UNDERSTANDING THE RELATIONSHIP BETWEEN GEOGRAPHICAL FACTORS AND DIRT ROAD DENSITY AT THE LOCAL LEVEL: A CASE STUDY OF ORKHON, KHONGOR, DARKHAN, SHARYN GOL AND JAVKHLANT SOUMS

Unurnyam Jugnee
Institute of Geography and Geoecology, MAS
Ulaanbaatar, Mongolia
unurnyamj@mas.ac.mn

Supervisor
Dr Benjamin David Hennig
University of Iceland
ben@hi.is

ABSTRACT

The Central Asian country of Mongolia is severely affected by desertification and land degradation. One of the main drivers of degradation is dirt road expansion, often informal and unplanned nature. A better understanding of the geographical factors that exacerbate dirt road expansion is helpful to develop management strategies to reduce dirt road-related land degradation. The aim of this study was to investigate the relationship between dirt roads and geographical factors at the local level using a case study area in the central northern province of Selenge. The study utilised geostatistical techniques to explore a total of 17 variables representing key natural and human characteristics. A total length of 2,998 km of dirt roads was determined in the case study area through satellite image analysis. Several dirt road clusters identified during the analysis indicated that dirt roads contribute significantly to land degradation in the study area. Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models showed some weaknesses in their capabilities. Multicollinearity was identified as one of the major challenges in such a local level analysis. The OLS model could explain up to 7% of the dirt road expansion in the study area, whereas the GWR model could explain 13%. The modelling results suggest that the effect of geographical factors on dirt road expansion varied throughout the study area. Rural settlement centres (soum and bag centre) and terrain slopes could influence an increase in dirt road density at the local level. Overall spatial analysis could identify some overall patterns in the data that may be valuable for identifying

the most pressing geographic areas and their specific characteristics that require addressing through specific planning and management approaches.

Key words: Geographically Weighted Regression, off-road driving, dirt road density, Geographic Information Systems, Mongolia

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
ALAMGC	Agency for Land Administration and Management, Geodesy and Cartography
DEM	Digital Elevation Model
GWR	Geographically Weighted Regression
LST	Land surface temperature
NDVI	Normalized difference vegetation index
OLS	Ordinary Least Squares regression
QGIS	Quantum Geographic Information System
SRTM	Shuttle Radar Topographic Mission
TGSI	Topsoil grain size index
TVDI	Temperature – Vegetation Dryness Index
USGS	United States Geological Survey

1. INTRODUCTION

Inappropriate management and overutilization of land resources are commonly considered the origin of land degradation. Vehicle off-roading is one of the direct impacts which lead to degradation (UNCCD 2017). Landscape alteration caused by off-road vehicles is being noticed not only in developed countries but also in developing countries (Keshkamat et al. 2013; Ploughe & Fraser 2022). Dirt roads are a result of the pressure of the wheels of vehicles which can cause soil erosion (Padgett et al. 2008; Al-Dousari et al. 2019; Zhang et al. 2019; Cao et al. 2021), sediment yield (Bravo-Linares et al. 2018; Ramos-Scharrón 2018; Nosrati & Collins 2019), changes in vegetation cover (Crisfield et al. 2012; Assaeed et al. 2019; Hogan & Brown 2021), and also social issues (Jackson 2015). Jackson (2015) showed that dust originating from unpaved mining roads in the South Gobi province in Mongolia endangers the health of local citizens and their animals.

Mongolia is a Central Asian country with a total land territory of 1,564,115.7 km². It is severely affected by desertification and land degradation (Han et al. 2021). According to the Desertification Atlas of Mongolia, 76.9% of the total land area has been affected by desertification (Baasandai 2020). Partly due to uncertain legal conditions of land utilization, improper land use has resulted in adverse effects on land resources (Bazarragchaa 2017). The main human activities which have led to land degradation are overgrazing, mining activity, dirt roads, crop cultivation, deforestation, and settlement area expansion (Ministry of Nature and Environment 1997; Doljin 2010; Ochirbat 2013).

Soil erosion caused by uncontrolled dirt roads is one of the main reasons for land degradation in Mongolia (Li et al. 2006; Ochirbat 2013). Due to the lack of paved roads and poor economic capacity to build paved roads, dirt roads play an essential role in transportation. They are broadly used as connections between settlement centres, mining sites, arable lands, winter camps, and other land uses in the rural areas (Byambaa & Murayama 2012). The total road network length in Mongolia is 49,200 km (UNECE 2018), of which 10,151 km are paved (National Statistics Office of Mongolia n.d.). This is, however, a general estimation due to the unmanaged nature of many dirt roads. The figure may increase if it also comprises the length of rural dirt roads (for example dirt roads that connect herders' winter camps, summer camps, wells, and springs) and the length of dirt tracks running parallel to main roads (for example main roads which connect province centres and soum centres) (Byambaa & Murayama 2012; Ochirbat 2013). The rapid expansion of dirt roads in the vast land area of Mongolia makes it difficult to detect and register temporal dirt roads in the official database (Dashpurev et al. 2020). Dashpurev et al. (2020) stated, for example, that temporal dirt roads caused by oil extraction in Eastern Mongolia have never been accounted for in official statistics due to the characteristics of those roads and they are often abandoned quickly after they are created.

Keshkamat et al. (2013) identified 37 main national dirt road corridors with a total length of 11,000 km running parallel to the main arterial road. Dashpurev et al. (2020) revealed that land area degraded by dirt roads and oil extraction infrastructure in the Menen Steppe and the Khalkh River area in Eastern Mongolia increased from 7,840 ha in 2005 to 14,730 ha in 2018. Ragchaadulam et al. (2020) estimated that the total length of the dirt roads caused by the Tavan Tolgoi coal mining in Tsogttsetsii soum in the Umnugobi province in Mongolia increased from 476.2 km in 2000 to 1,716 km in 2015.

The results of the different studies mentioned above that had been carried out in different parts of Mongolia (southern part: Ragchaadulam et al. 2020; eastern part: Dashpurev et al. 2020; and

national level: Keshkamat et al. 2013) and covered areas of different sizes, from 72.5 km² (Ragchaadulam et al. 2020) to 20,000 km² (Dashpurev et al. 2020)) suggest that dirt roads make a considerable contribution to land degradation in Mongolian rural areas. Moreover, these studies could support the assertion of Byambaa and Murayama (2012) that, if the official statistics would take rural dirt roads into account, the total length of the registered Mongolian road network might increase. Therefore, there is a need to conduct a comparative study on dirt road distribution and the contribution of dirt roads to land degradation in different geographical areas with various types of land utilization.

The studies by Dashpurev et al. (2020) and Ragchaadulam et al. (2020) mentioned above covered areas with intensive mining industry. Mining, including illegal gold mining, and tourism are the main origins for the formation of dirt roads (Yadambaatar & Sandag 2010), while the increase in the number of vehicles also contributes to the creation of new dirt roads (Damdinsuren et al. 2008; Ochirbat 2013). It is still unclear to what extent the land area is degraded by dirt roads. Comprehensive research on the impact of dirt roads on land degradation has not yet been conducted (Dashpurev et al. 2020) and there is little information available about the total amount of land area degraded by transportation. Zamba et al. (2006) estimated that nearly 0.7 million ha of land are affected by vehicle-induced degradation, whereas Ochirbat (2013) noted that approximately 1.5 million ha of land area are degraded by road erosion in Mongolia. According to the UNECE (2018), as a result of the dirt track corridor running parallel to national main routes, more than 3 million ha of land are degraded. The remaining rural dirt roads with little use intensity also have some noticeable contributions to land degradation (UNECE 2018).

Due to the significant impact of dirt roads on land degradation, there is a need to accurately determine the spatial distribution and size of land area that is affected by dirt roads and discuss their impact on the surrounding environment. Although some studies have focused on the ecological effect of dirt roads in Mongolia, only a few studies have made thorough investigations of the causal relationship between geographical factors and the width of dirt road corridors at the national level. Keshkamat et al. (2013) carried out an in-depth study on the causal relationship between geographical factors and dirt road corridors using Geographically Weighted Regression (GWR) and Ordinary Least Squares (OLS) Regression. Their results showed that correlation between corridor width and geographic variables varies throughout the country and that the corridor widths are not accidental but a result of drivers' reaction to geographical factors. It is still unclear, however, how geographical factors influence dirt road density at the local level.

Modelling and analysing the relationship between driving factors and environmental processes is essential to science-based decision-making for environmental protection. Global and local models can be used for modelling and analysing the relationship between driving factors and environmental processes (Du et al. 2014). Local spatial regression techniques were initially developed by Cleveland and Devlin in 1988 and the GWR is a subclass of the local spatial regression techniques (Cho et al. 2009). The GWR model may help to identify the emergence or existence of dirt roads in the study area. Conversely, in traditional global regression models such as OLR, the coefficients of the explanatory variables of the process are assumed to be constant throughout the whole study area, so the parameters estimated by such a regression model cannot give information on how the process being expressed by the parameter changes over space (Fotheringham & Sachdeva 2022).

Potential to investigate the spatially non-stationarity relationships between response variables and explanatory variables is the main advantage of GWR (Yang et al. 2019). Meanwhile, some shortcomings in relation to local multicollinearity (Wheeler & Tiefelsdorf 2005) and misspecification (Fotheringham & Sachdeva 2022) are commonly observed when applying GWR techniques. GWR techniques are widely used in spatial analysis (Chen et al. 2012) and they have wide-ranging applications in many research fields. For instance, they have been applied to detect the non-stationary relationships between dirt road corridors and geographic factors (Keshkamat et al. 2013), vegetation restoration and natural, policy, and socio-economic factors (Zhang et al. 2020), urban expansion and linear infrastructural, housing, and market forcing factors (Mondal et al. 2015), land consumption and geomorphologic, socio-demographic, economic structural, and institutional quality factors (Punzo et al. 2022), road fatality rates and economic, infrastructural, and social developmental factors (Wachnicka et al. 2021), and the relationship between infant mortality rates and economic and health care factors (Wang & Wu 2020).

It is essential to identify land areas that are potentially affected by dirt roads and develop management measures to mitigate and prevent dirt road-related land degradation in Mongolia. In order to provide accurate information for land management planning, we need to understand the factors that influence dirt road density at the local level and the spatial distribution of those factors throughout the local territory. The focus of this project is therefore on the relationship between geographical factors and the expansion of dirt road density that can lead to soil erosion and changes in vegetation cover.

The overall aim of this study was to investigate the relationship between geographical factors and dirt road density using a case study area in Mongolia to better understand these connections at a local level.

This will be investigated through the following specific objectives:

1. To identify key geographical factors that lead to expansion of dirt road density at the local level in Northern Mongolia
2. To explore the suitability of spatial statistics techniques, which can be used to determine the correlation between geographical variables
3. To estimate the relationship between geographical variables using a Geographically Weighted Regression model in order to determine spatial variation of the driving factors on dirt road density

The research questions were:

– What geographical factors can influence the increase of dirt road density at the local level?
– Are there any key variables that explain the highest dirt road density at the local level?

The project findings can be used as reference information for local decision-makers and land managers to make efficient decisions on land use planning and mitigate the negative influences of dirt roads. It can also provide the basis for monitoring the effectiveness of environmental protection policy implementation.

2. METHODS

2.1 Study area

The study area for this study covers the Darkhan, Khongor, Sharyngol, and Orkhon soums of Darkhan Uul province and Javkhlant soum of Selenge province in Mongolia. The study area has been chosen for several reasons. Firstly, this area has a combination of topographic features including medium and low mountains, plains, intermountain hollows, and river valleys. These various topographic features can help understanding of the relationship between natural factors and dirt road expansion. Secondly, this area plays a significant role in Mongolia's economy. A diverse range of land utilization such as open pit coal mining, gold mining, grazing, cropland, and urban (Darkhan city, the second largest city of Mongolia) and rural settlement centres (soum and bag centre) can be found in the area. These various land utilization types can help understanding of the relationship between land use factors and dirt road expansion. Thirdly, the study area belongs to the Orkhon river basin which has been defined as one of the land degradation hotspot areas in Mongolia (Nyamtseren et al. 2019). Therefore, the expected research results can help to address the contributions of dirt roads to land degradation in this hotspot area.

The study area is located in the northern part of Mongolia and covers a total land area of 4,465 km² which is 0.29 percent of the total territory in Mongolia (Fig. 1). The elevation in this area ranges from 700 to 2,600 m above sea level. The area is situated on the north-western side of the Khentii mountain range and belongs to the Selenge river basin area. The land surface is uneven, and the geomorphological features are characterised by a combination of medium and low mountains, hills, steppes, intermountain hollows, and river valleys. The absolute height of the mountains reach 1,000-2,000 m, rising 600-800 m above the surrounding landscape (Dechingungaa et al. 2013).

The area is subject to a harsh continental climate with cold winters and short summers. The temperature amplitude is high, and the growing season lasts from May to August. According to the weather observations in the period of 1991-2020, the monthly mean temperature in January, the coldest month of the year, is -23°C to -25.6°C. In July, the warmest month of the year, the mean monthly temperature is 19.6°C to 20.9°C. The mean annual precipitation is approximately 330 mm, and a large amount of the total precipitation falls in summer as rain (IRIMHE n.d.).

In terms of hydrology, the study area belongs to the Arctic Ocean Basin and is divided into separate catchment areas including the Sharyn Gol River catchment, the Yeroo River catchment, and the Kharaa River catchment. All these rivers originated in the Khentii mountain ranges and flow north and north-westward into the Orkhon River. Chestnut soil and kastanozem soils are commonly distributed in the study area. Common soil types in the mountain, hill, and intermountain hollow are kastanozem soil and chestnut soil. Alluvial meadow and meadow soils are distributed in the river valleys. The main soil textures are loamy, loamy sand, and sandy clay (Geobotanic 2015).

Mongolia has four levels of administrative units. Mongolia is divided into the administrative units of aimag (aimag is equivalent to province) and the capital city. The capital city and aimags therefore constitute the second-level administrative units. Aimags (province) are further divided into soums (soum is equivalent to district). Soum is the third level administrative unit. Soums are further divided into bags (bag is equivalent to county) as the fourth level of administrative

units. The study area focuses on the third and fourth level of the administrative units of Mongolia.
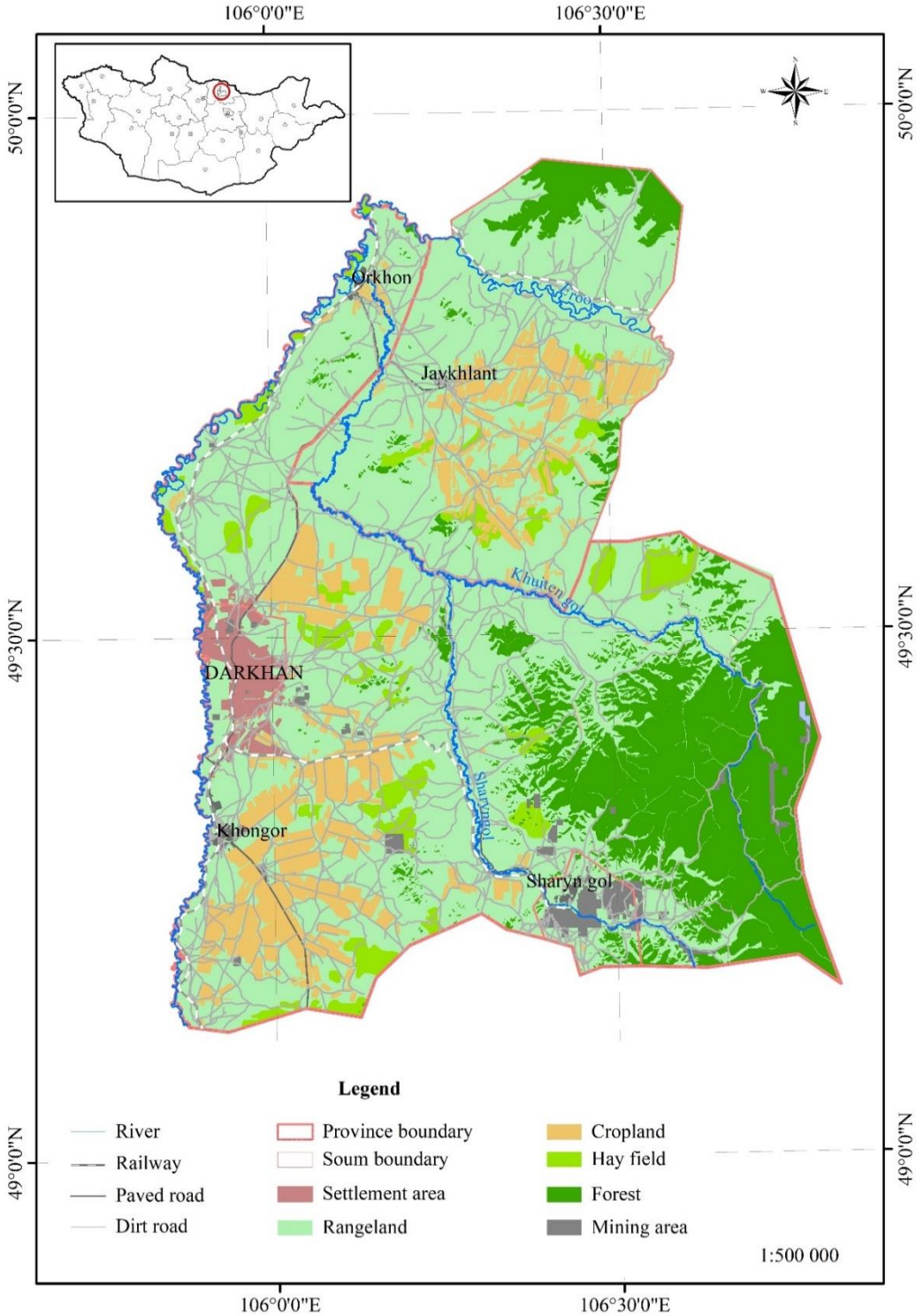

Figure 1. Location and land use of the Orkhon, Sharyn Gol, Darkhan, Khongor and Javkhlant soums in Mongolia. (Source: adapted from ALAMGC 2021).

