

Gully erosion susceptibility considering spatiotemporal environmental variables: Midwest U.S. region

Jeongho Han, Jorge A. Guzman*, Maria L. Chu

Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA



ARTICLE INFO

Keywords:

Machine learning
MaxEnt model
LiDAR
Gully erosion
Gully erosion susceptibility mapping
Soil erosion

ABSTRACT

Study region: The study was tested in Jefferson County in Illinois, USA, whose land use is a typical representation of row crop cultivation in the Midwestern USA.

Study focus: This study aimed to predict the gully erosion susceptibility in agricultural land using remote sensed environmental data (topographic, pedologic, land cover, precipitation, and vegetation development) considering their spatio-temporal variability in a modeling framework based on the maximum entropy model MaxEnt. The methodology thoroughly evaluated each environmental factor contributing to gully erosion prediction and used a set of rules based on accuracy, transferability, and efficiency to evaluate the model performance.

New Hydrological Insights for the Region: This study developed a data-driven modeling framework that can be applied across other regions. The modeling framework indicates that fifteen factors were the most relevant for developing the gully erosion susceptibility map, where 7.4% of the agricultural land in the study area was found at *elevated* risk of developing gully erosion. Slope, land cover, organic matter, seasonal LAI, and maximum daily precipitation were the most contributing environmental factors to the study area. Furthermore, this study identified the importance of high temporal resolution in varying seasonal factors (i.e., leaf area index and precipitation) to improve model predictability compared to annual temporal discretization.

1. Introduction

Soil erosion is the detachment, mobilization, and deposition of soil particles occurring asymmetrically across landscapes and streams by wind or water under the combination of complex processes. In general, soil erosion by water, also known as water erosion, can be classified into three processes following its development: sheet, rill, and gully erosion. Among the three, gully erosion has been regarded as the most severe phase (Frankl et al., 2013; Zegeye et al., 2017), forming large channels that are challenging to manage with conventional agricultural machinery decreasing crop production (Cama et al., 2020). As a result, once gullies are developed, their management is challenging, lengthy, and costly. In this regard, designing and implementing proactive measures to prevent gully formation became critical, thereby predicting the probable locations of gully formation, the first step in directing conservation strategies (Popp et al., 2000; Conoscenti et al., 2014).

A variety of studies on predicting gully erosion have been attempted through many physical-based models, including CREAMS (Chemicals, Runoff, and Erosion from Agricultural Management Systems), EGEM (Ephemeral Gully Erosion Model), MIKE SHE, and WEPP (Water Erosion Prediction Project) (Knisel, 1980; Flanagan and Nearing, 1995; Refsgaard and Knudsen, 1996; Woodward,

* Corresponding author.

E-mail address: jag@illinois.edu (J.A. Guzman).

1999). These models require extensive input data (e.g., climate, topography, soil, and time series of detailed farm operation information) that are not commonly available. In addition, these models cannot integrate the necessary spatio-temporal resolution of data and surface dynamics to simulate the entire process accurately on large scales nor evaluate gully erosion susceptibility due to the computational complexity and the insufficient understanding of gully formation (Pourghasemi et al., 2017; Vanmaercke et al., 2021). Due to these drawbacks of physical-based models, there is a need for computational tools to efficiently predict gully erosion across the local, regional, and global scales, finding a good balance between accurate process representation, feasible data requirements, and computational capacity (Vanmaercke et al., 2016; Bauer et al., 2021).

As alternatives, data-driven or probabilistic statistical approaches based on machine learning models have become widely applied to estimate gully susceptibility (Angileri et al., 2016; Rahmati et al., 2017; Arabameri et al., 2020a; Cama et al., 2020). These approaches can solve problems with scarce data with lesser expertise to reproduce the aimed outputs (Amiri et al., 2019). Among many machine learning models, the Maximum Entropy (MaxEnt) model has been widely used to predict environmental issues, including gully erosion (Leblanc et al., 2005; Davis and Sims, 2013; Dube et al., 2014; Xu et al., 2020), although it was initially developed to study habitat suitability by determining the probability of the presence of species (Phillips et al., 2006). Most of the previous research on gully susceptibility prediction mainly focused on steady-state variables in time (e.g., topographic and pedologic factors) without accounting for the temporal variability of significant factors pertinent to gully erosion, such as precipitation and surface cover (Bou Kheir et al., 2007; Li and McCarty, 2018; Li et al., 2018; Zabihi et al., 2018). Although some studies applied temporal factors, including precipitation and the normalized difference vegetation index (NDVI) (Botero-Acosta et al., 2017; Pourghasemi et al., 2017; Arabameri et al., 2020b), the values of temporal factors in their research were on an annual basis, and this temporal scale is inappropriate to reflect the seasonal variability of drivers that can exacerbate or mitigate gully development (Gomez Gutierrez et al., 2009).

In this regard, the main objective of this study was to develop a framework to predict the geospatial location of gully erosion in agricultural lands accounting for the spatio-temporal changes in environmental conditions. Specifically, this study was geared to (1) develop a modeling framework that integrates environmental factors driving gully erosion at a high spatial and temporal resolution and (2) develop and evaluate MaxEnt models to identify the environmental factors that are theoretical and computationally efficient to estimate gully susceptibility, and (3) create a gully erosion susceptibility map and evaluate its linkage with the environmental features.

2. Materials and methods

2.1. Study area

The modeling framework was developed, tested, and evaluated in Jefferson County, Illinois, USA, located in the southern part of Illinois, covering 1512 km² (Fig. 1a). Rolling topography is found across the County with local relief of about 15 m. Upland prairies are flat to moderately hilly and shallow valleys (Leighton et al., 1948). From east to the west, three-fourths of the county area lies in the Big Muddy River Watershed, draining towards the Rend Lake at Jefferson and Franklin Counties (Leighton et al., 1948; Fitzjerrells and Lueker, 2009). The Rend Lake Dam is used for flood control, water supply, recreation, fish, and wildlife. As a stimulus for economic redevelopment (Nieto and Donath, 1976), siltation and water quality have been of great importance. The climate in Jefferson County is characterized by hot, dry summers ranging from 15 °C to 32 °C and cold, wet winters ranging from –6–4 °C (NOAA, 2020).

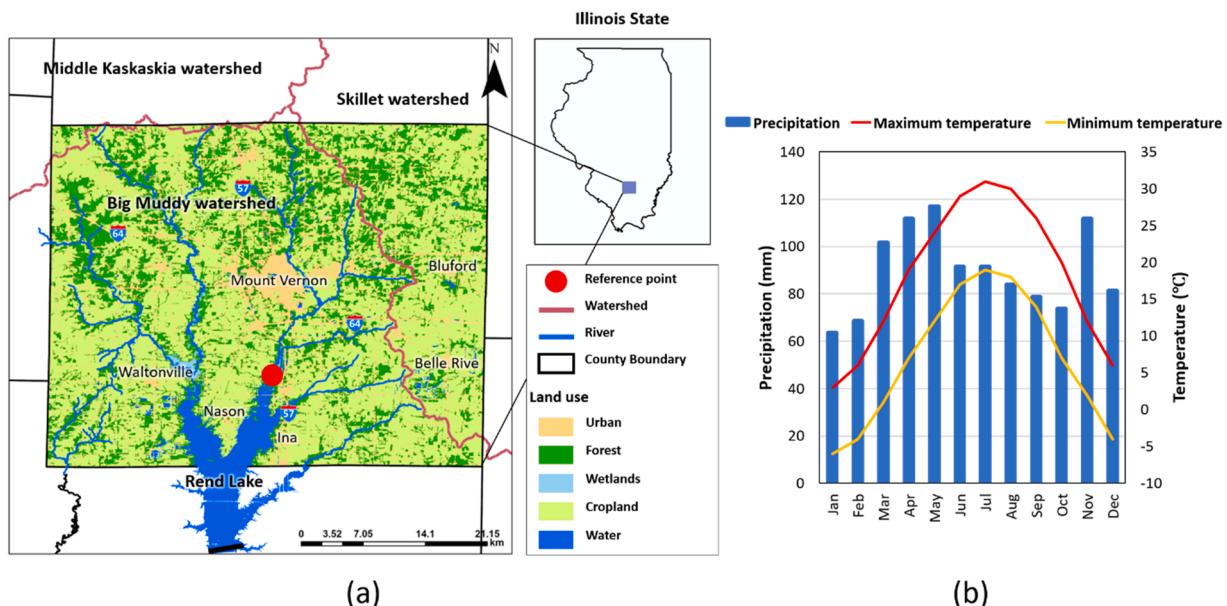


Fig. 1. Geographical location, land use, and climate of the study area.

Precipitation follows a bimodal pattern with peaks in May and November and a total annual average precipitation of 1117.6 mm (Fig. 1b). The most rainfall happens during spring (middle Mar-Jun; 30 %), while 20 % occurs during the fall season (middle Sep-Dec). The major land uses in Jefferson County are agriculture (59 %), forest (27 %), urban (10 %), and water (4 %) based on the 2012 Cropland Data Layer (CDL) product of the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) (Boryan et al., 2011), with agricultural areas dominated by row-crop production (corn and soybean). The area of this study serves to characterize the agricultural activity in the Corn Belt region dominated by intense row crop production across corn, soybean, and wheat crops, with similarities in soil genesis and land management practices.

