UTeM UNIVERSITI TEKNIKAL MALAYSIA MELAKA	**UNIVERSITI TEKNIKAL MALAYSIA MELAKA CENTRE FOR RESEARCH AND INNOVATION MANAGEMENT**
	PROJECT COMPLETION REPORT

PERHATIAN:

1. Laporanakhirprojekhendaklahdihantarke**CRIM dalamtempohdua (2)**bulanselepasprojektamat.

2. Silalampirkandokumen yang berkaitanseperti yang dinyatakan.

3. Senaraisemak:

No.	Perkara	Tandakan (√)
1.	Borang RND13 lengkap	☐
2.	Sertakan Template Profile Penyelidikan.	☐
3.	Salinan Softcopy (CD).	☐

A. PROJECT DETAILS

Principal Researcher : AziahBintiKhamis

Faculty/Centre : Faculty of Electrical Engineering

Project Title : A New Malaysian Reference MV Distribution Network Models for Energy Losses/Efficiency Model and Future Network Development

Project No.: RAGS/1/2015/TK0/UTEM/02/3

CoE: CERIA

RG :

Project Duration : Starts Date 01/12/2015 Final End Date:31/5/2019

Budget Approved: RM**72,900**Amount Spent Up to this period: **RM 45,497.35**

Project Members : MultidiciplinaryM culty

B. PROJECT ACHIVEMENT AND PERFORMANCE

OVERALL	**0 – 50%**	**51 – 75%**	**76 – 100%**
Work completion (please state %)			100%
Financial Utilization (please state %)		62.41	
RESEARCH OUTPUT			
I. PUBLICATION (Recorded at UTeMeRepository)	UTeM Press	Index Scopus/ISI	Others
a. No. of Journal Publication (Please attach the first page of publication)			
b. No. of Conference Proceeding (Please attach the first page of publication)			
c. No. of Other type of publication eg. monograph, books, chapters in book			

II. PROTOTYPE DEVELOPMENT	National	International
a. No. of Intellectual Property Rights		
b. Attended product exhibition & competition		
c. No. of Industrial Collaboration MoU/NDA/MoA		

III. HUMAN CAPITAL DEVELOPMENT

Number of Human Capital		On-Going		Graduated	
		Malaysian	Non-M	Malaysian	Non-M
1	PhD Student	0	0	0	0
2	Master Student	0	0	1	0
3	Undergraduate Student (SRA)	0	0	0	0
Total		0	0	0	1

IV. ASSETS AND INVENTORY PURCHASED (Cost more than RM 3000 per item)

1. High performance data processing system

2.

DECLARATION OF PRINCIPAL RESEARCHER

I acknowledged UTeM in providing the fund for this research work. (For University Grant only)

I certify that the information given in this final project report is true to the best of my knowledge.

Principal Researcher Signature :

Official stamp :
Name :
Designation :
Date :

ENDORSEMENT BY DIRECTOR OF CRIM

(Please state /comment on the performance of the project)

____

____

____

Signature & Official Stamp

Date

CRIM Revised date: 1 July 2018

LAMPIRAN A

RESEARCH PROFILE / TECHNICAL REPORT

Picture related to the research project

A New Malaysian Reference MV Distribution Network Models for Energy Losses/Efficiency Model and Future Network Development

PROJECT NO

Dr. Aziah binti Khamis[a]
PM. Dr Ir. Gan Chin Kim[a]
Dr. Khairul Anwar bin Ibrahim[a]
Prof. Dr Ir. Au Mau Teng[b]
Musa bin Yusup Lada[a]
[a]Centre for Robotics and Industrial Automation (CeRIA),
Faculty of Electrical Engineering,
UniversitiTeknikal Malaysia Melaka (UTeM),
76100 Durian Tunggal, Melaka, Malaysia

[b]Institute of Power Engineering, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

