Understanding Gully Erosion Vulnerability in Old Imo State Using Geographic Information System and Geostatistics

Josh U. Ogbonna

Department of Geography and Planning,Abia State University, Uturu, Nigeria

Abstract   Geostatistical model (GM) and the Geographic Information System (GIS) are used as the Decision Support System (DSS) for the examination of the vulnerability rate of gully erosion in Old Imo State. The gully erosion sites were identified using a combination of Landsat TM satellite imagery, interpretation of aerial photographs and field surveys using the Global Positioning Systems (GPS). Databases on the geological formations, and gully erosion were determined through fieldwork, and digitization. Factors which influence gully occurrence, such as slope, distance from drainage and terrain mapping unit were calculated from the topographic database, and the relationship between these factors determined.

Keywords   Geostatistic Model, Logistic Regression, Gully Erosion, GIS, GPS, Fuzzy, DSS, Buffering

1. Introduction

Gully erosion is a great environmental issue in the recent decades, resulting in land degradation and lowering agricultural productivity and promotes serious land degradation. In the Old Imo State, the impact of gully erosion has been unimaginable, displacing communities and causing destruction of farm lands. However, human activities like land- clearing, and deforestation, overgrazing, as well as the creations of firewood tracks accelerate the natural rates of these processes. The immediate site effect is the loss of soil while the off-site effects include the yield of sediment on the river network, which results in declining water quality and damage to the hydraulic structures[1]. It is therefore necessary to assess and to manage areas that are vulnerable to gully erosion in order to mitigate any damage associated with it. Among the many causes, gullies triggered by man's activities and rainfall are the most common in the Old Imo State. The resultant need to predict gully occurrences has led to the development of numerous stochastic and process-based models, with increasing emphasis on the use of the Geographic Information Systems (GIS), Satellite Remote Sensing (SRS) and Geostatistical models.[2],[3], [4],[5],[6],[7],[8],[9],[1], and[10].

* Corresponding author:
juogbonna@yahoo.com (Josh U. Ogbonna)

In developing a sustainable remediation procedure, an assessment of gully erosion through a spatial mapping of the areas vulnerable to gully formation is necessary. These assessments are carried out by applying the frequency ratio and logistic regression models, with verification of the results. Hence, the objective of this study is to apply and verify the models for gully erosion in Old Imo State using the GIS and geostatistics. The map below (Fig. 1) shows Old Imo State (The Study Area) with spatial distribution of gully. The study area is bounded by latitudes 4°45' and 6°15' North and longitudes 6°30' and 8°09' East.

In order to assess soil vulnerability to gully erosion, it is necessary to identify and analyse the factors leading to the development of gully erosion. The parameters used were slope, aspect, proximity to drainage, lithology, and proximity to major structures, land cover, and geomorphologic units.

