Applied Energy 96 (2012) 302–315

Contents lists available at SciVerse ScienceDirect

Applied Energy

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

Initial layout of power distribution systems for rural electrification: A heuristic algorithm for multilevel network design

Ayse Selin Kocaman [a,*], Woonghee Tim Huh [b], Vijay Modi [c]

[a] Department of Earth and Environmental Engineering, Columbia University, 918 Mudd, 500 West 120th Street, New York, NY 10027, USA
[b] Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC, Canada V6T1Z4
[c] Department of Mechanical Engineering and Earth Institute, Columbia University, 220 Mudd, 500 West 120th Street, New York, NY 10027, USA

ARTICLE INFO

Article history:
Received 16 July 2011
Received in revised form 18 January 2012
Accepted 14 February 2012
Available online 13 March 2012

Keywords:
Green-field network design
Rural electrification
Multi-level network

ABSTRACT

We describe the first heuristic algorithm that selects the locations and service areas of transformers without requiring candidate solutions and *simultaneously* builds *two-level* grid network in a green-field setting. The algorithm we propose minimizes overall cost of infrastructure costs; specifically the combined costs of transformers and the two-tiered network together by solving transformer location problem as well as network design problems in a single optimization framework. In addition, it allows one to specify different costs for the higher throughput lines upstream of the transformer as compared to downstream of the transformer. Simulations are carried out based on real-world spatial distributions of demand points from rural locations in Africa, specifically in places without any pre-existing infrastructure to test the algorithm and generalize the results.

1. Introduction

In a technological landscape that is altered by the emergence of off-grid and distributed approaches, there is a need amongst infrastructure planners to evaluate the costs of networked or grid approaches vis a vis off-grid approaches to be able to make rapid assessment of the progress in rural electrification.[1] The investment costs of networked approaches are more difficult to estimate than the costs of off-grid approaches because it takes into account both the spatial distribution of demand and the optimal placement of infrastructure to meet that demand. This paper through its algorithm provides a new methodology to estimate the cost of green-field networks rapidly and with high accuracy.

The algorithm we present in this paper combines the transformer location problem and the low voltage (LV) and medium voltage (MV) network design problem into a single problem and solves them in a single optimization framework. We propose a heuristic algorithm to design a two-level radial power distribution system. The *first level* includes the determination of the numbers, locations and capacities of transformers that feed an LV distribution network. The transformers represent load points for an upstream MV network and the MV network is also determined as a part of the first level. The *second level* includes the determination of the layout of the low voltage network between the transformers and the specified ultimate demand points. Note that the high voltage (HV) network (for that matter source points) further upstream of the MV network are assumed to be known.[2] One could have further generalized the problem to include the determination of the HV networks as well, making it a three level problem, but here we consider the HV network as pre-specified for simplicity.

The algorithm we propose does not require a set of candidate locations to be considered as transformer locations. The maximum service distance in a low voltage distribution network is also prespecified and determined from engineering practice. Given these costs, the demand points, the location of the HV network, and the maximum distance of the demand point from the transformer, the algorithm automatically finds the locations and service areas transformers as well as the LV and MV network layout with the goal of minimizing the total costs.

Understanding the cost involved with electrification is important in designing a proper smart grid structure. This algorithm can serve as a tool for network engineers and planners to make rapid assessments assisting them with (a) estimates of total cost of distribution, (b) layouts of initial designs and (c) breakdown of

* Corresponding author. Address: 918 Mudd, 500 West 120th Street, New York, NY 10027, USA. Tel.: +1 347 653 1964; fax: +1 212 854 3304.
E-mail address: ask2170@columbia.edu (A.S. Kocaman).

[1] The International Energy Agency estimates that the number of people who do not have access to electricity is nearly 1.6 billion and this number is expected to increase over the next 30 years unless investments in providing modern energy services are expanded significantly.

[2] The medium voltage network connects these transformers to electricity sources further upstream where they could either be sub-stations of a high voltage transmission network or power generating stations. Note that the costs associated with what we call here for convenience high voltage networks or transformations from high voltage to medium voltage are not considered here.

0306-2619/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2012.02.029

total costs into transformer cost and medium and low voltage line costs and giving them a good starting point for more detailed smart grid projects. The methodology we propose ignores transmission losses, load flow considerations and local topography and hence the proposed designs are not meant to replace detailed engineering analyses of grid rollout. Therefore, this tool should be considered as a guide for planning within utilities to be used with large datasets rather than a tool that provides every detail. However, for completeness possible extensions of the algorithm to include more details of power distribution systems are also discussed in Section 5.4.

The sections of this paper are outlined as follows: the remainder of this section provides background information including literature review; a more precise statement of the problem is given in Section 2; our approach to the problem is explained in Section 3; algorithm results are provided in Section 4; and a discussion of our algorithm is presented in Section 5.

1.1. Background

Electricity access is one of the most important components of rural developments. It has been shown that better living conditions in developing countries cannot be achieved without investments in electricity [1]. In rural areas where renewable energy resources are widely available, small off-grid standalone systems appears to be an attractive alternative [2]. Moreover, decentralized technologies seems to be more suitable for rural and remote areas due to the fact that it helps avoid long distribution lines with low load densities, underutilized transformers and losses in distribution. It also has been discussed that whether decentralized alternatives which use locally available resources provide more reliable supply of energy [2,3]. Thus, most of the earlier research aims to investigate primarily renewable energy alternatives and off-grid technologies [4–9]. It is worth noting that, although there has been a lot of attention to rural electrification projects, literature on the networked approaches is very limited. In this paper, an attempt is made on estimating the cost of rural networks to facilitate the rural energy decision making based on purely cost comparison without considering other consequences of off-grid and grid approaches.

The so-called power distribution system problem, in general, has been studied extensively in the literature [10–19]. Techniques developed in prior efforts for this complex problem usually divide the problem into sub-problems at each level and then solve each sub-problem separately using various optimization techniques [10–14]. These studies differ from each other in how they represent the problem components as well as in the algorithms utilized. None of these studies address the problem of designing both LV and MV networks in a single framework. However, dividing the problem into sub-problems and solving them separately reduces the probability of reaching an optimal solution and prevents us from seeing the effects of different cost parameters on the final network layout. The methods that have been proposed in the literature are based on either mathematical programming techniques such as Mixed Integer Programming, Branch and Bound Method [12,13,17] or heuristic algorithms such as Genetic Algorithm [11,18,19]. However, complexity of the models and the algorithms reduces their applicability to estimate the cost of networked approaches in rural electrification discussions when spatial distribution of a very large data set (demand points) is available. In addition, regardless of the solution methods, all studies mentioned here, except for [11], includes pre-assumption of candidate locations for transformers or feeders. These studies do not provide a method to update the candidate transformer locations during the search for an optimum solution. Therefore, the final feeder network is strictly dependent upon the initial selection of candidate locations. In practice, however, determination of candidate locations is not always a simple task, and if the methodology has to scale to a larger number of demand points, clearly the transformer locations should be an outcome of the optimization process.

2. Problem statement

Locations of ultimate consumers are called "demand points" in the rest of the paper. The cost parameters are (1) the cost per meter of LV line, C_{LV}; (2) the cost per meter of MV line, C_{MV}; and (3) the unit cost of a transformer, C_T. The maximum service distance, modeled as the radius of coverage of a transformer, is specified[3] and called $D_{\max}$. No similar constraint is placed on the length of the MV line from the source point. The unit costs are assumed to not vary with load, a clear simplification of the reality. In the same vein, each demand point is assumed to have the same load and the load is assumed not to change over time making the problem "static".

Distribution system is designed to be radial, to have one path between demand points and transformers, due to the fact that it is the most widely used form of distribution design and it is the cheapest and the simplest alternative compared to loop and networked designs [20]. In radial design, since there is only one path between demand points and transformers, power flow is certain and the system can be operated easily. The major drawback to radial feeder design is reliability. Any equipment failure will interrupt service to all customers downstream from it. However, low statistical rate of failure of equipment on the low voltage level makes the adaptation of radial systems easier [12].

Within the service areas of the transformers, the low voltage network is permitted to be multi-point, in that, in order to minimize costs the wire to a demand point further in distance can first go through one or more intermediate demand points. This architecture is called a "multi-point" LV network here (see Fig. 1b). Maximum distance capacity of an LV line is then defined as another design parameter and called $L_{\max}$ (i.e. the maximum LV line used to connect a demand point to the transformer directly or through other demand points should be less than $L_{\max}$[4]). $L_{\max}$ value should be used to limit the maximum total load on LV line and should be greater than or equal to $D_{\max}$ so that each demand point within the service area of a transformer gets connected.

Given the cost parameters and subject to the constraints described above, the desired outputs of the algorithm are:

- Number and locations of the transformers.
- Medium voltage (MV) network that connects a source point to the transformers; and
- Low voltage (LV) network between the demand points and transformers.

Our objective function is the minimization of total system cost, which includes cost of transformers, cost of low voltage and medium voltage networks. Schematic illustration of our problem formulation can be seen in Fig. 2.

3. Methodology

Given the difficulty of the problem, a heuristic algorithm is developed to place transformers and locate the networks. The algorithm relies on a "greedy" approach that starts with a stage that each demand point has one transformer (i.e. for n demand points, there would be n transformers) and iteratively decreases the number of transformers. Initially, transformers are connected to

[3] This distance would vary over the network with local geography and topography but is assumed constant here.

