(2023) 10:22

RESEARCH

Open Access

Stepwise integration of analytical hierarchy process with machine learning algorithms for landslide, gully erosion and flash flood susceptibility mapping over the North-Moungo perimeter, Cameroon

Alfred Homère Ngandam Mfondoum^{1,2*}, Pauline Wokwenmendam Nguet³, Dieudonné Tchokona Seuwui¹, Jean Valéry Mefire Mfondoum⁴, Henry Bang Ngenyam⁵, Ibrahima Diba⁶, Mesmin Tchindjang^{2,7}, Bertin Djiangoue⁸, Ali Mihi⁹, Sofia Hakdaoui¹⁰, Roseline Batcha², Frédéric Chamberlain Lounang Tchatchouang², Igor Casimir Njombissie Petcheu^{2,7} and Luc Moutila Beni¹¹

Abstract

Background The Cameroon Volcanic Line (CVL) is an oceanic-continental megastructure prone to geo-hazards, including landslide/mudslide, gully erosion and flash floods targeted in this paper. Recent geospatial practices advocated a multi-hazard analysis approach supported by artificial intelligence. This study proposes the Multi-Geoenvironmental Hazards Susceptibility (MGHS) tool, by combining Analytical Hierarchy Process (AHP) with Machine Learning (ML) over the North-Moungo perimeter (Littoral Region, Cameroon).

Methods Twenty-four factors were constructed from satellite imagery, global geodatabase and fieldwork data. Multicollinearity among these factors was quantified using the tolerance coefficient (TOL) and variance inflation factor (VIF). The AHP coefficients were used to weigh the factors and produce a preliminary map per Geoenvironmental hazard through weighted linear combination (WLC). The sampling was conducted based on events records and analyst knowledge to proceed with classification using Google Earth Engine (GEE) cloud computing interface. Classification and Regression Trees (CART), Random Forest (RF) and Gradient Boosting Regression Trees (GBRT), were used as basic learners of the stacked hazard factors, whereas, Support Vector Regression (SVR), was used for a meta-learning.

Results The rainfall was ranked as the highest triggering factor for all Geoenvironmental hazards according to AHP, with a coefficient of 1, while the after-learning importance assessment was more varied. The area under receiver operating characteristic (AUROC/AUC) was always more than 0.96, and F_1 -score is between [0.86–0.88] for basic classifiers. Landslides, gully erosion and flash floods showed different spatial distributions, confirming then their probability of co-occurrence. MGHS outputs clearly displayed two and three simultaneous occurrences. Finally, the human vulnerability assessed with population layer and SVR outputs showed that high human concentrations are also the most exposed, using the example of Nkongsamba's extract.

*Correspondence:

Alfred Homère Ngandam Mfondoum

stats.n.maps.expertise@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2023. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.



Conclusions Combining AHP with single learners, then a meta-learner, was efficient in modelling MGHS and related human vulnerability. Interactions among geo-environmental hazards are the next step and city councils are recommended to integrate results in the planning process.

Keywords Analytical hierarchy process, Cameroon volcanic line, Human vulnerability, Machine learning, Multigeoenvironmental hazards susceptibility, North-Moungo perimeter

Introduction

The Sendai Framework for Disaster Risk Reduction (SFDRR), stresses the need to understand disaster risks through identifying and assessing exposure to all hazards (UNISDR¹ 2015b). Geo-environmental hazards specifically, result from combined anthropogenic and natural factors (IRDR² 2014). They can occur suddenly (earth-quakes and volcanic eruptions) or over time (flash floods, drought, desertification, landslides, erosion), leading to fatalities, damages to property/environment and disruption of services (Bang 2022). A rapid response to these hazards is necessary to assess damages, save lives and mitigate reoccurrence consequences (Bang et al. 2019). This study focuses on landslides, gully erosion and flash floods.

Landslides are a sudden upward-downward mass movement of debris, soils or rocks due to gravity and other interaction of geo-environmental and humaninduced factors (Hungr et al. 2014). The majority of landslides originate from hydroclimatic events, specifically prolonged and/or huge rainfall (Park and Kim 2019) with several societal and environmental losses. An estimated 4.8 million peoples have been affected by landslides, and more than 18,000 deaths were recorded between 1998 and 2017.³

Gully erosion on its own, is an important phase/form of soil loss and land degradation especially in semi-arid and arid regions. Several processes, i.e., anthropogenic – agricultural practices (Zhao and Hou 2019), deforestation (Gholami 2013) – natural, i.e., wind (Skidmore 1982), as well as overland water flows (Xiong et al. 2018), easily trigger or promote soil erosion. According to literature, flowing water, i.e., torrential, represents the main natural driver, initiating or emphasizing gullies dynamic, which are particularly devastating for both environment and the society (Arabameri et al. 2020; Pourghasemi et al. 2020a).

Besides, flood refers to an overflow of water onto normally dry surrounding lands (NOAA/NWS).⁴ Both coast lands and inlands are affected. Depending on the triggering factor, the spatial location and the speed of the event/process, there three main types such as, coastal flood, fluvial flood and pluvial flood. Between 1995 and 2015, about 150,061 flood events happened worldwide, approximately 109 million people were impacted by the flood damages, causing 157,000 deaths (UNISDR 2015a),⁵ with direct costs of USD 75 billion per year, which were liable for 11.1% of the global disaster casualties (Alfieri et al. 2017). Several investigations report that floods affected around 200 million people each year on a global scale (e.g., Tien Bui et al. 2019). In this study, pluvial flood was studied, due to its suddenness compared to the other two types, affecting more the population preparedness/ resilience; while events of its (re)occurrence in the study area, were recorded during the rainy season, preceding or simultaneously with landslide and gully erosion (see Data and Methods section). Usually, the streams overflow from an existing waterway (river, stream, watershed, drainage ditch, etc.), happens between few minutes to less than 6 h, often as a result of heavy rainfall and inundate surrounding areas (Kron 2005). It is then commonly referred as flash flood, as used here.

Recently, the UNISDR has emphasized the significance of multi-hazard assessment and referred to it as an essential tool for a safer world in the twenty-first century (UNISDR 2015b, 2015c). In this regard, static and dynamic geospatial models have been developed and diversified for single hazard, with technical approaches applicable to simultaneous multiple hazards mapping.

The field inventory-based and plotting of data formerly used to produce hazard susceptibility map (HSM) (Varnes 1984; Colombo et al. 2005), was first completed by aerial photointerpretation, then satellite data, to build heuristic, statistical and data-driven models (Al-Juaidi et al. 2018; Khosravi et al., 2018; Meena et al. 2019; Rahmati et al. 2016; Mihi et al. 2020;). Then, artificial intelligence (AI) algorithms have been introduced in recent decades, to overcome performance limitations, especially artificial neural network (ANN) (Wang et al. 2019), and machine learning (ML) (Sadler et al. 2018; Pham et al. 2019; Chen et al. 2020; Towfiqul Islam et al. 2021), improving at the same time warning

¹ United Nations International Strategy for Disaster Reduction.

² Integrated Research on Disaster Risk.

³ https://www.who.int/health-topics/landslides#tab=tab_1.

⁴ https://www.weather.gov/mrx/flood_and_flash.

⁵ United Nations Office for Disaster Risk Reduction.

systems based on high computational approaches, for the monitoring and near real-time forecasting (Kirschbaum and Stanley 2018; Fayne et al. 2019; Ngandam Mfondoum et al., 2021; USGS 2022a, b, c).

Concerning multi-hazard mapping specifically, the Maximum Entropy (MaxEnt) algorithm has been combined with the technique for order of preference by similarity to the ideal solution and Mahalanobis distance (TOPSIS-MD), to harmonize the probabilistic context of multiple natural hazards (Sheikh et al. 2019). This ML has also been proven efficient in selecting predictors, assessing the degree of model fitting and predicting performance, for combined landslide, erosion/gully erosion and flood susceptibility assessment (Javidan et al. 2021). Another approach modelled different single hazards susceptibility through the simple additive weight (SAW) with the Analytical Hierarchy Process (AHP) coefficients, to prepare the path to the multi-hazard mapping of landslide, rockfall, goaf collapse and soil erosion (Sun et al. 2019). Further, the Boruta algorithm has been highly proficient in selecting and prioritizing factors per type of hazard in the same analysis; and then applying the Random Forest (RF) algorithm to train and validate flood, landslides and forest fire, to produce multi-hazard susceptibility map (Pourghasemi et al. 2020b). Moreover, a multiple and comparative learning, (i.e., boosted regression tree, random forest, support vector machine), was envisaged to produce a valid multi-hazard map of flooding, gully erosion, forest fires, and earthquakes (Pouyan et al., 2021).

However, most studied locations benefit from accurate and up-to-date databases on the occurrence and location of geo-hazards. This is not always the case in some areas. In addition, the socio-economic component is usually introduced as settlements (built-up, road), while the human dimension itself as the number of residents is not exclusively considered.

From the above statements, this study mixes field survey data and an analyst knowledge-based approach, with robust artificial intelligence processing, to predict landslide, gully erosion and flash flood distribution, from single to simultaneous occurrences. The main rationale of this work is to demonstrate how combining AHP to ML algorithms in the same modelling process, produces satisfactory performances to map the multi-geo-env hazards susceptibility (MGHS) and related human vulnerability. Specific goals are the following: *i*) use the AHP method to weigh the factors/predictors per model; ii) compare performances of different ML algorithms in accurately predicting occurrences on stacked-weighed factors; iii) assess the proficiency of a meta-learner in arbitrating the best distribution of single and combined geo-env hazards from stacked basic learners' output; iv) predict the human *direct threats to each and all occurrences from the population distribution perspective.*

The major novelty of this work, resides in performing ML over AHP's weighed factors, for multi-geo-env hazards modelling. The two have always been used separately according to the available literature that has been synthesized above, but not in the same modelling process, to the best of our knowledge. Furthermore, the population distribution layer is newly introduced in hazard multi-analysis.

Dataset and methods

Study area

The study was conducted in the Northern subset of the Moungo division (Littoral Region), embedded between three volcanic mountains, i.e., Manengouba (2396 m), Koupé (2070 m), Nlonako (1822 m), and belongs to the Cameroon Volcanic Line (CVL) (Elsheikh et al. 2014). The soils are essentially less evolved volcanic ones known as andosols (Gèze, 1942). It is slightly extended over the surrounding natural perimeter encompassing administrative divisions of, Haut-Nkam, Menoua (West Region), and Koupé-Manengouba (South-West Region), on the edges of the Bamileke Plateaus, which is constantly subject to different geo-env hazards (Nguimbous and Manguelle, 2010; Wotchoko et al. 2016; Ndonbou et al. 2022). The whole area is inside coordinates, North 4°54'-5°21' and East 9°46'30"-10°6'. The main cities covered are, Nkongsamba (Moungo's head of division), Melong, Kekem, then partially Santchou and Bangem (Fig. 1).