2.2. Maximum entropy model (MaxEnt)

The MaxEnt model is based on the maximum entropy theory with a machine learning approach applying a presence-only algorithm modeling phenomenon when information about their absence is insufficient or unattainable (Merow et al., 2013). MaxEnt uses a set of presence data of a phenomenon (e.g., gully erosion) and a set of environmental factors to spatially predict an unknown probability distribution. However, MaxEnt can suffer selection bias as sample collection is often spatially biased toward easily accessed areas to measure or observe a target variable here, gully erosion (Phillips et al., 2009). The approach proposed by this study is not affected by this issue as gully locations were detected using remotely sensed data, which is not limited to location access.

In MaxEnt, a given space X is represented in discrete grids where x_i represents the cell in the X domain where a targeting phenomenon (e.g., gully erosion) has been observed (Phillips et al., 2004). The most contributing environmental variable was allocated to each cell x_i , in which the probability of occurrence $P(x)$, called the Relative Occurrence Rate (ROR), was calculated using Eq. (1):

$$P_z(x_i) = \frac{e^{\lambda z(x_i)}}{\sum_i e^{\lambda z(x_i)}} \quad (1)$$

where z is a vector of the predictor at location x_i , and λ is a vector of regression coefficients. The calculated RORs are normalized values; consequently, their sum over the study area equals 1. The MaxEnt model employs the receiver operating characteristic (ROC) curves and the AUC (area under the ROC curve) to evaluate model performance. A ROC analysis shows how sensitivity (TRUE positive rate) changes against specificity (1 – FALSE-positive rate) over a series of thresholds, which can be summarized by an AUC (Muschelli, 2019). An AUC ranging from 0.5 to 0.6 is considered poor, while values closer to 1 are considered a more reliable metric performance when predicting the probability of the phenomenon's presence (here, gully erosion) (Rahmati et al., 2017). More details on the mathematical process can be found in Phillips et al. (2004) and Phillips and Dudík (2008a).

2.3. Methodology

The study was divided into four main parts (Fig. 2). First, a gully erosion inventory map was created by subtracting two light detection and ranging (LiDAR)-derived digital elevation models (DEMs). The differences in elevation were then filtered for noise, and those exceeding a given threshold were used as the “presence” of gully erosion in the MaxEnt model. Second, the environmental factors controlling gullies were identified, based on literature, and re-constructed using the different data sources. The importance of all factors was then investigated by performing the permutation test, and the major driving factors within each factor group (topography, pedology, land cover, and precipitation) were selected (objective 1). Third, a series of the MaxEnt models were developed with different combinations of the selected major factors, among which the best environmental factors were determined (objective 2). Lastly, using the model with the best factor set, a gully erosion susceptibility map was developed, and the geospatial attributes of

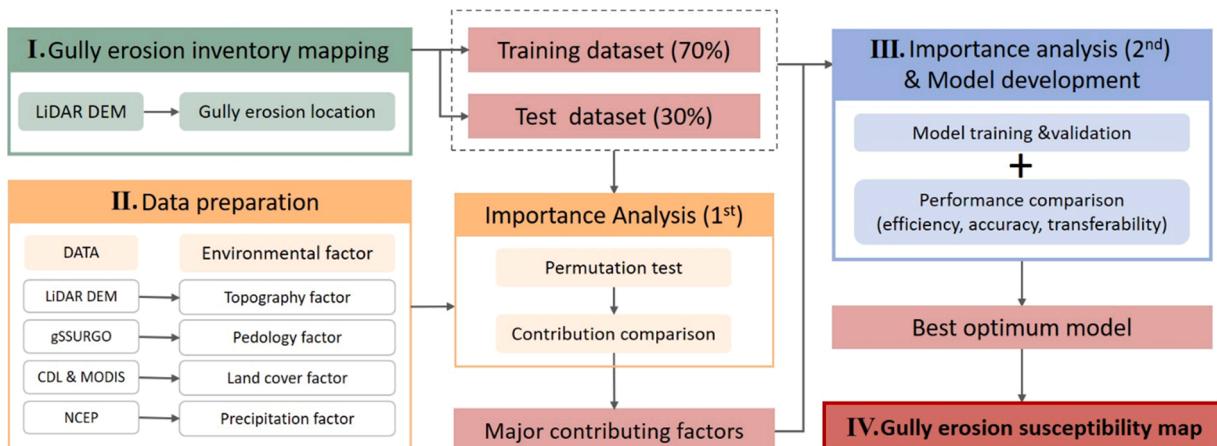


Fig. 2. Methodology and data flow used in the present study. LiDAR is light detection and ranging; gSSURGO is gridded soil survey geographic; CDL is cropland data layers; MODIS is moderate resolution imaging spectroradiometer; NCEP is the National Center for Environmental Prediction.

susceptible areas were explored (objective 3). The details about each process are described in the following sections.

2.3.1. Gully erosion inventory mapping

The gully erosion inventory in agricultural lands based on the Crop Data Layer (CDL) in 2012 was mapped using two LiDAR-derived DEMs at a 2 m (2012) and 1 m (2015) resolution, assuming that agricultural areas and cultivated crops in 2012 were maintained until 2015. The two DEMs, corrected to eliminate human-made structures or vegetations, were obtained from the Illinois State Geological Survey (ISGS; <https://clearinghouse.isgs.illinois.edu/>). The identification of gully erosion was derived from the difference in elevations between the two DEMs (Martinez-Casasnovas et al., 2002; Evans and Lindsay, 2010) using a 2-m grid resolution. The paved road (E. Bonnie Rd) elevation in Rend Lake (red dot in Fig. 1a) was used as a reference for the DEM calibration, assuming no elevation change happened on this road between 2012 and 2015. Ten points were randomly chosen, and the average difference in elevation between two DEMs was regarded as a noise caused by the inherent discrepancies between the two LiDAR datasets. The average elevation difference of the ten points was 6.2 cm and was subtracted from the difference between the two DEMs. A threshold value of 50 cm was applied to consider the presence of gully erosion only (Lafren et al., 1986; Lafren and Shaw, 1988) because the processed DEM map may also contain sheet and rill erosion. The 50-cm depth was estimated based on deep tillage depth (e.g., 33 cm; Raper et al., 2000) and the vertical accuracy of LiDAR DEMs (approximately 20 cm at a 95 % confidence level; Authority, 1998). As LiDAR's accuracy can be affected by shadows, affecting the wavelength intensity of the returned signal from the surface to a sensor, a thorough examination was conducted using aerial imageries of 2012 and 2015 acquired from the National Agriculture Imagery Program (NAIP) by USDA. This analysis aimed to remove locations with human activities that result in significant changes in elevation and cloud perturbation (e.g., shadows) in the LiDAR datasets (Guislain et al., 2016; Yüksel and Boyaci, 2018). The final estimated location of gully erosion was transformed into a point dataset to be used as the presence of gully erosion for the MaxEnt model and assumed to represent the probable location of gullies regardless of their size, type, or stage of development (i.e., both ephemeral and classical or permanent gullies). For cross-validation, 70 % of the gully sample points were randomly selected as a training dataset, and 30 % were used for model validation.

Table 1

Summary of the environmental factors used to predict the probability of the presence of gully erosion.

Factors (spatial resolution)	Description
Topography factors (2 m)	
Flow accumulation	Upslope number of cumulated cells
Slope-length factor, L.S.	Erosive power of the terrain
Topographic wetness index, TWI	Frequencies and duration of saturated conditions
Aspect	Angle between north and horizontal projection of a normal external vector at a given point (categorical factor)
Slope (degree)	Angle between a tangent and a horizontal plane at a given point
Plan curvature (m^{-1})	Curvature in a horizontal plane
Profile curvature (m^{-1})	Curvature of the surface in the direction of the steepest slope
Standard curvature (m^{-1})	Curvature of the surface itself
Stream power index, SPI	Erosive power of overland flow
Pedology factors (10 m)	
Clay content (%)	Soil particles that are less than 0.002 mm in diameter
Silt content (%)	Soil particles that are 0.002–0.05 mm in diameter
Sand content (%)	Soil particles that are 0.05–2 mm in diameter
Organic matter (%)	The plant and animal residue in the soil at various stages of decomposition
Available water capacity, AWC ($cm\ cm^{-1}$)	The quantity of water that the soil can store for use by plants
Saturated hydraulic conductivity, K_{sat} ($m\ s^{-1}$)	The ease with which pores in a saturated soil transmits water
Erodibility factor, K factor	The susceptibility of a soil to sheet and rill erosion by water
Bulk density ($g\ cm^{-3}$)	The oven-dry weight of the soil material less than 2 mm in size per unit volume of soil at a water tension of 1/3 bar
Soil texture	USDA textural classes which are determined by the size and proportion of the particles (clay, silt, and sand)
Soil loss tolerance factor, T factor	An estimate of the maximum average annual rate of soil erosion by a wind that can occur without affecting crop productivity over a sustained period (categorical factor)
Hydrologic soil group, HSG	Soils are assigned to one of four groups (A/B/C/D) according to the rate of water infiltration (i.e., runoff potential) with no vegetation (categorical factor)
Drainage class	The frequency and duration of wet periods under conditions similar to those under which the soil formed (categorical factor)
Land cover factors	
Annual leaf area index, LAI (500 m)	A dimensionless quantity that characterizes plant canopies; an average value during the corresponding period
Seasonal leaf area index, LAI (500 m)	Subclasses of agriculture classification (categorical factor)
Land use (30 m)	
Precipitation factors (1 km)	
Seasonal average daily precipitation, Ave_P (mm)	Average of the daily precipitation during each season
Seasonal maximum daily precipitation, Max_P (mm)	Maximum of the daily precipitation during each season
Seasonal total precipitation, Tot_P (mm)	The total amount of precipitation during each season
Annual precipitation, Annual_P (mm)	Annual total precipitation

2.3.2. Construction of environmental factors

To consider the spatial-temporal effects of the environmental features on gully erosion, topographic and pedologic factors were considered stationary factors, while land cover and precipitation were used as time-varying factors in daily, seasonal, and annual time series (Table 1). A set of nine topographic factors and a set of twelve pedologic factors were selected based on literature (Valentin et al., 2005; Bou Kheir et al., 2007; Tuo et al., 2015; Biddouci et al., 2017; Botero-Acosta et al., 2017; Lenhart et al., 2017; Li and McCarty, 2018; Li et al., 2018; Kariminejad et al., 2019; Cama et al., 2020). Topographic factors were extracted from the LiDAR-derived DEM of 2012. The pedology factors were extracted from the Gridded Soil Survey Geographic (gSSURGO) database (NRCS, 2021). The gSSURGO database, a gridded form of the standard USDA-NRCS soil survey geographic (SSURGO) database, contains the tabular data of various soil properties and derivatives to a state-wide extent. More details about this dataset can be found in NRCS (2021). The pedologic factors were mapped using the Soil Data Development Toolbox using the *muKey* gSSURGO database attribute that links soil attributes to their geolocation. Note that in this study, the representative soil depth for soil mapping was set to the upper 50 cm to consider the impacts of conventional tillage depth (Etana et al., 1999; Raper et al., 2000).