[4] $L_{\max}$ can also be considered as a constraint on distribution losses in LV level as the losses and wire lengths are linearly related.

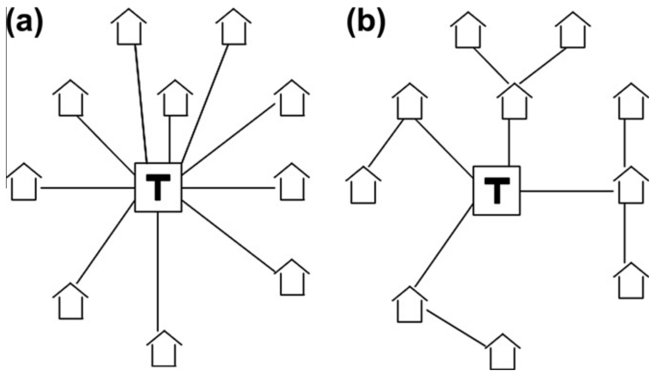

Fig. 1. Examples network configurations: (a) Point to point network (star configuration). (b) Multi-point network.

a prescribed single source point and to each other with a least cost medium voltage network; there is no LV network at this stage. The cost of this design is computed and provided as an initial upper bound to the cost of the network design since it is a feasible solution to the problem. It contains the maximum possible number of transformers and maximum length of MV line (which per unit length is more expensive than LV line). Then, one begins the process of eliminating transformers one at a time while removing some of the MV lines and adding new LV lines in the process to find a least cost network. The algorithm consists of the repeated applications of the following iterations:

- Search for the closest pair of transformers which can be replaced by a single transformer located at the centroid (center of mass) of the demand points without violating, D_{max} constraint.
- Build the MV network between the updated set of transformers and the source point (see Section 3.2 for details).
- Build the LV network between the demand points that are no longer served directly by transformers using LV line, ensuring constraint L_{max} (see Section 3.3).
- Compute the new overall cost.

The heuristic algorithm continues this iterative process until the number of transformers cannot be reduced any further without violating the D_{max} constraint (note that the solution with the least number of transformers is not necessarily the least cost since the design with the least number of transformers may have been obtained by adding more LV line length, and this trade-off may not be favorable to the total cost).

At this stage, all the computed costs during the transformer elimination process are compared and the least cost network design is selected. With one transformer replacing a "pair", and process repeated, one can think of the sets of demand points being served by one transformer as a "cluster". With this perspective the algorithm is an agglomerative algorithm, using a bottom-up approach to iteratively agglomerate (merges) the closest pair of points (see Section 3.1 for details).

The algorithm is further analyzed in its three components; (i) selecting the transformers to be removed, (ii) creating a MV network among the transformers and (iii) connecting the demand points and the transformers with an LV network. A flow chart of the algorithm can be seen in Fig. 3.

3.1. Locating the transformers

A capacitated agglomerative hierarchical clustering method is adapted to find the locations of transformers since it does not require pre-selected candidate locations for transformers. Note again that this is different than most of the work in the literature and is a must for our target demand points, households in rural Africa with little or none existing infrastructure.

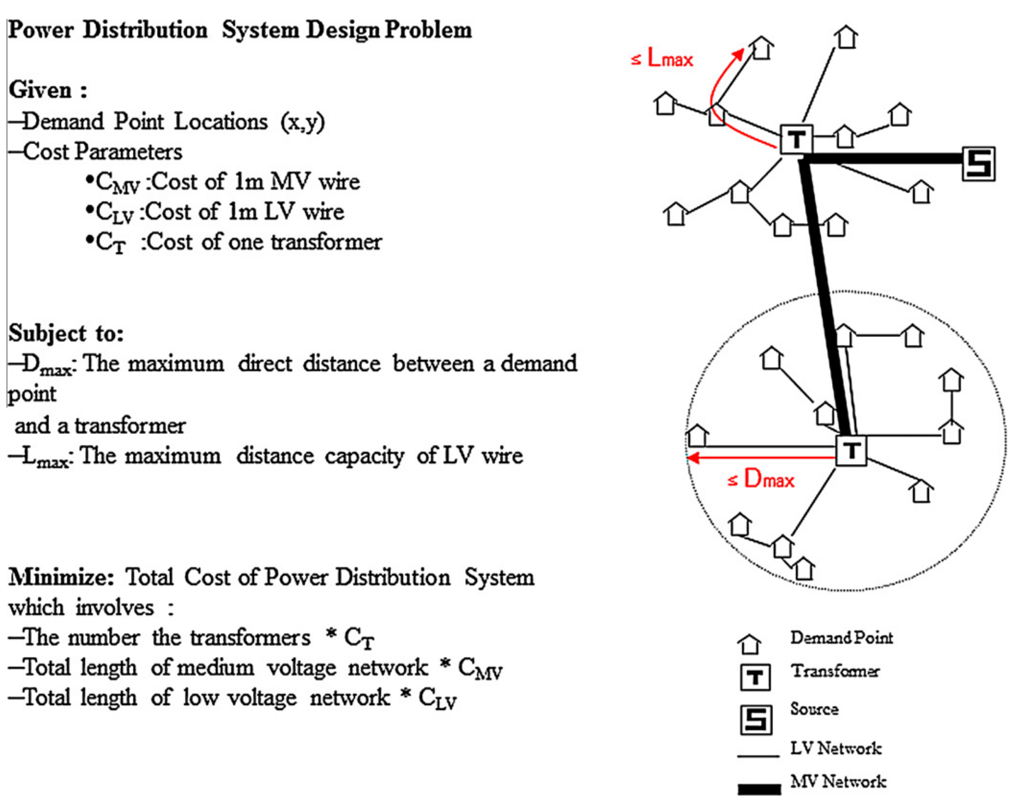

Fig. 2. Illustration of power distribution system.

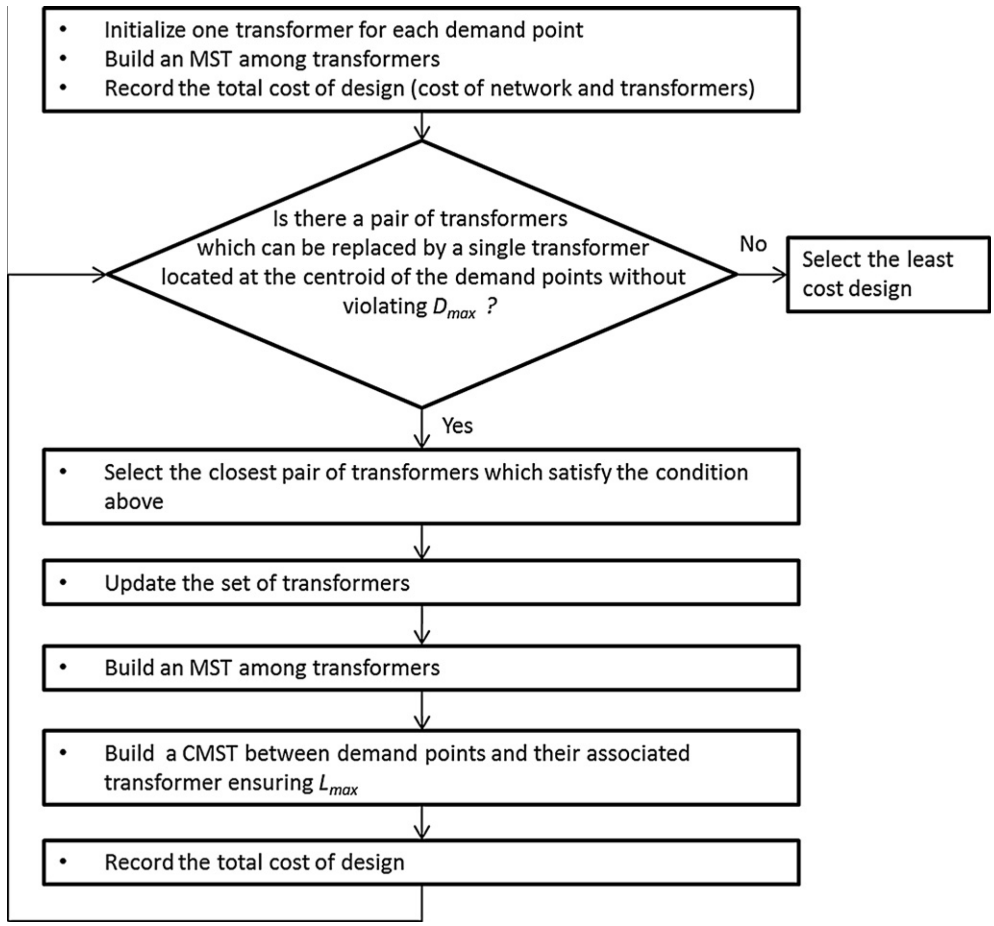

Fig. 3. Flow chart of the algorithm.

