

Article

## A Clustered, Decentralized Approach to Urban Water Management

Seneshaw Tsegaye<sup>1</sup>, Thomas M. Missimer<sup>2\*</sup>, Jong-Yeop Kim<sup>1</sup> and Jason Hock<sup>1</sup>

<sup>1</sup>Department of Environmental and Civil Engineering, U.A. Whitaker College of Engineering, Florida Gulf Coast University, 10501 FGCU Boulevard, Fort Myers, FL 33965-6565 USA

<sup>2</sup>Emergent Technologies Institute, U.A. Whitaker College of Engineering, 16301 Innovation Lane, Florida Gulf Coast University, Fort Myers, FL 33965-6565 USA

\*Correspondence: [tmissimer@fgcu.edu](mailto:tmissimer@fgcu.edu); Tel.: 239 745 4538

**Abstract:** Current models in design of urban water management systems and their corresponding infrastructure using centralized designs have commonly failed from the perspective of cost effectiveness and inability to adapt to the future changes. These challenges are driving cities towards using decentralized systems. While there is great consensus on the benefits of decentralization; currently no methods exist which guide decision-makers to define the optimal boundaries of decentralized water systems. A new clustering methodology and tool to decentralize water supply systems (WSS) into small and adaptable units is presented. The tool includes two major components: (i) minimization of the distance from source to consumer by assigning demand to the closest water source, and (ii) maximization of the intra-cluster homogeneity by defining the

cluster boundaries such that the variation in population density, land use, socio-economic level, and topography within the cluster is minimized. The methodology and tool are applied to Arua Town in Uganda. Four random cluster scenarios and a centralized system were created and compared with the optimal clustered WSS. It was observed that the operational cost of the four cluster scenarios is up to 13.9 % higher than the optimal, and the centralized system is 26.6% higher than the optimal clustered WSS, consequently verifying the efficacy of the proposed method to determine an optimal cluster boundary for WSS. In addition, optimal homogeneous clusters improve efficiency by encouraging reuse of wastewater and stormwater within a cluster and by minimizing leakage through reduced pressure variations.

**Keywords:** decentralized water systems; cluster approach; intra-cluster homogeneity; Water Supply Systems (WSS); Urban Water Systems (UWS)

## 1. Introduction

Human well-being and improvement are governed by the availability of water and energy [1]. These resources are either becoming less abundant relative to demand or are running the risk of critical scarcity in many places with recent crises occurring in Cape Town, South Africa and Chennai, India [2, 3]. Despite long- and short-term shortages, the conventional approach of designing centralized urban utility systems has been driven by 19<sup>th</sup> century technological principles, applied during a time when resources were abundant, demand was relatively small, and growth in urban areas was rapid [4]. This approach has led to a highly complex, interconnected, and dysfunctional systems that are a hindrance to efficient and sustainable resource management, particularly where growth patterns were uncontrolled and chaotic. With increasing global change pressures including population increases and climate change, there are increasing concerns about whether conventional centralized water systems will be able to manage scarcer and less reliable water resources in a cost efficient manner [5, 6]. These systems also require a huge management effort and small changes in one part of the system can cause change propagation through the whole system. For example, the rationing of water in areas with a scarce water supply is a complex problem. The problem is even greater in the case of large systems and when network distribution system pressure variation is high, and the treated water supply is intermittent. This can cause unacceptable variations in the

distribution system pressure with some cities having negative pressure during parts of high demand days (e.g., Jeddah, Saudi Arabia).

The conventional approach of design of infrastructure systems has been based on deterministic assumptions about the future despite the reality that such conditions are unknown. Urban systems face a range of future dynamic global change pressures such as population growth and urbanization, changes in socio-economic conditions, and technological developments. The current models of urban resource management and their corresponding complex infrastructure have already failed or are on the verge of collapse from the perspective of cost effectiveness, performance, and inability to adapt to the future change requirements, and loss of water-supply sources [7]. For example, the long-term planning of a large centralized water system under conditions of uncertain spatial growth of population may lead to consequences such as unnecessary investment, leading to system performance problems or other issues.

In order to cope with these challenges, various researchers have proposed decentralization of urban water systems. Decentralized water systems are small sub-systems (clusters) that have a large degree of autonomy within the overall system. According to Bieker et al. [8], future urban water systems are likely to be more decentralized than conventional systems because added water reuse systems require reducing the distance between water users and treatment locations. In addition, decentralization minimizes energy demand and infrastructure costs and maximizes the potential for recovery of heat energy if water is used close to where it is generated [8-12].

Besides efficient use and reuse of scarce and less reliable resources, there are additional reasons that support the shift from conventional centralized water systems to decentralized clustered water systems. Clustered systems could be adapted to future changes with low effort and without affecting the performance of the entire system [13, 14]. Their modular diversity exponentially increases the number of possible configurations that can be achieved for urban water systems from a given set of inputs. Decentralized/clustered water systems can be implemented in an incremental fashion, which reduces investment costs and makes the transition easier to manage [15-17]. In addition, decentralized water systems allow a staged development that traces the urban growth trajectory more closely and can be implemented more quickly than a conventional approach (as the planning and implementation process is easier to manage). According to Wang et al. [16], the gradual stepwise development of decentralized systems enables the expansion of urban water systems that follows the spatial growth, and hence, embeds flexibility to water systems. In addition, clustered systems provide a better capacity to reduce the risk associated with water-system contamination through biological or chemical ingress as well as malicious attacks with chemical, biological and/or radiological agents. This is because decentralized units are small and independent units where the effect associated with water contamination and malicious attacks will be contained within a cluster. However, in case of centralized urban water-supply systems any contaminant ingress and malicious attack could propagate to the whole systems. Internal peak demand storage of potable water, in the form of aquifer storage and recovery systems, can be integrated into decentralized systems wherein the injection and recovery wells can be placed

near ground storage tanks or pumping stations within the localized distribution and not only near the water treatment plants [18].

While there is great consensus on the benefits of decentralized water systems, one of the challenges for cities to move to a decentralized resource management is the lack of clear understanding on what clustered water systems should look like. To the best of our knowledge, no methods exist which guide planners in how to cluster/decentralize urban water systems. Thus, this paper presents a clustering methodology/tool that allows for better clustering of urban water systems into small and adaptable units. The clustering/decentralization approach proposed is based on two major optimization principles: minimization of the distance from source to consumer by assigning demand to the closest source and maximization of the homogeneity within the cluster by reducing the variation in population density, land use, socio-economic level, and topography. The efficacy of the developed clustering methodology/tool is also demonstrated in a real case study of Arua, Uganda.

## 2. Methods

### 2.1. Proposed Water Supply System Clustering Method

The proposed clustering methodology for water supply systems (WSS) is based on two major principles: (i) minimization of the distance from source to consumer, and (ii) maximization of the homogeneity within the cluster. In order to define an optimal cluster boundary that minimizes source-demand distance and maximizes the homogeneity within the

cluster, various parameters such as the location of water sources (surface water and ground water), topography (Digital elevation-DEM), spatial and temporal distribution of population, land use characteristics, and the socio-economic status of the area are considered [19]. These parameters are used to define source-demand distance, intra-cluster demand, and topographic homogeneity of the study area [20].

## 2.2. Source-Demand Distance

The sources-demand location plays an important role in reducing the transport distance of water and associated investment cost. Assigning demand to the nearest source location reduces the effort to collect and distribute water to the users. This reduces the cost of the pipe network (due to reduced pipe size/length) required and the energy needed for pumping long distances. Minimizing the transportation distance also increases the compactness of pipe and sewer networks, thereby maximizing resource conservation and minimizing losses (i.e. leakage) [21]. In addition, it improves the potential to reuse and recycle wastewater to the proximity within the cluster.

## 2.3. Intra-cluster Demand and Topographic Homogeneity

Understanding the water consumption and topography characteristics is extremely important for optimization of investment, operation costs and maximization of resource efficiency. Traditionally, analyses were performed for large regions which involved a variety of

topography, land uses, and associated demand characteristics. However, with the advent of clustering, study of the behavior of smaller areas has become necessary to allow for the creation of uniformity within the clusters. The population distribution, land use and socio-economic parameters are aggregated into a spatio-temporal demand distribution of the area. Intra-cluster demand homogeneity is used as one of the parameters to minimize the effort required to move water and wastewater. Intra-cluster homogeneity is the measure of the similarities or dissimilarities between parcels of the same cluster. Clustering of large and small demand areas together involves huge variations in consumption which can cause larger pressure fluctuations than areas with similar demand distribution. This causes additional efforts to supply and manage water and wastewater in the area. For example, areas with urban agriculture have different demand patterns than industrial or residential areas. Thus, maximizing the similarities by clustering residential and agricultural areas separately will improve the required efforts as compared to if they were clustered together. The clustering of different land uses into unique clusters will ensure multiple uses of water by cascading it from higher to lower-quality needs and through reclamation treatments for a return to the supply side of the other cluster. Water used by residential clusters can be re-used by industrial or agricultural clusters. Demand based clustering also improves the ability to implement relevant technology (i.e. water treatment and wastewater reuse recycling schemes) within a homogeneous cluster. This also allows better control of small and homogeneous cluster units.

