Physics and Chemistry of the Earth 34 (2009) 741–748

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

journal homepage: www.elsevier.com/locate/pce

Estimating gully erosion contribution to large catchment sediment yield rate in Tanzania

Preksedis Marco Ndomba [a,*], Felix Mtalo [a], Aanund Killingtveit [b]

[a] University of Dar es Salaam, Department of Water Resources Engineering, P.O. Box 35131, Dar es Salaam, Tanzania
[b] University of Science and Technology, Department of Hydraulic and Environmental Engineering, Faculty of Engineering, N-7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 21 January 2009
Received in revised form 16 June 2009
Accepted 30 June 2009
Available online 9 July 2009

Keywords:
Cubist
Data mining tool
Gully erosion
Sediment delivery ratio
Sediment yield

ABSTRACT

The objective of this paper is to report on the issues and proposed approaches in estimating the contribution of gully erosion to sediment yield at large catchment. The case study is the upstream of Pangani River Basin (PRB) located in the North Eastern part of Tanzania. Little has been done by other researchers to study and/or extrapolate gully erosion results from plot or field scale to large catchment. In this study multi-temporal aerial photos at selected sampling sites were used to estimate gully size and morphology changes over time. The laboratory aerial photo interpretation results were groundtruthed. A data mining tool, Cubist, was used to develop predictive gully density stepwise regression models using aerial photos and environment variables. The delivery ratio was applied to estimate the sediment yield rate. The spatial variations of gully density were mapped under Arc View GIS Environment. Gully erosion sediment yield contribution was estimated as a ratio between gully erosion sediment yield and total sediment yield at the catchment outlet. The general observation is that gullies are localized features and not continuous spatially and mostly located on some mountains' foot slopes. The estimated sediment yield rate from gullies erosion is 6800 t/year, which is about 1.6% of the long-term total catchment sediment yield rate. The result is comparable to other study findings in the same catchment. In order to improve the result larger scale aerial photos and high resolution spatial data on soil-textural class and saturated hydraulic conductivity – are recommended.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Gully erosion features can be categorised into two main types, namely classical gullies and ephemeral gullies. Ephemeral gullies will reform in the same location in a field where flow from land slope region concentrates (Chmelova and Sarapatka, 2002). Sometimes ephemeral gullies are called gullying sheet erosion, when the channels are between 10 and 20 cm deep but as much as several metres wide. Classical gullies are an advanced stage of channel erosion. They are formed when channel development has progressed to the point where the gully is too wide and too deep to be tilled across (Chmelova and Sarapatka, 2002). The channels are at least 50 cm deep and, more specifically, they can no longer be eliminated through cropping techniques. These channels carry large amount of water after rains and deposit eroded material at the foot of the gully. Gullies form in various shapes including V-shaped, with an even slope down to the lowest point; U-shaped, with vertical sides; and still others develop through tunnelling and subsidence.

Most of the available study methods (field measurements or modelling) of gully erosion work well at small basins and preferably at plot scales (Ndomba, 2007). They have the disadvantage that their results cannot easily be extrapolated to larger basins. For instance, field measurements as direct methods have been reported by other workers to give un-reliable estimates (Peart and Walling, 1988). Other workers such as Wasson (2002) have noted the transferability problem of plot or micro scale studies results to larger catchment.

Erosion models that have been developed for modelling gully erosion range from simple empirical models such as SedNet (Wilkinson et al., 2004) and physics based models such as SHETRAN (Bathurst, 2002) and Limburg Soil Erosion Model (LISEM) Gullies, European Soil Erosion Model (EUROWISE) (Jetten, 2002). In a physically deterministic way for instance, incision can be modelled by using shear stress of the flow along the sides and bottom of the gully and compare it to the shear strength of the profile (Jetten, 2002). This would require a complete knowledge of the soil cohesion and variation of cohesion with depth, and the capability of simulating the flow conditions in detail. Such kind of information is not readily available in ungauged (i.e. poorly gauged) and at a catchment scale. Therefore, in most cases the only option is to use simple tools and readily available data (Ndomba, 2007).

* Corresponding author. Tel.: +255 754 635 272.
E-mail address: pmndomba2002@yahoo.co.uk (P.M. Ndomba).

1474-7065/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pce.2009.06.009

Empirically, The USDA (1977) determined that at least six factors significantly influenced the rate of headcut advance: tributary watershed area, storm precipitation, soil erodibility, slope of approach channel above the headcut, groundwater level, and changes in runoff due to land use changes above the gully. The tributary watershed area and precipitation were found to best explain the rate of gully headcut advance. It should be noted that regression equations so far developed are meant for individual gullies within specific regions (Ndomba, 2007).

Gully erosion processes are not well understood in literature (Ndomba, 2007). Therefore, in this study gully erosion was modelled by data mining tools. This method usually involves the fitting of very complex "generic" models that are not related to any reasoning or theoretical understanding of underlying causal processes; instead, these techniques can be shown to generate accurate predictions in validation samples (Statsoft, 2006). Regression analysis was conducted between environmental variables (i.e. spatial data on landuse, mean annual rainfall, relief, soil characteristics such as soil texture) and gully density as dependent variable. It should be noted that the role of some of these environmental variables as degradation factors in the spatial and temporal domain of soil degradation intensity have been well documented by a number of workers (Hudson, 1971; and Yanda, 1995). However, Yanda (1995) cautioned that the interplay of these factors is site specific; in some places, the slope may be most important factor, while in others the human factors may be the most important. Rainfall is an important soil erosion factor as it determines vegetation cover and acts as a detaching as well as transporting agent (Yanda, 1995). Relief factors such as slope and aspect influence erosion. For instance, steep slopes render much of the area susceptible to erosion. This is because steep slopes increase the volume and velocity of surface runoff, which causes soil detachment and transportation. On steeper slopes the process can be intense enough to form gullies. Also, bare rock surfaces found at many hill crests permit instantaneous generation of surface runoff (Christiansson, 1972). Long slopes are more affected and particularly those, which drain a large subcatchment area. In Mwisanga catchment in Tanzania the total soil stripping on such steeply sloping parts of the catchment has been facilitated by the original relative shallowness of the soil and weathering profiles in such areas (Yanda, 1995). Landuse factors play an important role in land degradation. This is mainly through vegetation clearing, which exposes the soil surface to water impact.