Land use and leaf area index (LAI) were selected as land cover factors. The annual and seasonal LAI were considered to represent the spatio-temporal dynamics of the plant canopy. LAI maps were produced using the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI product (MCD15A2H) from 2012 to 2015, an eight-day composite LAI data with a spatial resolution of 500 m (Myeni et al., 2015). Precipitation factors were mapped using the National Centers for Environmental Prediction (NCEP) stage IV daily precipitation datasets, assimilating radar and gauge observations at a 1-km spatial resolution. The following factors were computed from 2012 to 2015 to account for the temporal variability of precipitation: (1) seasonal average daily precipitation (Ave_P), (2) seasonal maximum daily precipitation (Max_P), (3) seasonal total precipitation (Tot_P), and (4) annual precipitation (Ann_P). Seasons were defined based on the climate regime and cultural practices as Spring (March to May), Summer (June to August), Fall (September to November), and Winter (December to February). The annual and seasonal factors within the land cover and precipitation factors were compared to identify the temporal effects on gully erosion prediction.

2.3.3. Relative importance analysis of environmental factors

The relative importance of each factor within each factor group was analyzed through a permutation test. The permutation test measures the decrease in the accuracy of the final training result caused by the variable's permutation, which is normalized as an index for contribution analysis (Phillips and Dudík, 2008b). Regardless of the number of factors within factor groups, a threshold value of 0.05 was used to select the significant factors. The selection of a threshold value of 0.05 provides a satisfactory feature selection accuracy (Prasetyowati et al., 2021). In addition, the effects of the seasonal variance (i.e., temporal variability) in land cover and precipitation factors on gully erosion prediction were investigated. Based on the analysis of environmental factors, the major contributing factors were selected for each factor group.

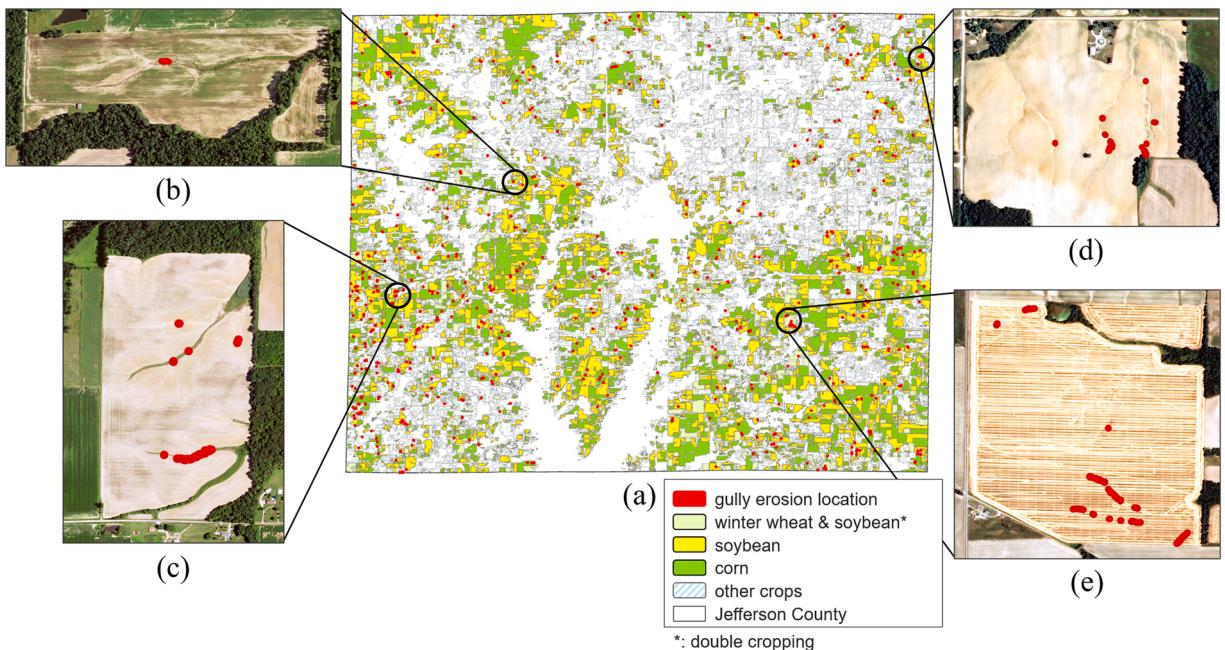


Fig. 3. (a) Gully erosion inventory map obtained from LiDAR-derived DEMs in agricultural lands. (b) - (e) aerial images indicating the detailed location of LiDAR estimated gullies (red dots).

2.3.4. Development and evaluation of the gully erosion susceptibility map

The multivariate model evaluation was built from baseline variables (i.e., land use, slope, Max_P, and organic matter) incremented by one variable per model and then performed a model evaluation, computed the relative degree of importance, and the model prediction accuracy. The best model was selected based on the following set of rules: (1) minimum number of applied variables (i.e., efficiency), (2) maximum model performance (i.e., accuracy), and (3) minimum AUC difference between the training and test (i.e., transferability). Following the best model selection, the probability of gully presence for agricultural lands was estimated and mapped in a 2×2 m spatial resolution and then classified into four susceptibility classes using the natural breaks (Jenks) method. The natural breaks method groups data to reduce the variance within classes and maximize the variance between classes (Jenks, 1967). Chalkias et al. (2014) demonstrated that the natural breaks method is more efficient in susceptibility mapping than other classification methods (e.g., equal interval and standard deviation classifications), and it has been the most used in susceptibility assessment (Chalkias et al., 2014; Chen et al., 2015; Sema et al., 2017; Hong et al., 2018). As a final step, each factor was evaluated through response curves to assess the change in the estimated probability of gully presence as a function of a change in each factor (sensitivity) while other factors remain fixed at their average values.

3. Results and discussions

3.1. Gully erosion location in the study area

A total of 5621 gully erosion locations in agricultural fields were identified over the study area (Fig. 3a). These gully locations were assumed as 'true' gully presences and used as the MaxEnt model's input. A visual evaluation of gully locations was conducted using the 2012 and 2015 aerial images to confirm their validity (Fig. 3b-e). Generally, gully locations were more pronounced in the southern part of Jefferson County, near the Rend Lake reservoir, where row-crop fields (i.e., corn and soybean) are more prevalent than in the northern part (Table 2).

3.2. Importance analysis of pedologic and topographic factors

The results of the importance analysis of pedologic factors showed that *K* factor, silt content, organic matter, and AWC had a contribution greater than 10 %, and the *K* factor had the highest contribution (19.1 %) to gully prediction (Fig. 4a). Among twelve factors, seven factors with a contribution greater than 5 % (*K* factor, silt content, organic matter, AWC, HSG, clay content, and *T* factor) were selected. In the case of the topography factors, the slope acquired the highest contribution (47.2 %) to gully prediction, followed by TWI, aspect, standard curvature, and profile curvature (Fig. 4b). Such high contributions of these factors were found in agreement with previous studies that the slope, curvature, aspect, and TWI are reported to show a high association with gully development (Li and McCarty, 2018; Li et al., 2018; Cama et al., 2020). The five topographic factors showing a contribution higher than 5 % were selected as the primary factors.

3.3. Temporal effects of the environmental factors on gully prediction

The contribution of the seasonal factors was compared with the non-seasonal factors (e.g., land use, annual LAI, and annual precipitation). In the importance analysis for land cover factors (Fig. 4c), the contribution of land use was higher than the annual LAI and seasonal LAI, accounting for approximately 70 % of the total contribution. Such significance of the land-use factor was reasonable since gully erosion (Fig. 3) was mainly found in areas with row-crop production of corn, soybean, and winter wheat (Table 2). Furthermore, the higher importance of land cover than LAI follows existing models, accounting for land use and land cover, rather than relative vegetation indices, such as NDVI and LAI. This might be because different crops require different cultivation practices (e.g., tillage, planting, irrigation), and some of this information is embedded in land cover data (Vanmaercke et al., 2021).