Using GIS as the basic analysis tool for gully erosion mapping, it presents an effective platform for spatial and database management and manipulation. Of the statistical methods available, the logistic regression method was applied to mapping impacts of gully erosion [11]. The adaptation of GIS and data mining techniques such as fuzzy logic, and artificial neural network constitutes a new horizon to impact evaluation which is applied in this work and constitutes a new vista to knowledge.
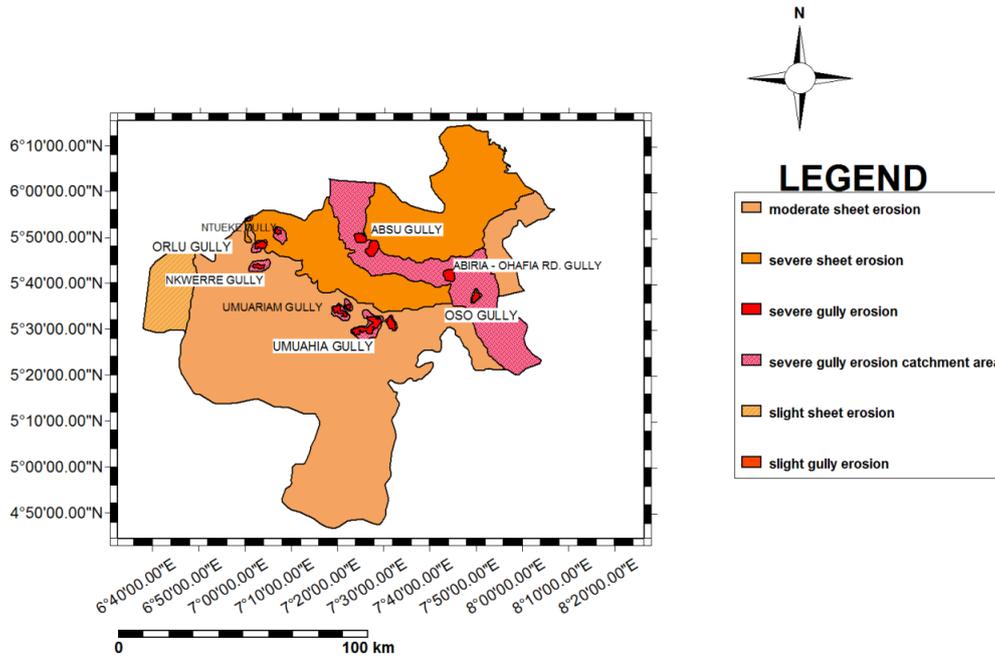
2. Conceptual Framework

Figure 1. Some Gully Erosion Sites in Old Imo State

Logistic regression is a multivariate analytic model useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables [12]. The advantage of logistic regression is that, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or a combination of both types, and they do not necessarily have normal distributions. In the present situation, the dependent variable is a binary variable representing the presence or absence of gully (0,1). Quantitatively, the relationship between the occurrence and its dependency on several variables can be expressed as:

$$p = 1 / (1 + e^{-z}) \qquad (1)$$

Where p is the probability of an event occurring. In the present situation, the value p is the estimated probability of gully occurrence. The probability varies from 0 to 1 on a sigmoidal (S) curve and z is the linear combination. It follows that logistic regression involves fitting an equation of the following form to the data:

$$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \qquad (2)$$

Where: b_0 is the intercept of the model, the b_1 (i = 0, 1, 2… n) are the slope coefficients of the logistic regression model, and the x_1 (i= 0, 1, 2… n) are the independent variables. The linear model formed is then a logistic regression of presence or absence of gullies (present conditions) on the independent variables (pre-failure conditions). The main advantage of logistic regression model is its robustness against violation of normality and homoscedasticity assumptions[9], which is often the case in most spatial analyses.

3. Data Capture and Methodology

Data preparation involved the digitization of GIS database using ILWIS 3.6 through an on-screen digitizing mode. This led to an interactive topographical, geomorphologic, and geological maps and land use/land cover data. A digitized map of the gully erosion sites captured from satellite imagery and ground truthed through field surveys was produced, and these digital data were input to a GIS platform. A vector-to-raster conversion was undertaken to provide raster data of gully erosion areas with 10 m by 10 m pixels. Factor maps related to gully formation were constructed in a vector-type spatial database. These included topographic maps and geological maps. A land cover map was extracted from Landsat TM satellite imagery with 30 m resolution. The factors such as slope, aspect, proximity to drainage, lithology, and proximity to major structures, land cover and geomorphologic units were also used.

Table 1. Sources and type of input data

Data Layers	Type	Data Capture	Attributes
Gully erosion	polygons	digitized	name, area
Streams	vectors	digitized	name, length
Roads	vectors	digitized	name, length
Towns	points, polygons	digitized	name, population
Landsat TM	raster, grey scale	Flash drive	7 spectral bands
Digital topography (DEM)	points, vectors, polygons	disks	x, y, z + planimetric data

Josh U. Ogbonna: Understanding Gully Erosion Vulnerability in Old Imo State Using
Geographic Information System and Geostatistics

68

The study area was divided into a grid with 1km by 1km cells, forming a total of 12,689 grid-cells and gullies occurring in 22 of these exhibiting a Poisson probability density function. Contour (5-meter interval) and survey base points that had an elevation value read from the topographic map were extracted, and a Digital Elevation Model (DEM) was constructed. Using the DEM, the slope gradient, slope aspect and curvature were calculated. The slope gradient of a surface refers to the maximum rate of change in z values across a region of the surface and the slope height of the surface show the height of z above sea level. The curvature represents the morphology of the topography. A positive curvature indicates that the surface is upwardly convex at that cell, and a negative curvature indicates that the surface is upwardly concave at that cell. A value of zero indicates that the surface is flat. A 1km buffering was carried out on the gully erosion sites (see Fig. 2).

