Fu *et al. Environmental Sciences Europe* (2022) 34:105
https://doi.org/10.1186/s12302-022-00660-w

Environmental Sciences Europe

RESEARCH Open Access

Identification of potential dam sites using OLS regression and fuzzy logic approach

Zhihao Fu[1], Imran Ahmad[2*], Assefa Fenta[2], Mithas Ahmad Dar[3], Afera Halefom Teka[2], Amanuel Zewdu Belew[2], Aserat Teshome[2], Minwagaw Damtie[2], Habtamu Nega[2], Tesfa Gebrie Andualem[2], Marshet Berhan[2], Diress Yigezu Tenagashaw[2] and Eusman Ebrahim Shumey[2]

Abstract

In this paper potential dam sites were identified using remote sensing and GIS. Determinant factors viz., precipitation, slope, flow accumulation, soil texture, land use, and geology were analyzed in the GIS domain. Each factor was reclassified and assigned the suitable fuzzy membership values depending on their influence on the dam site potential. All the fuzzified layers were overlaid using the "Fuzzy Overlay" tool in the GIS platform. Initially, a total of 26 dam sites were proposed. Only seven sites were selected depending on their proximity to nearby dams, settlements, and flow accumulation. The selected dam sites with their flow accumulation, elevation, precipitation, slope, stream order, maximum storage capacity, and the time of concentration were calculated. The determinant factors of suitable dam sites were subjected to the ordinary least squared (OLS) regression to understand the relationship of factors and the potential dam sites. The OLS regression model statistics showed that all factors are positively correlated with potential dam sites except slope (as low slopes are more suitable for dam construction). The OLS regression diagnostics showed that the multiple R squares values and the adjusted R-square values were found to be 0.835894 and 0.872153, respectively. In this study, Koenker's (BP) statistic was found statistically insignificant ($p > 0.01$), proving that the relationship model is consistent. Jarque–Bera statistic was conducted and also found to be statistically insignificant ($p > 0.01$) indicating the Gaussian distribution of residuals. This proves that the fuzzy logic approach coupled with OLS regression is a powerful tool in deciphering the potential dam sites and can be applied at a regional and continental scale.

Keywords: Fuzzy logic, GIS, OLS regression, Potential dam sites, Time of concentration

Introduction

People of the East African countries (Ethiopia, Tanzania, Kenya, and Uganda) are affected much by hunger or under-nourishment. The need to attain food security (by increasing agricultural productivity) is nowhere more pressing than in Ethiopia, which has become a typical case of recurring famines and food insecurity and is a major recipient of foreign food aid (Lire [1]). Agriculture is the backbone of the economy in Ethiopia and it relies largely on rainfall (rain-fed agriculture). Agriculture contributes approximately 75% of export commodity values, 43% of the GDP, and about 80% of employment [2]. The erratic nature of rainfall coupled with rain-fed agriculture is the main reason for widespread food insecurity in Ethiopia. Abduselam [3] has documented that recurrent drought in Ethiopia has a clear link with food security and famine. According to Haile [4], droughts occur every 8–10 years in Ethiopia leading to severe consequences for food production.

The small-scale irrigation development in Ethiopia pre-dates the Axum empire more than 2000 years ago [5]. However, it might have been less important because rain-fed agriculture, coupled with small-scale irrigation was found sufficient to sustain ancient populations [6]. However, with increasing food demand due to

---

*Correspondence: wonder_env@yahoo.com

[2] Department of Hydraulic and Water Resources Engineering, Debre Tabor University, Debre Tabor, Ethiopia
Full list of author information is available at the end of the article

increasing population coupled with increasing pressure on rain-fed agriculture the vital role of irrigation development is inevitable. Timely water supplies coupled with other appropriate agricultural inputs can push up agricultural productivity [7]. The government of Ethiopia has increased its focus on water resource utilization and development to curb food insecurity (FDRE, 2000). Its water policy stresses the increased and efficient use of small-scale irrigation through the building of small dams and the diversion of rivers.

Remote sensing offers valuable and huge data for hydrological studies [8–12]. The geospatial technology is strongly used in the dam site selection, however, its effectiveness in a geographical space may differ [13]. Singh et al. [14] used several remote sensing imageries to identify the potential dam sites. In their study, parameters like soil infiltration rate (moderate), slope (below 10%), soil type (sandy clay loam) and land use (shrubs and river beds) were used in selecting the sites. Thereafter, 14 potential dam sites were found which could practicably be used for water collecting and agriculture. Ibrahim et al. [15] identified potential sites for the possibility of constructing dams, including generating a model builder in GIS domain. This model combined various factors, such as land cover/use, slope, stream order, runoff potential, hydrology and soil quality to ascertain the suitability of the location for rainwater harvesting.

Forzieri et al. [9] assess the suitability of dam sites for water harvesting in arid areas. Geospatial technology has been used to assign the location of water harvesting structures across streams/watersheds (Kumar, 2009). Remote sensing images, digital elevation models, and topographic maps have been used successfully for the proposition of dam site locations [16]. Information related to terrain surface, slope, precipitation, drainage, land use, and watershed boundaries are important for dam site selection, which can be easily obtained from remotely sensed images.

This paper was motivated to contribute to this aspect and help with decision-making support for dam site selection in Farta Woreda, Ethiopia. The paper aims to develop a remote sensing and GIS-based network for a better understanding of various factors influencing the selection of the appropriate dam sites for irrigation purposes in the case of Farta Woreda, Ethiopia. In solving customary overlay analysis applications like models for site suitability and selection, fuzzy logic is an efficient overlay analysis technique. Fuzzy logic provides an effective technique for tackling inaccuracies arising in attributes and the geometry of spatial data. Integrating fuzzy overlay helps decision-makers in making productive decisions regarding this fuzziness. In this paper fuzzy overlay uses fuzzy membership classes to delineate suitable dam sites. Furthermore, OLS regression was done to understand the extent of relationships between the chosen factors/parameters with respect to the potential dam site locations.

Study area

Farta is one of the Woreda (District) in the Amhara region, Ethiopia. It is bordered on the west by Fogera, on the south by Misrak Este, on the East by Lay Gayint, and on the north by Ebenat. Farta Woreda covers an areal extent of 1358.249 km^2 with a perimeter of 171.280 km. Perennial and seasonal channels drain the Farta Woreda. The elevation of the study area varies from 1825 to 3830 m above mean sea level (Fig. 1). In the Farta district, agriculture contributes much to meeting the major objectives of farmers such as food supplies and cash needs. In Farta Woreda, crops are grown for food and cash, and live stocks are kept as security during food shortages to meet the cash needs. In terms of land use, an estimated 67% of the study area is cultivated with perennial and annual crops.

In the case of geology, volcanic rocks of the Recent Era dominate the surface geology of the study area. Debre Tabor basalts and trachyte with an areal extent of 650.61 km^2 dominate the geology of the study area followed by middle basalt flows (283.36km^2), Guna tuff (213.44 km^2), upper basalts, and trachyte (30.7km^2), quaternary lacustrine sediment (24.83 km^2), Guna trachyte (21.77 km^2), upper basalts and pyroclasts (17.63km^2), trachyte plug (3.85km^2), and the plateau basalts and pyroclasts (0.86km^2). The main objective of this paper is to identify suitable sites for dam construction based on topography,

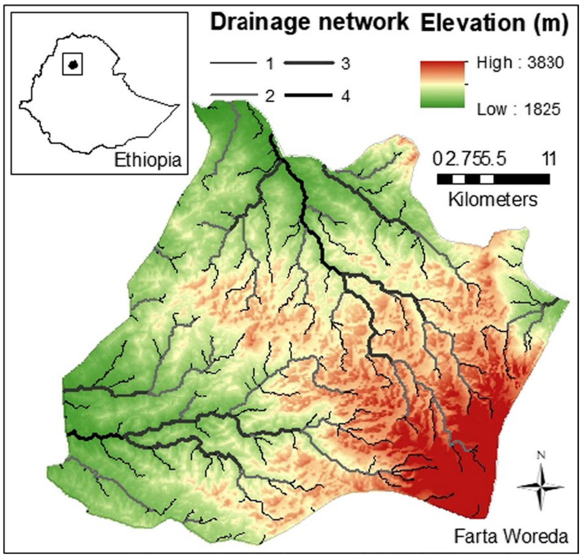


Fig. 1 Location of Farta Woreda (Ethiopia) with elevation

precipitation, geology, etc., based on remote sensing and GIS technology.

