou, Burkino Faso* (pp. 19-34).

Fan, W., Liu, Y., Xu, X., Chen, G., & Zhang, B. (2014). A new FAPAR analytical model based on the law of energy conservation: A case study in China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(9), 3945-3955.

FAO. (1976). A framework for land evaluation. Soils Bulletin 32, FAO, Rome. 72 p. Also, Publication 22, (R. Brinkman and A. Young (eds.)), ILRI, Wageningen, The Netherland.

FAO, 1978-1981. Report on the Agro-ecological Zones project. Vol. 1. Methodology and results for Africa. Vol.2. Results for Southwest Asia; Vol.3. Methodology and results for South and Central America; Vol. 4, Results for Southeast Asia. FAO World Soil Resources Report, 48/1, 4.

FAO. (1983). Guidelines: Land evaluation for rainfed agriculture, FAO Soils Bulletin 52.

FAO. (1984). Guidelines: Land evaluation for forestry. FAO soils bulletin 48.

FAO. (1989). Guidelines: Land evaluation for irrigated agriculture. FAO soils bulletin 57.

FAO. (1991). Guidelines: Land evaluation for extensive grazing. FAO soils bulletin 58.

FAO. (1993). Guidelines for land use planning. FAO Development Series 1.

FAO. (1996). Agro-ecological zoning guidelines. FAO Soils Bulletin 73.

FAO. (2007). Land evaluation: Towards a revised framework. Land and Water Discussion Paper 6, Rome 107 pp., ISSN: 1729-0554.

FAO (Food and Agriculture Organization) of the United Nations. (2023). Accessed on 25 May 2023, Available online at https://www.fao.org/land-water/land/en/.

Farifteh, J., Van der Meer, F., Atzberger, C., & Carranza, E. J. M. (2007). Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, *110*(1), 59-78.

Feilhauer, H., He, K. S., & Rocchini, D. (2012). Modeling species distribution using niche-based proxies derived from composite bioclimatic variables and MODIS NDVI. *Remote Sensing*, *4*(7), 2057-2075.

Fernandes, L. S., Fernandes, A. C. P., Ferreira, A. R. L., Cortes, R. M. V., & Pacheco, F. A. L. (2018). A partial least squares–path modeling analysis for the understanding of biodiversity loss in rural and urban watersheds in Portugal. *Science of the total environment*, *626*, 1069-1085.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302-4315.

Fischer, G., Nachtergaele, F. O., Van Velthuizen, H. T., Chiozza, F., Franceschini, G., Henry, M., ... & Tramberend, S. (2021). *Global agroecological zones v4–model documentation*. Food & Agriculture Organization.

Funk, C., Verdin, A., Michaelsen, J., Peterson, P., Pedreros, D., & Husak, G. (2015). A global satellite-assisted precipitation climatology. *Earth System Science Data*, *7*(2), 275-287.

Gallagher, S. J., Duddy, I. R., Quilty, P. G., Smith, A. J., Wallace, M. W., Holdgate, G. R., & Boult, P. J. (2004). The use of the Foraminiferal Colouration Index (FCI) as a thermal indicator and correlation with vitrinite reflectance in the Sherbrook Group, Otway Basin, Victoria.

Gallego-Elvira, B., Taylor, C. M., Harris, P. P., Ghent, D., Veal, K. L., & Folwell, S. S. (2016). Global observational diagnosis of soil moisture control on the land surface energy balance. *Geophysical Research Letters*, *43*(6), 2623-2631.

Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, *58*(3), 257-266. doi:10.1016/S0034-4257(96)00067-3.

Garand, L., Buehner, M., & Wagneur, N. (2004). Background error correlation between surface skin and air temperatures: estimation and impact on the assimilation of infrared window radiances. *Journal of Applied Meteorology*, *43*(12), 1853-1863.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, *185*, 1-17.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern recognition letters*, *31*(14), 2225-2236. doi:10.1016/j.patrec.2010.03.014.

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, *161*(2), 165-173. doi:10.1078/0176-1617-01176.

Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical research letters*, *32*(8).

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*, *58*(3), 289-298. doi:10.1016/S0034-4257(96)00072-7.

Gitelson, A. A., Peng, Y., Masek, J. G., Rundquist, D. C., Verma, S., Suyker, A., ... & Meyers, T. (2012). Remote estimation of crop gross primary production with Landsat data. *Remote Sensing of Environment*, *121*, 404-414. doi:10.1016/j.rse.2012.02.017.

Giuliani, G., Mazzetti, P., Santoro, M., Nativi, S., Van Bemmelen, J., Colangeli, G., & Lehmann, A. (2020). Knowledge generation using satellite earth observations to support sustainable development goals (SDG): A use case on Land degradation. *International*

Journal of Applied Earth Observation and Geoinformation, *88*, 102068. doi.org/10.1016/j.jag.2020.102068.

Gomez, C., Rossel, R. A. V., & McBratney, A. B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*, *146*(3-4), 403-411. doi:10.1016/j.geoderma.2008.06.011.

Gonzalez-Roglich, M., Zvoleff, A., Noon, M., Liniger, H., Fleiner, R., Harari, N., & Garcia, C. (2019). Synergizing global tools to monitor progress towards land degradation neutrality: Trends. Earth and the World Overview of Conservation Approaches and Technologies sustainable land management database. *Environmental Science & Policy*, *93*, 34-42.

Goodchild, M. F., Steyaert, L. T., Parks, B. O., Johnston, C., Maidment, D., Crane, M., & Glendinning, S. (Eds.). (1996). GIS and environmental modeling: progress and research issues.

Gungaadorj, Sh. (2009). Brought virgin lands under cultivation and it developed an independent agricultural cropland sector in Mongolia. *Proceeding of the national conference ATAR-50,* Ulaanbaatar, Mongolia, 23 November, 4-19.

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data*, *7*(1), 109. doi.org/10.1038/s41597

Haugum, S. V. (2021). Land-use and climate impacts on drought resistance and resilience in coastal heathland ecosystems.

Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., & Prentice, I. C. (2018). Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. *PeerJ*, *6*, e5457. doi.org/10.7717/peerj.5457.

Herrick, J. E., Urama, K. C., Karl, J. W., Boos, J., Johnson, M. V. V., Shepherd, K. D., ... & West, L. T. (2013). The global Land-Potential Knowledge System (LandPKS): Supporting evidence-based, site-specific land use and management through cloud computing, mobile applications, and crowdsourcing. *Journal of Soil and Water Conservation*, *68*(1), 5A-12A. doi.org/10.2489/jswc.68.1.5A.

Herrick, J. E., Beh, A., Barrios, E., Bouvier, I., Coetzee, M., Dent, D., ... & Webb, N. P. (2016). The land-potential knowledge system (LANDPKS): mobile apps and collaboration for optimizing climate change investments. *Ecosystem Health and Sustainability*, *2*(3), e01209.

Herrick, J. E., Neff, J., Quandt, A., Salley, S., Maynard, J., Ganguli, A., & Bestelmeyer, B. (2019). Prioritizing land for investments based on short-and long-term land potential and degradation risk: A strategic approach. *Environmental Science & Policy*, *96*, 52-58. doi.org/10.1016/j.envsci.2019.03.001.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *25*(15), 1965-1978.

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., & Henderson, S. B. (2014). Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, *154*, 38-45.

Hooker, J., Duveiller, G., & Cescatti, A. (2018). A global dataset of air temperature derived from satellite remote sensing and weather stations. *Scientific data*, *5*(1), 1-11.

Hopkins, L. D. (1977). Methods for generating land suitability maps: a comparative evaluation. *Journal of the American Institute of Planners*, *43*(4), 386-400.

Horning, N. (2010). Random Forests: An algorithm for image classification and generation of continuous field data sets. *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, Osaka, Japan.

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, *112*(5), 2232-2245. doi:10.1016/j.rse.

Huete, A. R., & Escadafal, R. (1991). Assessment of biophysical soil properties through spectral decomposition techniques. *Remote Sensing of Environment*, *35*(2-3), 149-159. doi:10.1016/0034-4257(91)90008-T.

Hutchinson, M. F. (1989). A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation, and wind run. *Need for climatic and hydrologic data in agriculture in South East Asia. In Proceeding UN University workshop*, Canberra, 95-104.

Immitzer, M., Stepper, C., Böck, S., Straub, C., & Atzberger, C. (2016). Use of WorldView-2 stereo imagery and National Forest Inventory data for wall-to-wall mapping of growing stock. *Forest Ecology and Management*, *359*, 232-246. doi:10.1016/j.foreco.2015.10.018.

An integrated methodology for assessing livestock grazing capacity [IMALGC]. (2019). [Бэлчээрийн даац тооцох нэгдсэн аргачлал*]. National Statistical Office of Mongolia, the*

Ministry of Food, Agriculture, and Light Industry of Mongolia, and the Ministry of Environment and Tourism of Mongolia, Ulaanbaatar, Mongolia.

Incerti, G., Feoli, E., Salvati, L., Brunetti, A., & Giovacchini, A. (2007). Analysis of bioclimatic time series and their neural network-based classification to characterize drought risk patterns in South Italy. *International Journal of Biometeorology, 51*, 253-263.

Inoue, Y., Sakaiya, E., Zhu, Y., & Takahashi, W. (2012). Diagnostic mapping of canopy nitrogen content in rice based on hyperspectral measurements. *Remote Sensing of Environment*, *126*, 210-221. doi.10.1016/j.rse.2012.08.026.

International Soil Reference and Information Centre [ISRIC]. (2022). Accessed on 22 April 2021, Available online at https://data.isric.org.

Information Research Institute of Meteorology, Hydrology, and the Environment [IRIMHE] of Mongolia. (2020). The land degradation and desertification process in Mongolia. *The fifth assessment report.* Ulaanbaatar, Mongolia.

Information Research Institute of Meteorology, Hydrology, and the Environment [IRIMHE] of Mongolia. (2021). Annual average near-surface air temperature and annual average precipitation. Mongolian statistical information service. Accessed 05 January 2023, Available online at http://www.1212.mn.

Information Research Institute of Meteorology, Hydrology, and the Environment [IRIMHE] of Mongolia. (2022). Monthly mean climate data for the period 1991-2021 at measured weather stations. Accessed 22 June 22, Available online at http://www.1212.mn.

International Resources Panel [IRP]. (2016). Unlocking the Sustainable Potential of Land Resources: Evaluation Systems, Strategies, and Tools. A Report of the Working Group on Land and Soils of the International Resource Panel. *Herrick, J.E., O. Arnalds, B. Bestelmeyer, S. Bringezu, G. Han, M.V. Johnson, D. Kimiti, Yihe Lu, L. Montanarella, W. Pengue, G. Toth, J. Tukahirwa, M. Velayutham, L. Zha.*

Ishida, T., & Kawashima, S. (1993). Use of cokriging to estimate surface air temperature from elevation. *Theoretical and Applied Climatology*, *47*, 147-157.

Ishwaran, H. (2007). Variable importance in binary regression trees and forests. Electron. J. Stat, 519–537.

International Soil Reference and Information Centre [ISRIC], FAO. (2002). Procedures for soil analysis. *ISRIC Technical Paper*.

Ivanov, N. N. (1962). Indicator of the biological effectiveness of the climate. *WGO, 94*(1), 65-70.

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote sensing of environment*, *34*(2), 75-91. doi:10.1016/0034-4257(90)90100-Z.

Jacquemoud, S., Bacour, C., Poilvé, H., & Frangi, J. P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, *74*(3), 471-481. doi:10.1016/S0034-4257(00)00139-5.

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, *37*(3), 1181-1194.

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote sensing of Environment*, *112*(10), 3833-3845. doi:10.1016/j.rse.2008.06.006.

Jigj, S. (1990). Regional characteristics of the impact of agricultural activities on nature, the scientific basis of appropriate utilization of natural resources of the Mongolian People's Republic. *Doctoral dissertation.* Ulaanbaatar, Mongolia.

Jin, M., & Dickinson, R. E. (2010). Land surface skin temperature climatology: Benefitting from the strengths of satellite observations. *Environmental research letters*, *5*(4), 044004.

Jordan, C. F. (1969). Derivation of leaf area index from quality of light on the forest floor. *Ecology*, *50*(4), 663-666. doi:10.2307/1936256.

Keramitsoglou, I., Kiranoudis, C. T., Sismanidis, P., & Zakšek, K. (2016). An online system for nowcasting satellite-derived temperatures for urban areas. *Remote Sensing, 8*(4), 306.

Khosbayar, Ch., & Narantuya, N. (2014). Changing biomass of flood meadow community under the influence of climate change [Татмын нугын бүлгэмдлийн биомасс цаг уураас хамаарч өөрчлөгдөх нь]. *Proceeding of the Mongolian Academy of Sciences 54*(01-209), 53-61. doi:10.5564/pmas.v54i1.660.

Kidd, C., Levizzani, V., & Laviola, S. (2010). Quantitative precipitation estimation from Earth observation satellites. *Rainfall: State of the science, 191*, 127-158.

Kilibarda, M., Hengl, T., Heuvelink, G. B., Gräler, B., Pebesma, E., Perčec Tadić, M., & Bajat, B. (2014). Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*, *119*(5), 2294-2313.

King, P. (2019). Sixth Global Environment Outlook (GEO6): Outcome of United Nations Environment Assembly 4. *Institute for Global Environmental Strategies (IGES)*.

Klingebiel, A. A., & Montgomery, P. H. (1961). *Land-capability classification* (No. 210). Soil Conservation Service, US Department of Agriculture.

Kobysheva, N. V. (2005). Encyclopedia of climatic resources of the Russian Federation.

Kogan, F., Stark, R., Gitelson, A., Jargalsaikhan, L., Dugrajav, C., & Tsooj, S. (2004). Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices. *International Journal of Remote Sensing*, *25*(14), 2889-2896. doi:10.1080/01431160410001697619.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3): 259-263.

Kumar, L., Sinha, P., Taylor, S., & Alqurashi, A. F. (2015). Review of the use of remote sensing for biomass estimation to support renewable energy generation. *Journal of Applied Remote Sensing*, *9*(1), 1-29, doi:097696-097696.

Kurtzman, D., & Kadmon, R. (1999). Mapping of temperature variables in Israel: as comparison of different interpolation methods. *Climate Research*, *13*(1), 33-43.

Kust, G. S., Andreeva, O. V., & Lobkovskiy, V. A. (2020). Land degradation neutrality: the modern approach to research on arid regions at the national level. *Arid ecosystems*, *10*, 87-92. doi:10.1134/S2079096120020092.

Laliberte, A. S., Goforth, M. A., Steele, C. M., & Rango, A. (2011). Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments. *Remote Sensing*, *3*(11), 2529-2551. doi:10.3390/rs3112529.

Lam, N. S. N. (1983). Spatial interpolation methods: a review. *The American Cartographer*, *10*(2), 129-150.

Lasaponara, R. (2006). On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecological modelling*, *194*(4), 429-434. doi: 10.1016/j.ecolmodel.2005.10.035.

Laurin, G. V., Chen, Q., Lindsell, J. A., Coomes, D. A., Del Frate, F., Guerriero, L., ... & Valentini, R. (2014). Above-ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *89*, 49-58. doi:10.1016/j.isprsjprs.2014.01.001.

Lawrimore, J. H., Menne, M. J., Gleason, B. E., Williams, C. N., Wuertz, D. B., Vose, R. S., & Rennie, J. (2011). An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *Journal of Geophysical Research: Atmospheres*, *116*(D19).

Leathwick, J. R., Overton, J. M., & McLeod, M. (2003). An environmental domain classification of New Zealand and its use as a tool for biodiversity management. *Conservation biology*, *17*(6), 1612-1623.

Li, F., Jiang, L., Wang, X., Zhang, X., Zheng, J., & Zhao, Q. (2013a). Estimating grassland aboveground biomass using multitemporal MODIS data in the West Songnen Plain, China. *Journal of Applied Remote Sensing*, *7*(1), 073546-073546. doi:10.1117/1.JRS.7.073546.

Li, Z. L., Tang, B. H., Wu, H., Ren, H., Yan, G., Wan, Z., ... & Sobrino, J. A. (2013b). Satellite-derived land surface temperature: Current status and perspectives. *Remote sensing of environment*, *131*, 14-37.

Li, L., & Zha, Y. (2019). Estimating monthly average temperature by remote sensing in China. *Advances in Space Research*, *63*(8), 2345-2357.

Li, M., Wu, P., & Ma, Z. (2020). A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *International Journal of Climatology*, *40*(13), 5744-5766. doi:10.1002/joc.6549.

Li, Z. W., Xin, X. P., Huan, T. A. N. G., Fan, Y. A. N. G., Chen, B. R., & Zhang, B. H. (2017). Estimating grassland LAI using the Random Forests approach and Landsat imagery in the meadow steppe of Hulunber, China. *Journal of Integrative Agriculture*, *16*(2), 286-297. doi:10.1016/S2095-3119(15)61303-X.

Li, Z., Lun, F., Liu, M., Xiao, X., Wang, C., Wang, L., ... & Sun, D. (2021). Rapid diagnosis of agricultural soil health: A novel soil health index based on natural soil productivity and human management. *Journal of Environmental Management*, *277*, 111402. doi:10.1016/j.jenvman.2020.111402.

Liang, T., Yang, S., Feng, Q., Liu, B., Zhang, R., Huang, X., & Xie, H. (2016). Multi-factor modeling of above-ground biomass in alpine grassland: A case study in the Three-River Headwaters Region, China. *Remote Sensing of Environment*, *186*, 164-172. doi:10.1016/j.rse.2016.08.014.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Liebig, M. A., Herrick, J. E., Archer, D. W., Dobrowolski, J., Duiker, S. W., Franzluebbers, A. J., ... & Strickland, T. C. (2017). Aligning land use with land potential: the role of integrated agriculture. *Agricultural & Environmental Letters*, *2*(1), 170007. doi:10.2134/ael2017.03.0007.

Liniger, H., Harari, N., van Lynden, G., Fleiner, R., de Leeuw, J., Bai, Z., & Critchley, W. (2019). Achieving land degradation neutrality: The role of SLM knowledge in evidence-based decision-making. *Environmental science & policy*, *94*, 123-134. doi:10.1016/j.envsci.2019.01.001.

Liu, H. Q., & Huete, A. (1995). A feedback-based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and remote sensing*, *33*(2), 457-465. http://ieeexplore.ieee.org/document/370361.

Liu, Y., Peng, J., & Wang, Y. (2018). Application of partial least squares regression in detecting the important landscape indicators determining urban land surface temperature variation. *Landscape Ecology*, *33*, 1133-1145.

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International journal of remote sensing*, *27*(7), 1297-1328. doi:10.1080/01431160500486732.

Lu, L., Zhang, T., Wang, T., & Zhou, X. (2018). Evaluation of collection-6 MODIS land surface temperature product using multi-year ground measurements in an arid area of Northwest China. *Remote Sensing*, *10*(11), 1852.