As of 2019, a total of 104,141 people lived in the study area, which was 3.25% of the total population of Mongolia. In other words, 3.25% of the total population lives in 0.29% of the total land territory of Mongolia. 81.9% of the total population in the study area lives in Darkhan city, 1.9% in Javkhlant soum, 3.1% in Orkhon soum, 5.6%t in Khongor soum, and 7.5% in Sharyngol soum (National Statistics Office of Mongolia n.d.). The total livestock number in the study area is 442,763, of which 24,314 are horses, 58,260 cattle, 205,535 sheep, 154,594 goats, and 60 camels (National Statistics Office of Mongolia n.d.).

2.2   Data

Two kinds of spatial datasets (vector and raster) were used in this study. Generally, the explanatory variables were identified based on the study of Keshkamat et al. (2013). Keshkamat et al. (2013) used 15 variables in total, including distance to province centre, distance to county centre, distance to main rivers, distance to secondary rivers, distance to main river crossings, distance to secondary river crossings, distance to lakes and marshes, south aspect, terrain slope, total traffic density, ratio of light-to-heavy traffic, soil grain size index, land surface temperature, soil moisture index, and vegetation index in their study. Due to lack of data availability at the local level, traffic related variables (total traffic density and ratio of light-to-heavy traffic) were not considered in this study. Since the research was conducted at the local level, it was assumed that the influence of the proximity to river crossing on dirt roads could be represented by the influence of the proximity to the main river and tributary river. Therefore, river crossing variables are not accounted for in this study. In order to reflect local geographic factors, the following additional variables were considered in this study: cropland, mining, paved road, herders' winter camps, and railroad. After confirming data availability, 17 variables were selected for the study (Table 1).

Land use datasets including settlement area, cropland, mining, paved road, unpaved road, railroad, and herder's winter camp location were provided by the National Geodatabase of the "Mongolian national unified land territory" of the Agency for Land Administration and Management, Geodesy and Cartography of Mongolia (ALAMGC 2021).

Variables representing soil and vegetation factors were generated from Sentinel 2 Level 1C data and Landsat 8 OLI Level 1 data. Keshkamat et al. (2013, p. 437) stated, "being the end of spring and onset of summer, it is the most optimal (balanced) period to derive variables such as MSAVI (Modified Soil Adjusted Vegetation Index), SMI (Soil Moisture Index), and GSI (Grain Size Index)". Therefore, cloud-free 4 tiles (T48UWA, T48UXA, T48UWV, and T48UXV) Sentinel 2 level 1C data acquired on 30 August 2021 and cloud-free 2 tiles (path 132, rows 025 and 026) Landsat 8 OLI level 1 data acquired on 21 August 2015 were downloaded.

Sentinel 2 satellites launched by the European Space Agency provide remotely sensed data with high temporal (5 days), spatial (10 m, 20 m, and 60 m), and spectral (13 bands) resolution to users (ESA 2015). Spectral bands (with 10 m spatial resolution) including band 2 (blue at 0.490 µm), band 3 (green at 0.560 µm), band 4 (red at 0.665 µm), and band 8 (near infrared at 0.842 µm) were used in this study.

Table 1. Geographic factors, variable layers and data sources used in the study.

Factor	Variable	Source
Proximity to settlements	Distance to province centre	Spatial datasets are acquired from the Geodatabase of the "Mongolian national unified land territory" of the ALAMGC (2021)
	Distance to soum centre	
	Distance to bag centre and recreation site	
Proximity to land use types	Distance to cropland	
	Distance to mining	
	Distance to paved road	
	Distance to railroad	
Proximity to a water body	Distance to main river	Vector data digitized from the Topography map of the NAGC (1984)
	Distance to tributary river	
	Distance to lake	
Soil	Soil moisture index (SMI)	Raster datasets are generated from the Landsat OLI 8 level 1 data (acquisition date: 21 August 2015, path 132/ row 025-026, spatial resolution 30m) from the USGS (2020)
	Land surface temperature (LST)	
	Topsoil grain size index (TGSI)	Raster datasets are generated from the Sentinel 2 level 1C data (acquisition date: 30 August 2021, tile: T48UWA, T48UWV, T48UXA, T48UXA) from the ESA (2021)
Vegetation	Vegetation index (NDVI)	
Terrain condition	Terrain slope	SRTM Digital Elevation Model (DEM) obtained from USGS (2018)
	Aspect	
Dirt road	Dirt road density	Vector file is digitized from the TGSI map generated from the Sentinel 2 level 1C data (acquisition date: 30 August 2021, tile: T48UWA, T48UWV, T48UXA, T48UXA) from the ESA (2021)

The Landsat 8 satellite was launched by the National Aeronautics and Space Administration (NASA) and United States Geological Survey (USGS) in 2013. It has two instruments, the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), that collect data with a spatial resolution of 15 m (panchromatic band), 30 m (visible, near infrared, and short-wave infrared, coastal/aerosol, cirrus bands), and 100 m (thermal infrared bands) (USGS 2019). Four spectral bands: the red band (0.636-0.673 µm), NIR band (0.851-0.879 µm), TIR-1 band (10.60-11.19 µm), and TIR-2 (11.50 -12.51 µm) of the Landsat 8 data were used in this study.

Terrain slope and aspect layers produced from USGS's Shuttle Radar Topographic Mission (1 Arc-Second Global, spatial resolution 30 m) DEM data using Spatial Analyst Toolbox in ArcGIS 10.8.1 software.

2.3 Geographic analysis

A spatial statistical analysis was used in this study. The GWR and OLS models were employed to investigate the relationship between dirt road density and geographical factors. The software used in this study were Microsoft Excel 365 (Microsoft Corporation n.d.) and R Studio (RStudio Team 2020) for data pre-processing and statistics, QGIS 3.26.0 (QGIS Development Team 2022) for data pre-processing, Arc GIS 10.8 (ESRI INC 2020), and ArcGIS Pro 2.9 (ESRI INC 2021) for spatial analysis and mapping.

2.3.1 *Data preparation*

All datasets were transformed into the same coordinate system (WGS84 geocentric coordinate system, UTM projection, central meridian 105°E) after they had been collected from the relevant sources. Vector data were converted into Euclidean distance raster data (spatial resolution is 30 m) and extracted by the study area using the Extraction and the Rasterize algorithm in QGIS 3.26.0 (QGIS Development Team 2022).

The dirt road density dataset was considered as a dependent variable in this study. Dirt roads were digitized from the TGSI map by visual interpretation. TGSI maps were generated from the Sentinel 2 Level 1C data obtained from ESA (2021). In order to accurately digitize the dirt road, both dirt road vector data layers which originated from the ALAMGC (2021) and Google Earth Pro were used as primary reference data. The dirt road shape file originating from Google Earth Pro was created by digitizing satellite images from Landsat / Copernicus. The dates of the mosaic satellite images were 5 June 2004, 9 August 2006, 22 July 2015, 23 June 2017, 23 April 2018, 9 March 2019, 10 April 2019, 4 May 2019, 2 July 2019, and 14 July 2019.

After the creation of the dirt road vector data, dirt road density was calculated by dirt road length per 1 km² land area. In order to obtain dirt road density values, grid cells with an area of 1 km² were generated all over the study area. The total length of dirt roads in each grid cell was determined using the Sum line lengths algorithm in QGIS 3.26.0 (QGIS Development Team 2022).

The remote sensing spectral indices Normalized Difference Vegetation Index (NDVI), Topsoil Grain Size Index (TGIS), Soil Moisture Index (SMI), and Land Surface Temperature (LST) were used as input parameters for spatial statistical analysis. Sentinel 2 level 1C products with radiometric and geometric correction were provided in Top-Of-Atmosphere reflectance (ESA 2015). All downloaded Sentinel 2 Level 1C images were atmospherically corrected using the Semi-Automatic Classification Plugin (SCP) of the QGIS 3.26.0 software.

NDVI was calculated by:

$$NDVI = (NIR - RED) / (NIR + RED) \tag{1}$$

The topsoil Grain Size Index was calculated as defined by Xiao et al. (2007):

$$TGSI = (RED - BLUE) / (RED + BLUE + GREEN) \tag{2}$$

where BLUE is the blue band with central wavelength 490 nm, bandwidth 65 nm (band 2 of the Sentinel product), GREEN is the green band with central wavelength 560 nm, bandwidth 35 nm (band 3 of the Sentinel product), RED is the red band with central wavelength 665 nm, bandwidth 30 nm (band 4 of the Sentinel product), and NIR is the near-infrared band with central wavelength 842 nm, bandwidth 115 nm (band 8 of the Sentinel product).

Sentinel 2 satellite products do not have a thermal band. Therefore, Landsat satellite products were used for calculation of Land Surface Temperature (LST) and Soil Moisture Index (SMI). Landsat Level 1 products were radiometrically and geometrically corrected products. These products can be rescaled to the Top-Of-Atmosphere reflectance or spectral radiance because they are represented by Digital Numbers (DNs) (USGS 2019). Before calculating spectral

indices, Landsat 8 Level 1 images were atmospherically corrected using the Semi-Automatic Classification Plugin (SCP) of the QGIS 3.26.0 software. In order to retrieve LST and SMI from Landsat products, Spectral Radiance and Top-Of-Atmosphere Brightness Temperature were calculated using the methods provided by Landsat 8 (L8) Data Users Handbook (USGS 2019).

Spectral Radiance was calculated by:

$$L_\lambda = ML * Q_{cal} + A_L \tag{3}$$

where L_λ is spectral radiance (W/(m² * sr * µm)), ML is radiance multiplicative scaling factor for the band, A_L is radiance additive scaling factor for the band, Q_{cal} is Level 1 pixel value in DN. Scaling factors can be found from the metadata file attached with the Landsat product.

The formula for Top-of-Atmosphere Brightness Temperature calculation is given below:

$$T = \left(\frac{K_2}{\ln\left(\frac{K_1}{L_\lambda}+1\right)} \right) \tag{4}$$

where T is top-of-atmosphere brightness temperature (K), L_λ is spectral radiance (W/(m² * sr * µm)), K_1 and K_2 is band-specific thermal conversion constant. Band-specific thermal conversion constant can be found from the metadata file attached with the Landsat product (USGS 2019, p. 54).

Temperature in degrees Kelvin (K) can be converted to degree Celsius through the following calculation:

$$\text{Celsius (°C)} = \text{Kelvin (K)-273.15} \tag{5}$$

Land surface emissivity and proportion of vegetation were calculated by following Haldar and Majumder (2022). The proportion of vegetation was calculated using the following equation:

$$P_v = \left[\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right]^2 \tag{6}$$

where Pv is proportion of vegetation, and NDVI is Normalized Difference Vegetation Index calculated by equation 1.

Land surface emissivity was computed by:

$$\varepsilon = 0.004 * Pv + 0.986 \tag{7}$$

where ε is land surface emissivity and Pv is proportion of vegetation calculated by equation 6.

Land surface temperature was determined by following Congedo (2016):

$$LST = \frac{T}{\left[1 + \left\{\left(\lambda * \frac{T}{\rho}\right) * \ln \varepsilon\right\}\right]} \qquad (8)$$

where LST is land surface temperature, λ is wavelength of emitted radiance (centre wavelength of band 10 of the Landsat satellite is 10.8 µm), T is top-of-atmosphere brightness temperature calculated by equation 4, ε is land surface emissivity computed by equation 7, ρ = h * c / s = 14388 µm K (h is Planck's constant = 6.626 * 10^{-34} J s; c is velocity of light = 2.998 * 10^8 m/s; s is Boltzmann constant = 1.38 * 10^{-23} J/K).

Soil moisture was calculated based on the *Temperature – Vegetation Dryness Index (TVDI)* defined by Sandholt et al. (2002). TVDI was calculated by the following equation:

$$TVDI = \frac{T_S - T_{Smin}}{a + b\ NDVI - T_{Smin}} \qquad (9)$$

where T_{Smin} is the minimum surface temperature in the triangle, defining the wet edge, T_S is the observed surface temperature at the given pixel, NDVI is the observed normalised difference vegetation index, a and b are parameters defining the dry edge modelled as a linear fit to data (T_{Smax} = a + bNDVI) where T_{Smax} is the maximum surface temperature observation for a given NDVI (Sandholt et al. 2002, p. 215).

The NDVI image generated by Landsat 8 data was clustered by an interval of 0.1 in order to determine the dry and wet edge of the LST – NDVI spectral space. The maximum and minimum LST values of each NDVI cluster were plotted against corresponding NDVI values. Then dry and wet edges were found from the scatterplot by a linear regression line that was drawn through extracted LST values. The dry edge line was found to be y = - 9.5262 * NDVI + 34.126 (R^2 = 0.5306) and the wet edge line was found to be y = 0.2621 * NDVI + 17.783 (R^2 = 0.0044) (Figure 2). The intercepts and slopes of the dry and the wet edge were used for estimation of TVDI.
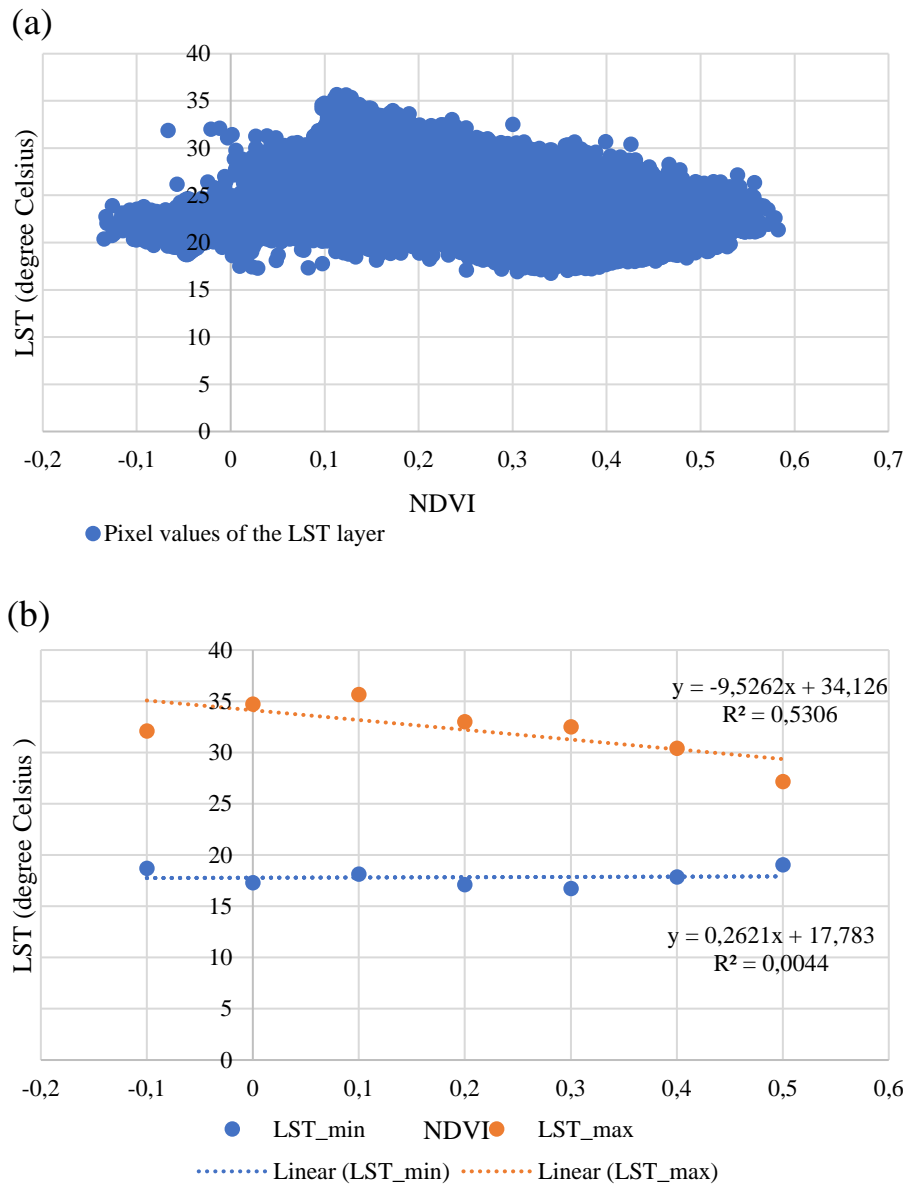
Figure 2. a) NDVI – LST scatter plot for the TVDI and b) the regression coefficients of the dry and wet edge obtained from the maximum and minimum LST values of corresponding NDVI clusters with 0.1 intervals. LTS_min and LST_max were indicated for the minimum and maximum values of the LST within the corresponding NDVI clusters. Linear (LST_min) was indicated by the Wet edge, Linear (LST_max) was indicated by the Dry edge. NDVI and LST were generated from the Landsat 8 OLI Level 1 data (acquisition data: 21 August 2015, path:132 row: 025-026, obtained from the USGS (2020).