Data and methodology

Criteria such as slope and hydrological conditions are indispensable criteria related to dam site selection, safety and dam construction [17]. In this study, criteria considered in identifying potential sites for dams based on the availability of data, literature appraisal and professional verdict. The methodology involves the following five steps:

1. Generation of essential parameters.
2. Assigning the fuzzy membership values.
3. Fuzzy overlay in GIS domain.
4. OLS regression analysis.
5. Spatial autocorrelation analysis.

Precipitation, slope, stream order, soil texture, and land use were analyzed in the GIS domain. The fuzzy logic approach has been used to evaluate the interrelationship of topographic features defining the dam site's prospectus. Suitable fuzzy membership values were assigned to the thematic layer according to their influence on dam site potential. All these thematic maps were assigned fuzzy logic membership values ranging from 0 (unlikely or unsuitable for potential dam site) to 1 (most likely or suitable potential dam site). Using the 'fuzzy overlay' tool, all these fuzzified thematic layers were overlaid to get the appropriate potential zones of dam sites. Fuzzy gamma operators have been used for factor theme integration. The OLS regression analysis has been done to understand and estimate the relationship between the various parameters with respect to the dam sites locations (Fig. 2). The OLS regression modeling was followed by "Spatial autocorrelation" analysis to check if the standardized residuals follow either a random or a Gaussian pattern.

Precipitation

Precipitation data (merged satellite gauge–precipitation estimate) for the year 2020 (Jan–Dec) were downloaded from global precipitation measurements with the spatial and temporal resolution of 0.1^0 and monthly, respectively. The precipitation data (mm/yr.) were projected using Empirical Bayesian Kriging (EBK) in the Adindan_UTM_Zone_37N coordinate system. The prediction errors of EBK semivariograms were checked for their validity. Empirical Bayesian Kriging (EBK) was chosen because it automatically adjusts parameters to receive accurate results through a process of sub-setting and simulations. EBK was also chosen because it accounts for the error introduced by estimating the underlying semivariogram, unlike other kriging methods. The prediction errors for mean standardized, root mean square standardized, and average standard for precipitation EBK are 0.01, 0.98, and 8.6, respectively; hence validating the EBK for precipitation.

Slope

A digital elevation model (DEM) was used for slope calculation. DEM was used for slope calculation. The data for slope calculation were acquired from the Shuttle Radar Topography Mission Digital Elevation Model (30 m × 30 m resolution) data for its better accuracy both in its horizontal and vertical resolution [18]. The slope was calculated in the "degrees" unit using the "slope" tool in the GIS domain.

Stream order

The drainage network (Stream order, SO) was also calculated from DEM. The DEM was sink-filled; then the "Flow direction" tool was run over the filled DEM followed by the flow accumulation. Accumulated flow accumulation is based on the concept that water in each cell will flow towards the steepest downward cell among its eight cells. A threshold value for flow accumulation defining the streams is very important. A threshold works as a divider to distinguish between the streams and small flows which will generally disappear either due to evapotranspiration or infiltration. A series of threshold tests were run to compare the extracted streams with the water features on the world topographic map provided by ESRI on ArcGIS online. Finally, a threshold of 3000 (3k) cells was found to be the most appropriate for the Farta Woreda. All the cells with an accumulated flow greater than or equal to 3k were taken as streams. The extracted streams were numbered based on Strahler stream ordering (Strahler 1957).

Soil texture, land-use, and geology

The soil map of Farta Woreda was modified from the FAO [19] soil map of Ethiopia. Land-use of Farta Woreda was prepared from Landsat 8 OLI/TIRS images (for the 2020 year) downloaded from the United States Geological Survey (http://earthexplorer.usgs.gov) with a spatial resolution of 30 m. A supervised learning classification was done in the ArcGIS 10.3 platform. The geology of the Farta Woreda was prepared from the Ethiopian geological survey of Ethiopia [20].

Time of concentration

Time of concentration (Tc) is the time taken by the overland flow to travel from the hydraulically distant point to the watershed outlet (where dam sites were selected).

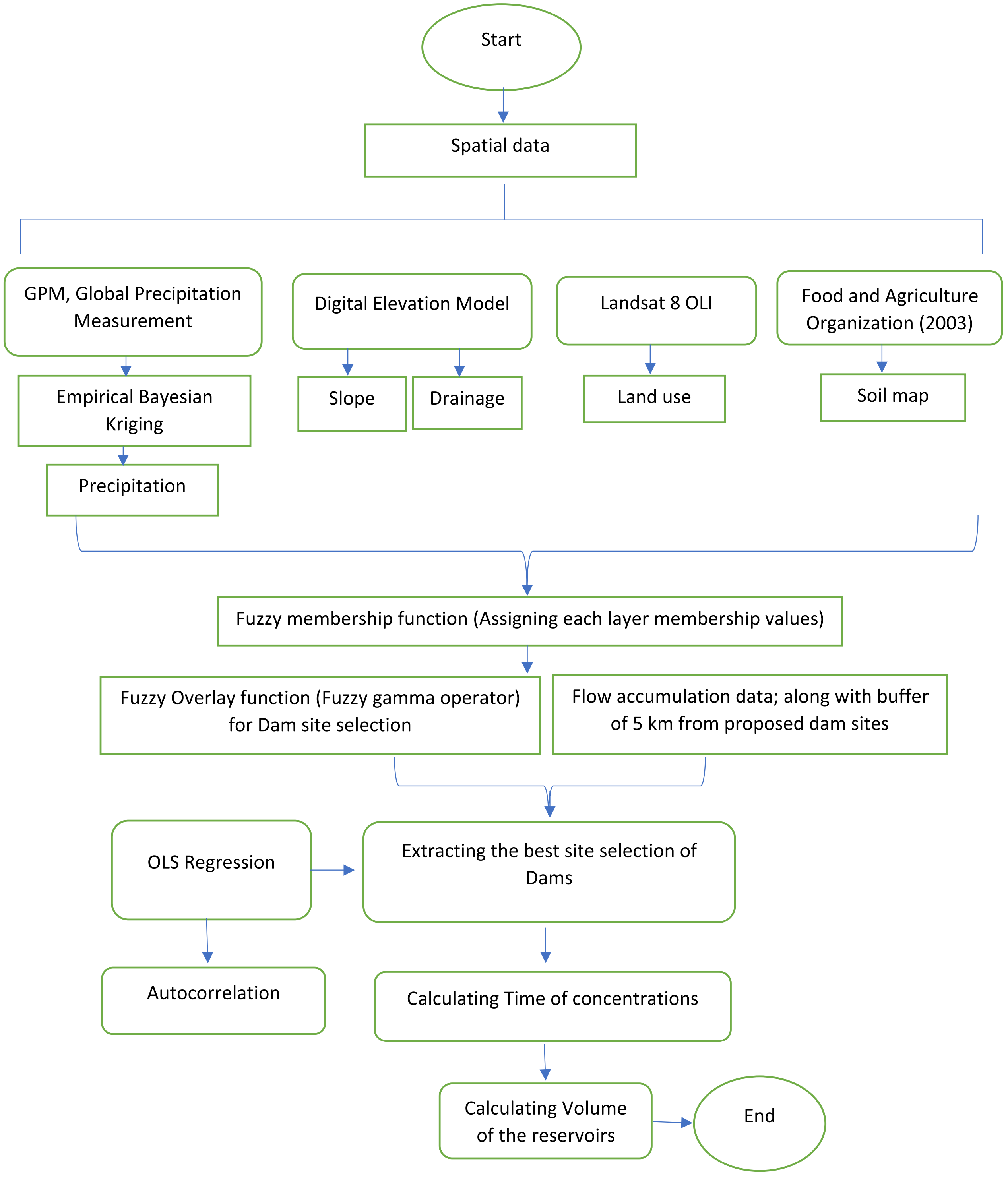


Fig. 2 Methodology used

Tc for all the seven watersheds (1–7) associated with the seven selected dam sites was calculated using the Kirpich equation [21]:

$$T_c = 0.007[L^{0.77}/S^{0.385}],$$

where Tc is the time of concentration (min); L is the length of the channel from headwater to outlet (ft), and S is the slope of the longest hydraulic length (ft/ft).