Maitra, S., & Yan, J. (2008). Principle component analysis and partial least squares: Two-dimension reduction techniques for regression. *Applying Multivariate Statistical Models*, *79*, 79-90.

Makhaya, Z., Odindi, J., & Mutanga, O. (2022). The influence of bioclimatic and topographic variables on grassland fire occurrence within an urbanized landscape. *Scientific African*, *15*, e01127. doi:10.1016/j.sciaf.2022.e01127.

Malczewski, J. (1996). A GIS-based approach to multiple criteria group decision-making. *International Journal of Geographical Information Systems*, *10*(8), 955-971.

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. John Wiley & Sons.

Malczewski, J. (2004). GIS-based land-use suitability analysis: a critical overview. *Progress in planning*, *62*(1), 3-65.

Malczewski, J. (2006). Multicriteria Decision Analysis for Collaborative GIS. *Collaborative geographic information systems*, 167-185.

Malczewski, J., & Rinner, C. (2015). *Multicriteria decision analysis in geographic information science* (Vol. 1, pp. 55-77). New York: Springer.

Marchi, M., Sinjur, I., Bozzano, M., & Westergren, M. (2019). Evaluating WorldClim Version 1 (1961–1990) as the baseline for sustainable use of forest and environmental resources in a changing climate. *Sustainability*, *11*(11), 3043.

Mariathasan, V., Bezuidenhoudt, E., & Olympio, K. R. (2019). Evaluation of earth observation solutions for Namibia's SDG monitoring system. *Remote Sensing*, *11*(13), 1612. doi 10.3390/rs11131612.

Martonne, E. (1925). Aréisme et indices article, Compte Rendu de L'Acad. Sci., Paris 182, 1395-1398

Mongolia Academy of Sciences [MAS]. (1990*). Mongolian People's Revolutionary (MPR): Information Mongolia*. Oxford: Pergamon Press.

Mattiuzzi, M., Verbesselt, J., Hengel, T., Klisch, A., Stevens, F., Mosher, S., Evans, B., Lobo, A., Hufkens, K., & Detsch, F. (2019). MODIS acquisition and processing package. R package version 1.1.5. Accessed on 29 September 2019, Available online at https://cran.r-project.org/web/packages/MODIS/MODIS.pdf

Mattiuzzi, M., Verbesselt, J., Klisch, A., Stevens, F., Mosher, S., Evans, B., Lobo, A., & Detsch, F. (2018). MODIS acquisition and processing package. R package version 1.1.4. Accessed on 25 October 2018, Available online at https://cran.r-project.org/src/contrib/Archive/MODIS/MODIS.pdf

Mongolian Customs [MC]. (2022). Mongolian customs: trade statistics data. Accessed 22 June 2022, Available online at http://www.customs.gov.mn/en/services/trade.

McHarg, I. L. (1996). *A quest for life: an autobiography*. John Wiley & Sons.

McKenzie, N. J., Grundy, M. J., Webster, R., & Ringrose-Voase, A. J. (Eds.). (2008). *Guidelines for surveying soil and land resources*. CSIRO publishing.

Mesquita, S., & Sousa, A. J. (2009). Bioclimatic mapping using geostatistical approaches: application to mainland Portugal. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *29*(14), 2156-2170.

Mevik, B. H., & Cederkvist, H. R. (2004). The mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *Journal of Chemometrics*, *18*(9), 422-429.

Meyer, H., Katurji, M., Appelhans, T., Müller, M. U., Nauss, T., Roudier, P., & Zawar-Reza, P. (2016). Mapping daily air temperature for Antarctica based on MODIS LST. *Remote Sensing*, *8*(9), 732.

Mezentsev, V. S. (1969). Humidity of the west Siberian plain. 168 p.

Mijiddorj, J. (2016). Climate change and its impact on agriculture, and food security. *Mongolian Journal of Agricultural Sciences*, *14*(1), 251-257.

Mijiddorj, S. (1985). Comprehensive mapping of the natural resources of the Mongolian People's Republic based on satellite data. *Doctoral dissertation*, Ulaanbaatar, Mongolia.

Mkhabela, M. S., Bullock, P. R., & Sapirstein, H. D. (2018). Characterizing the most critical climatic parameters that impact the quality of spring wheat (Triticum aestivum L.) on the Canadian Prairies using partial least squares (PLS) analysis. *Journal of Cereal Science*, *81*, 44-51.

Ministry of Food Agriculture and Light Industry [MoFALI] of Mongolia. (2021). Statistical information of the agricultural sector of Mongolia. Accessed on 22 June 2022, Available online at https://www.mofa.gov.mn/home

Montgomery, D. R., & Biklé, A. (2021). Soil health and nutrient density: beyond organic vs. conventional farming. *Frontiers in Sustainable Food Systems*, *5*, 417.

Mostovoy, G. V., King, R. L., Reddy, K. R., Kakani, V. G., & Filippova, M. G. (2006). A statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of Mississippi. *GIScience & Remote Sensing*, *43*(1), 78-110.

Mundava, C., Helmholz, P., Schut, A. G., Corner, R., McAtee, B., & Lamb, D. W. (2014). Evaluation of vegetation indices for rangeland biomass estimation in the Kimberley area of Western Australia. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*, 47-53. doi:10.5194/isprsannals-II-7-47-2014.

Munkhdulam, O., Jargaltulga, Ts., Mygmarjav, M., & Jigjidsuren, S. (2015). Assessment of land use suitability in Mongolia (cropland, pastureland, forest, urban, preserve and conserve, tourism and recreation, road and line network). *Scientific report.*

Munkhdulam, O., Clement, A., Amarsaikhan, D., Yokoyama, S., Erdenesukh, S., & Sainbayar, D. (2022). Detection of Anthropogenic and Environmental Degradation in Mongolia Using

Multi-Sources Remotely Sensed Time Series Data and Machine Learning Techniques. In *Environmental Degradation in Asia: Land Degradation, Environmental Contamination, and Human Activities* (pp. 17-47). Cham: Springer International Publishing.

Mutanga, O., Adam, E., & Cho, M. A. (2012). High-density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, *18*, 399-406. doi:10.1016/j.jag.2012.03.012.

Mysterud, A. (2006). The concept of overgrazing and its role in the management of large herbivores. *Wildlife Biology*, *12*(2), 129-141.

Nandintsetseg, B., Shinoda, M., & Erdenetsetseg, B. (2018). Contributions of multiple climate hazards and overgrazing to the 2009/2010 winter disaster in Mongolia. *Natural Hazards*, *92*, 109-126. doi.org/10.1007/s11069-017-2954-8.

National Statistics Office of Mongolia [NSOM]. (2016). Sown area. *Mongolian statistical information service.* Accessed on 22 June 2017, Available online at http://www.1212.mn.

National Statistics Office of Mongolia [NSOM]. (2022a). The unified land territory of Mongolia. *Mongolian statistical information service.* Accessed on 22 June 2023, Available online at http://www.1212.mn.

National Statistics Office of Mongolia [NSOM]. (2022b). The number of livestock for the period 1991-2021. *Mongolian statistical information service.* Accessed on 22 June 2023, Available online at http://www.1212.mn.

National Statistics Office of Mongolia [NSOM]. (2022c). The number population for the period 1991-2021. *Mongolian statistical information service.* Accessed on 22 June 2023, Available online at http://www.1212.mn.

National Statistics Office of Mongolia [NSOM]. (2022d). Gross domestic product, and foreign exports for the period 2002-2022. *Mongolian statistical information service.* Accessed on 22 June 2023, Available online at http://www.1212.mn.

National Statistics Office of Mongolia [NSOM]. (2022e). The number of households with livestock, and herder households. *Mongolian statistical information service.* Accessed on 22 June 2023, Available online at http://www.1212.mn.

Natsagdorj, L. (2003). Some aspects of assessment of the dzud phenomena. *Meteorology and Hydrology*, *23*, 3-18.

Natsagdorj, L., Munkhbat, B., & Gomboluudev, P. (2019). Climate biocapacity of Mongolia and its change. *Proceedings of the Mongolian Academy of Sciences*, 54-70.

Neupert, R. F., & Goldstein, S. (1994). Urbanization and population redistribution in Mongolia.

Nha, D. V. (2017). The role of land-use planning on socio-economic development in Mai Chau District, Vietnam. In *Redefining Diversity & Dynamics of Natural Resources Management in Asia, Volume 2* (pp. 87-111). Elsevier.

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E., & Stisen, S. (2011). Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. *Remote Sensing of Environment*, *115*(1), 107-116.

Nikolova, N., & Vassilev, S. (2006). Mapping precipitation variability using different interpolation methods. In *Proceedings of the conference on water observation and information system for decision support (BALWOIS)*, 25-29 May, Bulgaria, 25-29.

Noi, P. T., Kappas, M., & Degener, J. (2016). Estimating daily maximum and minimum land air surface temperature using MODIS land surface temperature data and ground truth data in Northern Vietnam. *Remote Sensing*, *8*(12), 1002.

Numata, I., Roberts, D. A., Chadwick, O. A., Schimel, J., Sampaio, F. R., Leonidas, F. C., & Soares, J. V. (2007). Characterization of pasture biophysical properties and the impact of grazing intensity using remotely sensed data. *Remote sensing of Environment*, *109*(3), 314-327. doi:10.1016/j.rse.2007.01.013.

Nyamtseren, M., Feng, Q., & Deo, R. (2018). A comparative study of temperature and precipitation-based aridity indices and their trends in Mongolia. *International Journal of Environmental Research*, *12*, 887-899. doi:10.1007/s41742-018-0143-6.

O'Connell, D., Abel, N., Grigg, N., Maru, Y., Butler, J., Cowie, A., ... & Stafford Smith, M. (2016). Designing projects in a rapidly changing world: guidelines for embedding resilience, adaptation, and transformation into sustainable development projects (Version 1.0). *Global Environment Facility, Washington, DC*.

O'Donnell, M. S., & Ignizio, D. A. (2012). Bioclimatic predictors for supporting ecological applications in the conterminous United States. *US geological survey data series*, *691*(10), 4-9.

Oke, T. R. (1988). The urban energy balance. *Progress in Physical Geography*, *12*(4), 471-508.

Ólafur, A. (2019). Land potential: What our land can do. Accessed 22 September 22, Available online at https://www.coursera.org/lecture/landscape-restoration-sustainable-development/land-potential-what-our-land-can-do-cuxRs

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., ... & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*(11), 933-938.

Openshaw, S., & Openshaw, C. (1997). *Artificial intelligence in geography*. John Wiley & Sons, Inc.

Otgonbayar, M., Atzberger, C., Mattiuzzi, M., & Erdenedalai, A. (2019). Estimation of climatologies of average monthly air temperature over Mongolia using MODIS land surface temperature (LST) time series and machine learning techniques. *Remote Sensing*, *11*(21), 2588.

Otgonbayar, M., & Sumya, E. (2020). Assessment of the bioclimatic potential in Mongolia [Монгол орны био-уур амьсгалын чадавхийн үнэлгээ]. *Landscape ecological potential of Mongolia [Монгол орны ландшафтын экологийн чадавх], Eds.* Erdenedalai, A (Chap 3, pp.139-161), Ulaanbaatar, Mongolia, Namnan design press. ISBN978-9919-23-153-8.

Otgonbayar, M., Sumiya, E., & Tovuudorj, R. (2021a). Estimating spatial distribution of aridity and moisture indices of Mongolia using remotely sensed time series data. *Proceedings of the Mongolian Academy of Sciences*, 20-29.

Otgonbayar, M., Jargalsaikhan, G., Tseveengerel, B., & Buyanaa, Ch. (2021b). Estimating the spatial distribution of grassland biomass using environmental variables and PLS regression model. *In MOGGA scientific magazine, Institute of Geography and Geoecology, Mongolian Academy of Sciences, №42,* 101-122. Available online: http://portal.igg.ac.mn/dataset/4f1f6da7-c323-435d-9517-07c8b179f6c9.

Otgonbayar. M., Erdenedalai, E., Chambers, J., & Tovuudorj, R. (2018). Assessment of the landscape-ecological potential of Khovd province, Mongolia using satellite imagery and the spatial multi-criteria decision-making method. *GSJ: Volume 6, Issue 10.*

Otgonbyar, M., Erdenedalai, A., & Dalantai, S. (2017). Assessment bioclimate potential of Mongolia based on satellite imagery and meteorological observation data. *Proceeding of Mongolian Academy of Sciences, 57, 03*(223). doi:10.5564/pmas.v57i3.885.

Oyler, J. W., Dobrowski, S. Z., Holden, Z. A., & Running, S. W. (2016). Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. *Journal of Applied Meteorology and Climatology*, *55*(7), 1441-1457.

Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222. doi:10.1080/01431160412331269698.

Paredes-Trejo, F. J., Barbosa, H. A., & Kumar, T. L. (2017). Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. *Journal of arid environments*, *139*, 26-40.

PCI Geomatica. (2015a). Color Balancing-New Automatic Color Balancing Algorithm (Bundle). Accessed on 08 May 2015, Available online at https://www.youtube.com/watch?v=sXOvaNZKkNo.

PCI Geomatica. (2015b). Color Balancing- New Interactive Mosaic Editing Tools. Accessed on 10 May 2015, Available online at https://www.youtube.com/watch?v=0_z7nmpdGtk.

PCI Geomatica. (2017). Geomatica training guide: Geomatica Orthengine course exercise. Available online: http://www.pcigeomatics.com/pdf/TrainingGuide-Geomatica-1.pdf.

Peng, C. (2000). From static bio-geographical model to dynamic global vegetation model: a global perspective on modeling vegetation dynamics. *Ecological modeling*, 135 (1): 33-54.

Peón, J., Recondo, C., & Calleja, J. F. (2014). Improvements in the estimation of daily minimum air temperature in peninsular Spain using MODIS land surface temperature. *International Journal of Remote Sensing*, *35*(13), 5148-5166.

Pinzon, J. E., & Tucker, C. J. (2014). A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote sensing*, *6*(8), 6929-6960.

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, *9*, 181-199.

Prata, A. J., Caselles, V., Coll, C., Sobrino, J. A., & Ottle, C. (1995). Thermal remote sensing of land surface temperature from satellites: Current status and prospects. *Remote sensing reviews*, *12*(3-4), 175-224.

Price, K., Purucker, S. T., Kraemer, S. R., Babendreier, J. E., & Knightes, C. D. (2014). Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. *Hydrological Processes*, *28*(9), 3505-3520.

Prihodko, L., & Goward, S. N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, *60*(3), 335-346.

Prince, S. D., & Tucker, C. J. (1986). Satellite remote sensing of rangelands in Botswana II. NOAA AVHRR and herbaceous vegetation. *International Journal of Remote Sensing*, *7*(11), 1555-1570. doi:10.1080/01431168608948953.

Prince, S. D. (1991). Satellite remote sensing of primary production: comparison of results for Sahelian grasslands 1981-1988. *International Journal of remote sensing*, *12*(6), 1301-1311. doi:10.1080/01431169108929727.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil-adjusted vegetation index. *Remote sensing of environment*, *48*(2), 119-126. doi:10.1016/0034-4257(94)90134-1.

R Core Team. A language and environment for statistical computing. Accessed on 31 October 2019, Available online at https://www.R-project.org/.

Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS journal of photogrammetry and remote sensing*, *57*(4), 241-262.

Ramoelo, A., Cho, M. A., Mathieu, R., Madonsela, S., Van De Kerchove, R., Kaszta, Z., & Wolff, E. (2015). Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International journal of applied earth observation and geoinformation*, *43*, 43-54. doi:10.1016/j.jag.2014.12.010.

Ren, H., Zhou, G., & Zhang, X. (2011). Estimation of green aboveground biomass of desert steppe in Inner Mongolia based on red-edge reflectance curve area method. *Biosystems Engineering*, *109*(4), 385-395. doi:10.1016/j.biosystemseng.2011.05.004.

Richter, K., Atzberger, C., Hank, T. B., & Mauser, W. (2012). Derivation of biophysical variables from Earth observation data: validation and statistical measures. *Journal of Applied Remote Sensing*, *6*(1), 063557-063557.

Ripley, B. D. (2001). The R project in statistical computing. *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network*, *1*(1), 23-25.

Rivera-Caicedo, J. P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., & Moreno, J. (2017). Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS journal of photogrammetry and remote sensing*, *132*, 88-101. doi:10.1016/j.isprsjprs.2017.08.012.

Robeson, S. M. (1994). Influence of spatial sampling and interpolation on estimates of air temperature change. *Climate Research*, *4*(2), 119-126. doi:10.3354/cr004119.

Roca, R., Alexander, L. V., Potter, G., Bador, M., Jucá, R., Contractor, S., ... & Cloché, S. (2019). FROGS: a daily 1× 1 gridded precipitation database of the rain gauge, satellite, and reanalysis products. *Earth System Science Data*, *11*(3), 1017-1035.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, *67*, 93-104. doi:10.1016/j.isprsjprs.2011.11.002.

Rossel, R. V., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near-infrared, mid-infrared, or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*(1-2), 59-75. doi:10.1016/j.geoderma.2005.03.007.

Rossiter, D. G. (1996). A theoretical framework for land evaluation. *Geoderma*, *72*(3-4), 165-190.

System for Automated Geo-scientific Analyses [SAGA]. (2013). Assessed 22 May 2020. Available online at www.saga-gis. org/en/index.

Salas, E. A. L., Seamster, V. A., Boykin, K. G., Harings, N. M., & Dixon, K. W. (2017). Modeling the impacts of climate change on Species of Concern (birds) in South Central US based on bioclimatic variables. *AIMS Environmental Science*, *4*(2), 358-385.

Santos, J. R. D., Lacruz, M. P., Araujo, L. S., & Keil, M. (2002). Savanna and tropical rainforest biomass estimation and spatialization using JERS-1 data. *International Journal of Remote Sensing*, *23*(7), 1217-1229. doi:10.1080/01431160110092867.

Green Gold Program, Swiss Agency for Development and Cooperation [GGP-SDC]. (2015). *National report on the rangeland health of Mongolia*. Accessed on 22 September 2022, Available online at https://jornada.nmsu.edu/files/Mongolia-Rangeland-health-Report_EN.pdf.

Selyaninov, G. T. (1937). Methods of climate description for agricultural purposes. *World climate and agriculture handbook. Gidrometeoizdat, Leningrad*, 5-27.

Seybold, C. A., Herrick, J. E., & Brejda, J. J. (1999). Soil resilience: a fundamental component of soil quality. *Soil science*, *164*(4), 224-234. doi:10.1097/00010694-199904000-00002.

Shamir, E., & Georgakakos, K. P. (2014). MODIS Land Surface Temperature as an index of surface air temperature for operational snowpack estimation. *Remote Sensing of Environment*, *152*, 83-98.