Input layers prepared for the subsequent analysis are shown in Figure 3. Here, soum centre and bag centre datasets were combined into a single input layer named by Soum and Bag centre.
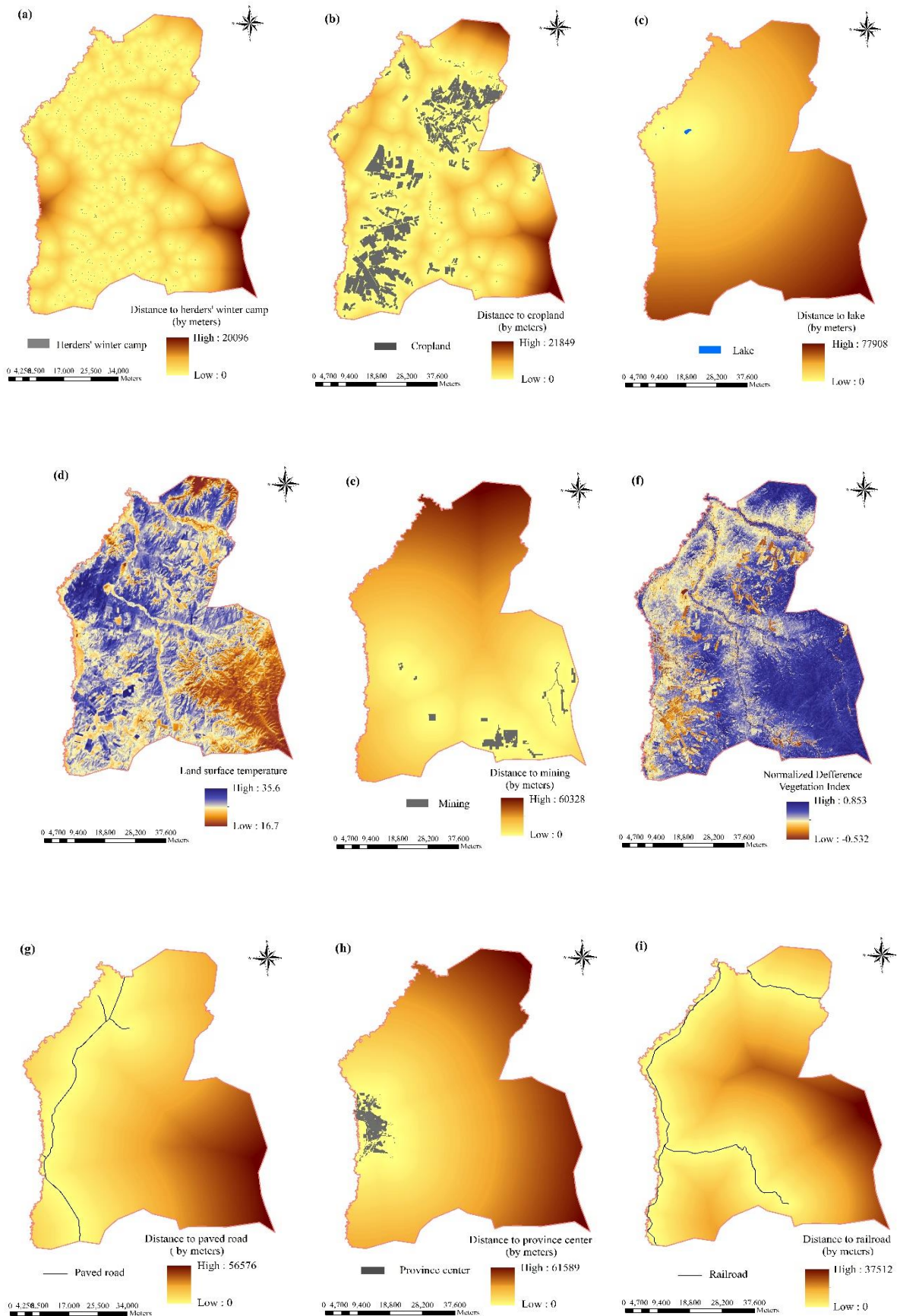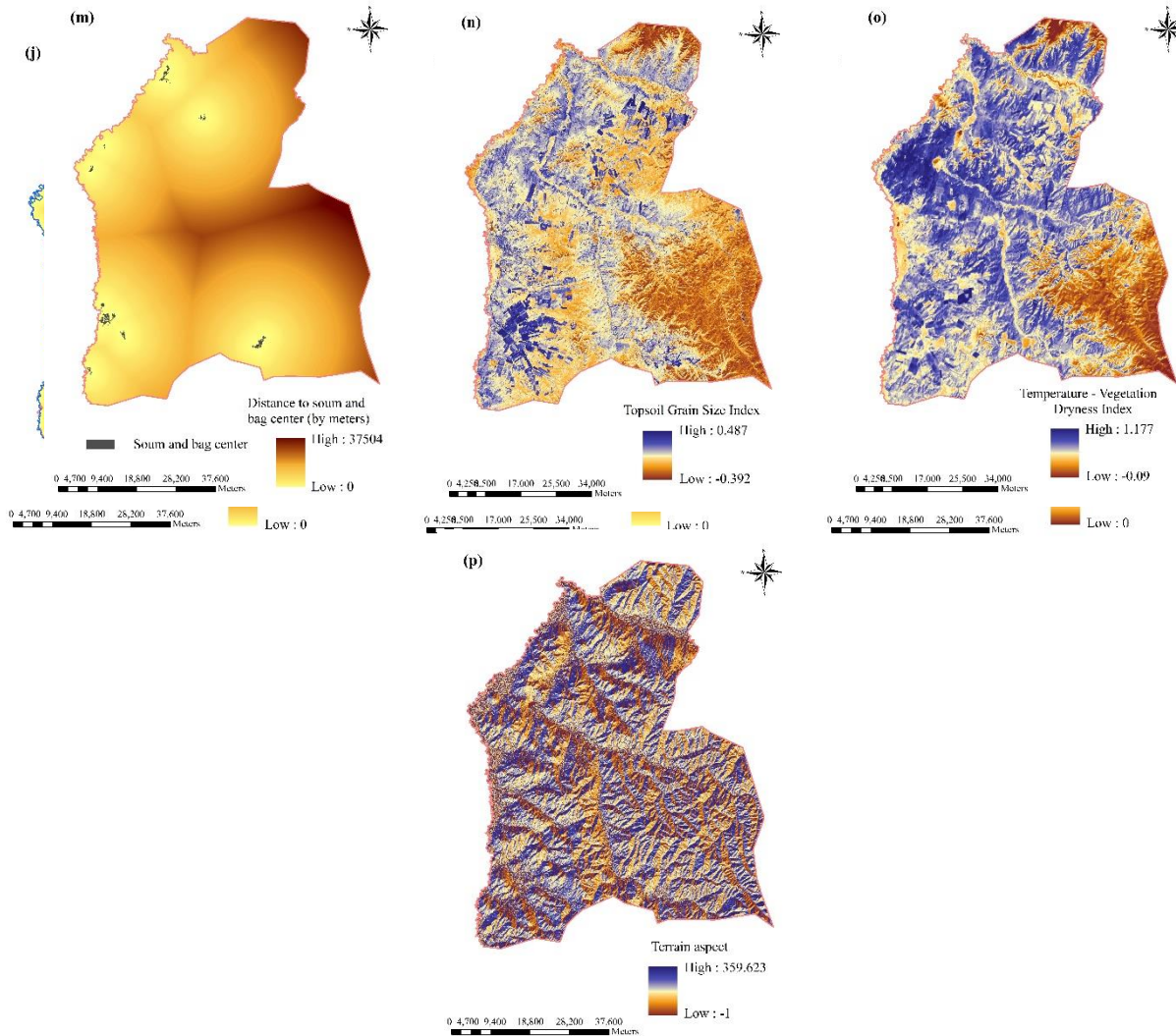
Figure 3. Input layers for the subsequent analysis in this study.

The various input layers use different units due to their various sources and were measured with different approaches. Therefore, in order to eliminate such differences, a data standardization procedure was used in this study. Input layers were individually standardized based on Otgonbayar et al. (2017). After the standardization, values of each input layer were ranged from 0 to 1. The standardization formula was given below:

$$E_i = (X_i - X_{min}) / (X_{max} - X_{min}) \qquad (10)$$

where: E_i is the value of standardized in pixel i, X_{min} is the minimum value of the individual layer, X_{max} is the maximum value of the individual layer.

Data points for the location and variables (independent and dependent) ought to be determined in order to carry out the regression analysis (Du et al. 2014). The systematic sampling method was used for the study. Grid points applied for further analysis were created from the grid cells (area of each grid cell was 1 km²) with the value of dirt road density. A total of 2,461 points

were selected for the analysis. In order to build regression points, pixel values of each input raster layer were extracted by grid points using the Extraction tool in ArcGIS 10.8.1 (ESRI INC 2020).

2.3.2   Variable screening and model selection

In order to define the statistical association between dependent and independent variables, a bivariate correlation analysis was employed. If an independent variable has no association with the dependent variable and a statistically non-significant relationship with the dependent variable, that independent variable was omitted from the subsequent analysis. The high values (more than 0.7) in the Pearson's correlation coefficient matrix indicate the multicollinearity of the variables (Tang et al. 2020). In this study, Pearson's product moment correlation coefficients between independent variables higher than 0.5 were considered one of the criteria to exclude an independent variable.

The Person's product-moment correlation coefficient, *r* can be calculated by the following formula:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{[\sum(X - \bar{X})^2][\sum(Y - \bar{Y})^2]}} \qquad (11)$$

where: r is the Pearson's correlation coefficient, X and Y are values of the variables, $\bar{X}$ and $\bar{Y}$ are means of the variables (Healey & Donoghue 2021, p. 369).

Pearson's product-moment correlation test was performed using the "sjPlot" package (Lüdecke et al. 2022) in Rstudio (RStudio Team 2020). The result of the Pearson's product-moment correlation test is shown in Table 6. There was not any variable that had no correlation (r = 0) with dirt road density, but some variables failed to reject the null hypothesis (Ho: there is no correlation between dirt road density and independent variables). Variables including *distance to lake* (r = 0.02, p = 0.3), *distance to mining* (r = 0.01, p = 0.66), and *terrain aspect* (r = - 0.02, p = 0.43) were not statistically significant in their association with dirt road density. Therefore, those independent variables that have non-significant relationships with the dependent variable were omitted from further analysis.

All the remaining variables had a statistically significant association with dirt road density at a 99% confidence level. These variables were: *distance to herders' winter camp* (r = 0.07), *distance to cropland* (r = 0.11), *Land surface temperature* (LST) (r = 0.12), *Normalized Difference Vegetation Index (NDVI)* (r = - 0.09), *distance to paved road* (r = 0.1), *distance to province centre* (r = 0.09), *distance to railroad* (r = 0.18), *distance to main rivers* (r = 0.06), *distance to tributary rivers* (r = - 0.09), *terrain slope* (r = - 0.24), *distance to soum and bag centre* (r = 0.16), *Top Grain Size Index (TGSI)* (r = 0.12), *Temperature – Vegetation Dryness Index (TVDI)* (r = 0.13).

Results of the Pearson's product-moment correlation test showed that there were intercorrelations between independent variables. Correlation coefficients above 0.5 were found between following variables: *NDVI* and *TGSI* (r = - 0.85, p < 0.001); *TVDI* and *LST* (r = 0.98, p < 0.001); *LST* and *distance to lake* (r = 0.50, p < 0.001); *distance to paved road* and *distance to cropland* (r = 0.5, p < 0.001); *distance to paved road* and *distance to lake* (r = 0.62, p < 0.001); *distance to province centre* and *distance to lake* (r = 0.5, p < 0.001); *distance to province*

centre and *distance to paved road* (r = 0.58, p < 0.001); *distance to tributary rivers* and *distance to mining* (r = 0.61, p < 0.001), *distance to tributary rivers* and *distance to paved road* (r = - 0.77, p < 0.001), *distance to tributary rivers* and *distance to railroad* (r = - 0.54, p < 0.001); *distance to soum and bag centre* and *distance to paved road* (r = 0.51, p < 0.001), *distance to soum and bag centre* and *distance to railroad* (r = 0.55, p < 0.001) and *distance to herders' winter camp* and *distance to cropland* (r = 0.57, p < 0.001). Based on Keshkamat et al. (2013), the *TGSI* and *LST* variables were omitted from the subsequent analysis. Keshkamat et al. (2013) noted that *TGSI* and *LST* were not locally significant variables in the GWR. As a result, variables including *distance to cropland, distance to paved road, distance to railroad,* and *distance to tributary rivers* were excluded from the analysis.

Multicollinearity can result in a biased model result through the inaccurate estimation of regression coefficients (Tang et al. 2020). Dormann et al. (2012, p. 28) noted "Perfect collinearity occurs if predictions are exact linear functions of each other and is simply a case of model misspecification – one variable needs to be omitted". Commonly used approaches for multicollinearity are variance inflation factor (VIF) and Pearson correlation coefficient matrix (Tang et al. 2020). The variance inflation factor value was calculated with the following formula (Daoud 2017):

$$VIF = 1 / (1 - R_i^2) \qquad (12)$$

where: R_i^2 is the coefficient of determination of the independent variable i.

It is considered as multicollinearity if the variance inflation factor value of the independent variable is higher than 7.5 and it should be removed from the regression model (Qiu et al. 2012). Variance inflation factor values of the independent variables were computed using the "car" package (Fox & Weisberg 2019) in RStudio (RStudio Team 2020). Results of the VIF test are shown in Table 2.

Table 2. VIF values of the independent variables.

Variable	VIF value	Variable	VIF value
Distance to herders' winter camp	1.215	Distance to province centre	1.271
NDVI	1.231	Terrain slope	1.109
Distance to main river	1.078	TVDI	1.271
Distance to soum and bag centre	1.188		

There was no independent variable with a high VIF value. VIF values of variables ranged from 1.176 to 2.166. The VIF test results showed that all variables can be well fitted into the regression model in terms of multicollinearity.

Three sub-models were run in parallel through the Ordinary Least Square (OLS) Regression and Geographically Weighted Regression (GWR). All independent variables were arranged into these three sub-models. The variables related to natural factors including *distance to main rivers, terrain slope, NDVI, and TVDI* were arranged into the first sub-model (Model 1), the variables representing land use factors (*distance to province centre, distance to herders' winter camp, and distance to soum and bag centre*) were arranged into the second sub-model (Model 2), and factors and all of variables together were arranged into the third sub-model (Model 3).

In order to find the appropriate models in which all variables are statistically significant to the model, the Exploratory Regression Tool in the Spatial Statistics Toolbox in Arc GIS Pro 2.9 (ESRI INC 2021) was used in this study. The Exploratory Regression Tool can find the best combination of candidate variables for the OLS regression model based on the evaluation of all possible combinations of the input independent variables within the threshold of user-specified criteria (ESRI n.d.). In the exploratory regression, the following search criteria were taken: 10 for the maximum number of explanatory variables, 5 for the minimum number of explanatory variables, 0.5 for the minimum acceptable adjusted R squares, and 0.05 for the maximum coefficient p-value cut-off. Potential combinations of candidate variables found by the exploratory regression were then compared with each other by the adjusted R squares and Akaike's Information Criterion (AIC).

Both R² and Akaike's Information Criterion (AIC) are measurements of model performance. Weisent et al. (2012, p. 7) noted that "based the AIC goodness-of-fit statistic for comparing models, the model with the lowest AIC statistic is the one with the best model fit". The goodness of fit and the complexity of the model can be measured by AIC (Du et al. 2014). The AIC can be calculated as follows:

$$AIC = 2n\,ln(\hat{\sigma}) + n\,ln(2\pi) + n\left(\frac{n=tr(S)}{n-2-tr(S)}\right) \qquad (13)$$

where: n is the number of the sample size, $\hat{\sigma}$ is the standard deviation of the error term, and tr(S) is the trace of the hat matrix.

Exploratory regression results showed that there was not any combination of candidate variables that met the criterion of adjusted R². Adjusted R² of all possible combinations was 0.06-0.07 in the natural factors data set (Table 3), 0.02-0.03 in the land use related factors data set (Table 4), and 0.07-0.08 in the integration of all data sets (Table 5).

Table 3. Summary of potential combinations of candidate variables from the natural factors data set (*SP* is *terrain slope*, *TVDI* is *Temperature – Vegetation Dryness Index*, *RV_I* is *distance to main river*, and *NDVI* is *Normalized Difference Vegetation Index*). (p-value: *** is p < 0.01, ** is p < 0.05, and * is p < 0.1).

Potential combination of candidate variables	Adjusted R²	AICc
- SP***	0.06	- 3052.48
+ TVDI*** - SP***	0.06	- 3071.26
- NDVI*** - SP***	0.06	- 3054.12
+ RV_I*** - SP***	0.06	- 3060.06
+ RV_I*** - SP*** - NDVI*	0.06	-3060.65
+ RV_I*** -SP*** + TVDI***	0.07	- 3079.88

Table 4. Summary of potential combinations of candidate variables from the land use related factors data set (*SC* is *distance to soum and bag centre*, *PC* is *distance to province centre*, and *WC* is *distance to herders' winter camp*). (p-value: *** is p < 0.01, ** is p < 0.05, and * is p < 0.1)

Potential combination of candidate variables	Adjusted R²	AICc
+ SC***	0.02	- 2971.33
+ SC*** + WC**	0.02	- 2972.52
+ SC*** + PC***	0.03	-2977.16
+ PC** + SC*** + WC*	0.03	- 2977.52

Table 5. Summary of potential combinations of candidate variables from the integration of all data sets (*SP* is *terrain slope*, *TVDI* is *Temperature – Vegetation Dryness Index*, *RV_I* is *distance to main river*, and *SC* is *distance to soum and bag centre*). (p-value: *** is p < 0.01, ** is p < 0.05, and * is p < 0.1)

Potential combination of candidate variables	Adjusted R²	AICc
- SP*** + SC***	0.07	-3085.35
+ SC*** - SP*** + TVDI***	0.07	- 3099.84
+ RV_I*** - SP*** + SC***	0.07	-3094.70
+ SC*** - SP*** + TVDI*** + RV_I***	0.08	-3110.09

Among the possible combinations, the following models were chosen for subsequent analysis in this study.