An agglomerative hierarchical clustering method starts with as many clusters as the number points to be clustered. At each step, the clusters are merged according to a rule and eventually only one cluster remains where all points are connected. In contrast, a *capacitated* agglomerative hierarchical clustering method has no assumption on the final number of clusters [21] and finds the minimum possible number of clusters that can be achieved with the given constraints. In clustering methods, many rules can be used depending on the problem definition. In our problem, to be able to incorporate the D_{max} conveniently, the closest pair based on the Euclidean distance between transformers has higher priority to be merged. The applications of clustering methods on similar problems using Euclidean distance can also be seen in [21–25]. Here, as opposed to stopping the process when the best possible agglomeration violates the capacity constraint (infeasible) [22], we choose the next best feasible agglomeration (if there exists one) as proposed by [24].

An illustrative example of our agglomerative clustering approach is presented in Fig. 4. In this example, we have five demand points and D_{max} is 2. Fig. 4a represents the initial configuration and Fig. 4b–d shows how the closest pairs of transformers which do not violate the capacity condition are connected one by one. No further change is possible in Fig. 4d since the agglomeration of the final two clusters would violate the D_{max} constraint.

3.2. Medium voltage line layout

At any iteration, when transformer locations are known, the problem is to find the least cost layout that connects the transformers and the given source point. This can easily be solved by using the well-studied Minimum Spanning Tree (MST) algorithms [26–29] which aim to find a tree (i.e. network containing no cycles) that spans all the points minimizing the total length of the network[5] with the guarantee of the exact optimal solution [29]. For the problem at hand, the points represent the transformers and the source point. Hence, goal is to find an MST such that all transformers and the source point are connected. Although there are different MST algorithms in the literature; in this paper, Prim's algorithm [26] is implemented as it has better running time performance for dense sites [30].

Prim's MST algorithm starts with choosing a starting point and adds the shortest segment of this point to the network. Until all nodes are spanned, the shortest segment emanating from the existing points on the network is added. The connections that would create cycles are avoided. Prim's algorithm is illustrated on an example in Fig. 5 on an example with one source and four transformers. We note that since the MST algorithm finds the optimal network, changing the starting point will not affect the result as all starting points will end up with the same network.

3.3. Low voltage line layout

As "clusters" emerge during each iteration cycle, an LV network needs to be laid out within each cluster. The total length of the connections is minimized while ensuring that the length of LV line between a transformer and the demand points are always less than a given L_{max} value.

[5] Since there is no constraint on the maximum length of MV line, the least cost solution is simply the least total length.

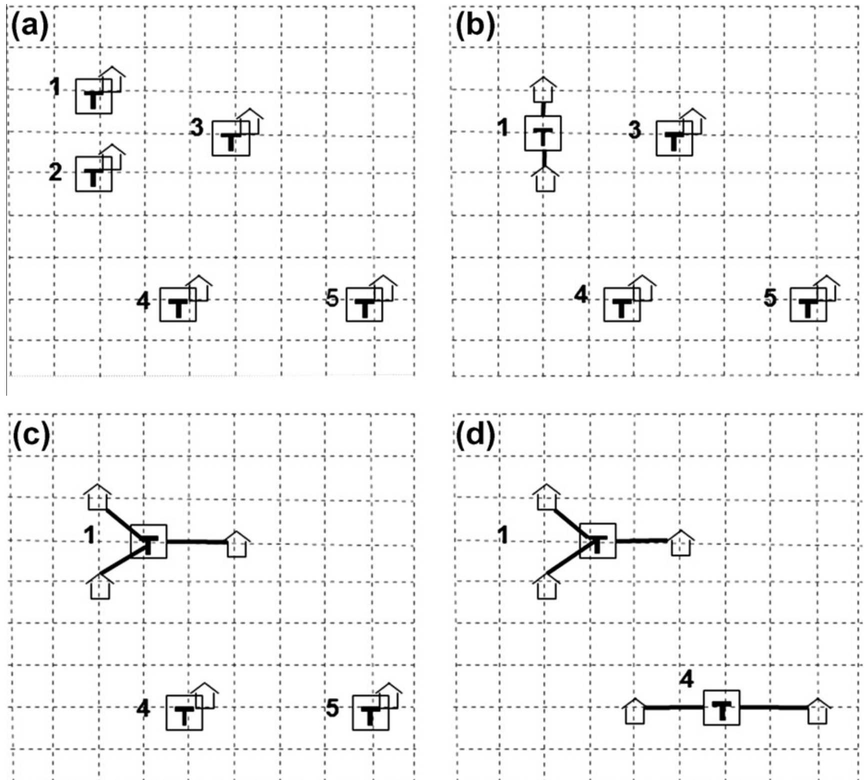

Fig. 4. Agglomerative clustering example: (a) Initial configuration where each demand point has one transformer. (b) The closest pair (1 and 2) gets connected. (c) Next closest pair is 1 and 3. (d) 4 and 5 are connected and no further change is possible without violating D_{max} constraint.

Constructing a least cost multi-point network (cost efficient compared to star configuration, see Fig. 1) is similar to the Minimum Spanning Tree (MST) problem, but with an additional L_{max} constraint. This extra condition converts the problem into a capacitated minimum spanning tree (CMST) problem. CMST is a minimal cost spanning tree which has a designated root (transformers) and a capacity constraint which ensures that the length of a sub-tree incident on the root does not exceed a certain distance (L_{max}). This problem is well studied for its applications to centralized communication design. Unlike MST, CMST cannot be efficiently solved in polynomial time. However, Essau and Williams [31] and Sharma and El-Baradi [32] developed heuristic algorithms to solve the CMST problem and Chandy and Russell [33] showed that these heuristics find near optimal solutions within 10% (often 5%) of the optimal solution.

Essau–William's heuristic algorithm is implemented in order to get the least total length LV layout within each cluster. A CMST algorithm starts with connecting all demand points to the transformers using the star configuration. The procedure to go from star configuration to a multi-point configuration is simply the successive iterations of calculating the trade-off values from removing the direct connections between the demand points and the transformer, as well as, adding indirect connections through their neighbors. At each iteration, the trade-off value is computed for every demand point and the largest trade-off value (i.e. the greatest improvement) that does not lead to a violation of the L_{max} constraint is used to update the network. The algorithm terminates when there is no further improvement possible.

An example of CMST algorithm with L_{max} value of 5 is presented in Fig. 6. It starts with the star configuration in Fig. 6a, and constructs the multi-point connections in Fig. 6b–d until there is no further change possible with the given L_{max}. Notice that although the trade-off value of point 1 is greater, its direct connection to the transformer is not removed due to the constraint.

4. Results on some Sub-Saharan Africa sites

The algorithm has been tested on household level data from nine sites in Sub-Saharan Africa shown in Fig. 7. The data were digitized from QuickBird satellite imagery of sites; details of the digitization method are discussed in [34]. For most of the sites, a QuickBird image covers an area of $10 \times 10\ \text{km}^2$ which covers a large representative area. Even though all structures do not necessarily correspond to households, we assume that each structure represents a demand point which needs to be electrified.

In rural electrification programs, the clusters of demand points and loads are small, so we use representative costs corresponding to the typical small 25 kVA transformer [35,36]. Cost parameters and other constraints consistent with rural electrification practice [37–40] are assumed to be:

D_{max} = 500 m.
L_{max} = 600 m.
C_{LV} = \$10/m.
C_{MV} = \$25/m.
C_T = \$5000

In Table 1, statistics and the outcome of our algorithm are shown for all sites. These results can help network engineers and planners estimate the number of transformers and the LV, MV line lengths easily for a particular location. In addition, by using our algorithm, they can also quantify their empirical observations. For example, from Fig. 7b–e, Mbola seems more dispersed than Bonsaaso (both have around 1000 households). Therefore the dynamics of their networks are expected to be different by observation. The algorithm indeed outputs 90 transformers for Mbola and 18 for Bonsaaso, and, the cost per household is more than 2.5 times for Mbola than for Bonsaaso. Similarly, Tiby (Fig. 7c) seems highly nucleated and this leads to shorter (cheaper)

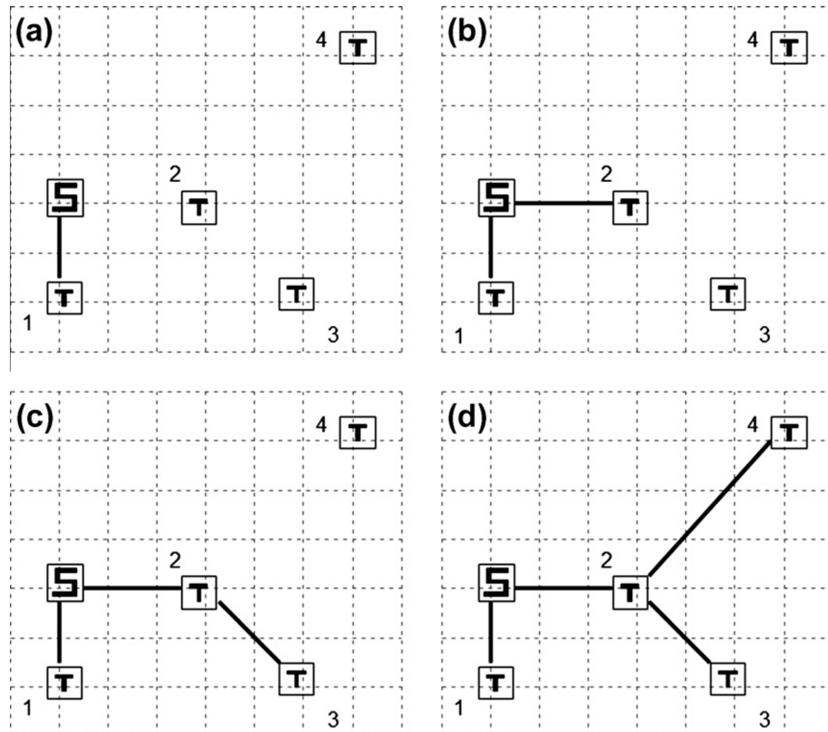

Fig. 5. A five-point (four transformers and a source point) example of Prim's algorithm: (a) S is the starting point and S–T1 is the shortest segment emanates from S. (b) S–T2 is the shortest segment to the existing network without creating loops. (c) T2–T3 is the next. (d) Finally T2–T4 gets connected and the MST is complete.

connections. Hence, from the algorithm results; Tiby has $691 cost per household which is the cheapest among all sites tested.