Sims, N. C., England, J. R., Newnham, G. J., Alexander, S., Green, C., Minelli, S., & Held, A. (2019). Developing good practice guidance for estimating land degradation in the context of the United Nations Sustainable Development Goals. *Environmental Science & Policy*, *92*, 349-355. doi.org/10.1016/j.envsci.2018.10.014.

Sohn, L. B. (1973). Stockholm declaration on the human environment, the. *Harv. Int'l. LJ*, *14*, 423.

Sombroek, W. G. (1993). Agricultural use of the physical resources of Africa: Achievements, constraints, and future needs. *Sustainable food production in sub-Saharan Africa*, *2*, 12-30.

Sombroek, W. G. (1997). Land resources evaluation and the role of land-related indicators. *FAO Land and Water Bulletin (FAO)*.

Sternberg, T., & Batbuyan, B. (2013). Integrating the Hyogo Framework into Mongolia's disaster risk reduction (DRR) policy and management. *International Journal of Disaster Risk Reduction*, *5*, 1-9.

Sternberg, T. (2010). Unravelling Mongolia's extreme winter disaster of 2010. *Nomadic Peoples*, *14*(1), 72-86.

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., & Eklundh, L. (2007). Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. *Remote sensing of Environment*, *110*(2), 262-274.

Stoll, M. J., & Brazel, A. J. (1992). Surface-air temperature relationships in the urban environment of Phoenix, Arizona. *Physical Geography*, *13*(2), 160-179.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, *56*(1), 79-107.

Sun, J., Passi, K., & Jain, C. K. (2016). Improved microarray data analysis using feature selection methods with machine learning methods. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1527-1534.

Sun, Y. J., Wang, J. F., Zhang, R. H., Gillies, R. R., Xue, Y. Y. C. B., & Bo, Y. C. (2005). Air temperature retrieval from remote sensing data based on thermodynamics. *Theoretical and applied climatology*, *80*, 37-48.

Suttie, J. M. (2005). Grazing management in Mongolia. *Grasslands of the World*, *34*, 265.

Sykes, M. T., Prentice, I. C., & Cramer, W. (1996). A bioclimatic model for the potential distributions of north European tree species under present and future climates. *Journal of Biogeography*, 203-233.

Sys, I. C., Vanranst, B., & Debaveye, J. (1991). The land evaluation part II, Methods in land evaluation Agric. *Pub. General administration for development co-operation, place, du, camp de Mars, 5bte*, 57-1050.

Teich, I., Gonzalez Roglich, M., Corso, M. L., & García, C. L. (2019). Combining earth observations, cloud computing, and expert knowledge to inform national-level degradation assessments in support of the 2030 development agenda. *Remote Sensing*, *11*(24), 2918. doi.org/10.3390/rs11242918.

Thiébaux, H. J. (1991). Statistics and the environment: The analysis of large-scale earth-oriented systems. *Environmetrics*, *2*(1), 5-24.

Thompson, R. S., Shafer, S. L., Anderson, K. H., Strickland, L. E., Pelltier, R. T., Bartlein, P. J., & Kerwin, M. W. (2004). Topographic, bioclimatic, and vegetation characteristics of three ecoregion classification systems in North America: comparisons along continent-wide transects. *Environmental Management, 34*, S125-S148.

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review, 38*(1), 55-94.

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature, 418*(6898), 671-677.

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves the performance of ecological niche modeling. *Ecography, 41*(2), 291-307.

Todd, S. W., Hoffer, R. M., & Milchunas, D. G. (1998). Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International journal of remote sensing*, *19*(3), 427-438. doi:10.1080/014311698216071.

Trends.Earth. Conservation International. Available online: http://trends.earth. 2022.

Trevor, H., Tibshirani, R., Friedman, J. (2011). The elements of statistical learning. *Springer: Stanford, CA*, USA.

Trudgill, S. T., & Briggs, D. J. (1979). Soil and land potential. *Progress in Physical Geography*, *3*(2), 283-299. doi:10.1177/030913338200600206.

Tserendash, S., & Altanzul, Ts. (2013). The rangeland management booklet [Бэлчээрийн менежментийн гарын авлага]. Accessed on 25 September 2017, Available online at http://crrl.mn/new/medee/buteeluud/Pasture_manegment_booklet.

Tserendash, S., Tserendulam, R., Buyan-Orshikh, H., Narantuya, N., Oyuntsetseg, Ch., & Khudulmur, S. (2000). Mongolia's pastureland potential, ecology, and quality assessment. *Edited by Tserendash, S. The Research Institute of Animal Husbandry Mongolian Academy of Sciences,* Ulaanbaatar, Mongolia.

Tse-Wei, C., Yi-Ling, C., & Shao-Yi, C. (2008). Fast Image Segmentation Based on K-Means Clustering with Histograms in HSV Color Space. *Multimedia Signal Processing, IEEE 10th Workshop, Cairns, Qld, Australia*, 8-10 October. doi:10.1109/MMSP.2008.4665097.

Tsogtbaatar, J., & Khudulmur, S. (2014). Desertification atlas of Mongolia. Admon Press, Ulaanbaatar, Mongolia.

Tsogtbayar, J., & Otgonbayar, M. (2014). GIS-based assessment of cropland suitability of optimal allocation in the Jargalant subprovince of Mongolia. *Proceeding of the Khurel Togoot-2014 scientific conference.* 141-147.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., ... & El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset is compatible with MODIS and SPOT vegetation NDVI data. *International journal of remote sensing*, *26*(20), 4485-4498.

Tyrwhitt, J. (1950). *Surveys for planning. In Town and country planning textbook*. Architectural Press, London.

Ullah, S., Si, Y., Schlerf, M., Skidmore, A. K., Shafique, M., & Iqbal, I. A. (2012). Estimation of grassland biomass and nitrogen using MERIS data. *International journal of applied earth observation and geoinformation*, *19*, 196-204. doi:10.1016/j.jag.2012.05.008.

United Nations [UN]. (1995). Chapter 40: Information for decision-making and Earthwatch. *Commission on Sustainable Development, Economic and Social Council E/CN*. United Nations, New York.

United Nations Conference on Environment and Development [UNCED]. (1993). Agenda 21: Program of action for sustainable development. *UN Conference on Environment and Development.* United Nations, New York.

United Nations Environment Program [UNEP]. (1993). Agro-ecological assessments for national planning: the example of Kenya. *FAO Soils Bulletin*, *67*, 154.

United Nations Convention to Combat Desertification [UNCCD]. (1994). Elaboration of an International Convention to Combat Desertification in Countries Experiencing Serious Drought and/or Desertification, Particularly in Africa. *UN. Doc. A/AC.241/27, 33 I.L.M. 1328,* United Nations.

United Nations Convention to Combat Desertification [UNCCD]. (2016). Achieving Land Degradation Neutrality at the country level. *Building blocks for LDN target setting.*

United States Department of Agriculture [USDA]. (1983). National agricultural land evaluation and site assessment handbook. *US Government Printing Office*, Washington DC.

United States Department of Agriculture-Natural Resources Conservation Service [USDA-NRCS]. (2013). The PLANTS Database. *National Plant Data Team,* Greensboro, North Carolina.

United States Department of Agriculture- Natural Resources Conservation Service [USDA-NRCS]. (2023). Accessed on 25 May 2023, Available online at https://www.nrcs.usda.gov/conservation-basics/natural-resource-concerns/land.

Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S. J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, *114*(2), 449-465.

Vanselow, K. A., & Samimi, C. (2014). Predictive mapping of dwarf shrub vegetation in an arid high mountain ecosystem using remote sensing and random forests. *Remote Sensing*, *6*(7), 6709-6726. doi:10.3390/rs6076709.

Vega, G. C., Pertierra, L. R., & Olalla-Tárraga, M. Á. (2018). Erratum: MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific Data*, *5*.

Verheye, W. H. (1997). Land use planning and national soil policies. *Agricultural Systems, 53*(2-3), 161-174.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273-290. doi:10.1016/j.isprsjprs.2015.05.005.

Verrelst, J., Schaepman, M. E., Koetz, B., & Kneubühler, M. (2008). Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sensing of Environment*, *112*(5), 2341-2353. doi:10.1016/j.rse.2007.11.001.

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P., ... & Moreno, J. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, *40*, 589-629. doi:10.1007/s10712-018-9478-y.

Vink, A. P. A. (1975). Land resources. *Land Use in Advancing Agriculture*, 66-130.

Viovy, N. (2018). CRUNCEP version 7-atmospheric forcing data for the community land model. *Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*, *10*.

Vogt, J. V., Viau, A. A., & Paquet, F. (1997). Mapping regional air temperature fields using satellite-derived surface skin temperatures. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *17*(14), 1559-1579.

Voropay, N. N., & Ryazanova, A. A. (2018). A comparative assessment of the aridity indices for analysis of the hydrothermal conditions. In *IOP Conference Series: Earth and Environmental Science*, *190*(1), 012041. IOP Publishing. doi:10.1088/1755-1315/190/1/012041.

Walker, B. H., Steffen, W. L., & Langridge, J. (1999). Interactive and integrated effects of global change on terrestrial ecosystems. *The terrestrial biosphere and global change: implications for natural and managed ecosystems*, 329-375.

Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, A. S., Perkins, S. L., & Guralnick, R. P. (2007). Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS one*, *2*(7), e563.

Waltari, E., Schroeder, R., McDonald, K., Anderson, R. P., & Carnaval, A. (2014). Bioclimatic variables derived from remote sensing: Assessment and application for species distribution modeling. *Methods in Ecology and Evolution*, *5*(10), 1033-1042.

Walther, G. R., Berger, S., & Sykes, M. T. (2005). An ecological 'footprint of climate change. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1571), 1427-1432.

Wan, Z. (2007). Collection-5 MODIS land surface temperature products users' guide. *ICESS, University of California, Santa Barbara*, *400*.

Wan, Z. (2008). New refinements and validation of the MODIS land-surface temperature/emissivity products. *Remote sensing of Environment*, *112*(1), 59-74.

Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote sensing of Environment*, *140*, 36-45.

Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and remote sensing*, *34*(4), 892-905.

Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. L. (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote sensing of Environment*, *83*(1-2), 163-180.

Wang, Y., Wu, G., Deng, L., Tang, Z., Wang, K., Sun, W., & Shangguan, Z. (2017). Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Scientific reports*, *7*(1), 6940. doi:10.1038/s41598-017-07197-6.

Weissmann, H., & Shnerb, N. M. (2014). Stochastic desertification. *Europhysics Letters*, *106*(2), 28004.

Willmott, C. J., & Robeson, S. M. (1995). Climatologically aided interpolation (CAI) of terrestrial air temperature. *International Journal of Climatology*, *15*(2), 221-229.

Willmott, C. J., Rowe, C. M., & Philpot, W. D. (1985). Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *The American Cartographer*, *12*(1), 5-16.

Wold, H. (1982). Soft modeling: the basic design and some extensions. *Systems under indirect observation, Part II*, 36-37.

Wold, S., Ruhe, A., Wold, H., & Dunn, Iii, W. J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, *5*(3), 735-743. doi:10.1137/0905052.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, *58*(2), 109-130.

World Meteorological Organization [WMO]. (2014). Climate Data Management System Specifications. Assessed on 22 May 2020.

Wu, C., Shen, H., Shen, A., Deng, J., Gan, M., Zhu, J., ... & Wang, K. (2016). Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. *Journal of Applied Remote Sensing*, *10*(3), 035010-035010. doi:10.1117/1.JRS.10.035010.

Wylie, B. K., Meyer, D. J., Tieszen, L. L., & Mannel, S. (2002). Satellite mapping of surface biophysical parameters at the biome scale over the North American grasslands: A case

study. *Remote sensing of environment*, *79*(2-3), 266-278. doi:10.1016/S0034-4257(01)00278-4.

Xiao, J., Shen, Y., Ge, J., Tateishi, R., Tang, C., Liang, Y., & Huang, Z. (2006). Evaluating urban expansion and land use change in Shijiazhuang, China, by using GIS and remote sensing. *Landscape and urban planning*, *75*(1-2), 69-80. doi:10.1016/j.landurbplan.2004.12.005.

Xiaoping, W., Ni, G., Kai, Z., & Jing, W. (2011). Hyperspectral remote sensing estimation models of aboveground biomass in Gannan rangelands. *Procedia Environmental Sciences*, *10*, 697-702. doi:10.1016/j.proenv.2011.09.112.

Xie, Y., Sha, Z., Yu, M., Bai, Y., & Zhang, L. (2009). A comparison of two models with Landsat data for estimating above-ground grassland biomass in Inner Mongolia, China. *Ecological Modelling*, *220*(15), 1810-1818. doi:10.1016/j.ecolmodel.2009.04.025.

Xu, Y., Knudby, A., & Ho, H. C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *International Journal of Remote Sensing*, *35*(24), 8108-8121.

Yan, H., Zhang, J., Hou, Y., & He, Y. (2009). Estimation of air temperature from MODIS data in East China. *International Journal of Remote Sensing*, *30*(23), 6261-6275.

Yang, Y. Z., Cai, W. H., & Yang, J. (2017). Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. *Remote Sensing*, *9*(5), 410.

Yoo, C., Im, J., Park, S., & Quackenbush, L. J. (2018). Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. *ISPRS journal of photogrammetry and remote sensing*, *137*, 149-162.

Young, A. (2000). *Land resources: now and for the future*. Cambridge University Press.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338-353.

Zadeh, L. A. (1990). The birth and evolution of fuzzy logic. *International Journal of General System*, *17*(2-3), 95-105.

Zadeh, L. A. (2007). Fuzzy logic is the logic of natural languages. In *Analysis and Design of Intelligent Systems Using Soft Computing Techniques*, 1-2. Berlin, Heidelberg: Springer Berlin Heidelberg.

Zakšek, K., & Schroedter-Homscheidt, M. (2009). Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(4), 414-421.

Zandler, H., Brenning, A., & Samimi, C. (2015). Quantifying dwarf shrub biomass in an arid environment: comparing empirical methods in a high dimensional setting. *Remote Sensing of Environment*, *158*, 140-155. doi:10.1016/j.rse.2014.11.007.

Zhang, H., Zhang, F., Zhang, G., He, X., & Tian, L. (2016). Evaluation of cloud effects on air temperature estimation using MODIS LST based on ground measurements over the Tibetan Plateau. *Atmospheric Chemistry and Physics*, *16*(21), 13681-13696.

Zhou, W., Peng, B., Shi, J., Wang, T., Dhital, Y. P., Yao, R., ... & Zhao, R. (2017). Estimating high-resolution daily air temperature based on remote sensing products and climate reanalysis datasets over glacierized basins: a case study in the Langtang Valley, Nepal. *Remote Sensing*, *9*(9), 959.

Zhu, W., Lű, A., & Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, *130*, 62-73.

Appendix A

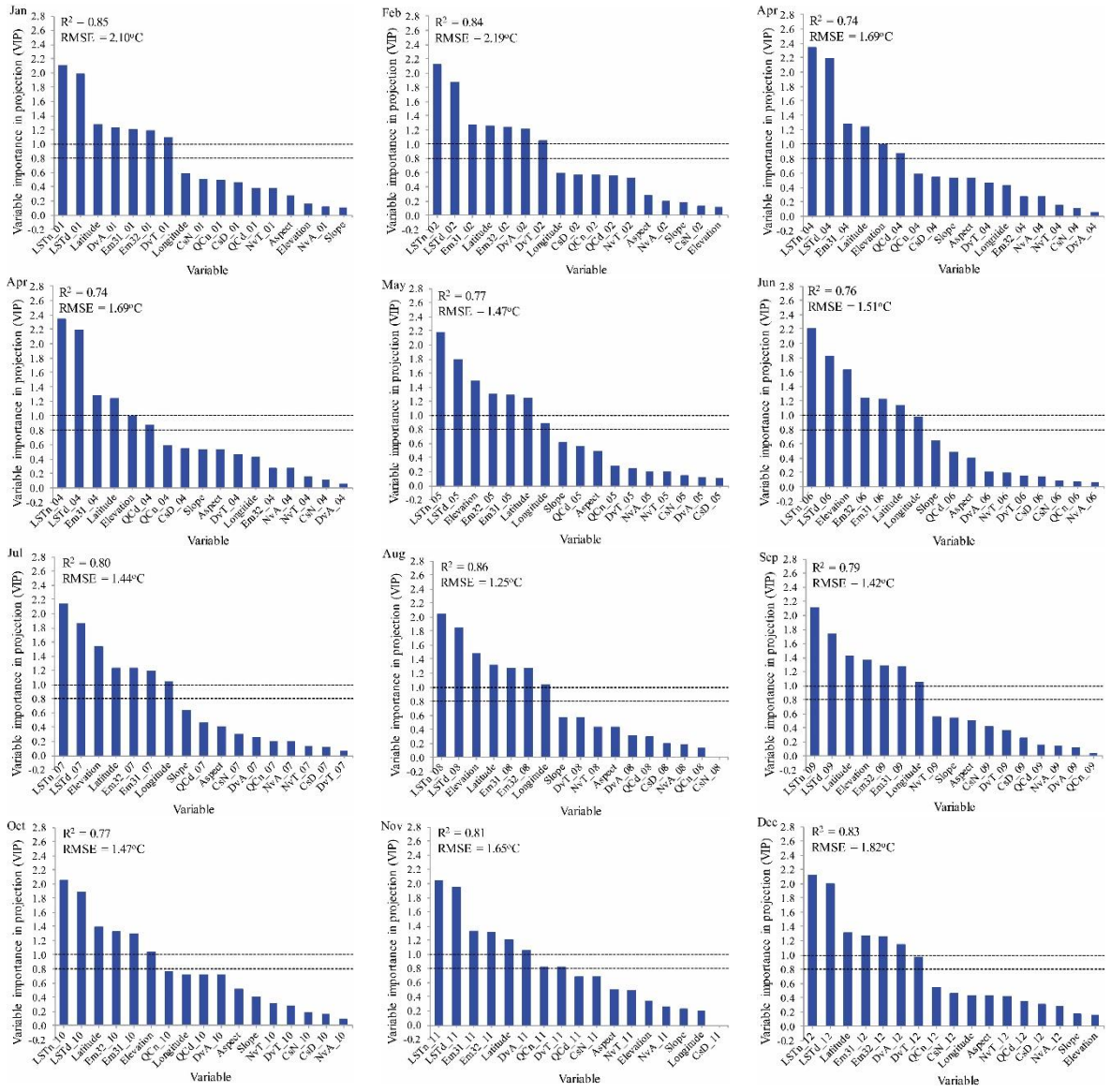


Figure A. 1 Variable importance in projections (VIPs), R^2, and RMSE for the 17 predictor variables in the twelve-monthly PLSR models.

Table A. 1 Model coefficients for the twelve-monthly PLS regression models including all 17 predictor variables. In the last column, the model of the annual average Ta is also included.