Reservoirs' volume

The volume of the selected dams has been computed in the GIS domain. According to the location of the dams in combination with the DEM (Digital Elevation Model), TIN (Triangulated Irregular Network), and 5 m and 2 m contour layers. The contour that marks the reservoir's maximum level has been chosen and converted into a polygon to calculate the total volume of the reservoir.

Ordinary least squares regression

The relationship between the various factors viz., precipitation, land-use, geology, etc. (independent variables) and potential site areas (dependent variable) has been carried out using ordinary least squares (OLS) regression in the GIS domain using the following equation:

$$\begin{aligned} A = {} & a0 + a1 \times 1 + a2 \times 2 + a3 \times 3 + a4 \\ & \times 4 + a5 \times 5 + a6 \times 6 + e, \end{aligned}$$

where "A" is the potential dam sites, dependent variable; X1 is the precipitation; X2 is the slope; X3 is the flow accumulation; X4 is the soil texture; X5 is the land use; X6 is the geology; α0 is the intercept; α1…α6 are the respective coefficients and e is the error/residuals. OLS was chosen because it creates easily interpretable output feature class and optional tables with coefficient information.

Spatial autocorrelation

A spatial autocorrelation tool was employed to check if the residuals exhibit a Gaussian pattern. A spatial autocorrelation tool (Global Moran's I) was run to check if the residuals of OLS Regression are clustered or dispersed. Furthermore, the Incremental autocorrelation tool was also run at distance bands of 466.359842 with the beginning distance of 9928 m to check if there is any possibility of clustering of residuals at any different distances.

Results

Precipitation in Farta Woreda varies from 1157.44 to 1701.97 mm/yr. The general distribution trend is low in the north and northeast part; while the distribution trend is increasing towards the south of the study area. The highest precipitation of 1599.48—1701.97 mm/yr. was found to cover an extent of 147.61 km^2 towards the south; While the lowest precipitation of 1157.44—1262.07 mm/yr. was found to cover an extent of 276.62 km towards the north. The spatial distribution map of the precipitation is shown in Fig. 3a.

The slope of the study area ranges from 0 to 67.18 degrees. The steep slopes of 38.41—67.18° were found to cover 14.63 km^2 of the study area. Whereas, flat to gentle slopes of 0—5.12° were found to cover an areal extent of 272.40 km^2.The spatial distribution of slope gradient in the Farta Woreda is shown in Fig. 3b.

The total number of streams found in the study area is 306. The number of streams of the first order is 168. The number of streams of 2nd, 3rd, and 4th order is 82, 41, and 15, respectively. The spatial distribution of stream order is shown in Fig. 3c.

In the case of soil texture, sandy loam was found to cover (257.0944 km^2) the eastern part of the study area. Whereas, loam was found to cover (1101.4946 km^2) the rest of the study area. The spatial distribution map of the soil texture is shown in Fig. 3d.

In the case of land use, annual cropland with an extent of 495 km^2 was found to dominate the study area followed by closed shrubland (299km^2), closed grassland (179km^2), open shrubland (159km^2), woodland (63km^2), dense forest (59km^2), perennial cropland (49km^2), sparse forest (34.35km^2), settlement (12.54km^2), wetland (0.38km^2), water body (0.052km^2), lava flow (0.02km^2), and salt pan (0.005km^2). The spatial distribution of land use in the study area is shown in Fig. 3e.

All these thematic layers were assigned the fuzzy membership values varying from 0 to 1. The fuzzified thematic layers are shown in Fig. 4a–e. These fuzzified layers were overlaid using the "Fuzzy overlay tool" in the GIS domain. The resulting map (Fig. 5a) shows the site selection areas for dam construction.

Initially, a total of 26 dam sites were proposed based on the precipitation, slope, stream order, soil, and land use of the study area (Fig. 5a). The proposed 26 dam sites with their flow accumulation, elevation, precipitation, slope, and stream order are shown in Table 1. Out of 26 proposed sites, 19 sites were discarded depending on their proximity to other dams, proximity to settlements. Flow accumulation and geology were also considered for dam site selection. The intersection tool was used to select the appropriate dam sites. The selected dam sites along with the underlying geology are shown in Fig. 5b. The selected dam sites with their flow accumulation, elevation, precipitation, slope, and stream order are shown in Table 2.

The Tc of the watersheds associated with the selected dam sites varies from 15.64 min (WS2) to 112.24 (WS7) (Fig. 6). The length of the channel