From the test results, it is also possible to reach some generalized conclusions. For all sites, overall transformer costs are between 8% and 15% of the total cost. In addition, for sites that has 1000–2500 household for 100 km^2, the bulk (~60%) of the total cost is composed of MV voltage lines. In the dense sites (Ruhiira, Mayange and Mwandama), there is a greater number of households per transformers and the total cost of LV line is comparable or even higher than the total cost of medium voltage lines despite the fact that the MV line is more expensive than LV line.

Furthermore, as a result of the rural electrification programs, some of the sites such as Mwandama, Pampaida and Mbola have already partial existing grid. When we compare our medium voltage network with the existing medium voltage line in these sites, we observe a highly good match. In Fig. 8, the overlap between the existing grid and the proposed grid is shown. This indicates that the planners may benefit from our algorithm in estimating the network structure and related costs also for these sites where there is an existing partial network.

5. Discussion

In this section, we first assess the sensitivity of the proposed networks in terms of the cost parameters for LV, MV lines and transformers (Section 5.1). This analysis can be significantly useful for policy planners to estimate the total cost fluctuations due to individual cost parameters. Next, considering that the complexity and the noise in the real data from Sub-Saharan Africa may complicate the understanding of our results, we also test our algorithm on simplified (simulated) datasets (Section 5.2) and see that we get consistent results for the artificial data. Then, we compare our algorithm with a sequential approach (Section 5.3) to see the relative performance and finally, we discuss the limitations and possible extensions of the algorithm for more detailed planning of power distribution systems and other parts of infrastructure problems such as siting of health or educational facilities (Section 5.4).

5.1. Network sensitivity analysis in terms of cost parameters

In our base case results, the cost parameters; C_{LV}, C_{MV}, C_T; are chosen as \$10, \$25 and \$5000 respectively. We perform a sensitivity analysis to understand whether there is a drastic change in the final network due to slight movements in these cost parameters. For the sake of generality, we present the general behaviors of the algorithm on uniformly distributed randomly generated points (1000 demand points on 10×10 km^2). At the end, we also present several runs of the algorithm on the data from Sub-Saharan Africa and verify the generalizations in real datasets.

5.1.1. Analysis with MV and LV line costs

First, we define a ratio, p, between cost parameters of MV and LV lines (i.e. $p = C_{MV}/C_{LV}$). When the transformer cost parameter, C_T, is set to zero, the differences in final networks help us understand the sensitivity of final network design to the ratio between cost of MV and LV lines. Initially every demand point has one transformer; therefore maximum amount of MV and zero LV lengths are used in the system. Fig. 9 shows how the total length of MV and LV lines change as the algorithm reduces the number of transformers from number of demand points to the minimum number of transformers subject to $D_{\max}$ constraint.

In Fig. 10a, we show the change in number of transformers as the p ratio is increased from 1 to infinity. We interestingly observe that there is a critical value p^* such that for all values less than p^* one transformer for each demand point (maximum number of transformer case) is the minimum cost design. For other values, greater than or equal to p^*, the solution includes almost the minimum possible number of transformers without violating the maximum distance ($D_{\max}$) constraint between demand points and transformers. Here, the critical p value (p^*) is observed around 1.70.

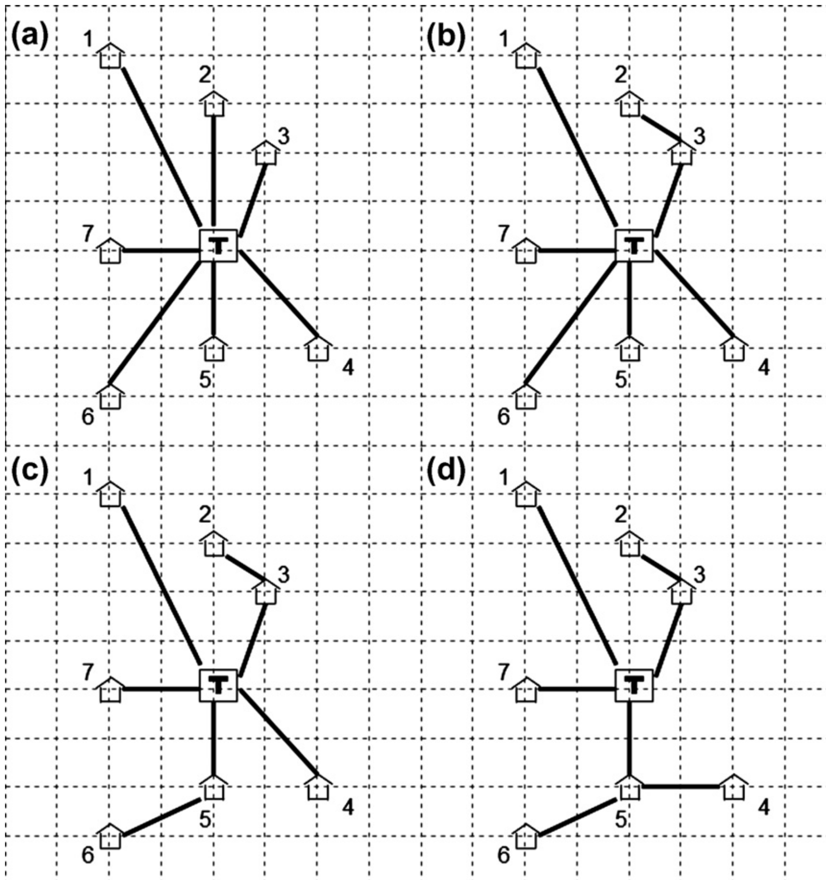

Fig. 6. A seven-point example of Essau–William's CMST algorithm: (a) Initial star configuration. (b) Maximum trade-off value is for Point 1 (2.23) however it violates $L_{\max}$. Point with the next best trade-off value (Point 2) is selected. (c) Next best trade-off value is for Point 6. (d) Final configuration is reached after Point 4 is connected.

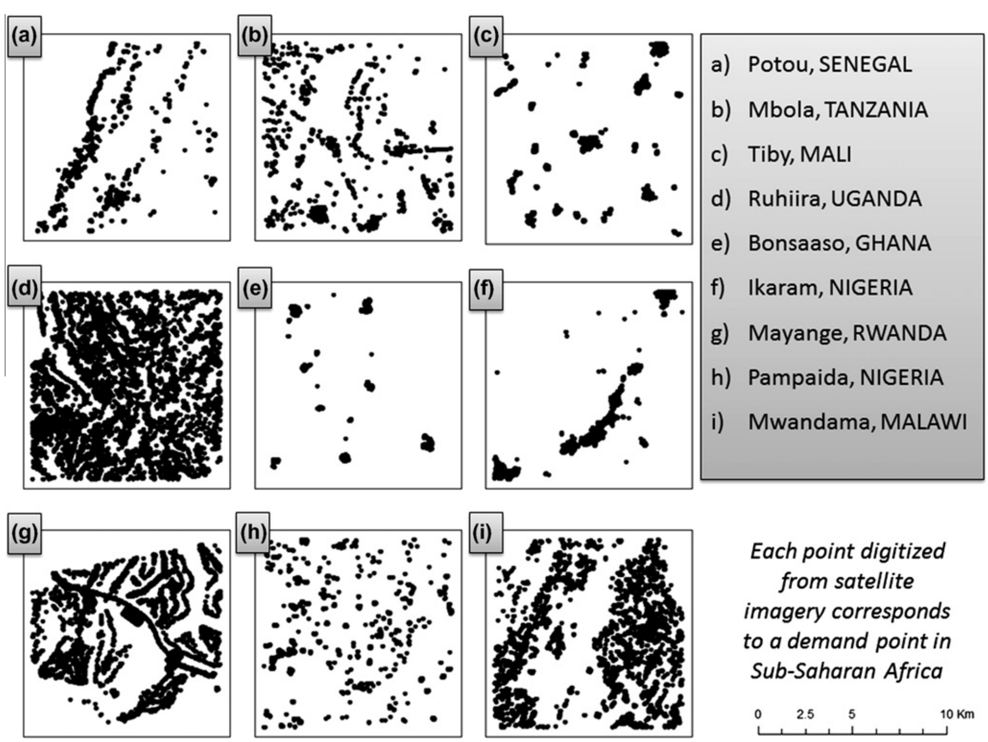

Fig. 7. Demand point locations for nine Sub-Saharan Africa sites digitized from satellite imagery.

