Energy 52 (2013) 77–88

Contents lists available at SciVerse ScienceDirect

Energy

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

A GIS (geographical information system)-based spatial data mining approach for optimal location and capacity planning of distributed biomass power generation facilities: A case study of Tumkur district, India

Deepak Paramashivan Kaundinya [a,b,c,*], P. Balachandra [b,c], N.H. Ravindranath [c], Veilumuthu Ashok [c]

[a] *Department of Mechanical Engineering, University of Alberta, Edmonton, Canada*
[b] *Centre for Sustainable Technologies, Indian Institute of Science, Bangalore 560012, India*
[c] *Department of Management Studies, Indian Institute of Science, Bangalore 560012, India*

ARTICLE INFO

Article history:
Received 21 May 2012
Received in revised form
31 December 2012
Accepted 2 February 2013
Available online 17 March 2013

Keywords:
Bioenergy
GIS (geographical information system)
Spatial clustering
Rural energy planning

ABSTRACT

This paper primarily intends to develop a GIS (geographical information system)-based data mining approach for optimally selecting the locations and determining installed capacities for setting up distributed biomass power generation systems in the context of decentralized energy planning for rural regions. The optimal locations within a cluster of villages are obtained by matching the installed capacity needed with the demand for power, minimizing the cost of transportation of biomass from dispersed sources to power generation system, and cost of distribution of electricity from the power generation system to demand centers or villages. The methodology was validated by using it for developing an optimal plan for implementing distributed biomass-based power systems for meeting the rural electricity needs of Tumkur district in India consisting of 2700 villages. The approach uses a *k*-medoid clustering algorithm to divide the total region into clusters of villages and locate biomass power generation systems at the medoids. The optimal value of *k* is determined iteratively by running the algorithm for the entire search space for different values of *k* along with demand–supply matching constraints. The optimal value of the *k* is chosen such that it minimizes the total cost of system installation, costs of transportation of biomass, and transmission and distribution. A smaller region, consisting of 293 villages was selected to study the sensitivity of the results to varying demand and supply parameters. The results of clustering are represented on a GIS map for the region.

1. Introduction

It is a well known fact that electricity provides impetus to economic as well as human development. However, in majority of the developing countries, the major section of the population, especially in rural areas, is deprived of the benefits of electricity access as well as economic development. India is not an exception to this with 364 million poor people lacking access to electricity [1]. On one hand, the conventional, fossil fuel based, centralized power generation systems have been catering mainly to meet the urban and industrial needs but have failed to address the energy needs of the rural poor [2]. On the other hand, there is an uncontrolled exploitation of the finitely available fossil fuel resources for power generation. This necessitates an urgent need to explore renewable energy options that can be operated on decentralized mode at smaller capacities and are available in plenty. Among all the renewable energy alternatives to generate power, biomass energy route is considered to be advantageous, because: (i) They could be set up at any location where plant vegetation and animal rearing are present, (ii) It is available all round the year and there are no seasonal variations, ensuring non-intermittent supply and (iii) It is cheap, easily portable and environmental hazards are minimal [2]. Woody biomass required for power generation can be generated without destroying the natural forests by growing dedicated plantations on abandoned and degraded land which do not have much competitive uses. Biomass can also be obtained from the crop residues from agricultural lands and plantations. The power generated from biomass offers other intangible benefits such as wasteland development, degraded land reclamation, environmental hazard reduction and local employment generation [3].

---

[*] Corresponding author. Centre for Sustainable Technologies, Indian Institute of Science, Bangalore 560012, India. Tel.: +91 9167520832; fax: +91 8023601428.
E-mail addresses: deepak@mgmt.iisc.ernet.in, deepak.sarangi@gmail.com, paramash@ualberta.ca (D.P. Kaundinya).

0360-5442/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.energy.2013.02.011

India has a large biomass resource potential in terms of agro and forest residues, and large tracts of approximately 40 million ha of wasteland for growing biomass. Current potential for power generation from Agro and forest residues alone is estimated to be 16,000 MW [4]. Therefore, for India, expanding the usage of locally available biomass resources to generate electricity is a logical strategy to address the rural power challenge. The distribution of biomass is not strictly uniform in a geographical region and when locally available biomass resources are insufficient to meet the local electricity needs in the region, biomass has to be imported. When biomass is collected from many villages and transported to a power generation facility, logistic systems have to be designed more efficiently and they should address the issue of optimal location of biomass power facility.

The location decisions with respect to biomass power generation facility mainly depend on two components of variable costs. The first component relates to cost incurred on biomass procurement and biomass transportation from the dispersed sources to the power generation facilities (or biomass power generation systems). Second component of the cost is with respect to evacuating the generated electricity and supplying it to various demand centers (households, agriculture, micro-industry, etc.). In this context, the relevant costs are related to local transmission and distribution systems. Since these biomass power systems are of small capacities and decentralized, the dominant cost will be for the distribution system rather than transmission. Therefore, biomass energy systems can become operationally cost effective by strategizing the location of the energy systems by minimizing transportation cost of biomass from the source, and transmission and distribution cost from energy production system to the demand centers. As an enabler to support such decisions, in this paper, we present a discussion on the development and validation of a mathematical model to determine the optimal location of biomass power systems to minimize the cost of transportation. In addition, the model also facilitates decisions with respect to optimal installed capacity of the biomass systems by matching the given demand for power and available quantity of biomass. The key element in the current study is to cluster villages in the study area, define biomass power generation centers within the clusters, collect biomass from various villages (supply points) and transport sufficient quantities of biomass to the biomass power generation center within the cluster, in order to satisfy the power plant requirement at the least cost so as to meet the dynamic demand for power. Tumkur district in India consisting of 2700 villages is selected for validating the model. Two scenarios, one medium-term (2015) and the other one long-term (2030) were developed for validating model with projected demand for energy, dynamic loads and biomass potential.

In the last few decades, owing to emergence of multitudes of innovative applications in both private and public sector enterprises, a strategic approach had to be adopted to locate the facilities such as warehouses, hospitals, schools and fire stations. A vast number of distribution models were formulated to locate a facility and allocate the demand and capacity to these facilities. The objective of these models was to minimize the total installation and operational costs of the facilities. Each one of them differed from the other in mathematical structure, computational time and complexity.

A summary of continuous location models, network location models, integer programming problems and other state-of-the-art problems was presented by Klose and Drexel (2005) [5]. Francis et al. (1983) [6] provided a review of literature of location analysis. Current et al. (1990) [7] reviewed the multi-objective aspects in the problem domain of facility location analysis. They found that most of the literature was comprised of cost minimization formulation, some of them dealt with demand oriented objectives, and only a few papers were of profit maximization type.

Melkote and Duskin (2001) [8] investigated a model that simultaneously optimizes the location and alters the network topology. They mention about different network location problems like set covering location problem, maximum covering location problem, P-median and P-center problems in which underlying network topology can have a significant impact on the optimal location decision. They also proved that their model can be successfully implemented in a number of applications like regional planning, power distribution, energy management and other areas. Syam (2008) [9] formulated a multiple server location-allocation model by considering most of the relevant costs and other parameters, namely, the transportation cost, facility cost, waiting time cost, queuing time, multiple server facilities and distance constraints.

Nema and Gupta (1999) [10] examined the waste-technology compatibility to locate waste treatment and disposal facilities. They formulated a multi-objective problem integrating both cost and risk parameters. Maniezzo et al. (1998) [11], have developed a decision support system for siting industrial waste management plants, minimizing the total costs and environmental impacts. The resulting formulation was NP-complete which could only be solved by adopting heuristic procedures. The optimal choice of location, technology, routing of hazardous waste was investigated by Alumur and Kara (2007) [12].

From the review of existing models it is evident that facility location planning models were not applied to locate decentralized energy systems. All the location—allocation studies found in the literature presuppose the set of demand points and end-use points, i.e., the number of locations is known a priori. When the number of locations is unknown, the classical location—allocation models cannot be applied. In the current study, we have proposed PAM (partitioning around the medoid) clustering algorithm (also known as *k*-medoid clustering) to tessellate the space or region and iteratively arrive at an optimal number of locations for power plants to minimize the total cost of transportation of biomass, transmission and system installation. The optimal tessellation is represented on a GIS (geographical information system) map. In the ensuing sections spatial tessellation using clustering algorithms and *k*-medoid algorithm based model formulation are explained. The key input parameters used in the proposed algorithm are biomass potential for power generation, demand for power, peak load and different cost parameters. The algorithm is developed in such a way that it simultaneously determines both the optimal number of biomass power systems and their locations, and right-size of the installed capacity of the system by minimizing the total costs of installation, transportation, and transmission and distribution.

