Journal of Geography, Environment and Earth Science International

Volume 28, Issue 11, Page 148-162, 2024; Article no.JGEESI.126403
ISSN: 2454-7352

Empirical and Machine Learning Models for Soil Erosion Risk Assessment: A Case Study of Tsageri Municipality, Georgia

Nika Tkeshelashvili [a*]

[a] *Department of Engineering Geodesy and Geoinformatics, Georgian Technical University, Tbilisi, Georgia.*

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: https://doi.org/10.9734/jgeesi/2024/v28i11843

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/126403

Original Research Article

Received: 09/09/2024
Accepted: 11/11/2024
Published: 16/11/2024

ABSTRACT

Soil erosion caused by water is one of the most common causes of land degradation worldwide. Within framework of this research soil erosion risk in Tsageri municipality, Georgia was evaluated using Revised Universal Soil Loss Equation (RUSLE) and a machine learning-based Random Forest (RF) model. Open access digital datasets and field observations collected in 2023-2024, which included visually identified erosion areas and GPS-recorded data on the presence or absence of erosion, were utilized in modeling process. Data processing and modeling conducted using ArcGIS Pro 3.0 and RStudio software. According to RUSLE results, 39.7% of the study area falls under the very low erosion risk zone, and 20.7% is in the very high risk zone. The RF model results indicated that 16.5% of the territory is under very low risk of erosion and 13.9% - very high

---

**Corresponding author: E-mail: tkeshelashvili.nika@gtu.ge;*

Cite as: *Tkeshelashvili, Nika. 2024. "Empirical and Machine Learning Models for Soil Erosion Risk Assessment: A Case Study of Tsageri Municipality, Georgia". Journal of Geography, Environment and Earth Science International 28 (11):148-62. https://doi.org/10.9734/jgeesi/2024/v28i11843.*

risk. It was observed that RUSLE model tends to overestimate erosion rates on steep, forested slopes, while the RF model, by incorporating additional variables, provided more accurate prediction. These findings suggest that combining RUSLE with machine learning improves soil erosion risk assessment, particularly in complex landscapes such as in Tsageri municipality. Future researche should focus on testing additional variables to refine the modeling process further and enhance predictions. The generated digital thematic maps offer valuable insights for understanding the spatial dynamics of soil erosion within the study area, analyzing the factors driving the process and developing effective mitigation strategies.

Keywords: Soil; erosion; machine learning; Georgia.

1. INTRODUCTION

Soil erosion caused by water is a complex process, influenced by natural and human-made factors and poses significant challenge to the sustainable economic development of many countries (Pimentel et al., 1995; Montgomery, 2007). In this regard Georgia is no exception, particularly in its mountainous regions, where soil erosion issues are amplified by the combination of geological, topographical, climatic, hydrological and pedological conditions. These natural factors, coupled with the effects of human activities, such as logging, overgrazing and poorly managed agricultural practices, create a highly favorable environment for the progression of soil erosion (Poesen, 2018).

To develop effective mitigation strategies for water-induced soil erosion and ensure the sustainable use of natural resources in the context of rapid climate change, it is essential to assess both the rate and extent of erosion quickly and efficiently (Panagos et al., 2015). This requires understanding erosion as a multi-dimensional process, identifying the key contributing factors, establishing clear cause-and-effect relationships and conducting spatial analyses at various scales (Borrelli et al., 2017).

The primary objective of this study is to model water-borne soil erosion and predict soil loss in the Tsageri municipality, a representative mountainous region of Georgia. Notably, for this region and much of the country, systematic soil erosion studies and assessment have not been conducted since the Soviet era, leading to lack of up-to-date data.

To achieve these goals, the study employs the Revised Universal Soil Loss Equation (RUSLE), a widely used empirical model for estimating long-term average annual soil loss (Renard et al., 1997). However, understanding the limitations of RUSLE, particularly in its reliance on empirical parameters that may not adequately capture the spatial heterogeneity of erosion process, the study integrates additional variables and incorporates the Random Forest (RF) algorithm, a machine learning model known for its ability to handle complex, non-linear relationships between the parameters of different phenomena (Breiman, 2001).

Given the advantages of erosion modeling, namely lower costs in terms of time and materials compared to field and laboratory studies, especially over large areas and the absence of in-situ measurements, the modeling process in this study predominantly relies on freely available digital data from scientific databases (e.g., Copernicus, Soil Grids, NASA), supplemented by field observations conducted during 2023-2024. This approach not only addresses the data gaps but also ensures that the methodology is scalable and replicable in other regions of country.

2. MATERIALS AND METHODS

2.1 Study Area

Tsageri municipality (Fig. 1) is located in Racha-Lechkhumi and Kvemo Svaneti region, between the southern branches of the Main Caucasus Range. Covering an area of just over 755 km^2, it constitutes approximately 1% of the country's total territory.

The study area exhibits a diverse soil cover, characterized by a wide variety of alluvial soils distributed throughout the Tsageri depression. On the slopes of the low mountains, brown forest and raw carbonate soils prevail. In the foothills and high mountain areas, brown forest and podzolic brown forest soils are predominant. In the subalpine and alpine zones, mountain-meadow soils are present, which transmits into more primitive mountain-meadow soils at higher elevations. The soil-forming rocks in these areas include crystalline shales, quartz diorites, limestones and gneisses.

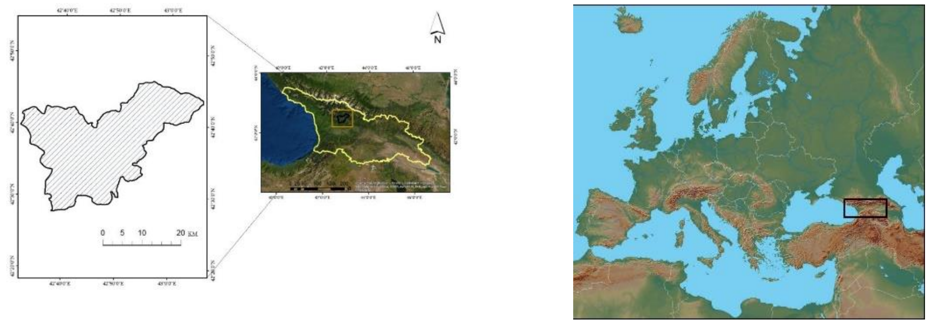


Fig. 1. Study area

Tsageri municipality has a humid subtropical climate with cold winters and long, warm summers. The avarage annual air temperature in the lowland areas is 10-11.5^0 C, while July temperatures range from 20-25^0 C. The recorded absolure minimum is -26^0 C. Annual precipitation is between 1200-1300 mm and reaching up to 2000 mm in the highlands. The primary components of the hidrographic network are the Rioni and Tskhenistskali rivers.

The administrative center of the municipality is the city of Tsageri. According to the 2014 census, the population is 10 387. The territory of the municipality is divided into 19 administrative units. Currently, according to the register of municipalities of Georgia, the number of populated areas is 58 villages and 1 city.