Model 1:
+ RV_I*** -SP*** + TVDI***

Model 2:
+ SC*** + PC** + WC*

 Model 3:
+ SC*** - SP*** + TVDI*** + RV_I***

2.3.3   Ordinary Least Square Regression

The Ordinary Least Square (OLS) Regression is a global linear regression model (Zhang et al. 2020). Fotheringham and Sachdeva (2022, p. 2) noted, "A global model produces a single estimate for each conditioned relationship represented in the model".

The OLS can be expressed as

$$y = \beta_0 + \sum_k \beta_k x_k + \varepsilon \qquad (14)$$

where: y is the dependent variable, β_0 is the intercept, x is the observed matrix about k independent variables, β_k is the kth coefficient, and ε is a vector of k random error terms with distribution N(0, σ^2 I) (Du et al. 2014).

2.3.4   Geographically Weighted Regression

Pimpler (2017, p. 138) stated, "Geographically Weighted Regression (GWR) is a local form of linear regression for modelling spatially varying relationships. GWR works by creating a local model of the variables or process that you are attempting to understand". In the GWR, the model parameters assume the coefficients of the independent variables to be non-stationary and to vary across each location in the study area. In the distinct locations, the parameters of the GWR model are various and each location has its own local regression parameters representing the relationship (Huang & Leung 2002).

The GWR model can be expressed by the following equation:

$$Z(s_i) = \beta_0\,(s_i) + \beta_1\,(s_i)\,x_1\,(s_1) + \ldots + \beta_k\,(s_i)\,x_k\,(s_i) + \varepsilon_i \qquad (15)$$

where s_i is the location at which the parameters are estimated

The parameters for GWR may be estimated by solving:

$$(s_i) = (X^T\,W\,(s_i)\,X)^{-1}\,X^T\,W\,(s_i)\,z \qquad (16)$$

where $W\,(s_i)$ is a n by n matrix, the diagonal elements of which are the geographical weightings of observations around point i:

$$\mathbf{W}(\mathbf{s}_i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix} \qquad (17)$$

where w_{in} is the weight assigned to the observation at location n (Lloyd 2010, p. 123).

The Gaussian function has been widely used in GWR contexts. With this function, a weight at the observation i is obtained with:

$$w_{ij} = \exp\,[\,-0.5\,(d\,/\,\tau)^2\,] \qquad (18)$$

where d is the Euclidean distance between the location of observation i, and location j and τ is the bandwidth of the kernel (Lloyd 2010, p. 74).

The goodness-of-fit of a GWR can be assessed using the geographically weighted coefficient of determination:

$$r^2\,(\,s_i\,) = (\,TSS^w - RSS^w\,)\,/\,TSS^w \qquad (19)$$

where TSS^w is the geographically weighted total sum of squares:

$$TSS^w = \sum_{j=1}^{n} w_{ij}\,(z_j - \overline{z_i})^2 \qquad (20)$$

and RSS^w is the geographically weighted residual sum of squares:

$$RSS^w = \sum_{j=1}^{n} w_{ij} (z_j - \hat{z}_j)^2 \qquad (21) \qquad (Lloyd\ 2010,\ p.\ 124)$$

GWR coefficients of each of the independent variables were examined by its t - value at the 95 % confidence level.

2.3.5 Spatial autocorrelation analysis

Mennis and Jordan (2005, p. 252) stated, "Spatial autocorrelation, also referred to as spatial dependency, occurs when the distribution of the values of georeferenced observations is not spatially random; rather, observations located near one another tend to have similar (or particularly dissimilar) values". Moran's I is a commonly used spatial autocorrelation test (Longley et al. 2001). Global and local Moran's I tests were used for verifying spatial stationarity of regression residuals.

Moran's I is global in the sense that it estimates the overall degree of spatial autocorrelation for a data set. Local spatial autocorrelation statistics provide estimates disaggregated to the level of the spatial analysis units, allowing an assessment of the dependency relationships across space. Moran's I behaves like a Pearson correlation coefficient. Its value is generally between – 1 and + 1. Positive values indicate positive autocorrelation and vice versa. Global Moran's I is calculated as follows:

$$I(d) = \frac{\frac{1}{W} \sum_i \sum_{i \neq 1} w_{ij}(z_i - \bar{z}) \times (z_j - \bar{z})}{\frac{1}{n} \sum_i (z_i - \bar{z})^2} \qquad (22)$$

where I(d) is the Moran coefficient for the distance class d, z_i's are the values of the variable, and i and j vary from 1 to n. w_{ij}'s take the value 1 when the pair of location (i,j) pertains to distance class d and 0 otherwise. W is the sum of the w_{ij}'s (Kalkhan 2011, p. 66).

The Anselin Local Moran's I can be calculated by the following equation:

$$Ii = \left(\frac{x_i - \bar{X}}{s_i^2}\right) \sum_{i=1, j \neq i}^{n} w_{ij} (x_j - \bar{X}) \qquad (23)$$

$$s_i^2 = \frac{\sum_{j=1, j \neq i}^{n}(x_j - \bar{X})}{n-1} - \bar{X}^2 \qquad (24)$$

where x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, w_{ij} is the spatial weight between feature i and j, n is the total number of features (Guo et al. 2021).

2.3.6 Spatial non-stationarity analysis

Selection of a bandwidth for the GWR is vital in terms of shaping the fit of the model (Charlton & Fotheringham 2009). The scale dependence of non-stationarity in the GWR parameter coefficients can be defined by following formula (Zhao et al. 2015):

$$SI = \frac{\beta_{GWR\_iqr}}{2 \times OLS_{se}} \qquad (25)$$

where SI is the stationary index, β_{GWR_iqr} is the interquartile range of standard errors for each independent variable's estimated coefficients in the GWR, and OLS_{se} is the standard error of the same variable's coefficients in the OLS regression.

Values of the stationary index of more than 1 (SI > 1) showed significant spatial non – stationarity at p < 0.05 level (Liu et al. 2019). Scale dependence of spatial non-stationarity in the local parameter coefficients were defined by iterated GWR analysis with increasing fixed kernel bandwidths from 1 km to 12 km at 1 km increments.

3. RESULTS

3.1 Spatial characteristics of dirt road distribution and bivariate correlations

A total of 2,998 kilometres of long dirt roads were determined in this study. Dirt road densities defined by total dirt road length per 1 km² land area varied all over the study area. The amount of dirt road densities ranged from 1.45 m/km² to 5965 m/km². The mean density of dirt road was 1218 m/km². A histogram of dirt road density distribution in the study area is shown in Figure 4 and the spatial distribution of the dirt road density is shown in Figure 5.

The results of Global Moran's I statistics (Moran's I 0.15, z - score 8.84, p < 0.000) showed that there was positive spatial autocorrelation in dirt road density. The spatial autocorrelation of dirt road density detected by Anselin Local Moran's I statistics (LISA) is shown in Figure 6. Confidence level for all clusters were 95%. High-High (HH) clusters indicated that area with a high value of dirt road density was surrounded by other areas with a high value of dirt road densities. HH clusters were mainly distributed near the settlement areas. Moreover, they were observed around the mining areas, not only the Shayn Gol open pit coal mining area, but also gold mining areas in the upper reach of the Sharyn Gol River. Some of the HH clusters observed along the Sharyn Gol River valley and the northern part of the Eroo river were formed in grasslands that are used for herders' summer camps. High-Low (HL) and Low-High (LH) outliers are produced when area with a high (or low) value of dirt road density is surrounded by other areas with a low (or high) value of dirt road density. Generally, the LH outliers were observed near the HH clusters, and the LL clusters were noticed in the mountains.
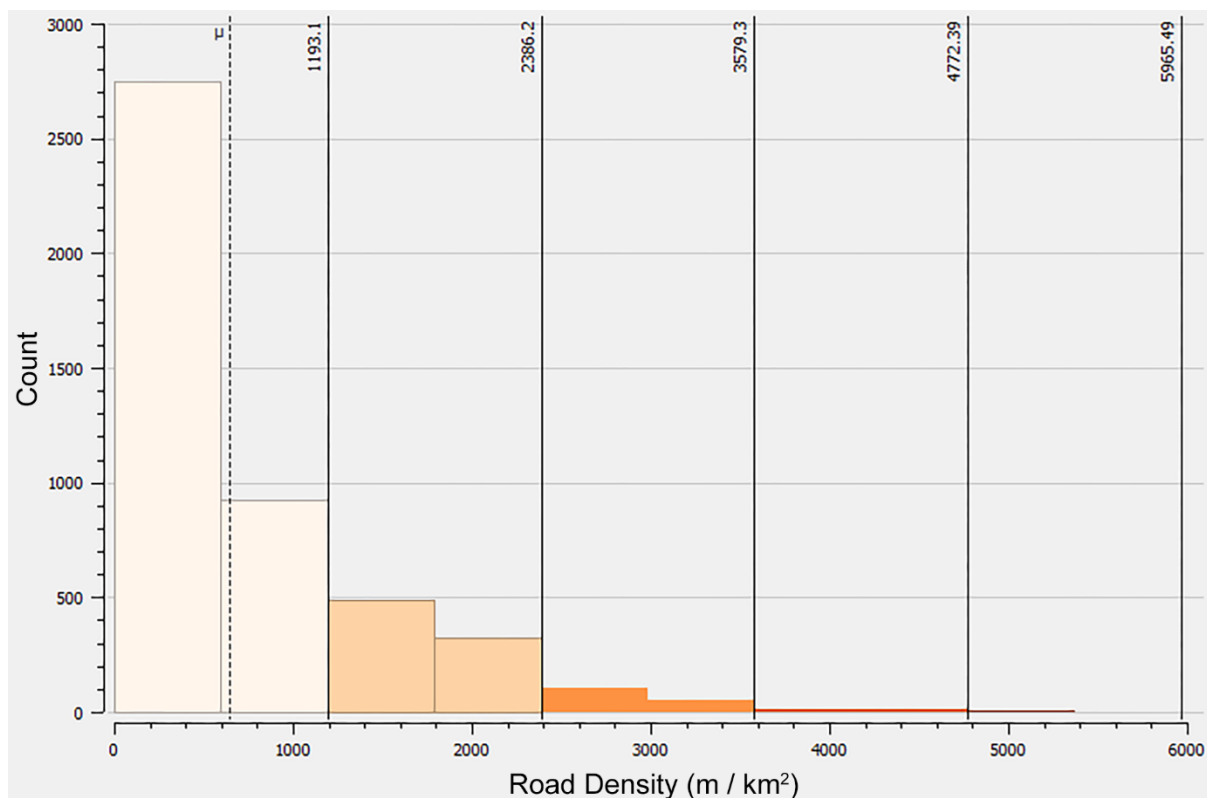
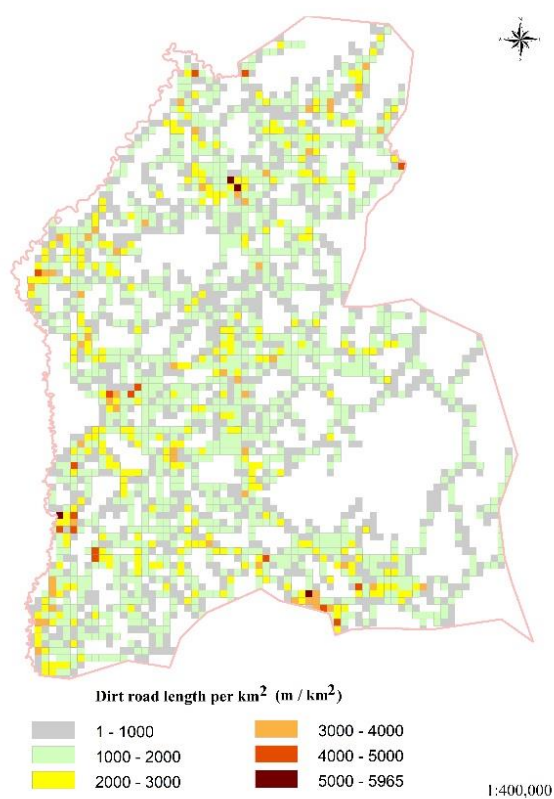Figure 4. Histogram of dirt road density (m/km²) in the study area.

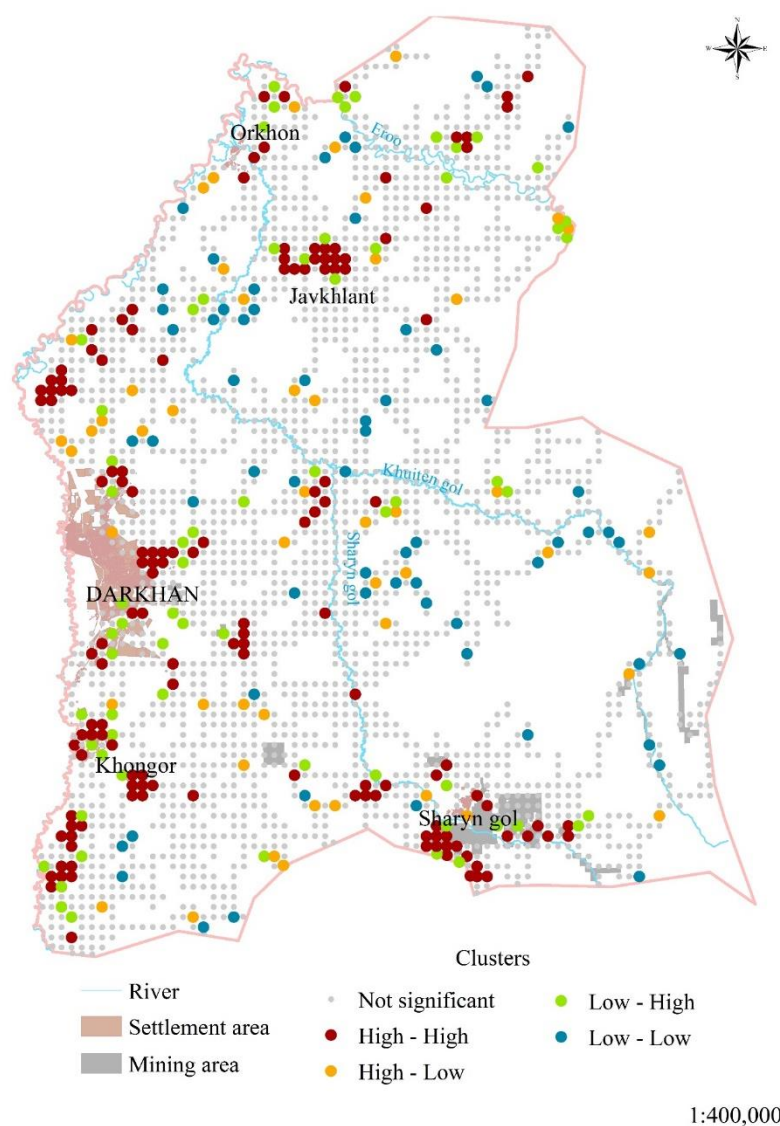Figure 5. Spatial distribution of the dirt road density in the study area.

Figure 6. Spatial autocorrelation of dirt road density detected by Anselin Local Moran's I statistics (p < 0.05). High-High and Low-Low clusters indicate positive spatial autocorrelation (an area neighbouring by area with a similar dirt road density value), High-Low and Low-High outliers indicate negative spatial autocorrelation (an area neighboured by an area with a different dirt road density value).

The result of Pearson's correlation test is shown in Table 6. As can be seen in table 6, the statistical association between dependent variable (*dirt road density*) and independent variables were positive; these include *distance to herders' winter camp, distance to cropland, distance to lake, distance to mining, distance to paved road, distance to province centre (city), distance to railroad, distance to main river, distance to soum and bag centre, Top Grain Size Index (TGSI), land surface temperature (LST), and Temperature – Vegetation Dryness Index (TVDI)*. In contrast, the statistical association between *dirt road density* and *Normalized Difference Vegetation Index (NDVI), distance to tributary river, terrain aspect* and *terrain slope* were negative. The relationship between *dirt road density* and *distance to mining, distance to lake*, and *terrain aspect* were statistically insignificant.

Table 6. Correlations between dependent (*dirt road density* (DR)) and independent variables, and inter-correlations between independent variables (n = 2461). Independent variables are *distance to soum and bag centre (SC), distance to herders' winter camp (WC), distance to cropland (CL), distance to lake (LK), land surface temperature (LST), distance to mining (MG), Normalized Difference Vegetation Index (NDVI), distance to paved road (PR), distance to province centre (PC), distance to railroad (RR), distance to main river (RV_I), distance to tributary river (RV_II), terrain slope (SP), terrain aspect (AP), Top Grain Size Index (TGSI)*, and *Temperature - Vegetation Dryness Index (TVDI)*. The computed correlation used Pearson-method with listwise-deletion. Statistical significance: *** is $p < 0.001$, ** is $p < 0.01$, * is $p < 0.05$, and ° is $p < 0.1$.