1.1. Clustering algorithms — spatial tessellation

Clustering is an important component of data mining. It refers to the process of combining any set of physical and abstract objects into groups or classes of similar objects. The objects are combined such that they are similar to each other within the cluster and dissimilar to objects from other clusters [13]. Most of the clustering algorithms applied on spatial data help in computation of quantitative results by tessellation of the data space. A tessellation is created by repeating a shape to cover a plane without any gaps or overlaps. In other words, tessellations are sets of mutually non-intersecting planes covering a region.

A tessellation is defined as follows [14]:

$$\text{Let the set } \mathfrak{I} \text{ be a closed set.}$$

$$\text{Let } \exists \mathcal{S}i \ni \bigcup \mathcal{S}i \supseteq \mathfrak{I} \text{ where } \mathcal{S}i \text{ are all open sets}$$

$$\text{In addition to this,}$$

$$\text{if } \bigcup \mathcal{S}i \equiv \mathfrak{I} \text{ and } \mathcal{S}i \cap \mathcal{S}j \equiv \emptyset \quad \forall\, i \neq j$$

$$\text{We say } \mathfrak{I} \text{ is tessellated by } \mathcal{S}i$$

Of the several clustering algorithms, partitioning method is widely used for large data bases and in particular GIS-based applications [15]. Partitioning clustering involves, partitioning the database *D* of *n* objects into a set of *k* clusters. Given *k*, find a partition of *k* clusters that meets the optimization criterion. In *k*-medoid clustering, the medoids of each cluster are taken as a reference point of cluster. The most widely acknowledged algorithm for *k*-medoid clustering is the PAM (partitioning around the medoids) in which *k* is specified in advance and the algorithm determines *k* partitions of *n* objects using medoids of the *k* cluster as the reference point for finding dissimilarity with other objects [13]. In the current study, the geographical entities (villages) with corresponding latitude and longitude values will become the database to be partitioned. The *k* biomass power systems are located at the *k* medoids.

1.2. Profile of the study location

The district of Tumkur adjoining Bangalore is located in the semiarid belt in the eastern half of the Karnataka state. It has ten talukas (blocks) and has an area of 10,596 Sq km. It lies between 13° 25′ and 13° 40′ north latitude and 76° 24′E and 77° 30′E longitudinal parallels. The study location is shown in Fig. 1.

The district has a number of local administration units namely 50 hoblies, 321 Gram Panchayats, 11 urban agglomerations, 2574 inhabited villages, 134 uninhabited villages and 1315 hamlets. The total population has grown by 11.87% between 1991 and 2001. According to 2001 census the total population is 2.58 million as against 2.30 million during 1991 [16].

The population density is 244 persons per Sq km and total literacy rate is 67%. More than 60% of the population is concentrated in rural regions. The landscape consists mainly of undulating plains with a sprinkling of hills. About 45,177 ha (4%) of land are classified under forest. Over exploitation of forests due to permissible free grazing has resulted in vagaries in rainfall pattern. The average annual rainfall in the district is highly variable and it decreased by 50% from 834 mm in 2005 to 470 mm in 2006. All the 10 administrative blocks exhibit similar demographic and geographic properties in Tumkur district [16]. Among them, Kunigal block was randomly selected to analyze the sensitivity of results to changes power demand and supply.

2. Biomass assessment for power generation

The important biomass resources available for power generation in the region are woody feedstock obtained by growing dedicated plantations in the wasteland and, crop residues and plantation residues. Crop residues include stalks of maize, maize cobs, groundnut shells and rice husk. Crop residues include stalks of maize, maize cobs, groundnut shells and rice husk. Common plantation residues include coconut husk, fronds and pith.

The finer details of the land-use characteristics at the village-level were obtained from the land-use/land cover GIS maps [17] on a scale of 1:1,134,063. It showed land-use characteristics ranging from cropland, degraded forest, tanks, coconut and areca nut plantations, scrub forest, fallow land and barren/stony wasteland. Cropland and degraded forest land were not considered for woody biomass extraction. From the cropland, surplus crop residues were considered as a source for biomass. Only wasteland was considered for growing biomass for power generation.

The intersection of Land Use Land Cover map and the village boundary map give rise to a new attribute table consisting of village names, land-use categories and area under different land-use categories for the entire district. Based on the previous literature [18–20], interview with villagers and land suitability index lookup table, the total technical biomass potential was estimated for each village.

A land suitability index was prepared and the total biomass availability under various land-use types was determined. These indices, based on expert knowledge in agronomy and previous literature, indicate the capability of a specific location to grow biomass for power in terms of maximum attainable yield (tonnes of biomass) per ha of a land-use category [18]. The maximum attainable yields under various land-use categories are shown for Tumkur region and Kunigal block in Table 1. A close examination of

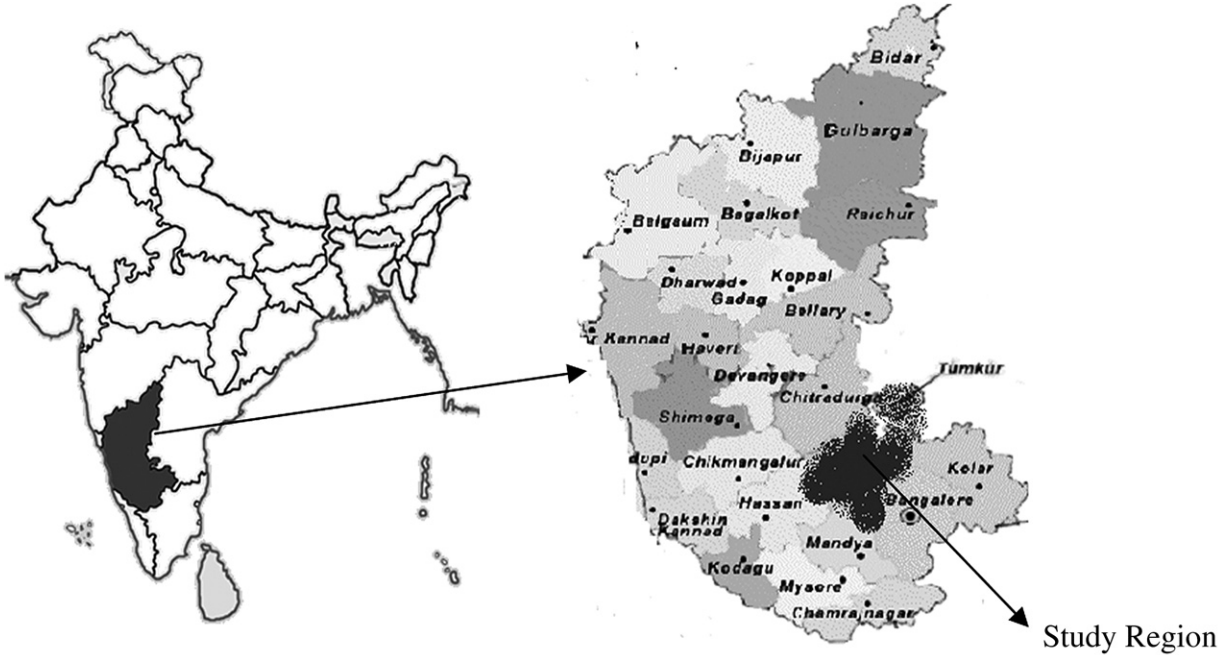

Fig. 1. Study location.

Table 1
Total biomass feedstock production from different land-use types.

