



The Application of a Data-Driven Method for Spatial Analysis and Prediction of Gully Erosion Susceptibility

3

Didams Gideon and Firuza Begham Mustafa

Abstract

Gully erosion (GE) is a major environmental problem that causes land degradation worldwide especially in arid and semi-arid regions. A crucial step to reduce and possibly reclaim areas degraded by GE is to predict potential susceptible areas using highly accurate predictive methods. This chapter demonstrates the suitability of a data-driven model (logistic regression) to this aim. The model was applied in South Gombe State, Nigeria, where land development is restricted by intense GE because of the semi-arid climate and physiographic conditions. A GE inventory was prepared from a total of 260 gully and non-gully locations compiled from interpretation of Google Earth images and field investigations. Besides, local environmental conditions and a 20 m DEM allowed the selection of soil texture, geology, land use, rainfall, and some topographical factors influencing GE susceptibility. Subsequently, the inventory data was randomly split into two datasets; 182% or 70%

training, and 78% or 30% validation, while influencing factors independence was assessed via multicollinearity scrutiny. Results of forward stepwise regression for the relationship between GE and selected factors indicated that distance from road is key to gully formation. After running the logit function, the resultant susceptibility map revealed that 3.1% of the study area was relatively safe, 50.1% less, 23.3% moderate, 19.2% high, and 4.3% extremely susceptible. Validation assessment using area under receiver operating characteristic curve provided 92.3% prediction accuracy. This study further confirmed logistic regression as an excellent and accurate data-driven method for spatial analysis and prediction of GE susceptibility. The method can be applied elsewhere with similar physiographic characteristics.

Keywords

Gully erosion susceptibility · Data-driven method · Spatial prediction · Logistic regression · Gombe State

D. Gideon
Department of Geography, University Malaya, Kuala Lumpur, Malaysia
e-mail: gdamsviva@gsu.edu.ng

F. B. Mustafa (✉)
Department of Geography, Gombe State University, Tudun-Wada, Gombe, Gombe State, Nigeria
e-mail: firuz@um.edu.my

3.1 Introduction

Gully erosion (GE) is the “erosion process whereby runoff water accumulates and often recurs in narrow channels and, over short periods, removes the soil from this narrow area to

considerable depths” (Poesen et al. 2003). The channel formed from this erosion process is generally classified as a “gully”. Because it is a geographically widespread phenomenon, it is also known by a variety of native names like Ravine, Uvrag, Wadi, Nuallah or ‘Cho, Hakisan Gaung, Carcava or Arroyo, Donga, and Kwari (in Hausa language) in France, Russia, Arabic, India, Malaysia (Malay), Spain, South Africa, and Northern Nigeria, respectively, (Castillo and Gómez 2016; Maria and Nicolae 2017).

Gullies are commonly classified as “permanent” and “ephemeral” gullies. A permanent gully is a wide and deep channel eroded by concentrated flow removing the upland soil and parent materials, which cannot be eliminated through conventional tillage operations. Whereas ephemeral gullies are formed by a concentrated overland flow which can be remediated by conventional tillage operations (Garosi et al., 2019). Ephemeral gullies are smaller than permanent gullies but larger than rill systems. Unlike rills, ephemeral gullies are formed in the same location each season.

GE is a threshold phenomenon that is controlled by numerous factors. Thus, gullies occur only after a threshold of runoff erosivity and soil erodibility has been crossed. In addition to rainfall, runoff erosive power depends on topography that controls the discharge, concentration, and velocity of overland flow (Conoscenti et al., 2013). Morphology, density, and development of gullies in a particular area are also substantially regulated by the underlying bedrock (Poesen 2011). The incidence of a gully is also governed by soil resistance, which is dictated by soil properties such as texture, bulk density, moisture conditions, and organic matter content (Poesen et al. 2003). GE is also related to the type and stage of crop production, as well as tillage direction and conservation practices (Li et al. 2016). Also, several studies have reported

triggering of gullies or increasing of GE rates as being caused by land use changes, intensification of farming activities, and overgrazing (Conoscenti and Rotigliano 2020).

Once formed, gullies become major sediment sources and often cause environmental problems within their reach (on-site effects) and downstream (off-site effects). Generally, GE results in different consequences: (i) significant land degradation and loss of productive capacity, (ii) high sediment yields and sediment discharge, which can transport both nutrients and pollutants, and (iii) sedimentation of reservoirs (reducing the water capacity of the reservoirs) and damage to the infrastructure and transport routes. Ecologically, GE can cause associated ecological problems such as eutrophication and acceleration of desertification processes. Therefore, the prediction of gully erosion susceptibility is the first and most important step in averting the undesired effects of gullies and achieving sustainable development (Rahmati et al. 2017b).

According to Domínguez-Cuesta (2013), susceptibility reflects the condition of becoming weak or easily influenced. In natural hazards terms, susceptibility is related to the spatial dimensions of hazards. It refers to the tendency of a region to experience the consequences of certain dangerous activity (e.g., floods, earthquakes, erosion, etc.) without considering either the moment of occurrence or possible casualties and financial impacts. Thus, GES may be defined as the probability of spatial occurrence of GE based on the relationships between the distribution of gullies in the past and factors that influenced their occurrence.

The prediction of GES involves several qualitative (Knowledge-driven) and quantitative (data-driven) methods. The knowledge-driven methods depend on expert opinions. The most common types of these methods simply use GE inventories to characterize sites of similar geographical properties that are susceptible to GE. Certain knowledge-driven approaches, however, incorporate the idea of ranking and weighting and may evolve to be semi data-driven in nature.

Examples are the use of the analytic hierarchy process (AHP) of Saaty and Vargas (2012) by Arabameri et al. (2019b) and weighted linear combination (WLC) by (Sujatha and Sridhar 2019). AHP involves building a hierarchy of decision elements (factors) and then making comparisons between possible pairs in a matrix to give a weight for each element and also a consistency ratio. It is sustained by three principles: decomposition, comparative judgment, and synthesis of priorities. WLC is a concept to combine maps of GE influencing factors by assigning a standardized score (primary-level weight) to each class of a particular parameter and a factor weight (secondary-level weight) to the parameters themselves. Being partly subjective, the results of these approaches vary depending on the knowledge of experts (Arabameri et al. 2019c).

Data-driven approaches rely on statistical description of the relationship between GE and influencing factors. They statistically assess the combination of the influencing factors that are more closely related to the spatial distribution of existing gullies. Therefore, according to the prevailing terrain conditions (e.g., slope angle, slope shape, lithology, land cover) it is possible to quantitatively predict the likelihood of future gully occurrence even in non-gully affected areas. In this process, data from past and present gullies are used to evaluate the relative importance of the influencing factors and respective classes. Data-driven GES models are sustained by three major assumptions: (i) gullies can be recognized, classified, and mapped; (ii) GE influencing factors can be identified, registered, and used to build GE predictive models; and (iii) future GE occurrence can be spatially predicted (Arabameri et al. 2019c; Zêzere et al. 2017). In other words, following the Uniformitarianism principle, the past and the present are considered keys to the future, hence it is assumed that future gullies are more likely to occur under the same geologic and geomorphologic conditions that led to past gullies.

Data-driven methods can be grouped in bivariate statistical analysis (BSA) and multivariate statistical analysis (MSA) (Lucà et al.

2011). The BSA compares independently each influencing factor with the GE distribution. Weights of the GE influencing factors are assigned based on gully density using different methods such as information value, weights of evidence, among others. Bivariate statistical models do not consider the interdependence of predictive variables, and this is a major drawback of the method (Arabameri et al. 2019c; Zêzere et al. 2017). The MSA evaluates the combined relationship between the dependent variable and a set of independent variables.

Numerous MSA methods exist, but those commonly used to predict GES include linear and quadratic discriminant analyses and logistic regression. Linear and quadratic discriminant analyses have been used by Arabameri and Pourghasemi (2019) to predict GES in Iran. The method was also reported to be significant to define GES classes in the Pathro River Basin of India (Gayen et al. 2019). Logistic regression has been applied for susceptibility mapping by various researchers including Lucà et al. (2011), Conoscenti et al. (2014), Reza and Ronak (2015), Arabameri et al. (2018), Razavi-Termeh et al. (2020), and Arabameri et al. (2020c). However, the logistic regression model is the most widely used among the MSA approach. Compared with other MSA methods, its independent variables do not need to be linearly related or normally distributed, and can be categorical, continuous, or their mixture. In addition to this, the logistic regression model can effectively combine with other statistical analysis methods and/or knowledge-driven methods to improve GES prediction. Hence, the main objective of this study is to extend the application of logistic regression as a technique within data-driven methods. The method was applied in South Gombe State, Nigeria, where land development is restricted by intense GE because of semi-arid climate and physiographic conditions. Also, a thorough literature review to date shows that despite the high sensitivity of this area to GE, no comprehensive studies have been carried out to identify areas that are particularly susceptible to GE at all.

3.2 Materials and Methodology

3.2.1 The Case Study—Southe Gombe State (SGS), Nigeria

SGS occupies about 8023 km² and is located within Gombe State, Nigeria. It is bounded by latitude 9°30' and 10°20' N and longitude 10°40' and 11°50' E (Fig. 3.1). SGS is characterized by a semi-arid type of climate marked by two distinct seasons: a rainy season from April to October and a dry season from November to March. The average annual temperature is 27 °C, while the average annual rainfall is about 835 mm with much of it falling between June and September. The rains come in the form of intensive, violent showers of short duration, especially at the beginning and end of the rainy season. The early

rains of April which come just after the dry season are very effective in the gullying process. The low relative humidity, characteristic of the dry season, leaves the surface of the soil dry and cracked at various points. These cracks and other human features like footpaths, are rapidly exploited by the runoff from storms of the early rains and greatly favor the inception and subsequent evolution of gullies. Rain in the area may fall continuously for two, three, or more hours—although most of it comes during the first 40 min of the period of fall. They are the type of rainfall that causes so much damage in a relatively short time, especially in places where the soil is bare or is partially covered by vegetation which is incapable of protecting it from the erosive impact of the rain.

The geology of the area comprises of the crystalline basement and Cretaceous sedimentary Formations (Ikusemoran et al., 2018) (Table 3.1).