	DR	CL	LK	LST	MG	NDVI	PR	PC	RR	RV_I	RV_II	SP	TGSI	TVDI	AP	SC	WC
DR	1	0.11***	0.02	0.12***	0.01	-0.09***	0.10***	0.09***	0.18***	0.06**	-0.09***	-0.24***	0.12***	0.13***	-0.02	0.16***	0.07***
CL	0.11***	1	0.25***	0.23***	-0.03	-0.24***	0.50***	0.44***	0.25***	-0.19***	-0.12***	-0.30***	0.30***	0.25***	-0.01	0.27***	0.57***
LK	0.02	0.25***	1	0.50***	-0.35***	-0.31***	0.62***	0.50***	0.12***	0.06**	-0.40***	-0.21***	0.27***	0.44***	-0.05*	0.07***	0.27***
LST	0.12***	0.23***	0.50***	1	-0.25***	-0.30***	0.47***	0.32***	0.22***	-0.04*	-0.35***	-0.16***	0.44***	0.98***	-0.12***	0.15***	0.34***
MG	0.01	-0.03	-0.35***	-0.25***	1	0.21***	-0.39***	0.41***	-0.12***	-0.12***	0.61***	0.10***	-0.16***	-0.21***	0.04*	-0.10***	-0.27***
NDVI	-0.09***	-0.24***	-0.31***	-0.30***	0.21***	1	-0.45***	-0.27***	-0.33***	-0.09***	0.40***	0.23***	-0.85***	-0.22***	0.01	-0.30***	-0.14***
PR	0.10***	0.50***	0.62***	0.47***	-0.39***	-0.45***	1	0.58***	0.46***	-0.23***	-0.77***	-0.32***	0.41***	0.41***	0.00	0.51***	0.40***
PC	0.09***	0.44***	0.50***	0.32***	0.41***	-0.27***	0.58***	1	0.35***	-0.20***	-0.20***	-0.23***	0.28***	0.30***	0.01	0.24***	0.14***
RR	0.18***	0.25***	0.12***	0.22***	-0.12***	-0.33***	0.46***	0.35***	1	0.15***	-0.54***	-0.17***	0.31***	0.19***	-0.02	0.55***	0.30***
RV_I	0.06**	-0.19***	0.06**	-0.04*	-0.12***	-0.09***	-0.23***	-0.20***	0.15***	1	0.06**	0.01	0.10***	-0.03	-0.05**	-0.04*	-0.09***
RV_II	-0.09***	-0.12***	-0.40***	-0.35***	0.61***	0.40***	-0.77***	-0.20***	-0.54***	0.06**	1	0.23***	-0.31***	-0.27***	-0.00	-0.41***	-0.19***
SP	-0.24***	-0.30***	-0.21***	-0.16***	0.10***	0.23***	-0.32***	-0.23***	-0.17***	0.01	0.23***	1	-0.24***	-0.17***	0.01	-0.18***	-0.13***
TGSI	0.12***	0.30***	0.27***	0.44***	-0.16***	-0.85***	0.41***	0.28***	0.31***	0.10***	-0.31***	-0.24***	1	0.39***	-0.03	0.27***	0.22***
TVDI	0.13***	0.25***	0.44***	0.98***	-0.21***	-0.22***	0.41***	0.30***	0.19***	-0.03	-0.27***	-0.17***	0.39***	1	-0.14***	0.12***	0.37***

AP	-0.02	-0.01	0.05*	-0.12***	0.04*	0.01	0.00	0.01	-0.02	0.05**	-0.00	0.01	-0.03	-0.14***	1	0.03	-0.01
SC	0.16***	0.27***	0.07***	0.15***	-0.10***	-0.30***	0.51***	0.24***	0.55***	-0.04*	-0.41***	-0.18***	0.27***	0.12***	0.03	1	0.23***
WC	0.07***	0.57***	0.27***	0.34***	-0.27***	-0.14***	0.40***	0.14***	0.30***	-0.09***	-0.19***	-0.13***	0.22***	0.37***	-0.01	0.23***	1

3.2 Spatial statistical analysis of the influencing factors on dirt road density

3.2.1 Scale dependence of spatial relationship

Results of the spatial non-stationarity analysis showed that there was a spatial non-stationary relationship between dirt road density and independent variables. Changes in spatially non-stationarity of the relationship between dependent and independent variables vary at a different scale. Figure 7 presents the stationary index response to the bandwidth changes of the GWR. As shown in Figure 7, the plotted stationary index values against the kernel bandwidth distances decline to a certain kernel distance and then become flat. Variables including *Normalized Difference Vegetation Index, Temperature – Vegetation Dryness Index, terrain slope* reached to spatially stationary within small spatial distances. These variables turned into stationary between 3 km and 4 km. Land use related variables had not only a sharp declining gradient but also relatively larger stationary index values. Distance thresholds were found at 10 km for the *distance to herder's winter camp*, 6 km for the *distance to soum and bag centre*, 7 km for the *distance* to *province centre*, and 5 km for the *distance to main river* variable. Although the threshold distances to become stationary vary for the variables, all the independent variables had reached a stationary state within 10 km. These different stationary index values suggested that independent variables could influence the dirt road density expansion at various spatial scales.

3.2.2 Analysis of the influencing factors on dirt road density based on Ordinary Least Square regression

Results of the OLS regression models in which the relationship between dirt road density and independent variables assumed spatially stationary throughout the study area are shown in Table 7. The sub-model (Model 3) with four independent variables were a slightly better fit to the data than other models as shown by the higher Adjusted R² and the lower AICc. Adjusted R² of the Model 3 was 0.078, and the AICc value was – 3,110. Model 2 with land-use related variables could explain only two percent of the dirt road density expansion. This model (Model 2) has the lowest Adjusted R² and the highest AICc value (AICc = - 2,977). The six percent of the dirt road density expansion in the study area could be explained by natural factors (Model 1). AICc value of the Model 1 was – 3,079.

All variables in Model 1 and Model 3 were statistically significant for the regression model with a 99 % confidence level. In Model 2, two from a total of three variables (*distance to herders' winter camp, distance to province centre, distance to soum and bag centre*) were statistically significant for the regression model. The variable representing *herders' seasonal*

location was statistically insignificant for Model 2. This variable might have an impact on dirt road expansion, but its contribution to the regression model is not as strong as other variables.
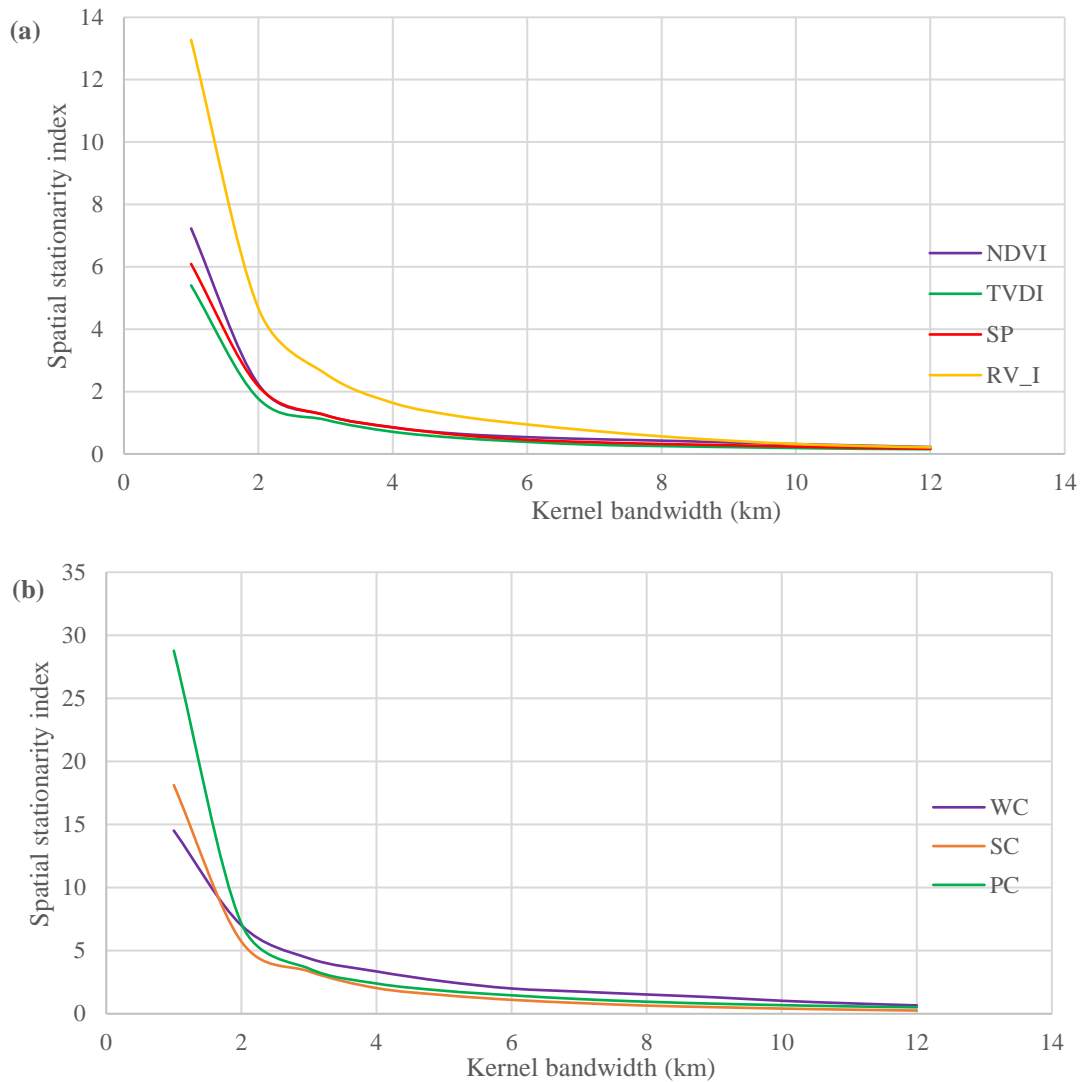

Figure 7. Variation in the spatial stationarity index values for (a) natural factors (*Normalized Difference Vegetation Index (NDVI), Temperature – Vegetation Dryness Index (TVDI), terrain slope (SP)*, and *distance to main river (RV_I))* and (b) land use related variables (*distance to herders' winter camp (WC), distance to soum and bag centre (SC), distance to province centre (PC))*. Spatial stationarity index (SI) values estimated by dividing the interquartile range of standard error for an independent variable's GWR coefficients by twice the standard error of the coefficient for the same variable in OLS. SI > 1 indicate spatial non-stationarity at a given kernel bandwidth in GWR.

Table 7. Summary of results for the Ordinary Least Square (OLS) regression. Models fit to explore relationship between dirt road density and independent variables (*distance to herders' winter camp (WC), Normalized Difference Vegetation Index (NDVI), distance to province centre (PC), distance to main rivers (RV_I), terrain slope (SP), distance to soum and bag centre (SC),* and *Temperature - Vegetation Dryness Index (TVDI).*

Variable	Coefficient	Standard error	T statistics	Probability	VIF
		Model 1			
Intercept	0.151	0.017	8.665	0.000*	
RV_I	0.04	0.01	3.262	0.001*	1.00
SP	-0.319	0.028	-11.240	0.000*	1.02
TDVI	0.112	0.024	4.679	0.000*	1.02
Adjusted R²					0.066
AICc					-3,079.884
		Model 2			
Intercept	0.098	0.020	4.888	0.000*	
WC	0.034	0.022	1.539	0.123	1.066
PC	0.031	0.011	2.648	0.008*	1.072
SC	0.085	0.013	6.447	0.000*	1.111
Adjusted R²					0.027
AICc					-2,977.522
		Model 3			
Intercept	0.106	0.019	5.561	0.000*	
RV_I	0.043	0.012	3.501	0.000*	1.00
SP	-0.292	0.028	-10.232	0.000*	1.056
TDVI	0.100	0.024	4.174	0.000*	1.037
SC	0.071	0.012	5.688	0.000*	1.043
Adjusted R²					0.078
AICc					-3,110.087

Global Moran's I for the residuals were 0.11 (z – score 6.623, p – value 0.000) in Model 1, 0.12 (z – score 7.34, p – value 0.000) in Model 2, and 0.1 (z – score 5.89, p – value 0.000) in Model 3. These Moran's I values indicated that there was a high spatial autocorrelation in the models. Given z – scores of the Moran's I were revealed there was a less than 1% likelihood these spatially clustered pattern could be the result of random change.

3.2.3 *Analysis of the influencing factors on dirt road density based on GWR*

Geographically Weighted Regression (GWR) models in which the relationship between dirt road density and independent variables assumed spatially non-stationary were a better fit to the data than the Ordinary Least Square regression (OLS) models. A decrease in the Akaike's Information Criterion (AICc) and an increase of the adjusted R² was observed in all GWR models. The adjusted R² increased from 0.066 to 0.122 in Model 1, from 0.027 to 0.087 in Model 2, and from 0.078 to 0.13 in Model 3. Differences in AICc values between the OLS and the GWR were 57.3 in Model 1, 95.54 in Model 2, and 38 in Model 3. Since differences in the AICc values between the GWR models and OLS models were more than 3, GWR models were considered a significantly better fit to the data. The model summary of the GWR results and its comparison with the OLS models were shown in Table 8.

Table 8. Model performance comparison between the Geographically Weighted Regression (GWR) and the Ordinary Least Square regression (OLS).

Comparison measure	Model 1		Model 2		Model 3	
	OLS	GWR	OLS	GWR	OLS	GWR
Adjusted R²	0.066	0.122	0.027	0.087	0.078	0.1308
AICc	-3,079.88	-3,137.18	-2,977.52	-3,073.06	-3,110.08	-3,148.08
Sigma	0.0166	0.0157	0.01739	0.0163	0.0164	0.0156
Global Moran's I	0.11	0.013	0.12	0.020	0.1	0.0017

Local R² values in the models revealed spatial variations through the study area (Fig. 6). Local R² values ranged from 0.04 to 0.42 in Model 1, from 0.005 to 0.25 in Model 2, and from 0.03 to 0.43 in Model 3. The percentage of local models that have R² values below the R² values of the corresponding OLS models was 3.12% in Model 1, 2.23% in Model 2, and 6.9% in Model 3. These regression points with local R² values lower than the corresponding OLS models were mainly distributed in areas where all local models fitted poorly to the dataset. Croplands in the west side of the Sharyn Gol River and gold mining areas in the upper reach of the Khuiten River were underestimated by GWR models compared to OLS models.
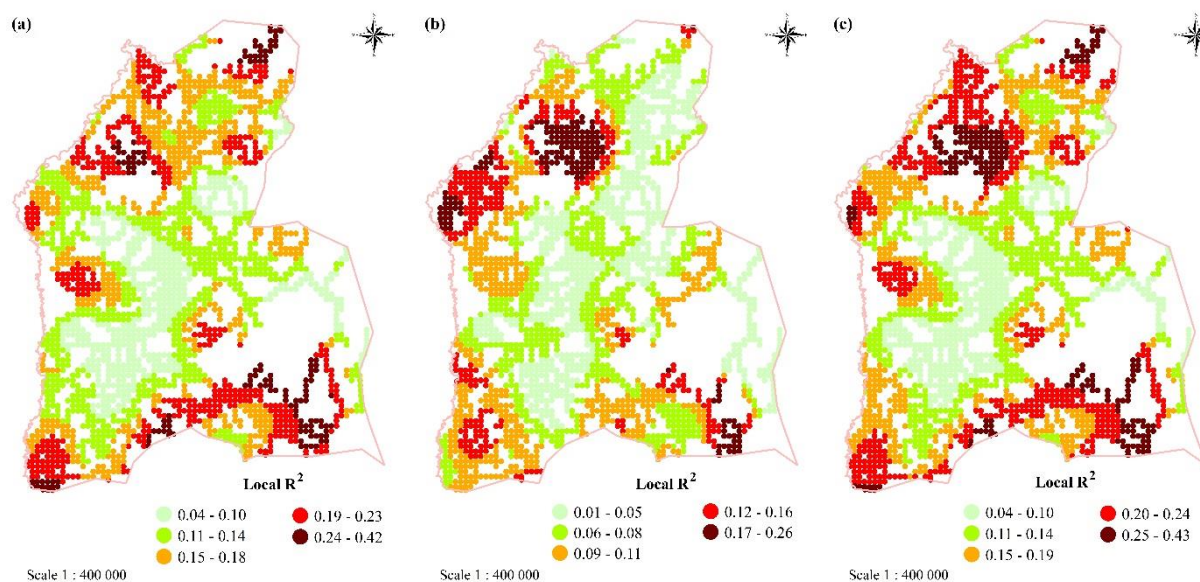

Figure 8. Spatial distribution of the local R² values of the GWR models: a) Model 1 fitted to the natural factors dataset, b) Model 2 fitted to the land use related variables dataset, and c) Model 3 fitted to the combination of the natural and land use related variables dataset. All models were computed by the Gaussian kernel weighting function with a fixed distance bandwidth of 4 km.