Land-use type	Biomass productivity (tonnes/ha)	Tumkur district		Kunigal block	
		Area in ha	Yield in tonnes	Area in ha	Yield in tonnes
Cropland	0.36[a]	696,429	250,714	60,663	21,838
With/without scrub	8	37,582	300,661	4023	32,183
Plantations	2.8[b]	142,811	399,870	13,157	36,840
Degraded forest	6	23,425	140,551	1540	9245
Forest plantations	6	13,555	81,330	1336	8016
Fallow	8	17,166	137,335	1923	15,384
Prosopis juliflora	4[c]	6039	24,157	0	0
Barren rocky/stony waste	1	24,933	24,933	2623	2623
Gullied/ravenous land	6	2974	17,849	228	1367
Salt affected land	1	7314	7314	0	0
Mining/industrial waste	4	6	24	0	0
Degraded	8	1207	96,556	38	305
Total		**1,059,592**	**1,394,399**	**94,686**	**127,499**

[a] Average crop residue was estimated to be 0.36 tonnes/ha. This was derived using 1.43 tonnes/ha of crop production. Biomass yield is 1.5 tonnes per every tonne of crop. Average surplus crop residue is 17% in Tumkur district [21]. Crop residue = 1.43*1.5*0.17 = 0.36.

[b] Both areca nut and coconut plantations generate an average 3.5 tonnes/ha of pith, frond and kernel. Fuel content will be 0.8 tonne per every tonne of biomass. Total biomass potential from areca plantation is 3.5*0.8 = 2.8.

[c] Ref. [20].

land types shows that among all the exploitable land types, plantations have maximum potential for biomass followed by land with/without scrub.

Given the *current growth* rates and land suitability indices for different land-use types, our study indicates that the overall *attainable* biomass potential in the medium-term, i.e., 2015 is 1.39 million tonnes in Tumkur district and 0.13 million tonnes in Kunigal block. For the long-term scenario, i.e., in 2030, *assuming* a 10% reduction in wasteland availability due to increase in cropland, the estimated total biomass supply is 1.1 million tonnes in Tumkur district and 0.11 million tonnes in the case of Kunigal block.

2.1. Assessment of demand for power in Tumkur district

The total annual peak load and aggregate demand for energy were calculated as a sum total of demands for household lighting and home appliances, street lighting, agricultural water pumping and industry sectors using literature. A primary survey was carried out during 2008–09 in the randomly chosen sample households from the villages again chosen randomly for collecting end-use-wise energy consumption data. Representative estimates of individual villages were used to calculate the aggregate demand for Tumkur district and Kunigal block in 2015 [20,22].

Since demand for electricity is strongly dependent on the socio-economic status of villages, an attempt was made to capture the variations in demand for electricity across villages belonging to various socio-economic categories. To facilitate such an assessment, the villages of Tumkur were classified into five socio-economic categories namely *Very high*, *High*, *Medium*, *Low* and *Very low*, based on socio-economic factors such as population, possession of domesticated animals, average literacy, income, land

Table 2
Representative peak loads for each socio-economic category.

Socio-economic category	Representative village	Per household peak load (W)
Very low	Kalnayakanahalli	44
Low	Kithnamangala	65
Average	Singonahalli Agrahara	114
High	Channapura	152
Very high	Hanchipura	256

Table 3
Annual demand in kWh for end-uses for different socio-economic categories [22].

End-use	Very high	High	Average	Low	Very low
Lighting and home appliances[a]	477	236	166	93	62
Street lighting[b]	17.5	17.5	17.5	17.5	17.5
Irrigation pump sets[c]	1350	1404	2214	3402	2808
Agro-processing[d]	16,000	16,000	8000	8000	4000
Industry[e]	12,000	12,000	6000	6000	3000
Domestic water pumping[f]	7300	7300	3650	3650	1825

[a] The values correspond to per household demand. For the entire village demand = number of households*per household demand for lighting and home appliances.

[b] The values correspond to per household demand at one street light per 10 households. For the entire village, demand = number of households*17.5.

[c] The values correspond to per household demand at 5400 kWh/pump/year. For the entire village, demand = number of households* no of irrigation pump sets per household*5400 kWh.

[d] The values indicate the demand for the whole village at 20 kWh/day for very low category, 40 kWh/day for low and average, and 80 kWh/day for high and very high categories for 200 days in a year.

[e] The values indicate the demand for the whole village at 10 kWh/day for very low category, 20 kWh/day for low and average category, and 40 kWh/day for high and very high categories 300 days in a year.

[f] The values indicate the demand for the whole village at 5 kWh for very low category, 10 kWh for low and average category and 20 kWh for high and very high categories of villages for 365 days.

Table 4
Total energy demand in Tumkur district (MWh).

Sl. no.	Block name	Base case scenario in 2015	ET[a] scenario in 2015	ET scenario in 2030[b]
1	Chikknayakanahalli	98,144	62,016	74,419
2	Gubbi	127,538	81,525	97,830
3	Koratagere	86,027	55,370	66,444
4	Kunigal	113,813	73,084	87,701
5	Madhugiri	120,756	76,581	91,898
6	Pavagada	95,431	58,440	70,129
7	Sira	112,936	70,487	84,584
8	Tumkur	140,336	89,650	107,580
9	Turuvekere	87,369	56,149	67,379
10	Tiptur	401,662	252,069	302,483
11	**Total**	**1,384,017**	**875,376**	**1,050,447**

[a] ET – efficient technology demand. Tube-lights are replaced by CFL (compact fluorescent lamp) reducing 2/3rd of the demand. Irrigation pump set valves replaced by efficient ones reducing 35% of the demand and street lights are replaced by efficient ones reducing demand by 50% [23–26].

[b] In the year 2030, 20% increase in demand was envisaged [26].

Table 5
Cost parameters in transportation analysis.

Sl. no.	Name of the item		Truck	Tractor (35 hp) + Wagon
1	Purchase price (Rs.)		505,368	379,026
2	Life (years)		10	15
3	Carrying capacity (tonnes)	Loose biomass	2.5	2.5
		Briquetted	8	6
4	Fuel consumption (L/km) (loaded)		0.22	3.5
5	Equivalent annualized cost (in Rupees) (at 10% cost capital)		82,246	49,831
6	Operation and maintenance cost (in Rupees) (at 10% of capital per year)		50,536	37,902
7	Average speed km/h		50	12.5
8	Cost of transportation per km (in Rupees)		30	20

Source: [28,29].

Table 6
Average cost of electricity distribution system for a typical village.

Sl. no.	Name of the item	Quantity	Unit rate (Rs. millions)	Total cost (Rs. million)
1	High transmission line	2.5 km	0.2	0.5
2	Low transmission line	1.5 km	0.2	0.3
3	Distribution transformers	3 Nos.	0.1	0.3
4	33 KV works – on prorate basis	–	–	0.2
	Grand total			1.3

Source: [31].

holdings, access to basic amenities, etc. The socio-economic factors were assessed by conducting a questionnaire-based survey in the study region during 2008–09. Households were asked to furnish details about whether the household is electrified or not, connection details, total number of end-use devices like lights, television, fans and other home appliances if any, and their daily power consumption pattern for these end-uses. Farmers were interviewed to elicit responses regarding the total rain-fed and irrigated land holdings, total area under different crops, number of irrigation pump sets, irrigation pattern and data on cropping pattern. Additional data on literacy levels and total bovine population in the household were also collected to determine the socio-economic status of the household and eventually the village. For each socio-economic category of villages, the representative power demand profiles were created, and peak load was determined by adopting Max (Max) function [22]. The representative peak demands are given in Table 2. The annual energy demand levels for villages of different socio-economic categories are shown in Table 3.

The estimated total energy demand in all the 10 blocks of Tumkur district is shown in Table 4.