PROJECT DURATION
01/12/2015 TO 31/5/2018

10 MARCH 2019

EXECUTIVE SUMMARY / ABSTRACT

Technical losses (TL) in distribution network is causing substantial economic and financial losses annually to utility companies. For effective mitigation measures and justify its return on investment, it is necessary to determine the level, location and source of TL in the distribution network. In a typical MV distribution network, there could be hundreds and thousands of MV feeders scattered over vast geographical areas. These MV feeders and its key parameters are never exactly the same for every feeder. The characteristics of the feeder varies significantly within the same network a utility and even more, among other network in other supply zone. Modelling each and every feeder requires tremendous amount of resources to collect data, analyze and update the network parameters which also changes periodically. To simplify analysis, MV feeders are classified into a limited number of types based on the identified key parameters. This is known as representative feeders (RF) model. In particular, these limited number of MV feeders are generic feeder models which are technically representative of the common and actual feeders found in real distribution system. Due to confidentiality, actual network information are sometimes difficult to obtain from utilities. Thus, with RF models, it is possible to build a synthetic MV distribution network which can generally used by researchers to perform various analysis, without sensitive information on the actual network. Although numerous works to develop RF has been found, none of the existing model are representative of Malaysian distribution network. There are no existing literature found which studies and establishes the generic MV feeder characteristics on the Malaysian distribution network. Case study performed on a sample MV network in Malaysia shows the total losses are found in the range between 1% and 4%, which are in reasonably accurate as they are in close agreement compared with TL calculation data obtained from TNB. The main advantage of the generic analytical approach presented in this project is its flexibility and applicability in estimating TL of a radial MV network of different sizes, configuration and supply zones. The TL estimation model could be developed into a useful tool for power system planners to evaluate TL of any MV network on a periodic basis and establish the distribution network TL trend over time.

1. INTRODUCTION

As the electric utility industry enters an era of competitive market environment, uncertain world economics, and increasing fuel and equipment cost, there are rising concerns in the utility’s affordability in operating and expanding the distribution network to meet the increasing energy demand. Furthermore, sustainability agenda are also driving the distribution network to accommodate sustainable, clean and efficient energy supply and delivery system in order to reduce greenhouse gas (GHG) emission and reliance of fossil fuel energy sources.

One of the most cost-effective ways to meet increasing demand, reduce operating cost and environmental impact is to improve energy efficiency of existing power delivery system (transmission and distribution) by moderating energy demand and reducing wastage of electrical energy [1]. This electrical energy waste or “losses” degrades the network performance and economic efficiency, resulting in higher cost of investment.

Studies have shown that the average TL in distribution network worldwide ranges between 5% and 10% of the total energy delivered [2],[3],[4],[5]. Considering that the amount of energy delivered through a typical utility distribution network is substantial, a reduction of even a fraction of that percentage could translate into reducing tonnes of GHG and financial savings of hundreds of millions of dollars annually. For example, in Malaysia, it is estimated that a 1% reduction in distribution losses equals to approximately RM200 to RM300 millions of saving a year [5]. In the UK, the total distribution TL cost around £1 billion a year and excessive GHG if it is produced by fossil-fueled power generation plant [2].

With distribution TL being so significant, it is no surprise that energy regulatory bodies are enforcing new policies for utilities to reduce TL and promote energy efficient distribution networks. In Malaysia, under the new IBR (Incentive Based Regulation), the cost impact contributed by TL now becomes one of the key performance indicator for determining the level of received financial incentives in Tenaga Nasional Berhad (TNB) [6]. Therefore, distribution TL has to be prudently managed, reduced and controlled to a reasonably practical level. For strategic and effective TL management program, it is important that the utility company develop clear business processes with the right information in mitigating TL.

Power distribution feeders (and network) stretches over extensive geographical areas and have a myriad of characteristics associated with different voltage levels, circuit lengths, installed transformation capacity, number of load points, circuit construction types (e.g., underground, aerial, mixed) and load segments served. This poses great challenges to analyze the power and energy flow, including TL as these parameters vary significantly from each feeders and networks. Conventional method (e.g. load flow or classical loss formulae) to analyze energy flow and TL of distribution networks would typically involve detail network and load modeling using commercial software. However, due to the extensiveness of distribution networks, it would take a significant amount of resources to accurately model a large distribution network for the purposes of determining its TL.

Load or power flow analysis and consequently, the TL associated with each MV distribution feeder, are influenced by a number of key parameters such as feeder length, load segment composition and distribution, topology, load profile of load points and cable size. In a typical MV distribution network, there could be hundreds and thousands of MV feeders scattered over vast geographical areas. These MV feeders and its key parameters are never exactly the same for every feeder. The characteristics of the feeder varies significantly within the same network a utility and even more, among other network in other supply zone. Modelling each and every feeder requires tremendous amount of resources to collect data, analyze and update the network parameters which also changes periodically.

To simplify analysis, MV feeders are classified into a limited number of types based on the identified key parameters. This is known as representative feeders (RF) model. In particular, these limited number of MV feeders are generic feeder models which are technically representative of the common and actual feeders found in real distribution system. Due to confidentiality, actual network information are sometimes difficult to obtain from utilities. Thus, with RF models, it is possible to build a synthetic MV distribution network which can generally used by researchers to perform various analysis, without sensitive information on the actual network. Although numerous works to develop RF has been found, none of the existing model are representative of Malaysian distribution network. There are no existing literature found which studies and establishes the generic MV feeder characteristics on the Malaysian distribution network.