Topography is the other major factor which affects the flow of water and wastewater. Areas with similar topographic characteristics (minimum differential relief) reduces cost associated with infrastructure and pumping of water and wastewater in the area. However, large



variations in topography increases the effort required to collect and supply water, discharge wastewater, and develop reuse systems. For example, water supply systems in areas with large topographic variations cause large pressure fluctuations and require a large amount of energy for pumping, as well as a large system capacity to satisfy the required level of service. Thus, partitioning the urban area based on improved intra-cluster topographic homogeneity will reduce the costs associated with water system investment and operation (energy). It allows for improved resource efficiency by encouraging reuse and recycling of wastewater within the cluster and by minimizing leakage (distribution system water loss) through reduced pressure variations.

The methodology uses two major steps for clustering water-supply systems (WSS); minimization of source-demand distance and maximization of intra-cluster homogeneity. The details of the proposed steps are shown in Figure 1. The starting point of the proposed method to cluster WSS is to consider all the input parameters of the study areas. This involves the location of water sources (surface water, groundwater, and stormwater collection points), topography, spatio-temporal population growth and associated demand, land use characteristics and socio-economic status. Thus, the proposed clustering method minimizes the source-demand distance by assigning demand to the source. The Euclidean norm minimization approach is used to minimize source-demand distance. This method also maximizes the homogeneity within the cluster so that source-demand distance, topography, and demand variations are minimized. A K-means algorithm is applied to maximize intra-cluster homogeneity [20].

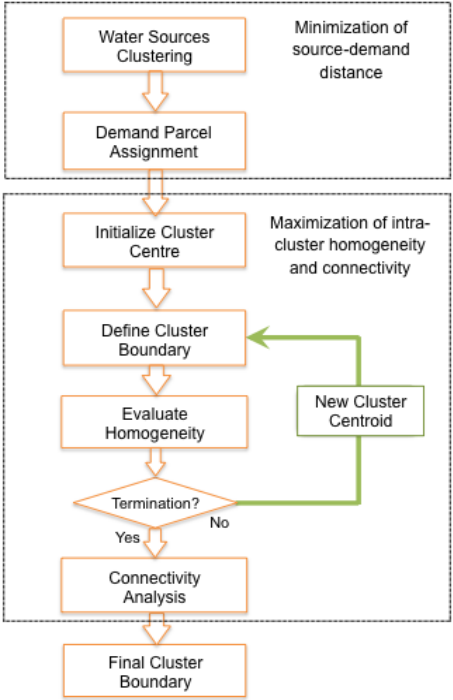


Figure 1. The proposed method for clustering WSS.

#### 2.4. Minimization of Source-Demand Distance

The first part of the proposed clustering method involves prior grouping of spatially distributed available water sources, such as surface and groundwater sources, stormwater, and greywater. This method involves grouping water sources and determining their group center such that the effort required for collection is minimized. Then each demand parcel is assigned to one source group center such that the distance between source and demand parcel (grid cell) is minimized. Grid parcels are square cells characterized by attributes of spatial location (X and Y coordinates), elevation, and demand. The source-demand distance for each parcel depends on the specified source center locations. Euclidean norm minimization is used to optimize the source-demand distance for all clusters. The formulation is done as a demand

assignment problem where each parcel is assigned to the nearest source. Then parcel membership is determined from the minimization process.

The determination of the optimal number of source centers is not the focus of this paper. The number of clusters for the area can be determined from the required size of a cluster. According to Bieker et al. [8], the size of a cluster must be guided by the principle “as small as possible, as big as necessary” to achieve the ecological, economic, and social interest. BMBF [22] compared different scales for areas which range from 10,000 up to more than 200,000 people and proposed a recommended size ranging from 50,000 to 100,000 people as a suitable scale for an integrated decentralized system for fast growing urban areas. Bieker et al. [8] argued that this scale offers huge opportunity in recovering heat from wastewater streams as the transport distance is short. The size of a cluster could be used to pre-determine an initial number of clusters or source groups and could be changed during the process of clustering. For example, considering an area with 8 water sources and 121 demand parcels (each representing a 100 m by 100 m area), the pictorial representation of the demand parcel assignment to the nearest source center is shown in Figure 2.

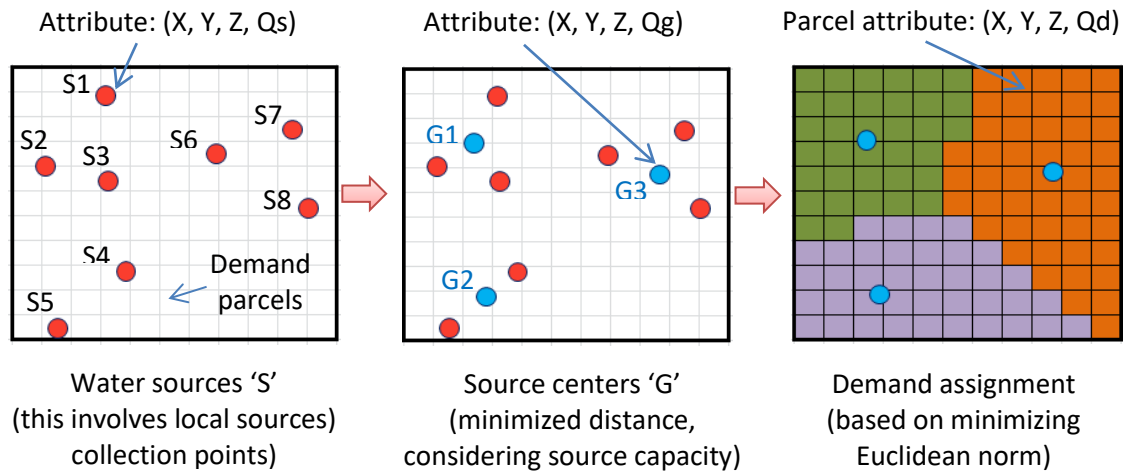


Figure 2. Assignment of parcels to the source center (X, and Y are location parameters, Z is elevation above sea level (asl), Q<sub>d</sub> is parcel demand, Q<sub>s</sub> and Q<sub>g</sub> are capacity of local water sources and group source respectively).

### 2.5. Identification of Source Centers: Water Source Clustering

Once the groups of sources are identified, a simple source center calculation is carried out to determine the centroid of the sources within the same group. Taking a similar approach as in determining a mass center, the source center is calculated using Eq. (1).

$$D_c = \frac{\sum_{s=1}^S D_i * Q_i}{\sum_{s=1}^S Q_i} \quad (1)$$

Where  $D_c$  is source center,  $Q_i$  and  $D_i$  are the supply capacity of the source and the distance from reference water source.

## 2.6 Assignment of Demand Parcel to the Nearest Source

Source allocation is a demand assignment problem where demand parcels are assigned to the nearest source center. The method employs a minimization of the sum of Euclidean norms within the cluster. Some researchers have proposed minimizing the sum of Euclidean distance for shortest-path optimization [23]. The theories and algorithms for minimizing Euclidean distance can be applied to many optimization problems. In this study, the sum of Euclidean norms is used to determine the membership of parcels based on the shortest distance to the water source centers. The same membership is given to the parcels that are assigned to the same source center. This increases the compactness [21] and reduces the cost of pipe networks and the energy needed for pumping long distances. Compacted networks with closer proximity also increase resource efficiency by reducing leakage that would be higher in large centralized systems. Given a set of parcels (representing the study area) with dimension vector  $P = \{P_1, P_2, \dots, P_n\}$ ,  $P \in \mathbb{R}^N$  Euclidean norm defines,  $\|P\| = (P * P)^{\frac{1}{2}}$ , if  $N = 1$  then  $\|P\| = |P|$ , the absolute value of  $P$ .  $\|P\|$  is the Euclidean norm of  $P$  that is used to measure the distance between points [24]. For example, suppose  $P = (X, Y) \in \mathbb{R}^2$  and the source centers are defined by  $C = (X_1, Y_1) \in \mathbb{R}^2$ . Then the shortest distance from the source to the parcel is determined using Eq. (2).

$$\min \|P\| = \sqrt{(X_1 - X)^2 + (Y_1 - Y)^2} \quad (2)$$

Given the Euclidean norm of each parcel (from each source center), distance minimization is performed using Eq. (3). Then each parcel will have membership (to the source center) based on the minimization of Euclidean norms. The membership defines grouping of similar parcels which are assigned to the same source center. The Euclidean norm minimization algorithm is shown in Figure 3.

$$\min d_{(P, P^c)} = \sum_{k=1}^c \|P\| = \sum_{k=1}^c \sqrt{\sum_{j=1}^n (P_j - P_j^c)^2} \quad (3)$$

Where  $d_{(P, P^c)}$  is the Euclidean norm from the source centers,  $\|P\|_{min}$  is the minimum Euclidean norm of each parcel from source centers,  $P$  is an attribute which is described by parameters where the variation needs to be minimized (i.e location and elevation parameters).