The Equatoguinean climate dominates the area, with eight months of rainfall (mid-March/mid-November) averaging 1400 mm/year, while peaks exceed 2000 mm in September (Kadjio Feudjio et al. 2021). Rainfall distribution impacts annual temperatures, which vary between 16 and 23°C, as well as relative humidity which varies between 55 and 99%. These climate parameters are deeply influenced by altimetry (Kadjio Feudjio et al. 2021) and in return, they influence the natural land cover distribution. The vegetation comprises degraded forest mixed with savannas (tree and grassy) on mountains top and hillslopes, forest galleries in valleys and spots of agroforestry (Tsewoue et al. 2020). From the compiled district council and national census statistics, about 288,604 inhabitants live in the studied subset, with an average population density of 142.9 per km².

Work environment and tools

The processing was conducted alternatively between the cloud environment and desktop tools. Google Earth Engine (GEE) platform was used for cloud computing (Gorelick et al. 2017). Python and JavaScript codes have been previously implemented for semi-automatic or fully



Fig. 1 Study location inside administrative (a) and altimetry (b) contexts

automatic socio-environmental monitoring and hazard susceptibility mapping (e.g., UN-SPIDER 2020; Elnashar et al. 2021; Handwerger et al. 2022). A JavaScript coding process was been used to select, filter, load, and preview the study area and subsequent data, as well as further machine learning.

The offline tools are licensed software such as Erdas Imagine 2020 version 16.6.0.1366, ArcGIS version 10.8.2, SAGA-GIS version 8.0.1, Super Decision Software (SDS) version 3.2 and, XLStats version 2020.1.64570. They were used for the preparation of factors, statistics and layouts.

Data preprocessing and factors preparation

Susceptibility modelling refers to the spatial prediction of a phenomenon by perception and formulating the link between the occurrence evidence and the influential thematic factors/predictors (Sheikh et al. 2019). To define these predictors and before any operation, each data was clipped at the extent of the subset.

Shuttle radar topography mission-digital elevation model imagery and direct layers

Direct layers refer to those prepared in a straight computation by the software. The Shuttle Radar Topography Mission-Digital Elevation Model, SRTM-DEM, was the only raw data provider of this group of layers. This data is freely available worldwide on the website of the United States Geological Survey Earth Resources Observation and Science (USGS/EROS), and offers a void-filled data at a resolution of 1 arc-second, i.e., approximately 30 m.⁶ It was mostly preprocessed with SAGA-GIS, to produce thirteen (13) raster layers, which are, elevation, slope degrees, slope aspect, plan curvature, profile curvature, relative slope position, terrain ruggedness index (TRI), slope length and steepness (LS), valley depth, drainage

⁶ https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-eleva tion-shuttle-radar-topography-mission-srtm-1.

density, channel network distance, stream power index (SPI) and topographic wetness index (TWI).

Other satellite data and indirect layers

Indirect layers are those prepared through several computations. The Sentinel-2 multispectral instrument (S2-MSI) surface reflectance image, was first used to perform the land use land cover (LULC) classification. In the collection, five cloudiness granules were available for the 22nd of December 2021 (Appendixes 1 & 2). They were loaded on the GEE interface, and then compiled, i.e., merged and stacked, by using the median reducer function (GEE 2022). Then, the image was converted from 32-bit to 8-bit unsigned integer, before proceeding to the learning process. Four classes of land features were defined, i.e., water, vegetation, soil and built-up. The Random Forest (RF) algorithm was applied (Number of Trees:100, Variables per split:5, Minimum Leaf Population:1, Seed:0.5) (Friedman 2001). Then, the Dynamic World dataset, which is a 10 m near-real-time (NRT) Land Use/Land Cover (LULC) product based on predictions of the Sentinel-2 L1C (Brown et al. 2022), was used for cross validation and accuracy assessment. The overall accuracy was 87% and the kappa coefficient was 0.82. (Appendix 3).

Further, the Shadow-Eliminated Vegetation Index (SEVI) was computed, to assess the natural coverage/ bareness, complementarily to the previous vegetation class from the RF. This index eliminates the effect of terrain shadow, including the self and cast shadows (Jiang et al. 2019). The algorithm is built around any conventional vegetation index (CVI), such as the reduced simple ratio index (RSR) used here, then the shadow vegetation index (SVI) and finally, the micro-topography correction factor ($f(\Delta)$). The latter adjusts the CVI and the corresponding SVI ratio, to avoid under-elimination or over-elimination. This study used, $f(\Delta) = 0.581$, as in the original formulation (Table 1).

Regarding the natural land cover, the soil typology layer was produced using the basic map (1:1 000 000) available (ORSTOM 1965). There are six types of soils depending on the altitude and the bedrock, i.e., black humus soils on basalts, red to yellowish ferrallisol on basic rock, red to yellowish ferrallisol on basic rock, red to yellowish ferrallisol on sediments, oligotrophic organic hydromorphic soil, organic gleysol and brown soil on basic rock (ORSTOM 1970).⁷ Next, the soil depth layer was prepared from the USDA soil texture categorization, known as, the USDA system (Hengl 2018). Version v02 was used, as it globally describes several textures and measures of six different depths (*0, 10, 30, 60, 100* and

200 cm), at 250 m of spatial resolution. However, because the study area mixes different materials according to the classes present in features of the source code, a combination was necessary to deduct new intervals of depths, with a probability, of *class* > 200 cm (Table 1).

Another factor is rainfall, which was composed of two sources. The first one is the TerraClimate data, which is a dataset of monthly climate and climatic water balance for global terrestrial surfaces (Abatzoglou et al. 2018). Its spatial resolution is about 4 km, all available scenes for the year 2021 were exported to the ArcGIS environment, and average values were computed. Complementarily, records from the rain gauges available at local offices of agriculture and livestock were collected, averaged and displayed on the TerraClimate data, to create shapefile in point features. From there, the mean annual rainfall (MAR) layer was created by using the Inverse Distance Weighted (IDW) interpolation for landslide and flash flooding models. Whilst the modified Fourier Index (MFI) was used to create the corresponding layer for the gully erosion model since this computing process better approximates the kinetic energy of rain over erosionprone soils, i.e., R-factor (Arnoldus 1980).

Additionally, some complementary pre-processing of the SRTM-DEM was performed in ArcGIS, using the spatial analysis set of tools. The fault density layer was prepared, following five steps: (*i*) producing the hillshade using surface tool; (*ii*) displaying the shapefile of existing faults⁸; (*iii*) conditioning the hillshade in fourth different azimuth/altitude angles of shaded combinations (i.e., $315^{\circ}/45^{\circ}$; $200^{\circ}/50^{\circ}$; $100^{\circ}/60^{\circ}$; $50^{\circ}/90^{\circ}$); *iv*) digitizing each shading layer to complete the existing shapefile; v) interpolating shapefile using the line density tool. The same shapefile was used as an entry to produce the distance-to-fault layer, with the Euclidean distance tool, fixing 1500 m as the influence zone, which is assumed to be the closest easily affected perimeter according to field knowledge.

Another shapefile created from the SRTM-DEM was the stream network, using the hydrology tool. After producing the flow direction and flow accumulation sub-components, the output raster is converted into a polyline feature. Because several rivers and small streams abound, the values selected were *stream* \geq 200. This feature was used to produce the distance-to-stream layer, following a Euclidean distance of 1500 m.

Furthermore, the road density layer was prepared based on the existing shapefile,⁹ also using the line density tool. Because the road degradation is very pronounced and

⁷ Translated from the original version in French.

⁸ Layer provided by the Institute of Mining and Geological Research.

⁹ Provided by the Ministry of Transportation.

Table 1 Conditioning factors in detail

Factor (unit)	ctor (unit) Basic expression		Scale (m): original / resampled	
Elevation (m)	N/A	Continuous	30/20	
Slope (angle) (°)	$\arctan \cdot \sqrt{f_x^2 + f_y^2}$	Continuous	30 / 20	
Slope (aspect) (°)	$270^{\circ} + \arctan \cdot \left(\frac{f_y}{f}\right) - 90^{\circ} \frac{f_x}{161}$	Continuous	30 / 20	
Plan curvature (m/100)	Convex value > 0; concave value < 0; Flat elsewhere	Continuous	30 / 20	
Profile curvature (m/100)	-//-	Continuous	30 / 20	
Relative slope position	Thresholding of Topographic Position Index (TPI), expressed as:	Categorical	30 / 20	
	$\overline{E_{Surrounding}}$ where E_{Pixel} is the elevation of the cell, and $E_{Surrounding}$ is the mean elevation of the neighbor pixels			
TRI	$\sqrt{ X (max^2 - min^2)}$	Continuous	30 / 20	
	where X is the elevation of each neighbor cell to a specific cell (0,0) (m), and <i>max</i> and <i>min</i> are the largest and smallest elevations among the nine neighboring pixels			
LS	$\left(\frac{A_{s}}{2213}\right)^{m} \cdot \left(\frac{\sin(\theta)}{0.0896}\right)^{n}$	Continuous	30 / 20	
	A_s is the unit contributing area (<i>m</i>), θ is the slope in radians, and <i>m</i> (0.4–0.56) and <i>n</i> (1.2–1.3) are exponents			
Valley depth (m)	N/A	Continuous	30 / 20	
Lithology	N/A	Categorical	1:5 00 000 / 20	
Fault density	$\begin{split} & [(L_1 \cdot V_1) + (L_1 \cdot V_1)] / \oplus, \text{ where:} \\ & \oplus = \text{radius;} \\ & \mathcal{L} = \text{portion of line within } \oplus; \\ & V = \text{field value} \end{split}$	Continuous	30 / 20	
Distance to fault (m)	Euclidean distance (two dimensions):	Continuous	30/20	
	$\Delta = \sqrt{[(x_2 - x_1)^2 +](y_2 - y_1)^2}, \text{ where:}$			
	$\Delta = \text{Euclidean distance;}$ (x ₂ - x ₁)= coordinate of 1 ^{rst} point, i.e., on one side of the fault; (y ₂ - y ₁)= coordinate of 2 nd point, i.e., on other side of the fault			
Soil depths (cm)	USDA global model	Continuous	250/20	
Soil types	N/A	Categorical	1:1 000 000 / 20	
Drainage density	Same as Fault density	Continuous	30/20	
Distance to stream (m)	Same as Distance to fault	Continuous	30/20	
Channel network distance (m)	Same as Distance to fault	Continuous	30/20	
SPI	$A_{\rm s} * tan \beta$	Continuous	30/20	
TWI	$Log_e\left(\frac{A_s}{tan\beta}\right)$	Continuous	30 / 20	
Mean annual rainfall (MAR)- IDW or (MFI)	$Z_{P} = \frac{\sum_{i=1}^{n} \left(\frac{z_{i}}{d_{i}^{P}}\right)}{\sum_{i=1}^{n} \left(\frac{1}{d_{i}^{P}}\right)} \text{ Or } MFI = \sum_{i=1}^{i=12} \frac{P_{1}^{2}}{P}$	Continuous	4 000 / 20	
Distance to road (m)	Same as Distance to fault	Continuous	20	
Road density	Same as Fault Density	Continuous	20	
LULC	Random Forest	Categorical	20	
SEVI	$\frac{NIR}{Red} * \left(\frac{SWIR1_{mean} - SWIR1_{min}}{SWIR1_{max} - SWIR1_{min}}\right) + 0.581 * \frac{1}{Red}$	Continuous	20	

consequently contributes to trigger hazards with multiple potholes on an extended perimeter of *1000* m on both sides of the pavement, the distance-to-road was equally prepared using the Euclidean distance tool.