The comparison of the seasonal LAI with the annual LAI showed that the former's contribution (24.3 %) was approximately five times greater than the latter's (4.6%) (Fig. 4c). This difference is because the seasonal LAI better represents the spatial heterogeneity of LAI distribution than the annual LAI. The LAI in spring and fall, in particular, had a more heterogeneous distribution than annual LAI, while LAI in summer and winter presented relatively monotonic distributions (Fig. 5). Given the dominant contribution of the land use

Table 2
Distribution of gully erosion location based on land use subclasses within agricultural lands.

Classification	Gully erosion location (%)	Ag. Practice
Soybean	3094 (55.04)	C.R. ^a , 5Y ^b
Corn	1695 (30.15)	C.R. ^a , 5Y ^b
Winter wheat & soybean	618 (10.99)	C.R. ^a , 1Y ^b
Other hay (non-alfalfa)	210 (3.74)	—
Winter wheat & corn	3 (0.05)	C.R. ^a , 1Y ^b
Winter wheat	1 (0.02)	—
Total	5621 (100)	

^a C.R. denotes crop rotation;

^b #Y denotes the average rotation period in years

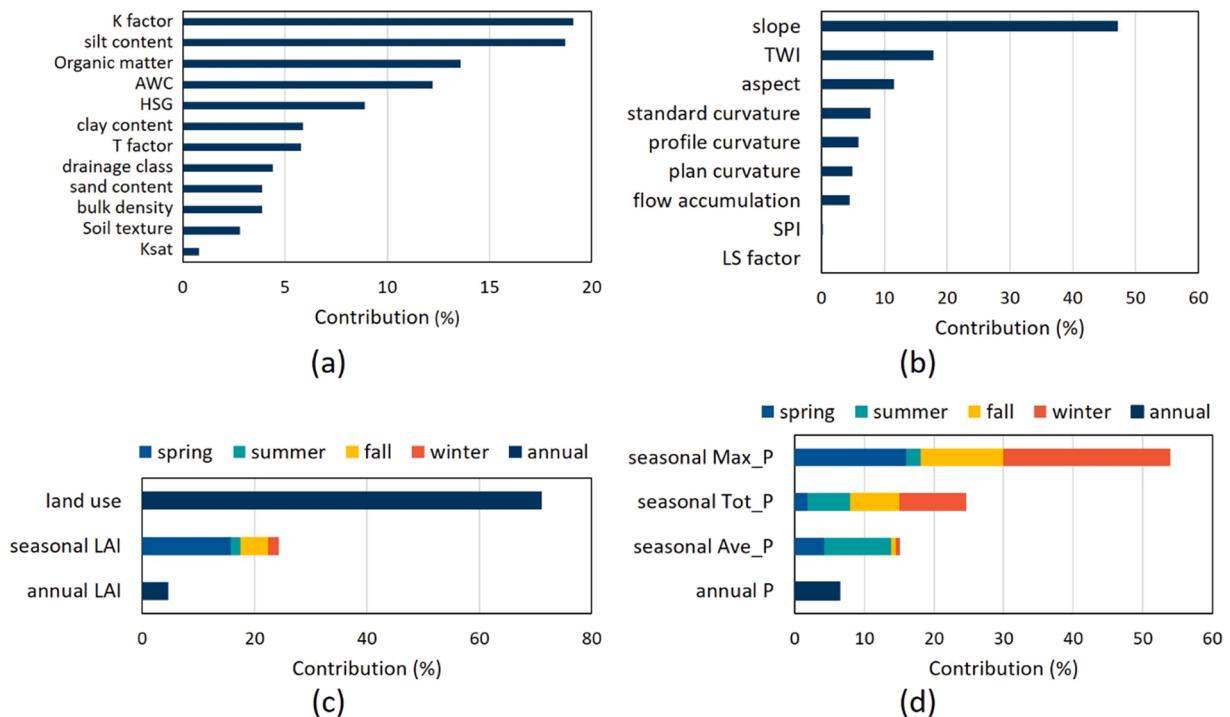


Fig. 4. Importance analysis results using the permutation test: (a) pedologic factors, (b) topographic factors, (c) land cover factors, and (d) precipitation factors (K factor is erodibility factor; AWC is available water capacity; HSG is hydrologic soil group; T factor is soil loss tolerance factor; K_{sat} is saturated hydraulic conductivity; TWI is topographic wetness index; SPI is stream power index; LS factor is a slope-length factor; LAI is leaf area index; Max_P is seasonal daily maximum precipitation; Tot_P is seasonal total precipitation, and Ave_P is seasonal daily average precipitation).

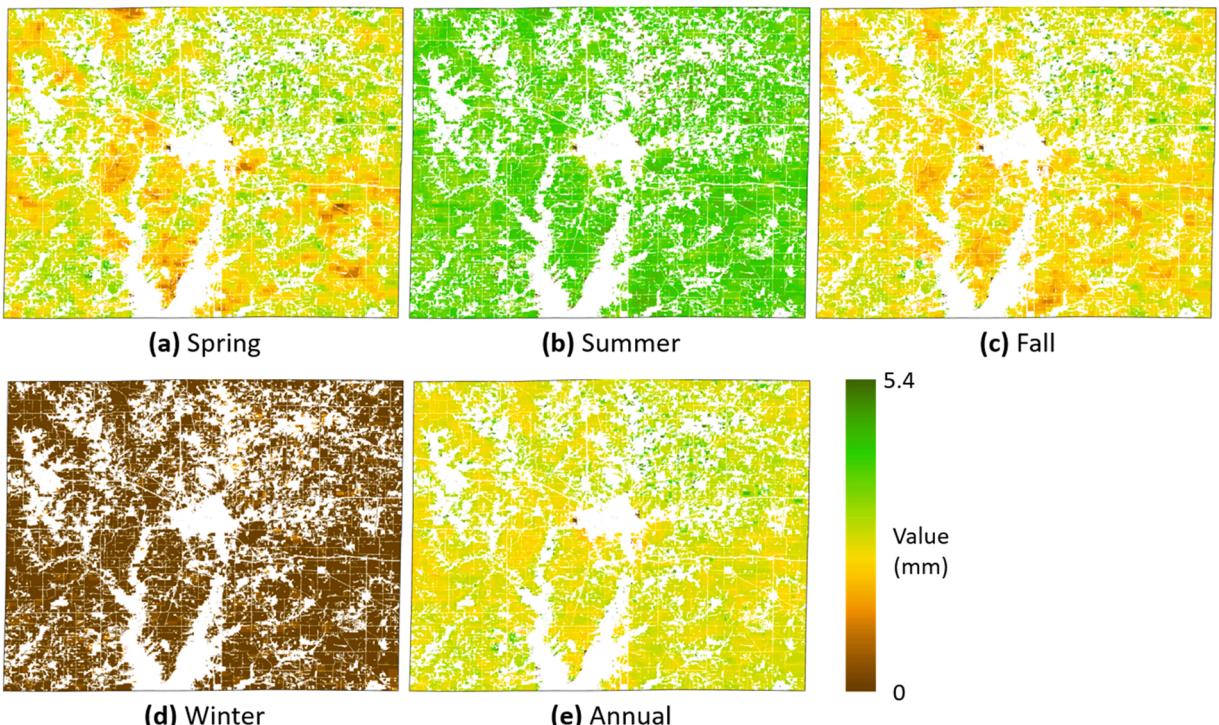


Fig. 5. Spatial distribution of the seasonal and annual leaf area index (LAI).

and its correlation with the LAI (Cohen et al., 2003), the seasonal LAI's high contribution implied that it encoded additional information on the spatial heterogeneity caused by vegetation, which the land use dataset does not provide. The importance of LAI was due in part to farming operations (e.g., tillage, planting, and harvesting) during spring (March and April) and fall (September and October). Crop growth, especially at the early stage, varies depending on the conditions of the surrounding environment, including soil moisture, temperature, solar radiation, and seed formulation (Sacks et al., 2010), and this led to the most noticeable spatial heterogeneity and the highest contribution (15.7 %) of the LAI in spring (Fig. 5a). In contrast, summer and winter presented less spatial variation. The reason is that row crops, in most cases, can reach maturity before the end of summer, covering the surface until harvest. Moreover, most agricultural fields are bare during winter except for areas planted with cover crops (Fig. 5b, d). Thus, seasonal LAI reasonably characterized such temporal and spatial changes in the land surface due to agricultural activities.