The movement of water is based on an absolute distance which depends on the link (pipe) layout and pressure distribution; this requires hydraulic simulation of the whole network. However, to simplify the clustering process, in this study the minimization of the Euclidean norm is employed by using the relative distance based on the coordinate of demand parcels and supply centers. Once the parcels are assigned to the source center by the minimizing Euclidean norm principle, the membership values are used in the maximization of cluster homogeneity.

### Minimizing Euclidean norm algorithm

- (1) For the given  $C$  source centers, the Euclidean norm of a parcel is determined with respect to their parameter  $P = \{P_1, P_2, \dots, P_n\}$ , yielding the distances  $d_{(p, pc)}$ .
- (2) Given the set of Euclidean norms  $\{d_1, d_2, \dots, d_c\}$  for each parcel, the total cluster Euclidean norm is minimized by assigning a parcel to the nearest source center.
- (3) Steps i and ii are repeated until all parcels are assigned to the closest source center (then a membership will be assigned to each parcel based on the source center to which they belong).

Figure 3. Basic minimizing Euclidean norm algorithm

### 2.7. Maximization Intra-cluster Homogeneity and Connectivity

Traditionally, the design of WSS has been performed for large spatial extent areas which involve a variety of topography, population distribution, land use, socio-economic, and associated demands. However, with the advent of decentralization, the study of the behavior of smaller areas has become a necessity so as to allow for uniformity within the clusters. In this section, clustering involving the maximization of intra-cluster homogeneity and connectivity analysis is used. Intra-cluster homogeneity is used to measure the similarity or dissimilarity between parcels of the same cluster. Maximization of intra-cluster homogeneity allows clustering the parcels so that parcel attributes within a cluster are closely related to one another [20]. Three major parameters are considered in the clustering process. These are membership (determined by Euclidean norm minimization), topography (elevation of the parcels), and spatio-temporal demand distribution (determined from the population

distribution, land use, and socio-economic parameters). The clustering process involves the grouping of similar parcels. An inter-cluster homogeneity is used as a measure of similarity between parcels and a K-means optimization technique is employed to maximize intra-cluster homogeneity by minimizing the total cluster variance with respect to the mean value. In addition, a connectivity analysis is proposed to ensure the linkage of parcels within clusters. The details of the proposed steps are discussed in the subsections.

## 2.8. K-means for Clustering WSS

“K-means clustering is a method of cluster analysis which aims to partition  $n$  observations into  $K$  clusters in which each observation belongs to the cluster with the nearest mean.” [25]. It is an evolutionary algorithm that minimizes the proximity to the mean of the cluster [25]. The name K-means comes from its method of operation in which it assigns observation on  $K$  clusters based on the observation proximity to the mean of the cluster. The squared Euclidean norm is used as a measure of homogeneity. A K-means algorithm is a commonly employed method that converges to a local optimum value for clustering. It is very popular because it is computationally fast and memory efficient. A K-means algorithm is used herein to cluster the WSS based on the principle of minimizing the dissimilarity of the three parameters: source-demand distance, topography, and demand within the cluster. Unlike topography and demand, the distance parameter is dependent on the source centers; thus, distance related membership value (discussed earlier) is used to identify to which source center each parcel is assigned.



Given a set of parcels  $p$  representing the study area  $\{X_1, X_2, \dots, X_p\}$ , where each parcel has  $n$ -dimensions (i.e. topography, elevation),  $K$ -means clustering aims to partition the parcels ( $p$ ) into  $K$  clusters ( $K \leq p$ ) within an assigned data-set  $S \{S_1, S_2, \dots, S_k\}$ . For the given cluster assignment  $A$  that involves  $K$  groups, the total cluster variance is minimized through minimization of the sum of the squares of Euclidean norm for all clusters using Eq. (4).

$$A = \arg \min_S \sum_{i=1}^K \sum_{X_j \in S_i} \|X_j - \mu_i\|^2 \quad (4)$$

$$\mu_i = \frac{1}{N_i} \sum_{X_j \in S_i} X_j \quad (5)$$

Where  $A$  is cluster assignment,  $K$  is the number of clusters,  $N_i$  is the number data-set assigned to  $S_i$ ,  $\mu_i$  is mean of parcels in cluster  $S_i$  and is calculated using Eq. (5) [26].

A  $K$ -means algorithm achieves optimal clustering assigning parcels so that the difference between parameters of the parcels and their centroids are as small as possible. Maximizing intra-cluster homogeneity for WSS involves several steps (as shown in Figure 1). Firstly, the optimizer selects initial cluster centroids (means) randomly. Secondly, an initial cluster boundary is defined by assigning demand parcels to the initial cluster centroids. Thirdly, the  $K$ -means optimization evaluates the difference between parameters of the parcels and their centroids (this is used as a measure of homogeneity). Fourthly, the optimization uses the homogeneity as a termination criterion. It uses an iteration based evolutionary optimization

which involves the assignment of parcels to the closest mean and calculating a new mean until the assignment no longer changes (means no improvement in homogeneity). Thus, the simulation stops. Otherwise the above steps repeat until there is no change in parcel assignment between subsequent simulations. Figure 4 shows the basic K-means algorithm used in clustering WSS.

#### Basic K-means Algorithm

- (1) Initialization of K means  $\{\mu_1, \mu_2, \dots, \mu_k\}$  where each mean is defined by d-dimension vector (n-parameters)
- (2) Given an initial set of K means, the algorithm assigns parcels to the closest mean so that the total variance is minimized with respect to the mean
- (3) Calculate a new mean to be the centroid of the cluster
- (4) Repeat steps (1) and (3) until the assignments do not change

Figure 4. Basic K-means Algorithm

Though the K-means algorithm maximizes the similarity within a cluster, it has some shortcomings. One of the limitations is that it does not consider the geospatial relative location of different neighboring parcels. The specific problem of clustering water systems requires the ability to handle not only the spatial extent, but also the geographic component with respect to neighboring parcels (i.e. the need to have the same membership parcels in the same spatial location). To avoid the possibility of detaching parcels of the same cluster in different spatial locations, intra-cluster parcel connectivity is used.

## 2.9. Intra-cluster Parcels Connectivity

Intra-cluster parcel connectivity, defined as the linkage of a parcel within a cluster, is used to check whether a parcel of one cluster is located in another cluster. Given the membership of parcel “P” defined as  $P_{(m,n)}$  and neighbor parcels as  $P_{(n\pm1,m\pm1)}$ , if parcel  $P_{(m,n)}$  of one cluster neighbors two or more parcels from another cluster, and has only one neighbor from its own cluster, the evaluation of the minimum Euclidean norm of the parcel  $P_{(m,n)}$  is performed with respect to the neighboring cluster centroid and is re-assigned to the closest one. In addition, the periphery parcels, which do not have many neighbors, are merged to the nearest cluster group in case they belong to another cluster. This connectivity analysis alone does not guarantee the existence of cluster members in another spatial location. One can use the smallest recommended size of cluster and/or the smallest demand that a cluster should supply to decide on merging isolated parcels to the neighboring cluster. An isolated parcel group will be kept as an independent cluster if the demand it supplies is greater than the required minimum size/demand within the cluster. However, a parcel group that does not satisfy the mentioned condition will be merged to the neighbor cluster. The decision of which cluster to combine is made by evaluating the minimum Euclidean norm value with respect to the centroid of neighboring clusters.

### 3. Results of Cluster Analysis Application to Arua, Uganda

#### 3.1 General Description of the Area

The developed clustering methodology was applied to a real case study in Arua, Uganda. Arua is located in the Northern Region of Uganda and lies between latitude 20° 30' N and 30° 50' N and longitude 30° 030' E and 31° 030' E (Figure 5). The Arua municipality is one of the fastest growing municipalities in the country. According to the statistical abstracts of the Uganda Bureau of Statistics [27], the population of the Arua municipality was 59,400 in 2011, with the population around the periphery of the municipality reported to be 49,893. With an annual growth rate of 3.4%, the total population in 2032 is estimated to be 220,887. The predicted spatial extent of Arua in 2032 is shown in Figure 6.

