Assessment of Household's Spatial Access to Water Points in Dodoma Rural Wards

Frank Mahuve1, Beatrice Christopher1, and Job Asheri Chaula2

1Department of Geospatial Sciences and Technology, Ardhi University, Dar es Salaam, Tanzania
2Department of Computer Systems and Mathematics, Ardhi University, Dar es Salaam, Tanzania

IJASR 2019
VOLUME 2
ISSUE 6 NOVEMBER - DECEMBER **ISSN: 2581-7876**

Abstract – Access to water points is often determined using cumulative based accessibility methods, particularly the coverage-based method. The coverage-based method fails to account for the effect of distance variation as well as the possibility of a household to be served by multiple, under-utilized, or over-utilized water points. As a result, uniform accessibility indices are reported regardless of the variation in distance to water points, whether households are served by multiple, under-utilized, or over-utilized water points. Thus, this paper integrated Fuzzy set operations and the Enhanced Two-Step Floating Catchment Area (E2SFCA) to address the limitations of the coverage-based methods. This new method, Fuzzy-based E2SFCA was employed in this study to assess the variation in household's spatial access to water points in Dodoma Rural Wards. The resulting accessibility indices were reported for each individual household, and vary from one household to another, except in some areas where there are water points of zero capacities. The accessibility indices range from 0 to 0.0204 within the nine selected rural Wards in Dodoma Urban district. The higher the index, the higher the household has access to water points. Households with the accessibility of 0.004 (100%) are considered to have full access to water points, while those with accessibility indices below 0.004 (below 100%) or above 0.004 (above 100%) are respectively considered to be underserved or over-served. The capability of the fuzzy-based E2SFCA method to eliminates the problem of reporting uniform accessibility indices encountered in the coverage-based method, and provide reliable access to water points is very apparent in this study. Thus, this approach can be adopted in accessibility studies of various service points such as education centres, health centres, and business centres, to mention few. Moreover, the findings of this study have implications for planning, decision making, and policy making regarding the improvement of access to water points in various areas. For further studies, when weighing the sub-catchment areas, it is recommended to generate pairwise comparison matrices based on the "household's utilization of water points at different distances" data whenever it is available.

Keywords: Water points, fuzzy set operation, E2SFCA, and spatial access.

Introduction

Access to services can be defined based on several specific dimensions, including availability, accessibility, accommodation, affordability, and acceptability, which describe the fit between the service consumers and the service delivery system [1]. The first two dimensions are spatial in nature, while the last three dimensions are non-spatial in nature [2]. The focus of this study is on integrating the first two dimensions to assess spatial access to water points. According to Guagliardo (2004), availability is defined as "the number of local service points from which a client can choose", and accessibility as "the travel impedance (distance or time) between client location and service point" [2]. The integration of these two spatial dimensions addresses two limitations of the conventional method. This conventional method, which is also referred to as the coverage-based accessibility measure defines access to water points by establishing the maximum number of people served by each water point and the maximum distance within which water points are located [3]. In the case of Tanzania, these maximum values would be 250 people served by each water point within a radius of 400m [4]. The first addressed limitation is the failure to account for the number of people served by each water point, and their respective spatial distribution [5]. That is, the possibility of water users to obtain water from multiple water points, and the fact that a number of water points might be operating beyond their capacity are not taken into account. The implication of this limitation is the uniform reporting of accessibility indices to all water users, regardless of whether they are served by a single water point or multiple water points. This might lead to over-reporting of accessibility indices for water users served by a single water point, and under-reporting for water users served by multiple water points. The second addressed limitation is the failure to take into account the water users' walk distance to water points [6]. This limitation might lead to the over-reporting of accessibility indices for water users walking longer distances, and underreporting for water users walking shorter distances.

To address these two limitations while retaining the definition of access to water points given by the coverage-based accessibility measure, an Enhanced Two-Step Floating Catchment Area (E2SFCA) method proposed by Luo and Qi, was integrated with the fuzzy set operation and employed in this study.

E2SFCA Method

The E2SFCA method is an extension of a coverage-based accessibility measure that makes use of a catchment area twice, once in the service points (i.e., water points) and once in the service users (i.e., water users) [7]. In the accessibility studies, a catchment area refers to a coverage area defined by a maximum established walk distance or time to service points or service users. A catchment area can be defined around service points, service users, or both service points and service users [7], [8]. The use of catchment area twice addresses the problem of uniform accessibility indices, which is attributed to the first limitation of a coverage-based accessibility measure by limiting the availability of water points based on the number water users they serve. That is, water points serving a large number of water users are made less available compared to water points serving the required or small number of water users. The E2SFCA method accounts for the distance impedance, a second limitation of a coverage-based accessibility measure by applying weights to different zones within a catchment area [7]. The implementation of the E2SFCA method is carried out in two steps.

In the first step, a catchment area defined around a service point is divided into several zones, which are assigned weights based on some functions of distance. Then, service users residing within each zone are searched, and the weighted service points – to – service users ratio within the catchment area is calculated using equation 1 [7].

$$R_j = \frac{S_j}{\sum_{k\in\{d_{kj}\in D_r\}} P_k W_r} = \frac{S_j}{\sum_{k\in\{d_{kj}\in D_1\}} P_k W_1 + \sum_{k\in\{d_{kj}\in D_2\}} P_k W_2 + \ . \ . \ .} \qquad (1)$$

where R_j is the weighted service points – to – service users ratio, S_j is the number of service points within the catchment j, P_k is the number of service users located within the sub-catchment (zone) r, D_r is the r[th] zone within the catchment, W_r is the weight for the r[th] zone, and d_{kj} is the travel distance measured from a service point j to a service user k.

In the second step, a catchment area defined around a service user (i.e., a household) is divided into several zones, which are assigned weights that are similar to those derived in the first step. Then, service points located within each zone are searched, and their respective weighted service points – to – service users ratio are summed up using equation 2 to determine service users' access to service points [7].

$$A_i = \sum_{j\in\{d_{ij}\in D_r\}} R_j W_r = \sum_{j\in\{d_{ij}\in D_1\}} R_j W_1 + \sum_{j\in\{d_{ij}\in D_2\}} R_j W_2 + \ . \ . \ . \qquad (2)$$

where $\mathrm{A_i}$ is the accessibility index for service user i, $\mathrm{R_j}$ is the weighted service points – to – service users ratio at service point j that falls within the catchment centred at service user i, $\mathrm{d_{ij}}$ is the travel distance measured from a service user i to a service point j, and $\mathrm{D_r}$ and $\mathrm{W_r}$ are as defined in equation 1.

The E2SFCA method reports accessibility indices, which are uniform within each zone but vary from one zone to another zone [7]. It assumes that distance variation within each zone is insignificant. Consequently, there is an abrupt change in accessibility indices when crossing the boundaries of neighboring zones. For instance, for a catchment area divided into two zones with distance breaks of 0m – 200m and 200m – 400m from a water point, the accessibility index of a water user located at 199m is the same as that of a water user at 0m, but very different to that of a water user at 201m. This contradicts with the first law of geography, which states that "everything is related to everything else, but near things are more related than distant things"[9]. That is, the accessibility index of a water user at 199m is expected to be more similar to the accessibility index of a water user at 201m than the accessibility index of a water user at 0m. In this study, the fuzzy set operation is integrated with the E2SFCA to account for the distance variation within each zone.

Fuzzy Sets Operations

The concept of the fuzzy set was first introduced by Zadeh in 1965 [10]. A fuzzy set is defined as a class of objects with a continuum of grades of membership ranging between zero and one, which are assigned by a membership function characterizing a fuzzy set [10]. That is, some objects partially belong to the set, and objects with more of the common property are allowed to belong to the set more strongly than others [11]. Consequently, the sharp transition from membership to non-membership is eliminated [12]. For further illustration, let $X = \{x\}$ be a collection of objects denoted generically by x, then a fuzzy set A in X is a set of ordered pairs expressed by equation 3 [12].

$$A = \{(x, \mu_A(x))\}, \qquad x \in X \qquad (3)$$

where $\mu_A(x)$ is a grade of membership of x in X, $\mu_A : X \longrightarrow [0, 1]$ is a membership function that maps x elements of X into grades of membership in the interval [0, 1], with 0 and 1 representing, respectively, non-membership and full membership of an element x to a fuzzy set A.