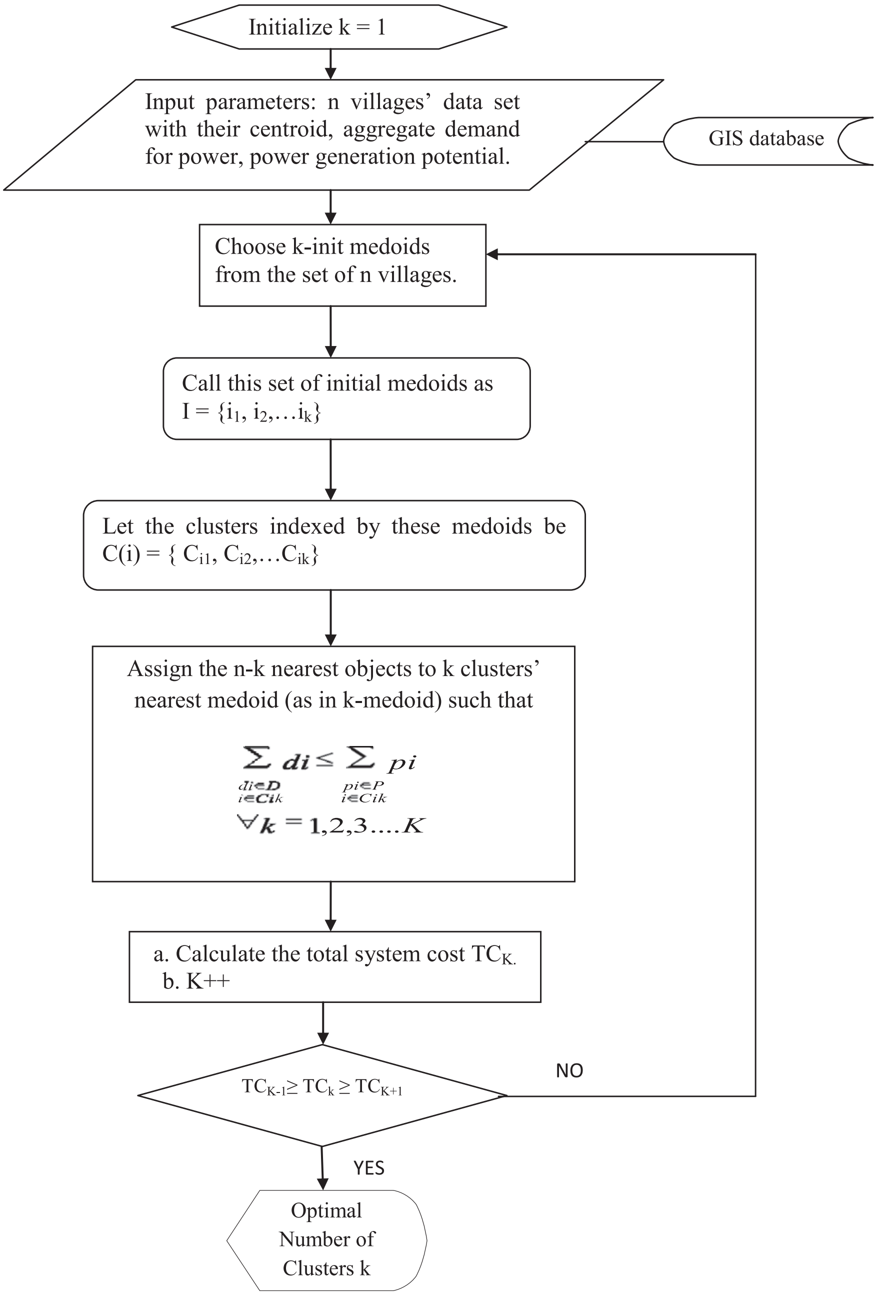

Fig. 2. Flow chart of the algorithm.

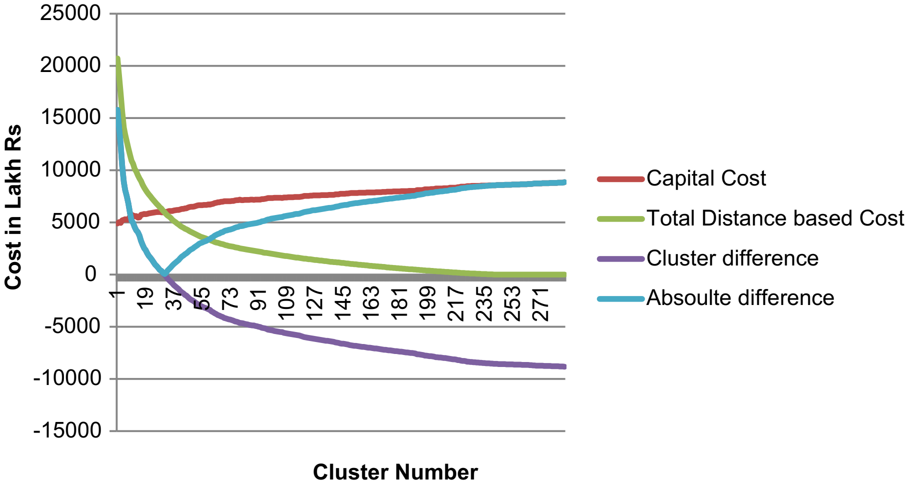

Fig. 3. Variation of overall cost components with cluster number.

For the medium-term, two cases were considered for estimating the energy demand. Base case scenario for 2015, assumes continuation of dependency on existing appliances and existing demand pattern. In the second case, the ET (efficient technology) scenario for 2015, the total energy demand was estimated by assuming higher energy efficiency levels (i.e., assuming that the existing appliances are replaced by efficient technology devices as explained in the footnote (a) of Table 4). For the long-term, only the ET scenario was considered and the estimated total energy demand for 2030 under this scenario suggests that year 2030 will experience an increase in energy demand of just 20% compared to ET scenario in 2015.

The study shows (Table 4) that Tiptur block has the maximum demand for electricity compared to rest of the blocks in Tumkur district. The high demand in this block can be correlated to the high population density in this block. It was found that out of 227

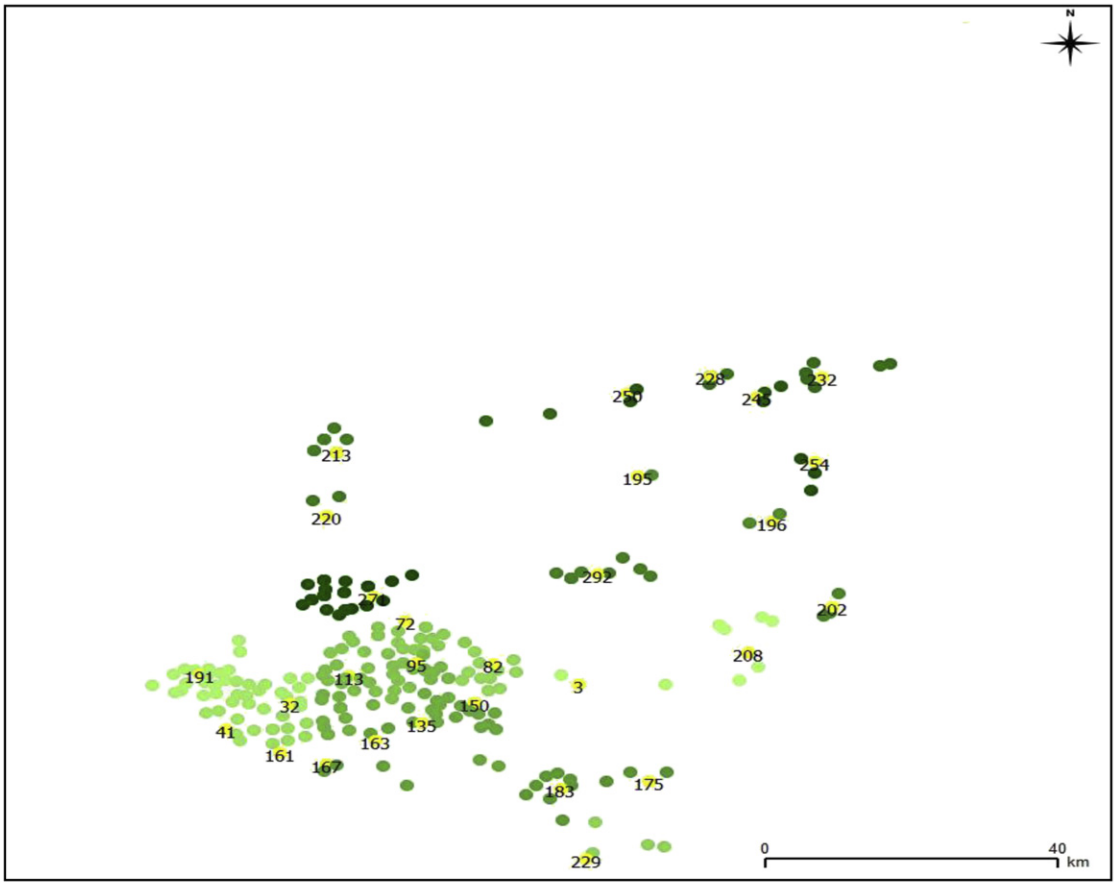

Fig. 4. GIS representation of medoids of different clusters for Kunigal block – ET scenario 2015.

villages in this block 123 villages have more than 500 households. Koratagere has the least demand for electricity followed by Turuvekere. From the biomass assessment results, considering conversion rates of 1.3 kg/kWh, it was estimated that the aggregate power generation potential using biomass sources (existing surplus crop residue and by growing biomass plantations on available wasteland) in Tumkur district by is 1180 GWh whereas the total base case demand in 2015 will be 1384 GWh.