Data mining is an analytic process designed to explore data usually large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data (Statsoft, 2006). Data mining is also known as knowledge discovery, machine learning or computational learning theory and the ultimate goal of data mining is prediction (Statsoft, 2006). The data mining tool used in this study is Cubist (Rulequest Research, 2004). Cubist is a tool for generating rule based predictive model from data (Rulequest Research, 2004).

Gully erosion in Mount Meru slopes in the Pangani River Basin (PRB) has been reported by Semu et al. (1992) as significant land degradation processes and sources of sediment to PRB. They reported presence of very big gullies up to 10 or more metres wide, with a depth of up to 10 m. Some of them are several 100 m long or more (Semu et al., 1992). However, little was known on the factors that influence the formation of gullies and contribution of gully erosion to the sediment yield in the study area.

Ideally, the data requirement to map and/or simulate gully size and morphological changes is huge. Some of data set used by others (Martinez-Casasnovas et al., 2003; Hughes and Prosser, 2003; and Yanda, 1995) includes: gully depth and width, thickness of soil horizon on the gully walls, documentation of gully morphologies and gullying processes based on photos, multi-temporal of black and white aerial photos taken during the dry season for the selected sites in the catchment with active growing gullies, ground truthing report, landform units, soil geological classification, soil characteristics such as soil thickness, A-horizon texture and B-horizon texture, climate indices such as temperature seasonality, minimum temperature-coldest period, temperature-annual change, mean annual precipitation, lowest period moisture index, moisture index seasonality, landuse map, Relief such as slope and hill slope length derived from Digital Elevation Model (DEM), High resolution DEM (i.e. 1 m by 1 m), annual average ground cover derived from Normalized Differential Vegetation Index (NDVI) based on Advanced Very High Resolution Radiometer (AVHRR) remote sensing. It should be noted that it was not possible to obtain all these data at a catchment scale for our study case. Thus, only 10 environmental variables as presented in Table 1 below were used.

Therefore, the objective of this paper is to report on the issues and proposed approaches in estimating the contribution of gully erosion to sediment yield at large catchment.

2. Materials and methods

2.1. Description of the study area

The study area is located in the upstream of Pangani River Basin (PRB), in the northeastern part of Tanzania (Fig. 1). The catchment outlet is located at Nyumba Ya Mungu reservoir (NYM) dam. The NYM reservoir catchment regulates about 12,000 km^2 area of upper PRB drained by rivers Kikuletwa (1DD1) and Ruvu (1DC1) (Ndomba, 2007). 1DD1 and 1DC1 subcatchments cover 7280 km^2 and 2590 km^2, respectively. The 1DD1 and 1DC1 gauging stations are located about 15 km and 5 km, respectively upstream of the reservoir (Ndomba, 2007). Gauging station at 1D8C near NYM dam is a most downstream boundary (Fig. 1). The NYM reservoir catchment comprises of complex geological formations such as North Pare Mountains, Mount Kilimanjaro, Mount Meru, Lakes

Table 1
Statistics of environmental variables data (i.e. 12179 data points) for Nyumba Ya Mungu (NYM) reservoir catchment as extracted from digital spatial data.

Statistic	Environmental variables									
	Elevation (masl)	Aspect (°)	SOL_WHC (mm)	SOL_texture (–)	SOL_k (mm/h)	SOL_Z (mm)	Slope (°)	Sll (m)	MAR (mm)	CN2 (–)
Minimum	737	0	10	2	0.1	10	0	61	372	57
Maximum	5825	360	154	3	0.1	185	26.1	122	2110	98
Mean	1243	162	86.2	2	0.1	96	2.3	95	699	77
Standard deviation	481	88	26.2	0	0	38	3.2	25	333	5
Coefficient of variation Cv (%)	38.7	54.3	30.4	0.0	0.0	39.6	139.1	26	47.6	6.5

Note: Elevation = ground elevation of the upland catchment, aspect = steepest down-slope direction with reference to north, SOL_WHC = soil water holding capacity, SOL_texture = textural class of the soil, SOL_k = saturated hydraulic conductivity of the soil, SOL_Z = soil depth, slope = upland ground slope, Sll = slope length, MAR = long-term mean annual rainfall, and CN2 = curve number for moderate moisture condition.