Seven (7) geologic formations cover the study area, these are: (1) *Niger Delta Formation* – consist of sands, gravels, clays mainly arenaceous with very low dip structure to the SSW;

(2) *Coastal Plain Sands Formation* – consists principally of sands, gravels, clays, shales (lignites). Topographycally, it consists of gently sloping plain; (3) *BendeAmeke Group Formation* – lithologies are clastic sedimentary rocks consisting mainly of lignites conglomerates with interbedded sandstones, and shales.. (4) Imo Shales Formation - *consists* of clay-shales with intra-formational sandbodies. It is typified with lowland ridges. (5) This formation is generally called The *Plateau and Escarpment* made up of three groups: *Upper Coal Measure, False Bedded Sandstone, and Lower Coal Measures* – lithologies are sands, gravels, siltstones, sandstones and grits. The topography consists of minor cuesta, mesas and escarpment. (6) This formation consists of – *Asata-Nkporo Shale, Awgu-Ndeaboh Shale Group, EzeAku Shale Group, and Asu River Group*. Lithologies are shale with lenticular sandbodies, thin limestones, and minor sandstone lenses with igneous bodies. (7) *Asu River Group Formation* – consist of crystalline rocks, with topography of mountain terrain with planatedmargins[1]. The type of geology and soil of an area plays an important factor in the development of gully erosion. Overlaying the gully erosion map on the soil map of the area show rate of gully erosion growth and the type of soil formation as tabulated values below show:

Table 2. Soil type / Gully erosion vulnerability

Gully Erosion site	Soil Type	Gully Erosion as at 1984 (in km^2)	Gully Erosion as at 2008 (in km^2)	Gully Erosion predicted for 2018 (in km^2)	Rate of growth
Orlu	RhodicFerrasols	6.58	31.01	34.07	1.02
Nkwerre	RhodicFerrasols	9.8	33.03	36.96	0.96
Ntueke	RhodicFerrasols	4.33	48.96	48.96	1.25
Umuariam	DystricFerrasols	12.8	19.79	27.8	0.29
Umuoriehi	DystricFerrasols	34.3	70.81	90.24	2.35
Amuzukwu	DystricFerrasols	12.0	42.74	48.96	1.28
MgbeluUmmunnekwu	RhodicFerrasols	22.69	29.55	38.75	0.29
Owutu	GleyicCambisols	18.3	33.03	42.13	0.76
Oso	GleyicCambisols	Recent development	41.42	47.2	NA
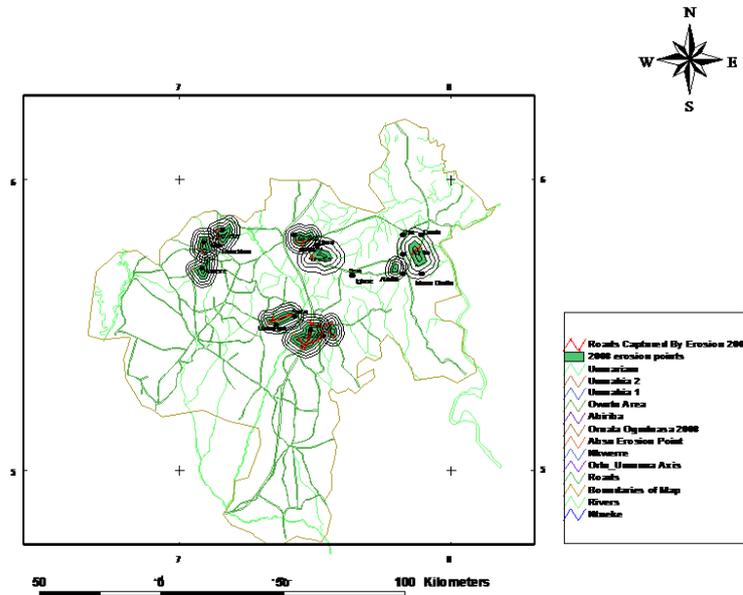
Figure 2. Buffered Gully Erosion sites to Estimate future Impact