2.2 Methods

Within the study, soil erosion modeling and risk assessment were conducted by integrating the RUSLE model and the Random Forest algorithm. The data required for the modeling were processed using ArcGIS Pro 3.0 and RStudio software tools.

The RUSLE (Revised Universal Soil Loss Equation) model predicts the average annual soil erosion rate for a given area by incorporating five key parameters: 1. Atmospheric precipitation, 2. Soil type, 3. Topography, 4. Vegetation cover, and 5. Erosion control measures.

By integrating the RUSLE with GIS platform, each element is represented in a separate raster data format. The expected average annual erosion is then calculated using the program's mathematical tools. This is expressed by the following equation:

$$A = R * K * LS * C * P \qquad (1)$$

Where - A is soil loss (t/ha per year); R represents the rainfall-runoff erosivity factor (MJ.mm/ha.h.yr); K - soil erodibility factor (t.ha.h/ha.MJ.mm); LS - slope length and steepness factor; C-vegetation cover factor; P - protective measures. LS, C and P factors have no dimension.

To calculate the R factor needed for modeling and assessing the erosive potential of atmospheric precipitation in the study area, 20-year average monthly precipitation data from the Global Precipitation Measurement (GPM) mission were used. This dataset comprises 240 files (GPM_3IMERGM v07) covering the years 2001 to 2020.

As illustrated in (Fig. 2), the GPM data reveal a trend of decreasing average annual precipitation in the study area over the reporting period.

Based on the received precipitation metrics, the R factor was calculated using the following formula:

$$R = 1.735 \times 10^{\left(1.50 \times log \sum_1^{12} \frac{p_i^2}{p} - 0.818\right)} \qquad (2)$$

Where, R represents the erosivity factor of precipitation, Pi and P - the monthly and annual amount of precipitation, respectively (Arnoldus, 1977).

As shown in the (Fig 3), the R factor values calculated using the formula are directly proportional to the amount of annual precipitation.

To calculate soil erodibility (K factor) in the RUSLE model, several physical soil parameters are considered, including texture (the distribution of soil particles by size), organic matter, structure and permeability. Among these, soil texture is the most influential on K factor values. Generally, the K factor reflects the soil's susceptibility to erosion (erodibility) and its conditions for sediment transport.

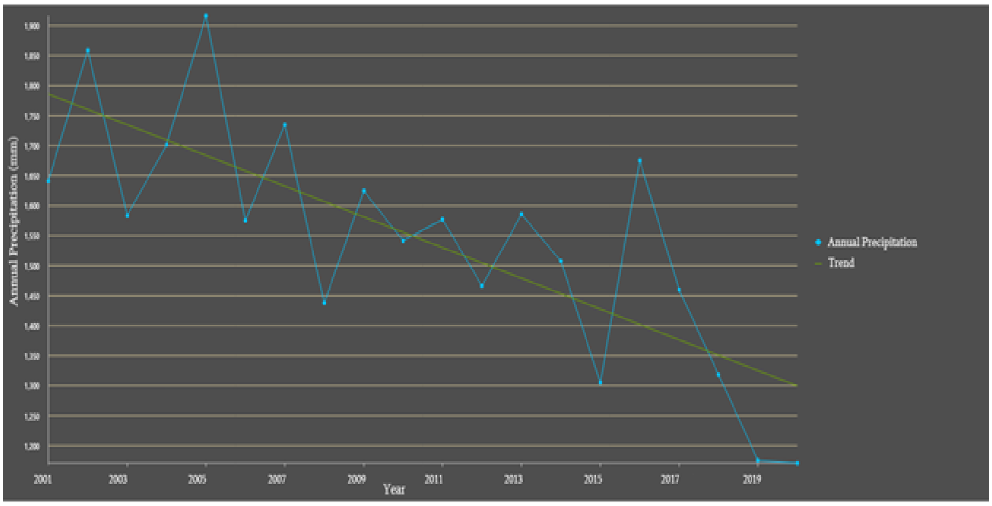


Fig. 2. Annual precipitation in the study area (GPM 2001-2020)

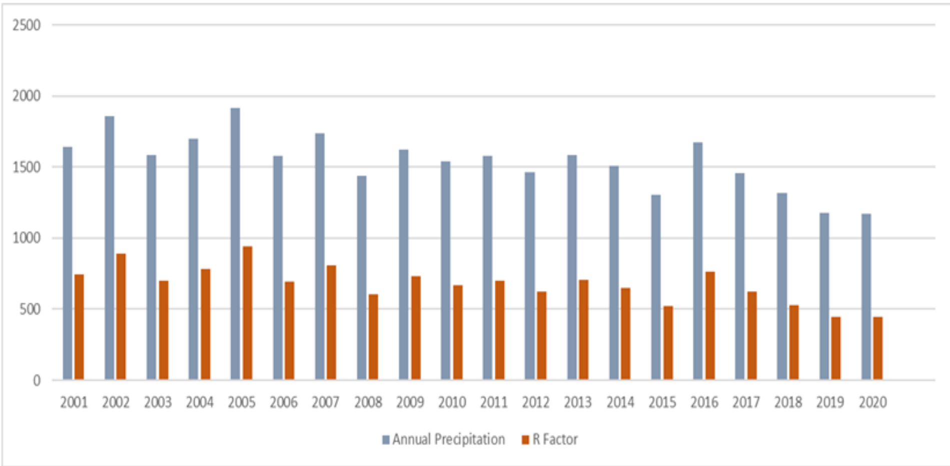


Fig. 3. Annual precipitation and obtained R factor values in the study area

Soil texture is determined by the proportional distribution of sand, silt, and clay particles in soil profiles. Surface texture, specifically within the top 20-30 centimeters, is especially important in terms of erosivity.

The soil parameters required to calculate the K factor for the study area, which in our case are expressed by the granulometric composition of the soil and the content of organic matter, were obtained from Soil Grids (Poggio et al., 2021) data for soil horizons up to 30 centimeters. This data includes indicators of silt, sand, clay and organic matter in the soils within the study area.

To calculate the K factor values, we use the following formula (Sharply & Williams, 1990):

$$K = \left(0.2 + 0.3e^{\left[-0.0256SAN\left(1-\frac{SIL}{100}\right)\right]}\right) \times \left(\frac{SIL}{CLA+SIL}\right)^{0.3} \times \left[1 - \frac{0.25C}{C+e^{(3.72-2.95C)}}\right] \times \left[1 - \frac{0.7SN_1}{SN_1+e^{(22.9SN_1-5.51)}}\right] \tag{3}$$

where, SAN is the sand content (%), SIL – the silt content (%), CLA – the clay content (%), C – organic substances (%), $SN_1 = 1 - SAN/100$.