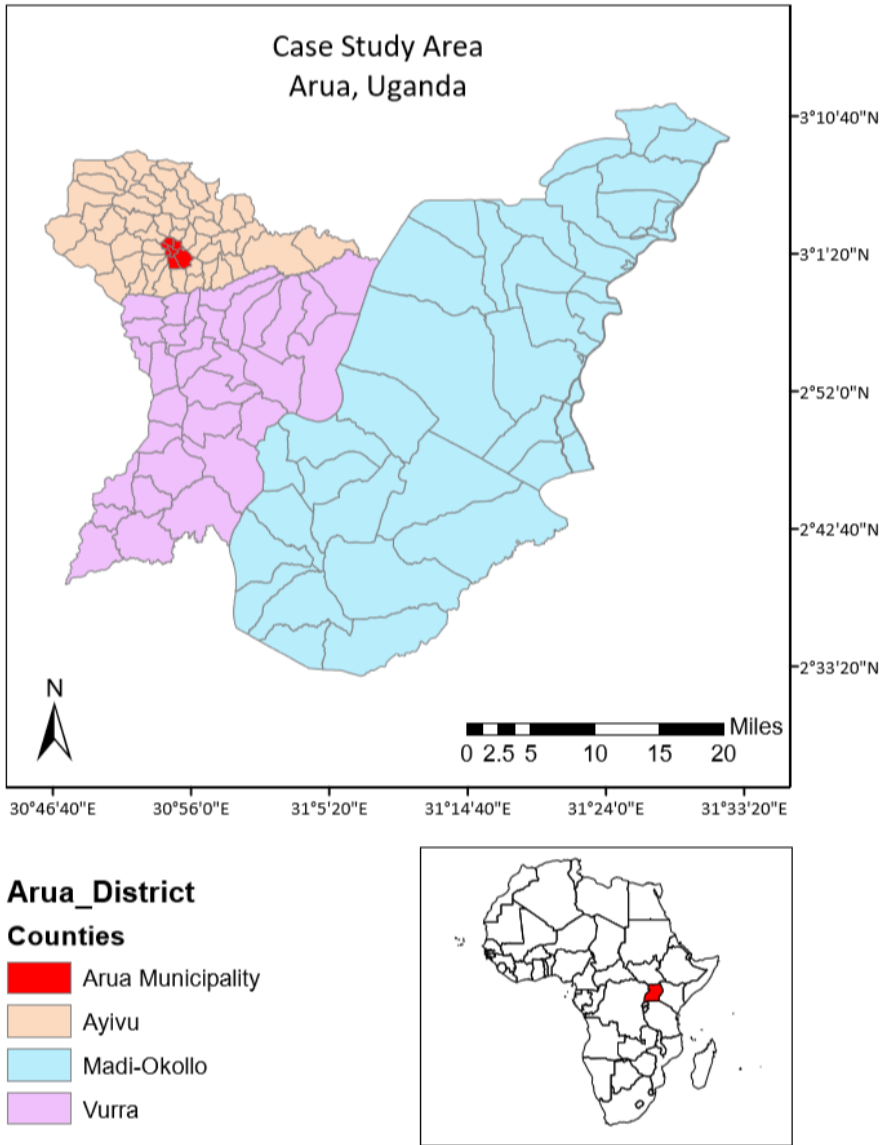


Figure 5. Geographic location of Arua, Uganda

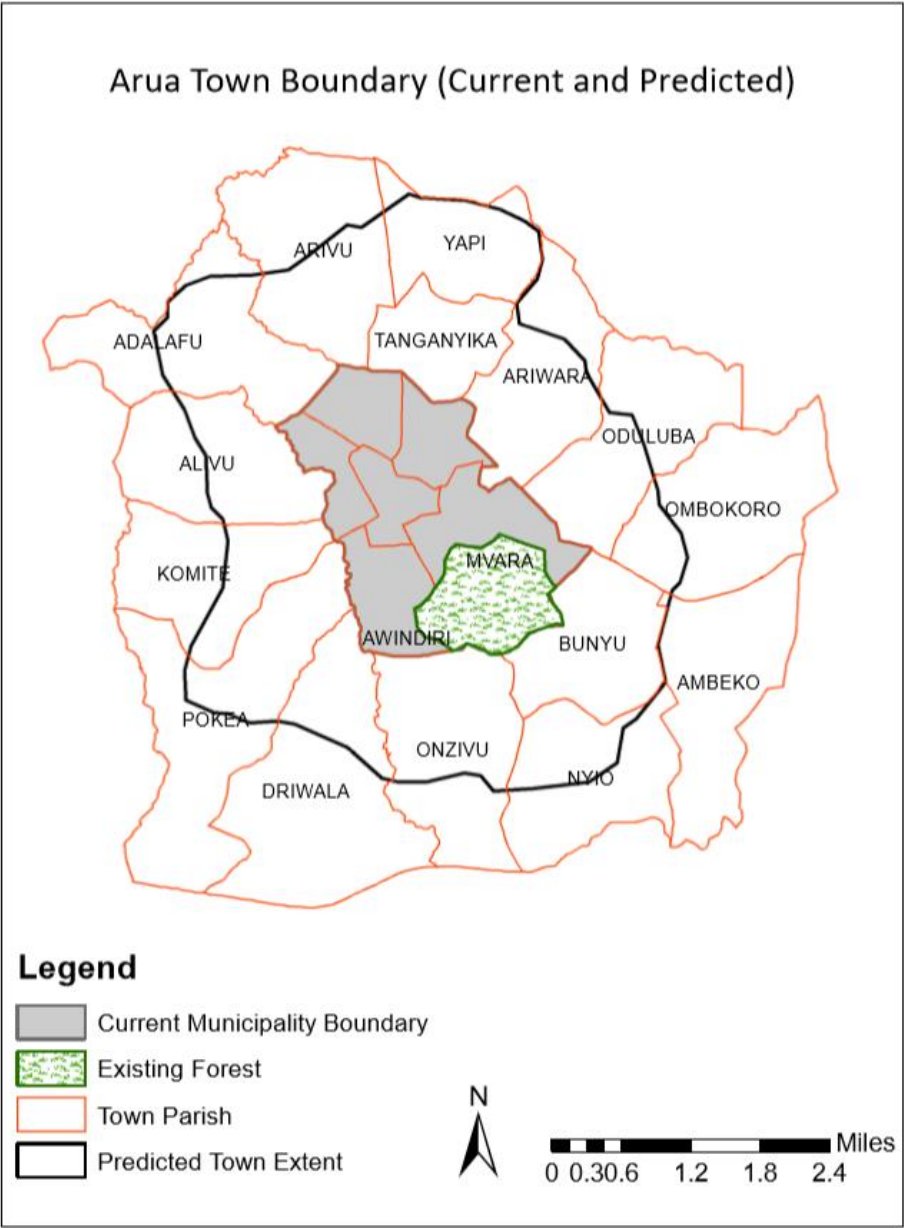


Figure 6. Predicted spatial extent of Arua in year 2032

The town of Arua is experiencing a critical shortage of water because it depends on only a small river (Enyau River) for its supply [28]. With population growth and increasing wealth it is predicted that the water demand will likewise rise to 17,217 m<sup>3</sup>/d in the year 2032, which

would worsen the water shortage. This predicted future demand takes into consideration the different population density and socio-economic status of each of the parish areas.

The current approach to water management in Arua is based on a conventional centralized approach where water is collected upstream, used, and discharged downstream and does not encourage the use of local sources such as groundwater, stormwater harvesting, or wastewater reuse and recycling. It has become obvious that the current practices of urban water management are not sustainable to meet the future challenges in Arua. However, the rapid urban growth in emerging areas coupled with the fact that those emerging areas do not have mature infrastructure and urban planning for the area has not yet occurred means that there are real opportunities to implement a clustered urban water system management system in Arua.

### *3.2. Application of the Proposed Clustering Method and Results*

One of the major initiatives of the Arua municipality is to degazette (repurpose) the forest area (called Barifa) in a 5-year time period and incorporate it into the central business district of the town. Since the forest area has a predefined boundary, the clustering processes in this study can isolate this area and consider it to be a pre-clustered unit. Additionally, prior to the clustering process, a decoupling of the existing central WSS from the emerging areas was performed by identifying the existing municipality boundary (Figure 6). Then the proposed WSS clustering technique which minimizes the source-demand distance and maximizes intra-cluster homogeneity was applied.

### 3.3. Source-Demand Distance minimization

In this case study, 10 groundwater sources and 4 potential surface-water abstraction locations were identified (Figure 7). Once the capacity and locations of available sources were identified, the water sources were merged into groups such that the distance between grouped sources was minimized. The study area was discretized into small parcels of the size 150 m by 150 m. The available water-source abstraction locations of the area were aggregated into seven groups. The decision to propose a number of groups might depend on the size of the area, the size of clusters required, the numbers of water source locations available, etc. Different researchers have highlighted the need for case-by-case analysis to determine the population number that should be supplied by a single source to determine the smaller cluster size [8, 22]. However, the determination of the number of groups required is not the focus of this paper. Thus, the minimum cluster size with a population of 10,000 was used in decentralizing the emerging area as suggested by Webster [5] to determine the number of source centers for grouping. The evaluation of the distance between sources was performed using Eq. (2). The output of source-group identification process is shown in Figure 7(a) and (b). Once the groups were identified the X, Y coordinate and supply capacity  $Q_s$  were used to calculate source-centers. The source and source-center information is summarized in Table 1.