Table 1
Algorithm results for nine Sub-Saharan Africa sites.

	Area of the image (km^2)	Number of demand points	Number of transformers	Total length		Average length per household		Cost distribution			Total cost (USD)	Average cost per household (USD)
				MV(m)	LV(m)	MV(m)	LV(m)	Transformers (%)	MV (%)	LV (%)		
Potou, SENEGAL	95.5	1797	71	57,644	63,742	32.08	35.47	14.6	59.2	26.2	2,433,515	1354
Mbola, TANZANIA	100	1175	90	70,762	81,320	60.22	69.21	14.8	58.3	26.8	3,032,239	2581
Tiby, MALI	100	2496	32	42,026	51,361	16.84	20.58	9.3	60.9	29.8	1,724,264	691
Ruhiira, UGANDA	100	6434	212	124,293	351,299	19.32	54.60	13.8	40.5	45.7	7,680,302	1194
Bonsaaso, GHANA	100	993	18	27,775	22,334	27.97	22.49	8.9	68.9	22.2	1,007,723	1015
Ikaram, NIGERIA	100	1484	33	33,572	46,417	22.62	31.28	11.2	57.2	31.6	1,468,479	990
Mayange, RWANDA	100	3909	114	68,746	185,987	17.59	47.58	13.7	41.4	44.8	4,148,506	1061
Pampaida, NIGERIA	100	1570	88	73,576	68,863	46.86	43.86	14.8	62	23.2	2,968,042	1890
Mwandama, MALAWI	100	4230	152	97,763	213,116	23.11	50.38	14.2	45.8	39.9	5,335,237	1261

To understand this sudden change further, we need to think about how the algorithm works in each step. The total cost is given by the following equation:

$$\text{Total Cost}_i = C_T \times i + C_{MV} \times \sum \text{MVLength}_i + C_{LV} \times \sum \text{LVLength}_i$$

And when we assume there is no transformer cost, it becomes

$$\text{Total Cost}_i = C_{MV} \times \sum \text{MVLength}_i + C_{LV} \times \sum \text{LVLength}_i$$

The algorithm compares the total cost values in determining the number of transformers. Therefore, the only way to decrease the number of transformers is if $\text{Total Cost}_i \leqslant \text{Total Cost}_j$, where $i < j$. This means

$$C_{MV} \times \sum \text{MVLength}_i + C_{LV} \times \sum \text{LVLength}_i \leqslant C_{MV} \times \sum \text{MVLength}_j + C_{LV} \times \sum \text{LVLength}_j$$

and

$$p\left(\sum \text{MVLength}_j - \sum \text{MVLength}_i\right) \geqslant \left(\sum \text{LVLength}_i - \sum \text{LVLength}_j\right)$$

Here, for small p values, the change in the transformer number is not profitable because it also makes LV longer and cost of the LV is not cheap enough to make above equation hold (see the black (darkest) line in Fig. 10d). Then algorithm results in one transformer for each demand point. However, after a critical p value ($p = 1.70$ in Fig. 10a) cost saving becomes possible by decreasing number of transformers within the allowed configurations. Once the algorithm favors less number of transformers, for all higher p values it goes all the way down to the minimum number of transformers subject to the $D_{\max}$ constraint because decrease in the MV length is faster than the increase in LV length; the fewer the number of transformers the less total cost (see Fig. 10b).

We perform the same analysis for some of the Sub-Saharan African sites and again obtain similar results; most notably, a sudden drop in the number of transformers. Table 2 summarizes the results. We observe that p^* differs based on the number of demand points and their spatial distributions. For example, for clustered sites (Bonsaaso and Ikaram) p^* value is smaller than the rest of the sites and p^* is highest for Mbola which is known with its dispersed settlement pattern. Since p^* is actually the ratio between the changes in LV and MV as the algorithm proceeds, it is expected to have smaller ratios for clustered sites where the decrease in total MV line is much faster than the increase in LV line for small number of transformers. Moreover, when $D_{\max}$ is set to 750 m, algorithm finds less number of transformers than the 500 m case and the sudden drop point is observed at smaller p^* values as expected.

5.1.2. Analysis with LV line and transformer costs

To be able to see the effect of transformer cost on the final network design, another ratio between C_T and C_{LV} is defined here as $q = C_T/(C_{LV} * D_{\max})$ and the cost of MV line is set to zero (since the cost contributions from C_T and C_{MV} are in the same way). Unlike the p ratio, $D_{\max}$ is introduced in the q ratio to make it dimensionless.

The total cost is given by

$$\text{Total Cost}_j = C_T \times j + C_{LV} \times \sum \text{LVLength}_j$$

To be able to decrease the number of transformers

$$\text{Total Cost}_i \leqslant \text{Total Cost}_j$$

$$C_T \times i + C_{LV} \times \sum \text{LVLength}_i \leqslant C_T \times j + C_{LV} \times \sum \text{LVLength}_j$$

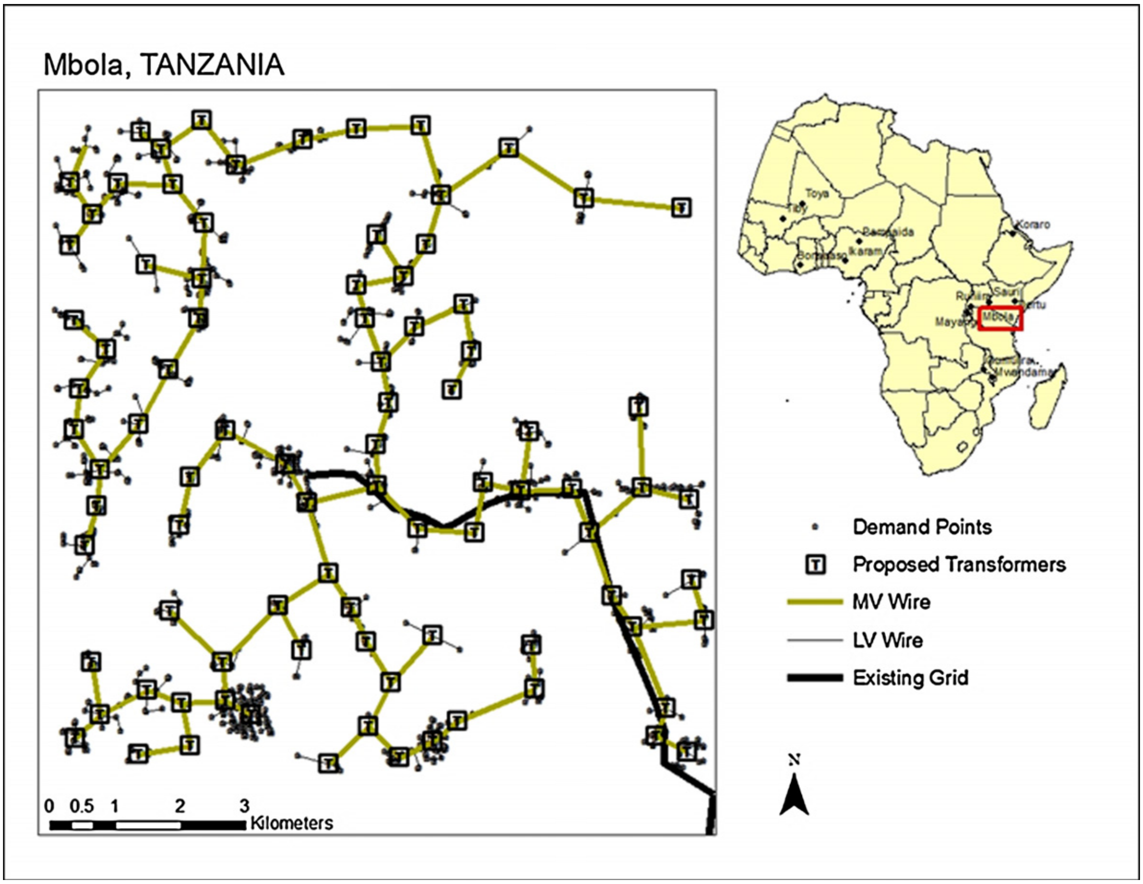

Fig. 8. Match between proposed network and existing grid. Proposed transformers, LV and MV networks compared to partial existing grid. Algorithm outputs 90 transformers for 1175 demand points.

$$\left(\sum \text{LVLength}_i - \sum \text{LVLength}_j\right) \leqslant qD_{\max}(j-i)$$

This time there is no sudden change and the number of transformers decreases gradually as q increases (Fig. 11a). The S-shaped behavior is due to the fact that closer nodes are connected first so the cost from LV length is initially smaller and the total cost is more sensitive to the change in transformer cost.

5.1.3. Analysis with all cost parameters

In Figs. 10c and 11b, we looked at more realistic cases where all the cost parameters are present. Fig. 10c shows p ratio analysis for different q values. For high q values, effect of S-shaped behavior of q dominates the sudden drop effect of p and we observe higher critical p values for small q values. In Fig. 11b, q ratio analysis with different p values is presented and for high p ratios the sudden drop effect of p ratio dominates the smooth decrease effect of q ratio.

In conclusion, depending on the price changes the design with minimum cost may change drastically. For the values ($p = 2.5$, $q = 1$) that we use for calculating the grid in the nine Sub-Saharan Africa sites, network generated by our algorithm is not sensitive. However, policy makers can use our algorithm as a tool to understand whether or not there is a possible design change with future prices.