LS factor values for Tsageri municipality are calculated by adapting the following formula (Moore & Burch, 1986):

$$LS = \left(Flow\ accumulation \times \frac{Cell\ Size}{22.13}\right)^{0.4} \times \left(\frac{\sin Slope}{0.0896}\right)^{1.3} \quad (4)$$

Where, LS is the slope length and inclination factor (dimensionless), Flow accumulation - A raster file with the flow accumulation values for each pixel, Cell size is the pixel size (in this case 12.5 m, depending on the spatial resolution of the DEM), sin Slope - sine of slope inclination (in degrees), and 0.4 and 1.3 are the empirical members of the formula (Moore & Wilson, 1992).

The C factor was quantified by applying the Normalized Difference Vegetation Index (NDVI) values derived from the Sentinel-2 imagery, into the following formula:

$$C = e^{\left[-a\frac{NDVI}{(\beta - NDVI)}\right]} \quad (5)$$

where a and β are dimensionless members that describe the shape of the curve representing the relationship between NDVI and the C factor. Their values are defined as 2 and 1, respectively.

The soil protection measures factor (P) is the most experimental component of the model and reflects the effect of various measures that reduce the volume and speed of surface runoff thereby decreasing the risk of erosion.

To determine the P factor over large areas, Wener's method is used. This is expressed by the formula (Wenner, 1981):

$$P = 0.2 + 0.3 \times S \quad (6)$$

where S is the inclination of the slope in percent. The P factor has no dimension.

The slope inclination (%) in the study area was determined based on the DEM, according to which P factor values was generated using the above formula.

The development and progression of soil erosion are affected by various geo-environmental parameters not considered in the RUSLE model. These include slope exposure, river network density, distance to roads and rivers, lithological characteristics, and others (Bouramtane et al., 2022).

To incorporate these parameters into the modeling process and enhance the accuracy of predicting the average annual soil loss in the study area, the Random Forest algorithm was further adapted.

Random Forest is an ensemble-based, non-parametric machine learning algorithm widely used to address ecological and geospatial problems such as water resources management, natural hazard control etc., in recent times, since its versatility allows for the analysis of various data types. For example, it can be satellite images, numerical and statistical data and others.

The Random Forest algorithm is based on decision trees and combines broad, regression and classification tree groups. Two fundamental parameters are required to construct a random forest model: 1. The number of trees in the forest, defined as "n-tree"; and 2. The number of features considered at each split, known as „m-try”. Classification trees provide the ability for individual trees to vote or make predictions, as a result of which one or the other class is determined according to the regulation of maximum rating across the entire forest.

In addition to the 5 factors used in the RUSLE model, digital maps of 11 factors (Fig. 5) were prepared based on available data for the study area, such as: Slope exposure, plane curvature, altitude, distance to road, distance to river, flow density, lithology, NDVI, slope inclination, stream power index (SPI) and topographic wetness index (TWI).

Slope exposure indicates the orientation of the slope relative to cardinal directions and is derived from processing the digital elevation model. Exposure determines the orientation from which the slope receives solar radiation, affecting temperature, evaporation, and moisture levels. South-facing slopes generally receive more direct sunlight, resulting in higher temperatures and evaporation rates compared to north-facing slopes. Soils on southern slopes typically have lower moisture content, making them more prone to erosion. Additionally, slope exposure influences vegetation growth, water flow formation, and accumulation, all of which can further impact erosion processes.

The curvature of the plane reflects the forms of terrain and is also derived from the processing of the DEM.

The altitude of given location affects the distribution of precipitation, the formation and accumulation of runoff. As elevation increases, the amount of atmospheric precipitation increases under the influence of orographic factors. This obviously leads to the presence of more water with potential to develop erosion. In such cases, the share of snowmelt water also increases, which influence seasonal changes in soil moisture content. The highly fragmented terrain and sparse vegetation, characteristic of high mountain areas, are also a contributing factor to water-borne soil erosion.

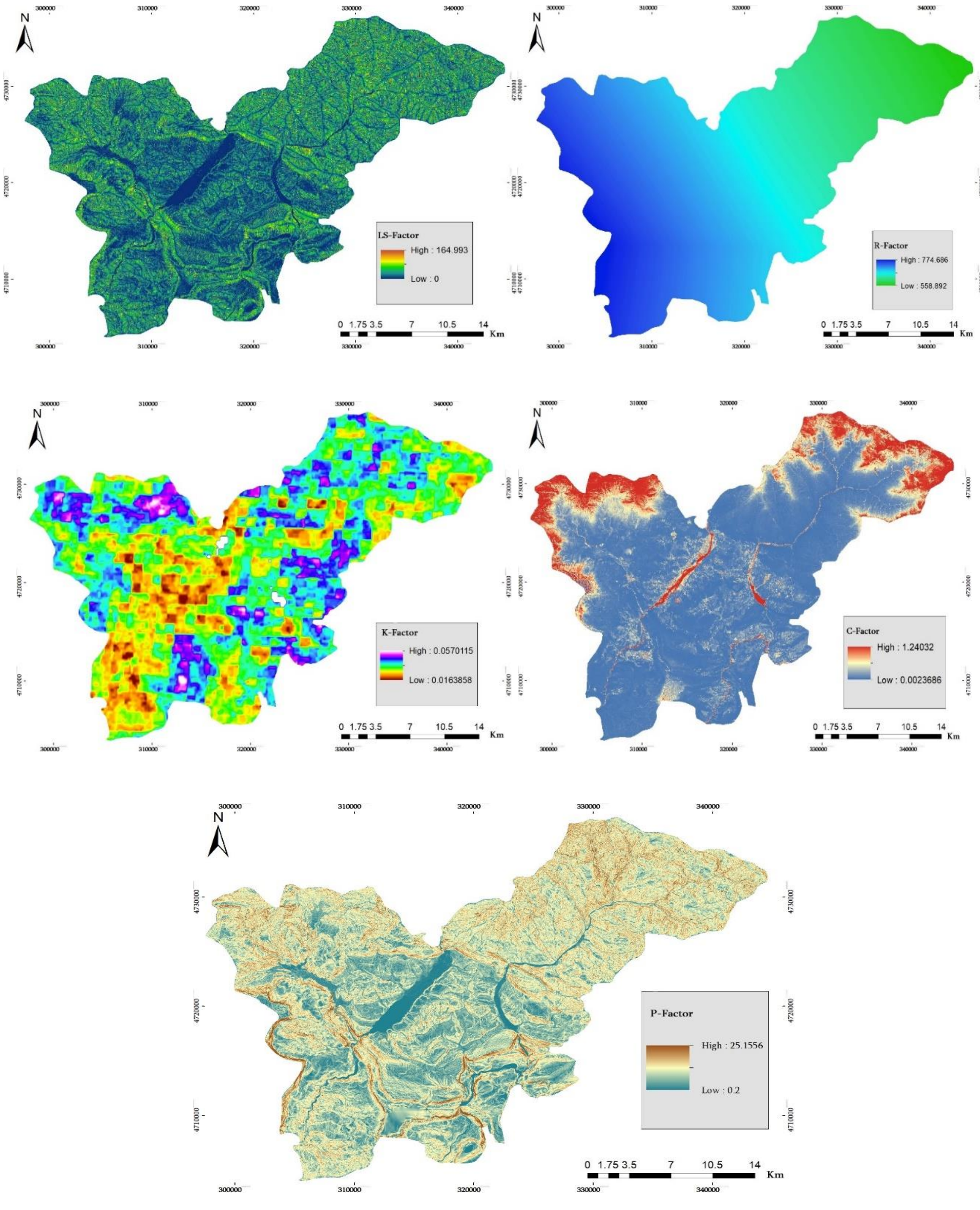


Fig. 4. RUSLE factors. from left to right: R factor; K factor; LS factor; C factor; P factor