Representative distribution feeder and network

Generally, each distribution feeder and its load characteristics varies significantly within the same network and supply zone, and even more among other network in different supply zone. With approximately more than five hundred (500) Main-In-Take substation and Primary Distribution Substations (PDS) in Malaysia, detailed computation and analyses of individual feeders would require enormous resources [7]. Hence, this has been a major obstacle for any strategic planning activity.

Another challenge is, while detailed distribution feeder models do reside in utilities database, complete and detail feeder dataset and as-build drawing/model may not be readily available to researchers or difficult to obtain [8],[9]. This may be due to confidentiality restrictions or poorly documented of feeder datasets and models [10]. Therefore, a reasonable and less computationally expensive approach is to obtain a few samples of actual network or using available standard test cases model, such as the widely recognized IEEE test feeders [11]. The aim is that results performed on these sample network model would be meaningful and its impact could be inferred and extrapolated to represent the other and/or entire part of the distribution network [12],[13].

One problem arising is that, using results performed on either samples or standard test cases feeder model to infer the entire network population may not be accurate if the network model is not reasonably representative of the actual or as-built network [14]. For example, as the IEEE test feeders is purely adopted based on the North American (U.S) power system, it is peculiar to distribution feeders from other countries with different network settings. Meanwhile, although different network in the same utilities might have the same type of equipment and component ratings, the use of samples of network might not be adequately representative of other network at different supply zone as each feeder are never similar.

In order to mitigate this impediment to research, there are growing research works found to generate/synthesize “generic distribution feeders”, which is closely similar to the real network and could be openly distributed to researchers. These feeders were constructed based on the statistical analysis performed on real distribution feeder models obtained from utilities. By analyzing the characteristics of existing radial distribution feeders, it is possible to construct a small number of synthetic feeder models with similar fundamental characteristics of the real as-built system, but without the sensitive utility specific information [9],[15],[16]. This is also known as representative feeders (RF). In most literature, these generic characteristics of RF is normally used to construct a complete distribution network, known as Representative/Reference Network (RN).

The adoption of RF and/or RN model brings numerous benefits. It facilitates researchers to better understand the characteristics of the actual and location-specific network population, and to simplify any study (either technical or economic) needed to be carried out on it [15]. These generic model also reflect the main cost and other significant aspects of network operation [16]. The resulting models then serve as a reference or benchmark for assessing the overall efficiency, i.e. the cost and quality of service, of the corresponding real networks. In other words, since each model can link its characteristics and behaviors to the population it represents, the results from any study performed on them (e.g. load flow and reliability) can confidently extrapolated/inferred for the entire feeders population in the network. Also, such a study is more efficient and less rigorous than analysis of every feeder in each network in the entire distribution system.

Escalating social, economic and environmental pressures are forcing utility companies to improve its network energy efficiency by reducing its distribution TL. However, before any mitigation measures are taken, it is necessary to determine the level, location and source of TL in the entire distribution feeders for a more targeted approach to reducing TL and justify the return on investment. This, however, poses a significant challenge as typical distribution feeders are large in number and have myriad of characteristics. Evaluating the energy flow and its associated TL in each feeder and transformers at every corner in the entire distribution network requires large amount of resources. Therefore, developing a set of generic feeder or network model could potentially be used to effectively estimate system wide TL as its results can be inferred/extrapolated to the entire distribution system it represents using appropriate weighting factors [13].

Overview of RF/RN development method

From literature, development of RF/RN involves three main stages, as shown in Figure 1 [14],[16],[17]. Note that, this overall process is nearly similar to load profiling methodology discussed previously– the difference is on the type of analyzed data. First, it involves collecting feeder data on all feeders for the entire population. The issue of data availability, quality and data consistency are found as one of the largest issues that had to be addressed in order to ensure acceptable quality of RF [9],[10].Therefore, data cleansing, verification and validation is required to eliminate erroneous or large outliers in the data.

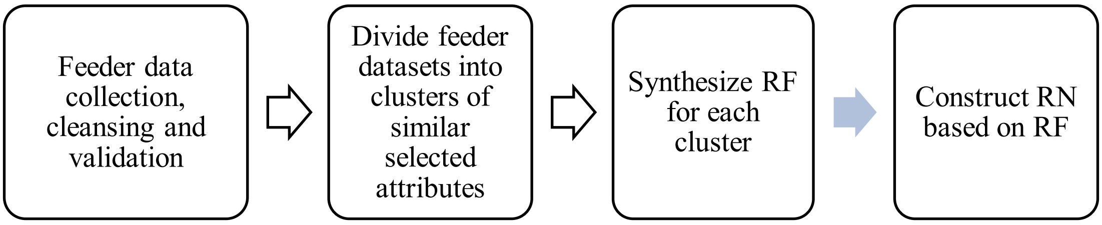

Figure 1 Process of establishing RF and RN

Next, using various statistical methods available (e.g. clustering), the feeder's population are "grouped/classified" into separate and optimized number of clusters, distinguished based on selected feeder attributes. The defined range of the feeder attributes determines the similarity. The selection of feeder attributes for clustering depends mainly on: (i) the available data, (ii) type of study to be conducted. For example, for load flow studies, the attributes include feeder type, impedances and length, load characteristic and voltage level [15]. Although there is no limit to the number of feeder attributes to be used to classify feeder cluster, the process becomes more complex and difficult if the number of attributes becomes excessive [17].