Thus, the estimated biomass potential is inadequate to meet the base case demand in. On the other hand, the total biomass potential is adequate to meet the demand in ET scenario of 875 GWh in 2015. It can also be observed that if biomass production drops down by 10% due to increase in cropland, urbanization, the quantity of biomass available is still sufficient to meet 20% increase in demand for power in ET 2030 scenario.

2.2. *Cost parameters*

Total biomass resource required for meeting the power needs in each cluster and their production costs were derived using the estimates presented earlier as well as additional parameters obtained from secondary sources of data. Transportation costs over a defined cluster were assessed by considering the average cost of transportation of all available modes of transportation [27–29]. The biomass farm-gate price was assumed to vary linearly with the quantity of biomass. Therefore, the effect of unit price of biomass on location is negligible, while the location decision depends strongly on transportation costs. Three modes of transportation are currently available within Tumkur district for transporting biomass namely bullock carts, tractors and trucks. Bullock carts have limited carrying capacities and constrained by the distance it can travel [30]. Therefore, for the current study, only tractors and trucks were considered for transporting biomass from origin to the facility. Each one of them is limited by different carrying capacities (volume of biomass). The costs include fuel cost (diesel for trucks and tractors), annual operations and maintenance cost, and labor cost. The assumptions made and cost parameters used in the calculations are presented in Table 5.

From Table 5 and previous literature [26], the average cost of truck and tractor modes of transportation was estimated to be *Rs. 25/km*. Apart from the capital cost and transportation cost, the average cost of transmission and distribution of electricity was also a deciding factor in location analysis. The average cost of setting up electricity transmission and distribution (T&D) system in a typical village is shown in Table 6. Using these estimates, the average cost of T&D system was estimated to be Rs. 0.2 million per km for low voltage LT (transmission lines) and an additional cost of Rs. 0.5 million (cost of 3 distribution transformers and a 33 KV works-on prorate basis) per village.

The biomass farm-gate price was assumed to vary linearly with the quantity of biomass and the cost of biomass thus was assumed to be Rs. 1000 per tonne.

The system capital cost and cost of T&D are one-time costs over the life of the project, whereas transportation cost is an annual variable cost. The optimal number of clusters was chosen as the cluster with the least absolute difference between capital cost, and

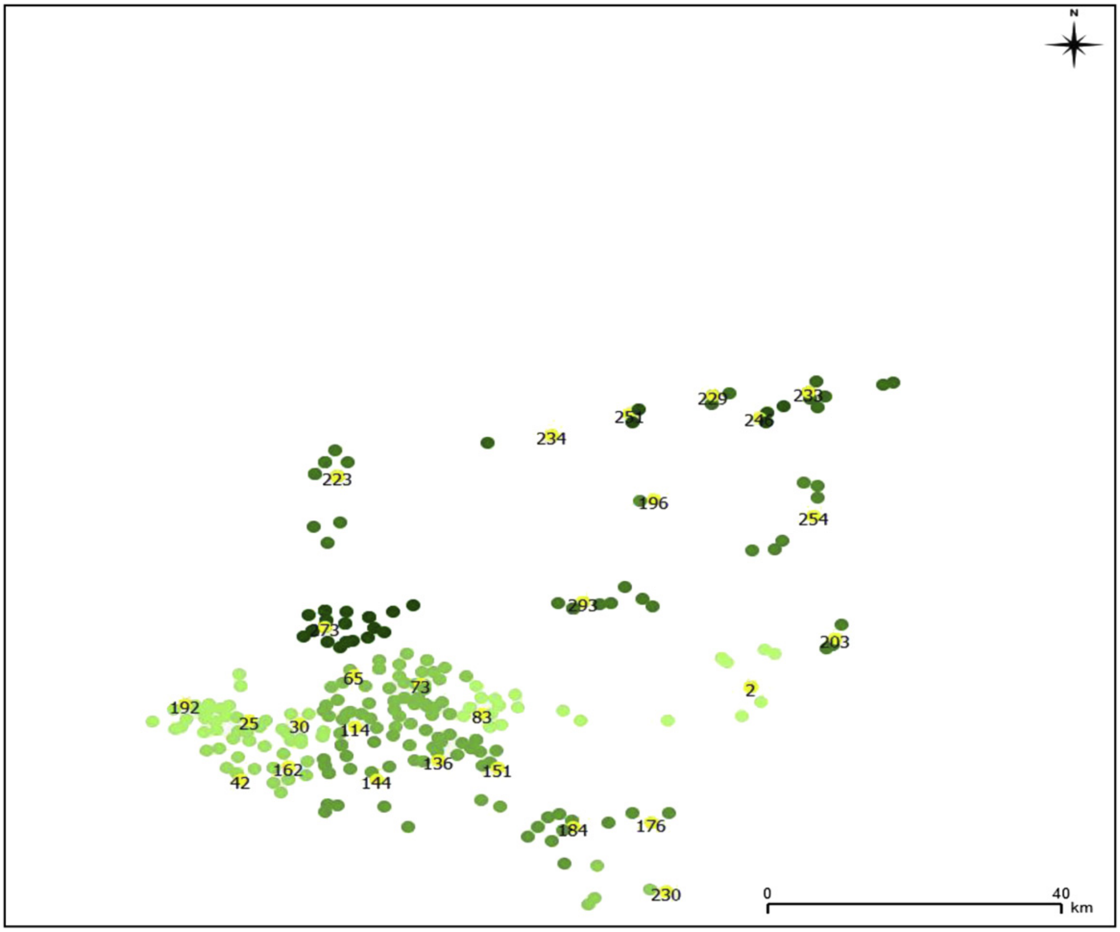

Fig. 5. GIS representation of medoids of different clusters for Kunigal block – ET scenario 2030.

transportation and T&D cost. This assumption was necessary to find the optimal cluster numbers because with the increase in the number of clusters, the total number of biomass energy systems also increased. As an outcome of which the total capital cost for the systems also increased. On the other hand, with number of clusters declining, total distance traveled within the system came down, reducing the cost of transportation of biomass and transmission of power.

3. Proposed algorithm

The *k*-medoid clustering algorithm was iteratively run for the entire range of values specified for *k* as input and modified algorithm would compute the system cost for different values of *k* and exhaust all the values of *k* specified in the range. Constraints were included within the algorithm to ensure matching of demand and supply in the algorithm. The villages were clustered using the proposed *k*-medoid clustering algorithm [14].

3.1. Algorithm parameters

Let

V $= \{v_1, v_2,\ldots..v_n\}$ be the set of villages with demand centroid (latitude, longitude values) coordinates $\{(x_{11},x_{12}), (x_{21},x_{22}),\ldots\ldots(x_{n1},x_{n2})\}$.

Let the aggregate demand for each village be

D $= \{d_1, d_2, \ldots.d_n\}$ *in kWh*

Let the power generation potential equivalent to the total biomass availability in each village be

P $= \{p_1, p_2,\ldots.p_n\}$

Let

K $= \{1, 2, 3\ldots k\}$ be the total number of clusters

T $= \{1, 2, 3\ldots t\}$ is the set of all possible standard capacities of biomass power plants

$\mathbf{C}_{tk}$ be the annualized total cost of operation of biomass power system with capacity t, $\forall t \in T$

$\mathbf{TC}_k$ be the total cost of the system for the entire region, $\forall k \in K$

Each village center was identified by as an ordered pair of latitude and longitude values. Using these parameters *k*-medoid algorithm was run with an additional constraint of matching the total power supply available with total demand from all the villages within the cluster. This is repeated for different values of *k*. Since the power systems are being designed for decentralized application keeping in mind the objective of meeting local demand, the standard capacities are restricted to be within 1 MW for these clusters. That is, the upper bound for cluster level installed capacity of the biomass power system is set to 1 MW. For each *k* (number of clusters in the region) the total system costs are compared and the *k* for which the total system cost is minimum represents the optimal number of clusters. The proposed algorithm is shown in Fig. 2.