The next parameter selected as an additional variable for the RF model is the site's distance from the road. The construction and operation of roads directly affect soil erosion, which is manifested in the destruction of vegetation cover, soil compaction, alterations in surface flow patterns, etc.

The distance from various streams can also have a potential influence on water erosion of the soil. Rivers, streams, canals, etc. represent a kind of natural drainage system that carries surface runoff and sediments to downstream areas. An area closer to a particular stream is hydrologically more connected to the hydrographic network, which can increase the risk of erosion due to concentrated surface runoff and sediment transfer to the riverbed. During heavy rains and snowmelt, sediments are mobilized from these areas and transferred to streams, which increases erosion rates and sediment accumulation in coastal areas.

Flow density expresses the degree of hydrological connectivity within an area. High values of this factor indicate the abundance of water flows, which impact water and sediment transport. Areas with high stream density values are characterized by more surface runoff and sediment movement. Flow density values also determine the morphology and stability of riverbeds within a watershed. The narrow riverbeds with steep gradients are often associated with higher flow density values. These conditions result in the high speed of the flow and potentially high erosion along the riverbed. Intense erosion, under very high flow rates, can cause riverbed rupture, widening, and meandering, which in return alters flow dynamics and sediment transport conditions.

Lithology plays a fundamental role in soil formation. Different types of rocks are characterized by different levels and rates of weathering. Relatively resistant, such as igneous and metamorphic rocks, produce coarser-textured and less erodible soils. In contrast, sedimentary rocks like slates or sandstones, which experience significant weathering, produce fine-textured soils that are more susceptible to erosion. Lithology affects the resistance of soil and substrate to water erosion. Rocks with high resistance to erosion, such as basalts and granites, provide a more stable substrate to erosion agents and contribute to soil stability. The raster file used in the model was created by digitizing the 1:500,000 scale geological map of Georgia (Gudjabidze & Gamkrelidze, 2003) for the study area.

NDVI (Normalized Difference Vegetation Index) is one of the main indicators of vegetation greenness (health) and density. Higher NDVI values indicate greater vegetation cover and biomass. Vegetation has a decisive importance in terms of soil erosion control, which is manifested in the retention of atmospheric precipitation, reduction of surface runoff, stabilization of the soil by the root system, etc. Areas with high NDVI values are usually covered with dense vegetation, which mitigates the impact of raindrop kinetic energy and soil particle detachment. Dense vegetation enhances water infiltration into the soil and increases soil moisture levels. Additionally, vegetation produces a dead cover that acts as a protective layer and enriches the soil with organic substances, further aiding in erosion control.

Slope inclination reflects the steepness or inclination of the surface and, in conjunction with the LS factor in the RUSLE model, affects soil water erosion. The slope inclination gradient is directly related to the volume and velocity of surface runoff. Steep slopes typically experience rapid surface runoff with high kinetic energy. During intense rainfall, concentrated flows on moderately to steeply inclined slopes can create microchannels, which facilitates the development of rill erosion. Furthermore, the slope gradient influences both soil moisture distribution and infiltration rates.

The Stream Power Index (SPI) is a geomorphological parameter used to measure the erodibility of flowing water. It is influenced by parameters such as slope inclination, stream floor morphology and water outflow. The SPI helps to quantify the potential for erosion by evaluating the energy and capacity of water to transport sediments. The concept of SPI was created by Luna Leopold and Ronald Miller in 1956 and is calculated using the following formula:

$$SPI = \frac{Q}{W} \times S \tag{7}$$

where, Q represents water outflow (cubic meters per second), W - width of the riverbed (meters) S - gradient of the streambed (dimensionless or in percent).

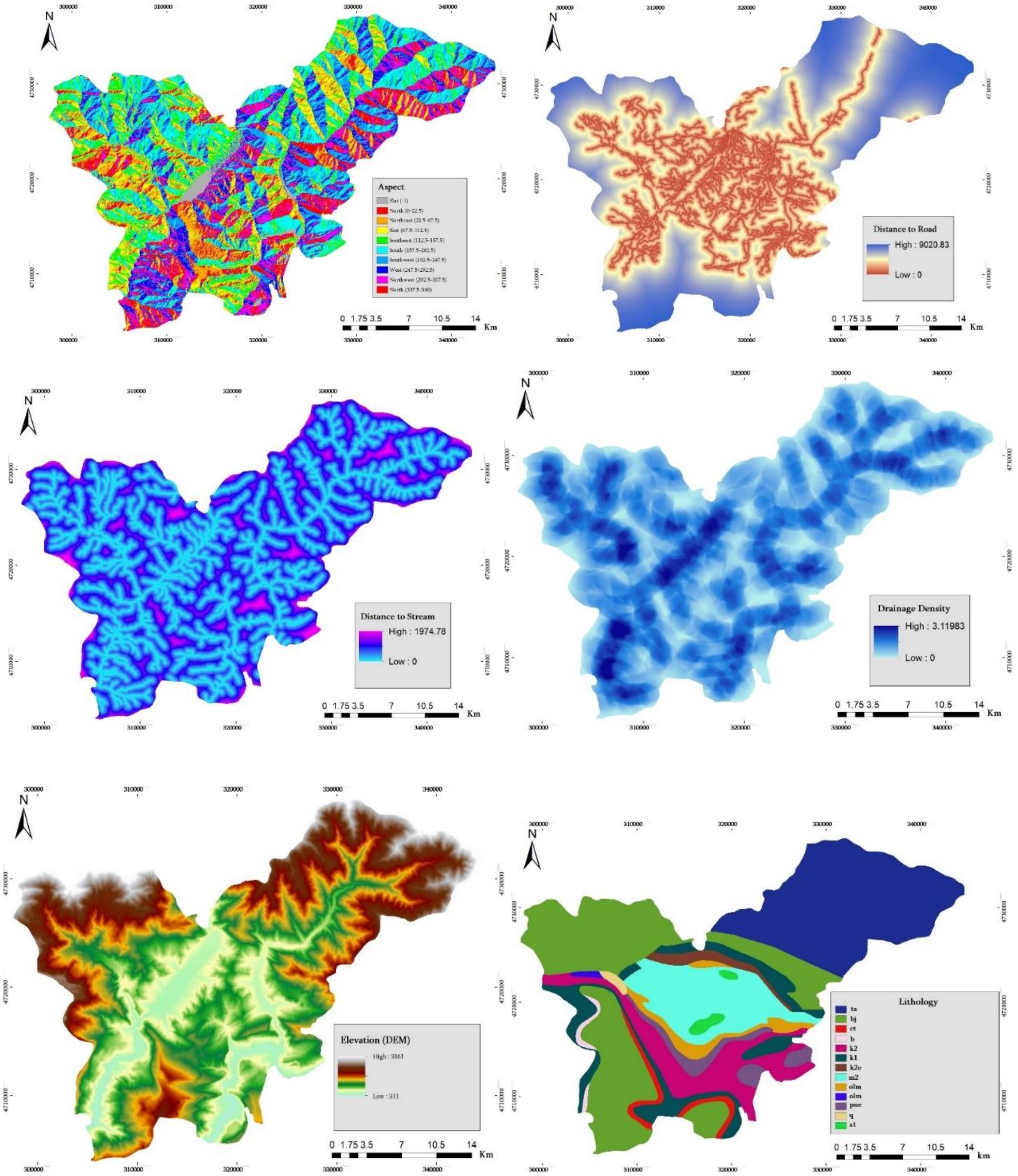
SPI indicators give an idea of the erosive potential of a stream. High values indicate high erodibility.