Variable	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Intercept	-52.83	-93.62	26.17	55.52	93.68	89.37	97.27	45.11	68.76	62.92	53.54	22.17	82.62
LSTd	0.255	0.219	0.208	0.185	0.159	0.146	0.137	0.178	0.150	0.128	0.197	0.234	0.319
LSTn	0.420	0.457	0.425	0.265	0.223	0.232	0.227	0.284	0.210	0.196	0.344	0.407	0.464
CsD	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
CsN	-0.014	-0.003	0.007	0.000*	0.001	0.001	0.002	0.000*	0.004	0.000*	0.000*	-0.011	0.000*
DvA	0.015	0.015	0.007	-0.002	0.005	0.008	0.011	0.021	0.006	0.022	0.003	0.004	-0.007
DvT	0.094	0.109	-0.015	-0.130	-0.072	-0.041	-0.019	0.003	-0.105	0.061	-0.044	-0.021	0.095
Em31	0.069	0.104	0.021	-0.152	-0.133	-0.126	-0.131	-0.050	-0.127	-0.125	-0.047	0.006	-0.079
Em32	0.101	0.158	0.003	-0.012	-0.208	-0.198	-0.209	-0.096	-0.199	-0.195	-0.087	-0.009	0.007
NvA	–0.019	-0.034	-0.034	-0.012	-0.009	-0.002	-0.007	-0.017	0.007	-0.004	-0.017	-0.028	-0.038
NvT	0.031	0.100	-0.114	0.045	0.053	0.050	0.034	0.029	0.169	0.083	-0.030	-0.038	-0.413
QCd	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
QCn	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
Elevation	0.000*	0.000*	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001	-0.001	-0.001	0.000*	-0.001
Slope	–0.021	0.028	-0.009	-0.061	-0.059	-0.062	-0.065	-0.047	-0.051	-0.037	-0.025	-0.031	0.026
Aspect	–0.002	-0.001	-0.001	-0.002	-0.002	-0.001	-0.002	-0.001	-0.002	-0.002	-0.003	-0.003	0.003
Latitude	–0.411	-0.401	-0.248	-0.268	-0.236	-0.214	-0.246	-0.244	-0.260	-0.248	-0.241	-0.393	0.336
Longitude	0.099	0.106	0.039	0.023	0.042	0.046	0.052	0.065	0.048	0.032	0.016	0.062	0.012

* <0.0001.

Table A. 2 PLS regression results. Summary statistics (R^2 and RMSE) for the monthly Ta prediction models, including six groups of variables.

	n	G1		G2		G3		G4		G5		G6		G7	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
January	712	0.87	1.96	0.87	1.95	0.34	4.34	0.82	2.26	0.39	4.17	0.86	2.01	0.85	2.10
February	712	0.83	2.19	0.83	2.19	0.32	4.40	0.79	2.45	0.37	4.26	0.85	2.05	0.84	2.19
March	712	0.80	1.89	0.74	1.94	0.36	3.40	0.77	2.04	0.39	3.32	0.81	1.84	0.81	1.83
April	712	0.79	1.51	0.79	1.53	0.48	2.39	0.77	1.58	0.32	2.73	0.75	1.67	0.74	1.69
May	712	0.76	1.48	0.79	1.41	0.74	1.54	0.80	1.37	0.31	2.52	0.76	1.50	0.77	1.47
June	712	0.78	1.44	0.80	1.38	0.75	1.52	0.79	1.41	0.26	2.63	0.78	1.45	0.76	1.51
July	712	0.83	1.33	0.86	1.20	0.79	1.48	0.84	1.28	0.28	2.71	0.82	1.34	0.80	1.44
August	712	0.84	1.36	0.87	1.23	0.81	1.44	0.85	1.28	0.30	2.80	0.83	1.37	0.86	1.25
September	712	0.81	1.35	0.84	1.24	0.82	1.30	0.83	1.26	0.32	2.53	0.80	1.37	0.79	1.42
October	712	0.83	1.27	0.83	1.26	0.68	1.73	0.82	1.31	0.41	2.36	0.81	1.32	0.77	1.47
November	712	0.83	1.54	0.83	1.57	0.37	2.97	0.79	1.70	0.40	2.92	0.82	1.59	0.81	1.65
December	712	0.86	1.68	0.86	1.67	0.36	3.53	0.82	1.89	0.34	3.58	0.85	1.74	0.83	1.82

Table A. 3 Model equations obtained from the PLS regression models using three variables: LSTd, LSTn, and elevation. The months are numbered from 01 to 12.

Regression Models	R^2	RMSE
$Ta_{01} = -2.137 + 0.347 \times LSTd + 0.497 \times LSTn + 0.00033 \times elevation$	0.87	1.95
$Ta_{02} = -3.037 + 0.297 \times LSTd + 0.493 \times LSTn + 0.00019 \times elevation$	0.83	2.19
$Ta_{03} = -1.986 + 0.252 \times LSTd + 0.477 \times LSTn - 0.001 \times elevation$	0.74	1.94
$Ta_{04} = 0.516 + 0.296 \times LSTd + 0.424 \times LSTn - 0.002 \times elevation$	0.79	1.53
$Ta_{05} = 3.863 + 0.272 \times LSTd + 0.383 \times LSTn - 0.002 \times elevation$	0.79	1.41
$Ta_{06} = 7.060 + 0.242 \times LSTd + 0.384 \times LSTn - 0.002 \times elevation$	0.80	1.38
$Ta_{07} = 8.440 + 0.241 \times LSTd + 0.398 \times LSTn - 0.002 \times elevation$	0.86	1.20
$Ta_{08} = 7.644 + 0.253 \times LSTd + 0.419 \times LSTn - 0.002 \times elevation$	0.87	1.23
$Ta_{09} = 5.294 + 0.291 \times LSTd + 0.407 \times LSTn - 0.002 \times elevation$	0.84	1.24
$Ta_{10} = 3.418 + 0.266 \times LSTd + 0.406 \times LSTn - 0.002 \times elevation$	0.83	1.26
$Ta_{11} = -0.912 + 0.271 \times LSTd + 0.411 \times LSTn - 0.001 \times elevation$	0.83	1.57
$Ta_{12} = -2.560 + 0.314 \times LSTd + 0.463 \times LSTn + 0.00037 \times elevation$	0.86	1.67

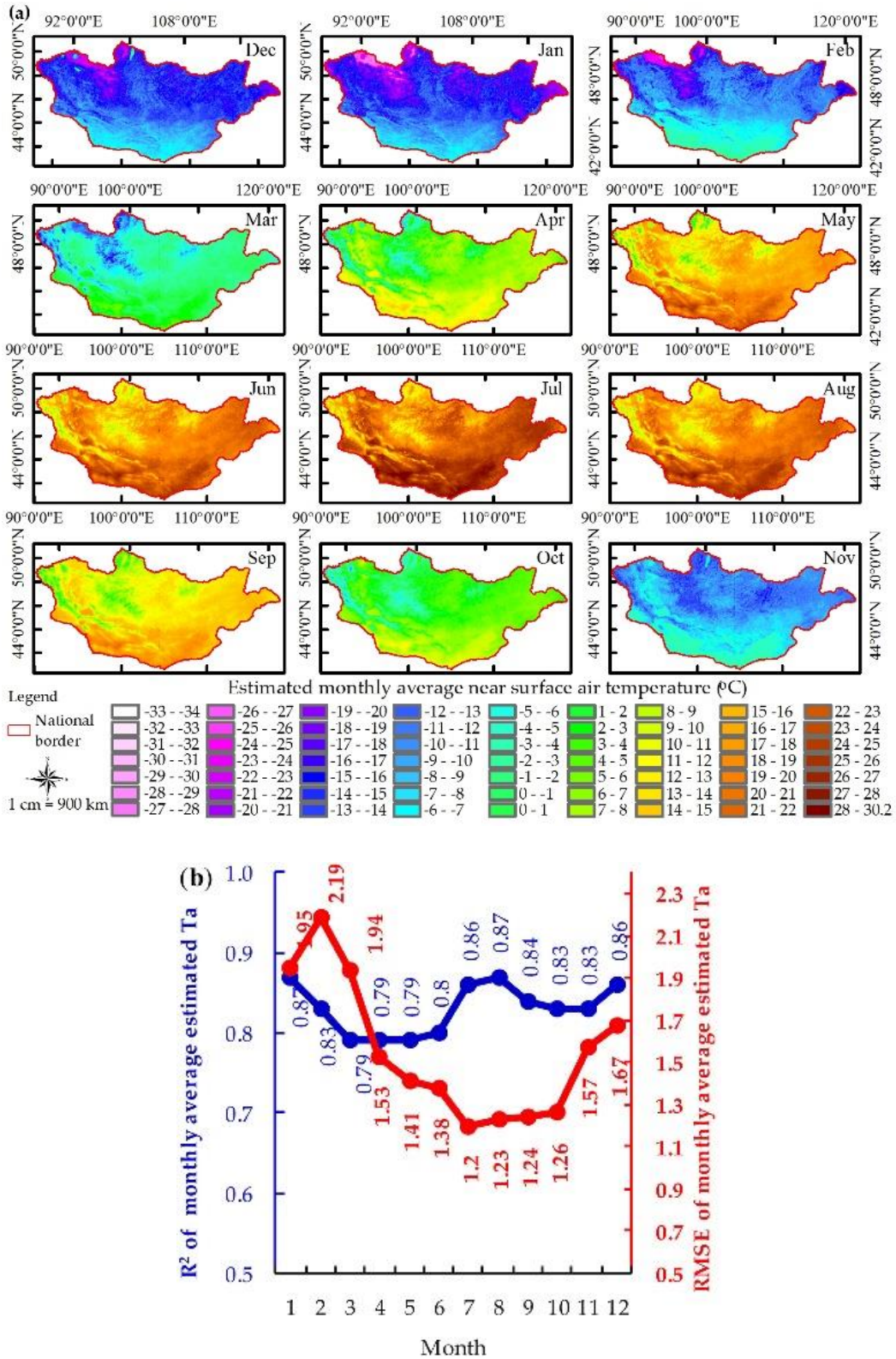


Figure A. 2 Estimated monthly average Ta based on the PLS regression model and using LSTd, LSTn, and elevation as predictor variables. (a) Spatial maps of estimated monthly average Ta over Mongolia at 1 km spatial resolution. (b) Monthly statistics of R^2 (blue) and RMSE (red) between observed and predicted air temperature.

Appendix B

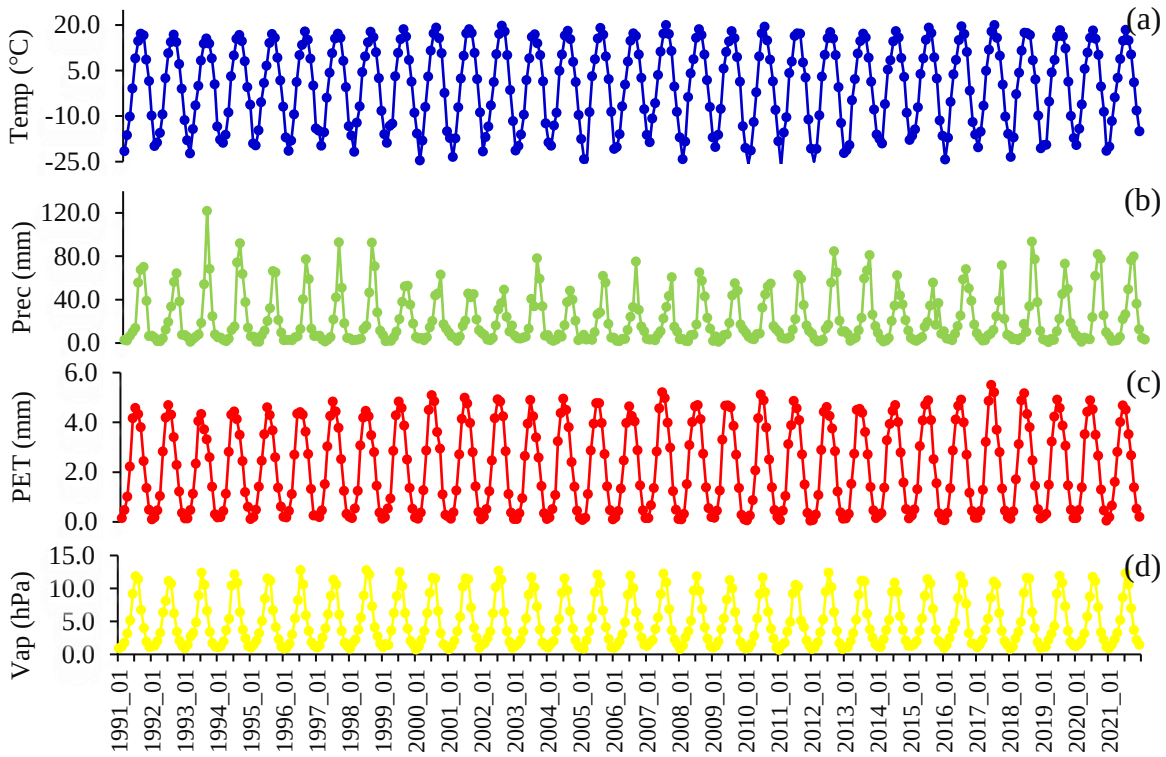


Figure B. 1 Variation of the (a) monthly mean air temperature, (b) monthly mean precipitation, (c) the monthly mean potential evapotranspiration, and (d) monthly mean vapor pressure derived from the CRU-TS v4.06 monthly climate dataset for the period 1991-2021 as an average of all weather stations (n=63) using a zonal statistics approach.

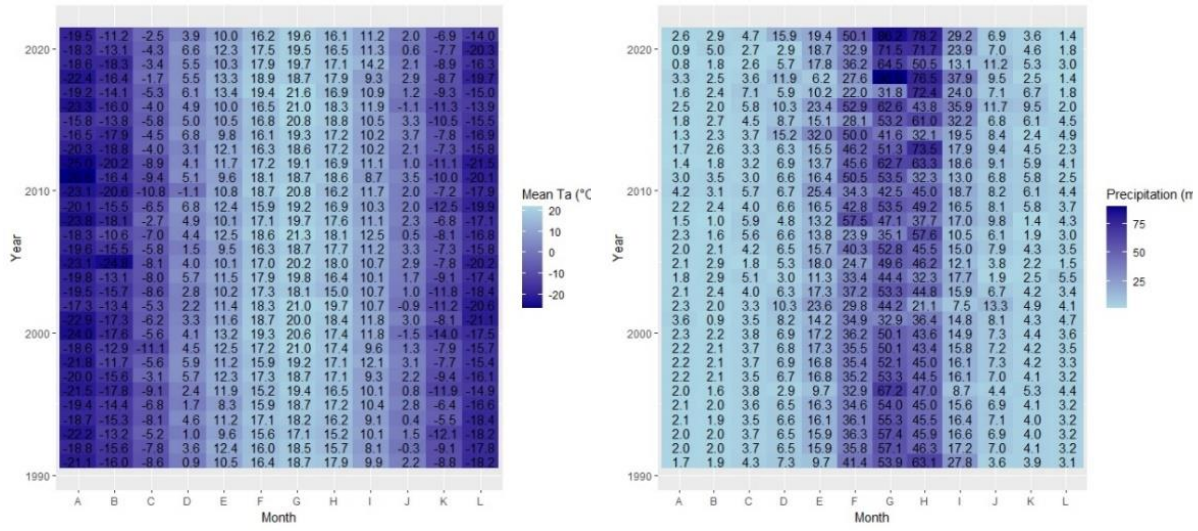


Figure B. 2 (left) Monthly mean temperature (Ta), and (right) monthly mean precipitation (Precip) for the period 1991-2021 as an average of all weather stations. Letters of the alphabet represent twelve months.

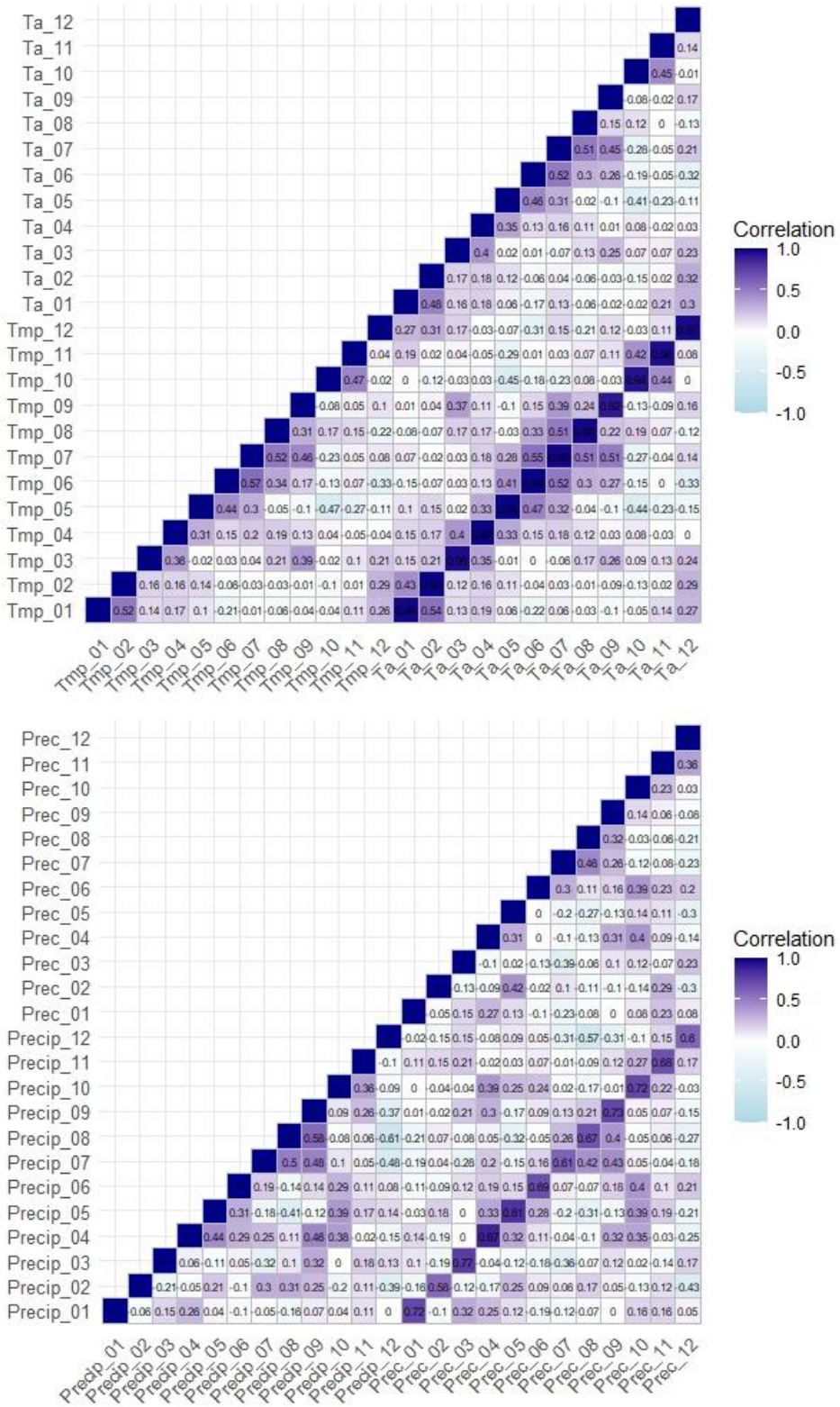


Figure B. 3 The correlation matrix between monthly mean Ta and Tmp (top), and monthly mean Precip and Prec (bottom). The monthly mean Ta and Precip data were obtained from weather stations for the period 1991-2021. Monthly mean Tmp and Prec data were extracted from CRU-TS for the same period. Extracted monthly mean Tmp and Prec data was an average of all weather stations. High correlations are highlighted in dark blue.