The algorithm initializes $k = 1$. Then based on the minimum number of clusters (such that the maximum installed capacity does not exceed 1 MW) the first set of clusters is obtained. Step 6 in the algorithm ensures that the total supply is always greater than the total demand within the cluster. Then the total system cost is calculated for this cluster configuration. The algorithm is run for all the values of *k*. The *k* with minimum total system cost is the optimal number of clusters. We have tried to arrive at the optimal *k*

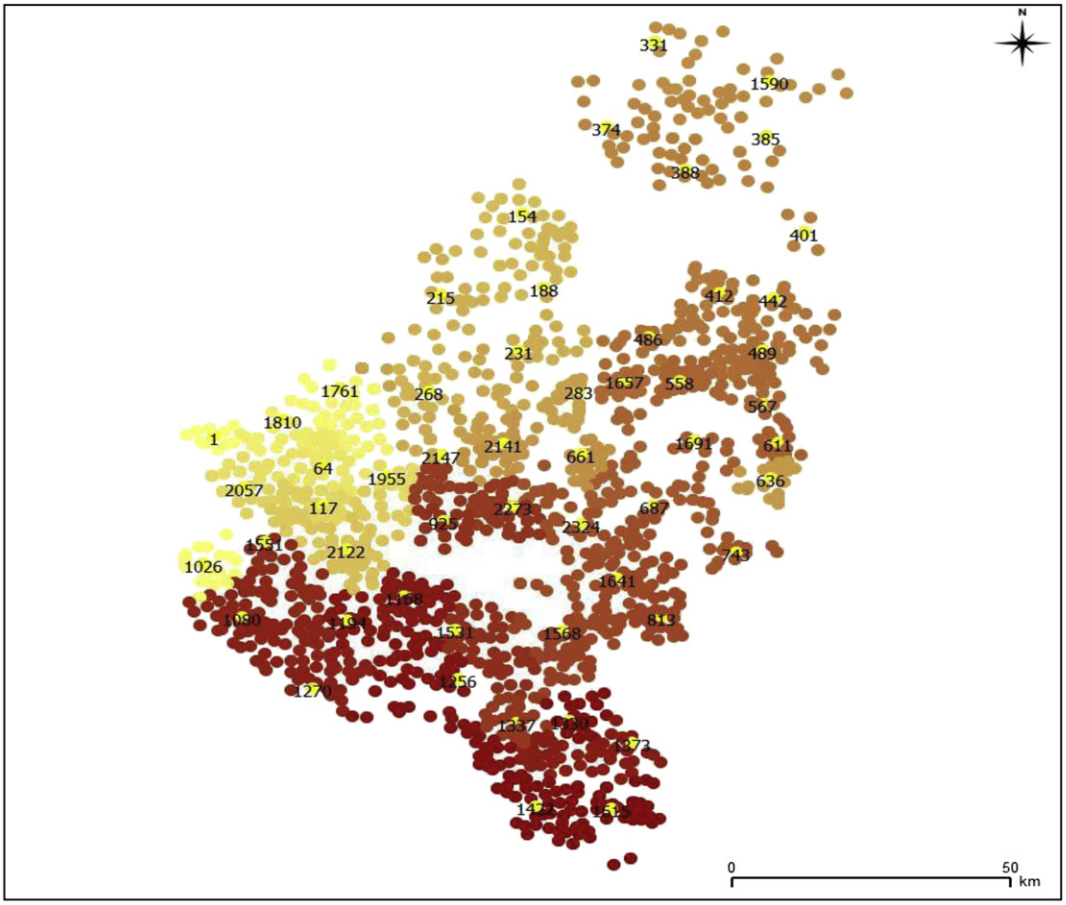

Fig. 6. GIS representation of medoids of different clusters for Tumkur district – ET scenario 2015.

Table 7
Summary of *k*-medoid clustering results for Kunigal block for current demand with efficient technology devices (2015).

Sl. no.	Name of the item	Description
1	Optimal value of *k*	37
2	Total annual biomass requirement (tonnes)	95,000
3	Total annual biomass feedstock cost (million Rs.)[a]	95
4	Total distance traveled in all the clusters (km)	745
5	Transportation cost of biomass per trip (Rs.)[b]	18,600
6	Local transmission and distribution cost (million Rs.) (LT lines + additional transformer cost of Rs. 0.5 million per village)[c]	290
7	Installed capacity	6.4 MW
8	System capital cost (million Rs.)	600

[a] Fuel wood only at Rs. 1000 per tonne, excluding transportation cost.
[b] At Rs. 25/km.
[c] For stand-alone systems only.

(number of clusters) iteratively by solving for all the *k*s because any lower and upper bound for *k* is still a theoretical bound and is of little significance while solving real world problems such as the current problem [13]. The algorithm gives the optimal value of *k* (The cluster corresponding to the least total system cost) that would minimize the total cost for the region, which includes total system installation cost, cost of biomass transportation and cost of T&D. The value of *k* indicates that the optimal solution which minimizes the total cost will have *k* clusters of villages. Within each cluster there will be many villages. The power generation centers were proposed to be located at the *k* medoids of the optimal clusters. The biomass collected from all the villages within the cluster has to be transported to the medoids of that cluster and biomass-based power generation system will be installed at the medoid. These medoids along with villages in the cluster are indicated on the GIS village boundary map (Figs. 4–6).

4. Implementation of *k*-medoid clustering algorithm and results

4.1. Clustering of villages in Kunigal block – efficient technology scenario 2015

The block level clustering was performed for Kunigal block consisting of 293 villages. For model validation, annual energy demand of 73 GWh under ET scenario for 2015 is considered (Table 4). The supply within the block from biomass resources using all the wasteland available for growing energy plantations and available crop residues was estimated to be 98 GWh (81 GWh from fuel wood and 16 GWh from crop residue constituting 83% and 17% of the total potential). The total required installed capacity of the biomass power system for the entire block (including the reserve margin of 20%) was 6.4 MW. From the clustering algorithm output, the optimal number of clusters for the changed demand and supply values was found to be 31.

The summary of variation of different cost components of the system with the number of clusters is shown in Fig. 3. Since the distance varies monotonically with number of clusters, the cost of intra-cluster distance reduces with the increase in number of clusters. In the same way, the capital cost also varies monotonically but in the opposite direction, therefore increasing with the increase in the number of clusters. The total distance based cost component refers to the sum of transportation cost and local T&D cost. The minimum absolute difference of costs occurred at number of clusters $k = 37$. These optimal medoids were identified along with their latitude and longitude values from the result. The corresponding latitude and longitude values were located on the Kunigal block GIS map. The GIS map with clusters and medoids is shown in Fig. 5.

Table 8
Summary of *k*-medoid clustering results for Kunigal block for ET scenario in 2030.

Sl. no.	Name of the item		Description
1	Optimal value of *k*		32
2	Total annual biomass requirement (tonnes)	Fuel wood	94,000
		Crop residue	19,000
3	Total annual biomass cost (million Rs.)	Fuel wood	94
		Crop residue[a]	28
4	Total distance traveled in all the clusters (km)		848
5	Transportation cost of biomass per trip (Rs.)[b]		21,000
6	Local transmission and distribution cost (million Rs.) (LT lines + additional transformer cost of Rs. 0.5 million per village)[c]		310
7	Installed capacity		7.8 MW
8	System capital cost (million Rs.)		670

[a] At Rs. 1000 per tonne and Rs. 500 processing cost for crop residue.
[b] At Rs. 25/km.
[c] For stand-alone systems only.

Table 9
Important results of *k*-medoid clustering for Tumkur district – ET scenario 2015.