The topographic wetness index (TWI) is a terrain-based parameter used to characterize the potential wetness of a landscape, given its topographic features. TWI is calculated from a digital elevation model using the following formula:

$$TWI = In\left(\frac{A}{tan(\beta)}\right) \quad (8)$$

where, A represents the catchment area of the upper reaches of the stream, the total area from which all streams flow into a particular point. β - local slope angle. The TWI is dimensionless.

The training data for the Random Forest algorithm consist of point inputs derived from the RUSLE, high-resolution satellite imagery and fieldwork observations (Fig. 6). A total of 250 points indicating either the presence or absence of erosion were collected. Of these, 70% were used to train the model, while 30% were reserved for testing.



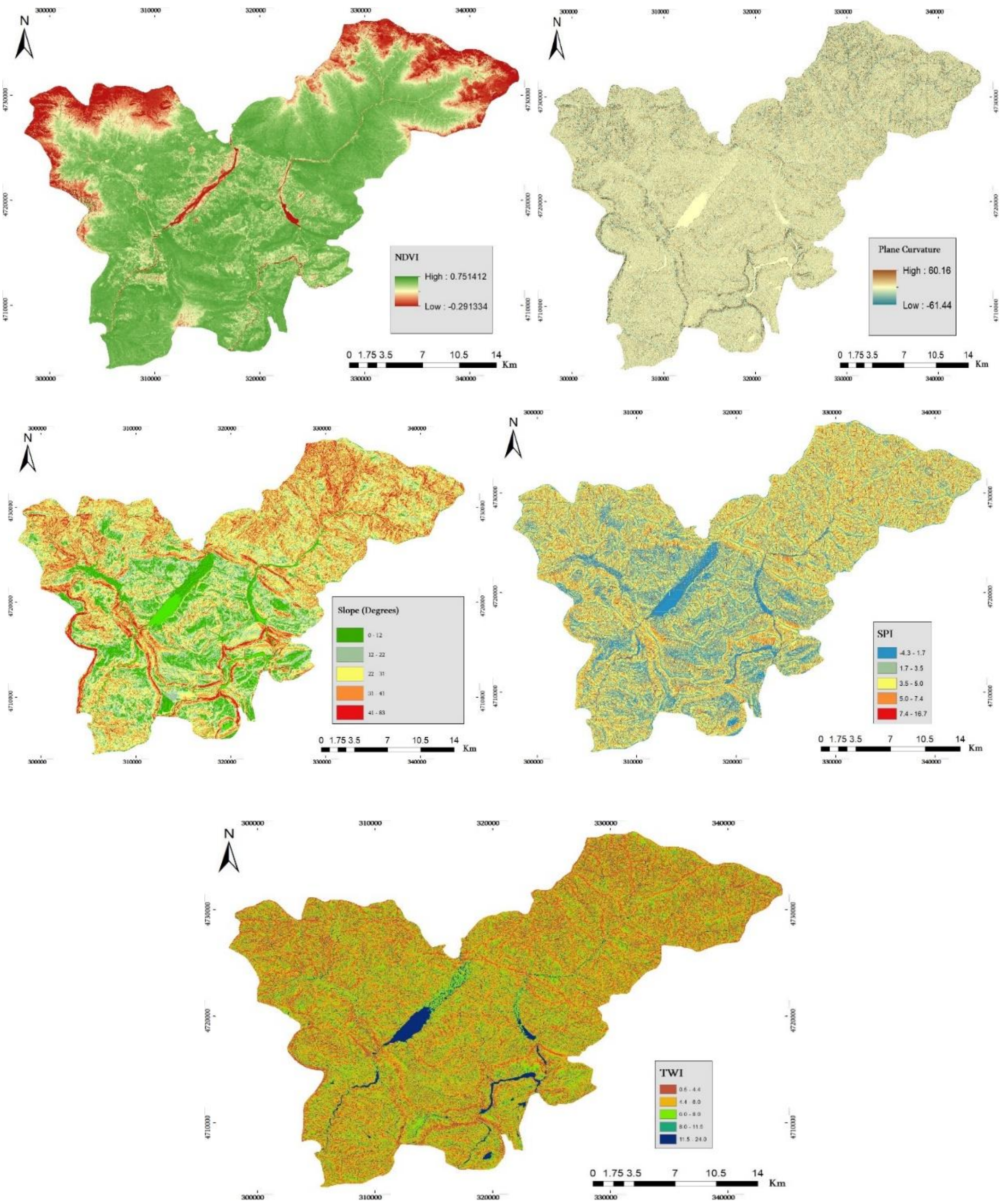


Fig. 5. Additional predictors for RF. From left to right: Aspect; Distance to Road; Distance to Stream; Draingae Density; Elevation; Lithology; NDVI; Plane Curvature; Slope; Spi; TWI

Model parameters were selected through cross-validation and model optimization. The number of trees in the forest - n-tree was set to 500, and the number of random variable considered at each node - m-try was set to 4.

Table 1 shows the statistics of the obtained model. F1-Score describes the correctness of the model, taking into account the precision of the predictions and recall. The achieved value - 0.77 indicates that the model attained a relatively high level of accuracy in predicting both the presence and absence of soil erosion.

Sensitivity, also known as the true positive rate, determines the proportion of true positive predictions from all cases of erosion. A sensitivity index of 0.79 indicates that the model effectively identified 79% of the actual occurrence of soil erosion.