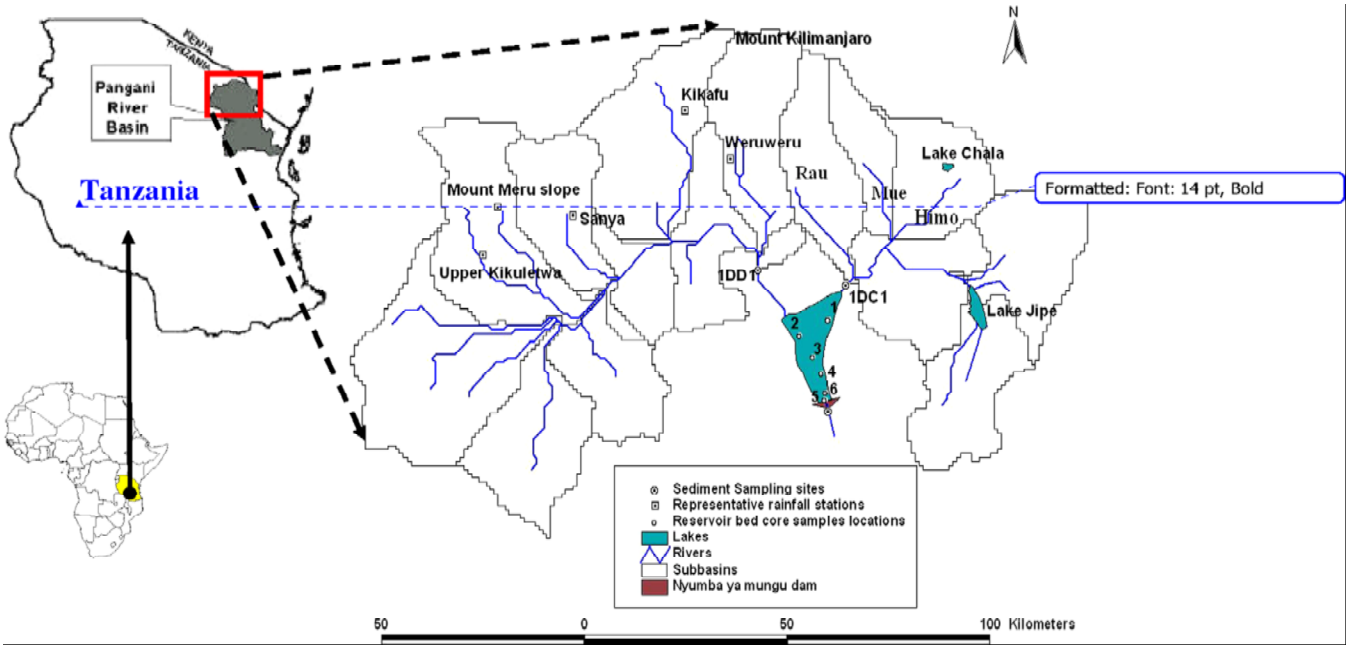


Fig. 1. Location map of Nyumba Ya Mungu (NYM) Reservoir catchment in the Pangani River Basin, Tanzania.

Challa and Jipe, etc. The geology of Kilimanjaro region is consisted of Neogene volcanic and pre-Cambrian metamorphic rocks, which are extensively covered by superficial Neogene deposits including calcareous tuffaceous materials, derived from the Kilimanjaro volcanic and the deposits around Lake Jipe (Geological Survey, 1960).

2.2. Data and data analysis

2.2.1. Data types and sources

The environmental variables used for this study are presented in Table 1 above. Most of the spatial data was sourced from Internet public domain website of United States Geological Survey (USGS, 2006). Mean Annual Rainfall data was derived from hardcopy isohyetal map of the basin as presented in the URT (1977). The digitized isohyets were converted to point features using ArcView software 3.2 extension called Vector Converter tool. The mean annual rainfall in points format (x, y, z) was rasterized using Kriging algorithm of gridding under Surfer software version 8.0 (Golden software, 2002). A FORTRAN code was developed by the authors to extract and analyze the spatial data.

Aerial photos used for this study were sourced from Ministry of Land, Surveys and Mapping Division of Republic Government of Tanzania. A time series of aerial photos used for determining gully size changes are R1: 6914, 6915, 6916, 6917, 6918 and 6919 with scale of 1:30000 aircraft run of 1987 and 60TN 5: 070, 071, 072, 073 and 074 with scale of 1:25000, a scene of 1962. These photos were taken from the dry month of January in the foot slopes of mount Meru, at Oldonyo Sambu village in Arumeru District.

Gully depth and width were measured on at least three gully cross-section profiles, on the upper, middle and lower part of the surveyed gullies at the Oldonyo Sambu village. A mean depth value from several measurements spanning at least 1 m was determined for each cross-section. Afterwards, the average of the mean values of gully depths for surveyed cross-sections was determined. Ground photos took some additional information such as dominant gully morphology and gullying processes.

2.2.2. Analysis of environmental variables

One would note from Table 1 above that some of this data such as elevation and curve number for moderate moisture condition (CN2) are not directly related to data used by other workers. As it was mentioned earlier that only readily available spatial data was used. CN2 which represents landuse type and soil characteristics in the particular soil unit is more informative than the landuse or land cover alone (Ndomba, 2007). One would also note from Table 1 above that there is high spatial variation for most of environmental variables with coefficient of variations of up to 139%, slope for instance. Exceptionally, two variables such as soil texture (SOL_texture) and soil saturated hydraulic conductivity (SOL_k) are spatially constant. There are only two soil-textural classes that are 2 and 3 with the former class dominating as demonstrated by the mean value of 2 in Table 1 above. According to (FAO, 2003) definition this soil is characterized as medium grained soils that may contain sandy loams, loams, sandy clay loams, silt loams, silt, silty clay loams and clay loams with less than 35% clay and less than 65% sand; the sand fraction may be as high as 82% if a minimum of 18% clay is present. Such a classification seems to compare with mapping in the Soil Atlas of Tanzania and Particle Size Distribution (PSD) of the transported suspended sediments in the fluvial system (Ndomba, 2007). Elevation was included in order to capture the high spatial variability of relief factors in the catchment (Table 1). With the exception of two parameters (SOL_texture and SOL_k) that are spatially constant, the table also characterises the data set (cases) that was used to predict gully densities across the catchment.

Based on morphological characterization analysis of the catchment using DEM Ndomba (2007) indicated that NYM reservoir catchment is complex. Most of the river tributaries main channels draining to NYM reservoir are more than 25 km long measured from the reservoir inlet. And two of them, Mount Meru slopes and Upper Kikuletwa river channels extend beyond 120 km. Rivers with their headwaters in the Mount Kilimanjaro slopes (Weruweru, Kikafu, Mue and Rau) are shorter compared to the rest. Besides, the slopes of the latter group of rivers are relatively steeper. About 73% of the basin area falls below elevation of 1200 masl. This portion of the catchment is relatively flat with ground slope of less than 3%. That is, few metres increase in the ground height results into larger gain in area coverage. Also about less than 20% of the area may be characterized as mild or steep slopes (i.e. ground slope of greater than 5%). Only freely downloadable Digital Elevation Models (DEM) of 1 km × 1 km resolution have been used to derive

most of these morphological factors, but as reported in Ndomba (2007), the findings are comparable to available literatures in the basin and field observations. For instance, slopes greater than 40% and that of less than 1% are typically found in the mountain slopes of Kilimanjaro and Ruvu River flood plains, respectively. The discussions in this section suggest that there is high chance for sediment eroded from gullies to be deposited before reaching the catchment outlet as alluvium deposits.