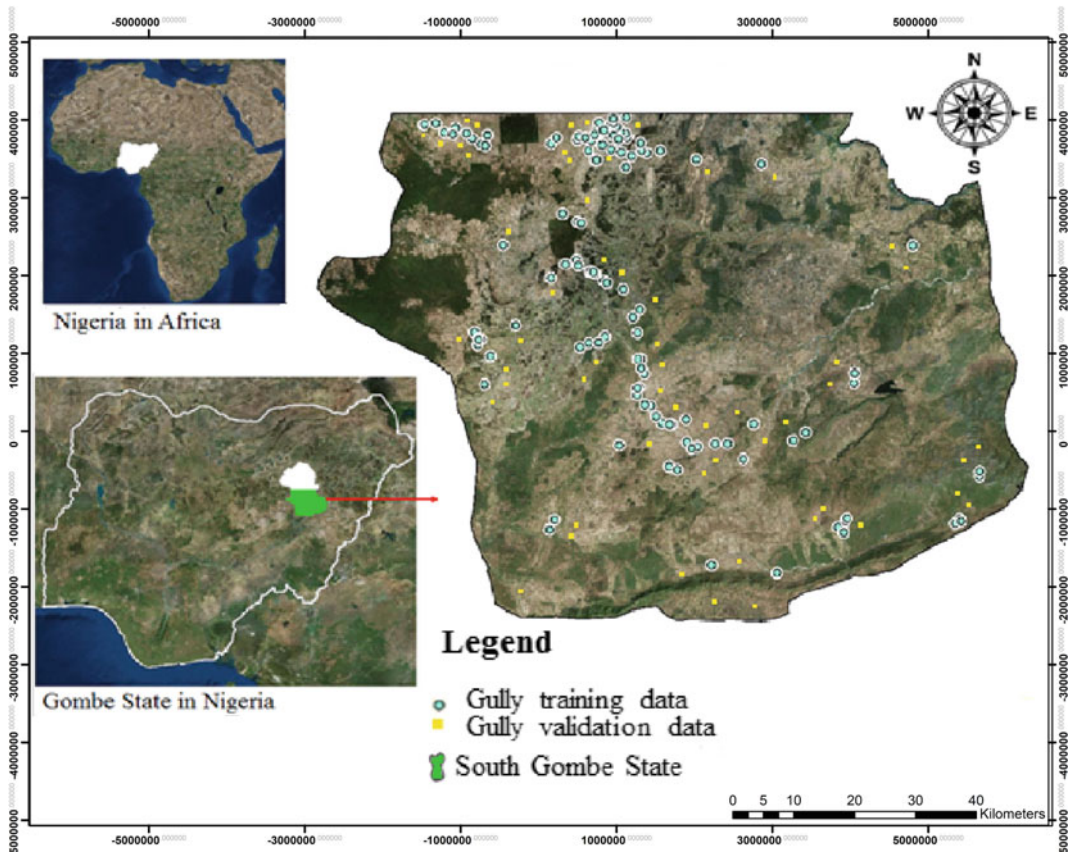


Fig. 3.1 Location of the study area and gully landforms distribution

Table 3.1 Geologic formations, lithological characteristics, and age of rock materials

Geologic formation	Lithologic units	Age
Alluvium	Alluvium	Holocene
Basalt	Basalt	Cretaceous
Kerikeri	Sandstone, shale, and clay	Palaeocene
Gombe	Sandstone, siltstone, shale, coal, and ironstone	Maastrichtian
Pindiga	Shale, limestone, and sandstone	Turonian
Yolde	Shale, limestone, and sandstone	Cenomanian
Bima	Sandstone, siltstone, and shale	Albian
Basement Complex	Porphyritic granite/coarse porphyritic biotite and biotite hornblende granite	Pre Cambrian
	Coarse, porphyritic hornblende granite	
	Undifferentiated granite, migmatite and granite gneiss	

The soils correlate with the underlying geology, and are mostly sandy, low in organic matter, and are characterized by low water-holding capacity. The only exception to this observation is the wetland (fadama) soils that are fine-textured with higher organic matter content and relatively higher water-holding capacity (Ikusemoran et al., 2016). Geomorphologically, the SGS is characterized by complex terrain and dominated by rugged hills of granite and sandstone, volcanic plugs, a sedimentary and volcanic plateau, and low, swampy plains; with elevation ranging from 200 m (a.s.l.) in the central parts to about 1170 m (a.s.l.) in the southern section. Land use in the area is characterized by livestock grazing, rain-fed farming, and dry season fadama (market gardening) cultivation.

3.2.2 Methodology

3.2.2.1 Spatial Data Used for the Study

The data used in this study was collected by integrating field data, remote sensing (RS) data, and geographic information system (GIS) applications. GIS data processing and computing can produce GE maps with low costs and acceptable accuracy. Also, GIS allows fast and easy representation and analysis of spatial data and can generally incorporate information layers from diverse sources (Arabameri et al. 2020d). Thus,

the basic maps used in this study were satellite imageries archived by Google Earth, Landsat 8 images (spatial resolution 30 m), soil and geological maps (scale 1:100,000), and a Digital Elevation Model (DEM) with a spatial resolution of 20 m. Table 3.2 presents a summary of the data types used in the present study.

3.2.2.2 Spatial Analytical Process

Figure 3.2 displays the flowchart of the approach followed in this study and consists of the following steps: Spatial data collection which involves identification and extraction of permanent GE features (gully inventory data) and assembly of GEIFs; spatial data analysis comprising independence scrutiny (multicollinearity) among GEIFs, and random partitioning of the gully inventory data; susceptibility spatial prediction which involves a determination of the spatial relationship between GE and GEIFs, and estimation of the probability of gully occurrence and generation of GE susceptibility map; and Validation of results.

3.2.2.3 Spatial Data Collection

As described above, the assumption of susceptibility to GE is that past and/or present gullies are important for understanding the occurrence of future gullies, therefore, the acquisition and construction of GE inventory is considered an important step in geomorphological analyses and

Table 3.2 Summary of data used for the study

Data type	Sources	Uses
Gully inventory	Field surveys, Google Earth imagery, and NEWMAP	For identifying actual gully erosion locations in the study area
Rainfall	Gombe State University weather station, and the Nigeria Meteorological Agency (NiMet)	For classifying the amount of rainfall of the study area
Soil	Federal Department of Agricultural Land Resources (FDALR) Abuja, Nigeria	The base map for extraction of soil texture map of the study area
Geology	Nigeria Geological Survey Agency (NGSA)	The base map for extraction of the Geology map of SGS
Land use	Landsat 8 images with 30 m resolution—United State Geographical Survey (USGS)	For classifying and mapping land uses of the study area
Digital elevation model (DEM)	Developed from topographic maps (1:50,000) of the study area	For extraction of topographical factors related to the gully erosion formation in the study area

stochastic modeling like the present study. In this study, gully landforms (Fig. 3.3) were identified through the interpretation of Google Earth image taken in 2018, and comprehensive field surveys conducted in early 2019 using a global positioning system (GPS) to record coordinates of gullies. The gullies were mapped as polygons but later converted to points by considering the locations of the head-cut portion of each gully. An equal number of non-gully point locations were randomly selected and later combined with the gully locations.

As previously mentioned, GES is regulated by a variety of factors. However, it is difficult to use all factors at once (Arabameri et al. 2019a, b, c) and thus it is important to pick key factors that could theoretically affect GE in a specific area. Although, there is no standardized guiding theory for the selection of factors (Arabameri et al. 2018), comprehensive literature review, local environmental conditions in the study area, and multicollinearity checks were utilized to select 14 factors: rainfall, aspect, slope angle (SA), length of slope (LS), elevation, plan curvature (PC), topographic wetness index (TWI), stream power index (SPI), drainage density (DD), distance from stream (DS), distance from road (DR), geology, soil texture (ST), and land use (LU). After selecting and analysis of the GEIFs, it is essential to classify their numerical values for effective modeling, and when

representing them on a map so that, visually, they can be as clearly understood as possible. For this reason, existing GISs are equipped with several methods for automatically performing the classification of susceptibility values. The commonly used ones include the quantile, equal interval, Jenks natural breaks, and geometrical interval classifications (Osaragi 2019). Those suitable for data used in this study have been applied.

3.2.2.4 Spatial Data Analysis

Partitioning of GE Inventory Data

Before its implementation, the logistic regression model must be trained and tested with two separate samples of data. But practice in early research on the implementation of logistic regression showed that for the percentage of training and testing samples, there is no clear rule of thumb. Nevertheless, it is suggested that the percentage for testing should be inversely proportional to the square root of the number of free adjustable parameters (Abdulkadir et al. 2020). Consequently, by using the unsupervised filtering procedures to avoid replications, the GE inventory data was divided into 70% training and 30% testing samples. This is consistent with the percentage of samples considered in related studies (Abdulkadir et al. 2020; Arabameri et al. 2020a; Rahmati et al. 2017a).