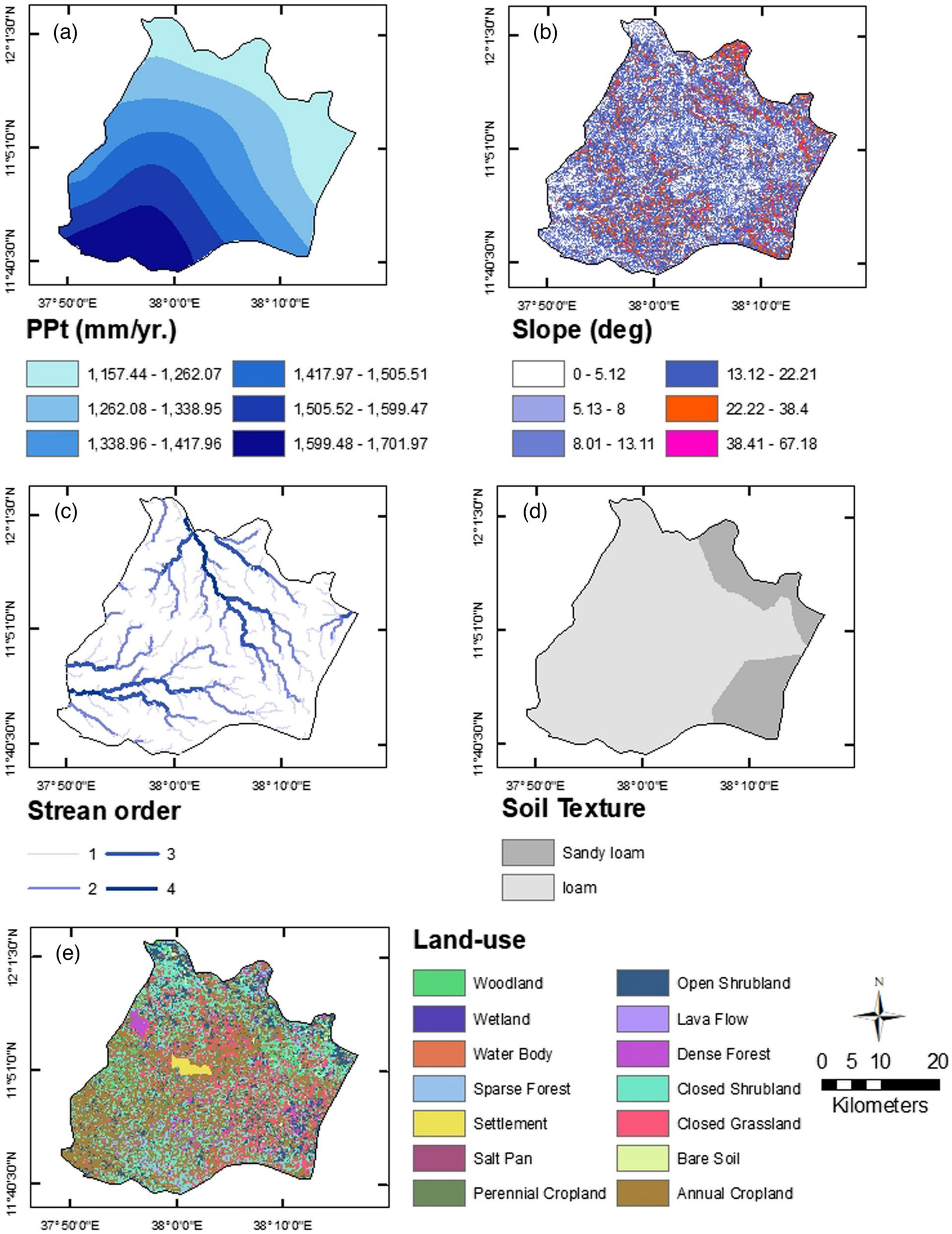


Fig. 3 **a** Precipitation (mm/yr.); **b** slope (degrees); **c** stream order; **d** soil texture, and **e** land-use of Farta Woreda
]

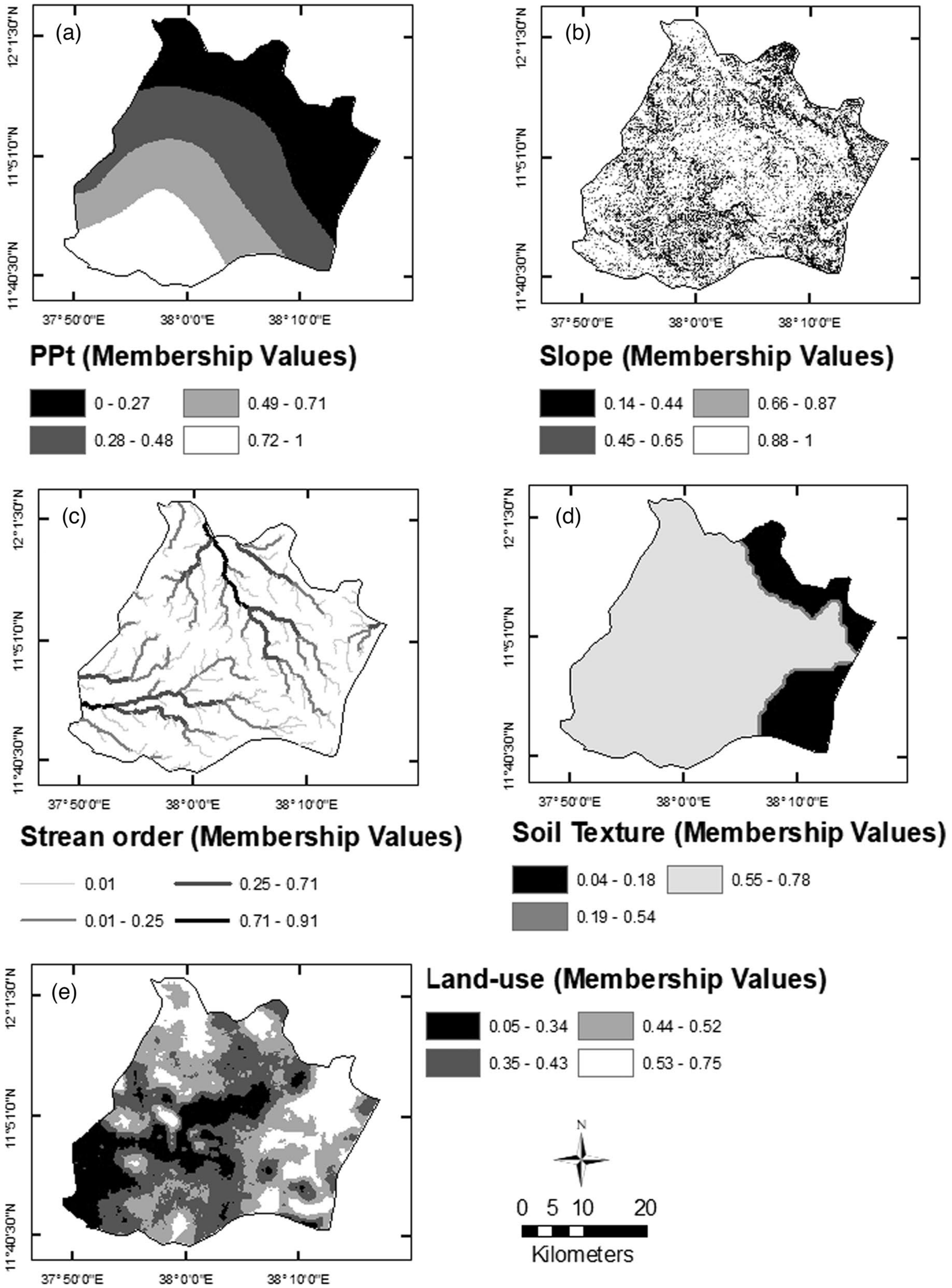


Fig. 4 Fuzzy membership values of, **a** precipitation; **b** slope; **c** stream order; **d** soil texture, and **e** land-use of Farta Woreda

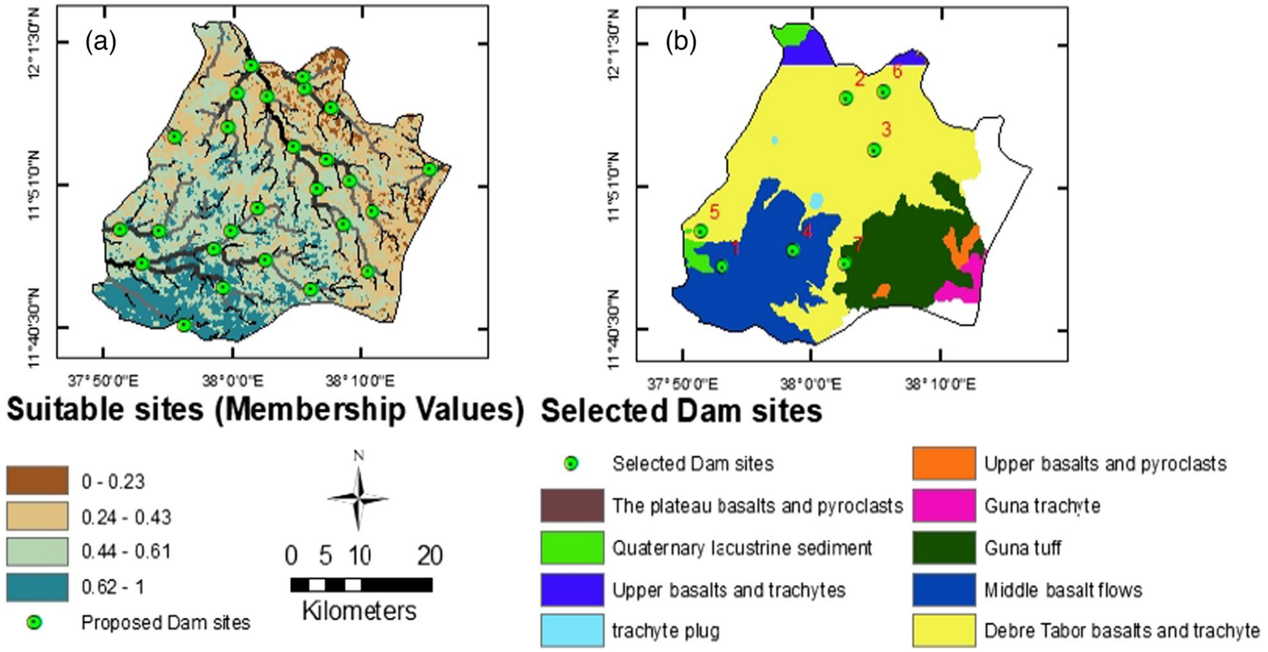


Fig. 5 **a** Fuzzy membership values of suitable areas for proposed dam sites; **b** selected dam sites super positioned on geology

from headwater to outlet (ft), the slope of the longest hydraulic length, and the Tc for all the seven watersheds associated with the selected dams were calculated in the GIS domain (Table 2). The maximum storage capacity of the selected dams is also calculated in the GIS environment (Table 2).