Examples of clustering approach used are classical k-means [14], using k-means cluster with discriminant analysis and cluster optimization [14], cluster analysis using decision tree technique [18] and k-means with cubic clustering criterion [19]. The result will be that, feeders in a given cluster have similar attributes to each other, and dissimilar attributes from feeders in other clusters. The number or percentage (with respect to total feeder datasets) of system feeders in each cluster is also determined.

From each cluster, RF model is synthesized to represent as “generic” or ”average” feeder model, i.e. it has close resemblance with the rest of the network/feeders within its group/cluster. Finally, RN is constructed based on established RF. Additionally, the quality of the clusters (optimum number and similarity) are validated based on statistical indices, such as Variance Ratio Criterion, Similarity Matrix Indicator and Global Silhouette Coefficient [15]. However, some research works perform validation by comparing power system performance studies (e.g. load flow, reliability, losses and capacitor placement) of RN against the actual network it represents [14],[16],[20],[21]. Figure 2 illustrated the overall development framework of RN. Note that A, B, C and D are RN models, which is used to construct the corresponding RN models.

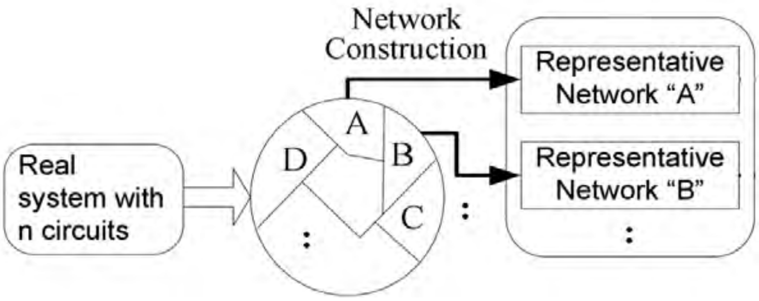

Figure 2Representative network development framework

Research works in representative feeder/network model

Earliest work in development of representative distribution feeders was carried out in 1985 by Willis, et al. at Westinghouse [14], where classical k-mean clustering algorithm was used to synthesize twelve (12) RN for the U.S. power distribution system. Each RN is an average feeder of each cluster. It was demonstrated that, performing analysis on RN yield reasonably accurate prediction of the behavior of the entire distribution systems, which includes voltage drop, losses and optimal number of capacitors. The resources (in man-hours) required to perform analysis using RN is approximately 100 times lesser that of those required for the analysis on entire actual system population. The results of each RN are extrapolated to the entire system based on the total load of all feeders in its cluster with an average error of less than 5%.

The Pacific Northwest National Laboratory (PNNL) conducted a study distribution feeders in the U.S. to develop a set of twenty-four (24) prototypical feeders, using k-mean clustering method to facilitate the analysis of new Smart Grid technologies [9]. A complete report can be found in [13]. The classification are based on voltage level, location of areas served and the network kVA density, i.e. light, moderate or heavy network. The RN model was determined by selecting one of the individual feeder, within a cluster, that is closest to the average center of each cluster. However, an extensive data set based on 37 parameters for each feeder of interest was required, that may prove difficult to collect in many, if not all utilities database. An example of PNNL's RN application can be found in [22] where the RN is used for evaluation of conservation voltage reduction (CVR) across the U.S. distribution network. Still in the U.S, another work by EPRI and California Solar Initiative Project classify twelve (12) RN using k-means clustering based on ten (10) feeder parameters for effective screening of new PV projects application[19]. Cubic clustering criterion is used to determine optimum number of clusters to accurately cluster the feeders. The network model is also publicly available at the GridLab website (http://www.gridlab.org).