2.2.3. Aerial photos interpretation

Desk mapping of aerial photographs of Oldonyo Sambu village using stereoscope, parallax bar and crack width microscope was done to measure the gully size and estimate the average morphological changes between 1962 and 1987 period. The crack width microscope is a tool to measure fiber crack width in concrete works and other materials; therefore this study borrowed the technology for purpose of gully width measurements only. The accuracy of the tool was validated before its use at the building material laboratory of the Faculty of Civil Engineering and the Built Environment, University of Dar es Salaam, Tanzania. It was possible to identify gullies on the aerial photos because severely eroded areas are lighter owing to the high reflectance associated with the presence of the exposed quartz gravels on the land surface. Besides, gullied areas are easy to distinguish on the aerial photos as they are characterized by dentritic and parallel features, which are indicated by a coarser texture and lighter tone than those of the ungullied part. These features emerge into the main drainage line, which often takes the form of "sand". The less eroded part of the catchment has a low reflectance and thus appears dark in tone. In contrast, foot slope colluvium has high reflectance due to the high sand content of the topsoil and the low organic matter content. Moreover, these areas are cultivated and the boundaries between the fields show higher reflectance as they are less covered by vegetation (Yanda, 1995; Ndomba, 2007).

2.2.4. Analysis of field-based gully geometric data

The deskwork measurements that are depths and widths were groundtruthed from fieldwork conducted on May 2006. Observation based on field survey revealed that the depth for most of the gullies goes up to 5 m in some places and in general the average measured depth and width of gullies is about 2.3 and 17.9 m, respectively. Gully width widening by undercutting and wall collapse are dominant gully erosion processes (Ndomba, 2007). The "location attribute" field in Table 2 represents the surveyed cross-section profile number. Therefore, one would note that a few locations such as (15*, 16* and 17*) as measured from fieldwork do not correspond with locations of laboratory measurements (aerial photo interpretation). The author would like to reveal the locations measured in the laboratory could not be reached because of accessibility problems; therefore, a randomly selected gully in the vicinity was measured. Also, the figures presented in Table 2 are averaged values and rounded to nearest digits. It should be noted that the gully sizes as measured from aerial photo interpretation were corrected (Table 2). The correction factor, for instance was computed as the ratio of average groundtruthed depth (2.3 m) of year 2006 to average laboratory measured depth (2.45 m) of 1987s aerial photo. The same applies to width correction factor.

2.2.5. Comparison between aerial photo laboratory interpretation and field-based gully geometric data results

The measurements of gully size changes based on aerial photo interpretation and fieldwork (Groundtruthing) are summarized and presented in Table 2 above. The deskwork measurements overestimated the fieldwork measurements by +3% and +5% for width and depth, respectively (Table 2). The authors attribute the errors to the coarseness of the scale of the aerial photos used. Therefore, the deskwork measurements of gully sizes for the two time sets of aerial photos were corrected accordingly (Table 2). It should be noted that the need of the correction factor has not been reported by other workers (Hughes and Prosser, 2003; Yanda, 1995; Manyatsi and Ntshangase, 2008). The net depth and width changes were computed as the difference between 1987 and 1962 averaged corresponding corrected measurements. For instance, in Table 2 the overall net depth of 1.09 m was derived as difference between 1987 (2.33 m) and 1962 (1.24 m) averaged corrected depths. Similarly, an overall net width change of 6.0 m was computed. This

Table 2
Measurements of Gully size changes (corrected measurements) based on Aerial photo interpretation and fieldwork (Groundtruthing).

Laboratory measurements-aerial photo interpretation					Field measurements (groundtruth)			Correction factors		Corrected measurements after groundtruthing			
Location	1962		1987		2006					1962		1987	
	Depth (m)	Width (m)	Depth (m)	Width (m)	Location	Depth (m)	Width (m)	Depth (–)	Width (–)	Depth (m)	Width (m)	Depth (m)	Width (m)
1	1.11	12.5	3.18	15	1	2	14			1.05	12.2	3.02	15
2	1.67	12.5	3.18	15	2	1.5	17			1.58	12.2	3.02	15
3	1.67	12.5	2.55	21	3	3	21			1.58	12.2	2.42	20
4	1.11	12.5	1.91	15	4	2.7	14			1.05	12.2	1.81	15
5	1.11	7.5	2.55	15	5	5.5	14			1.05	7.3	2.42	15
6	1.67	12.5	2.54	15	6	3.7	18			1.58	12.2	2.41	15
7	0.56	12.5	2.55	15	7	1.8	12			0.53	12.2	2.42	15
8	1.11	12.5	1.91	15	8	2.4	16			1.05	12.2	1.81	15
9	1.67	12.5	2.54	15	9	2.8	17			1.58	12.2	2.41	15
10	1.11	12.5	1.91	15	10	1.5	12			1.05	12.2	1.81	15
11	1.11	12.5	1.91	21	11	1.3	19			1.05	12.2	1.81	20
12	1.67	12.5	2.54	15	15*	1.2	25			1.58	12.2	2.41	15
13	1.11	25	1.91	51	16*	1.5	27			1.05	24.3	1.81	50
14	1.67	7.5	3.18	15	17*	1.7	25			1.58	7.3	3.02	15
Mean	1.31	12.7	2.45	18.4		2.3	17.9	0.95	0.97	1.24	12.3	2.33	17.9
Min	0.56	7.5	1.91	15.0		1.2	12.0			0.53	7.3	1.81	14.6
Max	1.67	25.0	3.18	51.0		5.5	27.0			1.58	24.3	3.02	49.6
STD	0.35	4.0	0.49	9.6		1.2	4.9			0.33	3.9	0.46	9.4
Cv (%)	26.93	31.4	19.93	52.2		50.4	27.5			26.93	31.4	19.93	52.2
SEM	0.09	1.1	0.13	2.6		0.3	1.3			0.09	1.0	0.12	2.5

Note: STD is standard deviation, Cv is coefficient of variation in percent, and SEM is the estimated standard error of the mean.

study used a number of aerial photos spatially distributed across the catchment as captured during the 1980s across the catchment in building up the rule based linear regression prediction models (Ndomba, 2007). Some of the photos were used for qualitative verification of predictive performance of the developed models.