Independent variables associated with Variance of Influence Factor (VIF) greater than 7.5 should be removed while processing OLS regression. It is because higher VIF values indicate redundancy. The selected variables viz., precipitation, slope, flow accumulation, soil texture, and geology were found to have low VIF values along with the significant robust probability statistics (Table 3). The multiple R squares values and the adjusted R-square values were found to be 0.835894 and 0.872153, respectively. The Akaike's Information Criterion (AICc) of the OLS was found to be 4630.433. This indicates that OLS has captured the heterogeneity of independent variables (Table 4). The spatial distribution of standardized residuals generated using the OLS tool is shown in Fig. 7. Using the spatial autocorrelation tool, the standardized residuals were found to exhibit a Gaussian distribution (Fig. 8) with the Global Moran's I summary shown in Table 5. Even the incremental autocorrelation carried shows the Gaussian pattern of standardized residuals. The summary of incremental autocorrelation is shown in Table 6.

Discussion

Precipitation has a direct relationship with the runoff water amount. Precipitation has a positive influence on the function of a dam except for landslides and floods caused by strong precipitation. Areas under high precipitation (found in the south and southwest of the study area) were found to have high fuzzy membership values (0.72–1). These areas are considered to be suitable for dam site construction. While areas with low precipitation (found in the north and northeast of the study area) were assigned a low fuzzy membership value (0–0.27) because these areas are less suitable for dam site construction. Flat to gentle slopes (0–5.12°) were considered to be the most suitable for dam site construction (due to stability of foundations) and hence were assigned the highest fuzzy membership values of 0.88–1; whereas, steep slopes (38.41–67.18°) were assigned the lowest fuzzy membership values of 0.14–0.44. This is because the steep slopes will cause dam instability. Drainage network/stream order (SO) provides the necessary runoff water for the dam function. Different levels (orders) of the drainage network provide a different amount of runoff water to the dam. The highest order or level provides the maximum amount of

Table 1 Proposed dam sites with their flow accumulations, precipitations, slope gradient, stream order, and their coordinates

Dam sites	Flow accumulation	Elevation (meters)	Precipitation (mm/yr.)	Slope (degree)	Stream order	*X* Coordinate	*Y* Coordinate
1	824308	1946	1557.54	0.71643	4	37° 53′ 7.325" E	11° 45′ 15.472" N
2	787957	1956	1268.53	3.07868	4	38° 2′ 45.971" E	11° 57′ 38.931" N
3	575443	2197	1332.99	7.816	4	38° 4′ 51.596" E	11° 53′ 54.808" N
4	506873	2069	1575.54	1.13269	3	37° 58′ 41.849" E	11° 46′ 26.984" N
5	282083	1921	1422.69	0	3	37° 51′ 24.297" E	11° 47′ 51.838" N
6	256439	1886	1234.77	0.506606	3	38° 1′ 30.824" E	11° 59′ 51.979" N
7	237329	2043	1279.69	3.57768	3	38° 0′ 27.761" E	11° 57′ 54.700" N
8	235823	2013	1234.85	0.506606	3	38° 5′ 38.840" E	11° 58′ 12.577" N
9	227807	2227	1498.4	2.1484	3	38° 2′ 37.233" E	11° 45′ 32.532" N
10	221888	2461	1308.03	11.6633	3	38° 7′ 23.840" E	11° 52′ 57.264" N
11	197299	2012	1505.13	2.08792	3	37° 54′ 22.703" E	11° 47′ 38.881" N
12	165981	2205	1523.15	1.13269	2	38° 0′ 2.960" E	11° 47′ 45.380" N
13	153985	2095	1245.17	9.52373	3	38° 7′ 43.971" E	11° 56′ 47.765" N
14	152606	2519	1287.32	8.1909	3	38° 9′ 16.450" E	11° 51′ 25.935" N
15	145132	2095	1188.27	0.71643	3	38° 15′ 34.733" E	11° 52′ 17.070" N
16	116380	2756	1318.4	0	2	38° 8′ 46.144" E	11° 48′ 13.019" N
17	107349	2270	1619.13	1.5195	3	37° 59′ 24.827" E	11° 43′ 32.731" N
18	75682	2221	1342.53	4.0776	2	37° 59′ 42.822" E	11° 55′ 20.741" N
19	75604	2429	1443.92	0.506606	2	38° 2′ 6.076" E	11° 49′ 22.738" N
20	73856	2618	1452.32	2.08792	2	38° 6′ 12.302" E	11° 43′ 23.628" N
21	72473	2712	1265.64	1.60165	2	38° 11′ 5.238" E	11° 49′ 12.746" N
22	50883	3049	1327.09	5.27431	2	38° 10′ 43.337" E	11° 44′ 48.406" N
23	29857	2155	1699.24	6.46037	2	37° 56′ 23.269" E	11° 40′ 46.774" N
24	25764	2089	1353.26	6.72805	2	37° 55′ 33.757" E	11° 54′ 40.886" N
25	16715	2555	1339.74	3.07868	1	38° 6′ 39.723" E	11° 50′ 50.430" N
26	9389	2061	1217.94	0.506606	1	38° 5′ 36.470" E	11° 59′ 4.546" N

runoff water to the dam, while the lower levels (orders) provide the minimum amount of runoff water. Hence, the highest stream order 4 was assigned the highest fuzzy membership values (0.71–0.91) and the lowest stream order 1 was assigned the lowest fuzzy membership values of 0.001. It is because the lowest stream order is less suitable for dam site construction due to less amount of runoff water. Different soil textures have different infiltration rates and hence influence the amount of runoff differently. Therefore, soil texture influences the dam function concerning the amount of runoff amount generated. Sandy loam is less suitable for dam site construction and hence was assigned the lowest fuzzy membership values of 0.04–0.18, while areas are underlain by loam soil texture was given the highest fuzzy membership values (0.56–0.78). In the case of land use, wetlands, water body, and settlements were restricted from dam site construction using restriction modeling in the GIS domain. Areas with perennial cropland and annual cropland were assigned the lowest fuzzy membership values (0.05–0.34) because we are selecting the dam sites for agriculture purposes without disturbing such land uses. Areas covered with bare soil were assigned the highest fuzzy membership values of 0.53–0.75. These fuzzified layers were overlaid to get suitable areas for dam site selections. Areas with the highest membership values are the most suitable for the dam sites and areas with the lowest fuzzy membership values are the least suitable for dam site selection.