Table B. 1 Climatic indices formulas and their sources that were used in this study

Indices	Acronym	Formula	Reference
Hydro-Thermal Coefficient	HTC	$HTC = \frac{\sum P}{[0.1 * \sum T_{>X^oC}]}$	Selyaninov 1937
Aridity Index	I_{ar}	$I_{ar} = \frac{P}{T + 10}$	De Martonne 1925
Humidity Factor	HF_{th}	$HF_{th} = \frac{P_{I-XII}}{E_o}$	Thornthwaite 1948
Moisture Index	MI	$MI = \frac{P}{[0.2 * \sum T_{>10^oC} + 306]}$	Mezentsev 1969

$\sum$P- the sum of precipitation in a warm period, mm, $\sum T_{>X^oC}$- the sum of air temperature >10^0C, P- monthly total precipitation (mm), T- monthly mean air temperature (^{o}C), P_{I-XII}- annual total precipitation (mm), E_o- annual total potential evapotranspiration (mm),

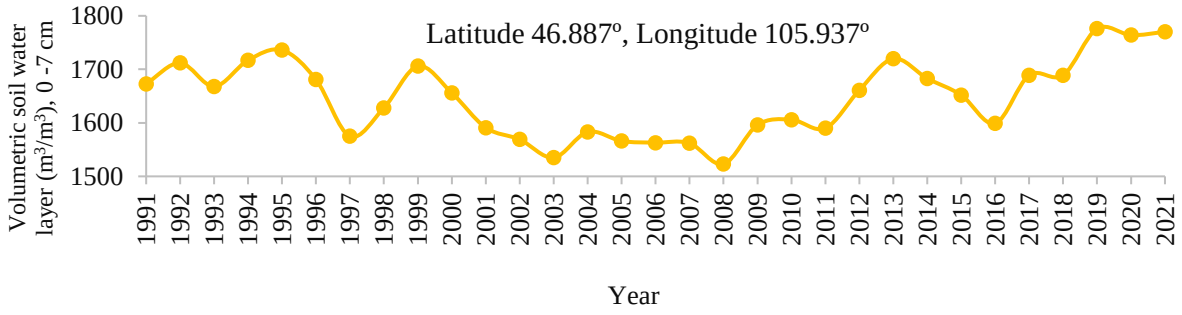


Figure B. 4 Variation of the annual mean soil moisture profile for the individual pixel was extracted from ERA datasets for the period 1991-2021.

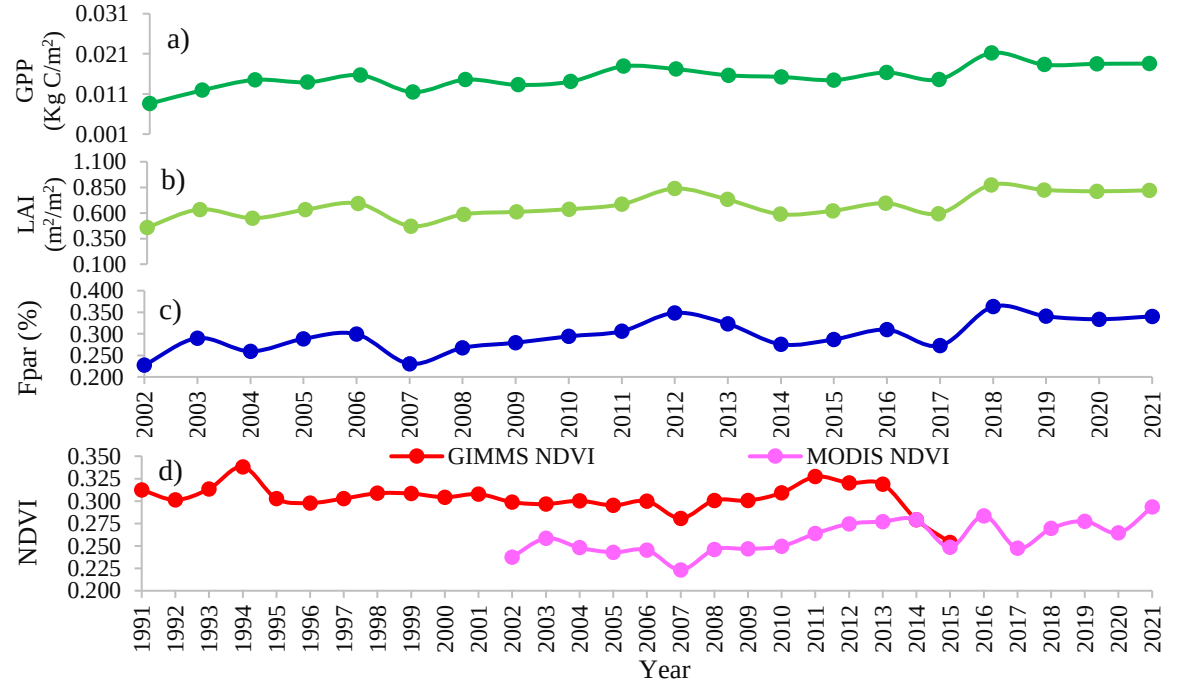


Figure B. 5 Variation of the mean of (a) GPP, (b) LAI, (c) Fpar, (d) NDVI for individual pixels from MODIS and GIMMS during the vegetation growing season (April-August) for the years 2002-2021, and 1991-2015, respectively. The sample pixels were selected from the objective variable (n=12988). Each variable of the curves was counted by averaging all objective variables.

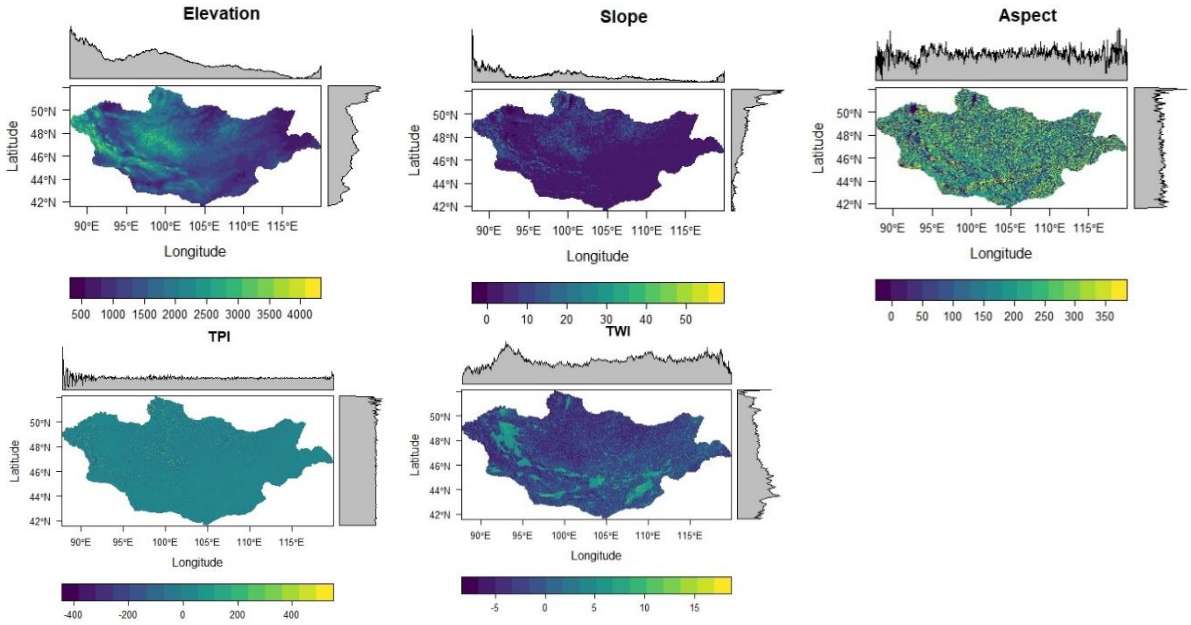


Figure B. 6 The spatial distribution maps of elevation (top left), slope (top center), aspect (top right), TPI (bottom left), and TWI (bottom center) derived from STRM DEM at a resolution of 90 m.

Table B. 2 The correlation matrix between topographic variables was derived from STRM DEM at a resolution of 90 m (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of the p-value (Pearson): "***" - <0.0001, "**" - <0.01, "." - <0.1.

Variables	Elevation	Slope	Aspect	TPI	TWI	Latitude	Longitude
Elevation	**1**	0.55***	0.00	0.12***	-0.38***	0.16***	-0.63***
Slope	0.55***	**1**	0.02***	0.16***	-0.45***	0.34***	-0.27***
Aspect	0.00	0.02***	**1**	0.01	-0.03***	-0.01**	0.02***
TPI	0.12***	0.16***	0.01	**1**	-0.19***	0.00	0.00
TWI	-0.38***	-0.45***	-0.03***	-0.19***	**1**	-0.21***	0.05***
Latitude	0.16***	0.34***	-0.01**	0.00	-0.21***	**1**	-0.12***
Longitude	-0.63***	-0.27***	0.02***	0.00	0.05***	-0.12***	**1**

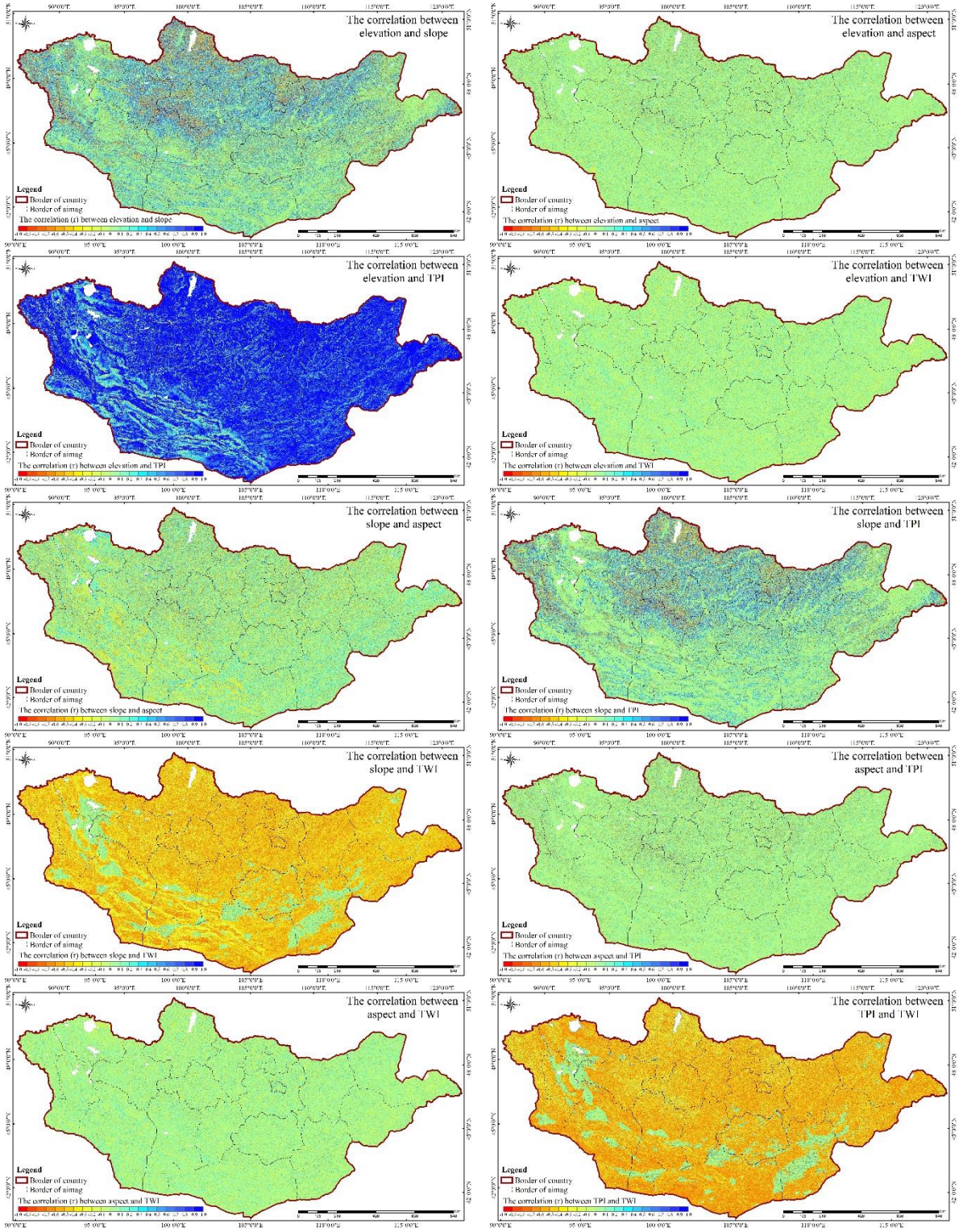


Figure B. 7 Spatial correlations between topographic variables.

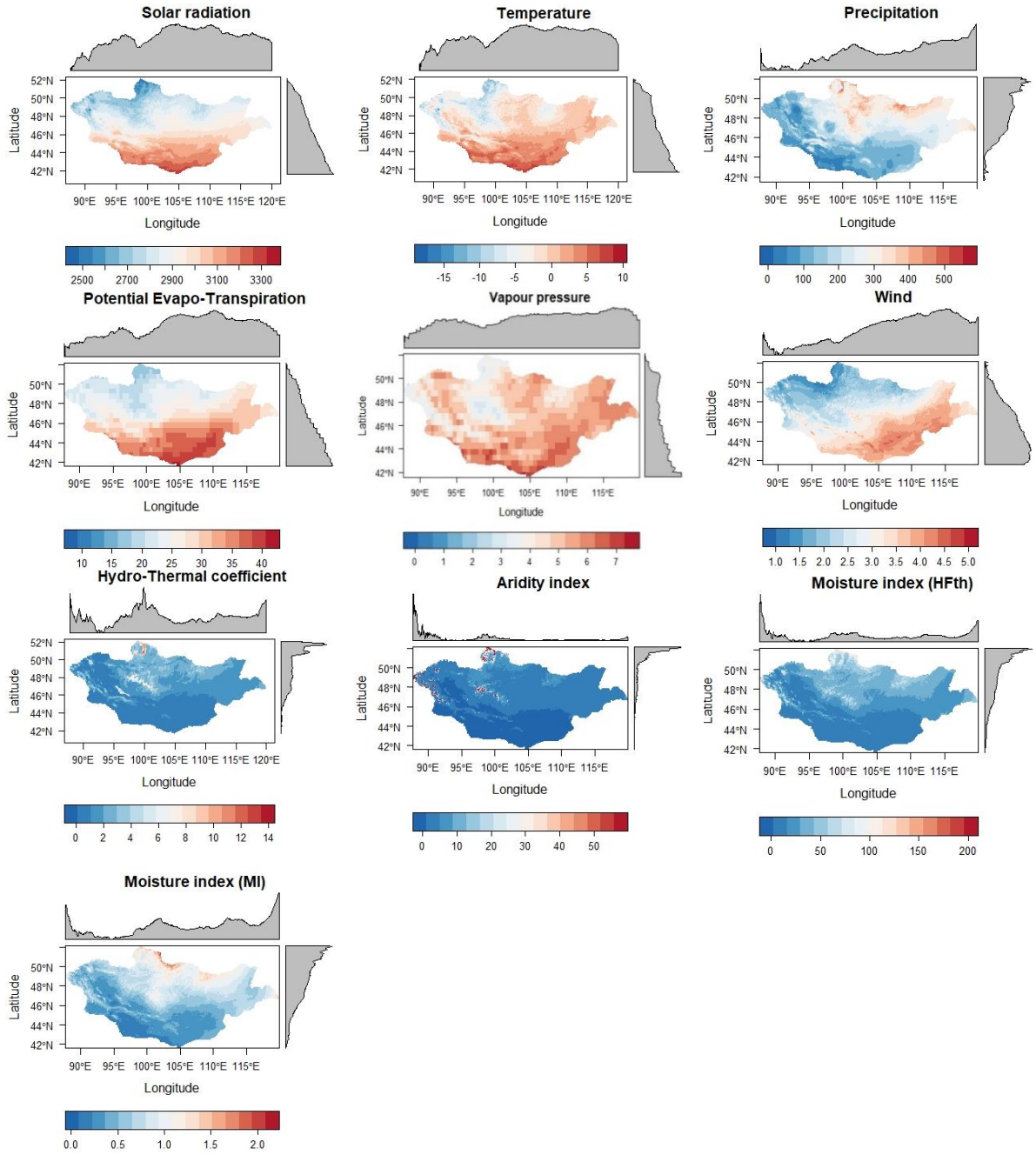


Figure B. 8 Spatial distribution maps of yearly mean Srad, Tmp, Prec, PET, Vap, and Wnd were extracted from CRU-TS and CRUNCEP monthly climate-gridded datasets for the period 1991-2021. The climate indices of HCT, Iar, HFth, and MI were calculated from yearly mean datasets of Tmp, Prec, and PET for the period 1991-2021 using empirical equations (see, Table B.1).

Table B. 3 The correlation matrix between yearly mean climatic variables was extracted from CRU-TS and CRUNCEP datasets for the period 1991-2021 (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of each p-value (Pearson) was noted as <0.0001.