All distributions of seasonal precipitation showed spatial and temporal differences (Fig. 6). For example, Max_P during spring and summer was likely to be more intense in the southeast part of the study area but relatively milder in fall and winter. Such spatio-temporal changes in precipitation patterns may lead to a higher contribution of seasonal precipitation factors than annual precipitation (Fig. 4d). The Max_P had the highest contribution among the seasonal factors as precipitation intensity is critical among the various rainfall characteristic descriptors (e.g., duration, total depth, and frequency). Such a significant effect of rainfall intensity on

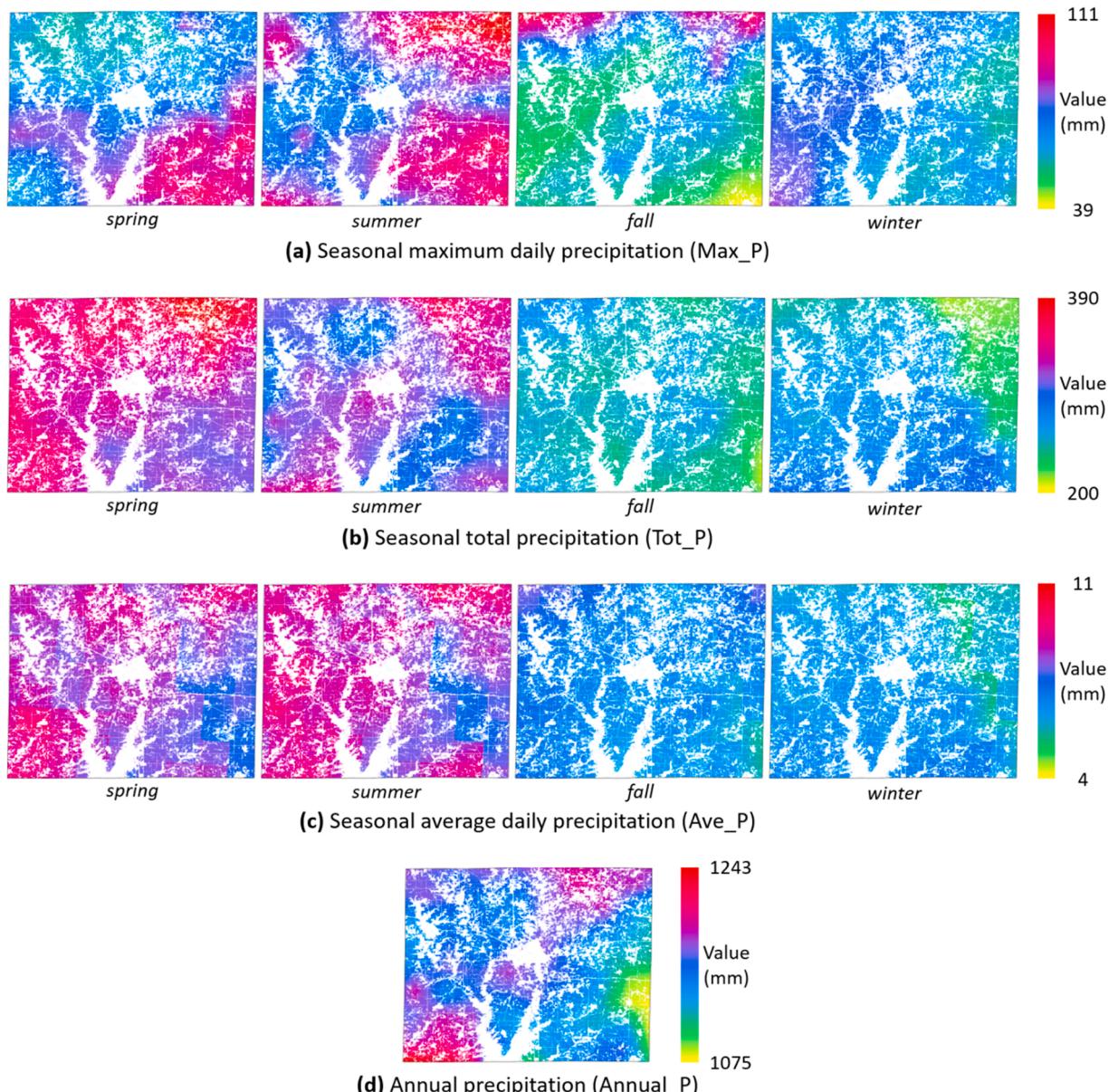


Fig. 6. Spatial distribution of the seasonal and annual precipitation factors.

gully development has been demonstrated by numerous studies (Vanmaercke et al., 2021). Rainfall intensity exacerbates particle detachment and mobilization due to increased raindrop energy and the likelihood of developing concentrated flow (Vanmaercke et al., 2016). In summary, land use and seasonal LAI in the land cover group and seasonal Max_P and seasonal Tot_P in the precipitation group were selected as major contributing factors in predicting the probability of gully erosion.

3.4. Environmental factors, model development, and model performance evaluation

Among all factors with a more significant contribution than 5 %, land use showed the highest contribution with 20.1 %, followed by slope, Max_P (fall and winter), organic matter, and AWC (Fig. 7). It is noted that each of these factors belongs to a different environmental group, implying that each environmental group carries unique information linked to gully erosion development. Another noteworthy finding is that different factor combinations can change the importance of factors. For example, the K factor had the lowest importance among the selected pedologic factors. However, it presented the highest importance when only pedologic factors were compared (Fig. 4a). This was explained as a result of collinearity or multicollinearity, a situation where two or more predictors are highly associated. This factor dependency suggests that even a factor relevant to a one-factor group (e.g., topography) can provide information to some degree to other factor groups (e.g., soil). However, it is challenging to reveal the underlying mechanisms and interactions that resulted in the change in the factor importance due to the black box nature of machine learning, a well-known constraint of machine learning techniques (Vanmaercke et al., 2021).

Table 3 summarizes the model evaluation and the criteria for model selection. It was considered that the higher the number of factors, the lower the model efficiency, the higher the AUC for training and testing, the higher the model accuracy, and the smaller the AUC difference, the higher the transferability. In general, factor incremental led to increased AUC for training and test, and thereby, the training AUC reached a plateau value of 0.828 at model 'MM_q'. Beyond the model 'MM_h', no significant improvement in accuracy was observed. Among the models, the 'MM_a' model indicated the highest transferability with an AUC difference of 0.004. Given the model selection criteria, the model 'MM_l' was selected as the best factor combination.

Although a direct comparison of the current study with the existing studies is challenging due to the differences in splitting strategies for training/test set partitioning, spatial resolution, and environmental conditions, the AUC metric can be used for comparison. The developed model MM_l shows better performance than other models developed in previous studies for gully susceptibility using an entropy-based model (e.g., MaxEnt) in most cases. For example, while AUC from the model MM_l was less accurate than the model developed by Kariminejad et al. (2019) (AUC of 0.958), it showed a higher AUC value than those of Arabameri et al. (2020a), Bernini et al. (2021), Pourghasemi et al. (2017) and Zabihi et al. (2018). Given that the mentioned studies used comparable predictors, the improved accuracy in the present study is likely attributed to the accounting of seasonal variability (i.e., LAI and precipitation).

3.5. Analysis of the gully erosion susceptibility concerning environmental factors

The response curves for the factors (i.e., sensitivity) with importance higher than 5% are indicated in Fig. 8. Note that there are

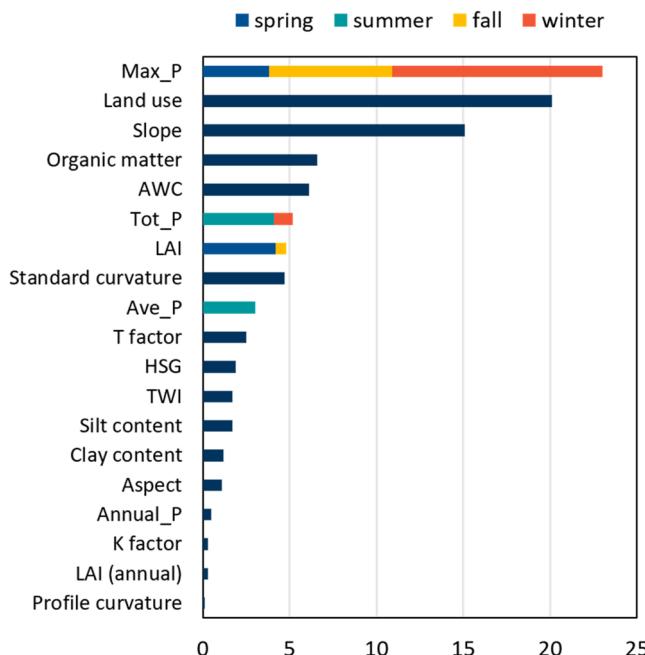


Fig. 7. Contribution analysis of the combined factors selected as a major contributing factor from each environmental factor group.

Table 3

Model performance evaluation based on a different factor combination. The blue box highlights the best model selected.

Model	Applied factor	Number of factors	AUC		AUC difference
			training	test	
MM _a	Land use, Slope, Max_P (winter), Organic matter	4	0.757	0.753	0.004
MM _b	MM _a + Max_P (fall)	5	0.779	0.773	0.006
MM _c	MM _b + AWC	6	0.781	0.775	0.005
MM _d	MM _c + Standard curvature	7	0.796	0.791	0.005
MM _e	MM _d + LAI (spring)	8	0.806	0.797	0.009
MM _f	MM _e + Tot_P (summer)	9	0.814	0.807	0.007
MM _g	MM _f + Max_P (spring)	10	0.818	0.810	0.008
MM _h	MM _g + Ave_P (summer)	11	0.820	0.813	0.007
MM _i	MM _h + T factor	12	0.820	0.813	0.006
MM _j	MM _i + HSG	13	0.820	0.813	0.007
MM _k	MM _j + Silt content	14	0.822	0.815	0.007
MM _l	MM _k + TWI	15	0.824	0.819	0.005
MM _m	MM _l + Clay content	16	0.825	0.819	0.006
MM _n	MM _m + Aspect	17	0.827	0.821	0.006
MM _o	MM _n + Tot_P (winter)	18	0.827	0.821	0.006
MM _p	MM _o + LAI (fall)	19	0.827	0.821	0.006
MM _q	MM _p + Annual_P	20	0.828	0.822	0.006
MM _r	MM _q + K factor	21	0.828	0.822	0.006
MM _s	MM _r + LAI (annual)	22	0.828	0.823	0.006
MM _t	MM _s + Profile curvature	23	0.828	0.823	0.005

The model MM_l, consisting of fifteen predictive factors, agrees with variables commonly found in other soil erosion models. For example, the Water Erosion Prediction Project Model (WEPP), a physical-based model, requires watershed topography (DEM, slope length and steepness, and aspect), soil properties (soil texture, depth, and erodibility), land use, and climate (precipitation and temperature).

irregular, abrupt peaks or troughs at specific factor values indicating overfitting or over-parameterization, one of the major concerns in machine learning models like MaxEnt, that can lead to predictive biases in the spatial distribution of environmental factors (Liu et al., 2018). A strategy to mitigate this problem is model simplification aimed at reducing the number of environmental factors with a sufficient sample size (Merow et al., 2013). However, results from this study indicated that the factor distribution presented discontinuities even though a set of rules were applied for model reduction. Therefore, it is recommended to focus on the overall trends indicated in the dashed lines in Fig. 8 when interpreting the sensitivity curves.