Operations on fuzzy sets are carried out to address the problems of uniform concentrations of attributes (i.e., uniform accessibility indices) within a class, vague classification of entities, and that of sharp delineation of classes (i.e., sudden change of accessibility indices across class (zone) boundaries) [13], [14]. These problems are apparent in the E2SFCA method [7], and are common in other models of spatial analysis that are based on crisp set theory [14]. Such fuzzy set operations combine grades with which an entity is a member of two or more fuzzy sets. That is, they reduce a set of such grades into one representative grade. Some useful operations that can be used to combine such grades include union (OR), intersection (AND), algebraic product, algebraic sum, arithmetic mean, and weighted mean [15]. The first four operations are non-compensative, while the last two are compensative. The first four operations can be categorized into two families, triangular norm (shortly: t-norm) and triangular conorm (shortly: t-conorm). The intersection (AND) and the algebraic product operations are among the basic t-norm operations. These t-norm operations produce the resulting grades that are lower or equal to the minimum of the aggregated grades. That is, the resultant grade will be high if and only if all the aggregated grades are of high values [16]. The union (OR) and the algebraic sum are among the basic t-conorm operations. These t-conorms operations produce the resulting grades that are higher or equal to the maximum of the aggregated grades. That is, the resultant grade will be low if and only if all the aggregated grades are of low values [16]. Arithmetic and weighted mean are among the averaging operations, which allow low grade on one fuzzy set to be compensated by high grade on another fuzzy set, so that the resulting grade becomes higher than the minimum grade, and lower than the maximum grade [16]. Although averaging operations are suitable and are preferred for combining scores in multi-criteria evaluation problems, the arithmetic mean is less preferred to be combined with the E2SFCA method as in some cases, it produces similar grades for two or more service users located in two different zones regardless of their location from a service point. For instance, the same grade of 0.35 is produced for a service user at 150m in zone one (0m – 200m), having grades of 0.5 on zone one and 0.2 on zone two (200m – 400m), and for another service user at 250m in zone two, having grades of 0.2 on zone one and 0.5 on zone two. That is, the two distances, 150m and 250m measured from a service point to service user are assumed to have the same influence on accessibility evaluation. Thus, in this study, the weighted mean fuzzy operation is combined with the E2SFCA method and employed to assess the household's spatial access to water points. The weighted mean, expressed in equation 4 addresses the limitation of the arithmetic mean by incorporating the weight of each zone [14], [16].

$$\mu_A(x) = \frac{\sum_{l=1}^{n} W_l \mu_{A_l}(x)}{\sum_{l=1}^{n} W_l} = \sum_{l=1}^{n} W_l \mu_{A_l}(x) = W_1 \mu_{A_1}(x) + W_2 \mu_{A_2}(x) + \cdots + W_n \mu_{A_n}(x) \qquad (4)$$

where $\mu_A(x)$ is the grade with which object x belongs to the universal set X, W_l is a non negative weight of zone l (i.e., $W_1, W_2, \ldots, W_n$ are non negative weights summing up to unity), $\mu_{A_l}(x)$ Is a grade with which object x belongs to a fuzzy subset of X, A_l (i.e., $A_1, A_2, \ldots, A_n$ are fuzzy subsets of X).

Each zone is given a weight basing on its influence on the evaluation of accessibility indices. Several methods, including ranking, rating, pairwise comparison and trade-off methods can be employed to determine the weight of each zone [17]. Among these weighting methods, a pairwise comparison method developed by Saaty in 1977 is the most promising approach [18].

Pairwise Comparison Method

The method of pairwise comparisons developed by Saaty in 1977 outside the GIS environment, also referred to as Analytical Hierarchy Process (AHP) in the context of a decision making[17] was first introduced in GIS applications by Rao *et al.* in 1991 [17], [18]. In this method, relative weights of criteria are derived from the principal eigenvector of a positive pairwise comparison matrix [19]. The derivation of these weights is carried out in three steps, the construction of a positive pairwise comparison matrix, the estimation of relative weights, and the estimation of the consistency ratio [17].

In the first step, a positive pairwise comparison matrix is constructed by defining its size using the number of criteria considered, and its elements using the rating values indicating the relative importance of the two compared criteria [18]. Since the rows and columns of this matrix are defined by the same criteria, then the position of any element of this matrix is the intersection of a row and a column defined by a respective pair of the two compared criteria. The rating values are provided on a nine-point continuous rating scale (Table 1) proposed by Saaty in 1977 [19].

Table 1: Nine-point continuous rating scale for pairwise comparison [19]

Intensity of Importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
2	Equal to moderate importance	
3	Moderate importance	Experience and judgement slightly favour one activity over another
4	Moderate to strong importance	
5	Strong importance	Experience and judgement strongly favour one activity over another
6	Strong to very strong importance	
7	Very strong importance	An activity is favoured very strongly over another
8	Very strong to the extreme importance	
9	Extreme importance	The evidence favouring one activity over another is of the highest order of affirmation

Since the pairwise comparison matrix is symmetrical and the positions of its diagonal elements are defined by the pairs of identical criteria, only rating values of the matrix elements on one triangle needs to be defined (i.e., rating values of the remaining elements are simply the reciprocal of the defined values), and the rating value of one indicating equal importance of the two compared criteria is defined for all diagonal elements [17], [18].

In the second step, the best fit set of the relative weights of criteria are produced by the principal eigenvectors of the pairwise comparison matrix [18]. A good approximation to these weights can be achieved by normalizing each column of the pairwise comparison matrix and taking the average of the matrix elements over the resulting rows [19]. The normalization of the columns of the pairwise comparison matrix is achieved by dividing each matrix element by the sum of all matrix elements in its column [17].

$$a_{ij_N} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} = \frac{a_{ij}}{a_{1j} + a_{2j} + \ldots + a_{nj}} \qquad (5)$$

where a_{ij_N} is the element i in column j of the normalized pairwise comparison matrix, a_{ij} is the element i in column j of the pairwise comparison matrix, $\sum_{i=1}^{n} a_{ij}$ is the summation of all elements in column j of the pairwise comparison matrix, and i and j are the rows and columns of both the pairwise and the normalized pairwise comparison matrices.

Lastly, since people's feelings and preferences are inconsistent and intransitive, a consistency ratio is estimated to determine how the judgement values (i.e., the matrix elements) are closer to being logically related than to being randomly chosen [19]. The estimation of the consistency ratio is done by comparing the consistency index (i.e., the logical one) with the random consistency index using equation 6.

$$CR = \frac{CI}{RI} \qquad (6)$$

and

$$CI = \frac{\lambda - n}{n - 1} \qquad (7)$$

where CR is the consistency ratio, CI is the consistency index, RI is the random index, λ is the average value of consistency vector, and n is the number of criteria.

The random index is the consistency index of a randomly generated pairwise comparison matrix, and is dependent on the number of criteria. The consistency indices for randomly generated pairwise matrices of various sizes (i.e., number of criteria) are provided in Table 2 [17].

Table 2: Random Indices for some randomly generated pairwise Matrices [17]

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

The average value of consistency vector, λ is estimated by computing the weighted sum vector by multiplying the pairwise comparison matrix with the vector of the corresponding relative weights of the involved criteria, dividing each of the components of the resulting weighted sum vector by the corresponding components of the corresponding relative weights of the involved criteria, and averaging the results [19].