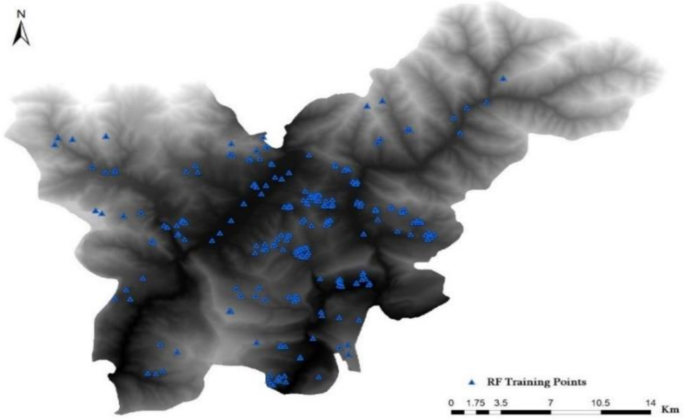


Fig. 6. RF training points

Table 1. Statistics of the obtained model

F1-Score	Sensitivity	Accuracy
0.77	0.79	0.91

Accuracy represents an indicator of overall accuracy of model's predictions, encompassing both true positives and true negatives. An accuracy of 0.91 indicates that the model correctly predicted 91% of all cases in the dataset.

Overall, considering the above parameters, it can be concluded that the obtained model demonstrates high performance in predicting soil erosion.

3. RESULTS AND DISCUSSION

After modeling all five factors included in the RUSLE (Fig. 4), thematic raster images were combined according to equation (1). As a result, the average annual rate of soil loss (t/ha per year) was obtained for the study area.

In (Fig. 7), accumulated lower numerical values reflect lower erosion rates at a given site and vice versa. The minimum value recorded is 0, and the maximum value is 1468.3.

We categorized the received erosion rates within the study area into 5 classes (very low, low, medium, high, very high), using the "natural breaks" classification method.

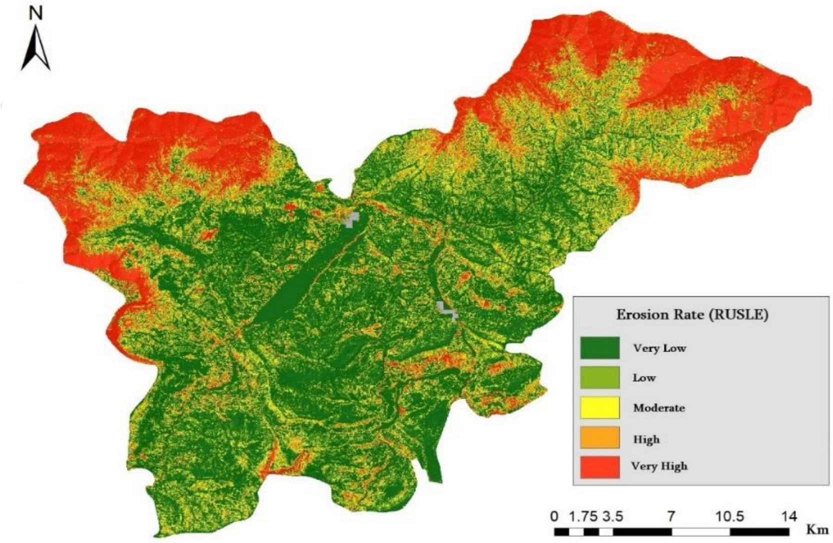
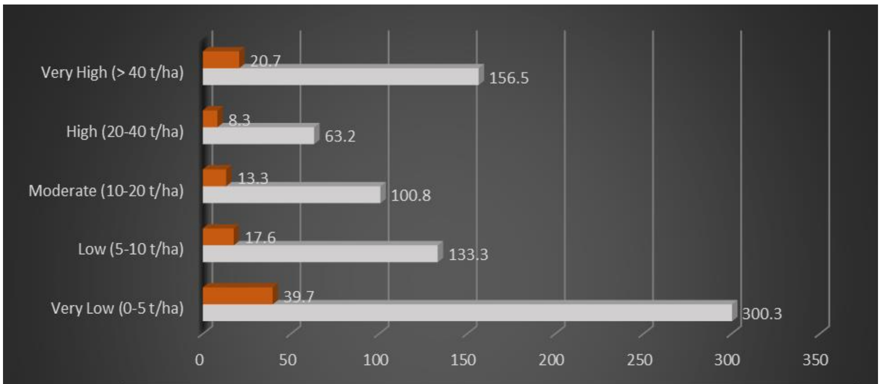


Fig. 7. Spatial distribution of soil erosion in the study area (as a result of RUSLE modeling)



	Very Low (0-5 t/ha)	Low (5-10 t/ha)	Moderate (10-20 t/ha)	High (20-40 t/ha)	Very High (> 40 t/ha)
%	39.7	17.6	13.3	8.3	20.7
Area (Km2)	300.3	133.3	100.8	63.2	156.5

% Area (Km2)

Fig. 8. Distribution of soil loss by area (RUSLE)

According to the RUSLE model results (Fig. 8), 39.7% of the study area falls within the very low erosion risk zone. Conversely, 20.7% of the area is classified as having a very high erosion rate. This is mainly associated with the high hypsometry and steep slopes of the Lechkhumi and Khvamli ridges, as well as the Askhi karst massif, which are either completely devoid of soil and vegetation cover, or are represented by primitive soils and sparse vegetation.

Similarly to the modeled data of the RUSLE, we also divided the erosion metrics obtained by the RF algorithm (Fig. 9) into five classes. Regarding the distribution of erosion risk classes, RF provided a somewhat different picture, which was expected. The Random Forest algorithm, leveraging its ability to process diverse data, can analyze the complex, non-linear relationships between soil erosion and its contributing factors. In contrast, RUSLE is an empirical model that assumes a linear relationship among its factors when predicting erosion. This linear approach may result in oversimplified or generalized outcomes, particularly in complex physical-geographical conditions. However, by this sign, its simplicity and practicality can also be considered a positive side, if the modeling is carried out in relatively less complex conditions.