Variables	Srad	Tmp	Prec	PET	Var	Wnd	HTC	Iar	HFth	MI
Srad	**1.00**	0.95	-0.58	0.96	0.73	0.80	-0.64	-0.46	-0.79	-0.70
Tmp	0.95	**1.00**	-0.46	0.92	0.84	0.75	-0.65	-0.52	-0.75	-0.58
Prec	-0.58	-0.46	**1.00**	-0.54	-0.32	-0.26	0.72	0.27	0.79	0.92
PET	0.96	0.92	-0.54	**1.00**	0.80	0.87	-0.61	-0.40	-0.73	-0.65
Var	0.73	0.84	-0.32	0.80	**1.00**	0.62	-0.59	-0.43	-0.58	-0.41
Wnd	0.80	0.75	-0.26	0.87	0.62	**1.00**	-0.37	-0.26	-0.49	-0.39
HTC	-0.64	-0.65	0.72	-0.61	-0.59	-0.37	**1.00**	0.54	0.80	0.72
Iar	-0.46	-0.52	0.27	-0.40	-0.43	-0.26	0.54	**1.00**	0.65	0.40
HFth	-0.79	-0.75	0.79	-0.73	-0.58	-0.49	0.80	0.65	**1.00**	0.92
MI	-0.70	-0.58	0.92	-0.65	-0.41	-0.39	0.72	0.40	0.92	**1.00**

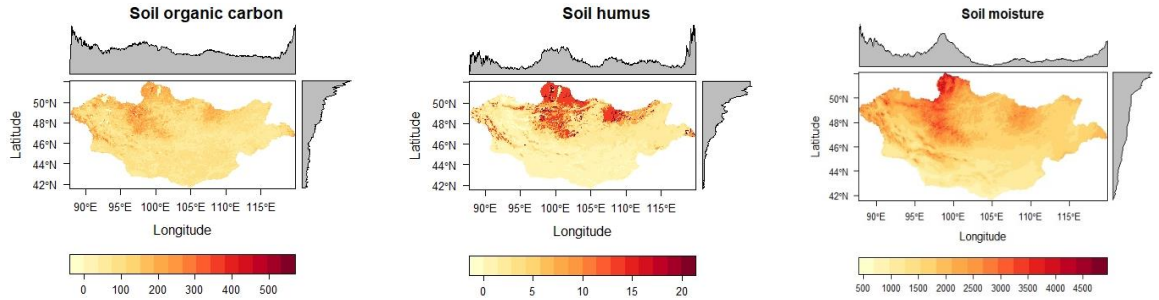


Figure B. 9 Spatial distribution maps of yearly mean SOC, sHumus, and sMoisture. (left) The SOC variable was extracted from the Trends.Earth system at a resolution of 250 m (Conservation International, 2022). (center) The vector data of soil humus were obtained from the IGG, MAS (right). The soil moisture data were extracted from the ECMWF database at a resolution of 0.28º for the period 1991-2021.

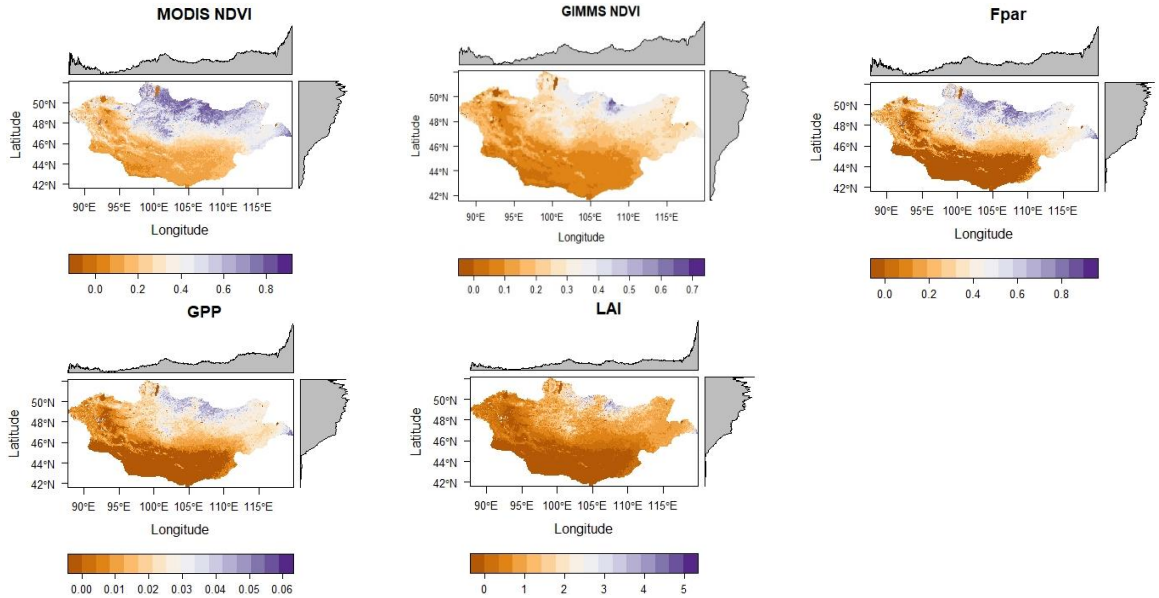


Figure B. 10 Spatial distribution maps mean modisNDVI, gimmsNDVI, Fpar, GPP, and LAI during the vegetation growing season (April to August). modisNDVI, Fpar, GPP, and LAI derived from the MODIS Aqua and Terra satellites at spatial resolutions from 250 to 500 m for the period 2002-2021 (LP DAAC, 2022). The mean gimmsNDVI derived from AVHRR/NOAA at a spatial resolution of 8 km for the period 1991-2015 during the vegetation growing season.

Table B. 4 The correlation matrix between soil variables (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of each p-value (Pearson) was noted as <0.0001.

Variables	SOC	sHumus	sMoisture
SOC	1	0.58	0.69
sHumus	0.58	1	0.62
sMoisture	0.69	0.62	1

Table B. 5 The correlation matrix between vegetation variables (n=19567425). Values are different from 0 with a significance level of α=0.05. The significance level of each p-value (Pearson) was noted as <0.0001.

Variables	modisNDVI	gimmsNDVI	Fpar	GPP	LAI
modisNDVI	**1**	0.95	0.96	0.95	0.93
gimmsNDVI	0.95	**1**	0.95	0.94	0.88
Fpar	0.96	0.95	**1**	0.97	0.93
GPP	0.95	0.94	0.97	**1**	0.95
LAI	0.93	0.88	0.93	0.95	**1**

Table B. 6 Summary statistics of the PLSR model (n=12988)

Observations (n)	12988
Coefficient of determination (R^2)	0.457
Adjected R^2	0.461
Standard (Std.) deviation	0.781
Mean Square Error (MSE)	0.608
Room Mean Square Error (RMSE)	0.780

Variable	t1*	u~1*	w1*	p1*	Coefficients of ß	VIP*	Std. deviation	Significance level	
								Lower (95%)	Upper (95%)
Intercept	-	-	-	-	0.501	-	-	-	-
Elv	-0.106	-0.227	-0.107	-0.034	0.000	0.536	0.011	0.511	0.562
Slp	0.267	0.130	0.061	0.085	0.004	0.306	0.013	0.277	0.334
Aspct	0.012	0.000	0.000	0.004	0.000	0.001	0.009	-0.020	0.022
TPI	-0.206	-0.129	-0.061	-0.066	0.000	0.305	0.020	0.261	0.350
TWI	-0.295	-0.164	-0.077	-0.094	-0.006	0.386	0.017	0.347	0.425
Lat	0.214	0.362	0.171	0.068	0.006	0.790	0.008	0.772	0.808
Long	0.712	0.335	0.158	0.227	0.019	0.854	0.009	0.834	0.875
Srad	-0.612	-0.201	-0.095	-0.195	0.000	0.473	0.009	0.452	0.494
Tmp	-0.411	-0.031	-0.014	-0.131	-0.001	0.072	0.012	0.046	0.098
Prec	0.886	0.663	0.312	0.282	0.001	1.562	0.007	1.546	1.578
PET	-0.591	-0.201	-0.095	-0.189	-0.004	0.473	0.008	0.456	0.490
Vap	-0.230	0.023	0.011	-0.073	0.003	0.055	0.013	0.026	0.085
Wnd	-0.401	-0.066	-0.031	-0.128	-0.009	0.157	0.007	0.140	0.174
HTC	0.745	0.463	0.218	0.238	0.058	1.091	0.008	1.073	1.109
Iar	0.657	0.330	0.156	0.209	0.021	0.778	0.015	0.743	0.812
HFth	0.944	0.590	0.278	0.301	0.006	1.390	0.005	1.380	1.401
MI	0.946	0.666	0.314	0.301	0.295	1.570	0.003	1.564	1.577
SOC	0.335	0.071	0.033	0.107	0.000	0.167	0.015	0.134	0.201
sHumus	0.680	0.439	0.207	0.217	0.011	1.035	0.017	0.996	1.074
sMoisture	0.520	0.140	0.066	0.166	0.000	0.331	0.010	0.307	0.354
modisNDVI	0.959	0.692	0.326	0.306	0.533	1.632	0.005	1.622	1.643
gimmsNDVI	0.940	0.666	0.314	0.300	0.421	1.572	0.003	1.565	1.578
Fpar	0.959	0.676	0.319	0.306	0.371	1.594	0.005	1.583	1.604
GPP	0.948	0.685	0.323	0.302	6.911	1.615	0.005	1.603	1.626
LAI	0.923	0.660	0.311	0.294	0.127	1.555	0.007	1.539	1.571

t1* and u~1*- Correlation matrix of the variables with the t and u~ components; w1* and p1*- vector; VIP*- Variable Importance in Projection

Table B. 7 Summary statistics of the PCR model (n=12988)

Observations (n)		12988
R^2		0.546
Adjected R^2		0.545
Std. error of the regression	MSE	0.510
	RMSE	0.714
	Mean Absolute Percentage Error (MAPE)	27.761
Akaike Information Criterion (AIC)		-8711.993
Schwarz Bayesian Criterion (SBC)		-8517.727

Analysis of variance (response variables)	DF	Sum of squares	Mean squares	F	Significance F (Pr > F)
Regression model	25	7963.607	318.544	624.239	**<0.0001**
Residual or error	12962	6614.404	0.510	-	-
Corrected total	12987	14578.011	-	-	-

	Principal Component Analysis			Model parameters for the components (response variables)					
Variable	Eigen value	Variance (%)	Cumulative variance (%)	Coefficients of ß	Standard error	t-statistic	Pr > |t|	Lower 95%	Upper 95%
Intercept	-	-	-	2.710	0.006	432.352	<0.0001	2.698	2.722
Elv	11.695	46.781	46.781	0.163	0.002	88.939	<0.0001	0.159	0.167
Slp	5.120	20.481	67.262	0.230	0.003	83.150	<0.0001	0.225	0.236
Aspct	1.633	6.533	73.795	-0.056	0.005	-11.502	<0.0001	-0.066	-0.047
TPI	1.090	4.361	78.157	-0.091	0.006	-15.215	<0.0001	-0.103	-0.080
TWI	1.001	4.003	82.160	-0.009	0.006	-1.441	0.150	-0.021	0.003
Lat	0.860	3.438	85.598	0.002	0.007	0.303	0.762	-0.011	0.015
Long	0.699	2.795	88.393	0.032	0.007	4.205	<0.0001	0.017	0.046
Srad	0.576	2.305	90.698	0.021	0.008	2.490	0.013	0.004	0.037
Tmp	0.490	1.961	92.659	0.072	0.009	8.000	<0.0001	0.054	0.089
Prec	0.421	1.682	94.341	-0.046	0.010	-4.716	<0.0001	-0.065	-0.027
PET	0.335	1.341	95.682	0.013	0.011	1.219	0.223	-0.008	0.034
Vap	0.312	1.248	96.930	-0.002	0.011	-0.144	0.886	-0.024	0.020
Wnd	0.180	0.722	97.652	-0.084	0.015	-5.678	<0.0001	-0.113	-0.055
HTC	0.151	0.605	98.257	-0.013	0.016	-0.794	0.427	-0.044	0.019
Iar	0.109	0.434	98.691	0.029	0.019	1.518	0.129	-0.008	0.066
HFth	0.077	0.309	99.000	0.063	0.023	2.800	0.005	0.019	0.107
MI	0.067	0.270	99.270	-0.101	0.024	-4.172	<0.0001	-0.148	-0.053
SOC	0.045	0.182	99.452	0.391	0.029	13.277	<0.0001	0.333	0.448
sHumus	0.043	0.173	99.624	0.051	0.030	1.678	0.093	-0.008	0.110
sMoisture	0.031	0.124	99.748	0.235	0.036	6.590	<0.0001	0.165	0.304
modisNDVI	0.022	0.086	99.834	-0.102	0.043	-2.379	0.017	-0.185	-0.018
gimmsNDVI	0.018	0.072	99.907	0.181	0.047	3.893	<0.0001	0.090	0.273
Fpar	0.014	0.056	99.962	0.000	0.053	-0.007	0.994	-0.105	0.104
GPP	0.009	0.034	99.997	-0.056	0.068	-0.821	0.412	-0.188	0.077
LAI	0.001	0.003	100.000	0.024	0.220	0.110	0.912	-0.407	0.456

Table B. 8 Summary statistics of the R^2 and RMSE for land potential prediction models, including 5 groups of variables using linear regressions (n=12988)

Model	G1		G2		G3		G4		G5	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
PLSR	**0.30**	0.88	**0.38**	0.83	**0.27**	0.90	**0.49**	0.76	**0.47**	0.77
PCR	**0.38**	0.83	**0.48**	0.76	**0.36**	0.85	**0.50**	0.75	**0.55**	0.71

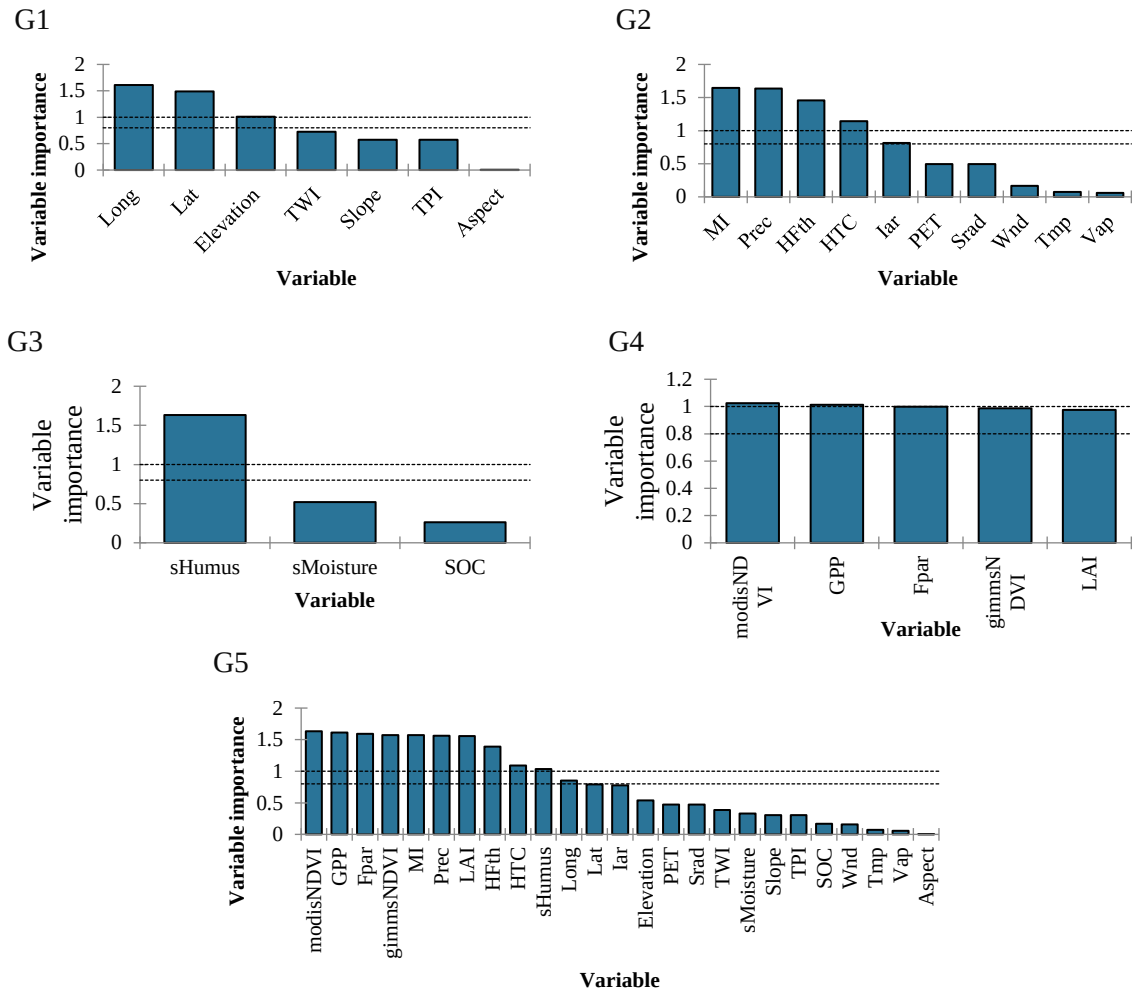


Figure B. 11 Estimated importance variables for the five groupings using the PLSR model (n=12988).

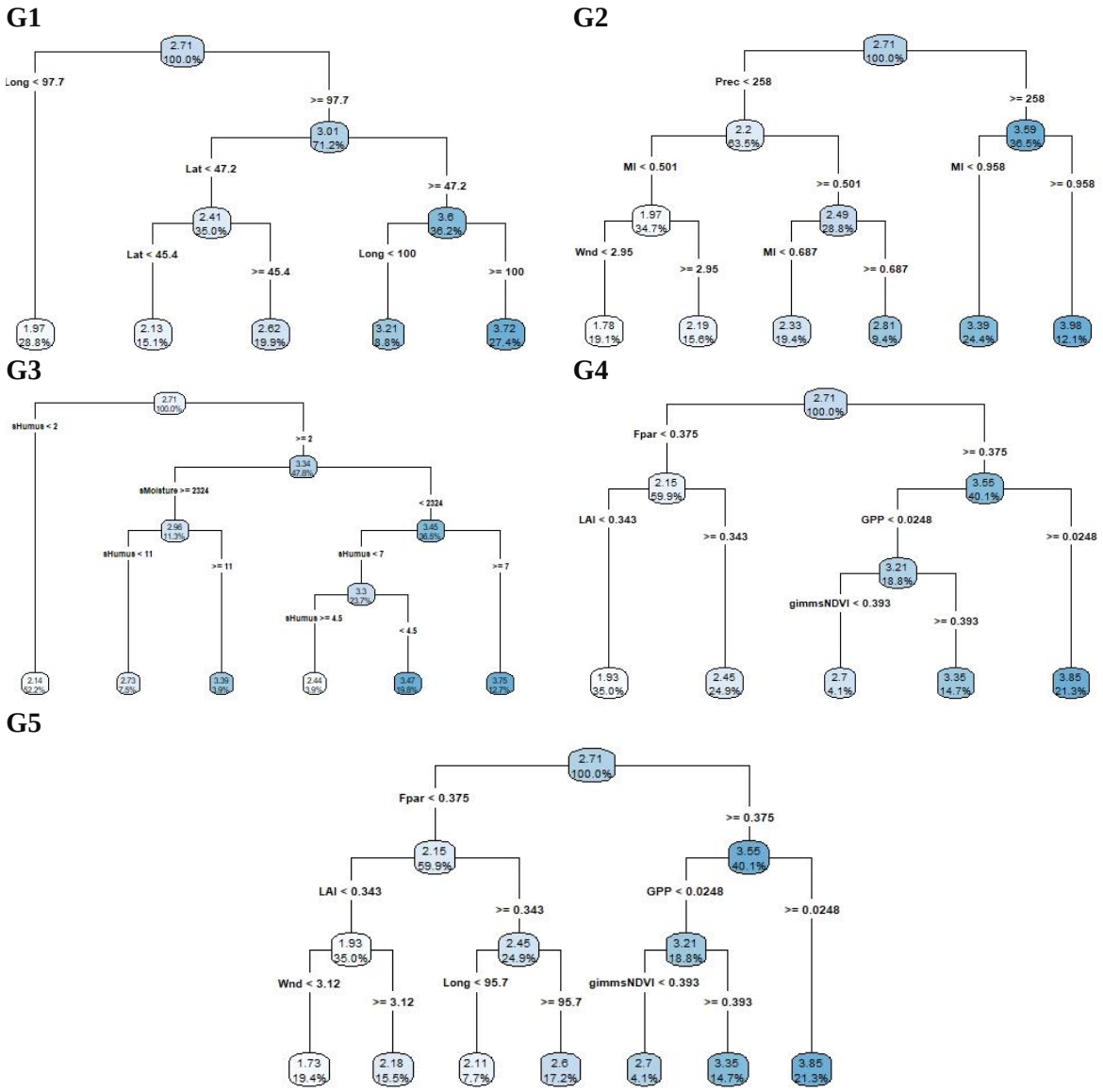


Figure B. 12 Regression tree of the CART was estimated from 23 explanatory variables with three steps of maximum tree depth (n=12988). The structure of each node is organized in a top-to-bottom direction. The right side of each level represents the important variables with a threshold value. In contrast, the left side of each level represents the less important variables with a threshold value.