In the UK, distribution RN model known as "UK Generic Distribution System" (UKGDS) specifies thirteen (13) representative (six extra high-voltage and seven high-voltage) feeder models using "decision tree" technique [20],[21]. UKGDS models are used for numerous reasons such as in network pricing mechanism, identify efficient investment strategies, study on network performance, evaluate demand management and energy storage, and more [17],[23],[24],[25]. The UKGDS model are characterized based on relevant disaggregation parameters, as shown in an example in Table 1. However, the detail on the implementation of "decision tree" approach is unavailable. An example of UKGDS model is shown in Figure 3. Although no statistical performance index was found to validate the quality of clusters, the RN performance were validated through comparing the reliability assessment results performed on the RN model and the actual network it represents – with average mismatch of less than 10%.[20]

Table 1 Example of feeder disaggregation parameters for UKGDS [20],[21]

No.	Area Served	Circuit Length	Customer Density	Construction type	Topology	Size	Branch Voltages
1	Rural	Long	Low	Overhead	Radial	Small	33
2	Rural	Long	Low	Mixed	Radial	Large	132/66/33
3	Suburban	Medium	Medium	Mixed	Radial	Large	132/33
4	Suburban	Medium	Medium	Mixed	Meshed	Small	132/33
5	Urban	Short	High	Underground	Meshed	Small	132/33
6	Urban	Short	High	Underground	Radial	Large	132/33/11/6.6

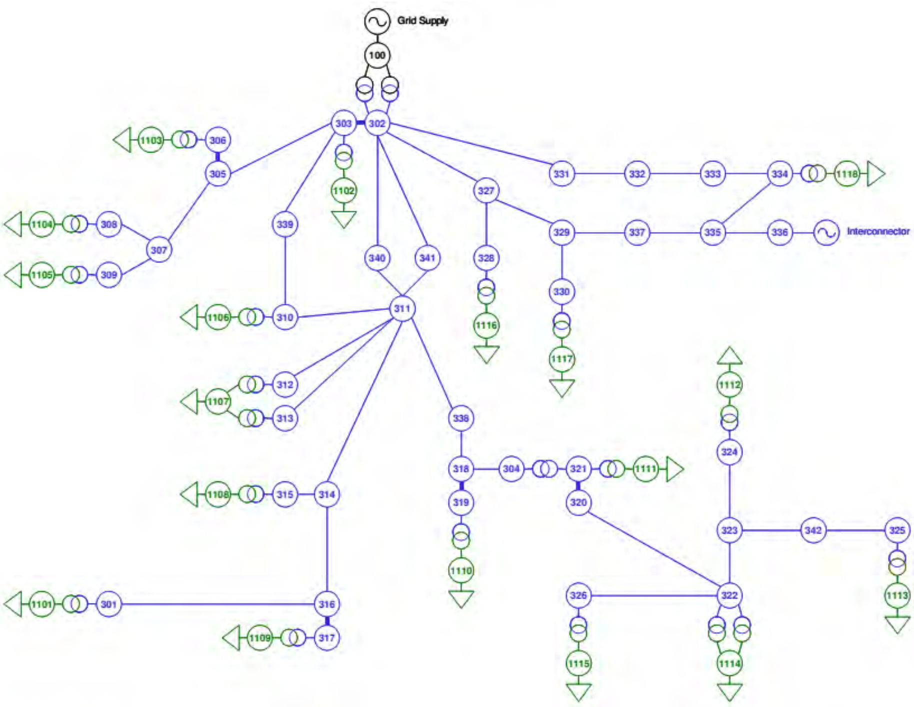

Figure3 Example of UKGDS RN model[20],[21]

In Australia, nine (9) MV and eight (LV) RN feeders were developed for the distribution network in the area of Perth for Smart Grid deployment using reduced dataset for Smart Grid deployment [18],[26]. The methods combine clustering and discriminant analysis technique for optimum clustering. The main contribution of the method is that, it relies upon only six (6) variables which are highly meaningful from an engineering perspective and readily available in most if not all distribution companies.

Authors in [15] developed eleven (11) representative LV feeders for North West of England with a high statistical accuracy based on several clustering performance index. Application for sensitivity of photovoltaic (PV) hosting capacity assessment is presented. It investigates and compare different clustering methods, which includes hierarchical clustering, k-medoids++, improved k-means++ and Gaussian Mixture Model (GMM); concluding that the improved k-means++ and GMM showed the best performances.

In Spain, a large scale and detailed RN model for large scale distribution planning with GIS and street map information is developed [27]. It serves as comprehensive planning tool to establish efficient costs distribution and incentive regulation for different DNO operating in different regions. It uses heuristic optimization algorithm which involves the minimum spanning tree and a branch-exchange technique. However, the works involves huge requirements regarding input data, for example, the geo-referred position of every single connection point, customer, and DER.

In Italy, in a project called ATLANTIDE, three (3) RN representing urban, rural and industrial supply zone were developed to to support case studies/scenarios of different emerging technologies integration (e.g. DER, DESS, active distribution network) [28],[29]. However, no specific type of clustering and statistical performance index method was mentioned. An example of the Italian RN are shown in Figure 4.