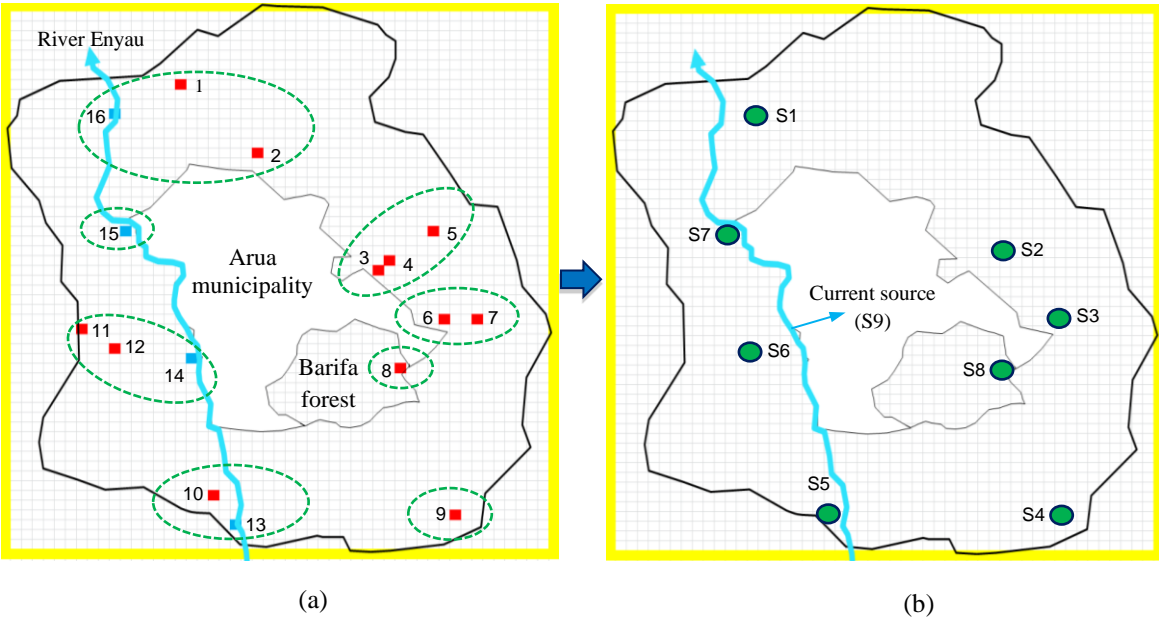


Figure 7. (a) Available water sources and their groups; (b) Water-source centers (based on minimized Euclidean distance)

Table 1. Source groups and location of source centers

Source Group	Source no.	Water source-center location	
		X (m)	Y (m)
1	1	1992	6772
	2		
	16		
2	3	5461	4650
	4		
	5		

3	6	6181	3600
	7		
4	9	6150	600
5	10	3110	510
	13		
6	11	2062	3108
	12		
	14		
7	15	1650	4950
Forest (8)	8	5400	2850
Municipality (9)		2400	3150

430

431 Once a source center was identified, the discretized square parcels (150 m by 150 m) were  
 432 assigned to the source centers. Each parcel has a location, topography, and demand attribute.  
 433 This stage used the location attribute (X, Y) coordinate of parcels and the centroid of the  
 434 available sources as inputs to minimize the source-demand location for each parcel. This case  
 435 study treats the forest areas (planned development) and the central municipality boundary as  
 436 an independent unit clusters wherein the boundary and inbound sources are pre-identified  
 437 prior to the clustering process. Eq. (2) is applied to each parcel of the emerging areas to  
 438 determine the Euclidean norm from the 7 source centers in the emerging area. Given the  
 439 Euclidean norm of each parcel (from the 7 source centers), the distance minimization was

performed using Eq. (3). Then, each parcel was assigned a membership value. Figure 8(a, b) shows the output of parcels assigned to the nearest source and the membership respectively using Euclidian norm minimization. The membership defines groupings of similar parcels which are apportioned to the same source center.

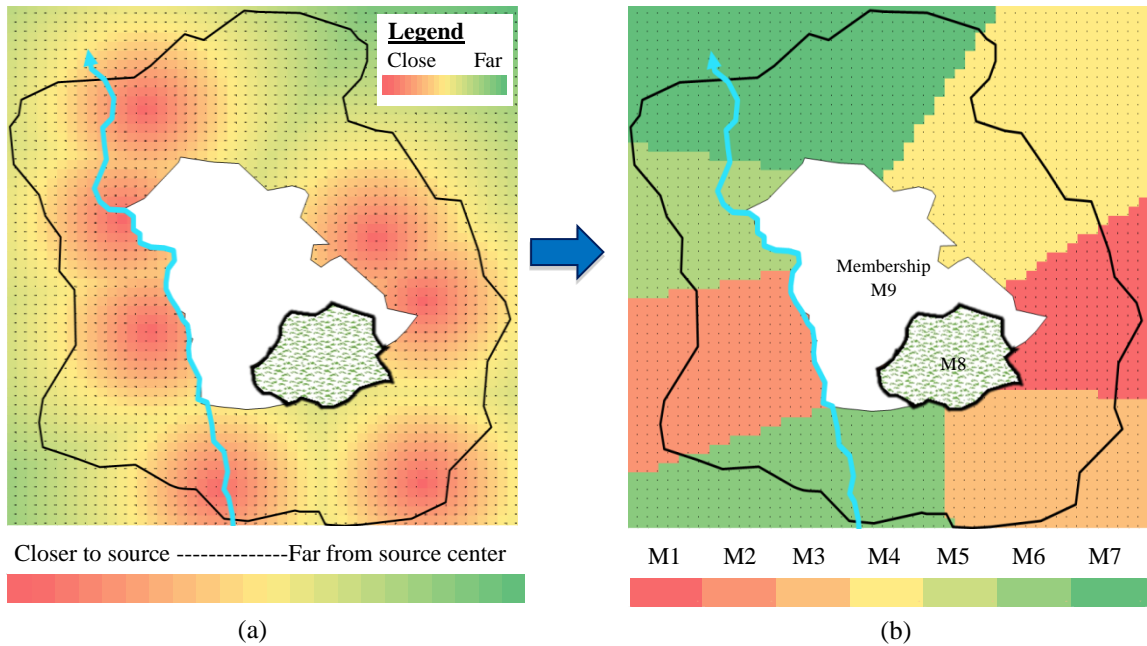


Figure 8. Maps showing parcel assignment (Minimized Euclidean norm) (a) and parcel membership-M based on source-demand distance (b)

The above clustering shown in Figure 8 is purely based on distance and does not include demand and topographic parameters. The next stage incorporates these parameters in addition to a membership value using intra-cluster homogeneity maximization.

3.4. Maximizing Homogeneity and Connectivity Analysis

Homogeneity maximization was applied to determine the final cluster boundaries for the study area. The distance-based membership value (determined in the distance minimization stage), topographic, and demand information were used as input parameters. The study area topography ranges from 1,160 m to 1,240 m asl, and the determination of demand was performed using the population, socio-economic status, and land use information. The input elevation and demand information are plotted for the case study area and shown in Figure 9(a, b).

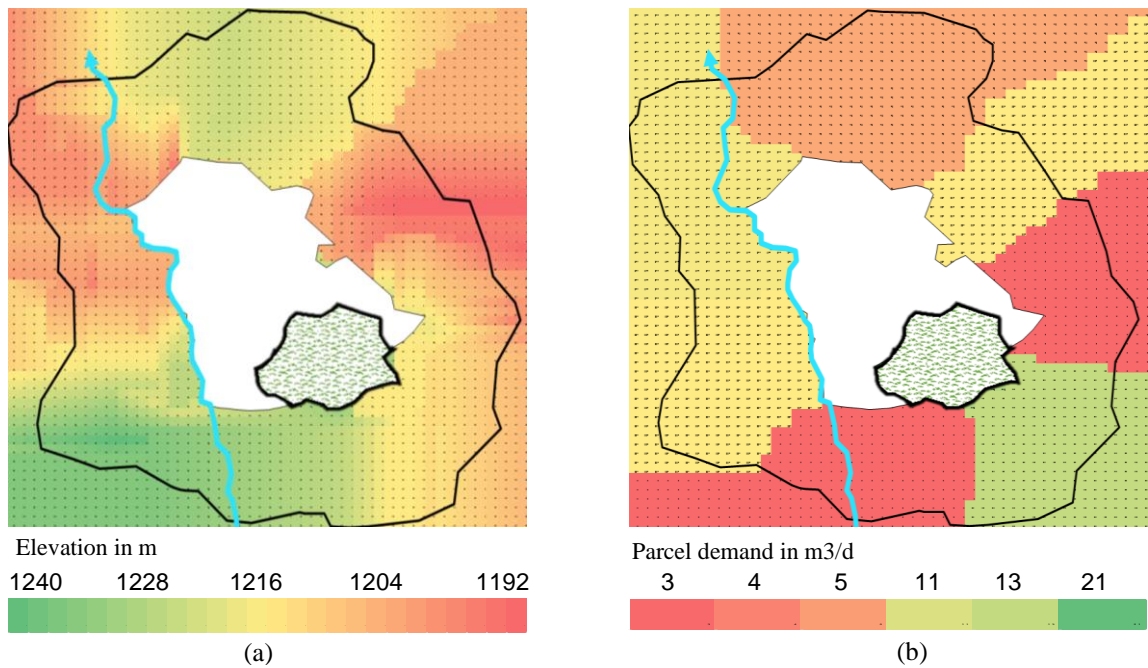


Figure 9. Maps showing elevation in m (asl) in the study area (a) and parcel water demand in m³/d (b)

Given the input parameters, a K-means algorithm was applied to maximize the intra-cluster homogeneity. Multiple runs of the K-means simulation were performed to avoid the problem associated with initialization, and the algorithm showed similar clusters. The final output of the clusters is shown in Figure 10(a).