Land use and land cover data were classified from Landsat TM satellite imagery. For calculating proximity to major structures, a distance buffer was carried out to see the future rate of devastation of gully erosion. Using the detected gully erosion locations and the constructed spatial database, gully analysis models were applied and verified. To represent the distinction quantitatively, frequency ratio and logistic regression models were used. For this analysis, the calculated and extracted factors were mapped to a 10 m resolution grid. The raster data were converted for the geostatistical analyses.

Then, using the frequency ratio and logistic regression models, the spatial relationships between the gully location and each gully causative factors, such as topography, soil, forest and land cover, were analyzed in the statistical program, and a formula of gully occurrence possibility was extracted using the relationships. The formula was used for calculating the gully erosion susceptibility index, which was mapped to each grid cell. In this study, ArcView 3.2a and ILWIS 3.6 Window versions were used as the basic analysis tools for spatial management and data manipulation.

4. Spatial Relationship between Gully Erosion and the Causative Factors

The relationship between areas where gully erosion has occurred and the causative factors were examined. To represent this distinction, the frequency ratio was used. The factors chosen, such as the slope, distance from drainage, lithology, land use, and vegetation cover were evaluated using the frequency ratio method to determine the level of correlation between the location of the gully erosion site in the study area and these factors. Probabilistic approaches were based on the observed relationships between each factor and the distribution of gullies. In the case of the relationship between gully occurrence and slope, below a slope of 12°, the ratio was <1, which indicates a low probability of gully formation as observed within Umuahia – Obowo axis. For slopes above 12°, the ratio was >1, which indicates a high probability of gully occurrence especially within Uturu – Isuikwuato axis. This means that the gully probability increases according to slope angle. As the slope angle increases, then the shear stress in the soil or other unconsolidated material generally increases. Gentle slopes are expected to have a low frequency of gully because of the generally lower shear stresses associated with low gradients. In the relationship between gully occurrence and soil, gully erosion was most abundant on 3 major soil types; these were *rhodicferrasols, dystricferrasols and gleyicferrasols*. The frequency of gully infestation was lowest on *gleyicferrasols* and highest on *rhodicferrasols.* In the case of the relationship between gully occurrence and curvature, it was observed that Aba – Ukwa axis which is flat areas had a low curvature value of 0.71 shows that it has a low probability of gully occurrence.

In the case of the relationship between gully formation and

lithology, the frequency ratio was higher in *BendeAmekeGroup* formations which consist principally of lignites conglomerates with interbedded sandstones, and shales, and was lower in *Niger Delta Formation* which consist mainly of sands, gravels, clays mainly arenaceous with polymictic conglomerates interbedded sandstones, siltstones, and shale.

For the relationship between gully formation and land cover, gully-infestation values were higher in agricultural lands, bushes and grass areas, and lower in broadleaf area. The reason for this is that gully erosion occurred mainly in none of the forested areas.

5. Applying the Geostatic Models: Frequency Ratio and Logistic Regression

The Frequency Ratio (GR) method is used for calculating gully erosion risk. This method measures the ratio of a landscape having a characteristic, where erosion has been mapped, as compared to where they have not. It can be described as:GR = (area of gully with characteristic / total gully area) /Total area with characteristic / total area.

Hence, ifGR> 1 implies proportionally greater area mapped with gully relative to the landscape as a whole with this characteristic.

If GR = 1 implies that proportions of the landscape with characteristic are equivalent or equal in areas with gullies and areas without.

But ifGR< 1implies that area mapped with gulliesis greater than the whole landscapes with this characteristic.