An attempt to develop RN was also found for distribution network in India for network pricing and Smart Grid [10],[16],[30],[31]. The adopted method is using “decision tree”. The methodology is based on repetitive stages of disaggregating real/actual network feeders into sub-groups based on certain “disaggregation parameters”, such as peak demand, voltage level and location of areas served (e.g. urban, or rural). The process is illustrated in Figure 5. Each feeder is classified depending on its conformance with the specified range of the disaggregation parameters. At each step of evaluation, the actual feeder is allocated to a sub-group based on the parameter related to that step. In the consecutive step, the feeders in each sub-group are further segregated. Each feeder finally joins a group that meets all the parameter criteria of that group. This process defines the feeders in each group. Each group developed during the disaggregation process contains feeders that are similar in terms of parameters, though not exactly identical. It can be seen that, although this process resembles clustering approach, it relies on simple statistical analysis and does not require complex and automated clustering algorithm like the ones presented in [14], [9],[15],[18],[26] and [27] earlier. The Indian RN model is show in Figure 6.

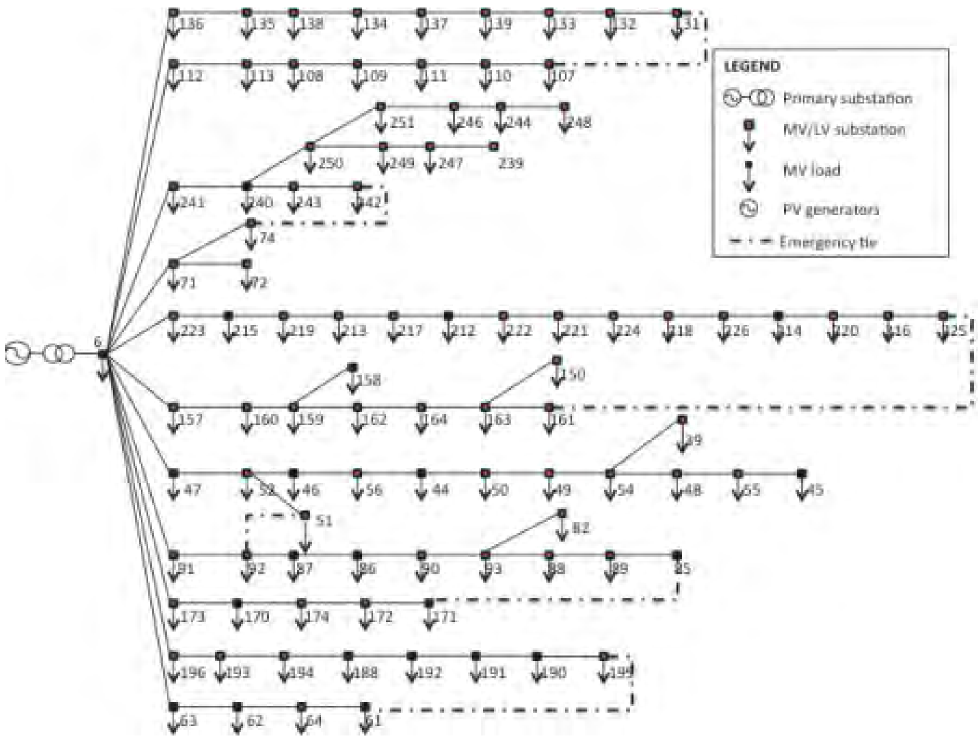

Figure 4 Example of RN for Italian urban distribution network [28],[29]

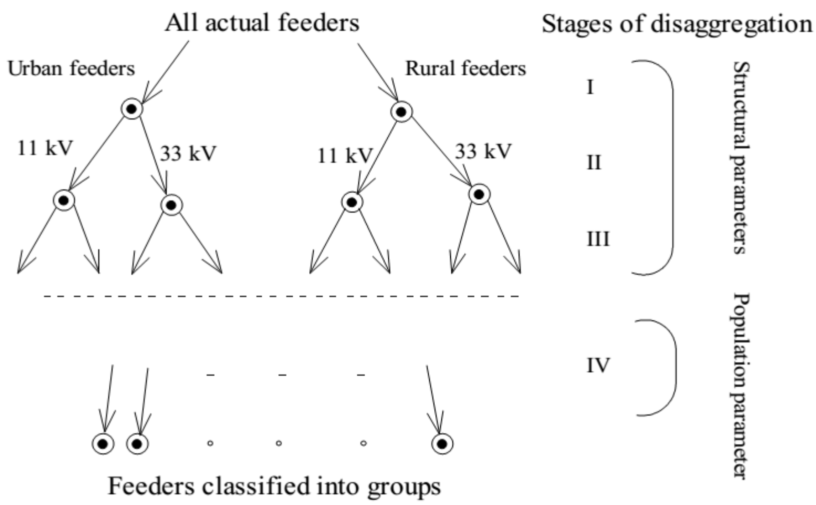

Figure 5 Feeder disaggregation process to develop Indian RN [16],[10]