The resistance of a geological layer influences the dam's safety. Geology is one of the many factors influencing dam construction. The most common cause of dam failure is the geology of the area underlying the dam construction. Geology with relatively high resistance to pressure, infiltration, and erosion are competent rock foundations. The most satisfactory materials for the desirable dam foundations are igneous rocks (e.g., granite), metamorphic rocks (e.g., quartzite), and sedimentary rocks (e.g., thick-bedded and flat-lying sandstones). In the Farta Woreda, the areas underlain by pyroclastics and lacustrine sediments give low resistance to the dam construction (hence less suitable for dam construction),

Table 2 Selected dam sites with their flow accumulations, precipitations, slope gradient, stream order, time of concentrations, storage, and their coordinates

Dam sites	Flow accumulation	Elevation (meters)	Precipitation (mm/yr.)	Slope (degree)	Stream order	Watershed area (km^2)	Tc (minutes)	Storage (10^{6}m^3)	X Coordinate	Y Coordinates
1	824308	1946	1557.54	0.71643	4	126.974	20.03	4.49994927	37° 53′ 7.325" E	11° 45′ 15.472" N
2	787957	1956	1268.53	3.07868	4	85.0056	15.64	2.21828482	38° 2′ 45.971" E	11° 57′ 38.931" N
3	575443	2197	1332.99	7.816	4	230.1776	23.93	4.82590126	38° 4′ 51.596" E	11° 53′ 54.808" N
4	506873	2069	1575.54	1.13269	3	111.6264	17.27	3.48587614	37° 58′ 41.849" E	11° 46′ 26.984" N
5	282083	1921	1422.69	0	3	112.8336	17.87	2.03720034	37° 51′ 24.297" E	11° 47′ 51.838" N
6	235823	2013	1234.85	0.506606	3	94.3296	16.11	1.032745456	38° 5′ 38.840" E	11° 58′ 12.577" N
7	227807	2227	1498.4	2.1484	3	91.1232	112.24	1.589764678	38° 2′ 37.233" E	11° 45′ 32.532" N

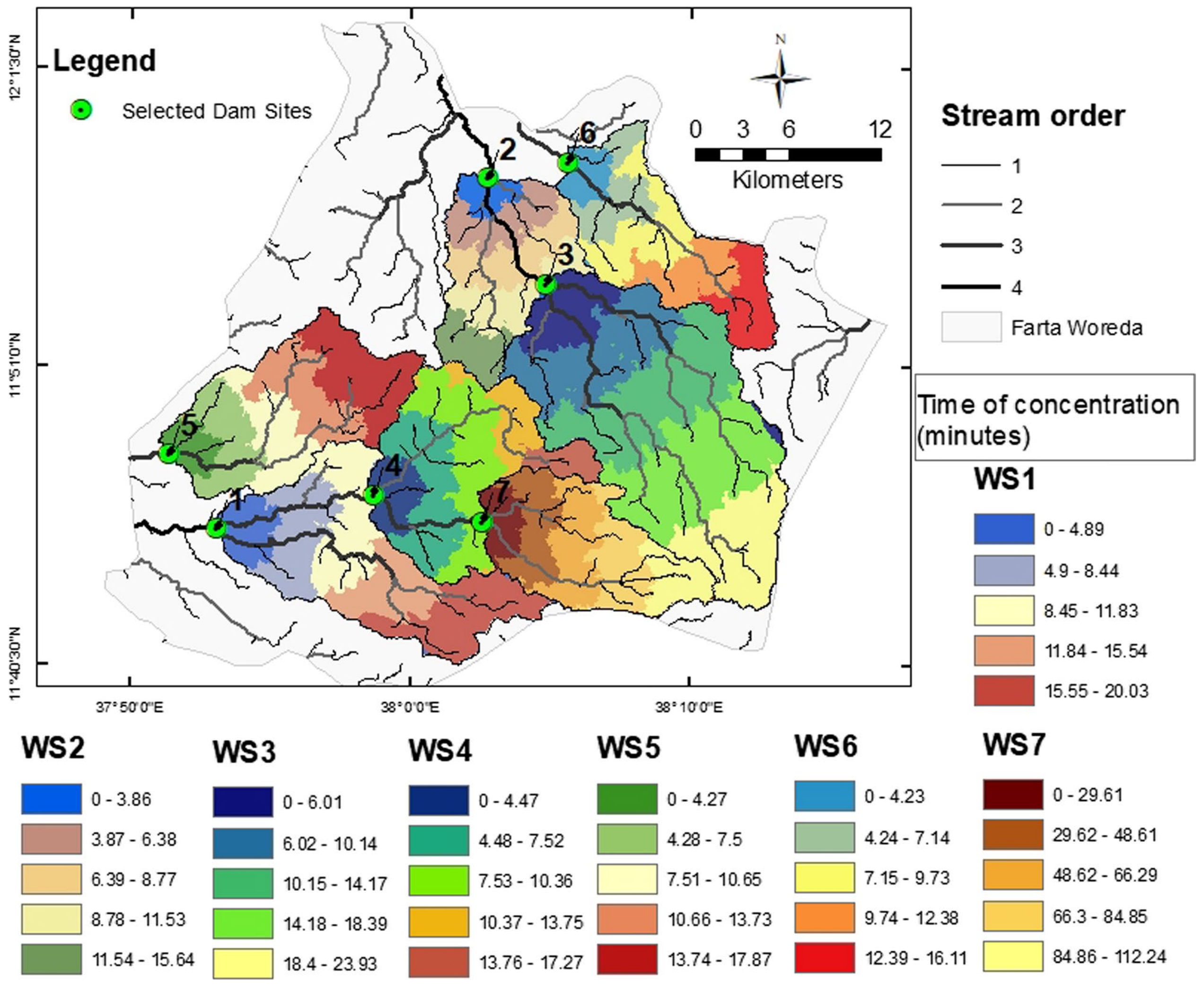


Fig. 6 Time of concentrations of the seven watersheds (WS1–WS7) associated with the selected dam sites (1–7)

Table 3 OLS model statistics (coefficient, robust probability and variance inflation factor of independent variables)

Variable	Coefficient	Robust probability	VIF[2]
Intercept	1.673	0.000012[1]	–
Precipitation	0.4525741	0.000171[1]	1.03476
Slope	−0.0020713	0.000006[1]	1.21112
Flow accumulation	0.2341236	0.000176[1]	1.09731
Soil texture	0.0003421	0.000237[1]	1.89732
Land-use	0.2483	0.000017[1]	1.71214
Geology	0.1511675	0.000283[1]	1.40911

[1] Statistically significant at the 0.05 level

[2] Large VIF (> 7.5) indicates independent variable redundancy

while areas underlain by basalts gives high resistance to the dam construction (hence more suitable for dam construction).

Finally, it can be seen that out of the seven watersheds associated with the selected dam sites; the time of concentration is the least in the case of watershed 2 associated with dam site 2 (15.64 min.), followed by watershed 6 (dam site 6, 16.11 min.; watershed 4 (dam site 4,17.27 min); watershed 5 (dam site 5, 17.87 min.); watershed 1 (dam site 1,20.03 min.); watershed 3 (dam site, 23.93 min.), and watershed 7 (dam site 7, 112.24 min.). Therefore, priority for the dam construction sites should follow the following order:

Dam site 2 > Dam site 6 > Dam site 4 > Dam site 5 > Dam site 1 > Dam site 7.