Methodology

This study employed the enhanced two-step floating catchment area (E2SFCA) method, integrated with the weighted mean operation (i.e., the fuzzy set operation) to assess access to water points in Dodoma rural wards. The procedures involved in this method are categorized into two groups, those which are carried out outside the GIS environment, and those which are carried out in a GIS environment. The former group includes sampling design, data collection, and evaluation of relative weights for each fuzzy set, which are presented in sections 2.2 to 2.4. The latter group includes computation of the weighted water points to population ratio, and evaluation of households' access to water points, which are presented in sections 2.5 and 2.6, and described by two portions of the Fuzzy-based E2SFCA model presented in Figures 2 and 3. The remaining section 2.1 describes the location of the study area.

Location of the Study Area

This study was carried out in Rural Wards of Dodoma Urban District, which is the national capital city of Tanzania and the administrative centre of the Dodoma Region. This district is one of the seven districts of Dodoma region bordered by Chamwino district on the eastern side and Bahi districts on the western side as shown in Figure 1.

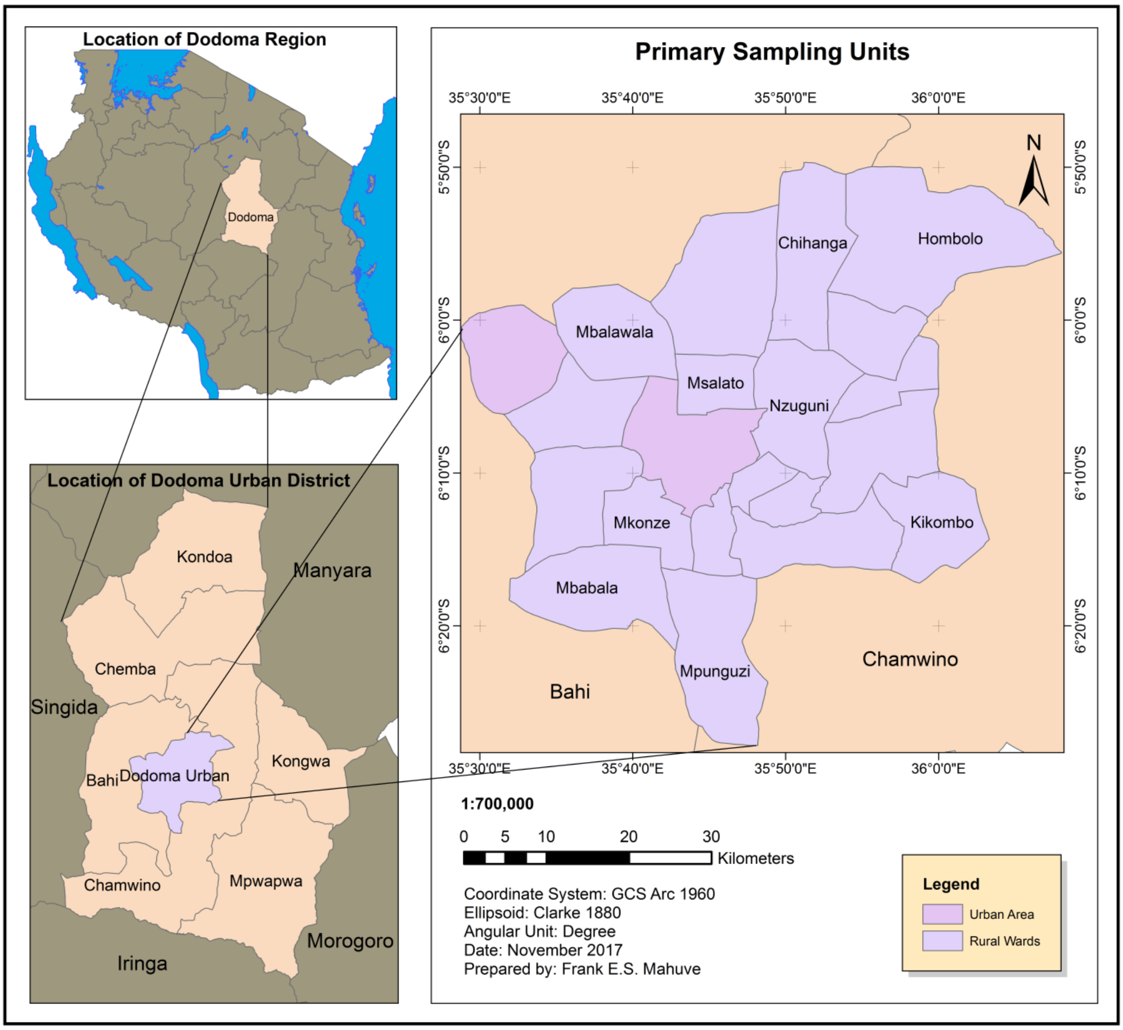

Figure 1: Location of the Study Area

The district is located at an altitude ranging between 1251 m and 1334 m above the mean sea level, covering an area of 2576 km² spanning between Latitudes 5˚48' and 6˚30' south of the Equator and Longitudes 35˚28' and 36˚10' east of the Greenwich meridian. Dodoma Urban district is characterized by a semi-arid climatic condition with little annual rainfall of about 400 mm to 600 mm falling between December and April. The average temperature in the study area ranges between 10ºC in July and 35ºC in September [20]. Administratively, the district is divided into 37 Wards of which 18 are urban Wards and 19 are rural Wards (Figure 1). According to census and housing survey conducted in 2012 by the National Bureau of Statistics (NBS), the population of Dodoma municipality is 410,956 inhabitants, of which 213,636 live in urban areas and 197,320 reside in rural areas [21]. Dodoma Urban District was selected to be a study area based on its semi-arid climatic condition, a condition that makes people residing in the Dodoma urban district to be dependent on groundwater for their daily activities. These people can obtain groundwater by subscribing to Dodoma Urban Water and Sanitation Authority (DUWASA) if their areas are covered by the DUWASA water supply network, or by fetching from either community or privately owned water points (i.e., boreholes) if their areas are not covered by DUWASA. Another reason for selecting Dodoma urban district is based on the expected increase in the demand of water, which could be attributed to the decision of the fifth Government of the United Republic of Tanzania to move the capital of Tanzania and Government institutions from Dar es Salaam to Dodoma.

Sampling Design

The evaluation of spatial access to water points requires two entities, water points and water users (i.e., households). A two-stage sampling technique was employed in the selection of both water points and households considered in this study.

The selection of water points was done in two-stages, clustering of water points and stratification of water points in each cluster. In the first stage, a sampling framework of 228 water points located in nine rural wards in the Dodoma urban district was clustered into nine clusters based on the wards where the water points are located. The sampling framework for water point selection was defined in two phases, the determination of a sample size of nine wards (i.e., labelled wards in Figure 1) from 19 rural wards of the Dodoma urban district, and a random selection of such nine wards. All 228 water points located in these nine wards were extracted from an excel file titled "Water Points Locations in Rural Water Supply 2015-2016" obtained from the Government Basic Statistics Portal. This file was created and updated by the Directorate of Rural Water Supply (DRWS) on August 18, 2015, and on February 23, 2017, respectively. In the second stage, water points in each cluster were stratified into two strata, functional and non-functional water points, making a total of 18 strata for all nine clusters. Finally, water points were randomly selected from each stratum, and 23 water points were selected for all 18 strata. To account for the possibility of a household to be served by multiple water points, 47 water points were selected from 205 water points remaining after randomly selecting the 23 water points. The coverage areas of these 47 water points overlap with the coverage area of any of the 23 water points.

The selection of households was based on the selected water points and involved two stages, clustering of households and stratification of households in each cluster. In the first stage, a sampling framework of 2463 households located in the catchment areas of 23 water points was clustered into nine clusters based on the wards in which households are located. These 2463 households were extracted from 5349 households, which are located in all 70 water points (i.e., the 23 and the 47 water points), and were obtained from Google Earth through digitization. In the second stage, households in each cluster were stratified into two strata, households in overlapping catchment areas and households in non-overlapping catchment areas, making a total of 18 strata for all nine clusters. Households in overlapping areas are served by multiple water points, while households in non-overlapping areas are served by one water point. Finally, households were randomly selected from each stratum, and 344 households were selected for all 18 strata.