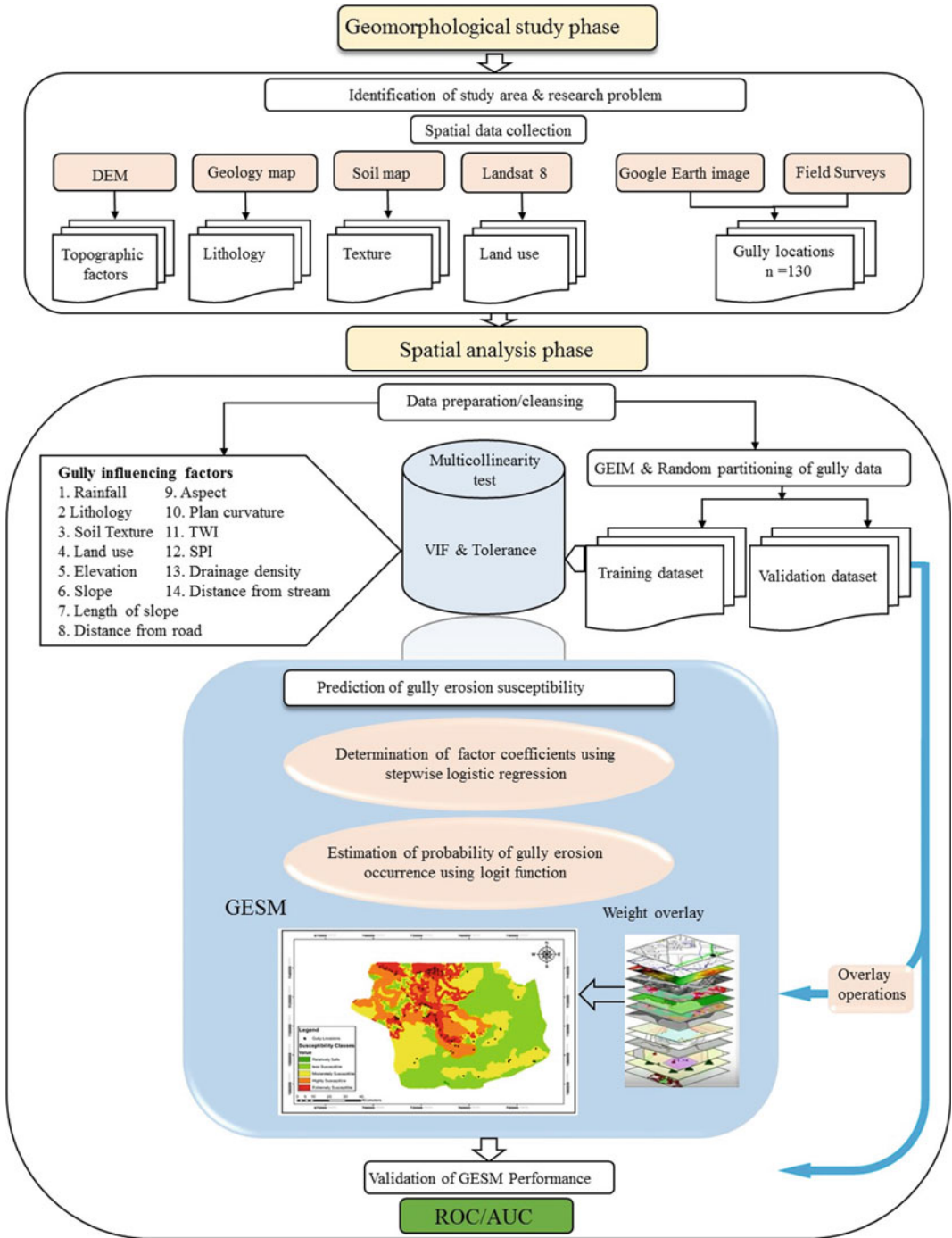


Fig. 3.2 Flow chart of data and procedures followed for the study

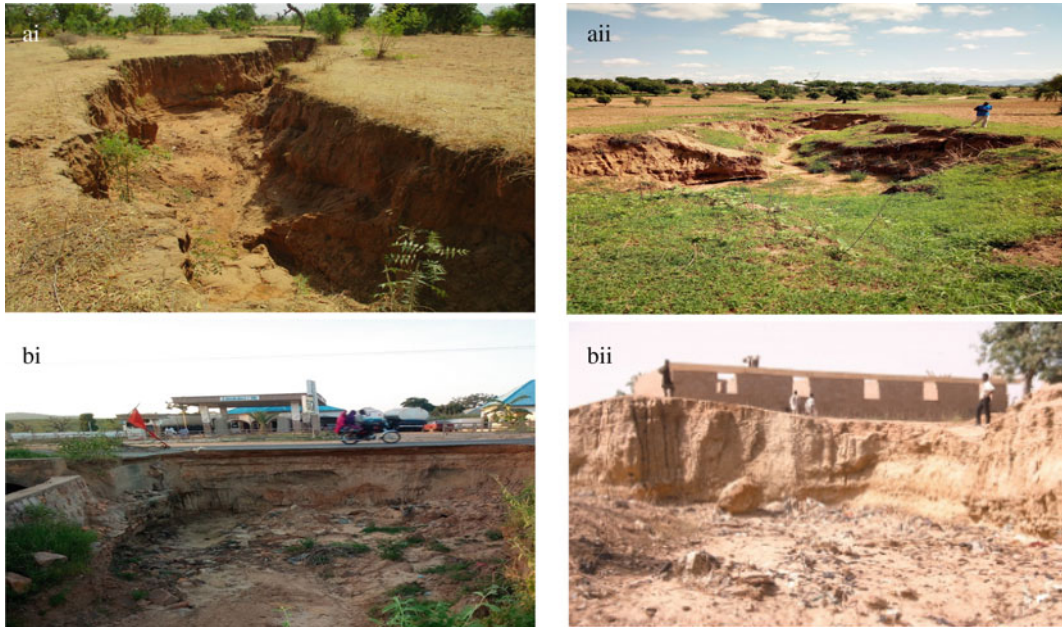


Fig. 3.3 Example of some identified gullies and their effects on (ai & ii) agricultural land (bi & ii), and on infrastructure in the study area

Extraction of GEIFs

Table 3.3 presents the methods of extraction and classification of GEIFs selected for the study. A Digital Elevation Model (DEM) with a spatial resolution of 20 m was produced from 1:50,000-scale topographic contour maps, and used for extraction of topographical factors: elevation, SA, LS, SPI, TWI, DD, DS, DR, aspect, and PC.

Elevation plays a significant role in determining vegetation cover type and to a large extent, precipitation characteristics (Gómez-Gutiérrez et al. 2015). Consequently, GE may occur in different elevations, depending on the initiating and developing mechanisms. SA plays a significant role in the dynamics of the processes controlling landscape development; it primarily affects surface runoff, DD, soil erosion, etc., (van der Meij et al. 2017). According to Conforti et al. (2010), steep SA facilitates high runoff velocity and consequent initiation of rill and gully. Hence, the SA is crucial for the prediction of GES. LS is the amalgamation of slope length (L) and steepness (S) of which L controls sediment detachment and generation, and S controls the movement of these sediments in response to

intense rainfall and related runoff (Pradeep et al., 2014). It is a key variable used in GES assessment. Theoretically, areas with high LS values are prone to GE activity (Conoscenti et al. 2014; Lucà et al. 2011).

Stream power is an important factor considered in the assessment of GES. It is the potential for flowing water to perform geomorphic work and is used to measure the erosive power of water flow based on the assumption that discharge is proportionate to the catchment area. The index SPI is one of the main factors controlling slope erosion processes since the erosive power of running water directly influences slope toe erosion and river incision (Conforti et al. 2010). It is also indicative of the potential energy available to entrain sediment so that areas with high SPI have a great potential for erosion.

TWI relates local topographic slope to the upslope contributing area at any given location within a watershed. It is commonly employed as a proxy for the potential for surface and subsurface water accumulation due to runoff and lateral transmissivity (Raduła et al., 2018). Since the dynamics, location, and size of saturated source

Table 3.3 GEIFs and methods used for their extraction and classification

Factor	Extraction method/source	Classification method	Classes
Rainfall	Ordinary kriging interpolation	Natural break	< 950, 951–1000, 1001–1050, and >1050 mm (Fig. 4a)
Geology	Digitized from NGSA map (1:50,000)	Geologic units	Alluvium, Basalt flow, Bima, Gombe, Kerikeri, Basement, Pindiga, and Yolde (Fig. 4l)
ST	Digitized from soil characteristics map prepared by FDALR	Supervised classification	silt clay loam, silt clay, sandy, sandy loam, sandy clay loam, and loam sand (Fig. 4m)
LU	Extracted from Landsat 8 imagery	Supervised classification	bare surface, farmland, forest, built-up area, shrubland, waterbody, wetland, and woodland (Fig. 4n)
Elevation	Extracted from DEM using ArcGIS software with spatial analyst tool	Natural break	<300, 300–500, 501–700, and >700 m (Fig. 4b)
PC	Extracted from DEM	Natural break	flat, Concave, and Convex (Fig. 4k)
SA	Extracted from DEM	Natural break	<5°, 5–10°, 11–30°, 31–60°, and >60° (Fig. 4c)
Aspect	Extracted from DEM	Equal interval	flat, north, northeast, east, southeast, south, southwest, west, and northwest (Fig. 4j)
LS	Extracted from DEM using Moore and Burch, (1986) formula $\left(\frac{faX_{cellsize}}{22.13}\right) \left(\frac{faX_{cellsize}}{22.13}\right) 0.4 \times \left(\frac{\sin\alpha}{0.0896} \frac{\sin\alpha}{0.0896}\right)^{1.3} \dots$ (1) fa = flow accumulation; α = slope angle	Natural break	<3, 3–5, 6–10, 11–50, and >50 m (Fig. 4d)
SPI	Extracted from DEM using $AS \times \tan\alpha \dots$ (2) AS = catchment area; α = slope angle	Natural break	<30, 30–50, 51–100, 101–200 and >200 (Fig. 4e)
TWI	Extracted from DEM using Moore et al. (1991) formula: $\ln \left(\frac{AS}{\tan\alpha} \frac{AS}{\tan\alpha}\right) \dots$ (3) AS = cumulative upslope area; α is slope angle	Natural break	<10, 10–12, 13–15, 16–18, and >18 (Fig. 4f)
DD	$L/A \dots$ (4) L = Total length of streams A = size of catchment (Line Density Tool in ArcGIS)	Natural break	<0.5, 0.5–1.0, 1.1–1.5, and >1.5 km/km ² (Fig. 4g)
DS	Euclidean distance tools in ArcGIS	Natural break	<50, 51–100, 101–300, 301–500 and >500 m (Fig. 4h)
DR	Euclidean distance and line density tools in ArcGIS	Natural break	<500, 501–1000, 1001–3000, 3001–5000 and >5000 m (Fig. 4i)

areas are controlled by subsurface hydraulic characteristics and watershed topography, TWI is frequently used to model GES. Technically, areas with larger upslope drainage areas and shallower slopes will produce larger TWI values, indicating a higher propensity for GE.

DD, also known as stream density, is the ratio of the total length of streams or channels in kilometer (km) to the size in km² of the area being studied. The measurement of DD is a useful numerical measure of landscape dissection and runoff potential. Hypothetically, a high DD