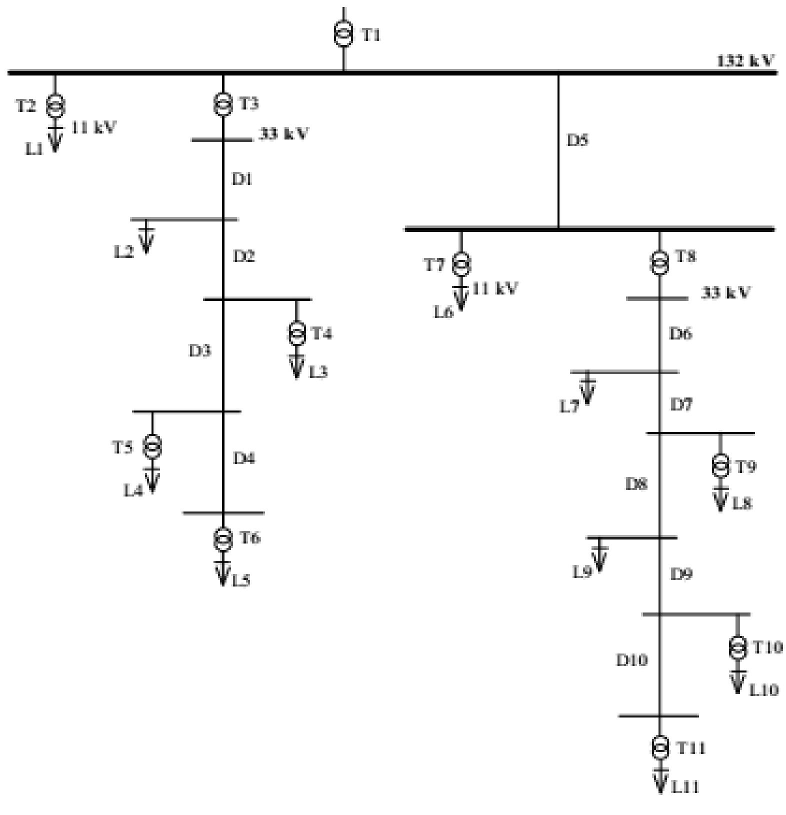

Figure 6 RN for Indian urban distribution network

Development of Malaysian distribution RF/RN model

From aforementioned literatures, there are growing interest in using RN as a system analysis & planning tool. Numerous evidences of publications were found related to RN development from different power utilities from different countries with different methodology and applications. The application of RN has been reported to be mainly in the areas of distribution network planning [27],[32], costs assessment under the incentive based regulation [33],[31],[34], to estimate TL [4],[14],[27],[30], and more recently, the assessment on the impact of integrating SG technologies [30] and DER in distribution networks [15],[35].

To date, there is no evidence of published work in establishing either RF or RN for the Malaysian distribution network. Most research works related to impact assessment are conducted for the Malaysian distribution network involves using samples of network, instead of representative network, such as found in the area of PV system [36],[37],[38], PEV charging [39] and embedded generation [40]. Even though the results were used to validate the presented methodology, the results could not be accurately extrapolated or inferred to a much broader group of network with based on network type, location of areas served (e.g urban, semi-urban or rural), feeder load characteristic, topology and configuration in Malaysia. Therefore, there is a research opportunity to investigate and develop either RF or RN the Malaysian distribution network. If made available, the Malaysian distribution RF/RN can be potentially used as common and generic test cases for numerous type of studies.

Estimating distribution TL using representative feeder/network model

Considerable amount of work are found in using RN to calculate TL, such as found in [14], [27],[30] and [41]. These references primarily uses load flow simulations to calculate the TL for each feeders in each RN. The TL for each RN is then assumed to be similar for each of the network under its clusters. Then, using extrapolation (based on appropriate weighing factors), the TL for the whole feeder population can be estimated effectively. The results is also more accurate than using random samples of network model as the samples might not be representative of the actual and diverse network. However, there are several drawbacks.

One problem is that, for effective and targeted approach of TL mitigation, it is important for utilities to be able to ascertain which feeders and/or transformers that contributes the highest and lowest level of TL. However, although RN is closely resembles the feeders under its clusters, each feeder in the RN and the feeders under its cluster might have different TL characteristics. Therefore, the feeder TL for RN model may have significant differences to the feeders in the network it represents. Therefore, one possible solution is to use a simplified and RF model, such as presented in [42], which can be adjusted to effectively estimate each feeder TL with different characteristics, such as feeder length, topology and load distribution.