Data Collection

The assessment of variation in spatial access to water points using the fuzzy-based 2SFCA method requires three types of datasets, coordinates of water points, coordinates of households, shapefile of various administrative areas, and household size.

The coordinates of each of the 70 water points were part of the variables of water points extracted from an excel file titled "Water Point Locations in Rural Water Supply 2015 – 2016", described in section 2.2. The spatial reference for these coordinates is a Geographic Coordinate System, GCS Arc 1960.

The coordinates of each of the 344 households were extracted from Google Earth through digitization. Google Earth, formerly known as Earth Viewer 3D is a virtual globe, map and Geographic Information program that was created by Keyhole, Inc, a Central Intelligence Agency (CIA) funded company acquired by Google in 2004. Google Earth, a free web-based geospatial system has an internal Geographic Coordinate System (GCS) on the WGS1984 datum and uses a Keyhole Markup Language (KML) file format to organize and display Google Earth data. Google Earth maps the Earth by superimposing satellite images, aerial photography and GIS data of varying spatial resolution onto a 3D globe, allowing users to see cities and landscapes from various angles [23]. The spatial resolution varies from one area to another basing on the types of Google Earth data available at such areas, but for most of the areas (except for some islands) it ranges between 15 cm and 15 m [24]. The process of digitization involved three steps. In the first step, the coordinate system and the file format of the shapefile of the coverage areas of 400 m around water points (i.e., the areas where households are located) were transformed from GCS Arc 1960 to WGS 1984 UTM, and from the ArcGIS file format (.shp) to Google Earth file format (.kml), respectively. The transformed file was opened in Google Earth and saved in My Places - a directory within Google Earth. In the second step, a folder named households was created to keep the digitized households. The creation of this folder was followed by the further zooming in of the Google Earth at areas of interest to load images of very high resolution (≈2.5 m), which are more clearer for the individual housing units to be easily discerned. Then, for each housing unit clearly seen on Google Earth, a placemark was added to the "households" folder, named and digitized by clicking a mouse on the respective housing unit displayed on the Google Earth map window. The digitized households are point data in WGS 1984 UTM coordinate system and Google Earth file format, .kml file format. Finally, the households point data in the .kml file format was converted into .xls file format using MyGeodata Converter, a free online geodata converter tool. The resulting excel file (.xls) consists of 344 rows and four columns Point Identifier (PID), Easting, Northing and Household Code (HH Code).

The shapefile of Tanzania regions, Dodoma districts, and Wards in Dodoma urban district were extracted from the shapefile of Tanzania wards, which was obtained from the website of the National Bureau of Statistics (NBS). The spatial reference of this shapefile is a Geographic Coordinate System, GCS Arc 1960.

Since the available data for the household size from the National Bureau of Statics (NBS) is at the Ward level and not the village or household levels, it was assumed that all households located within a certain Ward have the same household size, which equals the household size of that particular Ward. The household size of each of the nine wards where all 344 digitized households are located were extracted from the 2012 population and housing census, which was carried out by the NBS [21].

The three spatial datasets, coordinates of water points in GCS Arc 1960, coordinates of households in UTM WGS 84, and the shapefiles of administrative areas in GCS Arc 1960 were transformed to a common coordinate system, UTM Arc 1960.

Evaluating Relative Weights for each Fuzzy Set

Since the E2SFCA method requires a catchment area to be divided into several sub-catchment areas, in this study, catchment areas are divided into three zones, shorter, medium, and longer zones with distance breaks of 0m – 150m, 150m – 300m, and a 300m – 400m. Each of these fuzzy sets is assigned a weight basing on its influence on the household's water utilization or water point accessibility, which was determined using a pairwise comparison method in *MS Excel* environment. The determination of these relative weights (RWs) involved three stages, establishment of a pairwise comparison matrix, computation of relative weight for each fuzzy set, and the estimation of the consistency ratio (CR).

Stage 1: Establishment of a Pairwise Comparison matrix

In the first stage, a pairwise comparison matrix was established by comparing two fuzzy sets at a time in terms of their relative influence on the household's water utilization or water point accessibility. This comparison was based on the Saaty Rating Scale, presented in Table 1 in section 1.3. The comparison of all possible combinations of two fuzzy sets was made to establish a pairwise comparison matrix, presented in Table 3 (a). Along the diagonal of the pairwise comparison matrix, the value of the intensity of importance of *one* was assigned as two identical fuzzy sets which are of equal influence were compared.

Table 3: (a) Pairwise and (b) Normalized Pairwise Comparison Matrices

(a)

	Shorter	Medium	Longer
Shorter	1	5	9
Medium	0.2	1	3
Longer	0.11	0.33	1
Total	1.31	6.33	13

(b)

	Shorter	Medium	Longer	NPEV
Shorter	0.763	0.790	0.692	0.75
Medium	0.153	0.158	0.231	0.18
Longer	0.084	0.052	0.077	0.07
Total	1.000	1.000	1.000	1.00

Stage 2: Computation of Relative Weight for Each Fuzzy Set

In the second stage, a Normalized Principal Eigen Vector (NPEV) containing the relative weight (RW) of each fuzzy set was determined using equation 5 and involved three mathematical operations that were carried out in MS Excel. The first operation was the summation of the assigned numerical values along each column in a pairwise comparison matrix. In the second operation, each numerical value in a pairwise comparison matrix was divided by the sum of numerical values in its column to establish a normalized pairwise comparison matrix, presented in Table 3 (b). The final operation was the computation of the NPEV by averaging the normalized numerical values along each row in the normalized pairwise comparison matrix.

Stage 3: Estimation of Consistency Ratio

In the last stage, the consistency ratio (CR) was estimated using *equations 6* and *7* to measure the degree of consistency of the pairwise judgment relative to large samples of purely random judgment. An upper limit of CR equal to 10 percent has been suggested by Saaty as a measure of good consistency [25]. The estimation of the consistency ratio involved three operations, the computation of the consistency index, the selection of the random consistency index, and the computation of the consistency ratio. In the first operation, the consistency index was computed using *equation 7*, which requires the number of fuzzy sets, n, and the average value of consistency vector, λ as inputs. Since the number of fuzzy sets is known, n = 3, the only challenge in this operation was the computation of the average value of consistency vector, λ, which involved two phases. The first phase was the determination of the weighted sum vector (WSV) by multiplying the relative weight of the first fuzzy set times the numerical values of the first column in a pairwise comparison matrix, followed by relative weight of the second fuzzy set times the numerical values of the second column and the relative weight of the third fuzzy set times the numerical values of the third column, and finally adding the resulting values along each row, as presented in Table 4 (a). In the second phase, the consistency vector and its average value were determined. The consistency vector was determined by dividing the weighted sum vector (WVS) by the normalized principal eigenvector (NPEV) and the average value of the consistency vector was determined by averaging the numerical values in the consistency vector, as presented in Table 4 (b). Finally, the consistency index, CI = 0.022 was computed by plugging in the number of fuzzy sets, n = 3, and the average value of consistency vector, λ = 3.022 into *equation 7*.