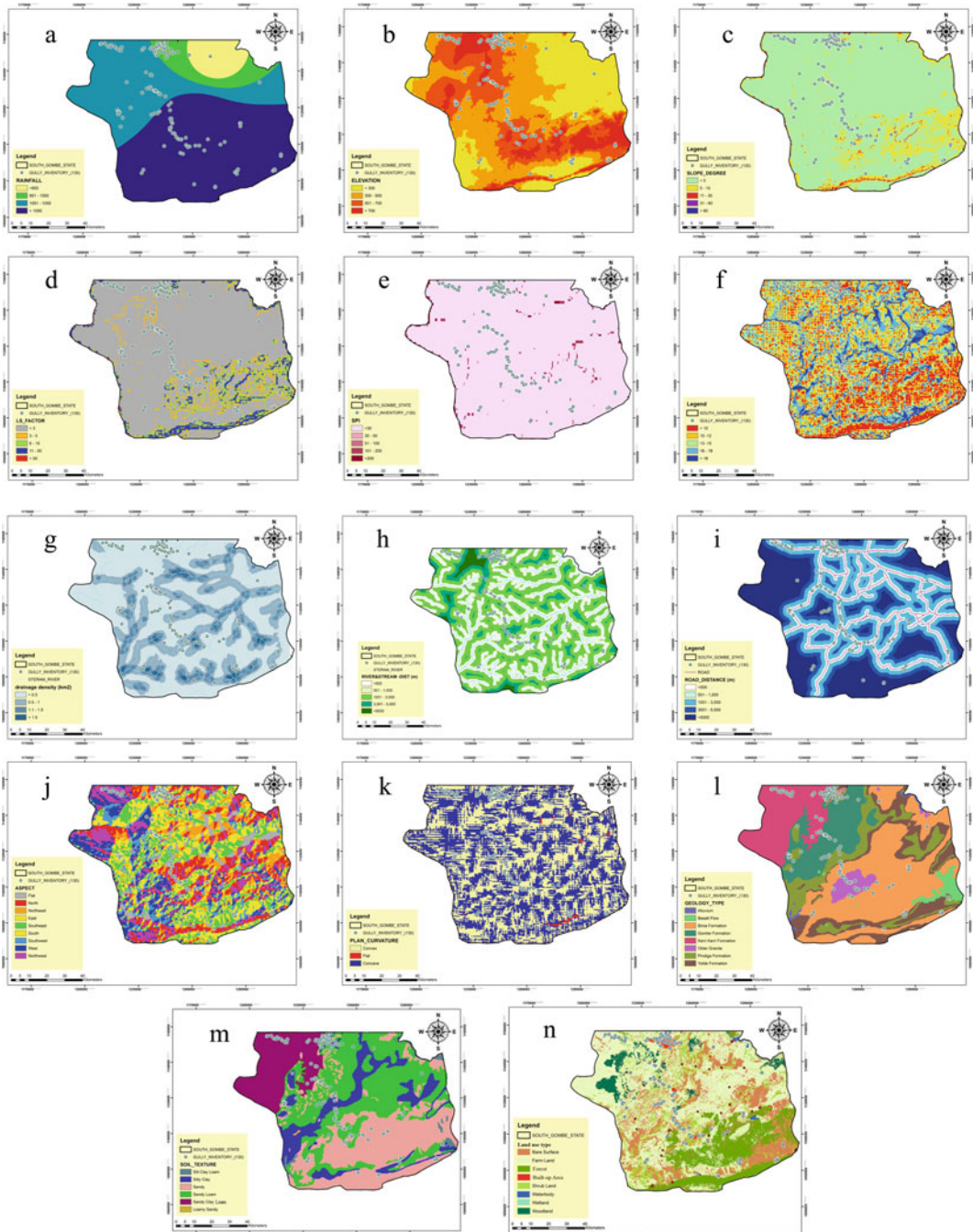


Fig. 3.4 GEIF Maps. **a** rainfall, **b** elevation, **c** SA, **d** LS, **e** SPI, **f** TWI, **g** DD, **h** DS, **i** DR, **j** Aspect, **k** PC, **l** geology, **m** ST, **n** LU

reflects a highly dissected drainage basin with a relatively rapid hydrological response to rainfall events, while a low DD entails a poorly drained

basin with a slow hydrologic response (Charizopoulos et al. 2019). However, the critical value of DD per square km that may cause soil erosion

by water is 0.90 km per square km of area (Surjit et al. 2015).

Gullies are generally connected to the stream network of an area, enabling the removal of the materials eroded from upland areas (Conoscenti et al. 2014). Joshi et al. (2019) observed that first- and second-order streams behave like gullies and accelerate soil erosion mechanism. These streams are normally located on greater elevations with steeper slopes and, thus, get conducive conditions for soil erosion. Further channel incision in those streams leads to their expansion and initiation of gullies in the nearby non-incised surface. For this reason, DS is recognized as a factor of gully development, with the assumption that the areas closer to existing first- and second-order streams are more susceptible to GE (Dube et al. 2014; Zakerinejad and Maerker 2015).

Roads play an important role in rural and urban development; however, they have a significant impact on gully occurrence and distribution (Pourghasemi et al. 2017). Roads induce concentration of surface runoff, divert concentrated runoff to other catchments, and increase catchment size, which eventually leads to gully development (Conoscenti et al. 2014). Consequently, the spatial distribution of road networks was examined.

Aspect is a topographical attribute that is considered crucial in GES assessment. It is commonly referred to as the direction to which the slope face. Aspect is expressed in degrees from north and clockwise, ranging from 0 to 360. The value of negative one (-1) is used to indicate flat surfaces such as flood plains and fluvial terraces. According to Conoscenti et al. (2014) and Rahmati et al. (2016), aspect indirectly influences GE processes given its relationship with the duration of sunlight exposure, evapotranspiration, moisture retention, vegetation cover type, and distribution on slopes.

PC is described as the curvature of a contour line formed by intersecting a horizontal plane with the surface. The use of the word curvature technically defines the rate of change of SA or aspect, usually in a particular direction (Conforti

et al. 2010). The influence of PC on slope erosion processes is the convergence or divergence of water during downslope flow. While assessing PC, positive values describe convexity, while negative values characterized the concavity of slope curvature. The values of the PC around zero indicate that the surface is flat.

Rainfall is a central factor that drives soil erosion and gully development through its potential ability (erosivity) to disintegrate soil aggregates and transport them downslope (Nearing et al. 2017). Rainfall erosivity is a function of the physical characteristics of rainfall (intensity and long duration). These aspects, in addition to those related to amount, drop size distribution, terminal velocity, and extraneous factors such as wind velocity and slope angle, determine rainfall erosivity (Rutebuka et al. 2020). Rainfall erosivity using annual rainfall distribution data from weather stations in SGS was analyzed.

Geological features and the weathering properties of material exposed or close to the earth's surface affect GE. Both soft and hard rocks are scoured by degradation through fluvial erosion, mass wasting, soil creeps, and landslides. Soft rocks are generally more susceptible to soil erosion than hard rocks. The extent of effectiveness of degradation is a function of the degree of consolidation or cementation among the different sedimentary types. GE in the areas underlain by rocks of the Basement Complex, including younger and older granites, usually occurs along modern and ancient water channels. These water channels have earlier been covered by alluvial deposits. Small scale gullies produced by fluvial action are generally associated with Basement Complex areas. Weathered zones of Basement Complex terrain are strewn with incipient and minor gullies running down slopes and generally oriented in all directions.

The Susceptibility of soil to GE is inclined to its characteristics which are closely associated with geologic settings and contribute expressively to soil infiltration, runoff rate, soil resistance to erosion, and gully occurrence (Rahmati et al. 2016). ST, organic content, structure, and

permeability have been shown to influence soil erodibility. Predominantly, ST has a strong correlation with soil erosion as it controls the erodibility and cohesiveness of the soil (Pal 2015). Thus, ST was assessed for the present study.

LUs are among the major environmental factors regulating hydrological and geomorphological processes by controlling overland flow, runoff generation, and sediment dynamics. Naturally, barren lands and sparse rangeland are more susceptible to erosion than pastures and forests with dense cover, where the vegetation cover greatly cuts the erosive action of surface runoff. In fact, there is a negative association between the rate of erosion and the density of vegetation (Hayas et al. 2017). Figure 3.4 shows the 14 selected factors used for the study.

Multicollinearity Scrutiny

Multicollinearity is a statistical phenomenon that describes interdependency conditions among multiple independent variables. In other words, it is the lack of independence which is indicated by high intercorrelations among a set of variables. In GES assessment, numerous methods have been used to test for multicollinearity. However, the variance inflation factor (VIF) and tolerance (TOL) are frequently used for this purpose (Cama et al. 2017). Hence, the VIF and TOL are used in this study. A $VIF \geq 10$ or $TOL \leq 0.1$ indicates serious multicollinearity (Guo-Liang et al. 2017).

3.2.3 Susceptibility Spatial Modeling Using MLR

3.2.3.1 Determination of Spatial Relation Between GE and GEIFs

GEIFs have been assessed and their independence confirmed. However, there is the need to determine the relationship between the GEIFs and GE to identify the most influential factors, and successfully predict GES. To achieve this, the forward stepwise logistic regression was employed. Usually, stepwise logistic regression

is most often used in situations where the important independent variables are not known, and their associations with the outcome are not well understood. It involves estimating the model with each variable entered in turn and looking at the change in the logarithm of likelihood when each variable is added. If the observed significance level (Wald test) is less than the probability for remaining in the model (0.05 in this study), the variable is entered into the model and the model statistics are recalculated to see if any other variables are eligible for entry. Finally, it becomes a model excluding all insignificant independent variables, and coefficients are allocated to the independent variable classes, correlated with the gully training data set. If a coefficient is positive, its transformed log value will be greater than one, meaning that the event is more likely to occur. If a coefficient is negative, the latter will be less than one, and the odds of the event occurring decreases.

3.2.3.2 Estimation of Probability of Gully Occurrence

The principle of logistic regression rests on the analysis of a problem, in which a result measured with dichotomous variables such as 0 and 1 or true and false, is determined from one to more independent factors. In the case of gully erosion susceptibility mapping, the goal of logistic regression is to find the best fitting model to describe the relationship between the presence or absence of gullies (dependent variable) and a set of independent parameters. Logistic regression generates the model statistics and coefficients of a formula useful to predict a logit transformation of the probability that the dependent variable is 1 (probability of occurrence of a gully event). Thus, in this study, binary Logistic regression estimates the probability (P) of the occurrence of a gully through the formula:

$$P = \frac{1}{1 + e^{-z}} \frac{1}{1 + e^{-z}} \quad (3.5)$$

where P is the probability of a gully occurring. Z is a value from $-\infty$ to $+\infty$, defined by the following equation;

$$Z = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (3.6)$$

where b_0 is the intercept (constant) of the logistic regression model, n is the number of independent variables, and b_1, b_2, \dots, b_n are coefficients, which measure the influence or contribution of independent variables (X_1, X_2, \dots, X_n). In addition to the model statistics and coefficients, the outcome of the logistic regression process was obtained in ArcGIS as a predicted map of probability defined by numbers that are confined between 0 and 1. The predicted probability values were subsequently reclassified into five susceptibility classes using the Jenks natural classification method to produce the final gully erosion susceptibility map.

3.3 Validation of the GESM

Validation is the task of demonstrating that the model is a suitable representation of the actual system: that it reproduces system behavior with enough reliability to satisfy analysis objectives (Rahmati et al. 2017b). Without validation, the prediction model and image are ineffective and lack scientific significance. As a result, the final users would not be sure of the model outputs and are unlikely to use them for planning and decision-making. Normally, the validation of predictions is based on the comparison between the prediction results and the unknown target pattern, the areas affected by future gully erosion. The unknown target pattern is usually represented by a part of the known gully pattern. Usually, the comparison of the model results and observed data is represented through a confusion matrix (Table 3.4).

According to Table 3.4, TP (true positive) and TN (true negative) also known as sensitivity are the numbers of gullies that are correctly

classified, whereas FN (false negative) and FP (false positive) referred to as specificity are the numbers of gullies that are incorrectly classified. Several approaches such as Efficiency, Kappa coefficient, Seed Cell Area Index (SCAI), Area Under Receiver Operating Characteristics (AUROC) among others have been used for validating GESMs. In this study, the AUROC is a graphical plot that illustrates the analytical ability of a model as its discrimination threshold is varied (Vakhshoori and Zare 2018). It is created by plotting sensitivity on the y-axis against specificity on the x-axis. The two parameters are derived as

$$\text{Sensitivity} = \frac{TP}{TP + FN} \frac{TP}{TP + FN} \quad (3.7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \frac{TN}{TN + FP} \quad (3.8)$$

The shape of the AUROC curve indicates the predictive performance of the model, where the predictive performance of the model is higher when the AUROC curve is closer to the upper left corner. The highest possible AUROC = 1 represents 100% specificity and 100% sensitivity. Furthermore, AUROC values of <0.6 indicate a poor, 0.6–0.7 a moderate, 0.7–0.8 a good, 0.8–0.9 a very good, and >0.9 an excellent model performance (Jiang 2020; Vakhshoori and Zare 2018).