The calculated Global Moran's I values for residuals in the GWR models were reduced from the same index values in the corresponding OLS models (see Table 8). Figure 9 shows the spatial autocorrelation of residuals in the OLS regression and the GWR models identified by the Anselin Local Moran's I statistics. Generally, spatial autocorrelation of the residual's clusters in the GWR models were declined compared to the corresponding OLS models, but there were still some areas that are under- or overestimated by all GWR models. These areas were mainly found around the settlement centres and near the junction of the Khuiten River and the Sharyn Gol River.
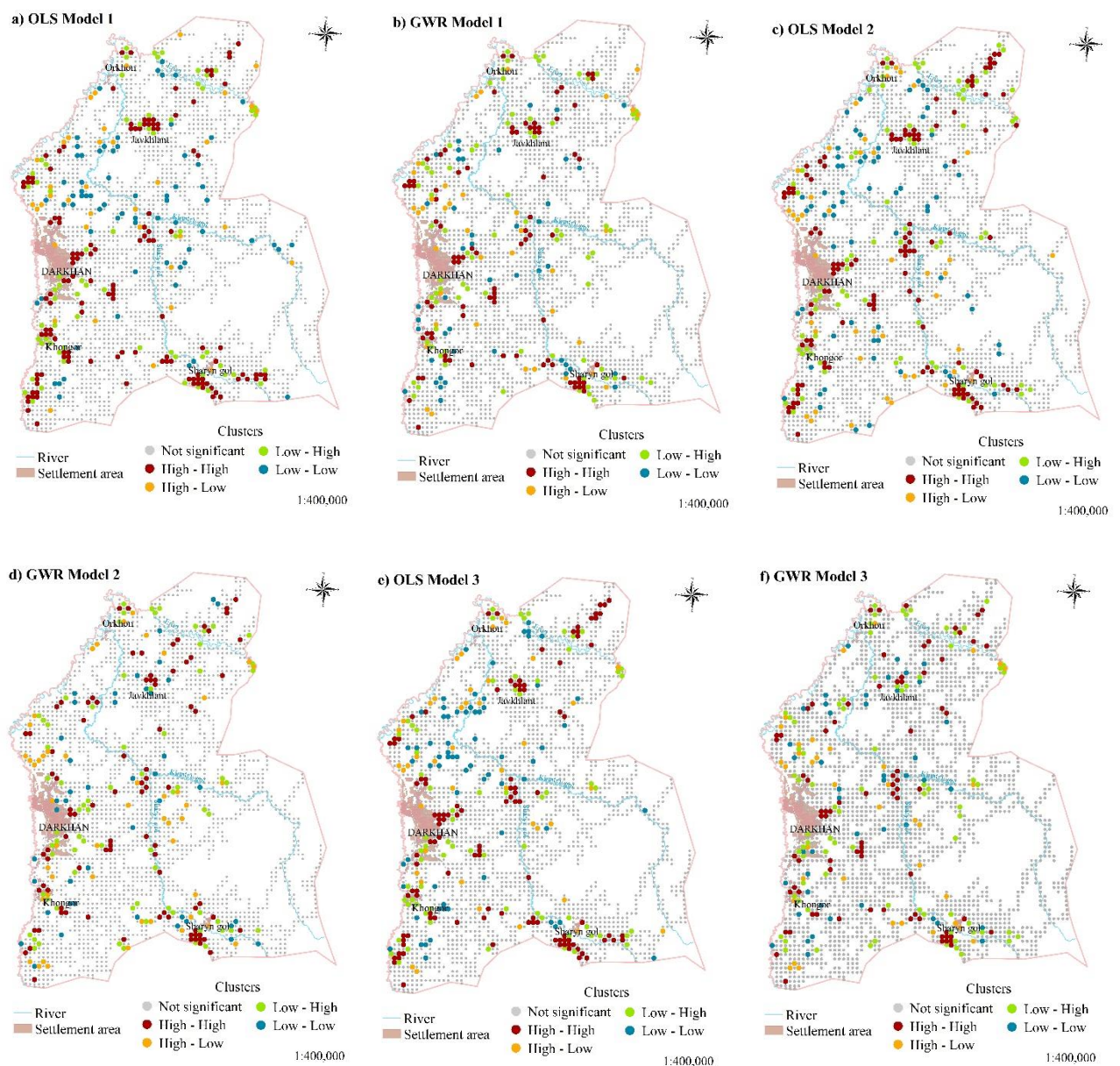
Figure 9. Spatial autocorrelation of the OLS and the GWR models' residuals detected by the Anselin Local Moran's I statistics algorithm in Arc GIS Pro 2.9 ($p < 0.05$). High-High clusters indicate spatial autocorrelation of the positive residuals (underestimation), Low-Low clusters indicated spatial autocorrelation of the negative residuals (overestimation), Low-High and High-Low outliers indicate the point with negative (or positive) residual value is surrounded by points with positive (or negative) residual values.

All variables except terrain slope had spatially varying positive and negative regression coefficients (Table 9). Regression coefficients of the terrain slope variable in both Model 1 and Model 3 were negative. These spatially varying negative and positive coefficients indicate that there were spatially non-stationary relationships between dirt road density and independent variables (Fig. 7, 8, and 9). Among the independent variables, terrain slope not only had a negative association with dirt road density, but also that these associated relationship covers relatively larger areas than other independent variables. Generally, the strong negative

correlation between dirt road density and terrain slope was found in areas with very little terrain slope gradient.

Table 9. Descriptive statistics of the estimated regression coefficients found through GWR modelling at the 95% confidence level. Models fit to explore relationship between dirt road density and independent variables (*distance to herders' winter camp (WC), distance to province centre (PC), distance to main rivers (RV_I), terrain slope (SP), distance to soum and bag centre (SC)*, and *Temperature - Vegetation Dryness Index (TVDI)).* GWR models were computed by the Gaussian kernel weighting function with a fixed distance bandwidth of 4 km.

Variable	Minimum	Q1	Median	Q2	Maximum
Model 1					
INTRCPT	-0.540	0.208	0.249	0.306	0.694
RV_I	-0.373	0.131	0.179	0.248	0.432
SP	-0.950	-0.548	-0.443	-0.340	-0.208
TDVI	-0.305	0.304	0.366	0.436	1.067
Model 2					
INTRCPT	-1.840	-1.134	-0.913	-0.571	0.680
PC	-2.358	0.304	0.480	0.659	1.813
SC	-0.567	0.306	0.511	0.696	1.364
WC	-0.676	0.428	0.612	0.801	1.452
Model 3					
INTRCPT	-0.893	-0.389	0.223	0.267	0.448
RV_I	-0.590	0.148	0.207	0.259	0.391
SP	-0.906	-0.484	-0.388	-0.292	-0.172
TVDI	-0.373	0.311	0.375	0.446	0.972
SC	-0.449	0.290	0.395	0.565	1.100

The positive association between dirt road density and proximity to the main river was observed at some part of the Sharyn Gol River, the junction of the Sharyn Gol and the Khuiten Gol, and the Kharaa River as well as a small part of the Eroo River. The negative association between variables mentioned above was found near the Sharyn Gol open pit coal mine and around the vegetable planting fields in the down reach of the Sharyn Gol River. Proximity to soum and bag centres has significant positive correlation with dirt road density expansion. This association was becoming stronger near bag and soum centres. The positive association between dirt road density and proximity to province centre was observed in some parts of the study area. Although some High-High clusters of the dirt roads shown in Figure 6 were found in the herders' seasonal location, the effect of the herders' winter camp location on dirt road density was significant in relatively small areas.
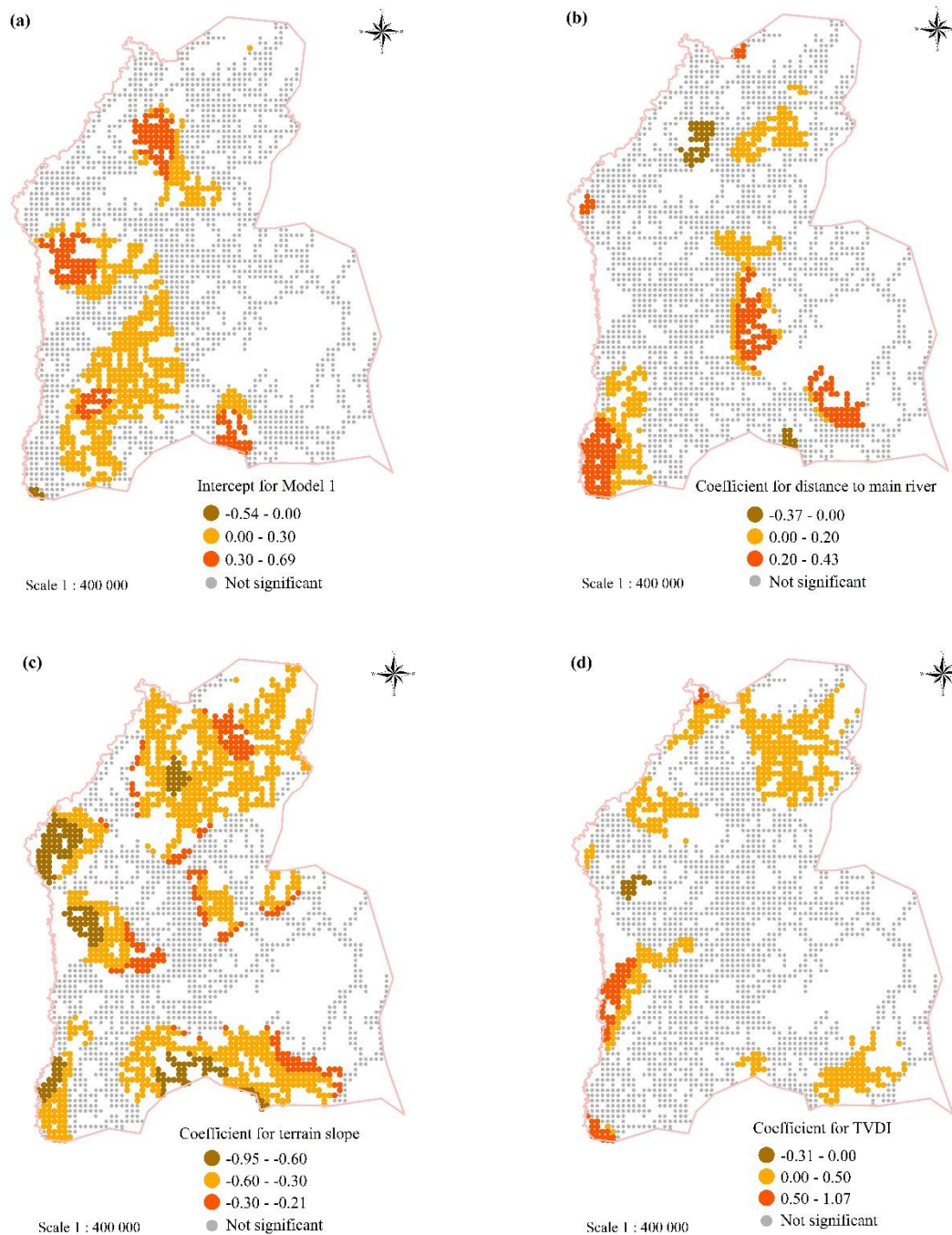
Figure 10. Spatial distribution of the Geographically Weighted Regression coefficients in Model 1; a) intercept, b) coefficient for *distance to main river*, c) coefficient for *terrain slope*, and d) coefficient for *Temperature – Vegetation Dryness Index*. Model 1 was computed with the Gaussian kernel weighting function with a fixed distance bandwidth of 4 km. Coefficient estimates are statistically significant at a 95% confidence level.
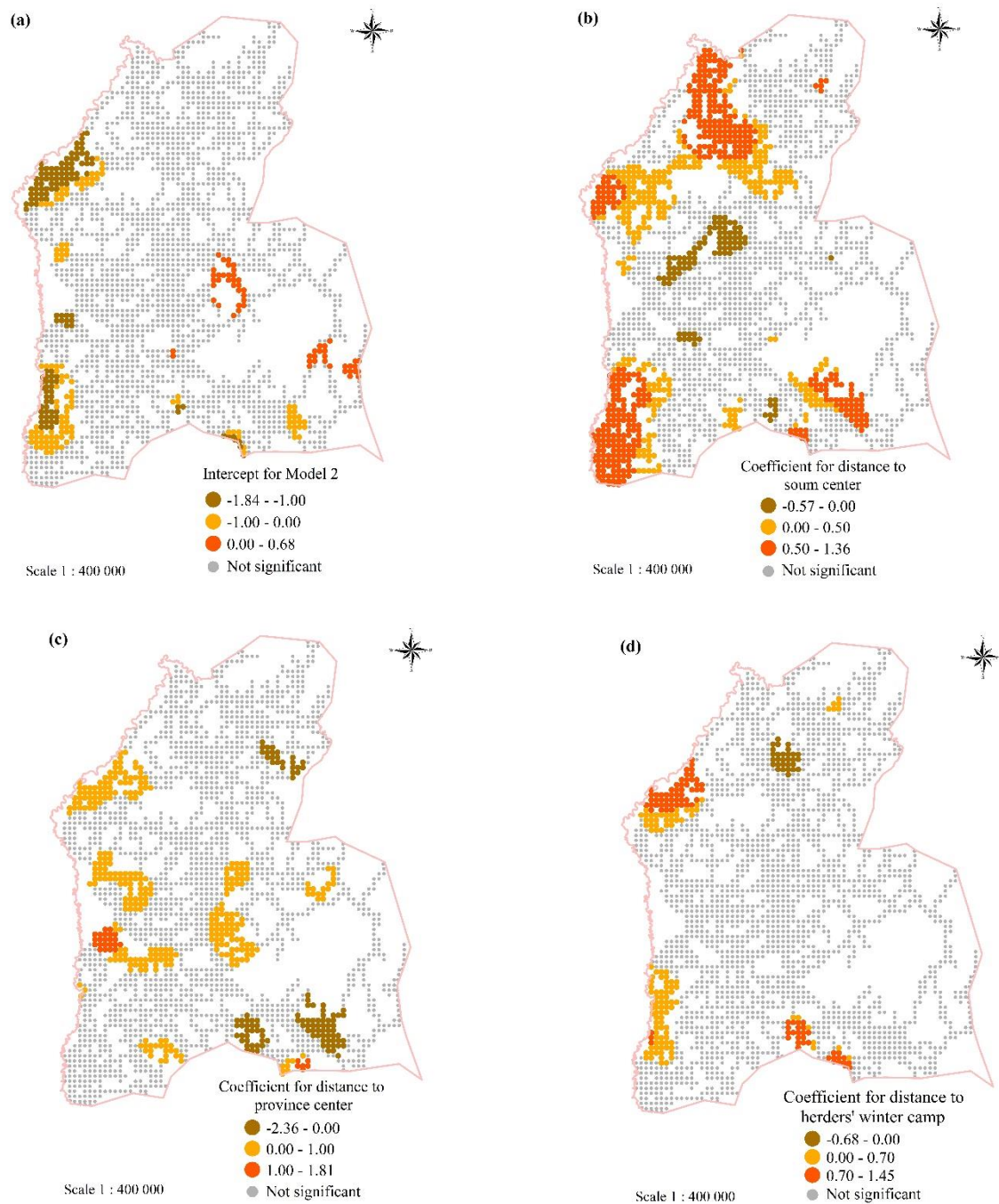
Figure 11. Spatial distribution of the Geographically Weighted Regression coefficients in Model 2: a) intercept, b) coefficient for *distance to soum* and *bag centre*, c) coefficient for *distance to province centre*, and d) coefficient for *distance to herders' winter camp*. The model was computed by the Gaussian kernel weighting function with a fixed distance bandwidth of 4 km. Coefficient estimates are statistically significant at a 95% confidence level.