Aerial photos sampled are located beyond the NYM reservoir water divide. Besides, some portions of the catchment are not represented such as Lake Jipe subcatchment. This is due to two main reasons. Firstly; the priority regions were those areas where growing gullies were reported by Semu et al. (1992) and are evident from available photos in the database and secondly; the photos were sampled in order to represent the high spatial variability of environmental variables as noted earlier. Besides, aerial photo interpretation exercises revealed that gullies are localized features in the catchment. Even, within a locality, the presence of gullies is not spatially continuous (Ndomba, 2007). The latter characteristics suggested that environmental variables that influence gully erosion might be varying in the same pattern/fashion and therefore landform unit was schematized as a grid of 1 km × 1 km. Other workers such as Bathurst (2002) proposed subgrid resolution of 100 m in SHETRAN model as typical gully scales. The authors would like also to note that the resolution adopted here compares to the one for rest of spatial data used in this study. The gully density was computed as the ratio of total length of gullies within one grid to the area of the grid or pixel in units of (km/km^2). Since the pixels are of the size of 1 km^2, basically the values are simply the total gully lengths. Other workers have successfully applied the same approach in their study case in Australia (Hughes and Prosser, 2003).

2.2.6. Determining a set of environmental variables and sample size for modelling

In Table 3 below in the last column, statistics of the data extracted from grids as explained above are presented. In order to build models under Cubist environment the gully density values must be greater or equal to unit (Rulequest Research, 2004). Therefore, the statistics are based on processed data whereby gully densities retain values greater than a unit. The parameters in the table are paired environmental variables that have been sampled from a finite population as presented in Table 1 above. This sample seems to be representative of the population of paired data points between environmental variables and gully densities based on estimated standard error of the mean (SEM). This study used only eight environmental variables for modelling since two variables: soil texture (SOL_texture) and soil saturated hydraulic conductivity (SOL_k) are spatially constant. The authors would like also to note that the sample size is adequate for eight independent environment variables as enlisted in Table 3. Most authors recommend at least 10–20 times as many observations (cases, respondents) as variables, for stable estimates of the regression line and replicability of the results (Statsoft, 2006). A high variation of elevation variable is observed, and this is explained by the fact that aerial photos are not available for all elevation bands in the study area.

2.3. Modelling approach

In this paper gully erosion modelling activity mainly entails estimation of contribution of gully erosion in the catchment sediment yield. Factors such as gully depth and width, bulky density of the soil, and erosion age (i.e. period between successive observations) and gully length per unit area (i.e. gully density) quantified gully erosion rates (Eq. (1)).

$$G_r = \frac{1000\rho_b\rho_G D_n W_n A_c}{T_a} \quad (1)$$

where G_r is the long-term gully erosion rate (t/year), D_n and W_n represent net depth and width change of gully feature (m), ρ_b and ρ_G are bulk density of the soil (t/m^3) and gully density (km/km^2), A_c is the catchment area (km^2), T_a is the time period spanning between multi-temporal aerial photos (year), and 1000 is a constant for unit conversion. Some of these parameters have already been estimated in previous section. For instance, the overall net change in depth and width across the catchment are 1.09 m and 6.0 m, respectively. Various soil bulk densities values that range between 1.3 and 1.5 have been assumed by other workers (Yanda, 1995; Wilkinson et al., 2004). A value of 1.5 t/m^3 was adopted in this study because it compares well with some typical values of measured bulk densities of the downstream reservoir bed material (Ndomba, 2007). It was also checked against the lowest possible value of 1.17 t/m^3 of the deposited sediments in the reservoir. A time period between multi-temporal aerial photos of 25 years was computed for this study. It should be noted that these values have been assumed constant for entire catchment. Estimation of gully erosion rate was lumped for all subcatchment on which gully features were predicted. Most of the modelling efforts were directed to estimating the gully density as it the most sensitive parameter in the Eq. (1) above (Ndomba, 2007). Other workers such as Hughes and Prosser (2003) adopted the same approach.

The prediction of gully density entailed building up statistical rule based models under Cubist tool environment. Stepwise rule based multiple regression models of the general form as presented in Eq. (2) were built by regressing all the environmental variables (X_i) as presented in Table 3 above with gully density as a dependent variable (Y).

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_pX_p \quad (2)$$

where the regression coefficients (or b_i coefficients) represent the independent contributions of each independent variable to the prediction of the dependent variable. And p denotes number of environmental variables that satisfies each condition or rule and a is regression constant. All eight predictive variables were used to construct the stepwise multi-regression models. The data was split into two samples of 70% and 30% for training and verification, respectively, as recommended by Statsoft (2006). The correlation coefficients and relative errors were used as criteria to quantify the performance of the developed rule based models.

Table 3
Statistics of data set (i.e. 200 data points) used for building up rule based multi-regression models.

Statistic	Environmental variables								Gully density (Gully) (km/km^2)
	Elevation (masl)	Aspect (°)	SOL_WHC (mm)	SOL_Z (mm)	Slope (°)	Sll (m)	MAR (mm)	CN2 (–)	
Minimum	809	0.0	67	70	0	61	492	63	1.0
Maximum	5825	359.0	154	185	28	122	2049	98	6.3
Mean	1885	216.0	108	99	5	70	891	77	1.5
Standard deviation	1099	107.1	39	32	5	16	260	7	1.1
Coefficient of variation Cv (%)	58	49.6	36	33	115	22	29	10	74.8
SEM	78	8	3	2	0	1	18	1	0.08

Gully density prediction to areas with no gully measurements was carried out using developed rule based models coded in FORmula TRANslation (FORTRAN) programming language. The algorithm developed in FORTRAN for new cases prediction had the three important general features as suggested by (Rulequest Research, 2004). Firstly; the rule or rules that cover the case were identified. Secondly: the values of the corresponding linear models for the case were calculated; and lastly, the values were averaged. The spatial mapping of gully density intensities and gully prone areas has been achieved using ArcView GIS version 3.2.

Since this study aimed at estimating the contribution of gully erosion to downstream catchment sediment yield, therefore, the estimated erosion rates were factored out using sediment delivery ratio as presented in Eq. (3) below.

$$S_G = \frac{\mathrm{SDR} * G_r}{100} \tag{3}$$

where S_G is the estimated long-term sediment yield rate from gully erosion (t/year), SDR is the Sediment Delivery Ratio (%), G_r is the estimated long-term gully erosion rate (t/year) and 100 is a constant that converts SDR to decimal number.