Table 4: (a) The Weighted Sum Vector and (b) The Consistency Vector and it's Average

(a)

	Shorter	Medium	Longer	WSV
Shorter	0.749	0.902	0.639	2.290
Medium	0.150	0.180	0.213	0.543
Longer	0.082	0.060	0.071	0.213

(b)

	NPEV	WSV	CV
Shorter	0.75	2.290	3.059
Medium	0.18	0.543	3.010
Longer	0.07	0.213	2.998
Average Value of CV, λ			3.022

In the second operation, the random consistency index, RI = 0.58 for the number of fuzzy sets, n = 3 was extracted from a list of consistency indices for various randomly generated pairwise matrices presented in Table 2. In the final operation, the consistency ratio, CR = 0.038 (3.8%) was computed using *equation 6* by plugging in the consistency index, CI = 0.022 and the random consistency index, RI = 0.58, which were obtained in the first and the second operations. Since the consistency ratio is less than 10 per cent, the pairwise judgments were consistent, and thus, the numerical values of 0.75, 0.18, and 0.07 in the normalized principal eigenvector (NPEV) in Table 3 (b) were accepted as the relative weights for the three fuzzy sets, shorter distance, medium distance, and longer distance, respectively.

Calculating the Weighted Water Points to Population Ratio

The weighted water points to population ratio were calculated using *equation 1* in a GIS environment using *ArcMap 10* software and involved three major processes, calculation of water point capacity, calculation of weighted population, and calculation of weighted water points to population ratio.

Process 1: Calculating Water Point Capacity

In the first process, water point capacity, the numerator in *equation 1* was determined for each catchment area by summing up the capacity of each water point within that catchment area. The capacity of *one* and *zero* was assigned to each functioning and non-functioning water points, respectively. The calculation of water point capacity in each zone involved three steps. In the first steps, water points located in each zone, defined by a threshold distance of 400 m from each water point were identified using *Point Distance* tool. The resulting output of this step is a table "WaterPointsInZones". In the second step, basing on the functionality, the capacity of each water point in each zone was defined by joining "WPCapacity" field from a layer "NearWaterPoints", a copy of the layer "WaterPoints_Arc_1960" to a table "WaterPointsInZones". The joining of such a field was done using the *Join Field* tool. In the final step, the water point capacity within each zone was determined by summing up the capacity of each water point within each zone using a *Summary Statistics* tool. The final output of this operation is a table "WaterPointCapacityInZones" with a field "SUM_WPCapacity" containing the water point capacity in each zone.

Process 2: Calculating Weighted Population

In the second process, the weighted population, the denominator in *equation 1* was determined for each catchment area by summing up the product of the household's size and the weight with which a household gets service in that catchment area. The process involved two operations.

In the first operation, the weight (i.e., travel impedance) with which a household is served in a catchment area was determined using weighted mean fuzzy operation given by equation 4. This weight was determined by integrating the degrees with which each distance to water point belong to each of the three fuzzy sets, shorter, medium and longer with the respective distance breaks of 0m – 125m, 125m – 275m, and 275m – 400m. The integration of these grades was carried out in four phases. In the first phase, the distance from each water point to all households within a catchment area was determined using a *Point Distance* tool, and two spatial datasets, a shapefile "AllHouseholds_Arc_1960" with 5349 households and a shapefile "WaterPoints_Arc_1960" with 70 water points. The resulting distances are presented on the table "DistancetoHouseholdsMatrix". In the second phase, three fields "SHORTER_D", "MEDIUM_D", "LONGER_D" for the degree with which a distance belong to a shorter distance, medium distance, and longer fuzzy sets were added to a table "DistancetoHouseholdMatrix". In the third phase, the degree with which each distance belongs to each of the three fuzzy sets was determined using membership functions (i.e., equations 8, 9, and 10), implemented in the corresponding fields using the *Calculate Field* tool. These membership functions were defined such that the cross over points equally belongs to the neighbouring zones. That is, a grade of 0.5 is assigned to all households at zone boundaries, 125m (i.e., households equally belong to shorter and medium zones) and 275m (i.e., households equally belong to medium and longer zones). In the final phase, a field "D_Grade" was added to a table "DtoHHFuzzyMembership" using the *Add Field* tool, and the degree of membership with which each distance belongs to both fuzzy sets was determined using the *weighted mean* fuzzy operation implemented using the *Calculate Field* tool.

$$\mu_{Shorter}(d_{kj}) = \begin{cases} 1,\ d_{kj} \leq 50m \\ \dfrac{200m - d_{kj}}{150m},\ d_{kj} > 50m \\ 0, d_{kj} \geq 200m \end{cases} \tag{8}$$

$$\mu_{Medium}(d_{kj}) = \begin{cases} 0, d_{kj} \leq 50m \\ \dfrac{d_{kj} - 50m}{150m}, d_{kj} > 50m \\ \dfrac{350m - d_{kj}}{150m}, d_{kj} \geq 200m \\ 0, d_{kj} \geq 350m \end{cases} \tag{9}$$

$$\mu_{Longer}(d_{kj}) = \begin{cases} 0, d_{kj} \leq 200m \\ \dfrac{d_{kj} - 200m}{150m}, d_{kj} > 200m \\ 1, d_{kj} \geq 350m \end{cases} \qquad (10)$$

In the last operation, the weighted population served by each water point was determined by summing up a weighted household size of each household in a catchment area, and was carried out in three phases. In the first phase, a "HS" field containing household size in the layer "AllHouseholds_Arc_1960" was joined to a table "DtoHHFuzzyMembership" using a *Join Field* tool. In the second phase, a new field "WEIGHTEDHS" was added using *the Add Field* tool and the weighted household size was determined by multiplying the household size in the "HS" field times the corresponding weights in the field "D_Grade" using the *Calculate Field* tool. Finally, the weighted population was determined by summing up a weighted household size of each household in a catchment area using a *Summary Statistics* tool, giving out the final output in a field "SUM_WEIGHTEDHS" in the table "WeightedPopulation".

Process 3: Calculating Weighted Water Points to Population Ratio

In the third process, the Water point to population ratio was calculated using *equation 1* by dividing the water point capacity by the weighted population. This process involved five operations. In the first operation, two fields, "SUM_WPCapacity" and "SUM_WEIGHTEDHS" from two respective tables "WaterPointCapacityInZones" and "WeightedPopulation" were joined into a layer "WaterPoints_Arc_1960". The joined result was converted into a table, "WaterPointtoPopulationRatio" using *Table to Table* tool. Then a new field, "Ratio" was added to this table using the *Add Field* tool, and in this new field, the water point to population ratio was calculated using the *Calculate Field* tool. Finally, the "Ratio" field from a table "WaterPointtoPopulationRatio" was joined into a layer "WaterPoints_Arc_1960" using the *Join Field* tool to facilitate the generation of a Point Map of Water Points to Population Ratio.