5.2. Testing of algorithm on simulated data

Our results for African sites are specific to each site because all sites have different number of demand points and spatial distribution patterns. The noise and the variety in the data from Sub-Saharan Africa sites can make it difficult to draw conclusions from the results of the algorithm. This is why we test our algorithm on simulated data which provide more intuitive sense of what results would be. Same cost parameters and constraints are used for the base case run of simulated data ($C_{LV} = 10$, $C_{MV} = 25$, $C_T = 5000$, $D_{\max} = 500$, $L_{\max} = 600$).

5.2.1. Multivariate normally distributed randomly generated data

We present six $10 \times 10\,\text{km}^2$ artificial sites with 1000 demand points that are randomly generated using multivariate normal distribution (generalization of Gaussian distribution in two dimensions). To filter out the noise, we stretch out the points in two dimensions by increasing standard deviations of the multivariate

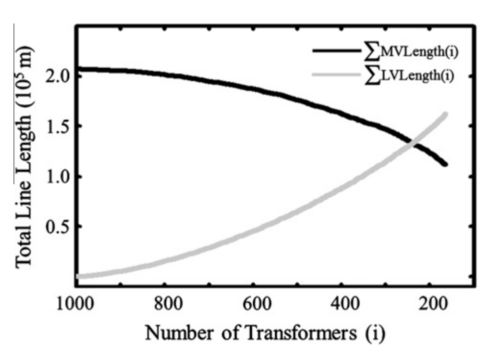

Fig. 9. Change in total length of MV and LV lines. As the algorithm decreases the number of transformers for 1000 uniformly distributed randomly generated demand points data within a $10 \times 10\,\text{km}^2$ area, total MV line length decreases, while total LV line length increases.

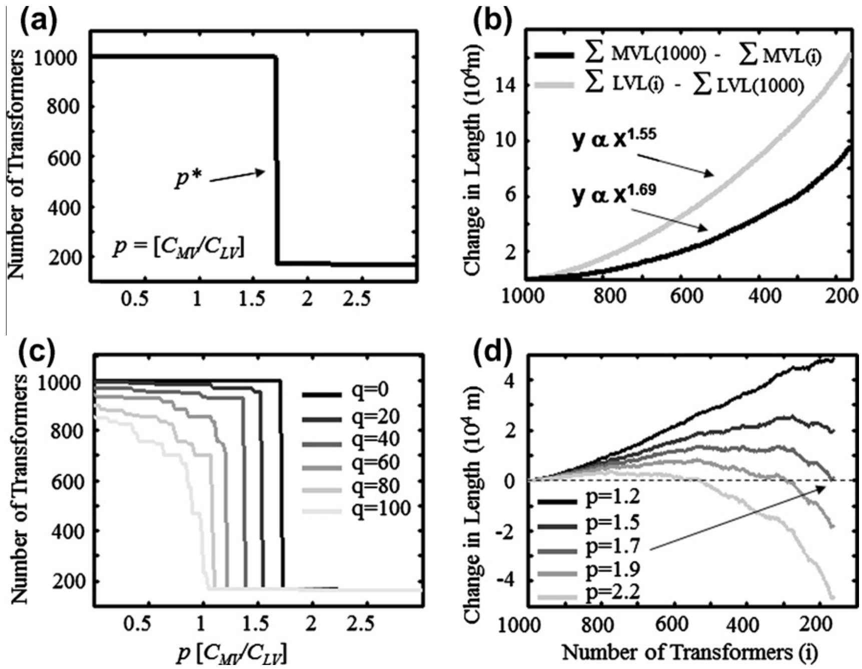

Fig. 10. Network sensitivity analysis: (a) The sudden change in the number of transformers as the ratio between C_{MV} and C_{LV} cost parameters increases. (b) Change in total MV and LV lengths from the 1000 transformer case. (c) The change in the number of transformers for different q ratios as p ratio increases (see Section 5.1.2 for q). (d) The difference between the curves in b with MV weighted with different p ratios.

normal distribution (from 250 to 1500). The generated sites are shown in Fig. 12a–c and e–g, and the results are presented in Fig. 12d and h. As expected, the number of transformers goes up with higher standard deviation (Fig. 12d) and as the number of transformers increases, the total MV line used increases, while the total LV decreases consistently (Fig. 12h).

5.2.2. Uniformly distributed randomly generated data

Next, we present six different sized areas (4×4, 6×6, 8×8, 10×10, 12×12, $14 \times 14\,\text{km}^2$) with again 1000 demand points that are generated randomly using uniform distribution (i.e. the likelihood of generating a demand point on any point of the site is same). Since these sites have different areas but the same number of points the mean distance between demand points is the highest for $14 \times 14\,\text{km}^2$ site and decreases as the site area gets smaller. The generated sites are presented in Fig. 13 and the results are summarized in Fig. 14.

To facilitate the quantitative comparison between sites, we refer to "average nearest distance" defined by Clark and Evans [41]. We calculate the average nearest distance using a Geographic Information System (GIS) tool for each different sized site and they are shown in the x-axis of Fig. 14a and b (60 for the densest site ($4 \times 4\,\text{km}^2$), around 260 for the largest site ($14 \times 14\,\text{km}^2$)). Due to the lower number of demand points within the radius of $D_{\max}$, number of transformers increases almost linearly (Fig. 14a) as the average nearest distance increases (same total number of demand points, same $D_{\max}$) and when we keep the ratio between the average nearest distance and the $D_{\max}$ same, we obtain the same number of transformers for each site (Fig. 14b).

In addition, we also test our algorithm on the same sites setting the $D_{\max}$ to various numbers between 50 and 1000 and present the results in Fig. 14c. As expected, the number of transformers increases as the $D_{\max}$ decreases and converges to the number of demand points. This is in general more sensitive in the dispersed areas making the slope steeper (slope in Fig. 14a) for lower $D_{\max}$ values.

Table 2
Critical p^* ratio for some of Sub-Saharan sites.

	p^*	$D_{\max} = 500$ m Drop From	To	Ratio	p^*	$D_{\max} = 750$ m Drop From	To	Ratio
Bonsaaso, GHANA	1.20	993	18	0.018	1.18	993	17	0.02
Ikaram, NIGERIA	1.31	1484	35	0.024	1.27	1484	30	0.02
Pampaida, NIGERIA	1.64	1570	288	0.183	1.63	1570	67	0.04
Mbola, TANZANIA	1.74	1169	93	0.080	1.64	1174	61	0.05
Random1000	1.70	1000	169	0.169	1.54	1000	91	0.09

5.3. Comparison of our algorithm with a sequential approach

As it was discussed in Section 1.1, to our knowledge, none of the previous studies in rural electrification and power engineering literature exactly matches with the objective of our paper. Complexity of the existing models [12,13,17] limits their suitability for the large data sets of demand points (up to 6500 households per site in our case). However; we can still compare our results with a relatively simpler sequential approach. In this approach, problem is divided into three sub-problems: transformer location problem, MV network design problem and LV network design problem. Then, each sub-problem is solved sequentially.

In the sequential approach, a greedy approach proposed for set covering problem[6] by Chvatal [42] is implemented to solve transformer location problem. Chvatal proves that the cost returned by the heuristic algorithm is at most $H(d)$ times of the cost of an optimal solution where $H(d) = \sum_{i=1}^{d} 1/i$ and d is the size of the largest set found by the algorithm. Once the locations of the transformer and their service areas are known, the MV and LV networks are found using MST and CMST algorithms, respectively. Then, total cost of transformers and cost of networks are calculated.

[6] Set covering problem aims to find the minimum number of sets subject to the constraint that each demand point should be covered (served) by a facility within a certain coverage criterion.

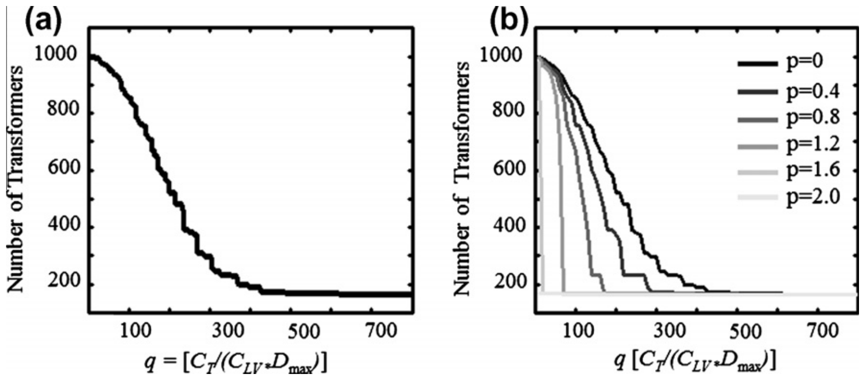

Fig. 11. Network sensitivity analysis: (a) The change in the number of transformers as the C_T/C_{LV} ratio increases. (b) The change in the number of transformers for different p ratios as q ratio increases.