This is because the lower the Tc, the quicker overland flow will reach the dam site without losing much water

Table 4 OLS regression diagnostics

Number of observations	1463	Akaike's information criterion (AICc)[1]	4630.433
Multiple R- squared[1]	0.835894	Adjusted R-squared	0.872153
Joint F statistic[2]	743.1133	Prob(> F), (6,1456) degrees of freedom:	0
Joint Wald statistic[3]	2300.732	Prob(> Chi-squared), (6) degrees of freedom:	0
Koenker (BP) statistic[4]	17.31002	Prob(> Chi-squared), (6) degrees of freedom:	0.003172*
Jarque–Bera statistic[5]	0.312731	Prob(> Chi-squared), (2) degrees of freedom:	0.813711

[1] Measures model performance/fit

[2] Significant *p*-value indicates model significance

[3] Significant *p*-value indicates robust significance

[4] When this test is statistically significant ($p < 0.01$), the relationships modeled are inconsistent

[5] Significant *p*-value ($p < 0.01$) indicates residuals deviate from a Gaussian distribution

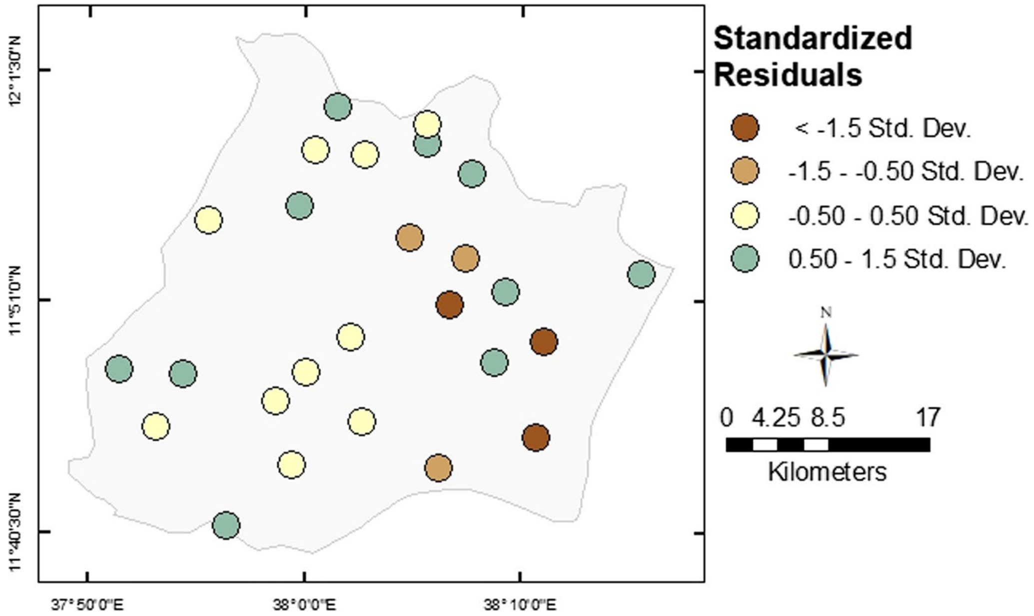
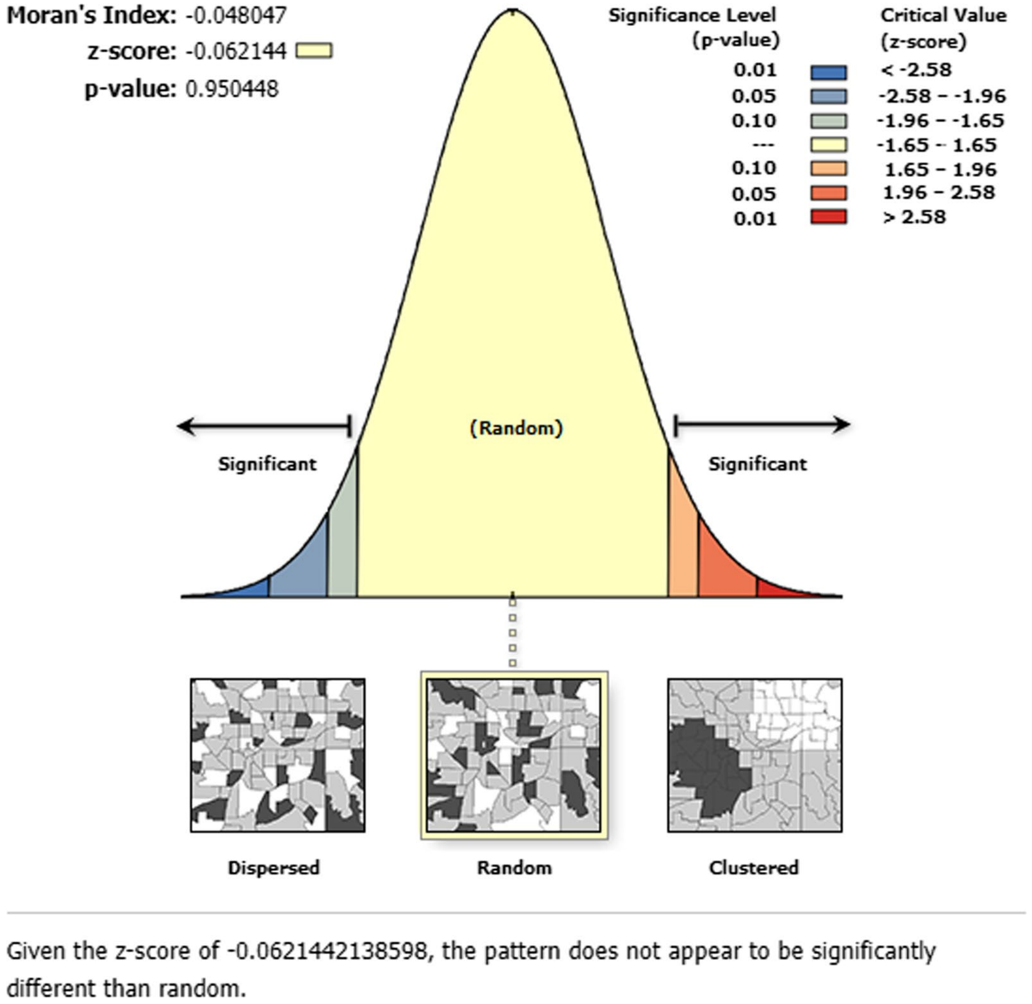


Fig. 7 Spatial distribution of standardized residuals

to evaporation and infiltration. These seven dams will help much for the irrigation of croplands in the associated watersheds. Furthermore, it's quite obvious from the table that dam site 3 is having a maximum storage capacity of 4.285*10^6 m^3 followed by dam site 1 (4.499*10^6 m^3), dam site 4 (3.485*10^6 m^3), dam site 2 (2.218*10^6 m^3), dam site 5 (2.037*10^6 m^3), dam site 7 (1.589*10^6 m^3), and dam site 6 (1.032*10^6 m^3).

The OLS results show that all the independent variables showed a positive coefficient relationship with the potential dam site variable except for the slope. The maximum coefficient has been observed in precipitation (0.4525741) followed by land-use (0.2483), flow accumulation (0.2341236), geology (0.1511675), and soil texture (0.0003421). While as, the slope has been observed to have a coefficient value of (−0.0020713). This is because low slopes represent the high dam site potential. It implies that the potential dam site favors the high precipitation over a low slope with hard rock terrain overlaying on loamy soil on the barren land. The coefficients obtained by each independent variable provide explanations for the OLS model. It can be seen that all the independent variables have a significant robust probability at a 0.05 level. Lower VIF values (<7.5) also indicate the absence of redundancy in the independent variables.

The appraisal of the model significance was done by employing Joint F-statistic and Joint Wald Statistics (OLS diagnostics). If Koenker (BP) statistic is statistically significant, then Joint F-statistic cannot be trustworthy. In our study, Koenker (BP) statistic was found statistically


Moran's Index: -0.048047

z-score: -0.062144

p-value: 0.950448

Given the z-score of -0.0621442138598, the pattern does not appear to be significantly different than random.

Fig. 8 The Gaussian pattern of standardized residuals (represented by the z-score)

insignificant ($p > 0.01$), proving that the relationship modeled was consistent. Jarque–Bera statistic was also found to be statistically insignificant ($p > 0.01$) indicating the gaussian distribution of residuals. Any regression modeling comes up with certain errors or residuals. These standardized residuals must exhibit a Gaussian pattern, then the only model is unbiased. In our study first, we carried out Spatial Autocorrelation on standardized residuals which yielded their Gaussian pattern. Furthermore, to further crosscheck, we employed incremental autocorrelation with a certain distance increment (466.359842); with a beginning distance of 9928.0 m; keeping the row standardization "true" to check if there might be a correlation at some distance, but the model was again found to be unbiased.