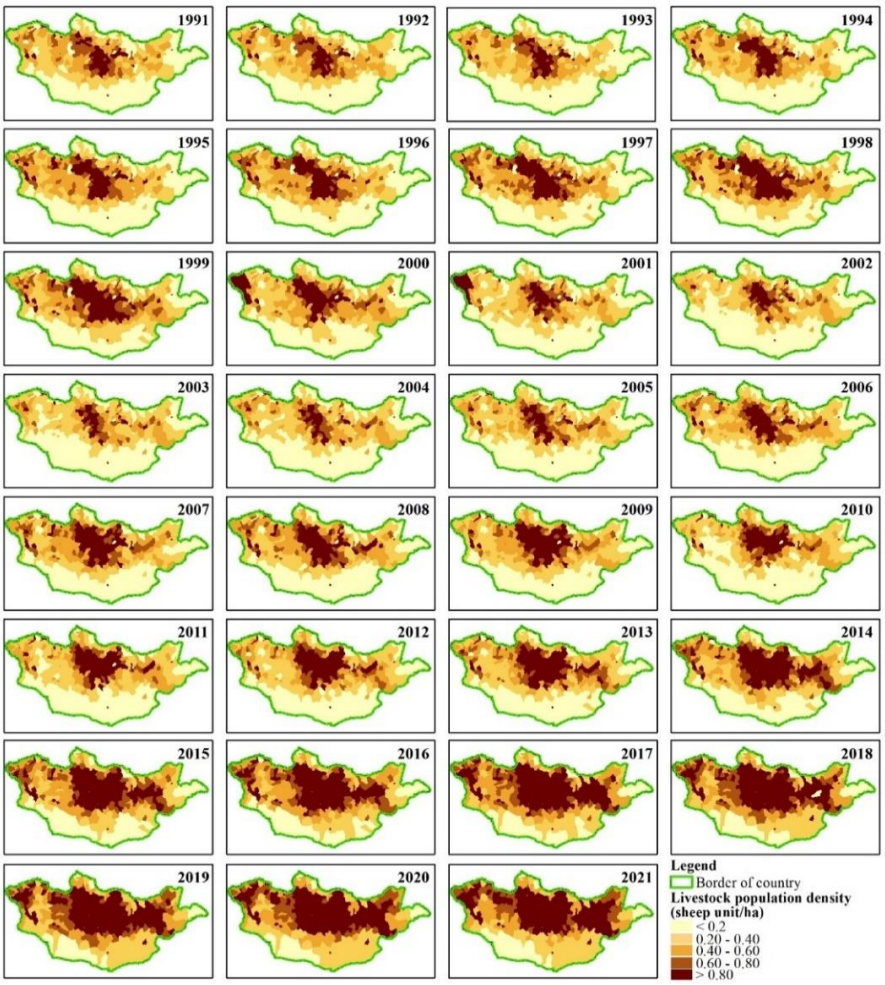


Figure B. 13 Spatial distribution maps of livestock population density (sheep unit ha^{-1}) in Mongolia for the period 1991-2021.

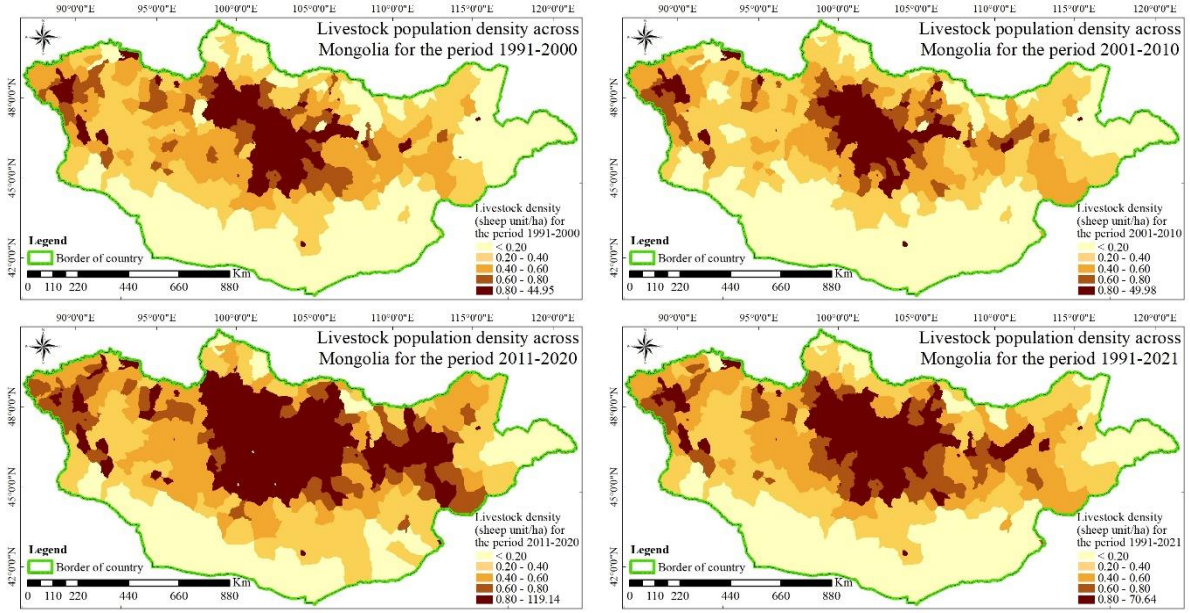


Figure B. 14 Livestock population density (sheep unit) in Mongolia for each decade from 1991, and the 31-year average.

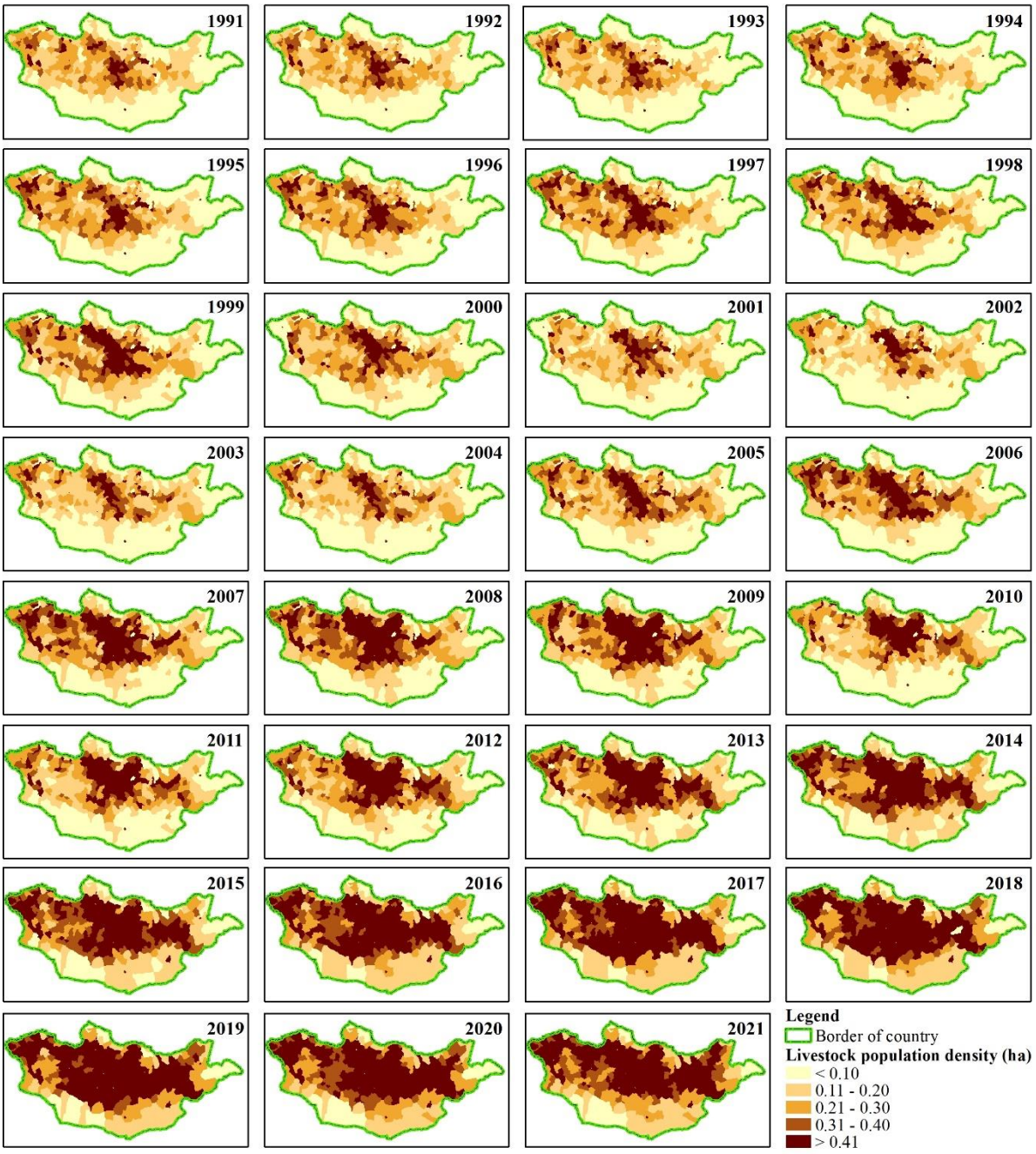


Figure B. 15 Spatial distribution maps of the total number of livestock population density (head ha^{-1}) in Mongolia for the period 1991-2021.

Appendix C

Detection of anthropogenic and environmental degradation in Mongolia using multi-sources remotely sensed time series data and machine learning techniques. (2022). *Appeared in Environmental Degradation in Asia: Land Degradation, Environmental Contamination, and Human Activities (pp. 17-47).* Cham: Springer International Publishing

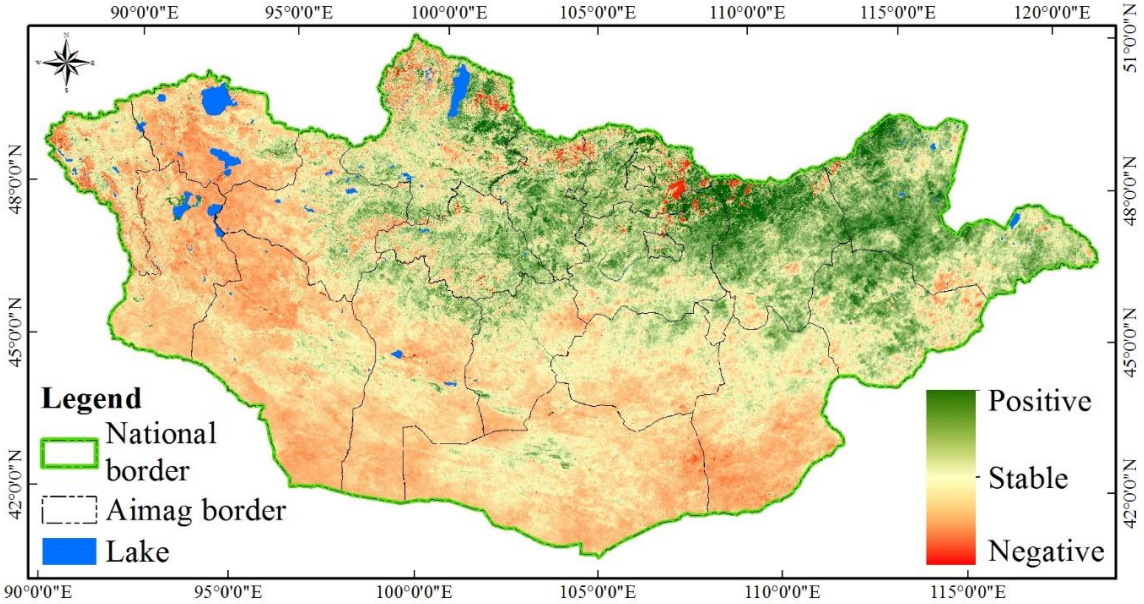


Figure C. 1 The trend of slope value of linear regression estimated from GIMMS and MODIS NDVI for the period 1990–2019.

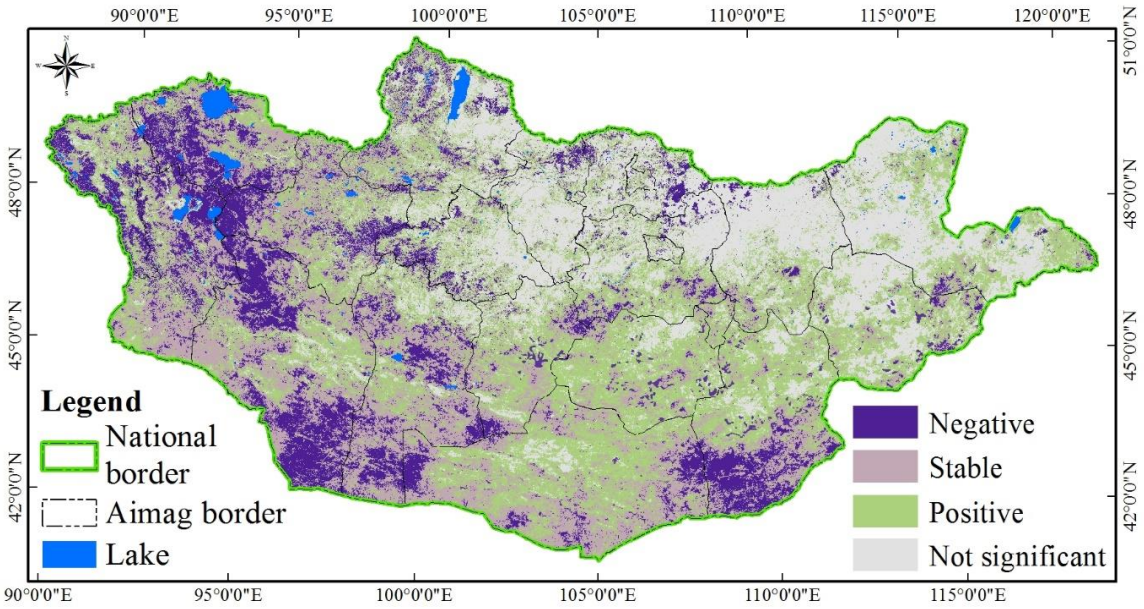


Figure C. 2 The trend of slope value of significant linear regression estimated from GIMMS and MODIS NDVI for the period 1990–2019.