In addition, a lithology layer was prepared by using the geological base map (1:500 000), scanned, reprojected,

orthorectified and digitized according to the rocks existing in the area. Those are, embrechite, syenite, rhyolite-trachyte and basalt (ORSTOM¹⁰ 1965 and 1970).

¹⁰ Office de la Recherche Scientifique et Technique d'Outre-Mer.

Therefore, a total of twenty-four (24) factors were prepared from the collected data (Appendix 4). Table 1 summarizes their expressions based on ArcGIS, SAGA-GIS, and additional literature.

Inventory of hazards

The most reliable records are found in annual reports of sub-division offices and councils, but without a proper Global Positioning System (GPS) location. Some other previous local events are mapped and documented at the Institute of Mining and Geological Research (IMGR-Yaoundé), but not regularly (e.g., Season, Year, Pace). The two sources were combined to schedule fieldwork and geolocate areas/ spots of occurrence. The most reliable records are between 1997 and 2022. Then, landslide/mudslide affected twentysix (26) points causing, 4 deaths, destroying 7 houses and making 14 homeless in Nkongsamba in 1997,11 Ideath, 12 houses, 103 homeless and 1 road in the district of Kekem in 2007 (at Moumé-Bafang), then traffic disruption toward Santchou (Ayonghe et al. 2002; Tchindjang 2013; Ngandam et al., 2021). Gully erosion is permanent on hillsides and beside paved roads along steep slopes, with aggravation and sudden collapses during every rainy season, on thirteen (13) known points. However, none of the gully's events has been declared deadly since then, although about 200 houses have been voluntarily abandoned or evicted by authorities; while more than eleven kilometers (11 km) of road became increasingly risky,¹² especially in Bangem and Melong. With regards to flash floods, it happens only during the rainy season in marshy pools and lowlands, with fifteen (15) records including about 198 abandoned or evicted houses recorded mainly in Santchou, Kekem and Nkongsamba.

The data once gathered were displayed on raw elevation and slope maps, and reclassified into five classes each, to allow comparison with the study subset. Then based on contour lines, altitudes and slope angle, the same number of points as records were created to build the most accurate and reliable geodatabase for each hazard (Appendix 5). This file has been useful to perform the learning process.

Multicollinearity and co-(in)dependency test

The multicollinearity among different conditioning factors is an indispensable step to assess their co-(in) dependency. As a basic assumption, a strong linear correlation among the factors indicates redundancy, which reduces the model's accuracy and performance, leading to errored results (Dormann et al. 2013). Among other tools, Relief-F, Chi-square statistics, Information, Linear

Support Vector Machine (LSVM), Variance Inflation Factor (VIF) and Tolerance coefficient (TOL), are dominant in the literature (Mind'je et al., 2020). The two latter, i.e., VIF and TOL were used here, as they have been previously successful in multi-hazard factors' selection (e.g., Javidan et al. 2021). They are expressed as a pair of reciprocals, and calculated as follows (O'Brien 2007) (Eq. 1 & 2):

$$VIF = \frac{1}{1 - R^2} \tag{1}$$

$$Tol = \frac{1}{VIF} \tag{2}$$

where, R^2 is the coefficient of determination of the regression between the factor of interest and the other landslide conditioning factors. As a rule of thumb to interpret these coefficients, when *VIF*>5 or 10 and *TOL*<0.1, there is a multicollinearity obstacle (O'Brien 2007; Dormann et al. 2013). As such, their values depend on the analyst and the goal of the research. In this study, since there is a high number of layers, *VIF* ≤ 3 and *TOL* ≥ 0.3 , were proposed as thresholds to maximize the probability of excluding any similarity in-between factors, per geoenv hazard. All factors were standardized, [0-1], before proceeding with the test, while the priority in including/ excluding a factor was given to the VIF values.

Core processing

Analytic Hierarchy Process

The Analytical Hierarchy Process (AHP) is a widespread multicriteria decision-making (MCDM) tool for various types of assessments, especially for geospatial applications. This approach was introduced in the 1980s by Thomas L. Saaty (1980 and 1988). It has been broadly used ever since, with successful applications in varied geospatial modelling around the world, for instance, site suitability selection (e.g., Ngandam Mfondoum et al., 2014 and 2019) or hazards prediction (e.g., El Jazouli et al. 2019; Nsangou et al. 2022).

AHP proceeds by a semiquantitative and flexible process, which involves a matrix-based pairwise comparison of different criteria or sub-criteria depending on the ramification of the tree, to solve a specific complex problem. The problem is decomposed into a hierarchical structure of several levels where, the target, main criteria, sub-criteria and alternatives are considered. A comparison is made between the criteria and the applicable alternatives, using the assessment scale from *1* to *9* based on the relative importance (Saaty 1988). Then, comparison matrices are formed and validated by using a coefficient, the consistency ratio (CR), that articulates the compatibility of all the parameters involved. The CR value should be, *< 0.1*,

¹¹ Field information and analyst knowledge.

¹² Mungo Division Transport authority.

 Table 2
 Random consistency index values

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.4	1.45

before the model is considered valid. This coefficient is defined as the consistency index (CI) divided by the random consistency index (RI), and they are expressed as follows (Saaty 1980) (Eq. 3 and Table 2):

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{3}$$

where λ_{max} is the largest eigenvalue of the considered matrix, and *n* is number of criteria.

Therefore, using the SDS v3.2, each hazard was considered as a problem to solve. Consequently, three different AHP were performed, to define a weight of evidence corresponding to each factor in triggering the targeted geoenv hazard. From the importance explained by a pairwise matrix and the structure of the hierarchy, absolute values of weights were obtained for further analysis.

Machine learning algorithmic

Three ML algorithms were used to classify the weighed and stacked factors, i.e., Classification and Regression Trees (CART; Breiman et al. 1984), Random Forest (RF; Breiman 2001) and Gradient Boosting Regression Trees (GBRT; Friedman 2002). Then, the Support Vector Regression (SVR) ML, was used for a meta-learning of the three previous.

The SVR especially is a variant of the support vector machine (SVM), and its principle is founded on the pattern's recognition. As a reminder, SVM constructs hyperplanes in a multidimensional space separating linear and nonlinear samples of different class labels (Cortes and Vapnik 1995). Special properties are built for the decision surface, to ensure high generalization ability of the learning. The output should present a separating hyperplane or decision boundary, that maximizes the distance between these hyperplanes and the classes sample, as summarized below (Eq. 4):

$$u = w \bullet x - b \tag{4}$$

where *w* is the normal vector to the hyperplane and *x* is the input vector. The separating hyperplane is the plane u = 0. The nearest points lie on the plane $u = \pm 1$. The margin is thus (Eq. 5):

$$m = \frac{1}{\|w\|_2} \tag{5}$$

The optimization parameters inside the margin *m* are cost, penalty and gamma (*C*, ξ , γ).

From the SVR perspective,¹³ there are three new considerations: an additional tunable parameter epsilon (ε) which determines the width of the margin *m*; support vectors which are the points that fall outside rather than just the ones at the margin *m*; penalty parameter ξ which measures now the distance to points outside the margin *m*, with the analyst controlling the regularization of parameter *C*. This may allow for nonlinear relationships by fitting smooth transformation functions and adjusting deviations between classes of hazard prediction by stacked factors. From these principles and functionalities, SVR appeared as a good alternative of metric-based meta-learning to combine and adjust CART, RF and GBR outputs.

First combination, stacking and sampling

The weighted linear combination (WLC), also known as Simple Additive Weight (SAW) (Hwang and Yoon 1981), was performed among all standardized factors weighted by their AHP coefficient, to obtain the first preview of each hazard. The output was reclassified into five classes using the Jenks natural breaks method. Then the sampling was conducted for automatic learning using the events' inventory and distributed following a 50/50 ratio as shown in Table 3 and Appendix 4.

Therefore, the scaled-weighted factors were stacked and reexported to the GEE platform to perform CART, RF and GBRT learning.

Further, to run the SVR meta-learning process, other stackings were done for each output of CART, RF and GBRT, i.e., landslide, gully erosion and flash flood. As such, a total of nine (09) SVR were conducted. At the difference of the basic learners, only the recorded events sample was used, for a smoothing process picturing real occurrences.

Multi-geoenvironmental hazard susceptibility (MGHS) model The final step was the MGHS model. Using a simple linear combination (SLC), all the hazards were added per primary ML algorithm, i.e., CART, RF and GBRT. Further, the meta-learning consisted in summing the three

¹³ https://towardsdatascience.com/support-vector-regression-svr-one-ofthe-most-flexible-yet-robust-prediction-algorithms-4d25fbdaca60. Visited on September 18, 2022.

Table 3 Sampling distribution for basic learners

Hazard	Sample	Total		
	Recorded	Generated	Both	
Landslide	26	26	52	
Gully Erosion	13	13	26	
Flood	15	15	30	

previous SVR outputs to produce the improved MGHS. Then, all the results were exported to the ArcGIS environment, to proceed with area calculation, hazards association preview, further analysis and layouts.

Human exposure mapping process

Beyond the artificial settlements and vegetation cover that involve human impact (road, LULC, SEVI), the number of residents has been integrated as the ultimate layer to cross with each hazard, and better express the life-threatening status. Human exposure is herein defined as the intersection of a hazard susceptibility map and a spatially distributed population layer (Bernhofen et al. 2021). As such, population density has been previously used from the settlements' perspective (e.g., Modugno et al. 2022).

Practically, the population shapefile per locality, was constructed in point features from the regional offices of the central bureau of the Census and population studies (BUCREP 2010), and then interpolated using the IDW tool. A simple linear combination (SLC) was performed with each output of the SVR. The results were reclassified to obtain the final classes of hazards' direct threat on human life, i.e., vulnerability.