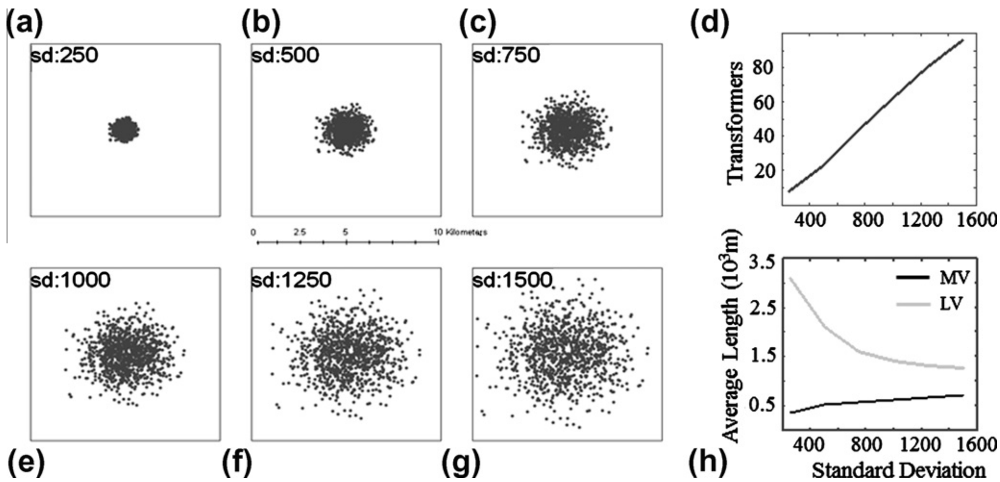

Fig. 12. Multivariate normally distributed random data and results: a–c and e–g, multivariate normally distributed $10 \times 10\,\text{km}^2$ sites with 1000 demand points. (d) The behavior of algorithm as the standard deviation increases. (h) Average LV line per transformer declines as the average MV line and the number of transformer increases.

Final number of transformers and total cost results for both sequential approach and our algorithm are presented in Table 3. It is shown that our algorithm tends to perform better than the sequential approach in terms of the total cost providing 4.5% improvement across all sites. We note that both approaches discussed here are based on polynomial time algorithms and provide reasonably good solutions to an NP-hard problem.

5.4. Limitations of our algorithm and possible extensions

The model described in this paper intended for obtaining quick estimates for the network structure and associated costs as a part of feasibility analysis, rather than being used for detailed implementation. Below we list a set of simplifications that we adopted:

- Our present model does not take some concepts into consideration such as power flow, power loss, voltage regulations, and transformer sizes for simplicity purposes.
- We use constant cost parameters throughout the entire system for transformers and wires. Thus, we limit our model to have single type of transformer (25 kVA) and single LV and MV technology (three-phase).
- We assume that our demand points are distributed with the similar loads and spatial characteristics and do not model bulk MV and non-homogeneous loads.
- Upfront capital costs dominate the operations and maintenance (O&M) costs and are treated as "overnight" costs (i.e. it is assumed that the entire system investment is made at once).
- We do not include in our model network control devices such as voltage regulators, and switches.

Our intent in this paper is to keep the model as simple as possible with the assumptions above while focusing on designing two-level network such that the overall distribution system is optimized in one framework as suggested by [43]. A multi-level network design problem which includes multi-point network configuration in one level has enough complexity; however our model can still be extended to include some of the important concepts mentioned above.

For example, based on the number of demand points served by a transformer, transformer size can also be determined by the model and different cost parameters can be used to calculate the total transformer costs. For example, instead of using 25 kVA for \$5000 each, we can prefer assigning cheaper 16 kVA transformers to the clusters which serve small number of points in sparse areas. Thus, we could decrease the total cost and avoid underutilized transformers.

Furthermore, it is also possible to put a capacity constraint (to incorporate the power losses and voltage regulations) in MV network even though this would make the problem even more complicated. Using the source as root, we can use CMST algorithm to design the MV level instead of using MST. For the LV level, $L_{\max}$ can be determined from the power loss and voltage drop constraints. Based on the number of transformers and the total length of the network, number of network control devices can be estimated and the cost of these devices can also be included in the objective function (total cost) of the system.

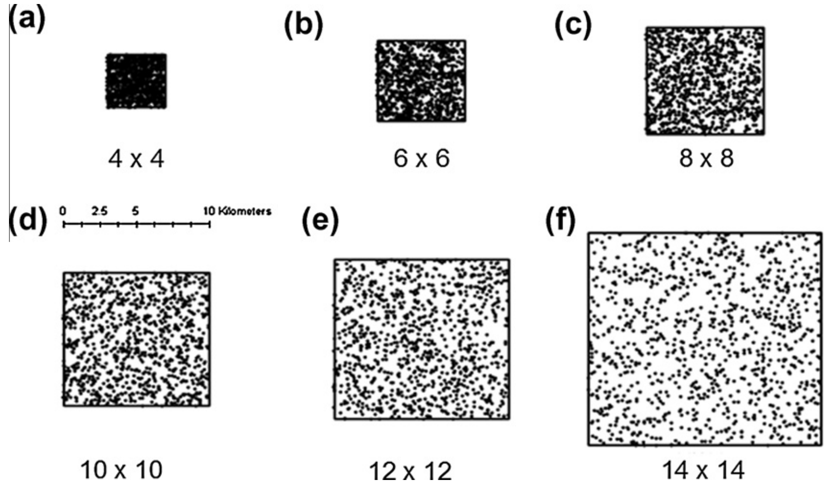

Fig. 13. Uniformly distributed randomly generated data. Uniformly distributed randomly generated points on different sized areas.

Another modification in the algorithm that could be done is that instead of using, multi-point network configuration in the LV level, star configuration can also be employed as it might be preferable for some situations. In this case, at each step of the algorithm CMST algorithm will be skipped and total cost of LV network will be found after calculating the direct distance between transformers and their associated demand points.

It is also possible to introduce time in our model with an assumption on the lifetime of a grid. For depreciation purposes, the lifetime of a grid is considered between 20 and 30 years but in reality it is usually more with a proper maintenance [20]. It is also acceptable to use a number between 1/8 and 1/30 of the capital cost for an estimate on the O&M costs in annual basis [20]. Thus, given annual costs and life time the grid, O&M costs can be discounted to the present value and be included in our objective function.

Furthermore, another potential application of our algorithm would be in facility location problem where given a set of household locations, planners are interested in finding how many schools or health facilities they need [44–48]. The unique situations of rural areas, in particular the sites in Sub-Saharan Africa as explained in the introduction; prevent many of the existing algorithms from being applicable as they usually require a set of candidate facilities as an input and there is no way to refine the candidate locations on the two dimensional coordinate system of the ground in these sites. By removing the cost of the network from the objective function, our algorithm can easily be modified for this purpose. This will simplify our algorithm to an agglomerative clustering algorithm that minimizes the total cost of opening facilities subject to the D_{max} constraint, which specifies the maximum distance between each household and the facilities (discussed in Section 4.1 in detail).

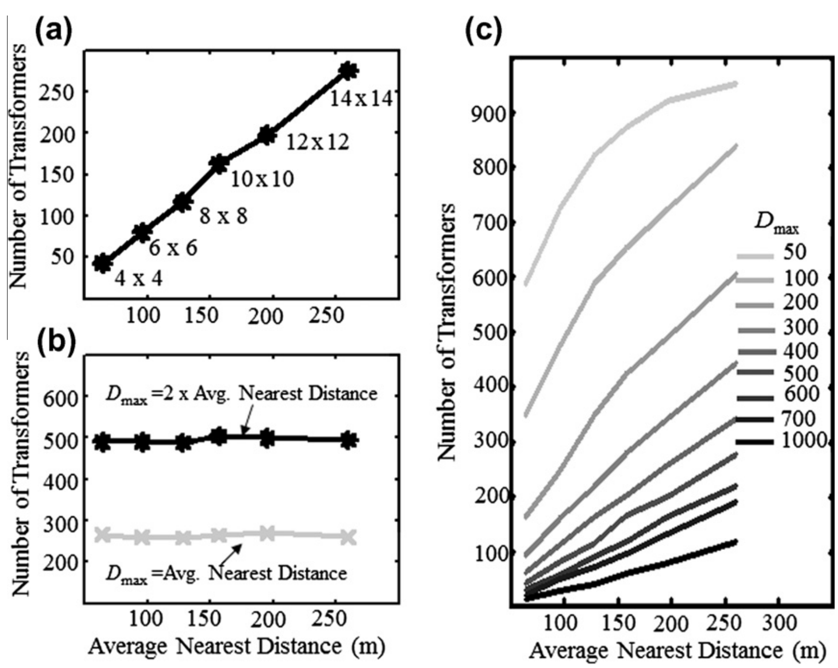

Fig. 14. Results for uniformly distributed random data. (a) Number of transformers outputted by algorithm for the sites which have same number of demand points but different average nearest distances. (b) The number of demand points when the ratio between D_{max} and average nearest distance is kept constant for each site. (c) Number of transformers versus average nearest distance for different values of D_{max} constraint.

Table 3
Comparison of our algorithm with a sequential approach.