The uncertainty involved in linking on-site rates of erosion and soil loss within a drainage basin to the sediment yield at the basin outlet using sediment delivery ratio has been reported by Walling (1983). Nevertheless, various SDR equations have been suggested in the literature but a typical value of 50% for the PRB suggested by Mtalo and Ndomba (2002) was adopted. However, based on morphological characteristics some subcatchments such as Lake Jipe catchment were considered by this study that do not contribute sediment to the main outlet of the catchment (Ndomba, 2007). This approach of applying various delivery ratios is well supported in literature. For instance, as reported in Walling (1983) it is suggested that each sediment source should be viewed as possessing a unique delivery potential and that the probability of sediment being exported from a particular source should be related to its relative position with respect to the stream and the basin divide. A similar approach as a special case is also adopted in SedNet model (Wilkinson et al., 2004), where sediments delivered by gullies as bedload in lakes and reservoirs are all deposited (i.e. zero downstream yield). It should be noted that in SedNet model delivery ratio is not applied to cases other than lake/reservoir linkages. However, recent studies indicate that not all eroded sediments leave the foot of the gully, for instance, Martinez-Casasnovas et al. (2003) have reported in their study case in Spain a sediment delivery ratio of 68.1% only in a catchment of 0.688 km^2 in size. As Bathurst (2002) acknowledges that this is still a poorly researched area, such that in the SHETRAN model the proportion of the eroded material transferred to the channel is determined arbitrarily according to the distance of the gully from the channel. Therefore, the discussions above justify the method adopted in applying the delivery ratio and probably to the values used for this study. This study finally compared the sediment yield rate from gully erosion to the actual long-term sediment yield rate of the entire basin as reported by Ndomba (2007).

3. Results and discussions

3.1. Developed rule based models for predicting gully density

Table 4 below presents seven rule based stepwise multi-regression models as developed in this study. As shown in the table each rule has three parts: some descriptive information, conditions that must be satisfied before the rule can be used, and a linear model (Rulequest Research, 2004). The output in Table 4 is organized in a such a way that the descriptive information notes the number of training cases covered by the rule, their mean and range of the dependent variable (target attribute) that is gully density or "gully", and a rough estimate of the error to be expected when this rule is used on new data. The conditions or rules constrain the values of some of the attributes by thresholding numeric attributes or restricting the values of nominal attributes. The conjunction of conditions establishes a context for the linear model (Rulequest Research, 2004). Lastly, a linear model, which is a simplified fit of the training data covered by the rule, is presented. The terms of the model are ordered with the most important attributes appearing first (Rulequest Research, 2004). Looking at Rule 4, for instance, we see that elevation has the most effect (negative) on gully density values for cases covered by this rule; Mean Annual Rainfall (MAR) is the next most important (positive effect). The rules are ordered by the average value of the target attribute (gully) for the cases that they cover. Rule 1 is therefore relevant mainly to low gully densities, Rule 7 to higher gully densities. The performance evaluation of the developed model is presented in Table 5. The model performs better in training than test cases, but the authors consider prediction accuracy based on relative error criterion is acceptable as compared to other workers findings on similar studies (Hughes and Prosser, 2003). Nevertheless, it should be noted that uncertainty in this study has been dealt qualitatively in a number of ways. The performance of the prediction was validated using aerial photos other than used in calibration and testing phases of the developed models across the catchment (Ndomba, 2007), Besides, the mapping result of gully density was compared with the one derived from independent analysis of erosion processes/sources identification (Ndomba et al., 2007).

As noted by Ndomba (2007) even unimportant parameters are used to build the regression models. However, in Table 4 some

Table 4
Predictive relationships of gully density, "Gully", (km/km^2).

Descriptive information					Condition or Rule	Stepwise linear regression equations
Rule no.	Cases	Target mean	Range	Estimated error		
1	73	1.0	1.0–1.0	0.0	Aspect ⩽ 284 SOL_WHC ⩽ 145	Gully = 1.0
2	18	1.0	1.0–1.0	0.0	Aspect > 312 SOL_WHC ⩽ 145	Gully = 1.0
3	9	1.233	1.0–1.716	0.1841	Aspect > 284 Aspect ⩽ 312 SOL_WHC ⩽ 145	Gully = 2.3217–0.055 * Slope −0.0016 * Aspect −0.0032 * SOL_WHC−0.00026 * MAR
4	4	1.2515	1.08–1.51	0.0219	MAR > 788 Elevation ⩽ 1507 SOL_WHC > 145	Gully = 6.8972–0.01469 * Elevation + 0.01978 * MAR
5	14	1.5272	1.0–2.724	0.3414	Elevation > 1507 SOL_WHC > 145	Gully = 6.4401–0.00309 * Elevation
6	13	3.2372	1.351–5.479	0.8044	MAR ⩽ 788 Elevation ⩽ 1507 Slope > 1.4 SOL_WHC > 145	Gully = 43.1022–0.03296 * Elevation + 3.405 * Slope + 0.0098 * Aspect
7	9	3.5956	1.621–5.465	1.2423	MAR ⩽ 788 Elevation ⩽ 1507 Slope ⩽ 1.4 SOL_WHC > 145	Gully = 6.1084–0.00172 * Elevation −0.004 * Slope
Total	140					

Table 5
Performance evaluation of the developed rule based linear models.

Evaluation criterion	Training data (140 cases as in Table 4)	Test data (60 cases)
Relative |error|	0.21	0.63
Correlation coefficient	0.93	0.55

important parameters such as soil depth do not feature in. An independent investigation carried out by the authors indicates that these parameters have high correlation with some of the parameters used. For instance, soil depth is highly correlated with water holding capacity ($r = 0.80$). It is known from literature that when there are very many variables involved, it is often not immediately apparent that this problem exists, and it may only manifest itself after several variables have already been entered into the regression equation (Statsoft, 2006). Nevertheless, when this problem occurs it means that at least one of the predictor variables is (practically) completely redundant with other predictors (Statsoft, 2006). And any attempts to include the redundant parameter in the multiple regression models will lead to highly inaccurate prediction. Although, the models are linear equations, it is difficult to draw a sweeping statement about the trend of variation of gully density from Table 4 above. Unlike traditional statistical data analysis, which is usually concerned with the estimation of population parameters by statistical inference, the emphasis in data mining (and machine learning) approach as adopted in this study is usually on the accuracy of prediction (predicted classification), regardless of whether or not the "models" or techniques that are used to generate the prediction is interpretable or open to simple explanation (Statsoft, 2006). Alternatively, the gully densities were mapped as one way of representing the spatial trend of variation of gully features. However, from these discussions one would note that the redundant parameters are indirectly represented.