Performance assessment and comparison of models

- Confusion matrix: The confusion matrix includes true positive (TP), false positive (FP), true negative (TN), and false negative (FN) categories (Frattini et al. 2010). The value calculated from the confusion matrix provides useful information on model performance and classification accuracy, based on probabilities. In this study, the balanced F-score (F₁-score), true positive rate (TPR) and false positive rate (FPR) were exploited to assess the performance of each learning from the final output (Eq. 6–9).
- Receiver Operating Characteristic: The receiver operating characteristic (ROC) or success rate curve, is a visualization tool which helps to validate the rationality and robustness of a probabilistic model (Pradhan 2013; Wu et al., 2016). The ROC curve is plotted by

statistical index value pairs, with the sensitivity or TPR on the x-axis and, "1-specificity" or FPR on the y-axis. These two values are used as cut-off-dependent metrics, associated with the area under ROC (AUC/AUROC) which is the cut-off-independent metric for accuracy and efficacy evaluation of predictive model outcomes (Eq. 9):

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + fn} \tag{6}$$

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

$$AUC = \int_{0}^{1} f(FPR)d(FPR) = 1 - \int_{0}^{1} f(TPR)d(TPR)$$
(9)

Figure 2 summarizes the main steps used to complete the methodology.

Results

Multicollinearity analysis

Globally, the VIF of different pixel sizes was around $1e5^{th}$, indicating no collinearity problem among factors, and consequently, they could contribute to subsequent modelling processes (Garosi et al. 2019). For landslide, lithology recorded the highest VIF, i.e., 2.223, and the lowest TOL value, i.e., 0.449. For gully erosion, the soil depth appears to be the most adequate factor, with a VIF value of 2.649 and a TOL value of 0.377. Then for flash flood, the plan curvature factor has the best fitting values to run the model, with 2.557 and 0.39 for VIF and TOL, respectively. While these values remained inside the targeted thresholds (i.e., $VIF \leq 3$ and $TOL \geq 0.3$), the relative slope position factor was excluded from the flash flood modelling, for a high VIF, i.e., 3.078, although a good TOL of 0.325.

To sum up, all 24 factors have been used in the modelling of landslide and gully erosion (Table 4). Contrarily, only 18 factors have been used for the flash flood modelling, with the exclusion of the relative slope position factor for high collinearity above the thresholds fixed (Table 4); while LS, fault density, distance to fault, distance to road and road density factors are usually not integrated with the flood modelling, according to the literature and confirmed by the analyst knowledge of this hazard's occurrence in the area.



Fig. 2 Flowchart summarizing the used methodology

Pre-/post-machine learning importance of conditioning factors

The pre-learning importance of factors is considered here as the AHP outputs. Whereas the post-learning importance refers to the ML explanations of each factor's contribution to the model.

The AHP values are shown by Fig. 3. From the analyst knowledge-based assessment of the study area, the rainfall is mainly responsible of all three hazards, with the highest value, 1, for all. The lowest coefficients are LULC for landslide (0.391) and flash flood (0.296), and

SEVI for gully erosion (0.346). These values were confirmed acceptable by the objectivity test introduced by the consistency ratio interval, $0.0209 \le CR \le 0.0256$, way below the allowed threshold, 0.1.

Automated learning gives different configurations of importance. For landslide susceptibility assessment, elevation has been the dominant factor following CART and RF learning, with 44.54% and 6.4%, versus rainfall factor (9.71%) according to GBRT learning. Concerning gully erosion, elevation (48.08%), TRI (6.51%) and slope (10.74%) are the three dominant

Table 4 Collinearity of factors

Landslide	Tol	VIF	Gully erosion	Tol	VIF	Flash flood	Tol	VIF
Elevation	0.685	1.459	Elevation	0.559	1.786	Elevation	0.413	2.415
Slope	0.699	1.429	Slope	0.679	1.472	Slope	0.391	2.556
Slope aspect	0.866	1.153	Slope aspect	0.875	1.142	Slope aspect	0.893	1.118
Plan curvature	0.871	1.147	Plan curvature	0.878	1.138	Plan curvature	0.39	2.557
Profile curvature	0.808	1.236	Profile curvature	0.806	1.239	Profile curvature	0.515	1.938
Relative slope position	0.472	2.115	Relative slope position	0.689	1.45	Relative slope position	Excl	Excl
TRI	0.494	2.022	TRI	0.544	1.835	TRI	0.4	2.496
LS	0.79	1.265	LS	0.78	1.281	LS	N/A	N/A
Valley depth	0.484	2.065	Valley depth	0.543	1.841	Valley depth	0.409	2.44
Lithology	0.449	2.223	Lithology	0.585	1.706	Lithology	0.482	2.073
Fault density	0.808	1.236	Fault density	0.792	1.261	Fault density	N/A	N/A
Distance to fault	0.835	1.196	Distance to fault	0.834	1.198	Distance to fault	N/A	N/A
Soil depth	0.6	1.664	Soil depth	0.377	2.649	Soil depth	0.654	1.528
Soil type	0.542	1.842	Soil type	0.387	2.578	Soil type	0.556	1.797
Drainage density	0.548	1.822	Drainage density	0.56	1.783	Drainage density	0.663	1.506
Distance to stream	0.751	1.331	Distance to stream	0.771	1.296	Distance to stream	0.763	1.309
Channel network distance	0.539	1.853	Channel network distance	0.527	1.896	Channel network distance	0.599	1.667
SPI	0.974	1.026	SPI	0.977	1.023	SPI	0.984	1.016
TWI	0.72	1.387	TWI	0.721	1.385	TWI	0.87	1.149
Rainfall-IDW	0.511	1.953	Rainfall-MFI	0.648	1.542	Rainfall-IDW	0.744	1.343
Distance to road	0.59	1.694	Distance to road	0.617	1.618	Distance to road	N/A	N/A
Road density	0.646	1.546	Road density	0.652	1.532	Road density	N/A	N/A
LULC	0.91	1.098	LULC	0.891	1.121	LULC	0.933	1.071
SEVI	0.903	1.106	SEVI	0.9	1.11	SEVI	0.908	1.101

The bold values has been used for all numbers representing the two collinearity metrics used here, i.e., Tolerance (Tol) and Variance Inflation Factor (VIF) Excl. = layer excluded for a too high VIF (3.078), although a good TOL (0.324); N/A=Not Applied to the model



Fig. 3 AHP coefficients for each factor per hazard

factors according to CART, RF and GBRT respectively. With regard to flash floods, elevation (26.64%), profile curvature (11.73%) and slope (20.78%) have been the most impacting factors, according to the same order of ML algorithms. CART depicted large gaps between the highest and lowest percentages of conditioning factors, with some values almost nil, i.e., 0.006%, contrary to RF and GBRT which showed a progressive difference between factors (Fig. 4).

Validation of models using ROC, AUC and F1-score

The ROC was plotted and the default cut-off value, **0.5**, was fixed for all the learners. It was completed by analyzing the area under the curve (AUC), which qualitatively assesses the prediction accuracy, i.e., the functionality of the models. A ROC value closer to **1** at the top left corner of the graph implies a larger proportion of the AUC and suggests that the model is more accurate. In contrast, a smaller AUC value indicates a poor accuracy of the



Fig. 4 Automatically computed importance of factors per hazard and per learning

model in predicting future occurrences of the phenomena, i.e., single or all hazards (Sun et al. 2019; Rehman et al. 2022). Statistics of the outputs in five classes have been used in this section.

At a general first glance, all the ROC are close to **1** (Fig. 5). For landslide learning, AUC values are between [0.967–0.972], with CART recording the lowest value and RF the highest one. The AUC values for gully erosion learning showed values between [0.976–0.978], with the lowest extremity for CART and the highest for RF. While for flash flood learning, AUC values reflected the highest accuracy, between [0.989–0.991], with GBRT and RF at low and high extremities.

Concerning the meta-learner, SVR showed the lowest values of AUC for landslide, i.e., 0.964, the highest for gully erosion (0.979) and the lowest value same as for GBRT, i.e., 0.989. But generally, there are very small differences among the AUC values of ML models for each

geo-env hazard occurrence, while the SVR accuracy and performance are the consensual accuracies among other learners. When it comes to the MGHS learning model, the maximum AUC, *1*, was reached by all the classifiers including SVR, for all classes.

Moreover, the interpretation of the F_1 -score allowed to show that single geo-env hazards were assessed high enough to be relevant, with basic ML algorithms of 0.86 for gully erosion and flash flood, and 0.87 for landslide (Table 5). The same values have been recorded by SVR for gully erosion and landslide but increased to 0.88 for flash floods. Concerning the MGHS, values were a little lower, between [0.82–0.84] (Table 5).

Spatial distribution and statistical trends of single geo-env hazards susceptibility

A total of twelve (12) maps were derived, i.e., CART-Landslide, RF-Landslide, GBRT-landslide, SVR-Landslide,



Fig. 5 ROC curves and corresponding AUC values

CART-Gully Erosion, RF-Gully Erosion, GBRT-Gully Erosion, SVR-Gully Erosion, CART-Flash Flood, RF-Flash Flood, GBRT-Flash Flood, SVR-Flash Flood (Fig. 6). Five classes of susceptibility were defined, such as "Very Low", "Low", "Moderate", "High" and "Very High". The assessment uses the preview with the main focus on the two latter classes (Fig. 7).

Landslide mostly affects the western half according to the three classifiers. However, the spatial distribution trends show more similarity between classes for RF and GBRT, than with CART. Further, at least 30% representing the largest percentage of the area is exposed to "Very High" susceptibility according to RF (30.9%) and GBRT (31.2%), versus only 21.6% for CART, which learning estimated the "High" susceptibility more distributed, i.e., over 28.7% of the area. Moreover, summing the two highest susceptibility classes gives 50.3% of the area according to CART, but only 47.3% and 49.1% for RF and GBRT respectively.

Table 5 F₁-score values

Model	Landslide	Gully erosion	Flash flood	MGHS
CART	0.87	0.86	0.86	0.82
RF	0.87	0.86	0.86	0.83
GBRT	0.87	0.86	0.86	0.84
SVR	0.87	0.86	0.88	0.83

Gully erosion on its side appears sparsely distributed. All classifiers share five major spots, which are located in the North-west, West, South, East and Centre. However, CART and GBRT properly exhibit the spots at the Centre, more than RF does. The "Very High" susceptibility class represents 15.8%, 18.5% and 20.2% of the area according to CART, RF and GBRT respectively, versus 19.2%, 22.1% and 18% for the "High" class, in the same order. In contrast to landslide, these two classes together



Fig. 6 Single geoenvironmental hazard spatial distribution

represent less than 50% of the area, i.e., 35%, 40.6% and 38.2% according to CART, RF and GBRT respectively. Then, between 60 and 65% of the area are lowly to moderately prone to gully erosion.

Concerning flash floods, it is demarcated to the East half of the study area according to all classifiers. The "Very High" susceptibility class occupies 20.9%, 22.3% and 22.5% of the area for CART, RF and GBRT

respectively. Whereas 10.6%, 13.8% and 14% represent the ratios of the "High" susceptibility class in the same order of learners. The total for these two classes is 31.6%, 36.1% and 36.6%, suggesting that 64% to 69% of the area are is lowly to moderately concerned by flash floods.