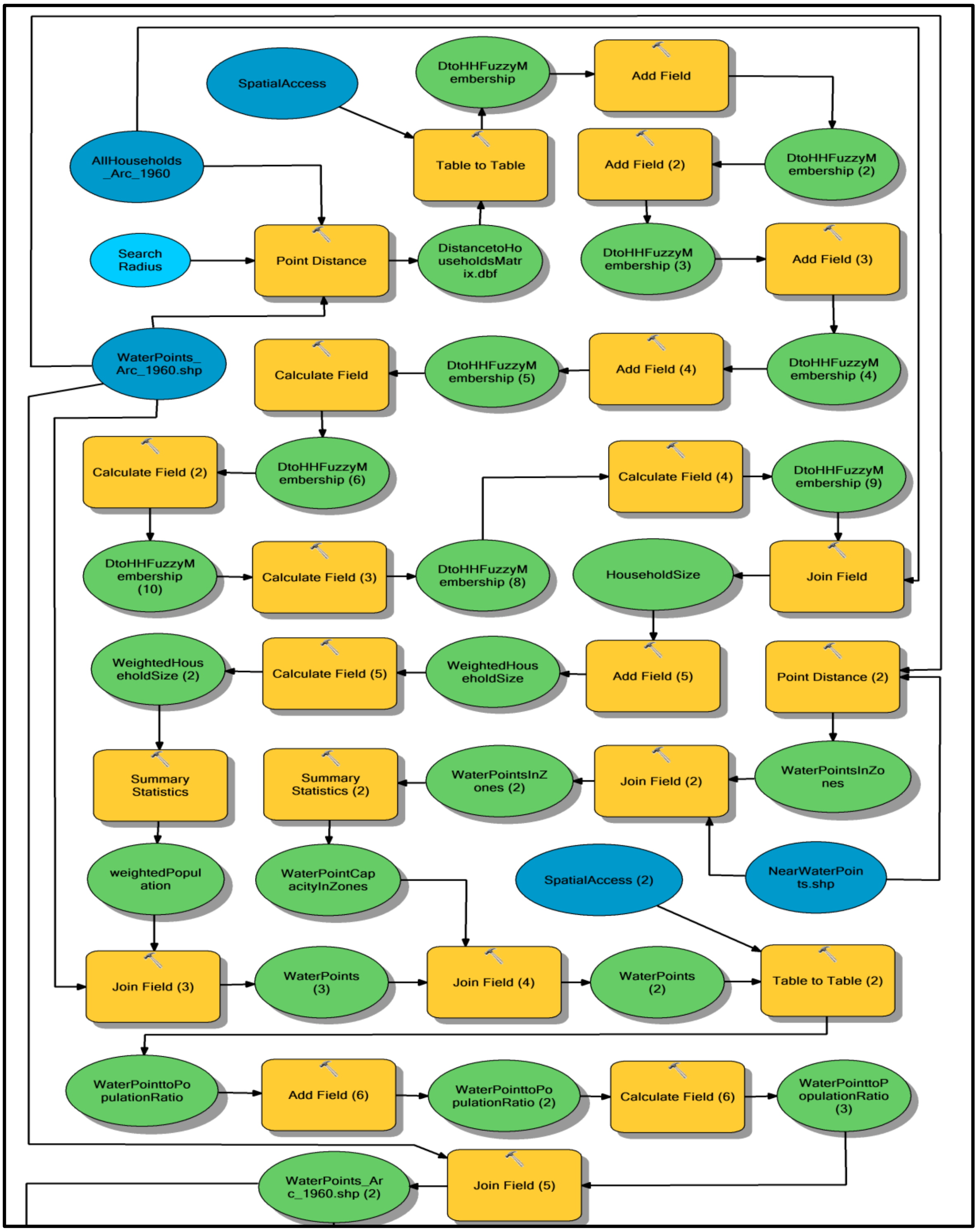

Figure 2: Fuzzy-based E2SFCA Model, Water Points to Population Ratio

Evaluating Access to Water Points

The evaluation of the household's spatial access to water points was done using *equation 2* by summing up the product of the weighted water point to population ratios times the corresponding weights of water points within a threshold distance of 400m from each household, and involved two processes, calculation of travel impedance and calculation of accessibility indices.

Process 1: Calculating Travel Impedance
In the first process, weights defined by grades with which each water point belongs to a catchment area of 400m radius from each household were determined using *weighted mean* fuzzy set operation and involved two operations. In the first operation, distance from each household to each water point located within a threshold distance of 400m was determined using a *Point Distance* tool and the resulting output was a table "DtoWPs". In the second operation, the travel impedance between each household and each water point within a threshold distance of 400m from each household was determined in the same manner as that determined in the first operation of process 2 in section 2.5 between each water point and each household within a threshold distance of 400m from each water point.

Process 2: Calculating Accessibility Indices
In the second process, four operations were carried out to calculate the household's access to water points. The first operation employed the *Join Field* tool to join the "Ratio" field from the layer "WaterPoint_Arc_1960" into a table "DtoWPsFuzzyMembership", the final output of process 1. The second operation employed the *Add Field* tool to add a new field "W_Ratio" into a table "DtoWPsFuzzyMembership", and the weighted water points to population ratio was calculated as the product of the values in the D_Grade field and those in the Ratio field using the *Calculate Field* tool. In the third operation, the *Summary Statistics* tool was employed to calculate household's access to water points by summing up the weighted water points to population ratio (i.e., values in the "W_Ratio" field) for each water point located within a threshold distance of 400m from each household. Finally, the *Join Field* tool was employed to join a field "SUM_W_Ratio" from the table "HHAccesstoWPs", the final output of process 2 into a layer "Households_Arc_1960" to allow the generation of a Point Map of Household's Access to Water Points.

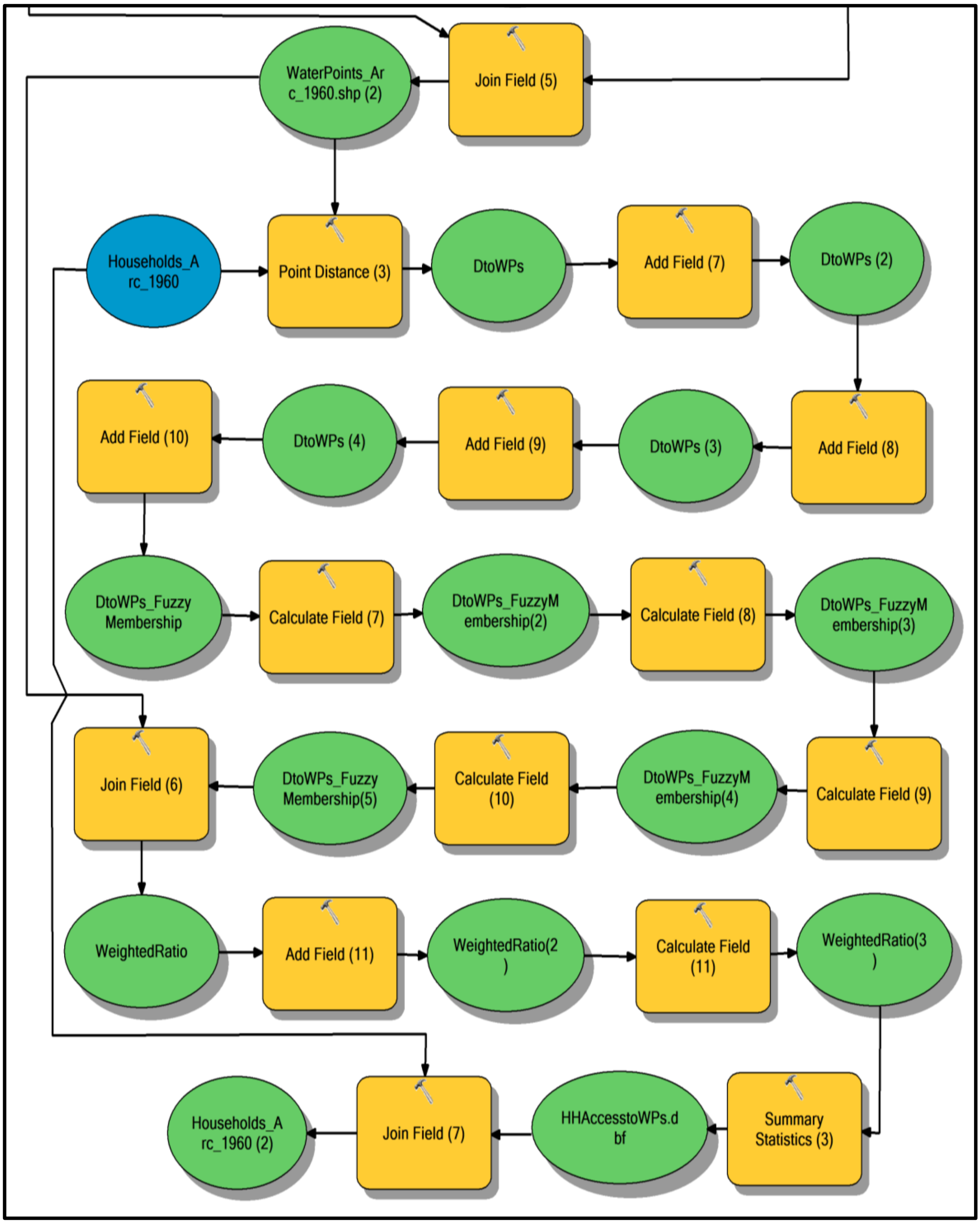

Figure 3: Fuzzy-based E2SFCA Model, Household's Access to Water Points

Results

The results of this study were generated from the final output of the Fuzzy-based E2SFCA Model, a Table "HHAccesstoWPS", and include a density map showing spatial variation in access to water points, and a vertical chart summarizing the variation in access to water points in each ward.