3.4 Results and Discussion

3.4.1 Gully Erosion Inventory

The results of the interpretation of Google Earth imageries and field investigations confirmed that the study area is indeed affected by the GE phenomenon. A total of 130 critical gully

Table 3.4 Confusion matrix used for the evaluation of models

Observed	Predicted	
	Non-gully (-)	Gully (+)
Non-gully (-)	(- -) True negative (TN)	(+ -) False-positive (FP)
Gully (+)	(- +) False-negative (FN)	(+ +) True positive (TP)

eroding sites, covering 34,499.2 or 4.3% of the study area were identified and classified based on their position on the landscape and morphological hydraulic and geometry characteristic into three (3) main types of permanent gullies: continuous (35%), discontinuous (55%), and bank (10%). Generally, the gullies are characterized by incisions having near-vertical banks and are mostly large. Gully depth ranges from 0.9 to 12.2 m, less frequently up to 20 m and exceptionally even greater. The length reaches a maximum of 10,857 m on continuous, 497.2 m on discontinuous, and 1092.3 m on the bank gullies. Cross sections of the gullies are mostly U-shaped even though V-shaped gullies are also found. Spatially, it was observed in the northern parts of the study area, where Gombe and Kerikeri geologic formations are massively exposed that gully channels are mostly wide, U-shaped, and lack vegetation cover on the side slopes signifying active gullying stage. While in the mountainous southern part, narrow and V-shaped gullies are common with some vegetation growing on slopes indicating stabilization. However, the presence of falls on some of the channels indicated gully rejuvenation. Table 3.5 presents the morphometric characteristics of the identified permanent gullies in the study area.

3.4.2 Multicollinearity Scrutiny

The results of multicollinearity scrutiny as shown in Table 3.6 indicate that the highest value of VIF is 4.854 and the lowest coefficient of tolerance is 0.206, respectively. Therefore, there is no collinearity between these factors, and this

allows the inclusion of all the factors in the final gully erosion susceptibility modeling process.

3.4.3 Susceptibility Spatial Prediction

3.4.3.1 Spatial Relationships Between GE and GEIFs

The results of the spatial relationship between GE and each GEIF obtained from the forward stepwise logistic regression are displayed in Fig. 3.5.

According to Fig. 3.5a, for all the rainfall classes, there is a positive relationship with GE occurrence since all the coefficient (β) values assigned to each class are positive. However, the rainfall classes 1001–1050 and >1100 with β values 2.261 and 1.183 exhibited a strong positive relationship when compared with the first two classes which were assigned weak positive association. The results suggest that the propensity of GE occurrence in the study area increases with an increase in rainfall amount. The result is in harmony with the findings obtained by Conforti et al. (2010) in the Turbolo catchment, Italy; Rahmati et al. (2016) in the Chavar region, Iran; and Makaya et al. (2019) in the upper uMgeni catchment in KwaZulu Natal, South Africa, who also showed that GE development generally occurs in regions with high rainfall amount.

Regarding elevation, regression analysis indicated that elevation has a positive but weak relationship with GE activity. From Fig. 3.5b it can be observed that an increase in elevation causes a corresponding increase in the probability of GE occurrence. This holds to the elevation class 300–500 m ($\beta = 0.369$) from which the

Table 3.5 Morphometric characteristics of the identified permanent gullies

Gully type	Length (m)			Width (m)			Depth (m)			Gullied area (ha)	Gully area (%)
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean		
Cont	2097	10,857	5550.7	6.4	68	31.4	1.4	20	12.2	21,950.1	64
Bank	777	1370	1092.3	4	43.1	6	0.9	13.4	1.8	8624.8	25
Discount	99.2	915	494.2	3.1	10.1	15.4	0.9	3.2	4.9	3924.3	11
Total										34,499.2	100

Table 3.6 Multicollinearity test results for the relationship among factors

Factor	Multicollinearity test		Factor	Multicollinearity test	
	Tol.	VIF		Tol.	VIF
Geology	0.578	1.730	Plan curvature	0.748	1.336
Soil texture	0.312	3.205	SPI	0.579	1.727
Land use	0.502	1.992	TWI	0.787	1.271
Aspect	0.906	1.104	Drain_density	0.206	4.854
Elevation	0.819	1.221	Dist_stream	0.872	1.147
Slope	0.613	1.631	Dist_road	0.287	3.484
LS	0.724	1.381			

relationship reverses and eventually becomes negative at the class >700 m (-0.299). This result agrees with the findings of numerous studies (Alireza et al. 2019; Arabameri et al. 2020d, 2019a; Rahmati et al. 2017b) who agreed that lower elevations are most susceptible to GE.

In the case of SA, a positive spatial correlation exists between gully formation and areas with SA below 30° . This is confirmed by the positive β values 1.387, 0.872, and 0.454 assigned to <5 , 5–10, and 11–30 classes, respectively (Fig. 3.5 c). Conversely, for SA classes above 30° , β values were negative, indicating weak relationships and low probability of gully occurrence. This result is in agreement with Lazarus (2012); Adediji et al. (2013); Rahmati et al. (2016), (2017a); Debanshi and Pal (2018), and Arabameri et al. (2020d) who found most gullies occurring on lower slopes. According to these studies, the reason for this could be because lower SAs have greater soil depth, intensive rainfall impaction and greater runoff from upslope will decrease soil strength resulting in the development and extension of the gully channels.

The influence of LS on GE as shown in Fig. 3.5d revealed that the probability of gully occurrence increased with increasing LS. The highest probability was found where LS was 6–10 m ($\beta = 1.262$). However, immediately after this maximum level, the probability to host a gully abruptly decreased. The reason for this sudden decline was unclear, but it might relate to land use, vegetation, and other factors in the area.

The result in this study confirms the findings of Bagio et al. (2017) and Zabihi et al. (2018) that a direct relationship exists between the LS and GE. Their result also implied that the higher the LS, the higher the probability of GE occurrence due to increased runoff velocity and a decreasing detachment and transport threshold of soil particles.

Concerning SPI, a significant positive association exists with GE. As shown in Fig. 3.5e a strong positive correlation was found in the class 51–100 ($\beta = 3.456$) and class 101–200 ($\beta = 1.356$). The remaining three classes had a low positive influence on GE. However, the results indicate that the probability to host a gully due to SPI increases with an increase in SPI. The outcome of this analysis is supported by the works of Shit et al. (2015); Rahmati et al. (2017a); Amiri et al. (2019); and Domazetović et al. (2019) who also observed that higher SPI values have higher erosion potential and often coincide with the appearance of soil removal and GE occurrence.

The bond between TWI and GE occurrence portrays an ascending pattern with the lowest class (<10) assigned the minimum β value (0.070) while the highest class (>18) got the maximum β value (3.112) (Fig. 3.5f). Thus, the result suggests that the probability of GE susceptibility intensifies with an increase in TWI. This result is similar to the study of Arabameri et al. (2020b) in the Chah Mousi watershed in Semnan province, Iran, that gully formation in the watershed is particularly favored in areas

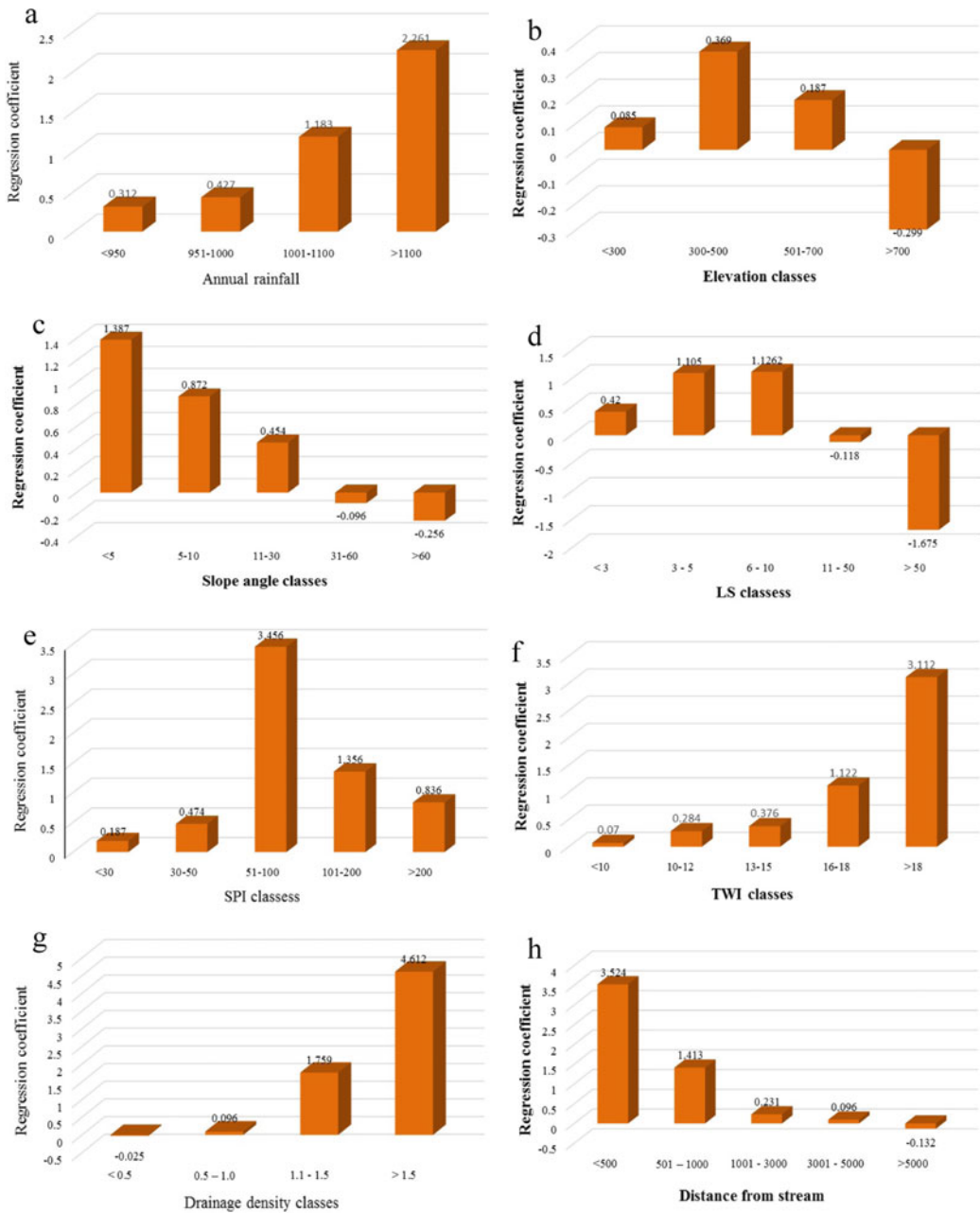


Fig. 3.5 Spatial relationships between GE and each class of GEIF **a** rainfall, **b** elevation, **c** SA, **d** LS, **e** SPI, **f** TWI, **g** DD, **h** DS, **i** DR, **j** Aspect, **k** PC, **l** geology, **m** ST, **n** LU

with high TWI values representing zones of saturation with high surface soil water along drainage paths. These saturated areas favor gully formation since the surface soils lose their strength as they become wet. Also, Zabihi et al.