While the logistic regression models the gully risk defining the relationship between predictor variables such as slope, rainfall, aspects /geology and gullies.

Hence, the frequency ratio, area ratio for gully occurrence and non-occurrence was calculated for each factor's class to the total area. The frequency ratios were calculated by dividing the gully occurrence ratio by the area ratio. The frequency ratios of each factor's type were summed to calculate the Gully Vulnerability Index (GVI), as in (3)

$$GVI = \Sigma GR \qquad (3)$$

Where GR = Frequency ratio of each factor's type or class.

The index was classified into equal areas and grouped into three classes for visual and easy interpretation. The minimum value of 3.35 and maximum value of 12.83, the mean value of 8.17 andthe standard deviation value of 1.51 were determined. Using the logistic regression model, expressed as an intercept beta coefficient plus all the predictor variables exhibits the spatial relationship between gully-occurrence and factors influencing gullies like slope, precipitation, aspects, annual geopotential energy from falling rain, and if gully location is within 85m from the road.

6. Verification of Gully Erosion Vulnerability Maps

In the verification of gully erosion vulnerability two basic assumptions are made:

→that gully formations are related to certain geo-based spatial attributes such as topography, soil, vegetation and land cover, and

→that future gullies impact will be influenced by the factors of rainfall and human interference.

These two assumptions were observed to be dominant in the study area. This is because, gullies were related to the spatial information and are influenced by rainfall and human interference through construction. The gully vulnerability analysis result was validated using known gully locations as in Map 1. Validation was performed by comparing the known gully location data with the gully vulnerability map buffered for 10 years (2018) using generated rate of gully growth. Each gully site with its associated factors was compared. The rate explains how well the model and factor predict the gully erosion. Map 2 shows the result of the buffered gully sites which forms a pathway for the explanation for the gully erosion vulnerability index.

In the case of logistic regression model, 0.9 to 1.0 (10%) class of the study area where the gully vulnerability index had a higher rank could explain 0.39 of all the gully sites. In addition, the 0.8 to 0.1 (20%) class of the study area where the gully vulnerability index had a higher rank could explain 57% of the gully sites. To compare the result in all the factors and logistic regression model used, an area ratio of 0.8 shows a prediction accuracy of 80%. While in the case of all factors and frequency ratio model, an area ratio of 0.7 shows a prediction accuracy of 70%. Factoring in the gully causative factors into derived logistic regression model show an average value of 0.7883 which tend to indicate the effectiveness of logistic regression model. This method is after [20].

Through an overlay, buffering and split operations, this work was able to predict the gully erosion vulnerability for 2018 as shown in Table 2.

7. Conclusions

Gullies are among the most hazardous of natural disasters. Government and research institutions worldwide have attempted for years to assess hazards and risks due to gully erosion and to show their spatial distribution [13],[14],[15],[16],[17],[18],[19], and[1]. In this study, a statistical approach for identifying the vulnerable area of gully sites using GIS shows considerable promise.

The result of validation of logistic regression and frequency ratio model, the logistic regression model showed the better prediction accuracy with more than 1.6%. The frequency ratio model is simple; the process of input, calculation and output can be readily understood. The large amount of data can be processed in the GIS environment quickly and easily. The logistic regression model requires conversion of the data to spread sheet formats for use in the statistical package, and later reconversion to incorporate it into the GIS database.

The integration of GIS with statistical analyses exposes the great utilities of database construction, display, printing, management and spatial analysis. Thus, it is necessary to integrate the GIS andstatistics to reduce the restrictions of using the two separate applications and increasing efficiency in gully vulnerability examination.

REFERENCES

[1]   OGBONNA, J.U. 2009. Macro-analytic study of the impact of gully erosion in Old Imo State of Nigeria. An Unpublished Ph.D. Thesis of the Department of Geography and Planning of Abia State University, Uturu. Nigeria.

[2]   FLORAS, S.A. and SGOURAS, I.D., 1999.Use of geoinformation techniques in identifying and mapping areas of erosion in a hilly landscape of central Greece.*International Journal of Applied Earth Observation and Geoinformation*, 1:68–77.