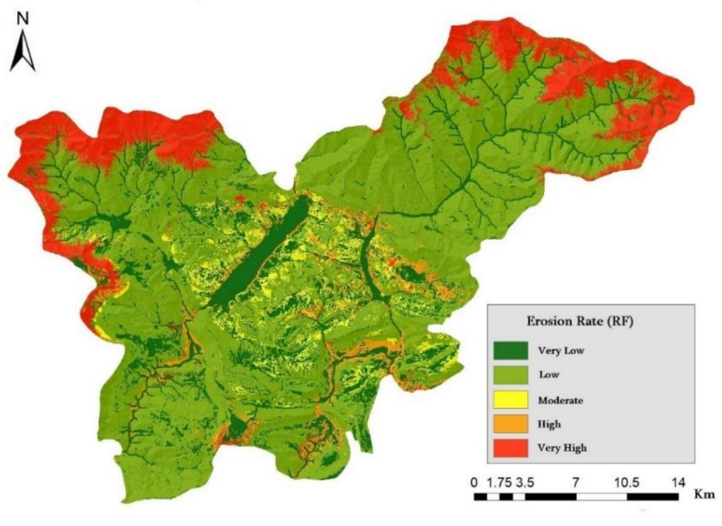
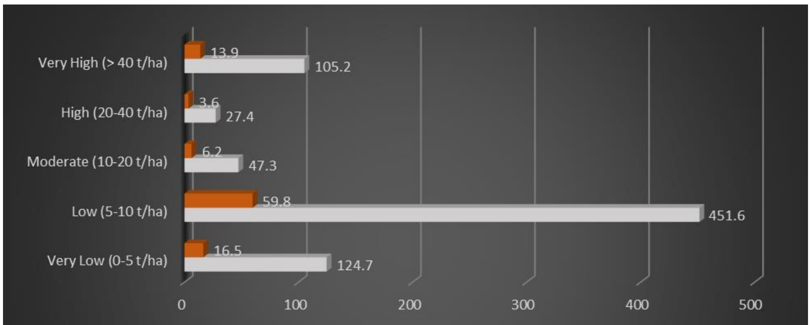


Fig. 9. Spatial distribution of soil erosion in the study area (as a result of RF modeling)



	Very Low (0-5 t/ha)	Low (5-10 t/ha)	Moderate (10-20 t/ha)	High (20-40 t/ha)	Very High (> 40 t/ha)
%	16.5	59.8	6.2	3.6	13.9
Area (km2)	124.7	451.6	47.3	27.4	105.2

Fig. 10. Distribution of soil loss by area (RF)

As shown in Fig. 10, the Random Forest modeling results indicate that 16.5% of the study area is classified as having very low soil loss and erosion risk, while 13.9% of the area is categorized as having very high risk.

Spatial analysis of the RUSLE model outcomes reveals that both C factor (cover management) and the LS factor (slope length and steepness) are important contributors to erosion processes. It should be noted that, due to the specific characteristics of the study area, the model inaccurately predicted high erosion rates on steep slopes, which are often covered by highly productive forests that serve soil protection and water regulation functions. Typically, in the case of high projection coverage of vegetation, soil erosion rates are reduced to a minimum. Nonetheless, the model struggled to compensate the high values of the LS factor in the study area with other factors (the main one being the C factor). This suggests that for extensive territories with diverse and complex terrain, it is essential to consider these limitations when interpreting the RUSLE model results. In particular, steep, forested slopes might require additional adjustments or alternative approaches to better capture the protective role of vegetation in reducing erosion risk.

A comparison of RUSLE results with those obtained from RF modeling reveals significant differences in the distribution of specific erosion risk classes between the models, across the study area (Fig. 11).

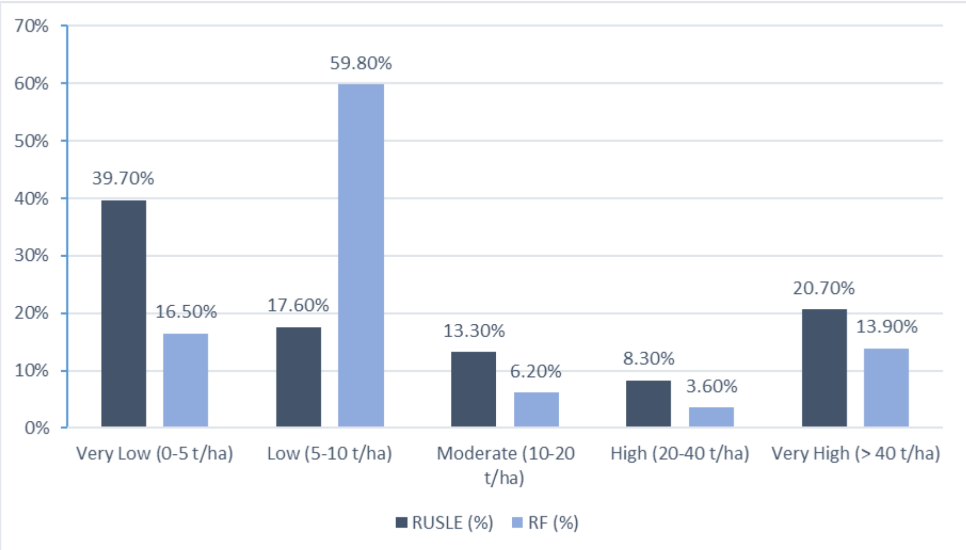


Fig. 11. Specific share of soil erosion classes in the study area for the RUSLE and RF models

The discrepancy is particularly notable in the class representing low erosion risk (5-10 t/ha). RUSLE identified this class for 17.6% of the territory, whereas the RF algorithm classified 59.8% of the area within the same risk zone. As mentioned above, in some cases, RUSLE overpredicts high erosion rates under conditions where the values of individual factors cannot compensate for each other. In this case, the random forest model assigned a low erosion risk class to areas where RUSLE identified medium or high erosion risk. This class is spatially related to slopes of different hypsometry and inclination in the study area. Although these features of the terrain contribute to erosion development, their influence is balanced by the dense vegetation, which significantly reduces the erosion risk. This characteristic of RUSLE is also noted by other authors (Liu et al., 2018). The difference between the modeling results is relatively minor for the specific share of medium, high, and very high erosion risk classes (about 5-7%). According to the Random Forest model, areas with very high potential for soil erosion are primarily associated with secondary meadows near villages, which are extensively used for grazing livestock (Fig. 12), also with quarries and exposed bare ground, such as arable fields. Comparing Fig. 9 to Fig. 7, difference in the spatial distribution of other classes is evident, which, as mentioned, is due to the differing evaluations of the relationships between individual variables by the models used.

The variable importance of the additional factors used in the Random Forest modeling process, indicates the influence each parameter had on predicting erosion. The variable importance is calculated by evaluating how much the inclusion of each variable contributes to the reduction in Gini coefficient (Gini impurity) across the trees in the random forest. Variables that lead to greater reduction in the Gini coefficient are considered more important for modeling the research phenomenon (Huffman et al.2023).