(2018) reported in their study in the Mazandaran province of Iran that, the greater the TWI factor, the greater is the potential for gully occurrence. High values of TWI increase the filtration rate and provide the conditions for piping and roof

of GE occurrence decreases. For aspect, positive correlation exists between east ($\beta = 2.737$), southeast ($\beta = 0.811$), northeast ($\beta = 0.644$), south ($\beta = 0.418$), north ($\beta = 0.225$), and southwest ($\beta = 0.087$) facing slopes. While west, northwest, and flat facing aspects exhibited negative relationships (Fig. 3.5j). This result implies that slopes facing north to southwest predominate and have a greater propensity to host gullies than slope aspect facing west. This is in agreement with Zabihi et al. (2018), who reported that eastward aspects in the Mazandaran province of Northern Iran are more susceptible to GE. This is so because the eastward facing slope aspects get more solar radiation in the northern hemisphere and, as a result, they experience more evaporation, higher soil porosity (total pore space), lower soil strength, and lower vegetation density.

The interpretation of the β values linking GE and PC classes revealed an all positive relationship (Fig. 3.5k). However, the concave class exhibited the strongest relationship with a β value of 0.287, followed by flat class (0.083), and convex class (0.021). There is a consensus between this result and the findings reported by Conforti et al. (2010) in the Turbolo catchment, Italy, and Rahmati et al. (2016) in the Chavar region, Iran, who also confirmed flat and concave curvatures are more prone to GE. About the connection between GE and geologic Formations, Fig. 3.5l showed that all Formations have a weak positive relationship with GE except Basalt Formation which had a negative relationship with the phenomenon. The positive relationship ranges from $\beta = 0.053$ on the Basement Complex to $\beta = 0.615$ on the Gombe sandstone. This confirms the study by Lazarus (2012) who concluded that the sandstones and shales that dominate Gombe States geology are more susceptible to GE than other geologic Formations in the area.

ST and GE revealed a significant relationship. However, among the six ST types (Fig. 3.5m), the sandy class had the highest positive β value (0.524), indicating the highest GE susceptibility, followed by sandy clay loam ($\beta = 0.467$), Sandy loam ($\beta = 0.214$), and Loamy sand ($\beta = 0.103$) accordingly. Contrarily,

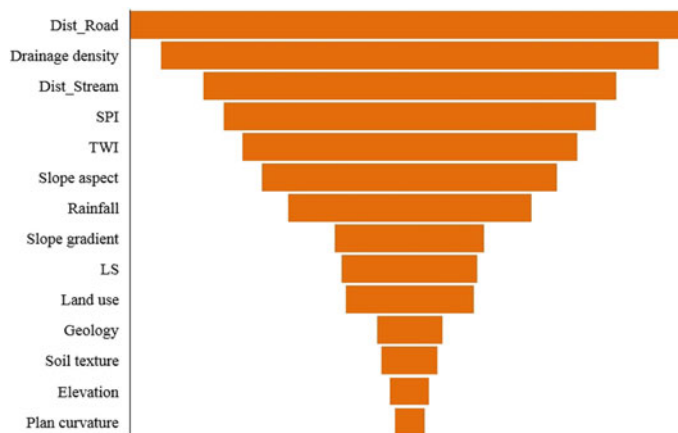
Silt clay loam and Silt clay obtained negative β values, showing the negative influence on GE incidence. A close look at the results also revealed that areas dominated by sandy loam and sandy clay loam hosted more gullies due to high sand contents that render them easily detachable. Similar work by Abdulfatai et al. (2014) and Igwe and Egbueri (2018) stated that the dominance of sand proportion in soils accelerates GE, while Silt clay loam and silt clay classes, due to sufficient and high clay content are resistant to soil erosion and gully development. Regarding LU (Fig. 3.5n), regression analyses established that bare surface, shrubland, farmland, built-up areas, and woodland have positive β values. The highest β value (1.197) was found on bare surfaces, followed by shrubland, which had a β value of 0.762. A negative relationship between LU and GE occurred in the wetland and waterbody classes, with waterbody having the lowest β value (-7.534). The finding is in harmony with previous studies (Amiri et al. 2019; Devátý et al. 2019; Shellberg et al. 2016), who concluded that protected locations like forest and rangelands experience less GE in comparison with bare surface regions.

3.4.3.2 The Relative Importance of the GEIFs

Following the stepwise regression allocation of β values to classes in all factors which confirmed their relationship with GES, the relative importance of each factor was determined based on the highest β value within the factor. The higher the β value, the stronger the effect of the given factor on GE occurrence. As shown in Table 3.7 and Fig. 3.6 DR, DD, DS, SPI, TWI, aspect, and rainfall had the strongest values and significantly contributed to GES whereas SA, LS, and LU exhibited moderate importance. In contrast, geology, ST, elevation, and PC were the least important and showed less impact in the gully occurrence in the study area, respectively. Therefore, this result suggests that among all the fourteen factors entered into the analysis, the DS factor by its highly significant β value is the most important single factor affecting GES in the study area.

Table 3.7 Relative importance and ranking of GEIFs

GEIF	β	Rank
DR	5.175	1
DD	4.612	2
DS	3.831	3
SPI	3.456	4
TWI	3.112	5
Aspect	2.737	6
Rainfall	2.261	7
SA	1.387	8
LS	1.262	9
LU	1.197	10
Geology	0.615	11
ST	0.524	12
Elevation	0.369	13
PC	0.287	14
<i>Constant</i>	-9.327	

**Fig. 3.6** Relative importance of GEIFs

3.4.3.3 GE Probability Mapping

After the influence of each factor class was determined, GES was predicted by computing the probability of gully occurrence in the study area. This was attained by executing Eq. 3.5 and 3.6. The value of z (Eq. 3.6) was computed by substituting b with the β values of the most important factors earlier determined. Because the highest positive significant coefficient in the

analysis ($\beta = 5.175$) belongs to the DR factor class < 500 , DR was introduced as the most significant determining factor for GE occurrence, thus assuming the rest of the factors are constant, for a unit change in DR, the probability of GE occurrence will be $e^{5.175}$ or 176.796 times. Hence, the β values of the most significant factors were inputted in the equation to compute z as

$$\begin{aligned}
 z = & -9.327 + 5.175 * DR(Class1) + 1.811 \\
 & * DR(Class2) + 4.612 \\
 & * DD(Class1) + 1.759 \\
 & * DD(Class2) + 3.831 \\
 & * DS(Class1) + 1.413 * DS(Class2) + 3.456 \\
 & * SPI(Class3) + 1.356 \\
 & * SPI(Class4) + 3.112 \\
 & * TWI(Class5) + 1.122 \\
 & * TWI(Class4) + 2.737 \\
 & * Aspect(east) + 2.261 \\
 & * rainfall(Class4) + 1.183 \\
 & * rainfall(Class3) + 1.387 \\
 & * SA(Class1) + 1.262 * LS(Class3) + 1.105 \\
 & * LS(Class2) + 1.197 * LU(baresurface)
 \end{aligned}
 \tag{3.9}$$

Once the value of z was obtained, it was inserted into Eq. 3.10 and the probability of GE occurrence was calculated as

$$\begin{aligned}
 P(\text{probability of gully occurrence}) & \\
 = & \frac{1}{1 + e^{-9.327 + 5.175 * DR(Class1) \\
 & + 1.811 * DR(Class2) + 4.612 * DD(Class1) + 1.759 * DD(Class2) \\
 & + 3.831 * DS(Class1) + 1.413 * DS(Class2) + 3.456 * SPI(Class3) \\
 & + 1.356 * SPI(Class4) + 3.112 * TWI(Class5) + 1.122 * TWI(Class4) \\
 & + 2.737 * Aspect(east) + 2.261 * rainfall(Class4) \\
 & + 1.183 * rainfall(Class3) + 1.387 * SA(Class1) + 1.262 * LS(Class3) \\
 & + 1.105 * LS(Class2) + 1.197 * LU(baresurface)}}
 \end{aligned}
 \tag{3.10}$$

The output of the above computation was subsequently transferred into the ArcGIS and the GESM with the probability (p) ranging from 0 to 1 was prepared. The produced map was then partitioned into 20 classes using a threshold of 0.05 equal probability intervals. After overlapping with the GE inventory map, a histogram representing the incidence of gullies and non-gully occurrence against the probability classes was plotted. Lastly, based on the histogram, the probability range was reclassified using the natural breaks classification method into five classes: 0–0.076, 0.076–0.494, 0.494–0.733, 0.733–0.898, and 0.898–1.00, representing relatively safe, less susceptible, moderately susceptible, highly susceptible, and extremely susceptible, respectively. Table 3.8 and Fig. 3.7 present the characteristics of the five susceptibility classes and the final GESM.

From a visual analysis of the final GESM (Fig. 3.7), it can be observed that most parts of the study area particularly the eastern and southwestern parts fall in the less and moderately susceptible classes, while the northwest and central parts fall in the highly and extremely susceptible classes. It was also observed from field observation that the areas within the high and extreme susceptibilities are mostly located near roads. This underscores the results of the relationship between GE and GEIFs of which DR

Table 3.8 Characteristics of the five GES classes

Susceptibility class	Susceptibility index	Surface area (ha)	Surface area (%)	No. gullies	% gullies
Relatively Safe	0–0.076	24,872	3.1	0	0
Less Susceptible	0.076–0.494	401,962	50.1	02	2.20
Moderately Susceptible	0.494–0.733	186,941	23.3	07	7.69
Highly Susceptible	0.733–0.898	154,045	19.2	16	17.58
Extremely Susceptible	0.898–1.00	34,500	4.3	66	72.53
Total		802,320	100	91	100

Table 3.9 Confusion matrix that presents results of the validation

Observed		Predicted		Total	% correct
		Gully (1)	Non-gully (0)		
Gully (1)		37 (TP)	02 (FN)	39	94.87
Non-gully (0)		04 (FP)	35 (TN)	39	89.74
Total		41	37	78	92.31