3.2. Predicted gully densities and mapped gully prone areas

The predicted gully density results are presented in Fig. 2 below. The average gully density for the entire study area is 0.016 km/km^2 or 16 m/km^2. From Fig. 2 one would note that gullies features are so localized, such that no continuous pattern is spatially depicted. The model has predicted that gully erosion is taking place mostly in the mountain foot slopes. However, it should be noted that not all mountain slopes are mapped with gullies, Weruweru and Rau catchments for instance. Lowland areas of some parts of Arusha and Lake Jipe catchment experience gully erosion as well.

3.3. Estimated sediment yield contribution from gully erosion

Based on the mapping in Fig. 2 above, only a few subcatchments experience gully erosion problem, namely Lake Jipe, Mue, Kikafu, Mount Meru slopes and Upper Kikuletwa. However, according to morphological factors as discussed in the data analysis section, only three catchments (Kikafu, Mount Meru slopes and Mue) with estimated total area coverage of 2158.7 km^2 are expected to deliver the eroded sediment to the downstream. Based on the morphological characteristics as analyzed above, not all sediment will reach the catchment outlet. Thus, a delivery ratio was applied. Therefore, the estimated erosion rate and sediment yield rate from gully erosion in the basin are 13,600 t/year and 6800 t/year, respectively. As reported in Ndomba (2007) the actual long-term sediment yield rate of the study area is 419,000 t/yr. Therefore, the estimated sediment yield contribution from gully equals to 1.6%.

It seems that the findings of this study do not compare to observations made by some workers (Martinez-Casasnovas et al., 2003; Manyatsi and Ntshangase, 2008) elsewhere, who have reported that gully erosion have a very significant contribution to total soil loss. However, the author would like to recall the views made by one of the erosion studies legends such as Hudson (1971), that splash and rill erosion are the most important forms of erosion on arable lands, forming a serious threat to the production of food crops. Christiansson (1981) in agreement with Hudson's view reported that gullies and stream banks only provide a fraction of sheet erosion in Msalatu catchment in central part of Tanzania. Glymph (1951) as reported in Christiansson (1981), in a study of 30 reservoirs and their watersheds in the loess hills region of the

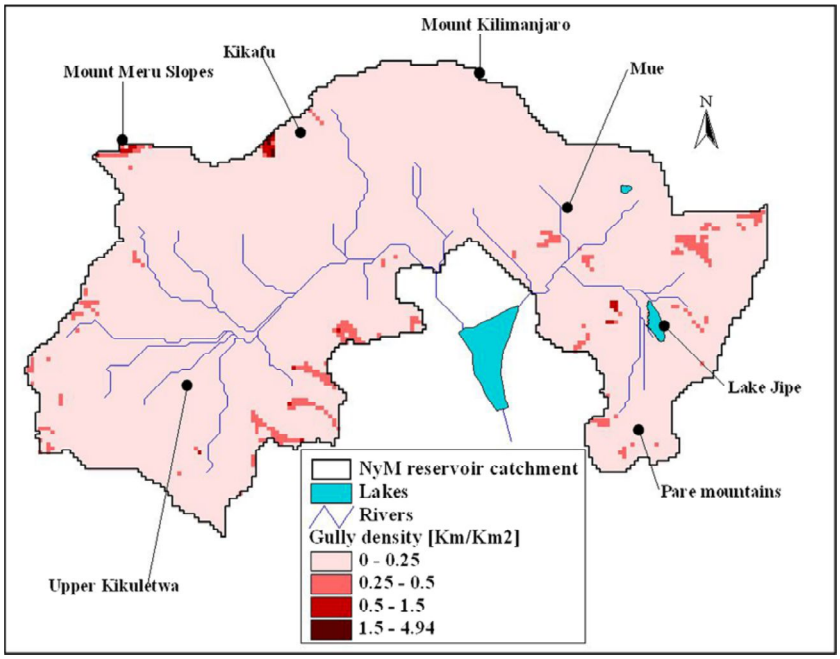


Fig. 2. Map of predicted gully density in the upstream of Nyumba Ya Mungu (NYM) reservoir.

Missouri River Basin, found that sheet erosion was by far the largest single source of sediment. Sheet erosion was the only source of sediment in nine of the reservoir watersheds (Glymph, 1951). It accounted for 95–98% of the total sediment in nine other watersheds (Glymph, 1951). In only five of the watersheds, did sediment derived from gully and channel erosion exceed 10% of the total (Glymph, 1951). It should be noted that this study dealt with incised growing gullies only; therefore other types of gully features such as ephemeral gullies were not modelled. Since ephemeral gullies are temporal features and smaller in size, they could not be monitored by multi-temporal aerial photos spanned over decades with coarser scales as applied in this study. Besides, this result is comparable to the findings of the some studies in the same catchment (Ndomba, 2007; Ndomba et al., 2007). For instance, Ndomba et al. used rating loops and fingerprinting techniques in identifying sediment sources and processes. It was concluded that sheet erosion was the dominant erosion process in the basin.

4. Conclusions and recommendations

The modelling exercise has revealed that gullying prone regions are located on some mountains's foot slopes in the basin. For instance, Mount Meru, Kikafu catchment in the Western region of mount Kilimajaro, Mue catchment in south eastern part of Mount Kilimanjaro, Eastern part of Pare Mountains in the Lake Jipe catchment, and Upper Kikuletwa scattered mountains slopes such as Lolkisale. The general observation is that gullies are localized features and not continuous spatially. The predicted average gully density and estimated sediment yield rate from gullies erosion are 0.016 km/km^2 (i.e. 16 m/km^2) and 6800 t/year, respectively. Therefore, gully erosion contribution to the total sediment yield in the catchment was estimated to be only 1.6%. With such a low contribution, probably other sources such as sheet or interill erosion are dominant erosion processes in the Pangani River Basin. The result is comparable to other study findings in the same catchment. In order to improve the result larger scale aerial photos and high resolution spatial data on soil-textural class and saturated hydraulic conductivity – are recommended.