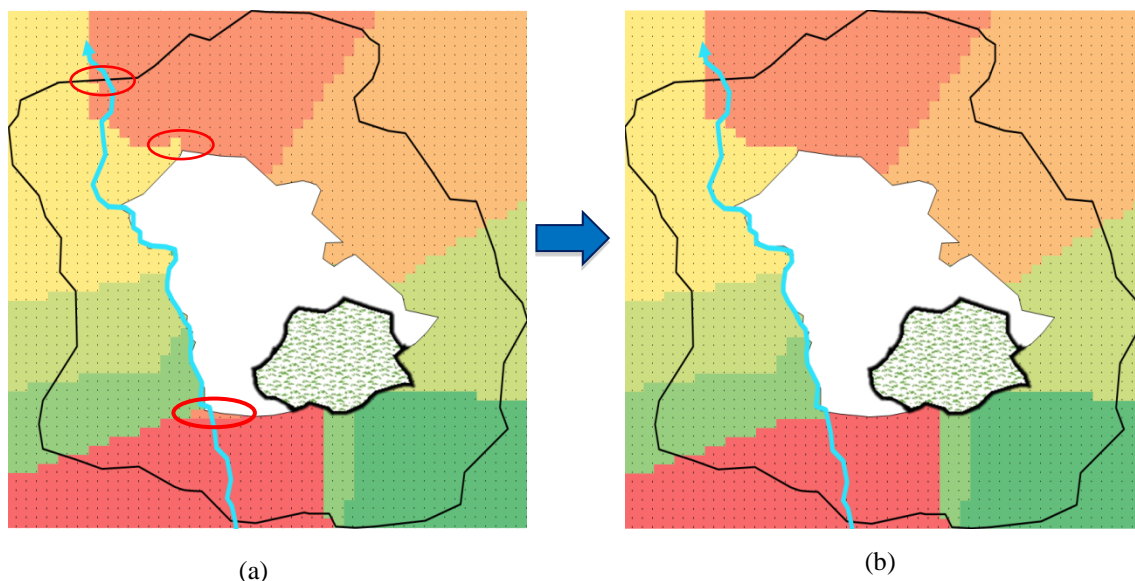


Figure 10. Maps showing (a) initial K-means clusters and (b) and clusters after merging isolated parcels

As shown in Figure 10(a), the same members could be in different spatial locations (see circled areas). To incorporate the spatial component of cluster location, intra-cluster parcel connectivity was applied. The parcel connectivity involves refining the boundary and merging parcels of one cluster which are located in a different cluster. First, simple parcel connectivity was done by considering the membership of each parcel. In this case study, parcels circled in red in Figure 10(a) are merged to their neighbors as shown in Figure 10(b). Secondly, if there is

a parcel group which is in another cluster, the size was used to decide whether to keep the group as a new independent cluster or to merge it with the nearest cluster. A group merging was performed if a cluster/group was too small. Groups with a size less than 20% of the maximum cluster size were distributed to the neighboring cluster to avoid large variation in cluster size. However, a recommended cluster size and/or the smallest demand that a cluster should supply were used to decide whether to merge isolated parcels. Figure 11(a) shows the final cluster boundaries after isolated neighboring parcels were re-distributed, and the final cluster boundary for the case study area is shown in Figure 11(b).

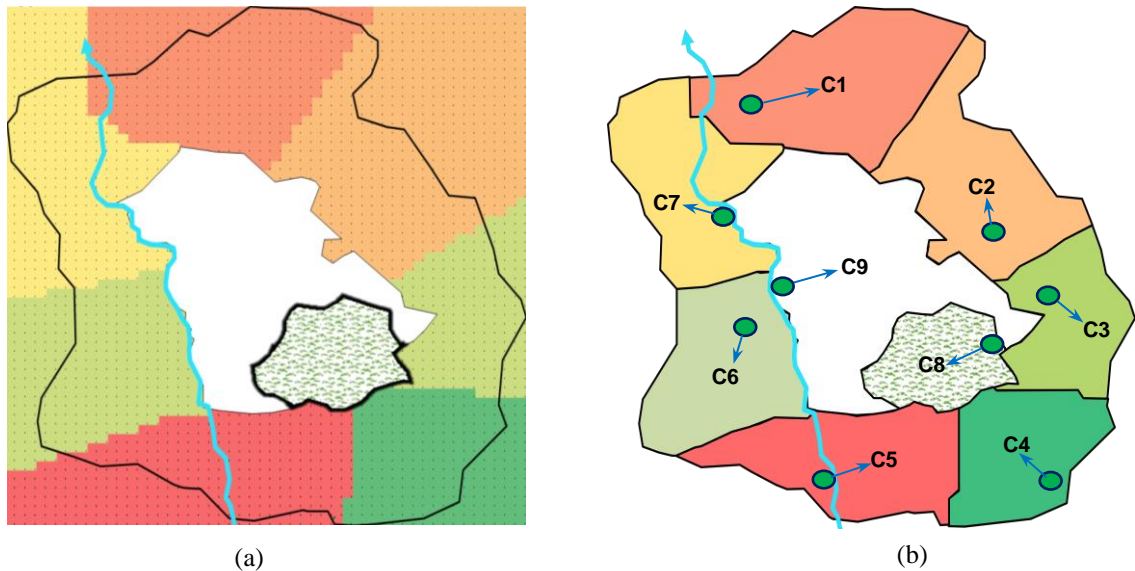


Figure 11. Maps showing clusters after re-distributing small groups (a) and WSS cluster boundaries with source centers (green circles) (b)

The developed clustering method offers an adequate solution to the decentralization paradigm through clusters that allow for improving the movement of water and wastewater in

the area. It divides urban area into clusters to allow for the provision of flexible, modular decentralized WSS.

## 4. Discussion

### 4.1. Performance Analysis

The proposed WSS clusters accomplishes two principles: (i) minimization of the distance from source to consumer, and (ii) maximization of the intra-cluster homogeneity. Testing to ascertain if these principles favorably impact cost is necessary to verify that the clustering method indeed produces an optimized WSS. Cost efficiency can be measured in a variety of ways, however, in this research a parcel-based power usage analysis was utilized. Several WSS are required for a suitable analysis; by assuming the same relative size and number of clusters, random WSS scenarios can be designed. By choosing four randomly clustered WSSs, a power footprint analysis can be conducted on each scenario and compared to the power footprint of the proposed optimized WSS clusters.

### 4.2. Method for Power Usage Analysis

Four decentralized WSS scenarios were considered when calculating power usage for each cluster. This was necessary to create an acceptable sample size for the analysis. To generate randomly assigned clusters, the cluster boundaries for each scenario were rotated



approximately 10° counter-clockwise from the previous scenario. This includes scenario-1 through Scenario-4. In addition, a typical centralized WSS was used as a benchmark for comparison. Since the design of WDS is not the focus of this paper, a parcel-based power footprint calculation was used for comparison between the scenarios. The power requirements of the four decentralized WSS boundaries were compared with the proposed decentralized WSS. Operational cost can be visualized by analysis of the power requirements for the various scenarios. Cost is directly proportional to power, therefore the cluster which demands the least power will be most cost efficient. The power provided by a pump is determined using Eq. (6).

$$P = \rho g Q h_p \quad (6)$$

$$h_p = h_v + h_f + h_E \quad (7)$$

Where  $P$  is the power to supply the fluid,  $\rho$  is the density of the fluid in kg/m<sup>3</sup>,  $g$  is the gravitational acceleration in m/s<sup>2</sup>,  $Q$  is the flow rate in m<sup>3</sup>/s, and  $h_p$  is the total head in meters. The total head is the summation of the velocity head ( $h_v$ ), friction head ( $h_f$ ), and elevation head ( $h_E$ ) are found in Eq. (7). The velocity head was calculated using  $\frac{v^2}{2g}$ . Over large distances, the friction loss of the fluid flowing through the pipes must be calculated. In 2015, Hunter Industries published the article *Friction Loss Tables* [29], which includes standard PSI loss for various types and sizes of piping. The article also contains Eq. (8), the Hazen-Williams formula used for pipe head loss calculations.

$$H_f = \left[ 0.2083 \left( \frac{100}{C} \right)^{1.852} \left( \frac{Q^{1.852}}{d^{4.866}} \right) \right] \times 0.433 \quad (8)$$



Where  $H_f$  is the friction loss (PSI) per 30.48 m (100 ft),  $C$  is the coefficient of retardation of pipe material,  $Q$  is the flow volume (GPM), and  $d$  is the inside diameter of pipe in imperial units. For example, assuming the water will be supplied by a 2.54 cm (1 in) schedule 80 pipe, the coefficient of retardation would be 150 with an inner diameter of 2.43 cm (0.957 in). The friction loss can be converted into a friction head  $h_f$  using Eq. (9). The calculation then has to be converted into metric units.

$$h_f = \left[ \frac{H_f \times 12^2}{\gamma} \right] / 100' \times D \quad (9)$$

In this equation,  $\gamma$  is the imperial unit weight of the fluid and  $D$  is the distance from the source center to the parcel. In some cases, the source center elevation is greater than the parcel being supplied with water, causing the elevation head to be negative. When this arises, the summation of the heads will most likely be negative; it is assumed that the power cannot be negative. This ensures only positive power requirements are considered for the total cluster. The last calculation is converting the power requirements to a yearly operational cost. The following formula can be used to estimate the cost.

$$Cost = R \frac{P(C)}{1000} \times 8760 \quad (10)$$

Where the cost is in USD per year,  $P$  is the power in Watts,  $R$  is the current conversion factor for Ugandan Shilling (UGX) to USD,  $C$  is the current price per kWh for Uganda in shs/kWh. This calculation was completed for each scenario in the power usage scenario analysis.