Fig. 12. Landslides and eroded secondary meadows near village Spatagori (2024)

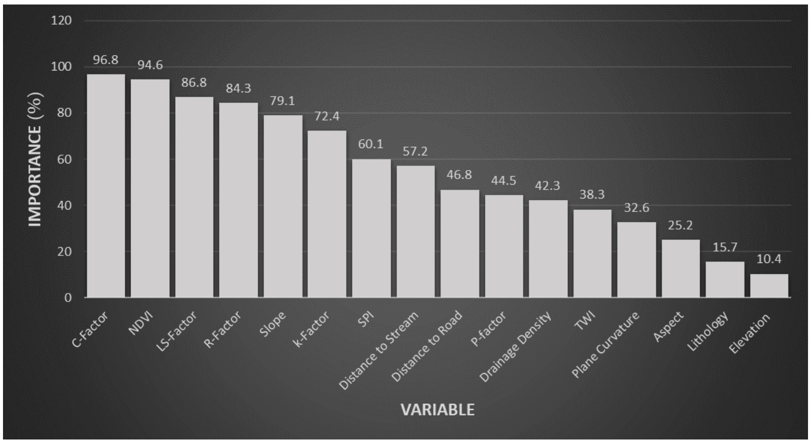


Fig. 13. Variable importance of the obtained model for soil erosion prediction

As shown in Fig. 13, the most important variables for soil erosion risk modeling in the study area include the C factor, NDVI, LS factor, R factor, slope inclination, K factor, SPI and distance to stream, each of which contributes more than 50% to the model's performance. Although each variable used in the modeling process has the potential to influence soil erosion development, their relative importance varies across different areas and depends on the conditions of interaction between them. In this case, the low value obtained for the lithology factor is noteworthy. As previously mentioned, the thematic raster image used in the modeling was derived from the digitization of the geological map of Georgia at a scale of 1:500,000. Therefore, the information available for the research area is highly generalized and lacks the detailed aspects important for accurate modeling, which is one of the main reasons for the low importance associated with the liyhology factor.

Spatial analysis of the thematic images of individual factors and the soil erosion map generated by the RF model reveals that areas with high values of the C factor and NDVI predominantly show a very low or low risk of erosion. In contrast, when high values of the LS factor and R factor coincide, the model indicates moderate to high rates of erosion.

4. CONCLUSION

In the present study, soil erosion risk by water in Tsageri municipality was assessed using two different approaches: an empirical model (RUSLE) and a machine learning model (Random Forest). The modeling process primarily relied on freely available digital data, supplemented by data collected through field expeditions in years 2023-2024. The results indicated that the distribution of erosion classes obtained from RUSLE and RF models differed across the study area. According to the RUSLE, 42.3% of Tsageri municipality is classified in the medium, high, or very high risk zones for soil water erosion. In contrast, the RF algorithm identified 23.7% of the area within these same risk categories. Comparing the modeling results with field data shows that the RUSLE model sometimes overestimates the risk of medium, high and very high erosion on high-gradient slopes covered with broad-leaved and mixed forests. In the case of the Random Forest model, this overestimation is largely addressed by incorporating additional variables into modeling process. In addition to this, the risk of erosion in the study area is primarily associated with cultivable land on slopes, quarry exploitation sites and meadows in the subalpine and alpine zones. It should be noted that in the future, the utilization of the largest part of Tsageri municipality's territory from an agricultural point of view is very limited due to the topographical conditions and the lack of available land.

The inclusion of additional variables in the modeling process represents an experimental aspect of this study. It is recommended to examine various sets of erosion-influencing variables to identify the most effective combinations for each specific study area.

While the quality of the data used affects the accuracy of the research findings, the results substantially address existing gaps in the field. They provide valuable insights into land degradation processes in the study area and the interrelationships among causative factors, which creates the basis for future studies.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

Arnoldus, H. J. (1977). *Methodology Used to Determine the Maximum Potential Range Average Annual Soil Loss to Sheet and Rill Erosion in Morocco.* Assessing Soil Degradation, FAO Soils Bulletin, FAO.

Borrelli, P., Robinson, D. A., Panagos, P., Lugato, E., Yang, J. E., Alewell, C., & Montanarella, L. (2017). An assessment of the global impact of 21st century land use change on soil erosion. *Nature Communications, 8*(1), 1-13. https://doi.org/https://doi.org/10.1038/s41467-017-02142-7

Bouramtane, T., Hilal, H., Rezende-Filho, A. T., Bouramtane, K., Barbiero, L., Abraham, S., . . . Torres-Rondon, L. (2022). Mapping Gully Erosion Variability and Susceptibility Using Remote Sensing, Multivariate

Statistical Analysis, and Machine Learning in South Mato Grosso, Brazil. *Geosciences, 12*(6). https://doi.org/https://doi.org/10.3390/geosciences12060235

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/https://doi.org/10.1023/A:1010933404324

Gudjabidze, G. E., & Gamkrelidze, I. P. (2003). *Geological Map of Georgia (Scale 1: 500 000).* Tbilisi: Georgian State Department of Geology and National Oil Company "Saknavtobi".

Huffman, G. J., Stocker, E. T., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2023). GPM IMERG Final Precipitation L3 1 month 0.1 degree x 0.1 degree V07. Retrieved 09 15, 2024, from 10.5067/GPM/IMERG/3B-MONTH/07

Liu, J., Gao, G., Wang, S., Jiao, L., Wu, X., & Fu, B. (2018). The effects of vegetation on runoff and soil loss: multidimensional structure analysis and scale characteristics. *Journal of Geographical Sciences, 28*(1), 59-78. https://doi.org/https://doi.org/10.1007/s11442-018-1459-z

Montgomery, D. R. (2007). Soil erosion and agricultural sustainability. *Proceedings of the National Academy of Sciences, 104*(33), 13268-13272. https://doi.org/https://doi.org/10.1073/pnas.0611508104

Moore, I. D., & Burch, G. J. (1986). Modeling Erosion and Deposition. Topographic Effects. *Transactions of American Society of Agriculture Engineering, 29*, 1624-1630. https://doi.org/http://dx.doi.org/10.13031/2013.30363

Moore, I., & Wilson, J. P. (1992). Length Slope Factor for the Revised Universal Soil Loss Equation: Simplified Method of Solution. *Journal of Soil and Water Conservation, 47*(4), 423-428.

Panagos, P., Borrelli, P., Poesen, J., Ballabio, C., Lugato, E., Meusburger, K., & Alewell, C. (2015). The new assessment of soil loss by water erosion in Europe. *Environmental Science & Policy*(54), 438-447. https://doi.org/https://doi.org/10.1016/j.envsci.2015.08.012

Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., Kurz, D., McNair, M., & Blair, R. (1995). Environmental and economic costs of soil erosion and conservation benefits. *Science, 267*(5201), 1117-1123. https://doi.org/https://doi.org/10.1126/science.267.5201.1117

Poesen, J. (2018). Soil erosion in the Anthropocene: Research needs. *Earth Surface Processes and Landforms, 43*(5201), 64-84. https://doi.org/https://doi.org/10.1126/science.267.5201.1117

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. https://doi.org/https://doi.org/10.5194/soil-7-217-2021

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., & Yoder, D. C. (1997). *Predicting soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE).* United States Department of Agriculture.

Sharply, A., & Williams, J. (1990). *EPIC-erosion/productivity impact calculator: 1. Model Documentation.* Washington DC: USA Department of Agriculture.

Wenner, C. G. (1981). *Soil Conservation in Kenya.* Nairobi: Ministry of Agriculture.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/126403