Land use had a range of p from 0.12 to 0.50, where 'other hay' has the highest probability of developing gully erosion, and grassland has the lower probability. All other land uses did not present prominent differences. Regarding organic matter and AWC, the model indicates that the risk of developing gully erosion is inversely proportional to their increase. Among other reasons, this behavior was found most especially for organic matter and explained by the linkages of soil organic matter with the reduction of runoff generation and increased rainfall infiltration due to its protecting capacity of soil sealing and crusting (Roth et al., 1985). In addition, the increase in biological activity increases macroporosity and improves the soil's water holding capacity (Bot and Benites, 2005). On the other hand, an opposite behavior (directly proportional increase) was found for slope and Max_P. Slope and Max_P are critical drivers in soil erosion, favoring the development of concentrated flow and particle detachment and runoff generation, respectively.

3.6. Development of the gully erosion susceptibility map

The gully erosion susceptibility map was developed based on the probability calculated by the selected model 'MM_l' (Fig. 9). The susceptibility was classified into four classes based on the natural breaks method, where the boundary values between classes were rounded for convenience' sake: (1) *low*: $0 < p \leq 0.15$, (2) *moderate*: $0.15 < p \leq 0.30$, (3) *high*: $0.30 < p \leq 0.50$, and (4) *elevated*: $0.50 < p \leq 1.0$. A total of 75.3 % of the agricultural areas fell in the *moderate* (29.4%) and *low* (46.0 %) categories (Table 4), while 24.6 % were found in the *high* (15.7 %) and *elevated* (8.9 %) categories. An evaluation of the identified gully erosion locations (i.e., 5621)

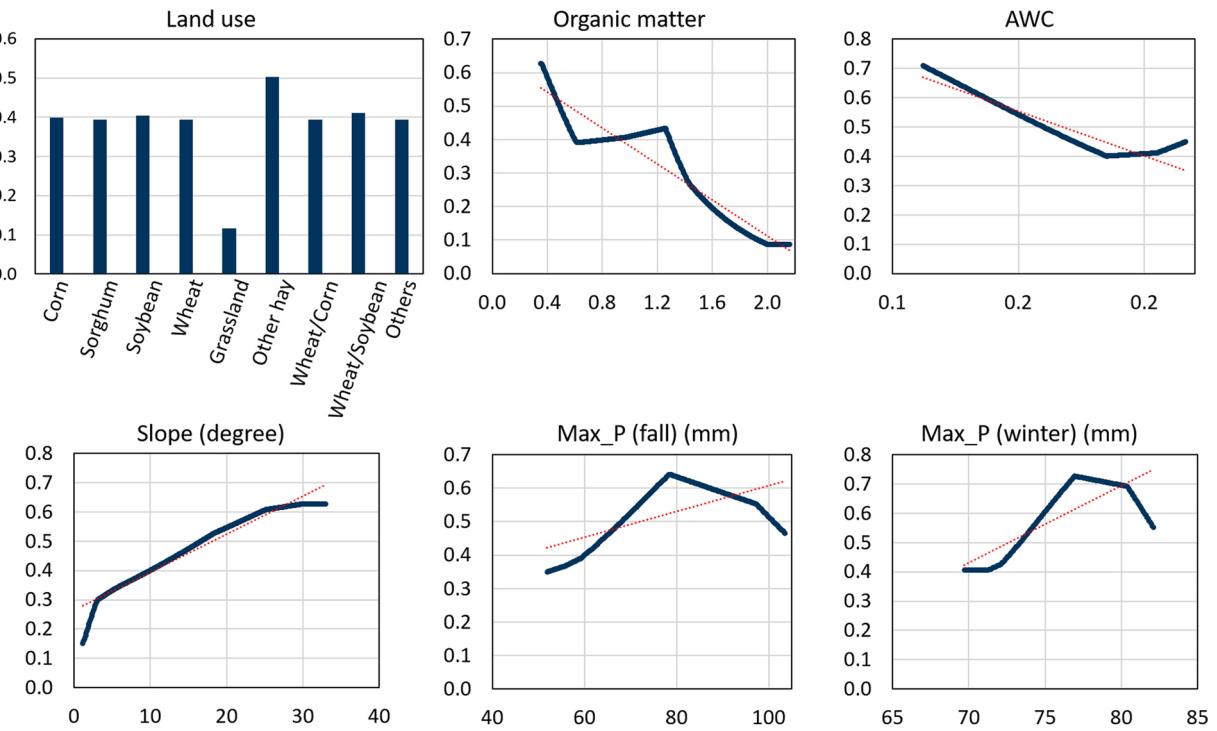


Fig. 8. Response curves of gully erosion susceptibility conditional to environmental factors. Y-axis is the probability of gully erosion. The X-axis is the range of values for the corresponding factor, and the red line is a trend line. Categorical variables are represented by bars and continuous variables by lines.

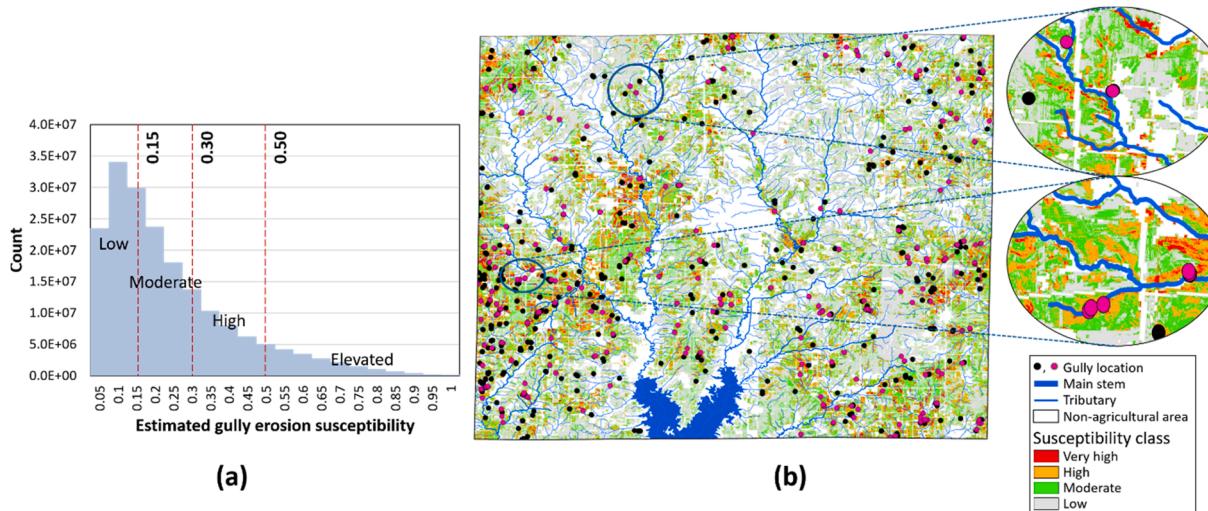


Fig. 9. Classes of gully erosion susceptibility (*low*, *moderate*, *high*, and *elevated*) based on the distribution of the predicted susceptibility. (a) frequency distribution of the estimated gully erosion susceptibility, and (b) gully erosion susceptibility map. Dots represent gully erosion locations obtained from LiDAR-derived DEMs, where green dots denote gully locations within 100 m from streams.

indicated that 90.5% of these locations (black and green dots in Fig. 9b) fell in the *high* (14.9 %) and *elevated* (75.6 %) categories. Note that as the erosion location was flagged as “*true*” gully erosion, it was expected that 100 % of the observed gullies were found in the *high* and *elevated* categories. However, 9.5 % of the gully locations (i.e., 533) were classified as *moderate* and *low* risk.

An analysis of the susceptibility map indicates that 73 % of the gully susceptible areas in the *high* and *elevated* categories are primarily located in corn and soybean fields (Fig. 3). In addition, fields near streams were identified with a higher vulnerability of gully development as the *elevated* category is mainly located adjacent to small streams (Fig. 9b), where approximately 50 % of gully locations

Table 4

The area of susceptibility classes and the number of gully erosion locations in the class.

Susceptibility class	Area (percent)	# LiDAR-derived gully sites (percent)
Low	410 km ² (46.0 %)	1199 (2.1 %)
Moderate	262 km ² (29.4 %)	414 (7.4 %)
High	140 km ² (15.7 %)	839 (14.9 %)
Elevated	79 km ² (8.9 %)	4249 (75.6 %)

were found nearby 100 m from small streams (e.g., green dots in Fig. 9b). Such patterns were found reasonable considering that erosion occurring near streams is driven by the occurrences of the high water table, seepage, and the development of concentrated flow in the proximity of water bodies. Furthermore, gully head retreat, streambank erosion driven by streamflow, and seepage fluxes can exacerbate gully formation (Allen et al., 2018; Botero-Acosta et al., 2017; Zabihi et al., 2018).