#### 4.3. Power Usage- Scenario Analysis

The design of water distribution network and optimization of capital cost is not the focus of this paper. Thus, to conduct a comparative analysis on the case study in Arua, a few assumptions had to be made and used for each scenario in a similar manner. Every cluster scenario of the WSS considered to have similar 15.24 cm (6-inch) schedule 80 distribution main with a length halfway to the farthest distance from the source center to the cluster boundary. In addition, a 2.54 cm (1-inch) schedule 80 pipe is to be used for all distribution from the main to customer networks, having a coefficient of retardation of 150 and an inner diameter of 2.43 cm (0.957 in) (used in Eq. (8)). This assumption is made so that each random WSS scenario is compared with optimized cluster using parcel-based energy usage. A few characteristics will change for each parcel of different scenarios, such as length, elevation and flow rate; the length, elevation and flow rate are taken from Figures 9(a, b). The same assumptions are applied to the centralized WSS scenario where the source center is located at the existing water source C9.

When performing the power usage analysis, all equations and concepts discussed in the methodology section were applied to each parcel within the enclosed area, excluding parcels in the existing municipality, forest area and outside the predicted town extent (Figure 6). This was

done for each of the five scenarios to produce an adequate sample for the power usage analysis. Table 2 shows the power requirements of each cluster within every scenario, the total power required by each scenario, the additional power usage each scenario has compared to the proposed WSS, and the operational cost per year for each scenario was calculated. It is also important to note that this is a comparative analysis as opposed to an absolute O&M cost estimate of an optimal WDS. The total and excess power usage from Table 2 is represented in Figure 12 with two separate axes.

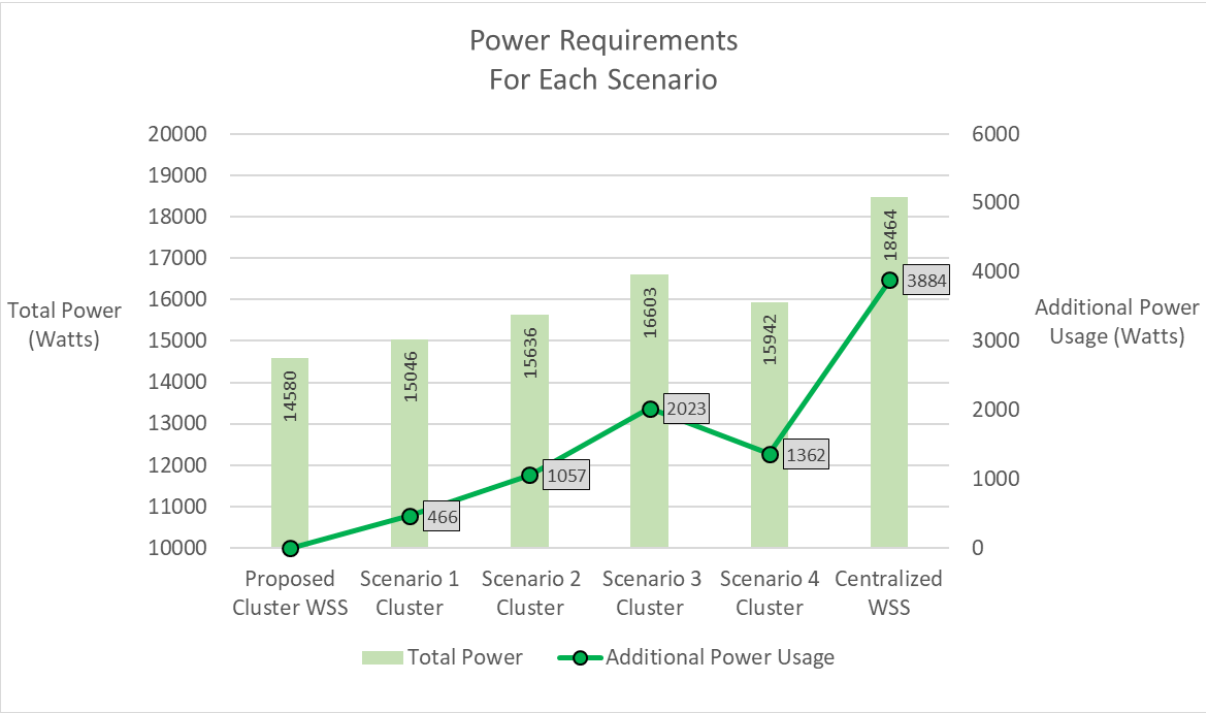
The yearly operational cost for each scenario was estimated using Eq. (10). To use this equation, the current conversion rate for UGX to USD and the price per kWh must be known; as of November 2018, the conversion rate was \$0.00027 USD/UGX (retrieved from Google Finance) and the price was \$0.185247 USD/kWh [30]. Table 2 contains the total cost of the proposed WSS and randomly clustered scenarios.

Table 2. Power requirements of each cluster, total power and operational cost per year for every WSS scenario

	<b>Proposed Cluster WSS</b>	<b>Scenario 1 Cluster</b>	<b>Scenario 2 Cluster</b>	<b>Scenario 3 Cluster</b>	<b>Scenario 4 Cluster</b>	<b>Centralized WSS</b>
<i>Cluster Number</i>	<i>Power Per Cluster(W)</i>	<i>Power Per Cluster(W)</i>	<i>Power Per Cluster(W)</i>	<i>Power Per Cluster(W)</i>	<i>Power Per Cluster(W)</i>	<i>Power for Centralized (W)</i>

1	852.14	804.53	739.50	934.16	956.58	18463.78
2	2967.08	3497.88	3990.10	4420.44	4742.18	
3	1154.50	971.45	820.32	868.80	941.99	
4	4597.13	4296.83	4435.66	4472.43	2986.95	
5	460.32	398.16	366.06	526.45	587.94	
6	4545.88	5074.51	5183.64	5124.28	5089.55	
7	2.46	2.38	101.09	256.46	636.46	
Total Power	14579.51	15045.74	15636.38	16603.01	15941.65	18463.78
Additional Power Usage	0.00	466.23	1056.87	2023.50	1362.14	3884.27
US\$ Per Year	\$23,659	\$24,416	\$25,374	\$26,943	\$25,870	\$29,962

591



592

593

Figure 12. The bar graph represents the total power requirements per cluster. The line graph represents the excess power of the WSS scenarios compared to the proposed WSS

The data shown in Table 2 evidently confirms that the proposed decentralized WSS requires less power and therefore money to operate. Figure 12 illustrates the proposed WSS scenario has the lowest power consumption compared to the conventional centralized WSS and the randomly generated clustered WSS. It was also observed that the operational cost of the random cluster scenarios is 3.2% to 13.9 % higher and centralized system is 26.6% higher than the optimal clustered WSS, consequently verifying the efficacy of the proposed method to determine an optimal cluster boundary for WSS.

## 5. Conclusions

Currently no method exists which guides planners and engineers on how to cluster Water Supply Systems (WSS). To address this need, a new optimization model that supports the development of clustered (decentralized) water systems has been developed and applied to a real case study in Arua, Uganda. The developed clustering methodology is based on two major principles: the minimization of the distance from source to consumer by assigning demand to the closest source center, and the maximization of the homogeneity within a cluster. Euclidean norm minimization was used to optimize the source-demand distance for all parcels to minimize the transportation distance and corresponding infrastructure requirements. Intra-cluster homogeneity was used to measure the similarity or dissimilarity between parcels of the

same cluster. Maximization of intra-cluster homogeneity allows clustering the parcels so that parcel attributes within a cluster are closely related to one another. Three major parameters considered in the clustering process, include membership (determined from Euclidean norm minimization), topography (elevation of the parcels), and spatio-temporal demand distribution (determined from the population distribution, land use, and socio-economic parameters). A K-means optimization technique were applied to maximize intra-cluster homogeneity to reduce the costs associated with water system investment and operation (energy and leakage) and to improve resource efficiency (recycling). The efficacy of the developed clustering method was tested in a real case study of Arua, Uganda. The WSS in Arua was divided into nine clusters, thereby reducing the effort required to move water and wastewater, as well as developing systems that offer opportunity to adapt to future changes. The case study demonstrated that it is possible to apply the developed methodology to delineate clusters based on minimizing distance between source and use and maximizing the intra-cluster homogeneity. An analysis using calculated power requirements showed that the clustered approach did provide lower cost and more efficiency.