	Number of demand points	Number of transformers		Total cost ($)		
		Our algorithm	Sequential approach	Our algorithm	Sequential approach	Difference (%)
Potou, SENEGAL	1797	71	70	2,433,520	2,534,654	−4.2
Mbola, TANZANIA	1175	90	96	3,032,250	3,180,655	−4.9
Tiby, MALI	2496	32	31	1,724,260	1,716,742	0.4
Ruhiira, UGANDA	6434	212	199	7,680,315	8,323,259	−8.4
Bonsaaso, GHANA	993	18	16	1,007,715	996,358	1.1
Ikaram, NIGERIA	1484	33	34	1,468,470	1,546,644	−5.3
Mayange, RWANDA	3909	114	107	4,148,520	4,327,479	−4.3
Pampaida, NIGERIA	1570	88	98	2,968,030	3,143,463	−5.9
Mwandama, MALAWI	4230	152	154	5,335,235	5,832,993	−9.3

6. Conclusions

A new heuristic algorithm for the design of two-level power distribution systems has been introduced. It has been presented that the algorithm finds the number and locations of MV/LV transformers without giving any candidate locations and finds a multi-point low voltage network between demand points and transformers. The algorithm is tested with the real household data digitized from satellite imagery of Sub-Saharan African villages and results are presented as an estimate for investment costs and financial requirements to support electrification problems. The proposed algorithm ignores transmission losses, load flow considerations and local topography. Hence it should be viewed as a quick tool which simplifies a complex problem and provides good starting point for decision makers and practitioners. However; our algorithm is flexible such that it can be simplified to other infrastructure problems (for example; facility location problem) or it can be extended to include more distribution system components such as transformer sizes.

Acknowledgments

Authors are grateful to the reviewers for their valuable comments on the content of this paper. Authors thank Serdar Kocaman, Susan Kum, Aly Sanoh and Daniel Soto for helpful discussions and comments on the manuscript.

References

[1] Modi V, McDade S, Lallement D, Saghir J. Energy and the millennium development goals. New York: Energy Sector Management Assistance Programme, United Nations Development Programme, UN Millennium Project, World Bank; 2005.
[2] WEC and FAO. The challenge of rural energy poverty in developing countries. London: World Energy Council; 1999.
[3] World Bank. Rural energy and development. Improving energy supplies for two billion people. Washington: World Bank; 1996.
[4] Herran DS, Nakata T. Design of decentralized energy systems for rural electrification in developing countries considering regional disparity. Appl Energy 2012;91(1):130–45.
[5] Bhandari R, Stadler I. Electrification using solar photovoltaic systems in Nepal. Appl Energy 2011;88(2). 458-46510.1016.
[6] Saheb-Koussa D, Haddadi M, Belhamel M. Economic and technical study of a hybrid system (wind–photovoltaic–diesel) for rural electrification in Algeria. Appl Energy 2009;86(7–8):1024–30.
[7] Urban F, Benders RMJ, Moll HC. Energy for rural India. Appl Energy 2009;86(1):S47–57.
[8] Silva Herran D, Nakata T. Renewable technologies for rural electrification in Colombia: a multiple objective approach. Int J Energy Sec Manage 2008;2:139–54.
[9] Kanniappan P, Ramachandran T. Goal programming model for sustainable electricity production from biomass. Int J Energy Res 2000;24:1–18.
[10] Hsu Y, Jwo-Hwu Y. Planning of distribution substations, feeders and sectionalizing switches using heuristic algorithms. Elect Power Energy Syst 1996;18(5):315–22.
[11] Hongwei D, Yixin Y, Chunhua H, Chengshan W, Shaoyun G, Jian X, et al. Optimal planning of distribution substation locations and sizes-model and algorithm. Elect Power Energy Syst 1996;18(6):353–7.
[12] Hamam YN, Hindi KS. Optimized design of low-voltage distribution networks a comprehensive algorithm. IEEE Trans Syst Man Cyber 1987;17:3.
[13] Aoki K, Nara K, Satoh T, Kitagawa W, Yamanaka K. New approximate optimization method for distribution system planning. IEEE Trans Power Syst 1990;5:3.
[14] Ramezani M, Falaghi H, Moghaddam MP, Haghifam MR. Genetic based algorithm for optimal placement of distribution transformers. Power Engineering Society General Meeting, IEEE; 2006. p. 5.
[15] Khator SK, Leung LC. Power distribution planning: a review of models and issues. IEEE Trans Power Syst 1997;12:3.
[16] Gonen T, Mahmoud AA. Bibliography of power distribution system planning. IEEE Trans PAS; 1983; Vol PAS-102, N0. 6.
[17] Gonen T, Ramirez-Rosado IJ. Optimal multi-stage planning of power distribution systems. IEEE Trans Power Delivery 1987;2(2):512–9.
[18] Turkay B, Artac T. Optimal distribution network design using genetic algorithms. Electr Power Compo Syst 2005;33:513–24.
[19] Miranda V, Ranito JV, Proença LM. Genetic algorithms in optimal multistage distribution network planning. IEEE Trans Power Syst 1994;9(4):1927–33.
[20] Willis HL, Scott WG. Distributed power generation. Marcel Dekker; 2000.
[21] Johnson RA, Wichern DW. Applied multivariate statistical analysis. 5th ed. Englewood Cliffs, NJ: Prentice Hall; 2002.
[22] Liao K, Guo D. A clustering-based approach to the capacitated facility location problem. Trans GIS 2008;12(3):323–39.
[23] Min HV. Consolidation terminal location-allocation and consolidated routing problem. J Bus Logis 1996;17:235–63.
[24] Barreto S, Ferreira C, Paixão J, Sousa Santos B. Using clustering analysis in a capacitated location-routing problem. Eur J Oper Res 2007;179(3):968–77.
[25] Lam M, Mittenthal J, Gray B. The impact of stopping rules on hierarchical capacitated clustering in location routing problems. Acad Inform Manage Sci J 2009.
[26] Prim RC. Shortest connection networks and some generalizations. Bell Syst Tech J 1957;36:1389–401.
[27] Kruskal J. On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 1956;7:48–50.
[28] Borůvka O. Příspěvek k řešení otázky ekonomické stavby elektrovodních sítí (contribution to the solution of a problem of economical construction of electrical networks) (in Czech). Elektronický Obzor 1926;15:153–4.
[29] Borůvka O. Ojistěm problěmu minimálním (about a certain minimal problem). Práce mor přírodově d spol v Brně III 1926;3:37–58.
[30] Skiena SS. The algorithm design manual. 2nd ed. Springer; 2010.
[31] Essau LR, Williams KC. A method for approximating the optimal network. IBM Syst J 1966;5(3):142–7.
[32] Sharma RL, El-Baradi MT. Suboptimal communications network synthesis. In: Proceedings of international conference on communications; 1970. p. 19.11–19.16.
[33] Chandy KM, Russell RA. The design of multi-point linkages in a teleprocessing tree network. IEEE Trans Comput 1972;21:1062–6.
[34] Zvoleff A, Kocaman AS, Huh WT, Modi V. The impact of geography on energy infrastructure cost. Energy Policy 2009;37:4066–78.
[35] Karhammar R, Sanghvi A, Fernstrom E, Aissa M, Arthur J, Tulloch J, Davies I, Bergman S, Mathur S. Sub-Saharan Africa: introducing low-cost methods in electricity distribution networks. ESMAP technical. Report 2006;104/06.
[36] Inversin AR. Reducing the cost of grid extension for rural electrification. ESMAP, Report 2000;227/00.
[37] ESMAP (Energy Sector Management Assistance Program, World Bank). Technical and economic assessment of off-grid, minigrid and grid electrification technologies, December, World Bank, Washington, DC. Technical paper 2007;121/07.
[38] KMOE (Kenya Ministry of Energy). Updating of the rural electrification master plan, vol. 1, prepared by DECON and re-engineering Africa consortium for the ministry of energy of Kenya and Ministry of Foreign Affairs of Finland, Nairobi; 2008.

[39] Parshall L, Pillai D, Mohan S, Sanoh A, Modi V. National electricity planning in settings with low pre-existing grid coverage: development of a spatial model and case study of Kenya. Energy Policy 2009;37:2395–410.

[40] Sanoh A, Parshall L, Sarr OF, Kum S, Modi V. Local and national electricity planning in Senegal: scenarios and policies. Energy Sustain Develop 2012. doi:10.1016/j.esd.2011.12.005.

[41] Clark PJ, Evans FC. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 1954;35(4):445–53.

[42] Chvatal V. A greedy heuristic for the set-covering problem. Math Oper Res 1979;3(4).

[43] Carter-Brown CG, Gaunt CT. Model for voltage drop apportionment in MV/LV rural electricity distribution. IEEE Africon 2002.

[44] Mahdian M, Ye Y, Zhang J. Improved approximation algorithms for metric facility location problems. In: Proceedings of the 5th international workshop on approximation algorithms for combinatorial, optimization; 2002. p. 229–42.

[45] Chudak FA, Shmoys DB. Improved approximation algorithms for a capacitated facility location problem. In: Proceedings of the 10th Annual ACM-SIAM symposium on discrete algorithms; 1999. p. 875–6.

[46] Chudak FA, Williamson DP. Improved approximation algorithms for capacitated facility location problems. Math Prog 2005;102:207–22.

[47] Barcelo J, Fernandez E, Jörnsten KO. Computational results from a new Lagrangean relaxation algorithm for the capacitated plant location problem. Eur J Oper Res 1991;53(1):38–45.

[48] Melcote S, Daskin MS. Capacitated facility location/network design problems. Eur J Oper Res 2001;129:481–95.