An evaluation of the model performance for each land use indicates that approximately 82 % of the total gully locations were well predicted across all categories (Fig. 10). However, the model poorly performed in 'other hay' and 'grassland' with a misprediction rate exceeding 30% (marked as a red square in Fig. 10). Given the limited area of these land use types within the study area, this was explained by insufficient information on these subcategories (i.e., poor quality or biased data), a common limitation of a data-driven model. For example, 'other hay' land use accounts for only 1.8 % of the total study area. Therefore, it is likely that this category does not carry enough information for model training and validation compared to the 'Corn' and 'Soybean' categories, which account for 39 % and 47 % of the total area, respectively.

4. Conclusions

Predicting areas vulnerable to gully erosion is essential to establishing practical and efficient mitigation plans (e.g., crop production, infrastructure, archaeology). In this study, twenty-eight gully erosion controlling factors were extracted from readily available data (LiDAR-derived DEM, gSSURGO, and remote-sensing data) and were used to develop a predictive model aimed at identifying vulnerable areas for gully development. Among all factors considered, the slope, land use, seasonal daily maximum precipitation, and organic matter indicated the highest contribution in predicting the presence of gullies. Furthermore, the spatio-temporal changes in land cover and precipitation were crucial across all seasons when predicting gully formation in agricultural areas. The proposed modeling framework was developed in Jefferson County, IL, and found that approximately 7.4 % of the agricultural land in the study area is at an *elevated* risk of gully erosion. This approach can be applied across the Midwest U.S. region, sharing similar land management and environmental variables.

Despite the model's reliable result, the locations of gully erosion regarded as the "true" presences datasets were extracted from the LiDAR-derived DEMs without field verification due to the gap in time between LiDAR acquisition, gully occurrences, and research. For this reason, the gully erosion locations were expected to include a degree of uncertainty, which may propagate to the model outputs, even though extensive scrutiny using high-resolution imageries was performed. In addition, the accuracy of land use data may add to the source of uncertainty given the discrepancy between the actual and LiDAR or remote sensing surveying conditions. Hence, quantifying data quality and uncertainty through field verification could enhance the validity of this study. This study proposed a cost-effective and time-efficient framework to develop a gully susceptibility mapping on large scales (e.g., regional) with readily available data. Various gully driving environmental factors were explored and associated with an area susceptible to gully erosion. The proposed framework will help policymakers and land managers focus their prevention or mitigation efforts, directing resources where they are most needed.

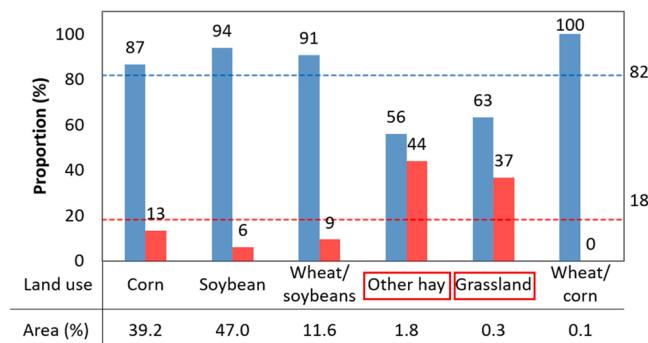


Fig. 10. Model performance evaluation based on land use. Dotted lines represented the average proportion of well-predicted (blue) and mispredicted (red) gully locations. The red squares represent subcategories with a more than 30 % mispredicted proportion.

CRediT authorship contribution statement

Jeongho Han: Conceptualization, Methodology, Software, Data curation, Writing – original draft preparation. **Jorge A. Guzman:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Maria L. Chu:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The authors do not have permission to share data.

Acknowledgments

The U.S. Department of Agriculture provided funding for this research - National Institute for Food and Agriculture (NIFA) award number 2019-67019-29884. NCAR/EOL provided Radar-derived QPE data (NCEP Stage IV) under the sponsorship of the National Science Foundation (<https://data.eol.ucar.edu/>). LiDAR data was provided by Illinois Geospatial Data Clearinghouse (<https://clearinghouse.isgs.illinois.edu/frontpage>), supported by the Prairie research institute, University of Illinois at Urbana-Champaign.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejrh.2022.101196](https://doi.org/10.1016/j.ejrh.2022.101196). These data include Google maps of the most important areas described in this article.

References

- Allen, P.M., Arnold, J.G., Auguste, L., White, J., Dunbar, J., 2018. Application of a simple headcut advance model for gullies. *Earth Surf. Process. Landf.* 43, 202–217. <https://doi.org/10.1002/ESP.4233>.
- Amiri, M., Pourghasemi, H.R., Ghanbarian, G.A., Afzali, S.F., 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.
- Angileri, S.E., Conoscenti, C., Hochschild, V., Märker, M., Rotigliano, E., Agnesi, V., 2016. Water erosion susceptibility mapping by applying stochastic gradient treeboost to the Imera Meridionale river basin, Sicily, Italy. *Geomorphology* 262, 61–76.
- Arabameri, A., Nalivan, O.A., Saha, S., Roy, J., Pradhan, B., Tiefenbacher, J.P., Ngo, P.T., 2020a. Novel ensemble approaches of machine learning techniques in modeling the gully erosion susceptibility. *Remote Sens.* <https://doi.org/10.3390/rs12111890>.
- Arabameri, A., Pradhan, B., Bui, D.T., 2020b. Spatial modelling of gully erosion in the Arid River Watershed using three statistical-based techniques. *Catena* 190.
- Authority, T.V., 1998. Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy. Natl Aeronaut Sp Adm Virginia, NV, USA.
- Bauer, P., Dueben, P.D., Hoefer, T., Quintino, T., Schulthess, T.C., Wedi, N.P., 2021. The digital revolution of earth-system science. *Nat. Comput. Sci.* 1, 104–113.
- Bernini, A., Bosino, A., Botha, G.A., Maerker, M., 2021. Evaluation of gully erosion susceptibility using a maximum entropy model in the upper mkhomazi river basin in south africa. *ISPRS Int. J. Geo-Inf.* 10 <https://doi.org/10.3390/IJGI10110729>.
- Biddoccu, M., Ferraris, S., Pitacco, A., Cavallo, E., 2017. Temporal variability of soil management effects on soil hydrological properties, runoff and erosion at the field scale in a hillslope vineyard, North-West Italy. *Soil Tillage Res.* <https://doi.org/10.1016/j.still.2016.07.017>.
- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring U.S. agriculture: the U.S. department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int.* 26, 341–358.
- Bot, A., Benites, J., 2005. The importance of soil organic matter: Key to drought-resistant soil and sustained food production, No. 80. Food & Agric Org.
- Botero-Acosta, A., Chu, M.L., Guzman, J.A., Stark, P.J., Moriasi, D.N., 2017. Riparian erosion vulnerability model based on environmental features. *J. Environ. Manag.* 203, 592–602. <https://doi.org/10.1016/j.jenvman.2017.02.045>.
- Cama, M., Schillaci, C., Kropáček, J., Hochschild, V., Bosino, A., Märker, M., 2020. A probabilistic assessment of soil erosion susceptibility in a head catchment of the Jemma basin, Ethiopian highlands. *Geosci.* <https://doi.org/10.3390/geosciences10070248>.
- Chalkias, C., Ferentinos, M., Polykretis, C., 2014. GIS-based landslide susceptibility mapping on the Peloponnese Peninsula, Greece. *Geosciences* 4, 176–190.
- Chen, X., Chen, H., You, Y., Liu, J., 2015. Susceptibility assessment of debris flows using the analytic hierarchy process method—a case study in Subao river valley, China. *J. Rock Mech. Geotech. Eng.* 7, 404–410.
- Cohen, W.B., Maiersperger, T.K., Yang, Z., Gower, S.T., Turner, D.P., Ritts, W.D., Berterretche, M., Running, S.W., 2003. Comparisons of land cover and LAI estimates derived from ETM+ and MODIS for four sites in North America: a quality assessment of 2000/2001 provisional MODIS products. *Remote Sens. Environ.* 88, 233–255.
- 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*. <https://doi.org/10.1016/j.geomorph.2013.08.021>.
- Davis, J.D., Sims, S.M., 2013. Physical and maximum entropy models applied to inventories of hillslope sediment sources. *J. Soils Sediment.* 13, 1784–1801.
- Dube, F., Nhapi, I., Murwira, A., Gumindoga, W., Goldin, J., Mashauri, D.A., 2014. Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District-Zimbabwe. *Phys. Chem. Earth Parts A/B/C* 67, 145–152.
- Etana, A., Håkansson, I., Zagal, E., Bucas, S., 1999. Effects of tillage depth on organic carbon content and physical properties in five Swedish soils. *Soil Tillage Res.* [https://doi.org/10.1016/S0167-1987\(99\)00062-8](https://doi.org/10.1016/S0167-1987(99)00062-8).
- Evans, M., Lindsay, J., 2010. High resolution quantification of gully erosion in upland peatlands at the landscape scale. *Earth Surf. Process Landf.* 35, 876–886. <https://doi.org/10.1002/esp.1918>.
- Fitzjerrells, B., Lueker, S., 2009. Hazard Mitigation Plan Jefferson County, Illinois Polis, 618, pp. 453–7370.
- Flanagan, D.C., Nearing, M.A., 1995. USDA-water erosion prediction project: hillslope profile and watershed model documentation. *Nserl Rep.* 10, 1–123.