Spatial Variation in Access to Water Points

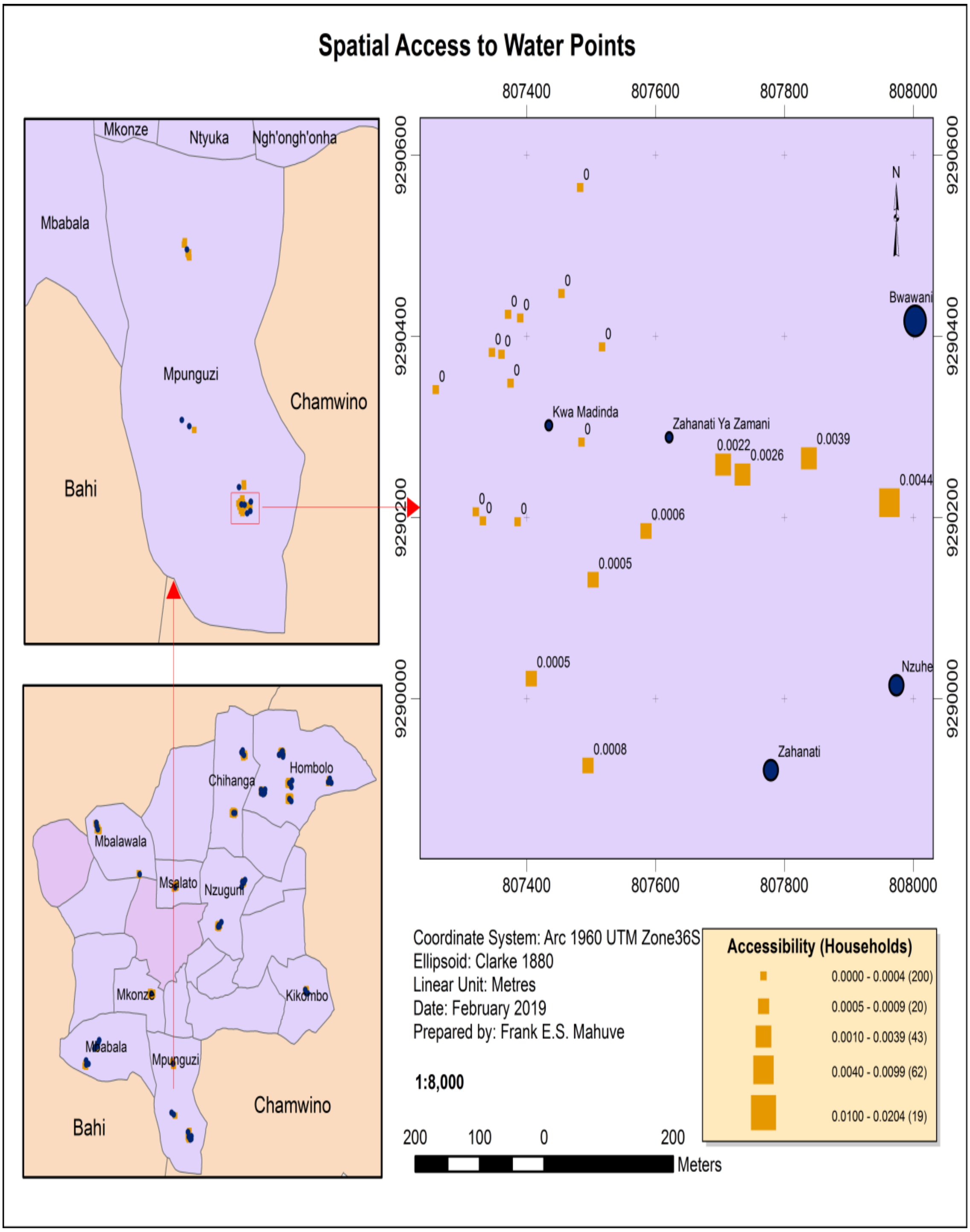

Figure 4: Spatial Variation in Access to Water Points

Variation in Access to Water Points in each Ward

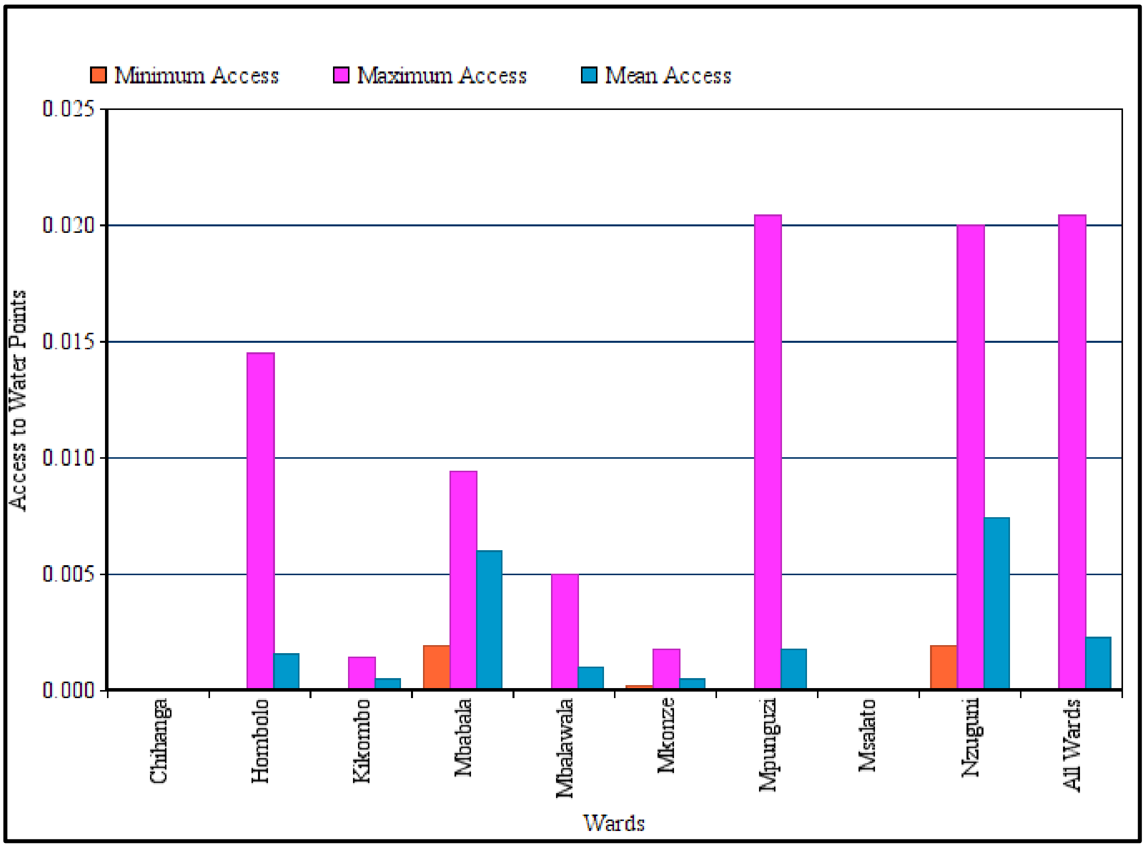

Figure 5: Variation in Access to Water Points in each Ward

Discussion

This study aimed at remedying the limitations of the conventional method, the coverage-based accessibility measure, which is often employed in evaluating access to water points. To accomplish such aim, the study adopted Luo and Qi E2SFCA method, and introduced the use of fuzzy set operation in defining travel impedance in the E2SFCA model. The resulting accessibility indices (i.e., the number of water points per person) generated from this Fuzzy-based E2SFCA model are presented in Figure 3 and Figure 4. Since in Tanzania, one water point serves a maximum of 250 people then one person must be served by 0.004 water points, and thus a household with an accessibility index of 0.004 is considered to have 100% access to water points. These accessibility indices in Figure 3 and Figure 4 vary from one location to another location within and across wards, but in some cases, these values remain uniform. For example, In Figure 3, in the selected portion of the Mpunguzi ward, accessibility index of *zero* is uniform to all households located in the northwest; while in the rest of the area accessibility indices vary from one household to another. The uniform accessibility value of *zero* in the northwest is attributed to the low values of water point capacity of *zero*, which is also apparent in the entire wards of Chihanga and Msalato presented in Figure 4. In the southwest, accessibility indices range from 0 to 0.0008, which is below 25% (i.e., below 0.0010). In the eastern part, accessibility indices range from 0.0022 to 0.0044 (i.e., 55% to 110%). This means in the eastern part households with access to water points of more than 100% are over served. In general, in all nine Wards, access to water points range from 0 to 0.0204 equivalents to 0% to 510% with a mean value of 0.0023, which is above 50%. The maximum access to water points is below 50% in four wards, Chihanga, Kikombo, Mkonze, and Msalato, and above 100% in the remaining five wards, Hombolo, Mbabala, Mbalawala, Mpunguzi, and Nzuguni. That is, all households in Chihanga, Kikombo, Mkonze, and Msalato are under-served. Also, some of the households in Hombolo, Mbabala, Mbalawala, Mpunguzi, and Nzuguni are under-served, and some are over-served. The minimum access to water points is below 50% in all nine Wards, ranging from 0% to 48%. Three Wards, Mbabala, Mkonze, and Nzuguni out of all nine wards have minimum access to water points that is above *zero*. Since accessibility indices of *zero* values are attributed by the water points of *zero* capacities, non functioning water points then households in these three Wards fetch water from at least one functioning water point. Generally, the capability of the Fuzzy-based E2SFCA method to address the problem reporting uniform household's accessibility indices regardless of their distance variation from water points, and regardless of whether they are served by a single water point, or multiple water points are apparent in this study. Thus, the Fuzzy-based E2SFCA method can be employed to assess the accessibility of various service points such as education centres, health centres, and markets in different areas, to mention few. Despite the capability and applicability of the Fuzzy-based E2SFCA method, three limitations are observed in this study. Two limitations are based on the threshold distance, and one on the weighing of the sub-catchment areas. Since the threshold distance is defined differently by different countries and organizations, this method cannot be used in comparison studies involving such countries. Second, since households are confined to water points within the threshold distance, water points beyond this distance are equally inaccessible regardless of how close they are to this distance. Third, in this study, the rating of one sub-catchment area with respect to another when generating a pairwise comparison matrix was based on the perception of the analyst, which vary from one analyst to another. Thus, this study, recommend the use of a household's utilization of water points at a different distance to generate a pairwise comparison matrix when such data are available.