(a)

(b)

(c)

Intercept for Model 3
● -0.89 - -0.40
● -0.40 - 0.00
● 0.00 - 0.45
● Not significant

Scale 1 : 400 000

Coefficient for distance
to soum center
● -0.45 - 0.00
● 0.00 - 0.50
● 0.50 - 1.10
● Not significant

Scale 1 : 400 000

Coefficient for TVDI
● -0.37 - 0.00
● 0.00 - 0.50
● 0.50 - 0.97
● Not significant

Scale 1 : 400 000

(d)

(e)

Coefficient for terrain slope
● -0.91 - -0.60
● -0.60 - -0.30
● -0.30 - -0.17
● Not significant

Scale 1 : 400 000

Coefficient for distance
to main river
● -0.59 - -0.30
● -0.30 - 0.00
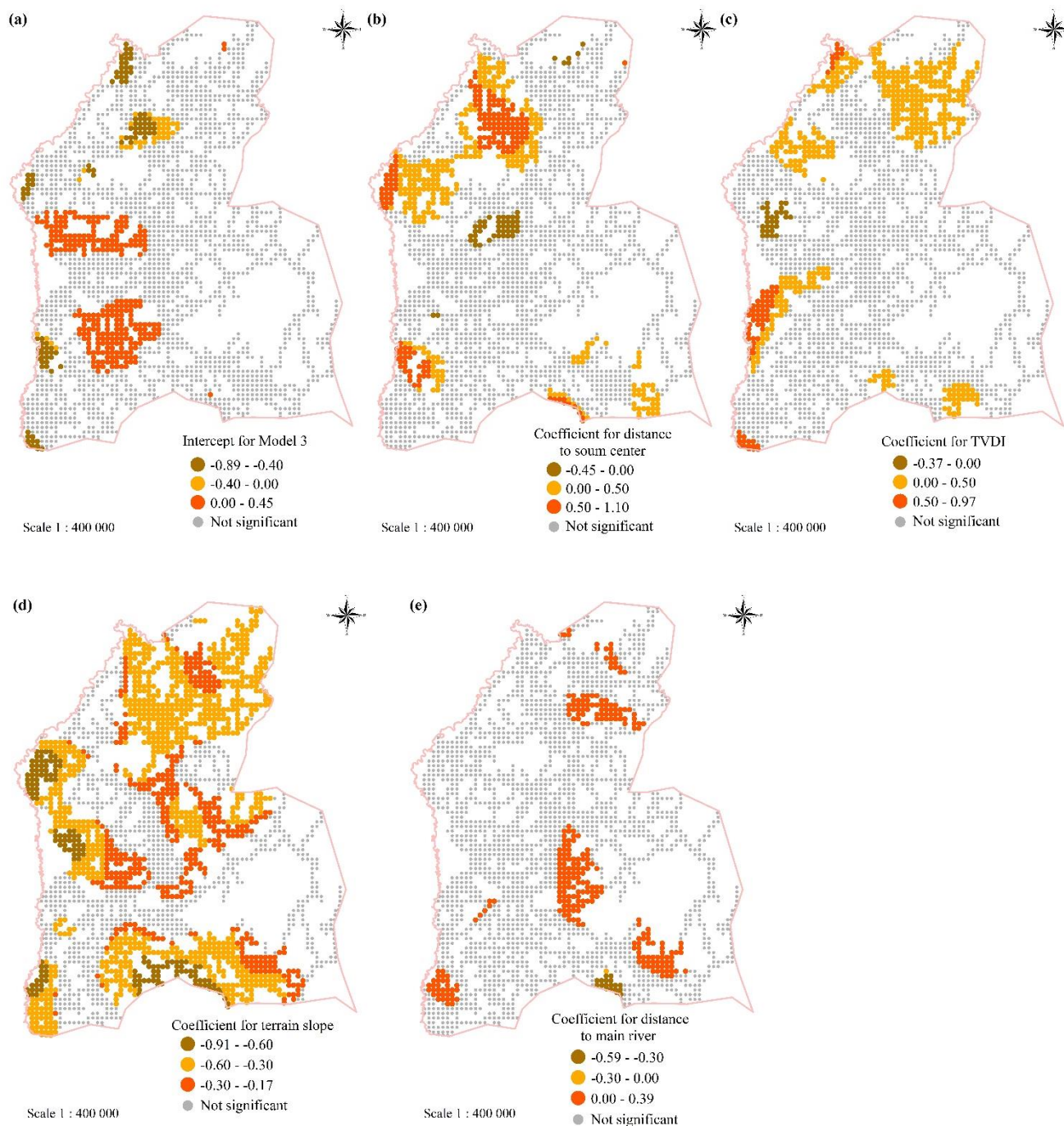● 0.00 - 0.39
● Not significant

Scale 1 : 400 000

Figure 12. Spatial distribution of the Geographically Weighted Regression coefficients in Model 3: a) intercept, b) coefficient for *distance to soum and bag centre*, c) coefficient for *Temperature – Vegetation Dryness Index*, coefficient for *terrain slope*, and d) coefficient for *distance to main river*. The model was computed by the Gaussian kernel weighting function with a fixed distance bandwidth of 4 km. Coefficient estimates are statistically significant at a 95% confidence level.

4. DISCUSSION

In order to get a better understanding of dirt road-related land degradation, this study carried out a geostatistical analysis of dirt road density and its influencing factors at the local level.

The study was carried out in the northern part of Mongolia and the total length of the dirt roads defined by this study was 2,998 km, which includes informal dirt roads not included in data from public authorities and is derived through satellite image analysis. A high concentration of dirt roads could be shown to have existed as clusters in several parts of the study area.

Results of the GWR models indicated that terrain slope plays a significant role in the dirt road density. As the terrain became steeper, the dirt road density decreased and vice versa. Proximity to soum centre and bag centre was a positive correlation with dirt road density. In other words, dirt road density tended to increase near the rural settlement centres. These findings were similar to research by Keshkamat et al. (2013) on the relationship between dirt road corridors and geographic factors at the national level. Moreover, they found the effect of soil moisture on dirt road corridor width. That effect (the wetter the soil, the wider corridor) was revealed in a small area.

In this study, the observed correlation between dirt road density and soil moisture (TVDI) was mainly positive. It was observed that dirt road density tends to increase when soil gets drier. A negative correlation (wetter soil and dirt road density expansion) was found in a relatively limited area. Wet soil might influence the increase of dirt road distribution in a particular condition.

Although GWR models showed relatively better results than the corresponding OLS models (see Table 8), there were some areas where the model was poorly fitted. Several limitations related to the nature of the geographic factors in this analysis could reduce the model fit and overall stability of models.

The main limitation was a multicollinearity. The research suggested that it is considered as multicollinearity if the variance inflation factor value is higher than 7.5 (Qiu et al. 2012) and high values (more than 0.7) in the Pearson's correlation coefficient matrix (Tang et al. 2020). As a result of the initial bivariate correlation screening, three insignificant associations could be observed between independent variables and the dependent variable and three multicollinearities (Pearson's correlation coefficient values more than 0.7) of the independent variables (see Table 6). After removing six variables from the analysis based on initial bivariate correlation screening, the remaining ten variables were tested by the variance inflation factor diagnose. The variance inflation factor values ranged from 1.23 to 2.26. Although all remaining variables met the thresholds suggested by Qiu et al. (2012) and Tang et al. (2020), there were found severe local multicollinearity of the independent variables (Fig. 13). Wheeler and Tiefelsdorf (2005) stated that such a collinearity situation can emerge in the GWR analysis even if it is not seen in global screening.

To reduce the local multicollinearity of the independent variables, a second bivariate correlation screening was carried out. The threshold criterium of Pearson's correlation coefficient value was set to 0.5 in the second bivariate correlation screening. After the second bivariate correlation screening, three variables were removed from the analysis. Even if multicollinearity of the independent variables was reduced by omitting some independent variables from the analysis, there was still collinearity in some independent variables. For example, the collinearity

problem could be observed in correlation between NDVI and distance to herders' winter camp (Fig 14).
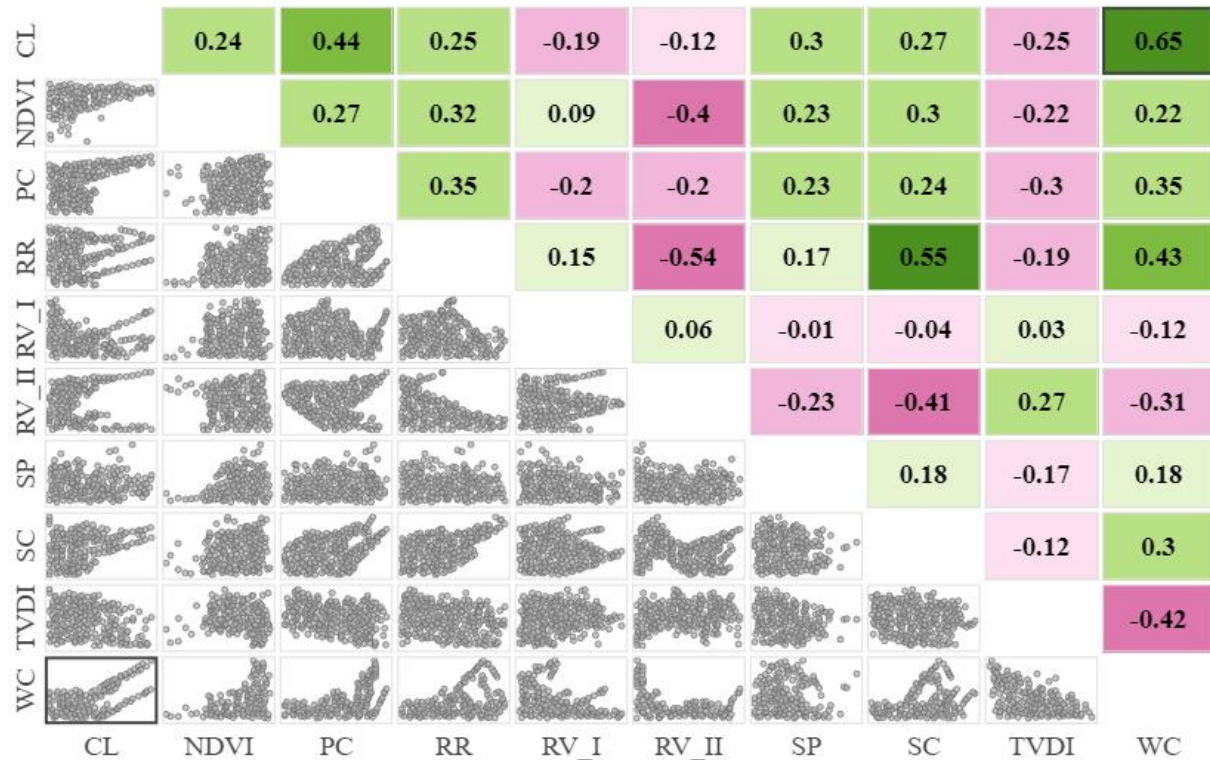

Figure 13. Bivariate scatter plots of pairs of independent variables. The Pearson's correlation coefficients were shown in the upper triangle. *CL* is *distance to cropland*, *NDVI* is *Normalized Difference Vegetation Index*, *PC* is *distance to province centre*, *RR* is *distance to railroad*, *RV_I* is *distance to main river*, *RV_II* is *distance to tributary river*, *SP* is *terrain slope*, *SC* is *distance to soum and bag centre*, *TVDI* is *Temperature – Vegetation Dryness Index*, and *WC* is *distance to herders' winter camp.* Scatter plot visualization and the Pearson's correlation coefficient were computed in ArcGIS Pro 2.9.

In reality, the livelihood of the herders depends highly on grassland quality. Herders tend to seek productive grasslands with dense and greener vegetation. This made it difficult to separate the herders' seasonal location factor from the vegetation factors in this analysis. Further research into adjusting the approach taken is needed to better address these special local characteristics.

Although multicollinearity of the independent variables was reduced by omitting nine independent variables, this resulted in other limitations with regards to misspecification. Issues caused by misspecification can happen when essential variables that have a significant role in the model calibration processes are not included in an analysis (Fotheringham & Sachdeva 2022).
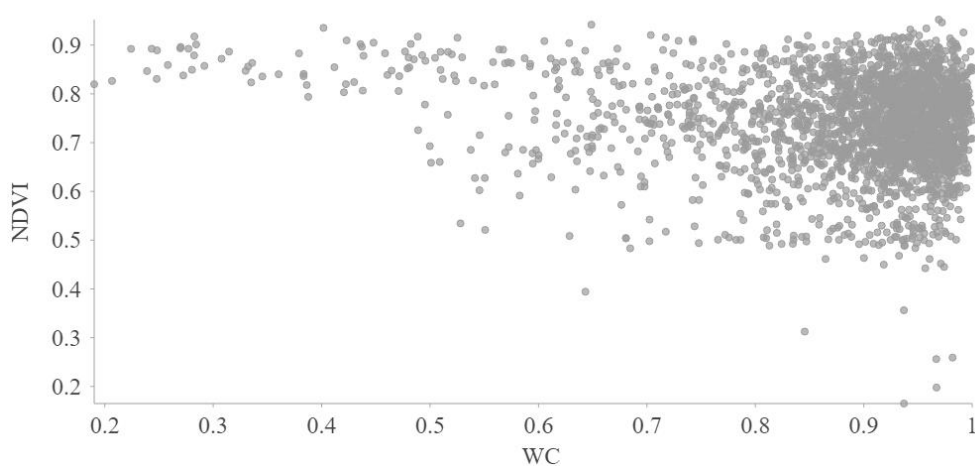
Figure 14. Bivariate scatter plot of the NDVI and distance to herders' winter camp location (WC).

As seen in Figure 8 (especially Fig. 8 b), there were some areas where all GWR models fit very poorly. Adjusted R^2 of these areas ranged from 0.04 to 0.19. These areas are mainly used for croplands in terms of land use. The variable representing cropland was omitted from the analysis due to severe local multicollinearity with some variables (distance to soum and bag centre, distance to province centre and distance to herders' winter camp). It remains a challenge to distinguish the influence of those factors on dirt road density expansion. All soums except the Darkhan (province centre) and the Sharyn Gol soum were originally established to develop agricultural production (crop and vegetable production). Since some crop fields are located very close to the soum centre, they share the same spatial characteristics. In addition to this complex situation, it was difficult to explain some of the dirt road outlier densities (3,455 m/km², 5,374 m/km², and 5,374 m/km²) found in between the Javkhlant soum centre and crop fields (see Fig. 3b, 3m, and Fig. 5). In other words, those outlier dirt road densities might be explained by a combination of circumstances, but not a single land use factor.

Moreover, in the intercept maps (Fig. 10a and 12a) that illustrated the influence of natural factors on dirt road density, those areas where the model fitted poorly could be easily detected. It might be related to the remaining selected variables as the adjusted variable screening processes did not have enough ability to explain the increase in dirt road density in those particular areas. In order to understand the relationship between dirt road density and independent variables in those areas, further studies to define the effect of omitted variables as well as the context effect on model fitting are needed.

Other signs of misspecification could be observed in maps of the spatial autocorrelation of the regression residuals (see Fig. 9). Statistically significant residual clusters (high or low) show that there is at least one essential independent variable absent from the model (Gao & Li 2011).

Results of Anselin Local Moran's I statistics indicated that there were statistically significant ($p < 0.05$) high clusters of dirt roads near the Sharyn Gol open pit coal mining and gold mining in the upper reach of the Sharyn Gol River valley (see Fig. 6). The highest dirt road density (5,965 m/km²) was observed in and near the Sharyn Gol open pit coal mining area. Results of the Pearson's product-moment correlation test showed a statistically insignificant association between dirt road density and distance to mining variables. The variable representing mining

35

was removed from the analysis based on insignificant correlation with the dependent variable. As a result, the effect of removing the mining variable was observed in all GWR models results. This effect was reflected by poor model fitting (adjusted R² was in the range 0.09-0.19) and High-High (i.e., underestimation) residual clusters in both OLS models and GWR models (see Fig. 9).

The sources of error in this study should be acknowledged and indicate where further research is needed to better adapt GWR approaches to local areas. The process of digitizing dirt roads itself could also affect the accuracy of the data and possible semi-automated processes using machine learning approaches might help to improve data accuracy and efficiency of this labour-intensive approach. Dirt roads were digitized from the TGSI maps by visual interpretation. The TGSI maps were generated from the Sentinel 2 Level 1C data obtained from ESA (2021). The TGSI map used by the base map was generated from Sentinel 2 Level 1C data recorded on 30 August 2021, and the additional TGSI maps that were used for validation were generated from the same satellite data recorded on 5 September 2020, 6 July 2021, 9 September 2021, and 22 May 2022. In order to accurately digitize dirt roads and reduce possible digitizing errors, two dirt road vector datasets from ALAMGC (2021) and Google Earth Pro were used as primary reference data. Although a dirt road digitizing process was validated through reference data and additional TGSI maps, errors are possible.

Moreover, some variables representing vegetation and soil moisture were generated from remote sensing data which was recorded on a particular date. This might be a source of error since this study did not consider seasonal variability in soil moisture and vegetation. This was partly due to time constraints, but also due to the pioneering character of this study with a focus on testing the overall approach first before making suggestions on how to operationalize it.

Furthermore, only dirt roads created by vehicles were considered in this study. The intensity of dirt road utilization and dynamic characteristics (newly created or abandoned) of the dirt roads were not considered and should be taken into account in future studies.

Finally future studies that focus on dirt road propagation modelling at the local level need to consider the potential multicollinearity and misspecification problems described above.

5. CONCLUSIONS

Mongolia is the 19th largest country in the world and 76.9% of the total land area has been affected by desertification (Baasandai 2020). Uncontrolled dirt roads are one of the main reasons for land degradation in Mongolia (Ochirbat 2013). Due to insufficient paved road supply, poor economic capacity to construct paved roads, and the vast territory, dirt roads play an essential role in the Mongolian transportation system. Detecting and registering temporary dirt roads in the official database is difficult due to the rapid increase of dirt roads in this large country (Dashpurev et al. 2020).

This study investigated dirt road distribution and the relationship between geographical factors and dirt road densities at the local level in the northern part of Mongolia. The study area covered 0.29% of the total land territory in Mongolia and the total length of dirt roads was 2,998 km. This means that approximately 7.6% of the currently available estimation of dirt roads in Mongolia were allocated to this 0.29% of the total land territory. The approach of analysing satellite images to identify dirt roads not accounted for in public datasets proved valuable to

gain a better understanding of the full dimension of dirt road distribution. The analysis was particularly important for understanding the local dimensions and characteristics that contribute to the informal part of dirt road expansion. The spatial autocorrelation analysis conducted in this study showed that a high concentration of dirt road clusters existed in several distinct parts of the study area. This finding indicates that the roads significantly contributed to land degradation in these areas.

It is necessary to take urgent management measures to mitigate the impact of dirt roads on the surrounding environment in areas that are highly affected by dirt roads. Establishing at least one main gravel road might, for example, help to mitigate the informal expansion of dirt roads.

The effect of geographical factors on an increase in dirt road density varied throughout the study area. Proximity to rural settlement centres, especially soum and bag centres, and terrain slopes have a significant influence on dirt road expansion. Due to several issues related to local multicollinearity and model misspecification, the results of this study were not sufficient to provide a more accurate conclusion regarding the main factors that can exacerbate an increase in dirt road expansion at the local level.

The GWR model might have some shortcomings in the study of dirt road density at the local level due to local multicollinearity. Special care also needs to be taken to diligently pre-process the data that goes into the model, which was not always possible in this study due to time constraints. Future studies could conduct a comparative analysis of the GWR modelling using the same variables but in different locations to confirm if the GWR model is suitable for predicting dirt road density expansion at the local level.