Table 5 Global Moran's I summary

Moran's index	−0.048047
Expected index	−0.040000
Variance	0.016769
z-score[1]	−0.062144
p-value	0.950448

[1] *z*-score value of 0.011335 shows that standard residuals exhibit a Gaussian pattern

Conclusion

Several factors (viz., precipitation, slope, flow accumulation, land use, and geology) were analyzed in the GIS domain. These factors were assigned with the appropriate fuzzy membership values and overlaid to get the map of potential dam sites (wherein areas with high fuzzy membership values are highly suitable for dam construction and vice-versa). Initially, a total of 26 dam sites were proposed. Dam sites with flow accumulation

Table 6 Global Moran's I summary by distance

[1]Distance	Moran's index	Expected index	Variance	*z*-score	*p*-value
9928	−0.035653	−0.04	0.018534	0.031927	0.97453
10394.35984	−0.047983	−0.04	0.015618	−0.06388	0.949066
10860.71968	−0.042666	−0.04	0.014733	−0.021965	0.982476
11327.07953	−0.071272	−0.04	0.01395	−0.264766	0.79119
11793.43937	−0.004242	−0.04	0.012883	0.315041	0.752731
12259.79921	0.009373	−0.04	0.011336	0.463718	0.642849
12726.15905	−0.040785	−0.04	0.010246	−0.007752	0.993815
13192.51889	−0.026187	−0.04	0.009603	0.140952	0.887908
13658.87873	0.022386	−0.04	0.008469	0.677931	0.497815
14125.23858	0.034498	−0.04	0.00786	0.840316	0.400731

[1] Distance measured in meters; beginning distance 9928.0; distance increment 466.359842

above 3k were selected to be the potential sites for dam construction. Of 26 dam sites, only 7 dam sites were selected depending on their proximity to nearby dams, settlements, flow accumulation, and geology. The factors chosen to decide the potential dam sites were subjected to OLS regression to understand the extent of the relationship of these factors with the potential dam sites. The coefficients obtained by these factors explain the model. A lower value of VIF confirms the absence of multicollinearity and redundancy among the factors chosen. Three kinds of tests were performed on residuals viz, Koenker's (BP) statistic, spatial autocorrelation, and incremental autocorrelation. All the tests revealed a Gaussian distribution of residuals, indicating that the model is not biased. The proposed sites are feasibly located which will yield multi-benefits such as flood reduction, water for agriculture, fish industry, and hydropower generation This study proves the importance of fuzzy logic coupled with OLS regression to be a powerful tool in deciphering the potential dam sites and can be used at a regional and continental scale.

Acknowledgements
The authors are grateful to the reviewers for their valuable comments. I (Imran Ahmad) am very much thankful to Henner Hollert, Editor (Environmental Sciences Europe) for handling this manuscript.

Author contributions
ZF, IA and MD conceptualized and drafted the methodology, interpreted the results. AF drafted the literature review. AT, AB, AT, MD, HN, TA, MB, DT, and ES reviewed the paper with necessary corrections. All authors read and approved the final manuscript.

Funding
Not available.

Availability of data and materials
The data will be provided based on a request.

Declarations

Ethics approval and consent to participate
We all authors approve the paper.

Consent for publication
We all authors have given consent for publication.

Competing interests
On behalf of all authors, the corresponding author states that there are no competing interests.

Author details
[1]China Water Resources Pearl River Planning, Surveying & Designing Co., Ltd., Guangzhou, China. [2]Department of Hydraulic and Water Resources Engineering, Debre Tabor University, Debre Tabor, Ethiopia. [3]Integrated Watershed Management Programme, Department of Rural Development and Panchayati Raj, Government of Jammu and Kashmir, Srinagar, India.

Received: 2 May 2022 Accepted: 13 August 2022

References

1. Lire Ersado (2005) World Bank Policy Research Working Paper 3494, January 2005.
2. CGIAR, Climate Resilient Green Economy Strategy. Sector-wise GTP II (2018) Implementation Monitoring Checklist. CGIAR Research Program on Climate Change; Agriculture and Food Security (CCAFS): Wageningen, The Netherlands.
3. Abduselam (2017) Food security situation in Ethiopia: a review study. Int J Health Econ Policy 2(3):86–96
4. Haile T (1988) Causes and characters of drought in Ethiopia. Ethiopian J Agric Sci 10(1–2):85–97
5. Kloos H (1991) Peasant irrigation development and food production in Ethiopia. Geogr J 157:295–306
6. Harrower MJ, Nathan S, Mazzariello JC, Kifle Z, Dumitru IA, Yemane M, Bongers JL, Gidey G, Zaitchik BF, Anderson MC (2020) Water, geography, and Aksumite civilization: the Southern Red Sea archaeological histories (SRSAH) project survey (2009–2016). Afr Archaeol Rev 37:51–67
7. Ayana M, Gebremeskel T, Mengistu A, Fasil E, Melkamu M (2015) Irrigation water pricing in Awash River Basin of Ethiopia: evaluation of its impact on scheme-level irrigation performances and willingness to pay. Afr J Agric Res 10:554–565
8. Abushandi E, Merkel B (2011) Rainfall estimation over the Wadi Dhuliel arid catchment, Jordan from GSMaP_MVK+. Hydrol Earth Syst Sci 8:1665–1704

9. Forzieri G, Gardenti M, Caparrini F, Castelli F (2008) A methodology for the pre-selection of suitable sites for surface and underground small dams in arid areas: a case study in the region of Kidal. Phys Chem Earth 33:74–85
10. Kite W, Pietroniro A (1996) Remote sensing applications in hydrological modelling. Hydrol Sci 41:563–591
11. Liu J, Chen JM, Cihlar J (2003) Mapping evapotranspiration based on remote sensing: an application to Canada's landmass. Water Resour Res 39:1189
12. Mahmoud S (2014) Investigation of rainfall-runoff modeling for Egypt by using remote sensing and GIS integration. CATENA 120:111–121
13. Jozaghi A, Alizadeh B, Hatami M, Flood I, Khorrami M, Khodaei N, Tousi EG (2018) A comparative study of the AHP and TOPSIS techniques for dam site selection using GIS: a case study of Sistan and Baluchestan Province, Iran. Geosciences 8:494
14. Singh P, Singh D, Litoria P (2009) Selection of suitable sites for water harvesting structures in Soankhad Watershed, Punjab using remote sensing and geographical information system (RS&GIS) approach—a case study. J Indian Soc Remote Sens 37(1):21–35
15. Ibrahim GRF, Rasul A, Hamid AA, Ali ZF, Dewana AA (2019) Suit- able site selection for rainwater harvesting and storage case study using Dohuk Governorate. Water 11:864
16. Youssef A, Pradhan B, Sefry S (2014) Remote sensing-based studies coupled with field data reveal urgent solutions to avert the risk of flash floods in the Wadi Qus (east of Jeddah) Kingdom of Saudi Arabia. Nat Hazards 75(2):1465–1488
17. Dai X. (2016) Dam Site Selection Using an Integrated Method of AHP and GIS for Decision Making Support in Bortala, Northwest China. Master's Thesis, Lund University, Lund, Sweden.
18. Odutola CA, Beiping W, Yao YZ (2013) Assessing vertical accuracy of SRTM ver 4.1 and ASTER GDEM ver 2 using differential GPS measurements—case study in Ondo State. Nigeria. International Journal of Scientific & Engineering Research 4 (12) 523.
19. FAO (2003) Digital Soil Map of the World. Rome.
20. GSE (2011) Geology, Geochemistry and Gravity Survey of West Gonder and Gonder Map Sheets (unpublished). Addis Ababa, Ethiopia.
21. Kirpich ZP (1940) Time of concentration of small agricultural watersheds. J Civ Eng 10(6):362
22. Mohamed AA (2017) Food security situation in Ethiopia: a review study. Int J Health Econ Policy. 2(3):86–96. https://doi.org/10.11648/j.hep.20170203.11

Publisher's Note