The overall output of the meta-learner SVR arbitrated the spatial distribution of susceptibility per geo-env hazard, by taking the most dominant features according to



Fig. 7 Areas of single geoenvironmental hazard per learner

the learning process and with consequences on the percentage of areas per class. For landslide, "Very High" and "High" susceptibility classes, represent 32.30% and 15.91% respectively, i.e., 48.21% of the area. Concerning gully erosion, 13.73% and 20.99% are the ratios for the same order of classes, i.e., 34.72% of the area. Whereas for flash floods, 17.31% and 16.10% are the ratios, i.e., 33.41% of the area.

In addition, discrepancies among classes are widely effective for flash floods, but less noticeable for landslide and gully erosion. As such, CART-Flash Flood, RF-Flash Flood and GBRT-Flash Flood previews, show the largest spatial distribution for the "Low" susceptibility class, confirmed by statistics, i.e., 39.37%, 39.92% and 43.47% respectively. Whereas, SVR-Flash Flood specifically highlights the "Very Low" susceptibility class to be the most extended, which is confirmed by its largest percentage, i.e., 40.75% of the total area.

Multi-geoenvironmental hazard susceptibility (MGHS)

Combinations of geo-env hazards were conducted according to the following SLC:

$MGHS_{ML} = Landslide_{ML} + GullyErosion_{ML} + FlashFlood_{ML}$

where *ML* refers to the targeted machine learning algorithm. To properly discriminate the different contributions, computations were conducted for different pairs and then for the three geo-env hazards, ML. The outputs combination obtained are therefore, "Landslide+Gully-Erosion", "GullyErosion+FlashFlood", "Landslide+Flash-Flood" and "Landslide+GullyErosion+FlashFlood", also named here as "All".

The analysis is based on the highest classes of the outputs (Figs. 8 & 9). At a glance, the susceptibility of combined landslide and gully erosion has the largest spatial cover, towards the northern, western and southwestern areas. From the basic ML, they are predicted to

simultaneously occur over [17.5-24.4%] of the subset, with the lowest percentage for CART and the highest for RF. Whereas, for SVR, they are estimated to occur over 21.4% of the subset (Fig. 10). The second obvious combination is the landslide and flash flood susceptibility of occurrence, which is most likely to happen in the center, western and south-western areas. They are extended between [5.3-8.9%] of the subset (Fig. 10), from RF to CART for basic ML; while SVR gives the same value as RF, i.e., 5.3% of the subset. The "GullyErosion+Flash-Flood" susceptibility combination is less spatially distributed, mostly at the center area, with a south-to-west expansion. The percentages vary between [1.2-3.4%], from RF to CART for basic ML; while SVR estimates that to 1.1%. At last, the combination of all geo-env hazards shows at first sight the same distribution patterns as the latter, with percentages [0.6-1.4%], from RF to CART/ GBRT; whereas SVR estimates this ratio at 0.7%.

Vulnerability status

Outputs for the vulnerability analysis are presented here as heatmaps, which have been suggested to better support rapid response or mitigation for hazards (Handwerger et al. 2022). Generally, the population distribution matched the hot spots of each geo-env hazard occurrence (Fig. 11). This may lead to assume that high human concentrations are globally the most exposed. This assumption has been verified in the main city of Nkongsamba, where an extract of the urban perimeter shows a concerning exposure of population to studied geo-hazards.

At the center of this urban sub-area, vulnerabilities to landslide and gully erosion are expanded in the South-West/North-East direction (Fig. 11). The sum of "Very High" and "High" hot-spot classes, represents 36.5% of the extract for landslide, and 28.95% for gully erosion (Fig. 12). However, the spatial distribution of the two selected classes is inverted for the vulnerability to these two geo-env hazards. While the vulnerability to



Fig. 8 MGHS distribution per combination of occurrence – Only the highest value is highlighted here



Fig. 9 MGHS summarized distribution – highest values are highlighted; lowest ones are combined into one class



Fig. 10 MGHS areas per learner. Ratios are calculated from the highest value area that highlight probability of occurrence on one hand, and the sum of "Very Low" to "Moderate" areas



Fig. 11 Vulnerability heatmap using SVR outputs for the whole subset and the extract of Nkongsamba



Fig. 12 Ratios of vulnerability for the extract of Nkongsamba. The class entitled "Other" includes the "Very Low", "Low" and "Moderate" classes

flash flood is highlighted in the southern center of the extract, at the complete opposite of the two previous, although still South-West/North-East orientated. The total area of flash floods vulnerability for the "Very High" and "High" hot-spots classes, represents *31.02%* of the extract. Eventually, the most vulnerable spaces are widely expanded in the South-West/North-East direction, with a more important visual matching for both landslide and gully erosion, versus smaller spots of the three simultaneous vulnerabilities in the south. The "Very High" and

"High" for all of them cumulate a total of *32.54%* of the urban sub-area. The three mainly concerned localities are Nkongsamba, Baréock and Baré (Fig. 11).

Discussions

Inference from the findings and comparisons to previous studies

The weighing of factors with the AHP forecasted their contribution per geo-env hazard's model, differently from usual empirical simple (factor * 1) or percentage (factor * 1/100) weighing in a raster calculation. For simultaneous occurrence, the same factor is assumed to not affect two or more, recorded or predicted geo-env hazards at the same level. Rather, the level of impact in triggering or aggravating may vary from nil (excluded) to fully, with different values. Moreover, with the usage of SLC in a single geo-env model and multi-hazards analysis, that impact always varies as previously proven (Javidan et al. 2021; Nsangou et al. 2022). As side mention, the seasonality of occurrences based on fieldwork and existing records state that landslides and flash floods only happen during the rainy season. While gully erosion takes place in any season, as they are more related to topography, lithology, and changes in human land use impacting land (vegetation) cover, and then easily triggered or



Fig. 13 Sampled illustrations. From left to right: Road National-5 collapsed after a rainfall-triggered landslide in Kekem (October 2007) (Balla et al. 2013); landslide in Nkongsamba, near administrative area sector-1 (September 2022); another section of Road National-5, flooded and collapsing in Santchou, following the overflows of the river Elienti after a huge rainfall (September 2020) (http://www.cameroon-info.net/article/cameroun-inondation-la-circu lation-est-perturbee-sur-la-route-santchou-dschang-a-la-suite-de-381012.html); houses and road flooded in Nkongsamba, sector Massa Ngah Junction-Market "C" after a huge rainfall (September 2022) (http://scores2000.info/2020/09/grandes-inondations-et-crues-exceptionnelles-sur-certa ines-zones-de-la-ville-de-nkongsamba.html)

emphasized by rainfall, rejoining previous remarks (Aber et al. 2010). Figure 13 illustrates some recent events in the area.

Further, performing basic ML algorithms over AHPweighed and stacked factors is another novelty of the proposed method. The advantage of this approach is the agreement of most ML for each model, then a potential accuracy improvement for a single modelling process, since each geo-env hazard susceptibility is highly dependent on the spatial distribution of real occurrence while including the importance of every weighed factor, according to an intelligent automatic process of detection. As such, there was a global spatial agreement of basic learners, i.e., CART, RF and GBRT per geo-env hazards, supported by values of AUC and F₁-score, higher than those obtained by Sheikh et al. (2019) by using the TOPSIS-MD, TOPSIS, Simple Additive Weight (SAW), and by Javidan et al (2021) by using MaxEnt. Likewise, probability of occurrence of pairs geo-env hazards is always higher than all at the same time, in accordance with the results obtained by Pourghasemi et al (2019), who assessed floods, earthquakes, and landslides.

Furthermore, introducing a regression meta-learning process over stacked basic learners' outputs to assess the spatial trends of geo-env hazards of occurrence, better displays spatial agreement than single learning. The results obtained and values of validation tools (AUC, F_1 -score) for SVR, are the same or higher than those recently obtained by Bordbar et al (2022), who combined ensembles learning and meta-heuristic techniques to flood, landslide and Earthquake occurrence.

At this point, a complementary validation of the basiclearner/meta-learner approach might be important, in addition to AUC and F_1 -score. Since SVR is assumed to arbitrate, improve and adjust the best spatial and statistical distributions of the CART, RF and GBRT in each case, likewise, the highest correlation of a basic learner with the meta-learner will indicates, the contribution and the reliability of the first to the second. Then, a correlation analysis was conducted on the classes value using the Spearman coefficient. Overall, the correlation coefficients are very high (0.921 $\leq \rho \leq 0.985$) (Fig. 14). Specifically, RF has the highest reliability to predict single geo-env hazard [0.933 $\leq \rho \leq 0.985$]. Whereas GBRT is equally performant as RF in mapping flood with ($\rho = 0.933$) and the highest proficiency when it comes to the MGHS modelling ($\rho = 0.933$). Then, there are different methodology options in the North-Moungo perimeter hazards study, and the analysis may select one of these learning processes, depending on the goal or accuracy, to reduce the computation length and steps.

In addition, geo-env hazard occurrence was predicted over all land covers, including vegetation (See Fig. 4). One logical explanation is that, topography (elevation, slope steepness), combined to volcanic lithology and pedology (less evolved soils; Gèze, 1942), globally impacts the landscape instability, more than the vegetation does. This has been highlighted and confirmed by the importance of factors analysis (See Fig. 6). Whereas on another hand, although afforestation is most of the time an important and suggested action for landslide, erosion and flood control, the vegetation composition of the area, which transits from degraded forest to shrubs and grasses savannas, deeply impacts the susceptibility of occurrence for each cover. Matter of fact, it has been previously demonstrated that in a space with mixed vegetation patterns, the geo-env hazards gradient of occurrence accordingly varies (e.g., Miles and Swanson 1986; Carone et al. 2015). As such, areas covered by grasses and herbs will be more affected, than those covered by shrubs and trees.

Another consideration is the spatial similarities pointed out on single susceptibility maps, between landslide and gully erosion expansions than with flash flood are to be examined (See Fig. 6). These trends were empirically proven in previous studies (e.g., Sun et al. 2019), and have direct implications on MGHS maps, with the largest space affectation to the couple "Landslide+Gully Erosion" and "Landslide+Flash Flood", than "Gully



Fig. 14 Correlation between meta-learner and single learners

Erosion + Flash Flood". As a consequence, landslide is to be ranked as the first danger, followed by gully erosion and flash flood. Likewise, the differences among single susceptibility maps, confirm not only the robustness of the model in accurately separating/connecting their areas of occurrence, but also the spaces needing more attention for the implementation of control/protection tools specific to each geo-env hazard.