- Frankl, A., Poesen, J., Haile, M., Deckers, J., Nyssen, J., 2013. Quantifying long-term changes in gully networks and volumes in dryland environments: The case of Northern Ethiopia. *Geomorphology*. <https://doi.org/10.1016/j.geomorph.2013.06.025>.
- Gomez Gutierrez, A., Schnabel, S., Felicísimo, Á.M., 2009. Modelling the occurrence of gullies in rangelands of southwest Spain. *Earth Surf. Process Landf. J. Br. Geomorphol. Res Gr.* 34, 1894–1902.
- Guislain, M., Digne, J., Chaine, R., Kudelski, D., Lefebvre-Albaret, P., 2016. Detecting and correcting shadows in urban point clouds and image collections. In: *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*. Institute of Electrical and Electronics Engineers Inc., 537–545.
- Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X., Xu, C., 2018. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. *The case of Dayu County, China. Sci. Total Environ.* 630, 1044–1056.
- Jenks, G.F., 1967. The data model concept in statistical mapping. *Int. Yearb. Cartogr.*, 7, pp. 186–190.
- Kariminejad, N., Hosseinielzadeh, M., Pourghasemi, H.R., Bernatek-Jakiel, A., Campetella, G., Ownegh, M., 2019. Evaluation of factors affecting gully headcut location using summary statistics and the maximum entropy model: Golestan Province, NE Iran. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2019.04.306>.
- Kheir, R.B., Wilson, J., Deng, Y., 2007. Use of terrain variables for mapping gully erosion susceptibility in Lebanon. *Earth Surf. Process. Landf.* <https://doi.org/10.1002/esp.1501>.
- Knisel, W.G., 1980. CREAMS: A field scale model for chemicals, runoff, and erosion from agricultural management systems. Department of Agriculture, Science and Education Administration.
- Laflen, J.M., Shaw, R.R., 1988. Ephemeral Gully Erosion Model, (EGEM) Version 1.1 User Manual. USDA Soil Conserv Serv Washington, DC, USA.
- Laflen, J.M., Watson, D.A., Franti, T.G., 1986. Ephemeral gully erosion. In: *Proceedings of the Fourth Federal Interagency Sedimentation Conference March 24–27, 1986, Las Vegas, Nevada*.
- Leblanc, S.G., Chen, J.M., Fernandes, R., Deering, D.W., Conley, A., 2005. Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agric. Meteorol.* <https://doi.org/10.1016/j.agrmet.2004.09.006>.
- Leighton, M.M., Ekblaw, G.E., Horberg, L., 1948. Physiographic divisions of illinois. *J. Geol.* <https://doi.org/10.1086/625474>.
- Lenhart, C., Gordon, B., Peterson, J., Eshenaur, W., Gifford, L., Wilson, B., Stamper, J., Krider, L., Utt, N., 2017. Agricultural BMP Handbook for Minnesota. Minnesota Dep Agric.
- Li, X., McCarty, G.W., 2018. Use of principal components for scaling up topographic models to map soil redistribution and soil organic carbon. *J. Vis. Exp.* <https://doi.org/10.3791/58189>.
- Li, X., McCarty, G.W., Karlen, D.L., Cambardella, C.A., 2018. Topographic metric predictions of soil redistribution and organic carbon in Iowa cropland fields. *Catena* 160, 222–232. <https://doi.org/10.1016/j.catena.2017.09.026>.
- Liu, Y., Zhou, K., Xia, Q., 2018. A MaxEnt model for mineral prospectivity mapping. *Nat. Resour. Res.* <https://doi.org/10.1007/s11053-017-9355-2>.
- Martínez-Casanovas, J.A., Ramos, M.C., Ribes-Dasi, M., 2002. Soil erosion caused by extreme rainfall events: Mapping and quantification in agricultural plots from very detailed digital elevation models. *Geoderma* 105, 125–140. [https://doi.org/10.1016/S0016-7061\(01\)00096-9](https://doi.org/10.1016/S0016-7061(01)00096-9).
- Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.
- Muschelli, J., 2019. ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. <https://doi.org/10.1007/s00357-019-09345-1>.
- Myneni, R., Knyazikhin, Y., Park, T., 2015. MCD15A2H MODIS/Terra+ Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.
- Nieto, A.S., Donath, F.A., 1976. Report of a Study of Structural Geology and Subsidence of the Rend Lake Dam Area, Franklin and Jefferson Counties, Illinois, for U.S. Army Engineer District, St. Louis Corps of Engineers, St. Louis, Missouri. CGS, Incorporated.
- NOAA, National Oceanic and Atmospheric Administration. (<https://www.ncdc.noaa.gov/>) (accessed 15 December 2020).
- NRCS, Natural Resources Conservation Service, 2021. Soil Survey Staff. Gridded Soil Survey Geographic (gSSURGO) database for the conterminous United States. *Nat. Resour. Conserv. Serv.* (<https://gdg.sc.egov.usda.gov/>) (accessed 10 January 2021).
- Phillips, S.J., Dudík, M., 2008a. A Brief Tutorial on MaxEnt. AT&T Research.
- Phillips, S.J., Dudík, M., 2008b. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>.
- Phillips, S.J., Dudík, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. In: *Twenty-first international conference on Machine learning - ICML '04*. ACM Press, New York, New York, USA, pp. 83.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19 (1), 181–197.
- Popp, J.H., Hyatt, D.E., Hoag, D., 2000. Modeling environmental condition with indices: a case study of sustainability and soil resources. *Ecol. Model.* 130, 131–143.
- Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerdà, A., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.*
- Prasetyowati, M.I., Maulidevi, N.U., Surendro, K., 2021. Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest. *J. Big Data* 8, 1–22. <https://doi.org/10.1186/S40537-021-00472-4/TABLES/2>.
- Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* 298, 118–137.
- Raper, R.L., Reeves, D.W., Burmester, C.H., Schwab, E.B., 2000. Tillage depth, tillage timing, and cover crop effects on cotton yield, soil strength, and tillage energy requirements. *Appl Eng Agric.*
- Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. *Water Resour. Res.* 32, 2189–2202.
- Roth, C.H., Meyer, B., Frede, H.G., 1985. A portable rainfall simulator for studying factors affecting runoff, infiltration and soil loss. *Catena* 12 (1), 79–85.
- Sacks, W.J., Deryng, D., Foley, J.A., Ramankutty, N., 2010. Crop planting dates: An analysis of global patterns. *Glob. Ecol. Biogeogr.* <https://doi.org/10.1111/j.1466-8238.2010.00551.x>.
- Sema, H.V., Guru, B., Veerappan, R., 2017. Fuzzy gamma operator model for preparing landslide susceptibility zonation mapping in parts of Kohima Town, Nagaland, India. *Model Earth Syst. Environ.* 3, 499–514.
- Tuo, D., Xu, M., Zhao, Y., Gao, L., 2015. Interactions between wind and water erosion change sediment yield and particle distribution under simulated conditions. *J. Arid Land* 7, 590–598.
- Valentin, C., Poesen, J., Li, Y., 2005. Gully erosion: impacts, factors and control. *Catena*. Elsevier, pp. 132–153.
- Vanmaercke, M., Poesen, J., Van Mele, B., Demuzere, M., Bruynseels, A., Golosov, V., Bezerra, J.F.R., Bolysov, S., Dvinskikh, A., Frankl, A., Fuseina, Y., Guerra, A.J.T., Haregewyn, N., Ionita, I., Makanzu Imwanga, F., Moeyersoons, J., Moshe, I., Nazari Samani, A., Niacsu, L., Nyssen, J., Otsuki, Y., Radoane, M., Rysin, I., Ryzhov, Y.V., Yermolaev, O., 2016. How fast do gully headcuts retreat? *Earth-Science Rev.*
- Vanmaercke, M., Panagos, P., Vanwallegem, T., Hayas, A., Foerster, S., Borrelli, P., Rossi, M., Torri, D., Casali, J., Borselli, L., Vigiak, O., Maerker, M., Haregewyn, N., de Geeter, S., Zglobicki, W., Bielders, C., Cerdà, A., Conoscenti, C., de Figueiredo, T., et al., 2021. Measuring, modelling and managing gully erosion at large scales: a state of the art. *Earth-Sci. Rev.* 218, 103637 <https://doi.org/10.1016/J.EARSCIREV.2021.103637>.
- Woodward, D.E., 1999. Method to predict cropland ephemeral gully erosion. *Catena* 37, 393–399.
- Xu, N., Meng, F., Zhou, G., Li, Y., Wang, B., Lu, H., 2020. Assessing the suitable cultivation areas for *Scutellaria baicalensis* in China using the Maxent model and multiple linear regression. *Biochem. Syst. Ecol.* 90, 104052 <https://doi.org/10.1016/j.bse.2020.104052>.

- Yüksel, S.E., Boyaci, M., 2018. Effect of LiDAR sensor on the success of shadow detection from hyperspectral data. *Pamukkale Univ. J. Eng. Sci.* 24, 198–204. <https://doi.org/10.5505/pajes.2016.13281>.
- Zabih, M., Mirchooli, F., Motavalli, A., Khaledi Darvishan, A., Pourghasemi, H.R., Zakeri, M.A., Sadighi, F., 2018. Spatial modelling of gully erosion in Mazandaran Province, Northern Iran. *Catena*. <https://doi.org/10.1016/j.catena.2017.10.010>.
- Zegeye, A.D., Steenhuis, T.S., Mekuria, W., Dagnaw, D.C., Addisse, M.B., Tilahun, S.A., Kasse, T.A., 2017. Effect of gully headcut treatment on sediment load and gully expansion in the sub humid ethiopian highlands. *Environ. Ecol. Res.* <https://doi.org/10.13189/eer.2017.050208>.