Sl. no.	Name of the item		Description
1	Optimal value of *k*		96
2	Total annual biomass requirement (million tonnes)	Fuel wood	0.97
		Crop residue	0.16
3	Total annual biomass cost (million Rs.)	Fuel wood[a]	970
		Crop residue[b]	160
4	Total distance traveled in all the clusters (km)		7158
5	Transportation cost of biomass per trip (Rs.)[c]		1,78,000
6	Local transmission and distribution cost (million Rs.) (LT lines + additional transformer cost of Rs. 0.5 million per village)[d]		2600
7	Installed capacity		71 MW
8	System capital cost (million Rs.)		571

[a] At Rs. 1000 per tonne fuel wood only.
[b] At Rs. 1000 per tonne and Rs. 500 processing cost for crop residue.
[c] At Rs. 25/km.
[d] For stand-alone systems only.

Table 10
Transportation cost from source to the power plant.

Census name	Easting	Northing	Distance to medoid (km)	Cost of transportation (Rs.)
Chikka Malalavadi	656119.4	1463715	1.9	48
Singanhalli	653260.7	1463182	4.6	115
Dodda Malalavadi	654824.9	1463973	3.2	80
Settibidu	656733.7	1463126	1.1	29
Hedigere	654876.2	1461577	3.3	83
Shettikere	650351.9	1461184	7.7	193
Bagenahalli	660831.8	1463756	3.0	76
Gunnagere	660270.3	1462102	2.5	63
Chotnahalli	657435.9	1461722	1.3	33
Rayagonahalli	658878	1461177	2.0	51
Kodihalli	659481.8	1463261	1.6	40
Kurudihalli	660270.3	1462102	2.5	63
Kottekere	654188.4	1460398	4.5	112
Kannasandra	653447.2	1460054	5.3	132
Huralibarasandra	657961.9	1463895	0.9	23
Nilattahalli	657961.8	1463896	1.0	25
Sulekuppe	662273.4	1466818	5.8	145
Taredakuppe	662087.6	1468783	7.1	179
Kamanahalli	662273.6	1466818	5.7	144
Chikka Malalavadi	656733.7	1463126	1.1	29
Naganahalli	657890.4	1462978	Medoid (system location)	Medoid (system location)
Vaddarakuppe	658971	1459926	3.2	80
Vanegere	657398.3	1459505	3.5	87
Total			73	1827

Table 11
Supply–demand–capacity-T&D costs for the cluster.

Census name	Economic category	Annual energy demand (MWh)	Generation potential from biomass (MWh)	Demand for power (kW)	T&D capital cost (million Rs.)
Chikka Malalavadi	3	364.9	281.29	34	0.9
Singanhalli	3	333.33	379.42	31	1.4
Dodda Malalavadi	3	286.73	359.68	27	1.1
Settibidu	3	443.07	486.81	42	0.7
Hedigere	1	199.88	301.11	6	1.2
Shettikere	2	67.17	73.09	2	2.0
Bagenahalli	3	381.44	298.94	36	1.1
Gunnagere	4	372.36	470	67	1.0
Chotnahalli	1	196.17	113.1	6	0.8
Rayagonahalli	3	352.88	359.84	33	0.9
Kodihalli	3	328.82	165.74	31	0.8
Kurudihalli	3	378.43	262.88	36	1.0
Kottekere	2	121.2	217.07	4	1.4
Kannasandra	2	85.18	149.21	3	1.6
Huralibarasandra	2	42.41	61.09	1	0.7
Nilattahalli	3	328.82	340.82	31	0.7
Sulekuppe	1	160.92	288.78	5	1.7
Taredakuppe	4	412.37	796.83	75	1.9
Kamanahalli	1	262.94	230.75	8	1.7
Chikka Malalavadi	2	67.17	225.65	2	0.7
Naganahalli	1	216.57	160.52	6	0.5
Vaddarakuppe	1	223.99	250.95	7	1.1
Vanegere	1	227.7	367.58	7	1.2
Total		5854.45	6641.15	500	26.1

The system design and other important results from clustering analysis are summarized in Table 7. The costs presented in the table refer to the overall system cost. For example, total cost of biomass transportation for the entire system is the cost of transporting biomass from each village to the medoid of the cluster to which the village belongs.

4.2. Clustering of villages in Kunigal block – efficient technology scenario 2030

For the long-term (ET scenario in 2030), as stated earlier, the energy demand is assumed to increase by 20% and the biomass supply is assumed to decrease by 10%. Under these assumptions, the total energy demand in 2030 is estimated to be 87.7 GWh and the total supply available is 88.480 GWh, which is still adequate to meet the increased demand. The total required installed capacity is 7.8 MW. The optimal number of clusters in this case is 32.

When the demand–supply parameters were changed, the minimum absolute difference occurred at cluster number $k = 32$. The difference between the costs initially increased and changed sign with the increase in the number of clusters. At this point absolute difference reached its minimum. This was chosen as the optimal cluster number. The important results of clustering analysis are presented in Table 8.

For these assumed values of demand and supply, the optimal k medoids were identified on the GIS map and locations of the biomass power systems were located. The GIS representation of the optimal clusters and medoids is shown in Fig. 6.

4.3. District level clustering analysis

The aggregate ET scenario demand for energy at the district level is estimated at 875 GWh in 2015 and the biomass requirement to meet this demand is estimated at 1.13 million tonnes. At the district level the aggregate biomass supply is likely to be 0.97 million tonnes of fuel wood and 0.21 million tonnes of crop residue. The key results of k-medoid clustering analysis at the district level are presented in Table 9.

The results show that the optimal number of clusters is 96. Total system capital cost to install power systems at all the medoids (with varying capacities within each cluster) was found to be Rs. 571 million. The clusters and their respective medoids are indicated on a GIS map shown in Fig. 6. Each cluster is marked in a different color and the optimal location (medoid) for siting biomass power systems are marked yellow (in the web version).

4.4. Analysis of the clustering results

For a detailed representation of the structure of k-medoid clustering results and their usefulness in drawing conclusions, results at the block level under current demand and supply values (Fig. 5) are expanded. Among the set of 37 optimal clusters, the biggest cluster with its medoid located at Naganahalli was selected for further analysis. The cluster consisted of 23 villages. The village names, their latitude and longitude, distance to the medoid and transportation costs are shown in Table 10.

From the medoid (Naganahalli), the nearest village (Huralibarasandra) and the farthest village (Shettikere) are at distances of 0.9 km and 7.7 km respectively. The total cost of one trip of transportation from all the sources to the medoid is Rs. 1827. The cost of local transmission and distribution, power generation and peak demand is shown in Table 11.

Table 12
Economic and design parameters of the installed capacity.

Sl. no.	Name of the item	Description
1	Fuel wood requirement	6579 tonnes
2	Crop residue requirement	1031 tonnes
3	Cost of biomass Fuel wood	Rs. 6.6 million
	Crop residue + cost of briquetting	Rs. 3.1 million
4	Installed system capacity	500 kW
5	System capital cost (at 0.87 Lakhs per kW + 25 Lakhs building cost)	Rs. 46 million
6	Annual labor costs (3 persons*Rs. 1500 per person*12 months	Rs. 54,000
7	Operation and maintenance cost (@10% of capital cost)	Rs. 4.6 million

Table 13
Summary of clustering results at block and district levels.

	Optimum number of clusters	Investment cost (million Rs.)	Inter-cluster ranges of installed capacity of biomass power system		Inter-cluster range of distances	
			Minimum capacity	Maximum capacity	Minimum distance traveled	Maximum distance traveled
Kunigal block – 2015 ET scenario with efficient devices	37 clusters	600	20 kW	886 kW	0 km	73 km
Kunigal block – with 2030 ET scenario	32 clusters	670	28 kW	1066 kW		
Tumkur district – 2015 ET scenario	96 clusters	5710	–		–	

In some of the villages, energy supply potential is not adequate to meet the demand (Tables 10 and 11). However, additional constraints were imposed in the algorithm such that the aggregate supply at the cluster level is always sufficient to meet the cluster aggregate demand during cluster formation.

A summary of the results, total crop residue and fuel wood requirement, cost of biomass and cost of local transmission and distribution is presented in Table 12.