From a spatial-scale viewpoint, it has been stated that multi-hazard susceptibility assessments are more efficient on larger areas (small scale), while multi-hazard risk assessments are practical at the site scale and on an event-based scheme (Motamedi and Liang 2014). This has been verified with the vulnerability assessment. From the extract of Nkongsamba, the population vulnerability to each geo-env hazard shows three different extents. But when it comes to simultaneous exposure, landslide and gully erosion visually represent a higher direct risk of occurrence, while there is a spatially and statistically small exposure of the population to both, plus flash flood at the same time. The outputs of population exposure support the idea of a one-per-one vulnerability analysis, so as not to neglect a geo-env hazard impact on the population (e.g., flash flood vulnerability in the South-Center of Nkongsamba), when focusing on the other that shows a different spatial distribution (e.g., landslide and flood vulnerability in the northern center of Nkongsamba).

Limitations of the study

Accuracy of the inventory was the first obstacle to building the geodatabase. In written reports, there are a lot of gaps in the inventories that lack a systematic follow-up. This certainly leads to some misanalysis if the disaster has not been exactly located, as well as different deviations in the spatial density distribution. Then, because the existing events are, for some, inaccurately distributed, the sampling had to be completed by creating new points based on environmental similarities than real event occurrences, according to the analyst's knowledge of the study area, before proceeding to the learning. Consequently, we were not also able to assess the accuracy in terms of the deviation between the real and the estimated location of the event.

Moreover, most satellite ready-to-use geodatabases of natural hazards are not extended everywhere or have not considered the study subset as exposed. For instance, a detailed landslide geodatabase exists on the USGS website, but only for the United States, while we were also not able to access the global landslide susceptibility prepared by Felsberg et al (2022), as well as the global sensitivity index proposed by Van Natijne et al (2022). Another example is the global soil erosion modelling (GloSEM) proposed by Borrelli et al (2017), using the Revised Universal Soil Loss Equation (RUSLE), but our request form was not answered. As a last illustration, the global flood database is more general and targets very large areas which do not include the Moungo's studied subset of the CVL as (enough) exposed. This may find an explanation in its small scale (continental), based on MODIS imagery which is coarse.¹⁴ The availability of such data would have been helpful to conduct a proper accuracy assessment of each classifier, based on the error matrix, as a plus to the process validation.

Outlooks and local mitigation orientations

The main technical activity envisages the direct interactions between pairs of geo-env hazards, as well as all together. However, this study has not considered the aspect of their direct chronological occurrence, as well as their (spatial, mechanical) relationship, i.e., one could trigger the other(s). For instance, on a flow continuum scale and under heavy rainfall, the usual sediment transport of a river's/stream's flow, can quickly escalate a flash flooding, that in its turn, will trigger huge debris flow, i.e., mudslides/landslides.¹⁵ It is more common that gully needs then to be established in a fully different modelling

process. Other future activities involve supporting the cities' council to build a flexible and updateable database of local disasters. This includes the exact geolocation of different events, with instant mapping, to complete the reports and avoid potential misanalysis or other biases. Practically, mitigation actions should be adapted to the highest scale, i.e., local context, more than the regional one. Based on the case of Nkongsamba, the purpose of mitigation is to properly and efficiently support the efforts of the Department of Civil Protection (DCP)¹⁶ in the actions to prevent disasters. Among others, the most used is the population eviction from sites judged prone to disasters (hillslopes, marshy lowlands, etc.). Whereas, structures, documents and conventions such as the National Observatory of Hazards (ONR) and National Plan of Contingency (PNC) (MINATD 2007). This should be preceded or supported by a proper mapping integrating all the variables used in this study or even more at the city scale, then, stacked/joined to the urban planning. Unfortunately, the process is not harmonized and has not been set as mandatory for the councils, in such a way very few of them are moderately ready to face the occurrence of disasters.

Above all, the scope of this study has been to produce visual support for policymakers, in reducing losses and damages, building resilience and targeted mitigation actions with a more explicit focus on populations, their health and livelihoods (e.g., forecasting, warning system, relocation) (UNISDR 2015b). This is challenging at the national scale, i.e., Cameroon, considering the very varied panorama of geo-env hazard risks and implications requiring details for efficient management (Bang 2022).

Conclusion

Susceptibility mapping is a fundamental tool for disaster management and planning development activities in hilly regions of tropical and subtropical environments. This study has proposed to combine AHP with ML algorithms, to produce better performances in building a tool named the multi-geoenvironmental hazards susceptibility (MGHS), as well as assessing human vulnerability. Twentyfour factors were selected to maximize the accuracy, all were used for landslide and gully erosion modelling, while only eighteen were integrated into the flood modelling according to the multicollinearity analysis (VIF and TOL)

¹⁴ https://global-flood-database.cloudtostreet.ai/#interactive-map.

¹⁵ https://www.usgs.gov/programs/landslide-hazards/landslides-101.

¹⁶ Ministry of Territorial Administration and Decentralization (MINATD).

and the literature review. Based on the validation of ROC trends, AUC and F₁-score, the model demonstrates a certain robustness. The approach of using three single learners, CART, RF and GBRT, as well as a meta-learner, i.e., SVR, has also been beneficial to assess the most dominating trends. Whereas, the MGHS tool stands for different combinations of two or three simultaneous occurrences. Nevertheless, with the limitations related to this study, such as, the location of some events or the non-coverage by internationally reliable satellite geodatabases for a complementary validation, accuracies record of the outcomes/ outputs suffers drawbacks and caveats that are still in work to be addressed. Overall, based on the findings, the process of weighing factors before learning has been beneficial to avoid the arbitrary preeminence of one on another. Likewise, and compared to studies based on ensemble learning, SVR demonstrated the ability to better predict multi-geohazards, according to its regressive process and using samples of real events, all of which make it testable for many other applications. In addition, by integrating the population density layer at last, the result of the urban sub-area of Nkongsamba clearly stated that the space planning did not previously take the risk of geo-env hazards, explaining a predicted large human exposure. Since the inhabitants are constructing everywhere, they are susceptible to disasters and vulnerable to death or eviction. Future planning is encouraged to imperatively integrate the geo-env hazard aspect, to prevent all inconvenience and threats to locals.

Appendix 1: Sentinel2 granules information sorted by recording time

Granule Full Identification	Time
	09:33:19
S2B_MSIL1C_20211222T093319_N0301_R136_ T32NPL_20211222T113237	09:33:19
S2A_MSIL1C_20211220T094411_N0301_R036_ T32NNM_20211220T113438	09:44:11
S2A_MSIL1C_20211220T094411_N0301_R036_ T32NNL_20211220T113438	09:44:11

Granule Full Identification	Time
	09:49:22

The sensing orbit direction is descending.

Appendix 2: Sentinel2 bands information

Name	Range (µm)	Bands (Resolution/m)	
Coastal aerosol	0.421—0.457	B1(60)	
Blue	0.439—0.535	B2(10)	
Green	0.537—0.582	B3(10)	
Red	0.646—0.685	B4(10)	
Red_edge1	0.694—0.714	B5(10)	
Red_edge2	0.731—0.749	B6(10)	
Red_edge3	0.768—0.796	B7(10)	
NIR _{wide}	0.767—0.908	B8(10)	
NIR _{narrow}	0.848-0.881	B8A(20)	
Water vapor	0.931—0.958	B9(60)	
Cirrus	1.338—1.414	B10(60)	
SWIR1	1.539—1.681	B11(20)	
SWIR2 2.072-2.312		B12(20)	

Appendix 3a: Validation error matrix

	Water	Open Vegetation	Built-up	Soil	Total
Water	28	0	0	2	30
Open Vegetation	0	75	0	0	75
Built-up	0	1	33	16	50
Soil	0	3	2	33	35

Appendix 3b: Consumer's accuracy

Class	Value
Water	1
Open Vegetation	0.95
Built-up	0.94
Soil	0.625



Appendix 4: The conditioning factors map from original/unscaled values



Appendix 5: Samples distribution for the learning process per geoenvironmental hazard

Acknowledgements

This paper was initiated by the first author, AHNM, remembering that, on Sunday 7th of September 1997, in Nkongsamba-sector "11-Bis", a huge rainfall-triggered landslide/mudslide, killed the four children of his neighbor and elementary school friend, named Dorethe Yonga. In support of the Nkongsamba city councils' local efforts, for a proper integration of the geoenvironmental hazards component in urban planning. In memory of all the deaths due to disasters in Cameroon. As another scientific contribution to the knowledge of challenges along the Cameroon Volcanic Line. Thank you to the heads of our respective laboratories and departments for the pieces of advice. Special gratitude to the reviewers for their insightful comments and suggestions for the improvement of the manuscript.

Author contributions

AHNM initiated the study, wrote the first manuscript and designed main mapping; AHNM, PWN, DTS and JVMM contributed to data collection and curation, method conception and validation; AHNM, HBN and ID discussed contents, literature review and references sourcing; MT, BD, AM, SH, RB, FCLT, ICNP and LMB joined the study and provided complementary data and improvement comments. All the authors contributed to manuscript editing, structuring and critical reviewing. All authors read and approved the final manuscript.

Funding

The Non-Profit Scientific Association Stats'NMaps, through its 'Research Support Funds' (S'NM-RSF), financially supported the purchase of articles and book chapters on sale, as well as all the fieldwork investigations for complementary data acquisition. The licensed software, Erdas Imagine version 16.6.0.1366, ArcGIS version 10.8.2, SAGA-GIS version 8.0.1, Super Decision Software (SDS) version 3.2 and, XLStats version 2020.1.64570, used in this research are properties of Stats'NMaps and available upon request.

Availability of data and materials

All the data and material used for this research were incorporated as needed to the text or added as appendixes. All other detailed data and script links are available upon request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Stats'NMaps, Non-Profit Scientific Association, Dallas, TX, USA. ²Department of Geography, University of Yaoundé I, Yaoundé, Cameroon.³Institute of Mining and Geological Research, Yaoundé, Cameroon. ⁴Department of Statistics and Econometrics, University of Toulouse Capitol 1, Toulouse, France. ⁵Department of Disaster and Emergency Management, School of Energy, Construction and Environment, Sir John Laing Building, Coventry University, Coventry CV1 5FB, UK. ⁶Department of Physics, Laboratory of Oceanography, Environmental Sciences and Climate (LOSEC), Sciences and Technologies Research and Formation Unit (UFR-ST), Assane SECK University of Ziguinchor, 270000 Ziguinchor, Senegal. ⁷Global Mapping and Environmental Monitoring-Cameroon Branch, Street of Melen, Yaoundé, Cameroon. ⁸Department of Geography, University of Maroua, Maroua, Cameroon.⁹Department of Natural and Life Sciences, Faculty of Exact Sciences and Natural and Life Sciences, Larbi Tebessi University, 12002 Tebessa, Algeria. ¹⁰Geoscience, Water and Environment Laboratory, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco. ¹¹Department of Geography, University of Douala, Douala, Cameroon.