Acknowledgements

This work was fully funded by The Norwegian Programme for Development, Research and Education (NUFU) – the Postgraduate Programme for Water Management Project and Water Management in Pangani River Basin Tanzania Project at University of Dar es Salaam. We also acknowledge the anonymous reviewers whose comments shaped substantially the presentation of this paper.

References

Bathurst, J.C., 2002. Physically-based erosion and sediment yield modelling: the SHETRAN concept. In: Summer, Wolfgang, Walling, Desmond E. (Eds.), Modelling erosion sediment transport and sediment yield. IHP-VI Technical Documents in Hydrology, No. 60, pp. 47–68.

Chmelova, R., Sarapatka, B., 2002. Soil erosion by water: contemporary research methods and their use. Geographica 37, 23–30. Acta Universitatis Palackianae Olomucensis Facultas Rerum Naturalium 2002.

Christiansson, C., 1972. Notes on morphology and soil erosion in Kondoa and Singida district, central Tanzania. Geografiska Annaler 54A (3–4), 319–324.

Christiansson, C., 1981. Soil Erosion and Sedimentation in Semi-arid Tanzania. Studies of Environmental Change and Ecological Imbalance. Scandinavian Institute of African Studies, Uppsala, Sweden. 208p.

FAO, (2003). The Digital Soil Map of the World. Land and Water Development Division, FAO, Rome. Food and Agriculture Organization of the United Nations Data CD-ROM. Version 3.6, completed January 2003.

Geological Survey, 1960. Summary of the geology of Tanganyika. Part II: geological map. Dar es Salaam, Geological Survey, Ministry of Mines and Commerce.

Glymph, L.M., 1951. Relation of sedimentation to accelerated erosion in the Missouri River Basin. US Soil Conservation Service. Technical Paper No. 102.

Golden Software, 2002. Surfer 8, user's guide: contouring and 3D surface mapping for scientists and engineers. <www.goldensoftware.com> (accessed October 2006).

Hudson, N.W., 1971. Soil Conservation. BT, Batsford Ltd., London, p. 320.

Hughes, A.O., Prosser, I.P., 2003. Gully and River bank Erosion mapping for the Murray Basin. CSIRO Land and Water, Canberra Technical Report 3/03, March 2003.

Jetten, V., 2002. LISEM user manual, Version 2.x. Draft version January 2002. Utrecht Centre for Environment and Landscape Dynamics, Utrecht University, The Netherlands.

Manyatsi, A.M., Ntshangase, N., 2008. Mapping of soil erosion using remotely sensed data in Zombodze South, Swaziland. Journal of Physics and Chemistry of the Earth 33, 800–806.

Martinez-Casasnovas, J.A., Anton-Fernandez, C., Ramos, M.C., 2003. Sediment production in large gullies of the Mediterranean area (NE Spain) from high-resolution Digital Elevation Models (DEM) and geographical information systems analysis. Earth Surface Processes and Landforms, vol. 28, pp. 443–456. doi:10.1002/esp.451. Published online in Wiley Interscience <http://www.interscience.wiley.com> (accessed October 2006).

Mtalo, F.W., Ndomba, P.M., 2002. Estimation of soil erosion in the Pangani Basin upstream of Nyumba ya Mungu reservoir. Water resources management: the case of Pangani Basin. In: Ngana, J.O. (Ed.), Issues and Approaches, Proceedings on Workshop. Dar es Salaam University Press, pp. 196–210 (Chapter 18).

Ndomba, P.M., 2007. Modelling of erosion processes and reservoir sedimentation upstream of Nyumba ya Mungu Reservoir in the Pangani River Basin. A PhD Thesis, Water Resources Engineering, University of Dar es Salaam.

Ndomba, P.M., Mtalo, F., Killingtveit, A., 2007. A proposed approach of sediment sources and erosion processes identification at large catchments. Journal of Urban and Environmental Engineering 1 (2), 79–86. ISSN 1982-3932.

Peart, M.R., Walling, D.E., 1988. Techniques for establishing suspended. In: Bordas, M.P., Walling, D.E. (Eds.), Sediment Budgets Proceedings of the International Symposium. IAHS Press, Wallingford, UK, pp. 269–279 (IAHS Publ. 174).

Rulequest Research, 2004. Rulequest Research Data mining tools with cubist. <http://www.rulequest.com/cr113.html> (accessed January 2006).

Semu, E., Göran, B., Skoglund, E., 1992. Evaluation of Soil Conservation and Agroforestry Programme – Arusha (SCAPA) in Tanzania.

StatSoft, Inc. 2006. Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: <http://www.statsoft.com/textbook/stathome.html> (accessed October 2006).

URT (The United Republic of Tanzania), (1977). Water Master Plan, Kilimanjaro Region, Main Report. Final report vol. I: Water Master Plan Team, Ministry of Water, Energy and Minerals, The United Republic of Tanzania.

USDA (United States Department of Agriculture), (1977). Procedure for determining rates of land damage, land depreciation and volume of sediment produced by Gully erosion. In: Guidelines for Watershed Management, FAO Conservation Guide No. 1, Rome, pp. 125–141.

USGS (United States Geological Survey), (2006). The HYDRO1k data for Africa. Earth Resources Observation and Science (EROS). <http://edc.usgs.gov/products/elevation/gtopo30/hydro/africa.html> (accessed January 2006).

Walling, D.E., 1983. The sediment delivery problem. Journal of Hydrology 65, 209–237.

Wasson, R.J. (2002). What approach to the modelling of catchment scale erosion and sediment transport should be adopted? In: Wolfgang, Summer, Walling, Desmond E. (Eds.), Modelling erosion, sediment transport and sediment yield. IHP-VI Technical Documents in Hydrology, No. 60, pp. 1–12.

Wilkinson, S., Henderson, A., Chen, Y., Sherman, B., 2004. SedNet User Guide. Client Report, CSIRO Land and Water, Canberra.

Yanda, P.Z., 1995. Temporal and Spatial Variations of Soil Degradation in Mwisanga Catchment, Kondoa, Tanzania. PhD Dissertation Series No. 4. Department of Physical Geography, Stockholm University, Stockholm, p. 136.