[3]   KUBAT, M., HOLTE, R., MATWIN, S., 1998. Machine learning for the detection of oil spills in satellite images. *Machine Learning*, 30, 195–215.

[4]   FORBES, A.D., 1995. Classification-algorithm evaluation: five performance measures based on confusion matrices. *Journal of Clinical Monitoring*, 11(3): 189-206.

[5]   ZINCK, J.A., LÓPEZ, J., METTERNICHT, G.I., SHRESTHA, D.P., and VÁZQUEZSELEM, L., 2001. Mapping and modelling mass movements and gullies in mountainous areas using remote sensing and GIS techniques. *International Journal of Applied EarthObservation and Geoinformation*, 3:43—53.

[6]   HABOUDANE, D., BONN, F., ROYER, A., SOMMER, S. and MEHL, W., 2002. Land degradation and erosion risk mapping by fusion of spectrally-based information and digital geomorphometric attributes. *International Journal of Remote Sensing*, 23: 3795–3820.

[7]   GIANNETTI, F., MONTANARELLA, and L., SALANDIN, R., 2001. Integrated use of satellite images, DEMs, soil and substrate data in studying mountainous lands.*International Journal of Applied Earth Observation and Geoinformation*, 3:25–29.

[8]   HANLEY, J.A., McNEIL, B.J., 1983, A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148: 839–843.

[9]   SYMEONAKIS, E. and DRAKE, N., 2004. Monitoring desertification and land degradation over sub-Saharan Africa.*International Journal of Remote Sensing*, 25:573–592.

[10] OGBONNA, J.U. and Ijioma, M.A., 2010. Mapping gully erosion susceptibility in Old Imo State, Nigeria using probability and statistics model.

[11] GARCÍA-MELÉNDEZ, E., MOLINA, I, FERRÉ-JULIÁ, M and AGUIMET, J., 1998.Multisensor data integration and

GIS analysis for natural hazard mapping in a semi-arid area (Southeast Spain).*Advanced Space Research*, 21: 493–499.

[12] HOSMER, D.W. and LEMESHOW, S., 2000, *Applied logistic regression* (Wiley: New York), 373 pp.

[13] KOOPMANS, B.N. and FORERO, R.G., 1993. Airborne SAR and Landsat MSS as complementary information source for geological hazard mapping. *Photogrammetry &Remote Sensing*, 48: 28–37.

[14] TRIPATHY, G.K., GHOSH, T.K., and SHAH, S.D., 1996. Monitoring of desertification process in Karanataka state of India using multi-temporal remote sensing and ancillary information using GIS.*International Journal of Remote Sensing*, 17:2243–2257.

[15] SINGHROY, V., MATTAR, K.E. and GRAY, A.L., 1998, Landslide characterisation inCanada using interferometric SAR and combined SAR and TM images.*Advanced Space Research*, 21:465–476.

[16] KOCH, M., 2000. Geological controls of land degradation as detected by remote sensing: a case study in Los Monegros,

north-east Spain. *International Journal ofRemote Sensing*, 21:457– 473.

[17] MA, J. W., XUE, Y., MA, C.F., WANG, Z.G., 2003. A data fusion approach for soil erosion monitoring in the Upper Yangtze River Basin of China based on Universal Soil Loss Equation (USLE) model. *International Journal of Remote Sensing*, 24: 4777– 4789.

[18] SINGHROY, V., 1995. SAR integrated techniques for geohazard assessment. *Advanced Space Research*, 1:67–78.

[19] SHRIMALIL, S.S, AGGARWALZ, S.P. and SAMRAL, J.S., 2001. Prioritizing erosion-prone areas in hills using remote sensing and GIS - a case study of the Sukhna Lake catchment, Northern India. *International Journal of Applied Earth ObservationandGeoinformation*, 3:54–60.

[20] LEPORE, C., SHANAHAN,P, BRAS, R.L.. Rainfall-induced landslide susceptibility zonation of PuertoRico.Environment al Earth Science, 2011. DOI 10,1007/s12665-011'0976-1.