To meet the cluster energy demand, annual biomass requirement was estimated to be 7610 tonnes at the rate of 1.3 kg/kWh. Within the cluster the total biomass potential was found to be 2053 tonnes (26%) of crop residue and 6579 tonnes (74%) of fuel wood. Fuel wood was preferred over crop residues because fuel wood does not require additional cost of briquetting. Therefore fuel wood was assumed to be used to meet the demand to the maximum extent possible and the remaining unmet demand for feedstock was assumed to be met through crop residue obtained from cropland.

4.5. Discussion

For both the medium-term and long-term scenarios, if the efficient devices are used in lieu of the conventional devices, the available supply of biomass is adequate to meet the power needs of the entire block. But the underlying assumption enforces that all the devices be of efficient technology type. The optimal number of clusters *k* has reduced with increase in demand and decrease in biomass supply (Fig. 6). This is because; the number of villages being constant, with decrease in biomass supply, the clustering mechanism distributes the limited resources among maximum number of villages there by reducing the clustering size. The clustering algorithm will try to accommodate more number of villages within each cluster. This increased the intra-cluster traveling distance eventually leading to increase in transportation and transmission costs. Clearly, with a projected 10% increase in demand in 2030, the total capital cost increased by 11%, but the number of clusters reduced. All other cost parameters remaining the same, the increase in capital cost could be therefore attributed to increase in the installed capacity within clusters rather than the clustering effect itself.

A consolidated representation of the results for various demand and supply values is shown in Table 13. Inter-cluster capacities range indicates the minimum and maximum installed system capacities among the clusters. The inter-cluster distance range indicates the minimum and maximum distance traveled within the clusters to transport biomass from the origin to power generation facility.

5. Conclusion

A major component of cost of delivered power through biomass route is the cost of transportation of biomass from source to power plant. It is therefore important to determine the optimal location when the potential locations for installing biomass power plants are many. A GIS-based *k*-medoid data mining algorithm adopted in the current study is effective for selecting suitable locations to install biomass power plants in rural regions. The current approach being generic in nature can be applied to all the regions where there is a potential for energy production from biomass. This approach helps to select sites to locate biomass power plants and allows for matching of biomass supply with the demand for power. The approach was validated for Tumkur region in India. The methodology of arriving at village-level demand for power was explained. Assessment of demand for power for villages belonging to various socio-economic categories was presented. The demand assessment was extended further by assuming changes namely; (i) Due to replacement of existing end-use devices with more efficient ones in 2015, (ii) Due to increase in consumption in the year 2030. Next, assessment of village-level biomass potential for power generation was presented under current growth rates and case in which the cropland was assumed to increase resulting in reduction of wasteland potential for biomass production. The objective of the current study was to determine the optimal locations for biomass power generation plants that match the total demand with the total supply options in the region. It was observed that the current biomass potential is adequate to meet the demand by using more efficient end-use devices. Using *k*-medoid clustering algorithm the optimal locations are arrived at by minimizing the total cost: a) cost of installation of power system b) cost of transportation of biomass and c) cost of transmission of power. Results were presented at block and district levels. Results of one cluster were discussed in detail to explain the mechanism of clustering. The nature of the current location–allocation problem with supply–demand matching constraints required an integrated approach to handle, model and analyze a large quantity of heterogenous input data. The current approach being generic in nature can be applied to all the regions where there is a potential for energy production from biomass. This approach helps to select sites to locate biomass power plants and allows for matching of biomass supply with the demand for power.

References

[1] Balachandra P. Dynamics of rural energy access in India: an assessment. Energy – The International Journal 2011;36(9):5556–67.
[2] Kaundinya DP, Balachandra P, Ravindranath NH. Grid-connected versus stand-alone energy systems for decentralized power – a review of literature. Renewable and Sustainable Energy Reviews 2009;13(8):2041–50.
[3] Balachandra P, Ravindranath D, Ravindranath NH. Energy efficiency in India: assessing the policy regimes and their impacts. Energy Policy 2009;38(11):6428–38.
[4] MNRE (Ministry of New and Renewable Energy India). Annual report, 'renewable energy in India – business opportunities'. New Delhi: Government of India; 2009.
[5] Klose A, Drexl A. Facility location models for distribution system design. European Journal of Operational Research 2005;162(1):4–29.
[6] Francis RL, McGinnis LF, White JA. Locational analysis. European Journal of Operational Research 1983;12(3):220–52.
[7] Current J, Min H, Schilling D. Multiobjective analysis of facility location decisions. European Journal of Operational Research 1990;49(3):295–307.

[8] Melkote S, Daskin MS. An integrated model of facility location and transportation network design. Transportation Research Part A: Policy and Practice 2001;35(6):515–38.

[9] Syam SS. A multiple server location-allocation model for service system design. Computers & Operations Research 2008;35(7):2248–65.

[10] Nema AK, Gupta SK. Optimization of regional hazardous waste management systems: an improved formulation. Waste Management 1999;19(7–8):441–51.

[11] Maniezzo V, Mendes I, Paruccini M. Decision support for siting problems. Decision Support Systems 1998;23(3):273–84.

[12] Alumur S, Kara BY. A new model for the hazardous waste location-routing problem. Computers & Operations Research 2007;34(5):1406–23.

[13] Han J, Kamber M. Data mining – concepts and techniques. In: Gray Jim, editor. The Morgan Kaufmann series in data management systems. Morgan Kaufmann Publishers, ISBN 1-55860-901-6; 2004. March 2006.

[14] Okabe A, Boots B, Sugihara K, chiu NS. Spatial tessellations – concepts and applications of Voronoi diagrams. 2nd ed. New Jersey: J Wiley; 2000.

[15] Fan B. A hybrid spatial data clustering method for site selection: the data driven approach of GIS mining. Expert Systems with Applications 2009;36(Part 2): 3923–36.

[16] Tumkur District Hand Book. Tumkur at a glance 2006–2007. Jilla Panchayat, Tumkur: Government of Karnataka; 2008.

[17] Natural Resource Data Management System (NRDMS) report 2009. NRDMS, Karnataka State Council for Science and Technology, IISc Campus; Bangalore.

[18] Brown SL. Guidelines for inventorying and monitoring carbon offsets in forest-based projects 1999. Manual Prepared for the World Bank.

[19] Sudha P, Subhashree D, Khan H, Hedge G, Murthy I, Shreedhara V, et al. Development of regional climate mitigation baseline for a dominant agro-ecological zone of Karnataka, India. Mitigation and Adaptation Strategies for Global Change 2007-07-01;12(6):1051–75. Springer Netherlands.

[20] Ravindranath NH, Hall DO. Biomass, energy and environment – a developing country perspective from India. Oxford University Press; 1995.

[21] http://www.cgpl.iisc.ernet.in [accessed December 2011].

[22] Deepak P. Distributed bioenergy systems for expanding rural electricity access in Tumkur district, India: a feasibility assessment using GIS, heuristics and simulation models. PhD Thesis, Indian Institute of Science (IISc); India: 2011, p. 125.

[23] Tanatvanit S, Limmeechokchai B, Chungpaibulpatana S. Sustainable energy development strategies: implications of energy demand management and renewable energy in Thailand. Renewable and Sustainable Energy Reviews 2003;7(5):367–95.

[24] Kumar A, Jain SK, Bansal NK. Disseminating energy-efficient technologies: a case study of compact fluorescent lamps (CFLs) in India. Energy Policy 2003;31(3):259–72.

[25] Ravindranath NH, Balachandra P. Sustainable bioenergy for India: technical, economic and policy analysis. Energy 2009;34(8):1003–13.

[26] Balachandra P, Kristle Nathan HS, Reddy BS. Commercialization of sustainable energy technologies. Renewable Energy 2010;35(8):1842–51.

[27] http://www.kredl.kar.nic.in/ [accessed December 2011].

[28] Singh Jagtar, Panesar BS, Sharma SK. A mathematical model for transporting the biomass to biomass power plant 2010;vol. 34(4). p. 483–8.

[29] Modernization of Draught Animal Power. Report. http://www.heritage-cartman.co.in; 2009 [accessed December 2011].

[30] Tripathi AK, Iyer PVR, Kandpal TC, Singh KK. Assessment of availability and costs of some agricultural residues used as feedstocks for biomass gasification and briquetting in India. Energy Conversion and Management 1998;39(15): 1611–8.

[31] Crisis Research. Impact analysis on CERC regulations. New Delhi, India: Central Electricity Regulatory Commission; 2010.