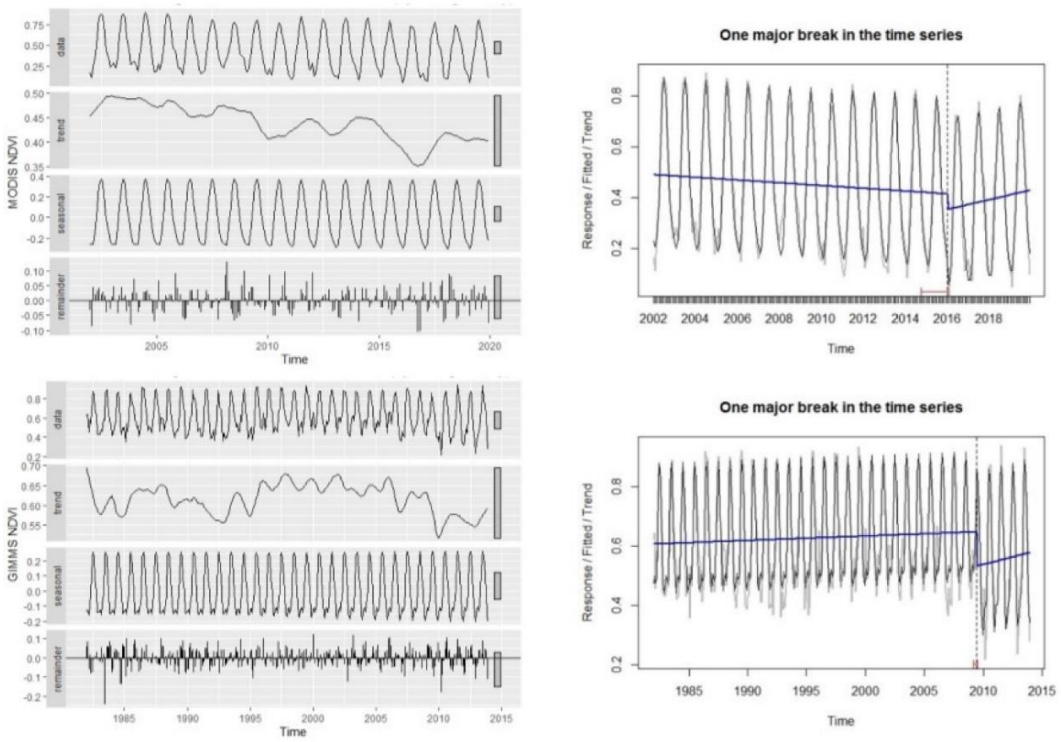
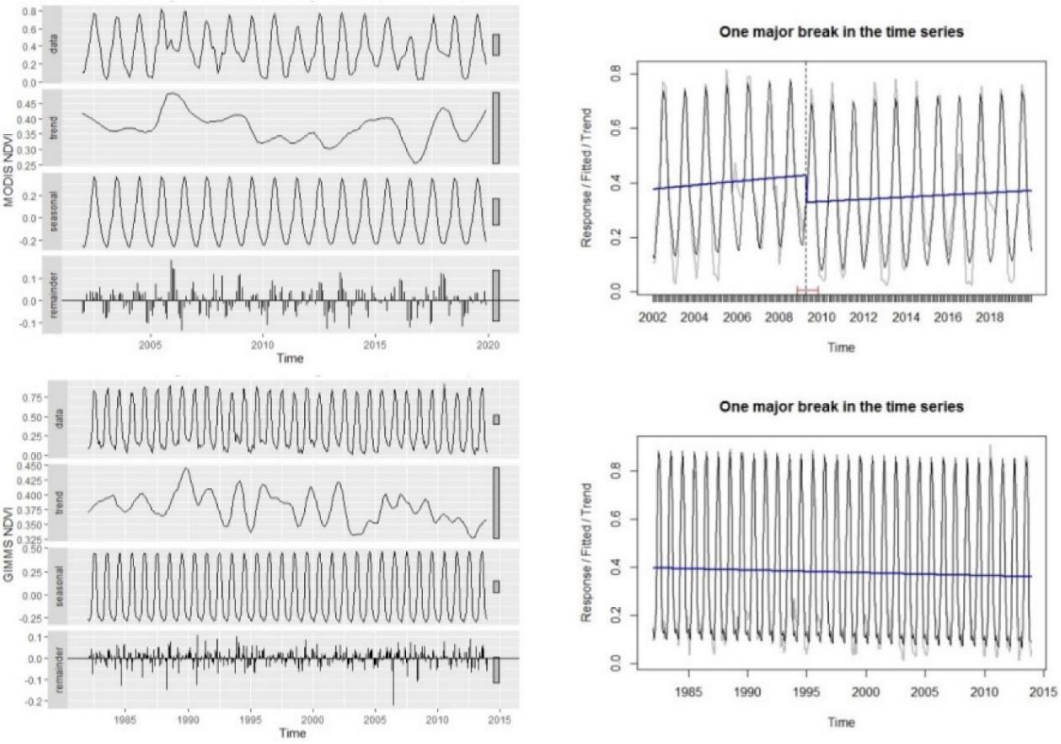
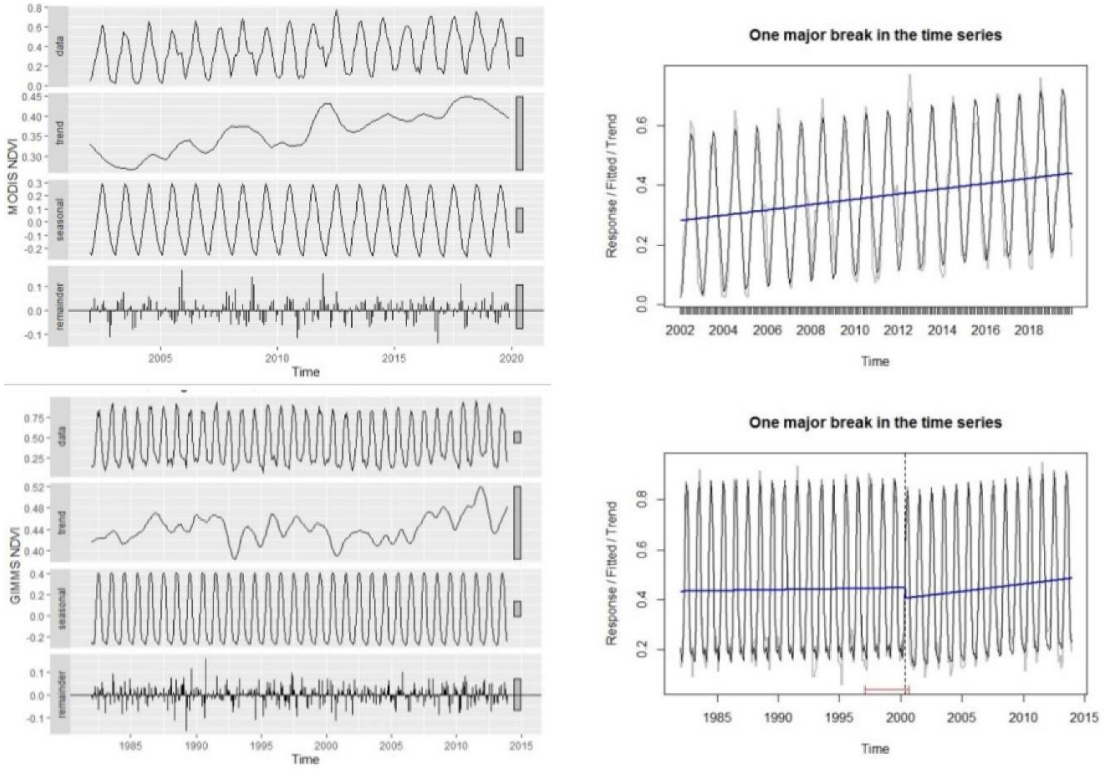


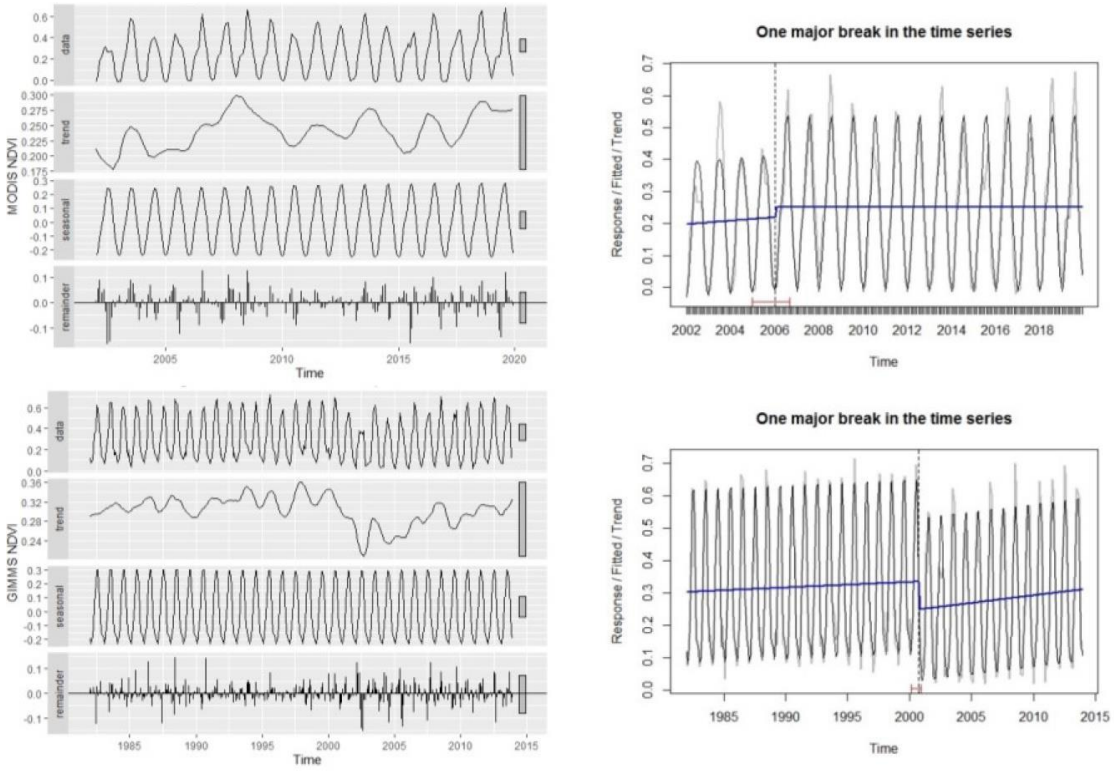
Figure C. 3 (a-f) Here, fitted components of trend, seasonal, and the remainder (evaluated noise) for the time series monthly NDVI of MODIS (left above) and GIMMS (left below). The abrupt change was observed in the trend component of the NDVI time series (right above and below). The grey, black, red, and blue lines represented a primary NDVI curve, fitted NDVI curve, direction and magnitude of abrupt change, and trend before and after the change, respectively. (a) Forest to barren land by mining activity (Latitude: 49.640, Longitude: 107.721).



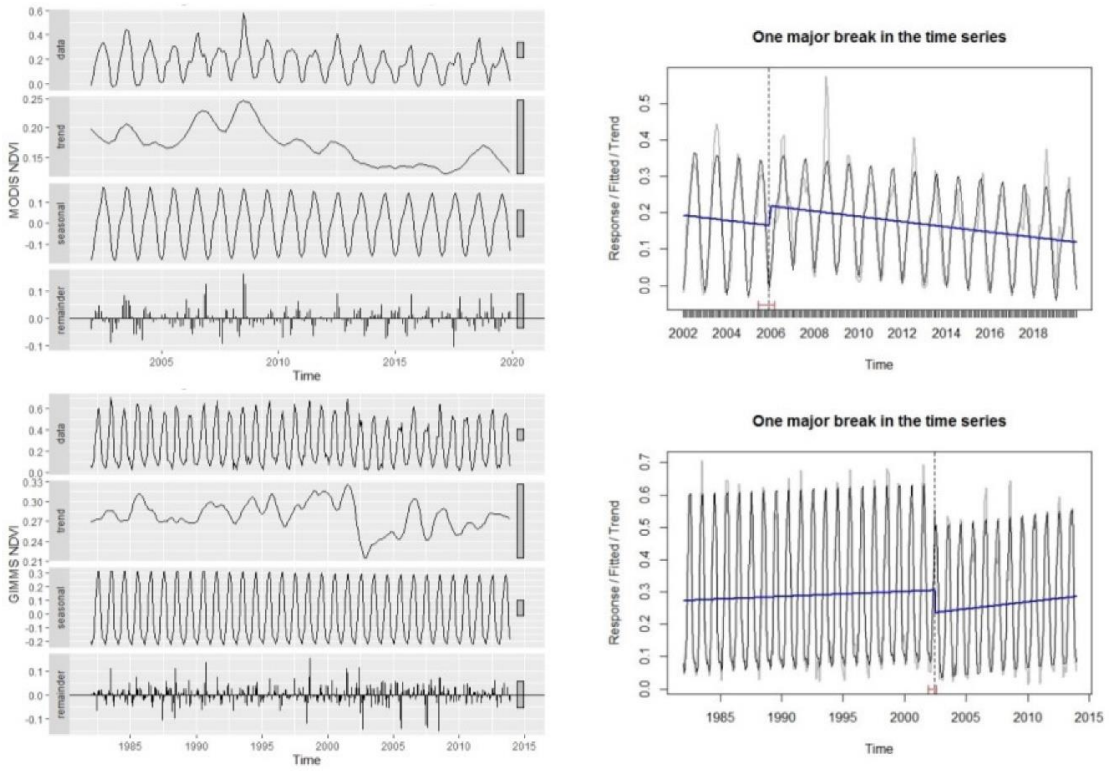
(b) Forest to grassland by fire (Latitude: 49.124, Longitude: 112.061). Explanations are found in Figure C.3 (a).



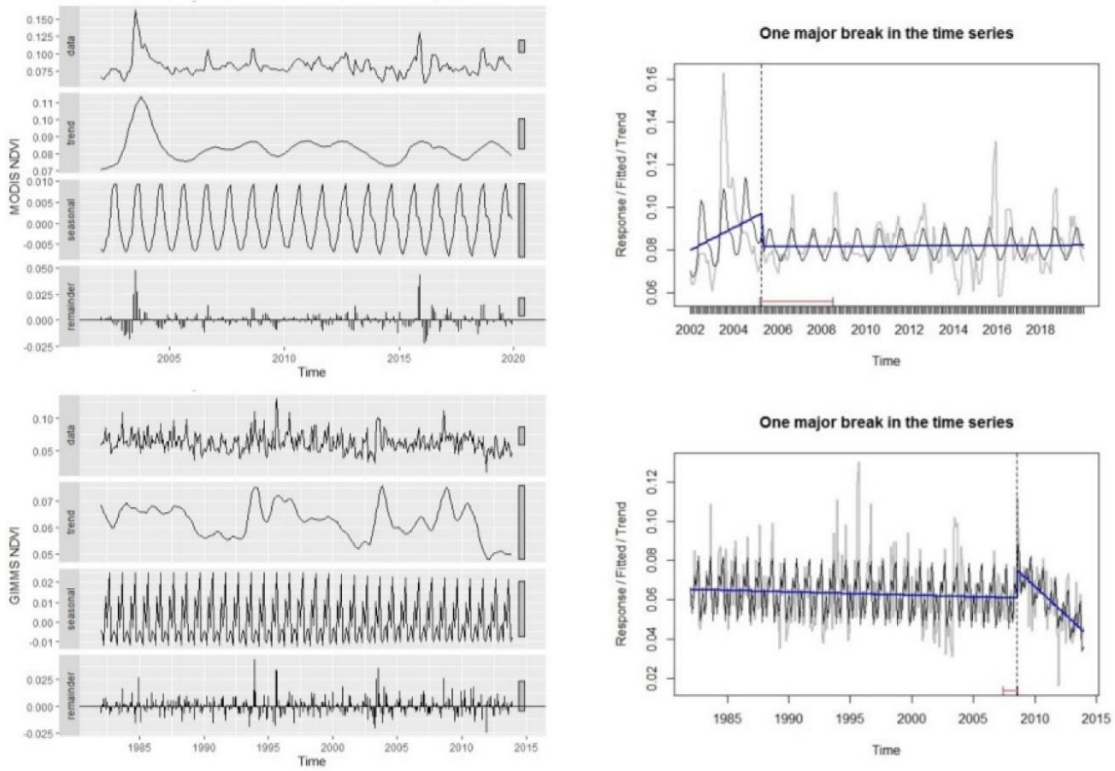
(c) Grassland (vegetation coverage >50%) to forest (Latitude: 50.167, Longitude: 106.498). Explanations are found in Figure C.3 (a).



(d) Grassland (vegetation coverage >50%) to cropland (Latitude: 49.415, Longitude: 105.628). Explanations are found in Figure C.3 (a).



(e) Grassland (vegetation coverage 25-50%) to barren land (Latitude: 48.423, Longitude: 104.529). Explanations are found in Figure C.3 (a).



(f) Grassland (vegetation coverage <25%) to urban land (Latitude: 43.040, Longitude: 106.84). Explanations are found in Figure C.3 (a).

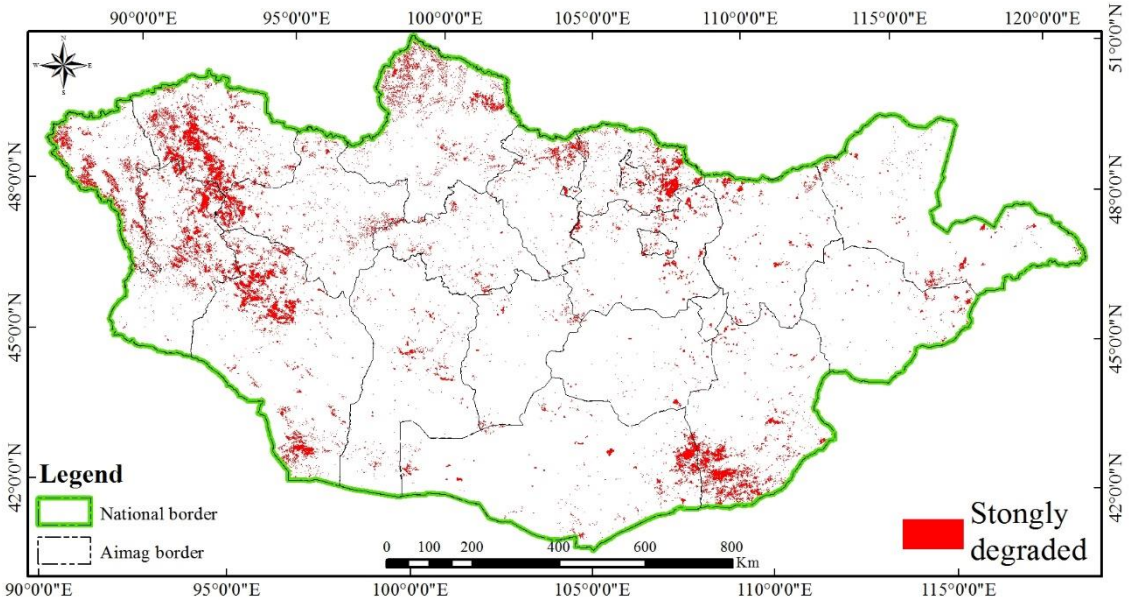


Figure C. 4 Estimated strongly degraded area from MODIS and GIMMS NDVI time series for the period 1990–2019 using trend, BFAST, and RESTREND analysis methods.

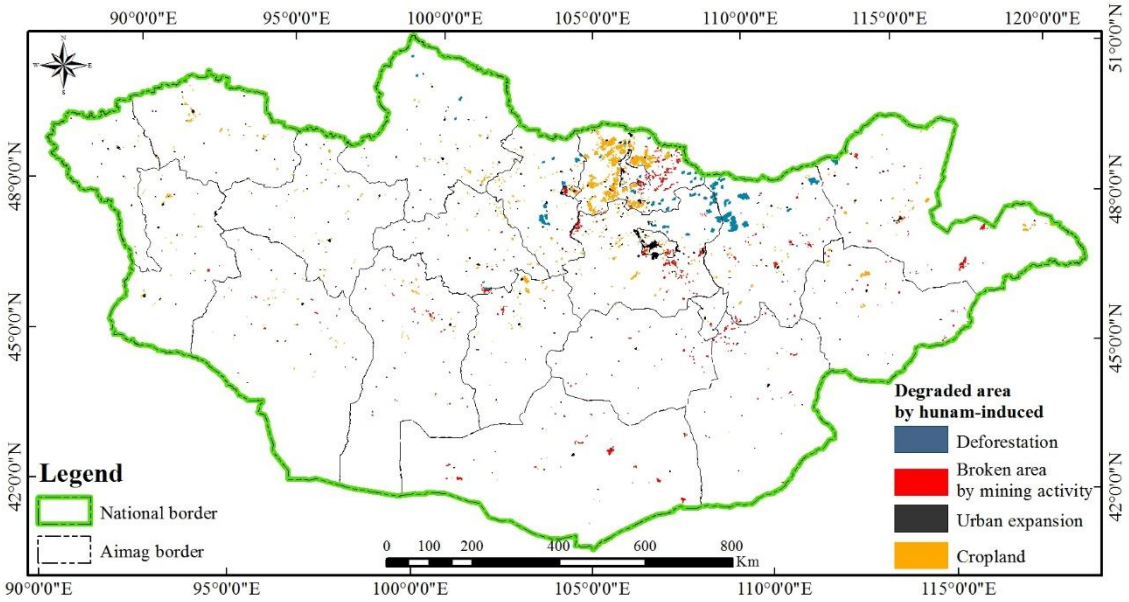


Figure C. 5 Generated strongly degraded area by human activity from Landsat OLI imagery, Google Earth map, and environmental geo-database of Mongolia.