This study demonstrated that a geostatistical approach to analysing dirt road distribution and their natural and human environments merits further investigation. The scope of the project was too limited to derive more advanced results and demonstrated some limitations in some of the commonly used regression models. However, preliminary results from this study revealed some valuable findings that show the impact of dirt roads in relation to land degradation. Some promising aspects of the work are, however, already useful to inform policy making and regional planning efforts with regards to where most urgent interventions are needed and to which geographical factors contribute to dirt road expansion at a local level.

ACKNOWLEDGEMENTS

LITERATURE CITED

Al-Dousari AM, Alsaleh A, Ahmed M, Misak R, Al-Dousari N, Al-Shatti F, Elrawi M, William T (2019) Off-Road Vehicle Tracks and Grazing Points in Relation to Soil Compaction and Land Degradation. Earth Systems and Environment 3:471-482

ALAMGC [Agency for Land Administration and Management Geodesy and Cartography] (2021) Mongolian national unified land territory geodatabase. Ulaanbaatar

Assaeed AM, Al-Rowaily SL, El-Bana MI, Abood AAA, Dar BAM, Hegazy AK (2019) Impact of off-road vehicles on soil and vegetation in a desert rangeland in Saudi Arabia. Saudi Journal of Biological Sciences 26:1187-1193

Baasandai E (2020) Монгол орны цөлжилтийн зураг [Desertification map of Mongolia]. Pages 67-68. In: Ganjuur S (ed.) Desertification AtlasToonot Print, Ulaanbaatar. (in Mongolian)

Bazarragchaa A (2017) Бэлчээрийн газрын хариуцлагатай засаглалыг сайжруулах боломж [Opportunity to enhance pastureland's governance responsibility]. Pages 5-10. In: Erkhembayar E and Erdenechandmani J (eds.) Proceeding of the Research Conference on Land Communication 2017, Ulaanbaatar, Mongolia, 5 May 2017. Munkhiin Useg Press, Ulaanbaatar (in Mongolian)

Bravo-Linares C, Schuller P, Castillo A, Ovando-Fuentealba L, Muñoz-Arcos E, Alarcón O, de los Santos-Villalobos S et al. (2018) First use of a compound-specific stable isotope (CSSI) technique to trace sediment transport in upland forest catchments of Chile. Science of the Total Environment 618:1114-1124

Byambaa G, Murayama S (2012) Монгол орны шороон замын эвдрэл, доройтол ба нөхөн сэргээлтийн судалгаа [Study on dirt road erosion and natural vegetation recovery in Mongolia]. Geographical Issues of Mongolia 8:39-45 (in Mongolian)

Cao L, Wang Y, Liu C (2021) Study of unpaved road surface erosion based on terrestrial laser scanning. Catena 199:105091

Charlton M, Fotheringham S (2009) Geographically weighted regression white paper. Maynooth, Ireland. https://www.geos.ed.ac.uk/~gisteac/fspat/gwr/gwr_arcgis/GWR_WhitePaper.pdf (accessed 20 August 2022)

Chen VYJ, Deng WS, Yang TC, Matthews SA (2012) Geographically weighted quantile regression (GWQR): An application to U.S. mortality data. Geographical Analysis 44:134-150

Cho S, Lambert DM, Kim SG, Jung & S (2009) Extreme coefficients in geographically weighted regression and their effects on mapping. GIScience & Remote Sensing 46:273-288

Congedo L (2016) Semi-automatic classification plugin documentation. https://semiautomaticclassificationmanual-v5.readthedocs.io/en/latest/index.html# (accessed 16 August 2022)

Crisfield V, MacDonald S, Gould A (2012) Effects of recreational traffic on alpine plant communities in the northern Canadian Rockies. Arctic, Antarctic, and Alpine Research 44:277-287

Damdinsuren B, Herrick JE, Pyke DA, Bestelmeyer BT, Havstad KM (2008) Is rangeland health relevant to Mongolia? Rangelands 30:25-29

Daoud JI (2017) Multicollinearity and regression analysis. Journal of Physics: Conference Series 949:012009

Dashpurev B, Bendix J, Lehnert LW (2020) Monitoring oil exploitation infrastructure and dirt roads with object-based image analysis and random forest in the Eastern Mongolian Steppe. Remote Sensing 12:144

Dechingungaa D, Dangaa E, Baast O, Erdenedalai A, Tovuudorj R, Serd-Yanjiv N, et al. (2013) Монгол орны төв бүсийн физик газарзүй, нийгэм эдийн засгийн газарзүйн иж бүрэн тодорхойлолт [A comprehensive study on the description of physical geography and socio-economic geography of the central region of Mongolia]. Ulaanbaatar, Mongolia (in Mongolian)

Doljin D (2010) Монгол орны ландшафт экологийн зарим асуудлууд [Landscape-ecological issues in Mongolia]. 2nd edition. Munkhiin Useg Press, Ulaanbaatar (in Mongolian)

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. (2012) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27-46

Du S, Wang Q, Guo L (2014) Spatially varying relationships between land-cover change and driving factors at multiple sampling scales. Journal of Environmental Management 137:101-110

ESA [European Space Agency] (2015) Sentinel-2 User Handbook. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000 (accessed 25 July 2022)

ESA [European Space Agency] (2021) Sentinel-2 image courtesy of the U.S. Geological Survey. https://earthexplorer.usgs.gov/ (accessed 10 July 2022)

ESRI (n.d.) Exploratory regression (Spatial statistics). https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/exploratory-regression.htm (accessed 25 July 2022)

ESRI INC [Environmental Systems Research Institute] (2020) ArcGIS Ddesktop. Redlands, California

ESRI INC [Environmental Systems Research Institute] (2021) Arc GIS Pro. Redlands, California

Fotheringham AS, Sachdeva M (2022) On the importance of thinking locally for statistics and society. Spatial Statistics:100601

Fox J, Weisberg S (2019) An R companion to applied regression. 3rd edition. Sage Publications, Thousand Oaks, California

Gao J, Li S (2011) Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. Applied Geography 31:292-302

Geobotanic (2015) Дархан Уул аймгийн газар зохион байгуулалтын ерөнхий төлөвлөгөө [General land use planning of the Darkhan Uul province]. Ulaanbaatar, Mongolia. https://www.gazar.gov.mn/storage/reports/July2019/wiEW07triFOr3LUNpjs3.pdf (in Mongolian)

Guo B, Wang X, Pei L, Su Y, Zhang D, Wang Y (2021) Identifying the spatiotemporal dynamic of PM2.5 concentrations at multiple scales using geographically and temporally weighted regression model across China during 2015-2018. Science of the Total Environment 751:141765

Haldar S, Majumder A (2022) Changing nature of Land surface temperature and transformation of vegetation cover and water bodies in the 2nd largest urban agglomeration of West Bengal, Eastern India. Remote Sensing Applications: Society and Environment 27:100811

Han J, Dai H, Gu Z (2021) Sandstorms and desertification in Mongolia, an example of future climate events: a review. Environmental Chemistry Letters 19:4063-4073

Healey JF, Donoghue C (2021) Statistics: a tool for social research and data analysis. 11th Edition, Cengage Learning, Boston

Hogan JL, Brown CD (2021) Spatial extent and severity of all-terrain vehicles use on coastal sand dune vegetation. Applied Vegetation Science 24: e12549

Huang Y, Leung Y (2002) Analysing regional industrialisation in Jiangsu province using geographically weighted regression. Journal of Geographical systems 4:233-249

IRIMHE [Information and Research Institute of Meteorology Hydrology and Environment] (n.d.) Climate service system [Уур амьсгалын үйлчилгээний систем]. http://climate-service.mn/climateservice/index.php?menuitem=2&product=1 (accessed 30 May 2022) (in Mongolian)

Jackson SL (2015) Dusty roads and disconnections: Perceptions of dust from unpaved mining roads in Mongolia's South Gobi province. Geoforum 66:94-105

Kalkhan MA (2011) Spatial statistics : geospatial information modeling and thematic mapping. CRC Press, Taylor & Francis Group, Florida

Keshkamat SS, Tsendbazar NE, Zuidgeest MHP, Shiirev-Adiya S, van der Veen A, van Maarseveen MFAM (2013) Understanding transportation-caused rangeland damage in Mongolia. Journal of Environmental Management 114:433-444

Li SG, Tsujimura M, Sugimoto A, Davaa G, Sugita M (2006) Natural recovery of steppe vegetation on vehicle tracks in central Mongolia. Journal of Biosciences 31:85-93

Liu C, Liu J, Jiao Y, Tang Y, Reid KB (2019) Exploring spatial nonstationary environmental effects on Yellow Perch distribution in Lake Erie. Peer Journal 7:e7350

Lloyd CD (2010) Local models for spatial analysis. 2nd edition. CRC Press, Taylor & Francis Group, New York

Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2001) Geographic information systems and science. John Wiley & Sons, New York

Lüdecke D, Bartel A, Schwemmer C, Powell C, Djalovski A, Titz J (2022) sjPlot: Data visualization for statistics in social science. R package version 2.8.11

Mennis JL, Jordan L (2005) The distribution of environmental equity: exploring spatial nonstationarity in multivariate models of air toxic releases. Annals of the Association of American Geographers 95:249-268

Microsoft Corporation (n.d.) Microsoft Excel 365. Redmond, Washington

Ministry of Nature and Environment (1997) National plan of action to combat desertification in Mongolia. Ulaanbaatar. https://knowledge.unccd.int/sites/default/files/naps/mongolia-eng2000.pdf (in Mongolian)

Mondal B, Das ND, Dolui G (2015) Modeling spatial variation of explanatory factors of urban expansion of Kolkata: a geographically weighted regression approach. Modeling Earth Systems and Environment 1:29

NAGC [National Agency for Geodesy and Cartography of Mongolia] (1984) Topography map (scale 1:100,000). Ulaanbaatar, Mongolia. (in Mongolian)

National Statistics Office of Mongolia (n.d.) Key indicators of auto transport. https://1212.mn/tables.aspx?tbl_id=DT_NSO_1200_012V1&TRANSPORT_select_all=0&TRANSPORTSingleSelect=_7_7.1_7.2_7.3&YearQ_select_all=0&YearQSingleSelect=_202101&YearY_select_all=0&YearYSingleSelect=_2020&viewtype=table (accessed 28 May 2022 a)

National Statistics Office of Mongolia (n.d.) Resident population in Mongolia. https://1212.mn/tables.aspx?tbl_id=DT_NSO_0300_004V5&BAG_select_all=0&BAGSingleSelect=_34316_3431651_3431653_345_34501_34504_34507_34510&RESIDENT_select_all=0&RESIDENTSingleSelect=_1_12_11&YearY_select_all=0&YearYSingleSelect=_2019&viewtype=table (accessed 4 April 2022 b)

National Statistics Office of Mongolia (n.d.) Number of livestock. https://www.1212.mn/tables.aspx?tbl_id=DT_NSO_1001_021V1&BAG_select_all=0&BAGSingleSelect=_181_182_183_184_185_261_262_263_264_265_267_341_342_343_344_345_346_348_421_422_423_5&TYPE_OF_LIVESTOCK_select_all=0&TYPE_OF_LIVESTOCKSingleSelect=_01&YearY_select_ (accessed 4 April 2022 c)

Nosrati K, Collins AL (2019) Investigating the importance of recreational roads as a sediment source in a mountainous catchment using a fingerprinting procedure with different multivariate statistical techniques and a Bayesian un-mixing model. Journal of Hydrology 569:506-518

Nyamtseren M, Nasanbat E, Otgon I, Chuluunbat J (2019) Монгол Улсын газрын доройтлын суурь үнэлгээ [Base assessment of the land degradation in Mongolia]. Pages 357-366. In: Nyamdavaa G, Battogtokh D, Mendsaikhan B, Mandakh N, Javzan C, and Tsagaantsooj N (eds.) Natural condition and territorial location aspects influencing in socio-economic development, Proceeding of the 4th International Conference, Ulaanbaatar, Mongolia, 21 January 2019. Soyombo Printing, Ulaanbaatar (in Mongolian)

Ochirbat B (2013) Human impact and land degradation in Mongolia. Pages 265-282. In: Chen J, Wan S, Henebry G, Qi J, Gutman G, Sun G, and Kappas M (eds.) Dry land East Asia: land dynamics amid social and climate change, ecosystem science and application. Higher Education Press and Walter de Gruyter GmbH, Berlin

Otgonbayar M, Atzberger C, Chambers J, Amarsaikhan D, Böck S, Tsogtbayar J (2017) Land suitability evaluation for agricultural cropland in Mongolia using the spatial MCDM method and AHP based GIS. Journal of Geoscience and Environment Protection 5:238-263

Padgett PE, Meadows D, Eubanks E, Ryan WE (2008) Monitoring fugitive dust emissions from off-highway vehicles traveling on unpaved roads and trails using passive samplers. Environmental Monitoring and Assessment 144:93-103

Pimpler E (2017) Spatial analytics with ArcGIS. 1st edition. Packt Publishing, Birmingham, UK

Ploughe LW, Fraser LH (2022) Find new roads TM? A systematic review on the impacts of off-road vehicle activity on soil, vegetation, and wildlife. Frontiers in Ecology and Evolution 9:805707

Punzo G, Castellano R, Bruno E (2022) Using geographically weighted regressions to explore spatial heterogeneity of land use influencing factors in Campania (Southern Italy). Land Use Policy 112:105853

QGIS Development Team (2022) QGIS Geographic Information System. Open Source Geospatial Foundation Project

Qiu B, Zhong M, Zeng C, Tang Z, Chen C (2012) Effect of topography and accessibility on vegetation dynamic pattern in Mountain-Hill Region. Journal of Mountain Science 9:879-890

Ragchaadulam B, Sainbuyan B, Boldbaatar N, Boldbayar R (2020) Уул уурхайн ашиглалтаас үүдэлтэй авто замын өөрчлөлт (Өмнөговь аймгийн Цогтцэций сумын жишээн дээр) [Influence of mining operation on road and its adverse impacts: A case study of Tsogttsetsii soum, Umnugobi province]. Geographical-Geoecological Issues of Mongolia 41:462-466 (in Mongolian)

Ramos-Scharrón CE (2018) Land disturbance effects of roads in runoff and sediment production on dry-tropical settings. Geoderma 310:107-119

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, Boston

Sandholt I, Rasmussen K, Andersen J (2002) A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. Remote Sensing of Environment 79:213-224

Tang J, Gao F, Liu F, Han C, Lee J (2020) Spatial heterogeneity analysis of macro-level crashes using geographically weighted Poisson quantile regression. Accident Analysis and Prevention 148:105833

UNCCD [United Nations Convention to Combat Desertification] (2017) Global land outlook. 1st edition. UNCCD, Bonn

UNECE [United Nations Economic Commission for Europe] (2018) Enviromental performance reviews: Mongolia. UNECE, New York and Geneva.

USGS [United States Geological Survey] (2018) Shuttle radar topography mission 1 arc-second global. https://earthexplorer.usgs.gov/ (accessed 13 August 2022)

USGS [United States Geological Survey] (2019) Landsat 8 (L8) Data users handbook. https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/LSDS-1574_L8_Data_Users_Handbook-v5.0.pdf (accessed 17 August 2022)

USGS [United States Geological Survey] (2020) Landsat 8-9 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Collection 2 Level-1 multispectral data. https://earthexplorer.usgs.gov/ (accessed 13 August 2022)

Wachnicka J, Palikowska K, Kustra W, Kiec M (2021) Spatial differentiation of road safety in Europe based on NUTS-2 regions. Accident Analysis & Prevention 150:105849

Wang S, Wu J (2020) Spatial heterogeneity of the associations of economic and health care factors with infant mortality in China using geographically weighted regression and spatial clustering. Social Science and Medicine 263:113287

Weisent J, Rohrbach B, Dunn JR, Odoi A (2012) Socioeconomic determinants of geographic disparities in campylobacteriosis risk: a comparison of global and local modeling approaches. International Journal of Health Geographics 11:45

Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems 7:161-187

Xiao J, Shen Y, Tateishi R, Bayaer W (2007) Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. International Journal of Remote Sensing 27:2411-2422

Yadambaatar B, Sandag B (2010) Хээрийн бүсийн бэлчээрийн экологийн зарим асуудал [Ecological issues of grassland in steppe zone]. Khukh Sudar printing, Ulaanbaatar (in Mongolian)

Yang SH, Liu F, Song XD, Lu YY, Li DC, Zhao YG, Zhang GL (2019) Mapping topsoil electrical conductivity by a mixed geographically weighted regression kriging: a case study in the Heihe River Basin, northwest China. Ecological Indicators 102:252-264

Zamba B, Dulam J, Chung Y (2006) Dust storms are an indication of an unhealthy environment in East Asia. Environmental Monitoring and Assessment 114:447-460

Zhang D, Jia Q, Wang P, Zhang J, Hou X, Li X, Li W (2020) Analysis of spatial variability in factors contributing to vegetation restoration in Yan'an, China. Ecological Indicators 113:106278

Zhang Y, Zhao Y, Liu B, Wang Z, Zhang S (2019) Rill and gully erosion on unpaved roads under heavy rainfall in agricultural watersheds on China's Loess Plateau. Agriculture, Ecosystems and Environment 284:106580

Zhao Z, Gao J, Wang Y, Liu J, Li S (2015) Exploring spatially variable relationships between NDVI and climatic factors in a transition zone using geographically weighted regression. Theoretical and Applied Climatology 120:507-519