Received: 30 October 2022 Accepted: 30 September 2023 Published online: 13 October 2023

References

Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC (2018) Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci Data 5:170191. https://doi.org/10.1038/sdata.2017.191

- Aber JS, Marzolff I, Ries JB (2010) Gully erosion monitoring. Small-Format Aerial Photogr. https://doi.org/10.1016/B978-0-444-53260-2.10013-4
- Alfieri L, Bisselink B, Dottori F, Naumann G, De Roo A, Salamon P, Wyser K, Feyen L (2017) Global projections of river flood risk in a warmer world. Earth's Fut 5(2):171–182. https://doi.org/10.1002/2016EF000485
- Al-Juaidi AEM, Nassar AM, Al-Juaidi OEM (2018) Evaluation of flood susceptibility mapping using logistic regression and gis conditioning factors. Arab J Geosci 11(24):765. https://doi.org/10.1007/s12517-018-4095-0
- Arabameri A, Chen W, Loche M, Zhao X, Li Y, Lombardo L, Cerda A, Pradhan B, Bui DT (2020) Comparison of machine learning models for gully erosion susceptibility mapping. Geosci Front 11(5):1609–1620. https://doi.org/10. 1016/j.gsf.2019.11.009
- Arnoldus H (1980) An Approximation of the Rainfall Factor in the Universal Soil Loss Equation. In: De Boodt M, Gabriels D (eds) Assessment of Erosion. John Wiley and Sons, New York, pp 127–132
- Ayonghe SN, Suh CE, Ntasin EB, Samalang P, Fantong W (2002) Hydrologically, seismically and tectonically triggered landslides along the Cameroon volcanic line. Cameroon Afr Geosci Rev 9(4):325–335
- Balla A, Kagou Dongmo A, Nkouathio DG, Ngapgue F (2013) Land instability in the highlands of western Cameroon: geological and geotechnical characterization of the Kekem landslide. Bull Inst Sci 35:39–51
- Bang HN (2022) A concise appraisal of Cameroon's hazard risk profile: multihazard inventories, causes, consequences and implications for disaster management. GeoHaz 3(1):55–87. https://doi.org/10.3390/geohazards 3010004
- Bang HN, Miles L, Gordon R (2019) Evaluating local vulnerability and organisational resilience to frequent flooding in Africa: the case of Northern Cameroon. Foresight 21(2):266–284. https://doi.org/10.5897/JASD2018.0510 Bernhofen MV, Trigg MA, Sleigh PA, Sampson CC, Smith AM (2021) Global flood exposure from different sized rivers. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-2021-102
- Bordbar M, Aghamohammadi H, Pourghasemi HR, Aziz Z (2022) Multi-hazard spatial modeling via ensembles of machine learning and meta-heuristic techniques. Sci Rep 12:1451. https://doi.org/10.1038/s41598-022-05364-y
- Borrelli P, Robinson DA, Fleischer LR, Lugato E, Ballabio C, Alewell C, Meusburger K, Modugno S, Schutt B, Ferro V, Bagarello V, Van Oost K, Montanarella L, Panagos P (2017) An assessment of the global impact of 21st century land use change on soil erosion. Nat Comm 8(1):2013. https:// doi.org/10.1038/s41467-017-02142-7

Breiman L (2001) Random forests. Mach Learn 45(1):5-32

- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Trees. Wadsworth
- Brown CF, Brumby SP, Guzder-Williams B, Birch T, Hyde SB, Mazzariello J, Czerwinski W, Pasquarella VJ, Haertel R, Ilyushchenko S, Schwehr K, Weisse M, Stolle F, Hanson C, Guinan O, Moore R, Tait AM (2022) Dynamic World, Near real-time global 10 m land use land cover mapping. Sci Dat 9(1):251. https://doi.org/10.1038/s41597-022-01307-4
- BUCREP (2010) Recensements de la population camerounaise. Presses du BUCREP, Yaoundé
- Carone T, Gioia M, Ferretti E, Marincioni MF (2015) Linking Vegetation Patterns and Landslide Occurrence An Empirical Method. Engineering Geology for Society and Territory. Springer, Cham, pp 1195–1198
- Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H, Wang X, Bian H, Zhang S, Pradhan B (2020) Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Sci Total Environ 701:134979. https://doi.org/10.1016/j.scitotenv.2019. 134979
- Colombo A, Lanteri L, Ramasco M, Troisi C (2005) Systematic GIS-based landslide inventory as the first step for effective landslide-hazard management. Landsl 2(4):291–301. https://doi.org/10.1007/s10346-005-0025-9
- Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Marquez JRG, Gruber B, Lafourcade B, Leitao PJ et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecogr 36(1):27–46
- Erdas Imagine (2020). https://supportsi.hexagon.com/s/article/ERDAS-IMAGI NE-2020-Release-Announcement-v16-6-0-1347?language=en_US. 16.6.0.1366

- El Jazouli A, Barakat A, Khellouk R (2019) GIS-multicriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco). Geoenviron Disast 6(3):1–12
- Elnashar A, Zeng H, Wu B, Fenta AA, Nabil M, Duerler R (2021) Soil erosion assessment in the Blue Nile Basin driven by a novel RUSLE-GEE framework. Sci Total Environ 793:148466. https://doi.org/10.1016/j.scitotenv. 2021.148466
- Elsheikh AA, Gao SS, Liu KH (2014) Formation of the Cameroon volcanic line by lithospheric basal erosion: insight from mantle seismic anisotropy. J Afr Earth Sci 100:96–108. https://doi.org/10.1016/j.jafrearsci.2014.06. 011
- Fayne JV, Ahamed A, Roberts-Pierel J, Rumsey AC, Kirschbaum DB (2019) Automated satellite-based landslide identification product for Nepal. Earth Interact 23(3):1–21. https://doi.org/10.1175/EI-D-17-0022.1
- Felsberg A, Poesen J, Bechtold M, Vanmaercke M, De Lannoy GJM (2022) Estimating global landslide susceptibility and its uncertainty through ensemble modeling. Nat Haz Earth Syst Sci 22:3063–3082. https://doi. org/10.5194/nhess-22-3063-2022
- Frattini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. Eng Geol 111(1–4):62–72. https://doi.org/10.1016/j.enggeo.2009.12.004
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. The Ann Stat 29(5):1189–1232
- Friedman JH (2002) Stochastic gradient boosting. Comp Stat Data an 28:367–378. https://doi.org/10.1016/S0167-9473(01)00065-2
- Garosi Y, Sheklabadi M, Conoscenti C, Pourghasemi HR, Van Oost K (2019) Assessing the performance of GIS- based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Sci Tot Environ 664:1117–1132. https://doi.org/10.1016/j. scitotenv.2019.02.093

GEE (2022) Reducer overview. https://developers.google.com/earthengine/guides/reducers_image_collection. Accessed November-December 2021

- Geze B (1942) Observations sur les sols du Cameroun Occidental. Ann Agro 12(1):103–131
- Gholami V (2013) The influence of deforestation on runoff generation and soil erosion (Case study: Kasilian Watershed). J for Sci 59(7):272–278. https:// doi.org/10.17221/20/2013-JFS

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031

Handwerger AL, Huang M-H, Jones SY, Amatya P, Kerner HR, Kirschbaum DB (2022) Generating landslide density heatmaps for rapid detection using open-access satellite radar data in Google Earth Engine. Nat Haz and Earth Syst Sci 22(3):753–773. https://doi.org/10.5194/nhess-22-753-2022

Hengl T (2018) Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m (Version v02) . 10.5281/zenodo.1475451.

- Hungr O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. Landslides 11(2):167–194. https://doi.org/10.1007/ s10346-013-0436-y
- Hwang CL, Yoon K (1981) Multiple attribute decision making: methods and applications. Springer, New York
- Ionita I, Fullen MA, Zgłobicki W, Poese J (2015) Gully erosion as a natural and human-induced hazard. Nat Hazards 79:1–5. https://doi.org/10.1007/ s11069-015-1935-z
- IRDR (2014) IRDR Peril Classification and Hazard Glossary, Integrated Research on Disaster Reduction. https://www.irdrinternational.org/knowledge_ pool/publications/173.
- Javidan N, Kavian A, Pourghasemi HR, Conoscenti C, Jafarian Z, Rodrigo-Comino J (2021) Evaluation of multi-hazard map produced using MaxEnt machine learning technique. Sci Rep 11:6496. https://doi.org/10.1038/ s41598-021-85862-7
- Jiang H, Wang S, Cao X, Yang C, Zhang Z, Wang X (2019) A shadow- eliminated vegetation index (SEVI) for removal of self and cast shadow effects on vegetation in rugged terrains. Int J Digit Ear 12(9):1013–1029. https://doi.org/10.1080/17538947.2018.1495770
- Kadjio Feudjio AS, Tchana CB, Amougou JA, Batha AR, Shegwe TQ and Yungsi WD (2021) Impacts de la dynamique des paramètres climatiques sur la production des cultures saisonnières (Mais, Haricot et Arachides) dans les

Arrondissements de Nkongsamba (1983 à 2016). Rev Esp Geo Soc Maroc, 50. https://revues.imist.ma/index.php/EGSM/article/download/26800/ 14025.