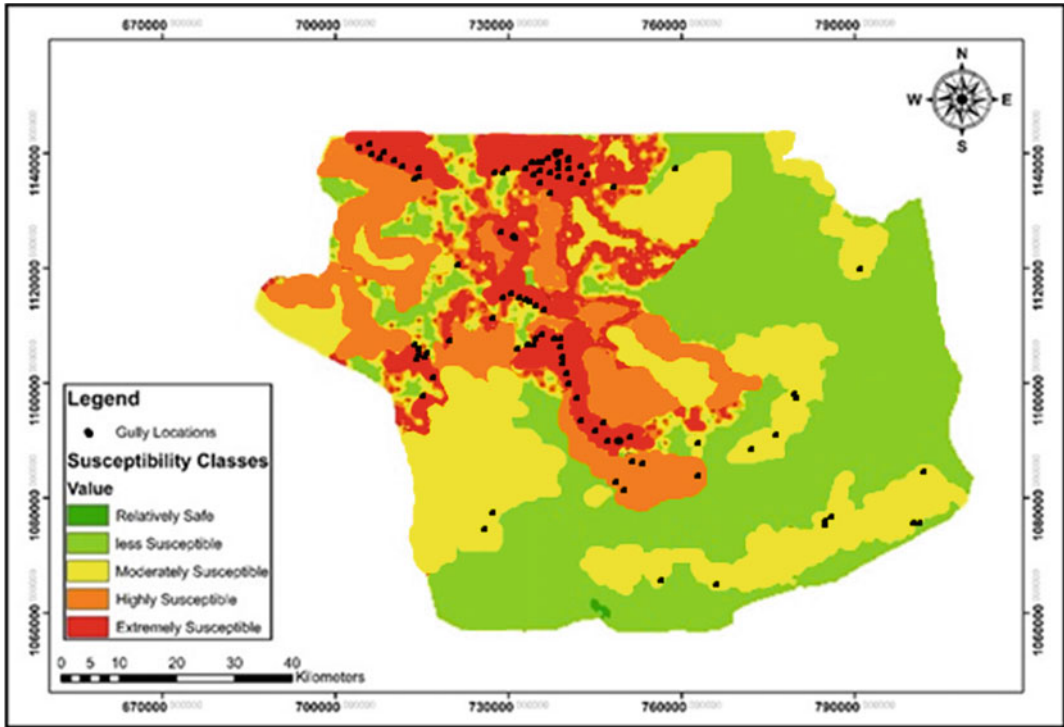


Fig. 3.7 Predictive GESM derived using the logistic regression model

(<500 m) is the most important single factor determining GE susceptibility in the study area

3.5 Validation

The predictive performance of the MLR model was evaluated using the validation data set (78). This was performed by calculating the values of sensitivity and specificity and then drawing the AUROC curve. As earlier indicated, sensitivity was computed using Eq. 3.7 as the fraction of locations

hosting gullies that were correctly classified as susceptible, while specificity is derived from the fraction of areas free of gullies that were correctly classified as not susceptible using Eq. 3.8. Table 3.9 and Fig. 3.8 present the validation results and the AUROC curve. From Table 3.9 it can be seen that the model correctly classified 37 gully locations and 35 non-gully locations, representing 94.87% and 89.74%, respectively. Also, the overall predictive performance of the model as given by the value of AUROC (0.923% or 92.3%) indicates excellent performance.

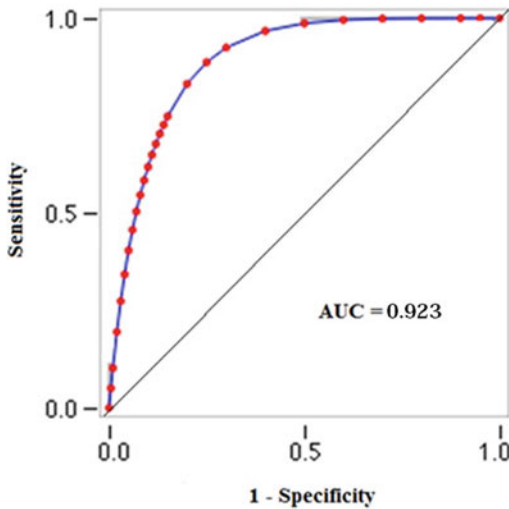


Fig. 3.8 AUROC curve for the final GESM

3.6 Conclusion

GE is the water erosion process that cuts soils and forms permanent gullies on the landscape that cannot be eliminated through conventional tillage operations. The formation and development of gullies is an important environmental threat throughout the world since it is responsible for land degradation, increase in sediment delivery, and reduction of water quality. It is also responsible for a decreased water travel time to rivers (and hence increased flooding probabilities), for the filling up of ponds and reservoirs, and for the destruction of infrastructure (buildings and roads). Therefore, the prediction of areas susceptible to GE is a crucial issue for environmental scientists, land managers, and decision-makers. To tackle this problem, researchers usually employ Knowledge-driven and Data-driven predictive methods. This study demonstrated the ability of data-driven logistic regression to accurately predict GES in SGS, Nigeria.

A GE inventory was prepared from a total of 260 gully and non-gully locations compiled from the interpretation of Google Earth images and field investigations. Besides, local environmental conditions and a 20 m DEM allowed the

selection of soil texture, geology, land use, rainfall, and some topographical factors influencing GE susceptibility. Subsequently, the inventory data was randomly split into two datasets; 182 or 70% for training the logistic regression model, and 78 or 30% validation for validation of prediction results, while influencing factors independence was assessed using multicollinearity scrutiny. Results of forward stepwise regression for the relationship between GE and selected factors indicated that distance from road is key to gully formation. After running the logit function, the resultant susceptibility map revealed that 3.1% of the study area was relatively safe, 50.1% less, 23.3% moderate, 19.2% high, and 4.3% extremely susceptible. Validation assessment using area under the receiver operating characteristic curve provided 92.3% prediction accuracy. This study further confirmed logistic regression as an excellent and accurate data-driven method for spatial analysis and prediction of GE susceptibility. The method can be applied elsewhere with similar physiographic characteristics.

References

- Abdulfatai IA, Okunlola IA, Akande WG, Momoh LO, Ibrahim KO (2014) Review of gully erosion in Nigeria: causes, impacts and possible solutions. *J Geosci GeomatS* 2(3):125–129
- Abdulkadir TS, Muhammad RUM, Okeola OG, Khamaruzaman WY, Adelodun B, Aremu SA (2020) Spatial analysis and prediction of soil erosion in a complex watershed of Cameron Highlands, Malaysia. In: Shit P, Pourghasemi H, Bhunia G (eds) *Gully Erosion Studies from India and Surrounding Regions*. Advances in Science, Technology & Innovation (IEREK Interdisciplinary Series for Sustainable Development). Springer, Cham
- Adediji A, Jeje LK, Ibitoye MO (2013) Urban development and informal drainage patterns: gully dynamics in Southwestern Nigeria. *Appl Geogr* 40:90–102
- Alireza A, Artemi C, John PT (2019) Spatial pattern analysis and prediction of gully erosion using novel hybrid model of entropy-weight of evidence. *Water* 1129(11)
- Amiri M, Pourghasemi HR, Ghanbarian GA, Afzali SF (2019) Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma* 340:55–69

- Arabameri A, Pourghasemi HR (2019) Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R. In: Pourghasemi HR, Gokceoglu C (eds) *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier, pp 299–321
- Arabameri A, Pradhan B, Lombardo L (2019a) Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling. *CATENA* 183:104223
- Arabameri A, Pradhan B, Rezaei K, Conoscenti C (2019b) Gully erosion susceptibility mapping using GIS-based multi-criteria decision analysis techniques. *CATENA* 180:282–297
- Arabameri A, Rezaei K, Cerdà A, Conoscenti C, Kalantari Z (2019c) A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Sci Total Environ* 660:443–458
- Arabameri A, Chen W, Lombardo L, Blaschke T, Bui DT (2020c) Hybrid computational intelligence models for improvement gully erosion assessment. *Remote Sensing* 12(1):25
- Arabameri A, Pradhan B, Bui DT (2020d) Spatial modelling of gully erosion in the Ardeb River Watershed using three statistical-based techniques. *CATENA* 190:104545
- Arabameri A, Pradhan B, Rezaei K, Yamani M, Pourghasemi HR, Lombardo L (2018) Spatial modelling of gully erosion using Evidential Belief Function, Logistic Regression and a new ensemble EBF–LR algorithm. *Land Degrad Dev*
- Arabameri A, Blaschke T, Pradhan B, Pourghasemi HR, Tiefenbacher JP, Bui DT (2020a) Evaluation of recent advanced soft computing techniques for gully erosion susceptibility mapping: a comparative study. *Sensors* 20(2)
- Arabameri A, Chen W, Blaschke T, Tiefenbacher JP, Pradhan B, Tien Bui D (2020b) Gully head-cut distribution modeling using machine learning methods—a case study of N.W. Iran. *Water* 12(16)
- Azareh A, Rahmati O, Rafiei-Sardooi E, Sankey JB, Lee S, Shahabi H, Ahmad BB (2019) Modelling gully-erosion susceptibility in a semi-arid region, Iran: Investigation of applicability of certainty factor and maximum entropy models. *Sci Total Environ* 655:684–696
- Bagio B, Bertol I, H, W. N., Schneiders, D., & Santos, M. (2017) Water erosion in different slope lengths on bare soil. *Rev Bras Cienc Solo* 41:e0160132
- Cama M, Lombardo L, Conoscenti C, Rotigliano E (2017) Improving transferability strategies for debris flow susceptibility assessment: application to the saponara and Itala catchments (Messina Italy). *Geomorphology* 28852–65. <https://doi.org/10.1016/j.geomorph.2017.03.025>
- Castillo C, Gómez JA (2016) A century of gully erosion research: Urgency, complexity and study approaches. *Earth Sci Rev* 160:300–319
- Charizopoulos N, Mourtziou P, Psilovikos T, Psilovikos A, Karamoutsou L (2019) Morphometric analysis of the drainage network of Samos Island (northern Aegean Sea): insights into tectonic control and flood hazards. *CR Geosci* 351(5):375–383
- Conforti M, Aucelli PPC, Robustelli G, Scarciglia F (2010) Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Nat Hazards* 56(3):881–898
- Conoscenti C, Rotigliano E (2020) Predicting gully occurrence at watershed scale: comparing topographic indices and multivariate statistical models. *Geomorphology* 359:107123
- Conoscenti C, Angileri S, Cappadonia C, Rotigliano E, Agnesi V, Märker M (2014) Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). *Geomorphology* 204:399–411
- Conoscenti C, Agnesi V, Angileri S, Cappadonia C, Rotigliano E, Märker M (2013) A GIS-based approach for gully erosion susceptibility modelling: a test in Sicily, Italy. *Environ Earth Sci* 70(3):1179–1195
- Debanshi S, Pal S (2018) Assessing gully erosion susceptibility in Mayurakshi river basin of eastern India. *Environ Dev Sustain*
- Devátý J, Dostál T, Hösl R, Krása J, Strauss P (2019) Effects of historical land use and land pattern changes on soil erosion—case studies from Lower Austria and Central Bohemia. *Land Use Policy* 82:674–685
- Dewitte O, Daoudi M, Bosco C, Van Den Eeckhaut M (2015) Predicting the susceptibility to gully initiation in data-poor regions. *Geomorphology* 228:101–115
- Domazetović F, Šiljeg A, Lončar N, Marić I (2019) Development of automated multicriteria GIS analysis of gully erosion susceptibility. *Appl Geogr* 112:102083
- Domínguez-Cuesta MJ (2013) Susceptibility. In: Bobrowsky PT (ed) *Encyclopedia of Natural Hazards*. *Encyclopedia of Earth Sciences Series*. Springer, Dordrecht
- Dube F, Nhapi I, Murwira A, Gumindoga W, Goldin J, Mashauri DA (2014) Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District - Zimbabwe. *Phy Chem Earth Parts A/B/C* 67(69):145–152
- Garosi Y, Sheklabadi M, Conoscenti C, Pourghasemi HR, Van Oost K (2019) Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci Total Environ* 664:1117–1132
- Guo-liang D, Yong-shuang Z, Javed I, Zhi-hua Y, Xin Y (2017) Landslide susceptibility mapping using an integrated model of information value method and logistic regression in Bailongjiang watershed, Gansu Province, China. *J Mountain Sci* 14(2):249–268. <https://doi.org/10.1007/s11629-016-4126-9>.
- Gomez-Guitierrez A, Conoscenti C, Angileri SE, Rotigliano E, Schnabel S (2015) Using topographical attributes to evaluate gully erosion proneness