Reference

1. R. Penchansky and J. W. Thomas, "The Concept of Access.pdf," Med. Care, vol. XIX, no. 2, pp. 127–140, 1981.
2.] M. F. Guagliardo, "Spatial accessibility of primary care : concepts, methods, and challenges," Int. J. Health Geogr., vol. 3, no. 3, 2004.
3. A. Jiménez and A. Pérez-Foguet, "Access to safe and year round functional water : an estimation of coverage for three central regions in Tanzania," in Water, Sanitation and Hygiene: Sustainable Development and Multisectoral Approaches, 2009.
4. United Republic of Tanzania, "National Water Policy," 2002.
5. E. A. Kosoe and I. K. Osumanu, "Water is Life: Situation Analysis of Access to Household Water Supply in the Wa Municipality, Ghana," Int. J. Environ. Prot. Policy, vol. 3, no. 1, pp. 1–13, 2015.
6. R. Ntozini et al., "Using Geographic Information Systems and Spatial Analysis Methods to Assess Household Water Access and Sanitation Coverage in the SHINE Trial," Clin. Infect. Dis., vol. 61, no. suppl 7, pp. S716–S725, 2015.
7. W. Luo and Y. Qi, "An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians," Health Place, vol. 15, pp. 1100–1107, 2009.
8. R. C. Lee, "Current Approaches to Shortage Area Designation," J. Rural Heal., vol. 7, no. 4, pp. 437–450, 1991.
9. W. R. Tobler, "A Computer Movie Simulation Urban Growth in Detroit Region," Econ. Geogr., vol. 46, pp. 234–240, 1970.
10. L.. Zadeh, "Fuzzy Sets," Inf. Control, vol. 8, pp. 338–353, 1965.
11. T. L. Saaty, "Measuring the Fuzziness of Sets," J. Cybern., vol. 4, no. 4, pp. 53–61, 1974.
12. R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," Manage. Sci., vol. 17, no. 4, p. B-141-B-164, 1970.
13. M. S. Mesgari, A. Pirmoradi, and G. R. Fallahi, "Implementation of Overlay Function Based on Fuzzy Logic in Spatial Decision Support System," World Appl. Sci. J., vol. 3, no. Supple 1, pp. 60–65, 2008.
14. P.. Burrough, "Fuzzy mathematical methods for soil survey and land evaluation," J. Soil Sci., vol. 40, pp. 477–492, 1989.
15. S. Rafati and M. Nikeghbal, "GROUNDWATER EXPLORATION USING FUZZY LOGIC APPROACH IN GIS FOR AN AREA AROUND AN ANTICLINE , FARS PROVINCE," in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, vol. XLII, no. 4/W4, pp. 441–445.
16. M. Grabisch, S. A. Orlovski, and R. R. Yager, "FUZZY AGGREGATION OF NUMERICAL PREFERENCES," in Fuzzy Sets in Decision Analysis, Operations Research and Statistics, R. Słowiński, Ed. New York: Springer Science+Business Media, 1998, pp. 31–68.
17. S. Drobne and A. Lisec, "Multi-attribute Decision Analysis in GIS: Weighted Linear Combination and Ordered Weighted Averaging," Informatica, vol. 33, pp. 459–474, 2009.
18. R. J. Eastman, W. Jin, P. A. K. Kyem, and J. Toledano, "Raster Decisions Multi-Criteria/Multi-Objective Decisions," Photogramm. Eng. Remote Sens., vol. 61, no. 5, pp. 539–547, 1995.

19. T. L. Saaty, "A Scaling Method for Priorities in Hierarchical Structures," J. Math. Psychol., vol. 15, pp. 234–281, 1977.
20. I. J. E. Zilihona, H. K. Mayaya, G. Mnyone, and J. Nkonoki, "Perceived Environmental Risks and Challenges of Urban Waste Management in Planned and Unplanned Settlements of Dodoma Municipality in Tanzania," Soc. Sci., vol. 8, no. 1, pp. 55–63, 2013.
21. National Bureau of Statistics and Office of Chief Government Statistician, "2012 POPULATION AND HOUSING CENSUS: Population Distribution by Areas," 2013.
22. United Republic of Tanzania, "Dodoma Municipal Council: Strategic Plan II for the Year 2012/13 to 2016/17," 2012.
23. N. Z. Mohammed, A. Ghazi, and H. E. Mustafa, "Positional accuracy testing of Google Earth," Int. J. Multidiscip. Sci. Eng., vol. 4, no. 6, pp. 6–9, 2013.
24. G. Pulighe, V. Baiocchi, and F. Lupia, "Horizontal accuracy assessment of very high resolution Google Earth images in the city of Rome, Italy," Int. J. Digit. Earth, vol. 9, no. 4, pp. 342–362, 2016.
25. R. Banai, "Fuzziness in Geographical Information System: contributions from the analytic hierarchy process," Int. J. Geogr. Inf. Syst., vol. 7, no. 4, pp. 315–329, 1993.