- Khosravi K, Pham BT, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Prakash I, Bui DT (2018) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Sci Total Environ 627:744–755. https://doi.org/10.1016/j.scitotenv.2018.01.266
- Kirschbaum D, Stanley T (2018) Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. Earth's Fut 6(3):505–523. https://doi.org/10.1002/2017EF000715
- Kron, W (2005) Flood risk = Hazard × Values × Vulnerability. Wat Int, 30(1):58– 68. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.594.999& rep=rep1&type=pdf.
- Meena SR, Ghorbanzadeh O, Blaschke T (2019) A comparative study of statistics-based landslide susceptibility models: a case study of the region affected by the Gorkha earthquake in Nepal. Int J Geo-Inform 8(2):94
- Mihi A, Benarfa N, Arar A (2020) Assessing and mapping water erosion-prone areas in northeastern Algeria using analytic hierarchy process, USLE/ RUSLE equation, GIS, and remote sensing. Appl Geomat 12:179–191. https://doi.org/10.1007/s12518-019-00289-0
- Miles DWR, Swanson FJ (1986) Vegetation composition on recent landslides in the Cascade Mountains of western Oregon. Can J for Res 16:739–744
- MINATD (2007) Première session de la plateforme mondiale pour la réduction des risques de catastrophes. Genève, Suisse, du 5 au 7 Juin. 5p. https://www.preventionweb.net/files/2209_CameroonStatementGP07.pdf
- Mind'je R, Li L, Nsengiyumva JB, Mupenzi C, Nyesheja EM, Kayumba PM, Hakorimana E (2020) Landslide susceptibility and influencing factors analysis in Rwanda. Environ Dev Sustain 22:7985–8012. https://doi.org/ 10.1007/s10668-019-00557-4
- Modugno S, Johnson SCM, Borrelli P, Balzter H (2022) Analysis of human exposure to landslides with a GIS multiscale approach. Nat Haz 112:387–412. https://doi.org/10.1007/s11069-021-05186-7
- Motagh M, Roessner S, Akbari B, Behling R, Stefanova Vassileva M, Haghshenas-Haghighi M, and Ulrich-Wetzel H (2020) Landslides triggered by 2019 extreme rainfall and flood events in Iran: Results from satellite remote sensing and field survey. EGU2020-10715. https://doi.org/10.5194/egusp here-egu2020-10715.
- Ndonbou RM, Nkouathio DG, Zangmo Tefogoum G, Guedjeo CS, Tematio P, Djukem Fenguia SN (2022) Mass movements susceptibility analysis along the Southern Escarpment of the Bamileke Plateaus (Western Cameroon Highlands) using a GIS-based analytical approach. Environ Earth Sci 81:154. https://doi.org/10.1007/s12665-022-10240-z
- Ngandam Mfondoum AH, Etouna J, Hakdaoui M (2014) Contribution of the analytic hierarchy process (AHP) in decision-making to (Re)define protected areas boundaries – case study in the national park of Campo Ma'an (South-Cameroon, Central Africa). Int J Sci Res 3(11):1842–1849
- Ngandam Mfondoum AH, Ngouyamsa Mfondoum RB, Wokwenmendam Nguet P, Gbetkom PG, Ntengo M (2019) Modelling best oil palm site planting in Njimom, West-Cameroon: a gis-analysis combining weighted linear combination, fuzzy analytical hierarchy process and utility function. J Geogr Inf Syst 11(2):138–165
- Ngandam Mfondoum AH, Wokwenmendam Nguet P, Mefire Mfondoum JV, Tchindjang M, Hakdaoui S, Cooper R, Gbetkom PG, Penaye J, Bekoa A, Moudioh C (2021) Adapting sudden landslide identification product (SLIP) and detecting real-time increased precipitation (DRIP) algorithms to map rainfall-triggered landslides in Western Cameroon highlands (Central-Africa). Geoenviron Disast. https://doi.org/10.1186/ s40677-021-00189-9
- Nguimbous-Kouoh JJ and Manguelle-Dicoum E (2010) Contribution of topographic and penetrometric measurements to a site characterization, case of the Kekem Landslide, National Road N°5 (Western Cameroon). Earth Sci Res S J 14(2):135–144
- Nsangou D, Kpoumié A, Mfonka Z, Ngouh AN, Fossi DH, Jourdan C, Zobo Mbele H, Mouncherou OF, Vandervaere J-P, Ndam Ngoupayou JR (2022) Urban flood susceptibility modelling using AHP and GIS approach: case of the Mfoundi watershed at Yaoundé in the South-Cameroon plateau. Sci Afr 15:e01043. https://doi.org/10.1016/j.sciaf.2021.e01043
- O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41(5):673–690. https://doi.org/10.1007/ s11135-006-9018-6

- ORSTOM (1970) Carte pédologique du Cameroun occidental echelle1 :1 000 000.
- Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. Appl Sci. https://doi.org/10.3390/app9050942
- Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019) Landslide susceptibility modeling using reduced error pruning trees and different ensemble techniques: Hybrid machine learning approaches. Cat 175:203–218. https://doi.org/10.1016/j.catena.2018.12.018
- Pourghasemi HR, Gayen A, Panahi M, Rezaie F, Blaschke T (2019) Multi-hazard probability assessment and mapping in Iran. Sci Total Environ 692:556–571. https://doi.org/10.1016/j.scitotenv.2019.07.203

Pourghasemi HR, Sadhasivam N, Kariminejad N, Collins AL (2020a) Gully erosion spatial modelling: Role of machine learning algorithms in selection of the best controlling factors and modelling process. Geosci Front 11(6):2207–2219. https://doi.org/10.1016/j.gsf.2020.03.005

- Pourghasemi HR, Kariminejad N, Amiri M, Edalat M, Zarafshar M, Blaschke T, Cerda A (2020b) Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Sci Rep 10:3203. https://doi.org/10. 1038/s41598-020-60191-3
- Pouyan S, Pourghasemi HR, Bordbar M, Rahmanian S, Clague JJ (2021) A multihazard map-based flooding, gully erosion, forest fires, and earthquakes in Iran. Sci Rep 11:14889. https://doi.org/10.1038/s41598-021-94266-6
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using gis. Comput Geosci 51:350–365. https://doi. org/10.1016/j.cageo.2012.08.023
- Rahmati O, Haghizadeh A, Pourghasemi HR, Noormohamadi F (2016) Gully erosion susceptibility mapping: The role of GIS-based bivariate statistical models and their comparison. Nat Haz 82(2):1231–1258. https://doi.org/ 10.1007/s11069-016-2239-7
- Rehman A, Song J, Haq F, Mahmood S, Ahamad MI, Basharat M, Sajid M, Mehmood MS (2022) Multi-hazard susceptibility assessment using the analytical hierarchy process and frequency ratio techniques in the Northwest Himalayas. Pakistan Remote Sens 14(3):554. https://doi.org/10. 3390/rs14030554

Saaty TL (1980) The Analytic Hierarchy Process. McGraw-Hill, New York Saaty TL (1988) The analytic hierarchy process, 2nd edn. McGraw-Hill, New York

- Sadler J, Goodall J, Morsy M, Spencer K (2018) Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and random forest. J Hydrol 559:43–55. https://doi.org/10.1016/j.jhydrol.2018. 01.044
- Sheikh V, Kornejady A, Ownegh M (2019) Application of the coupled TOPSIS– Mahalanobis distance for multi-hazard-based management of the target districts of the Golestan Province Iran. Nat Haz 96:1335–1365. https://doi. org/10.1007/s11069-019-03617-0
- Skidmore E (1982) Wind erosion control. Clim Change 9:209–218. https://doi. org/10.1007/BF00140537
- Sun L, Ma C, Li Y (2019) Multiple geo-environmental hazards susceptibility assessment: a case study in Luoning County, Henan Province, China. Geomatics Nat Haz Risk 10(1):2009–2029. https://doi.org/10.1080/19475 705.2019.1658648
- Tchindjang M (2013) Mapping of natural hazards in Cameroon. International Cartographic Association. https://icaci.org/files/documents/ICC.../403_proceeding.pdf.
- Tien Bui D, Shirzadi A, Shahabi H, Chapi K, Omidavr E, Pham BT, Talebpour Asl D, Khaledian H, Pradhan B, Panahi M (2019) A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (iran). Sens 19(11):2444. https://doi.org/10.3390/s19112444
- Towfiqul Islam ARM, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Thuy Linh NT (2021) Flood susceptibility modelling using advanced ensemble machine learning models. Geosci Front 12(3):101075. https://doi.org/10.1016/j.gsf.2020.09.006
- Tsewoue MR, Tchamba M, Avana ML, Tanougong AD (2020) Dynamique spatio-temporelle de l'occupation du sol dans le Moungo, Région du Littoral, Cameroun : influence sur l'expansion des systèmes agroforestiers à base de bananiers. Int J Biol Chem Sci 14(2):486–500. https://doi.org/10. 4314/ijbcs.v14i2.15

- UNISDR (2015a) Global Assessment Report on Disaster Risk Reduction. 316p. https://www.preventionweb.net/english/hyogo/gar/2015/en/gar-pdf/ GAR2015_EN.pdf?_gl=1*mj441*_ga*MTMzNjUwNzk0NS4xNjY0ODExO DE0*_ga_D8G5WXP6YM*MTY2NDgxMTgyOC4xLjAuMTY2NDgxMT gyOC4wLjAuMA.
- UNISDR (2015b) Sendai Framework for Disaster Risk Reduction 2015–2030. 32p. UNISDR/GE/2015 - ICLUX EN5000 1st edition.
- UNISDR (2015c) The Human Cost of Natural Disasters: A Global Perspective. 57p. https://reliefweb.int/report/world/human-cost-natural-disasters-2015-global-perspective.
- UN-SPIDER (2020) Step-by-Step: Recommended Practice: Flood Mapping and Damage Assessment Using Sentinel-1 SAR Data in Google Earth Engine. Office for Outer Space Affairs UN-SPIDER Knowledge Portal. https:// www.un-spider.org/advisory-support/recommended-practices/recom mended-practice-google-earth-engine-flood-mapping/step-by-step.
- USGS (2022a) https://www.usgs.gov/media/images/flood-inundation-mappi ng-interactive-floodplain-mapping-tool. Accessed between September-October 2022.
- USGS (2022b) Flood Inundation Mapper. https://fim.wim.usgs.gov/fim/. Accessed between September-October 2022.
- USGS (2022c) USGS Water Alert System. https://maps.waterdata.usgs.gov/ mapper/wateralert/). Accessed between September-October 2022.
- Van Natijne AL, Bogaard TA, Van Leijen FJ, Hanssen RF, Lindenbergh RC (2022) World-wide InSAR sensitivity index for landslide deformation tracking. Int J Appl Earth Ob Geoinf 111:102829. https://doi.org/10.1016/j.jag.2022. 102829
- Varnes DJ (1984) Landslide Hazard zonation-a review of principles and practice. IAEG-UNESCO, Commission on Landslides, 1–6
- Wang Y, Hong H, Chen W, Li S, Panahi M, Khosravi K, Shirzadi A, Shahabi H, Panahi S, Costache R (2019) Flood susceptibility mapping in dingnan county (china) using adaptive neuro-fuzzy inference system with biogeographybased optimization and imperialistic competitive algorithm. J Environ Manag 247:712–729. https://doi.org/10.1016/j.jenvman.2019.06.102
- Wotchoko P, Bardintzeff J-M, Itiga Z, Nkouathio DG, Suh Guedjeo C, Ngnoupeck G, Kagou Dongmo A, Wandji P (2016) Geohazards (floods and landslides) in the Ndop Plain, Cameroon Volcanic Line. Open Geosci 8:429–449. https://doi.org/10.1515/geo-2016-0030
- Wu X, Zhan BF, Zhang K, Deng Q (2016) Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China. Environ Earth Sci 75(2):146. https://doi.org/10.1007/s12665-015-5022-2
- Xiong M, Sun R, Chen L (2018) Effects of soil conservation techniques on water erosion control: a global analysis. Sci Total Environ 645:753–760
- XLStats (2020). https://www.xlstat.com/en/download?gclid=CjwKCAjw4P 60BhBsEiwAKYVkq8swDJIAVEj88gI5bAE4ahWhwTJXqiW0SsGTkWI8anm_ UJBKJRvTdhoCYOAQAvD_8wE. 2020.1.64570
- Zhao L, Hou R (2019) Human causes of soil loss in rural karst environments: a case study of Guizhou China. Sci Rep 9:3225. https://doi.org/10.1038/ s41598-018-35808-3

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com