- (susceptibility) in two mediterranean basins: advantages and limitations. *Nat Hazard* 79(1):291–314. <https://doi.org/10.1007/s11069-015-1703-0>
- Gayen A, Pourghasemi HR, Saha S, Keesstra S, Bai S (2019) Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci Total Environ* 668:124–138
- Hayas A, Poesen J, Vanwalleghem T (2017) Rainfall and Vegetation effects on temporal variation of topographic thresholds for gully initiation in mediterranean cropland and olive groves. *Land Degrad Dev* 28(8):2540–2552
- Igwe O, Egbueri JC (2018) The characteristics and the erodibility potentials of soils from different geologic formations in Anambra State, Southeastern Nigeria. *J Geol Soc India* 92(4):471–478
- Ikusemoran M, Didams G, Michael A (2018) Analysis of spatial distribution of geology and pedologic formations in gombe state, North Eastern Nigeria. *J Geography Geol* 10(1):83–108
- Ikusemoran M, Wanah B, Mbaya L (2016) Geospatial techniques for terrain analysis of Gombe State, Nigeria. *J Geography Environ Earth Sci Int* 6(1):1–20
- Jiang Y (2020) Receiver operating characteristic (ROC) analysis of image search-and-localize tasks. *Acad Radiol*
- Joshi V, Susware N, Sinha D (2019) Estimating soil loss from a watershed in Western Deccan India using Revised Universal Soil Loss Equation. *Landscape & Environ* 10(1):13–25. <https://doi.org/10.21120/le/10/1/2>
- Lazarus MA (2012) An assessment of gully erosion in Gombe Town, Gombe State, Nigeria. Unpublished Ph. D thesis. Department of Geography. University of Maiduguri, Borno State, Nigeria
- Li H, Cruse RM, Bingner RL, Gesch KR, Zhang X (2016) Evaluating ephemeral gully erosion impact on Zea mays L. yield and economics using AnnAGNPS. *Soil Tillage Res* 155:157–165
- Lucà F, Conforti M, Robustelli G (2011) Comparison of GIS-based gully susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. *Geomorphology* 134(3–4):297–308
- Makaya N, Dube T, Seutloali K, Shoko C, Mutanga O, Masocha M (2019) Geospatial assessment of soil erosion vulnerability in the upper uMgeni catchment in KwaZulu Natal, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C*
- Maria R, Nicolae R (2017) Gully erosion. In: Radoane M, Vespremeanu-Stroe A (eds) *Landform Dynamics and Evolution in Romania*. Springer Geography, Switzerland, pp. 371–396
- Moore ID, Bruch GJ (1986) Physical basis of the length-slope factor in the universal soil loss equation. *Soil Sci Soc America J* 50:1294–1298
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling: a review of hydrological, geomorphological and biological applications. *Hydrol Process* 5:3–30
- Nearing MA, Yin S-Q, Borrelli P, Polyakov VO (2017) Rainfall erosivity: an historical review. *CATENA* 157:357–362
- Nyssen J, Poesen J, Moeyersons J, Luyten E, Veyret-Picot M, Deckers J, Govers G (2002) Impact of road building on gully erosion risk: a case study from the Northern Ethiopian Highlands. *Earth Surf Proc Land* 27(12):1267–1283
- Osaragi T (2019) Classification and space cluster for visualizing geo information. *Int J Data Warehousing Min* 15(1):19–38
- Pal S (2015) Identification of soil erosion vulnerable areas in Chandrabhaga river basin: a multi-criteria decision approach. *Modeling Earth Syst Environ* 2(1):5. <https://doi.org/10.1007/s40808-015-0052-z>
- Pradeep GS, Krishnan MVN, Vijith H (2014) Identification of critical soil erosion prone areas and annual average soil loss in an upland agricultural watershed of Western Ghats, using analytical hierarchy process (AHP) and RUSLE techniques. *Arabian J Geosci* 8(6):3697–3711
- Poesen J (2011) Challenges in gully research. *Landf Anal* 17:5–9
- Poesen J, Nachtergaele J, Verstraeten G, Valentin C (2003) Gully erosion and environmental change: importance and research needs. *CATENA* 50(2):91–133
- Pourghasemi HR, Yousefi S, Komejady A, Cerda A (2017) Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci Total Environ* 609:764–775
- Rahmati O, Haghizadeh A, Pourghasemi HR, Noormohamadi F (2016) Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. *Nat Hazards* 82(2):1231–1258
- Rahmati O, Tahmasebipour N, Haghizadeh A, Pourghasemi HR, Feizizadeh B (2017a) Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: an integrated framework. *Sci Total Environ* 579:913–927
- Rahmati O, Tahmasebipour N, Haghizadeh A, Pourghasemi HR, Feizizadeh B (2017b) Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* 298:118–137
- Radula MW, Szymura TH, Szymura M (2018) Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. *Ecol Indic* 85:172–179
- Razavi-Termeh SV, Sadeghi-Niaraki A, Choi S-M (2020) Gully erosion susceptibility mapping using artificial intelligence and statistical models. *Geomat Nat Haz Risk* 11(1):821–845
- Reza E, Ronak S (2015) Evaluation of gully erosion susceptibility using logistic regression. in *Salavat Abad, Kordestan Province. Arid RegNs Geogr Stud*, 5(20)
- Rutebuka J, De Taeye S, Kagabo D, Verdoodt A (2020) Calibration and validation of rainfall erosivity estimators for application in Rwanda. *CATENA* 190:104538

- Saaty TL, Vargas LG (2012) Models, methods, concepts, and applications of the analytic hierarchy process, 2nd edn. Springer, US, p 346
- Seutloali KE, Beckedahl HR, Dube T, Sibanda M (2016) An assessment of gully erosion along major armoured roads in south-eastern region of South Africa: a remote sensing and GIS approach. *Geocarto Int* 31(2):225–239
- Shellberg JG, Spencer J, Brooks AP, Pietsch TJ (2016) Degradation of the Mitchell River fluvial megafan by alluvial gully erosion increased by post-European land use change, Queensland, Australia. *Geomorphology* 266:105–120
- Shit PK, Paira R, Bhunia G, Maiti R (2015) Modeling of potential gully erosion hazard using geo-spatial technology at Garbhatta block, West Bengal in India. *Model Earth Syst Environ* 1(1–2)
- Sujatha ER, Sridhar V (2019) Mapping of erosion susceptibility using a weighted linear combination model: a case study of a hill sub-watershed in Kodaikkandal, Western Ghats, South India. *Remote Sens Appl: Soc Environ* 14:34–45
- Surjit S, Ravinder J, Satya PK (2015) Vulnerability assessment of soil erosion using geospatial techniques—a pilot study of upper catchment of Markanda River. *Int J Adv Remote Sens, GIS Geogr* 3(1):9–21
- Vakhshoori V, Zare M (2018) Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? *Geomat Nat Haz Risk* 9(1):249–266
- van der Meij WM, Temme AJAM, Wallinga J, Hierold W, Sommer M (2017) Topography reconstruction of eroding landscapes—a case study from a hummocky ground moraine (CarboZALF-D). *Geomorphology* 295:758–772
- Zabihi M, Mirchooli F, Motevalli A, Khaledi Darvisahan A, Pourghasemi HR, Zakeri MA, Sadighi F (2018) Spatial modelling of gully erosion in Mazandaran Province, northern Iran. *CATENA* 161:1–13
- Zêzere JL, Pereira S, Melo R, Oliveira SC, Garcia RAC (2017) Mapping landslide susceptibility using data-driven methods. *Sci Total Environ* 589:250–267
- Zakerinejad R, Maerker M (2015) An integrated assessment of soil erosion dynamics with special emphasis on gully erosion in the Mazayjan basin, southwestern Iran. *Nat Hazard* 79(1):25–50. doi: <https://doi.org/10.1007/s11069-015-1700-3>

Didams Gideon is a Lecturer in the Department of Geography, Gombe State University, Tudun-Wada Gombe, Nigeria. He obtained both his Bachelor and Master of Geography degrees from the Federal University of Technology (now Modibbo Adama University of Technology (MAUTECH)), Yola Adamawa State Nigeria, and PhD in Physical Geography (Geomorphology) from the University of Malaya, Kuala Lumpur, Malaysia. He has over 10 years of teaching and research experience during which he has graduated several research students. He has also published research articles in reputable national and international journals and conference proceedings. His key research interests include soil erosion and modeling, watershed management and conservation, landscape reconstruction, application of machine learning algorithms, GIS, and remote sensing techniques in environmental geomorphology.

Firuza Begham Mustafa is an Associate Professor at the Department of Geography, Faculty of Arts and Social Sciences, University of Malaya. She specialized in agriculture geography, environment management and geophysical fields. She has produced more than 110 publications including in journals, conferences papers, IPs and posters. She has also written chapters on many international books and published several scientific books. She served in the Editorial Board Water Conservation Management (WCM), Editorial Board of Geology, Ecology and Landscapes (GEL), International Journal of Creative Industries (IJCREI) and Editorial Board of the Malaysian Journal of Tropical Geography (MJTG), University of Malaya. She was appointed as Visiting Fellow of Qinghai National University, China. She actively participated in Theo Murphy High Flyers Think Tank of Australia in 2011 and 2012. She is also Council Member of the International Geographical Union Commission on Commission Marginalization, Globalization, and Regional and Local Responses- C16.29, a member of International Geographical Union for Commission on Geographical Education, a member of the Southeast Asian Geography Association (SEAGA) and the Centre for Global Geography Education (CGGE) training by Association of American Geography (AAG). She is an Associate Fellow (AFMSA) of the Malaysian Scientific Association (MSA). She served as Auditor for Malaysia Qualification Agency (MQA) since 2009 for Geography and Environmental subjects.