**Acknowledgments:** The authors thank the U. A. Whitaker of Engineering, Emergent Technologies Institute at Florida Gulf Coast University for providing funding to publish this research.

636 **Author Contributions:** S.T. developed the concepts, preformed the modelling and wrote the  
637 initial manuscript. T.M.M. checked the methodology and revised the final manuscript. J.K  
638 edited the final manuscript. J.H helped with the power usage analysis.

640 **Author Conflicts of Interest:** The authors declare no conflicts of interest.

## 642 References

- 644 1. United Nations Educational, Scientific and Cultural Organization (UNESCO). *Water A Shared*  
645 *Responsibility: The United Nations World Water Development Report 2*; Berghahn Books:  
646 New York, NY, USA, 2006.
- 647 2. Poplak, R. What's actually behind Cape Town's water crisis. The Atlantic. 15 February 2018.  
648 Available online: [https://www.theatlantic.com/international/archive/2018/02/cape-town-water-](https://www.theatlantic.com/international/archive/2018/02/cape-town-water-crisis/553076/)  
649 [crisis/553076/](https://www.theatlantic.com/international/archive/2018/02/cape-town-water-crisis/553076/) (accessed on 13 August 2019).
- 650 3. Masih, N.; Slater, J. As a major Indian city runs out of water, 9 million people pray for rain.  
651 The Washington Post. 28 June 2019. Available online:  
652 [https://www.washingtonpost.com/world/2019/06/28/major-indian-city-runs-out-water-](https://www.washingtonpost.com/world/2019/06/28/major-indian-city-runs-out-water-million-people-pray-rain/?noredirect=on)  
653 [million-people-pray-rain/?noredirect=on](https://www.washingtonpost.com/world/2019/06/28/major-indian-city-runs-out-water-million-people-pray-rain/?noredirect=on) (accessed on 13 August 2019).
- 654 4. Bairoch, P.; Goertz, G. Factors of urbanisation in the nineteenth century developed countries:  
655 A descriptive and economic analysis. *Urban Studies* **1986**, 23, 285-305.
- 656 5. Webster, M.; Jacobsen, M.; Vairavamoorthy, K. *The Future of Water in African Cities: Why*  
657 *Waste Water?* International Bank for Reconstruction and Development/International

- 658 Development Association (World Bank), Washington, D. C., USA, 2012.
- 659 6. Valerie, N. *New Approach in Decentralized Water Infrastructure*. Coalition of Alternative  
 660 Wastewater Treatment, Gloucester, MA, USA, 2008. Available online:  
 661 [https://www.csu.edu/cerc/researchreports/documents/NewApproachesInDecentralizedWat](https://www.csu.edu/cerc/researchreports/documents/NewApproachesInDecentralizedWaterInfrastructure2008.pdf)  
 662 [erInfrastructure2008.pdf](https://www.csu.edu/cerc/researchreports/documents/NewApproachesInDecentralizedWaterInfrastructure2008.pdf) (accessed 14 August 2019).
- 663 7. Grigg, N. S. *Total Water Management: Practices for a Sustainable Future*. American Water  
 664 Works Association, Denver, CO, USA, 2008.
- 665 8. Bieker, S.; Cornel, P.; and Wagner, M. (2010) Semicentralised supply and treatment systems:  
 666 integrated infrastructure solutions for fast growing urban areas. *Water Science and*  
 667 *Technology* **2010**, 61(11), 2905-2913.
- 668 9. Chen, R.; Wang, X. Cost-benefit evaluation of a decentralized water system for wastewater  
 669 reuse and environmental protection. *Water Science & Technology* **2009**, 59(8), 1515-1522.
- 670 10. Verstraete, W.; Van de Caveye, P.; Diamantis, V. Maximum use of resources present in  
 671 domestic “used water”. *Bioresource Technol.* **2009**, 100(23), 5537–5545.
- 672 11. Newman, P. (2001) Sustainable urban water systems in rich and poor cities--steps towards a  
 673 new approach. *Water Science and Technology* **2001**, 43(4), 93-99.
- 674 12. Cornel, P.; Meda, A.; Bieker, S. Wastewater as a source of energy, nutrients and service  
 675 water. *Treatise on Water Science* **2011**, 3, 92.
- 676 13. Kluge, T.; Libbe, J. Transformation of network infrastructure—strategies for communities: the  
 677 example water. *Difu-Beitrage zur Stadtforschung*, Berlin, Germany, 2006.
- 678 14. Böhm, H. R.; Schramm, S.; Bieker, S.; Zeig, C.; Anh, T. H.; Thanh, N. C. The semicentralized  
 679 approach to integrated water supply and treatment of solid waste and wastewater—a



- flexible infrastructure strategy for rapidly growing urban regions: the case of Hanoi/Vietnam. *Clean Technologies and Environmental Policy* **2011**, 13(4), 617-623.
15. Weber, B.; Cornel, P.; Wagner, M. Semi-centralized supply and treatment systems for (fast growing) urban areas. *Water Science & Technology* **2007**, 55(1-2), 349-356.
16. Wang, X.; Chen, R.; Zhang, Q.; Li, K. Optimized plan of centralized and decentralized wastewater reuse systems for housing development in the urban area of Xi'an, China. *Water Science & Technology* **2008**, 58(5), 969-975.
17. Tsegaye, S., Eckart, J., and Vairavamorthy, K. Urban water management in cities of the future: Emerging areas in developing countries. In *On the Water Front*; Lundqvist, J., Ed.; SIWI, Stockholm, Sweden, 2012, pp. 42-48, 2012.
18. Maliva, R. G.; Missimer, T. M. *Aquifer Storage and Recovery and Managed Aquifer Recharge Using Wells: Planning, Hydrogeology, Design, and Operation*. Schlumberger, Houston, Texas, USA, 2010.
19. GAUFF Consulting Engineers. *Arua Emergency Water Supply Project: Inception Report*. GAUFF Consulting Engineers, 2011. Available online: [http://www.amk.rwth-aachen.de/uploads/media/RECLAIM\\_WATER\\_Publishable\\_Final\\_Activity\\_Report.pdf](http://www.amk.rwth-aachen.de/uploads/media/RECLAIM_WATER_Publishable_Final_Activity_Report.pdf) (accessed 15 July 2012).
20. Herrera, M.; Canu, S.; Karatzoglou, A.; Perez-García, R.; Izquierdo, J. An approach to water supply clusters by semi-supervised learning. In *Proceedings of the International Congress on Environmental Modelling and Software 190*, Ottawa, Canada, July, 2010.

- 700 21. Dopp, K. A.; Godfrey, N. *Legislative Redistricting-Compactness and Population Density*  
 701 *Fairness*, 2011. Available online at SSRN: <https://ssrn.com/abstract=1945879> or  
 702 <http://dx.doi.org/10.2139/ssrn.1945879> (accessed 14 August 2019).
- 703 22. BMBF. *Semicentralized Supply and Disposal Systems for Fast Growing Urban Regions in*  
 704 *China*. Final report of Sino-German Research Project. German Federal Ministry of Science  
 705 and Technology, 2006.
- 706 23. Andersen, K. D.; Christiansen, E.; Conn, A. R.; Overton, M. L. An efficient primal-dual  
 707 interior-point method for minimizing a sum of Euclidean norms. *SIAM Journal on Scientific*  
 708 *Computing* **2000**, 22(1), 243-262.
- 709 24. D'Agostino, M., & Dardanoni, V. (2009). What's so special about Euclidean distance?. *Social*  
 710 *Choice and Welfare*, 33(2), 211-233.
- 711 25. Singh, K.; Malik, D.; Sharma, N. Evolving limitations in K-means algorithm in data mining  
 712 and their removal. *International Journal of Computational Engineering & Management*  
 713 **2011**, 12, 105-109.
- 714 26. Al-Saleh, M. F.; Yousif, A. E. Properties of the standard deviation that are rarely mentioned  
 715 in classrooms. *Australian Journal of Statistics* **2009**, 38, 193-202.
- 716 27. Uganda Bureau of Statistics (UBOS). *Mid-Year Projected Population for Town Councils*.  
 717 Uganda Bureau of Statistics (UBOS), Kampala, Uganda, 2011.
- 718 28. COWATER. *District Town Sewerage, Sanitation Feasibility study: Arua Master Plan, National*  
 719 *Water and Sewerage Corporation*, Uganda, 2005.
- 720 29. Hunter. *The Handbook of Technical Irrigation Information. A Complete Reference Source for*  
 721 *the Professional*, 2015. Available online:

- 722 [https://www.hunterindustries.com/sites/default/files/tech\\_handbook\\_of\\_technical irrigati](https://www.hunterindustries.com/sites/default/files/tech_handbook_of_technical_irrigati)  
723 [on\\_information.pdf](#) (accessed 30 August 2019).
- 724 30. UMEME. *The Electricity End-User Retail Base Tariffs For 2019 As Set and Announced by The*  
725 *Electricity Regulatory Authority*. Uganda, 2018.
- 726