Representative feeder/network development

From the aforementioned literature, most RF/RN development uses statistical approach, mainly clustering method. However, despite a large and evolving number of clustering algorithms, this approach has several drawbacks. Firstly, determining the optimum number of clusters as well as which parameters/attributes is most suitable for clustering has been one of the most difficult to solve [43],[44]. This is because, clustering of data involves iteration process of trial and error to discover the natural grouping(s) of a set of patterns in a particular group/cluster [45]. Secondly, the results obtained from these techniques are often proven to be good, they require voluminous dataset to accurately cluster the network and for data training purposes. In addition, reliable and large number of data requirements also poses significant impediment for effective clustering process.

An example of a simple statistical tools in statistics is related to data "sampling". In statistics, sampling is a widely-used simple and systematic method of selection of a subset of individuals from within a statistical population of data to estimate characteristics of the whole population. There are many types of sampling technique, such as simple random sampling, systematic sampling, stratified sampling and clustered sampling. Among all sampling techniques, "stratified sampling" techniques have been widely used because of its superior efficiency [46]. The main advantage of this approach is that, apart from being simple, unbiased and representative, it has higher probability to produce a weighted mean and has less variability and deviation. Some application of stratified sampling technique has also been applied in power estimation of circuits [47], operational planning of generation system [48], load profiling [49],[50], and in TL calculation [51],[52],[53].

Before establishing a set of RF model, it is important to obtain statistical analysis of the datasets to in order to classify/disaggregate the datasets into groups of feeders with similar or generic characteristics. Many previous works have utilized data mining approaches (e.g. clustering, classification) and artificial intelligences (AI) techniques to statistically group the feeders which have similar characteristics. However, these methods involves complex mathematical manipulations and intensive data input for learning. To solve this, a simple classification based on "stratification" method is found to be an acceptable alternative approach.

In basic definition, network disaggregation or classification is the process of combining and classifying the datasets into groups of identical selected parameters. Many previous works disaggregate their network based on voltage levels, location of areas served, equipment, load served and type. There's no standard on how to classify the network since it all depends on the available network and load information, and the intended application.

Data stratification technique, which is part of a representative sampling method, is a simple process of clearly dividing members of the population into homogeneous subgroups (i.e. 'strata') before the data in each final stratum is sampled (or just use all data in the strata) and analyzed [46]. Many applications of data stratification is found used as part of calculating TL, such as in [54],[55],[51].

The stratification is carried out in such a way that each strata are homogeneous within themselves with respect to the characteristic under study as well as heterogeneous among other strata. In other words, each data in the same strata have at least one similar characteristics. Furthermore, data stratification can be conducted as many stages, depending on the number of intended type of data classification. Then, each strata will be analyzed to synthesize the prominent and representative characteristics using standard statistical analysis tools, such as mean, variance and standard deviation.

In stratified sampling technique, the basic idea is, the dataset population, N, is partitioned into L disjoint subpopulations, called "strata" as shown in Equation 1. Figure 7 illustrates the stratification concept of data population. Each stratum is to be mutually exclusive with no overlap of dataset. Note that, the stratification process can be conducted at multiple stages. The main advantage of this approach is that, apart from being simple, it has higher probability to produce a weighted mean and has less variability and deviation. Thus, this guarantees each group is truly represented in the sample.

Thus, for example, N_1 represents the total number of datasets in the first strata. The stratification involves two most important steps: (i) determining the number of strata, L and (ii) determining the appropriate strata boundaries. In this study, one may choose any number of strata, based on what data available and the characteristics which have the dominant impact on circuit design and performance [17].

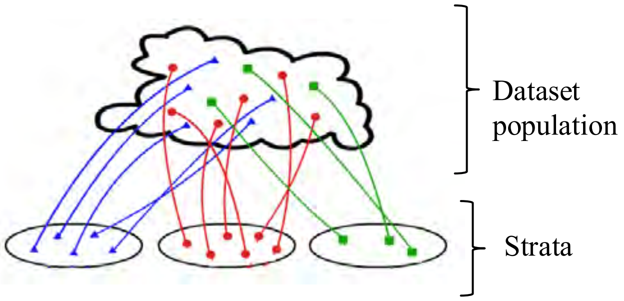

Figure 7 Stratification concept

$$N = N_i + N_{i+1} + \cdots + N_L \;,\;\; i = 1, \dots, L \qquad \text{(Equation 1)}$$

Where:
N = total no of dataset for entire population
i = index for stratum number
L = total number of stratum

If all characteristics were used as criteria to stratify the feeder datasets, the result would be a set of representative feeders that will be as large as the number of distribution feeders in the network itself, which is impractical. In general, the number of types or classification of RF depends on the number of specified parameters. There is no limit to the number of network classification or stages of stratification that can be adopted, but complexity of the process and size of the RF/RN would increase exponentially with an increase in the number of parameters, without any commensurate increase in accuracy [34].

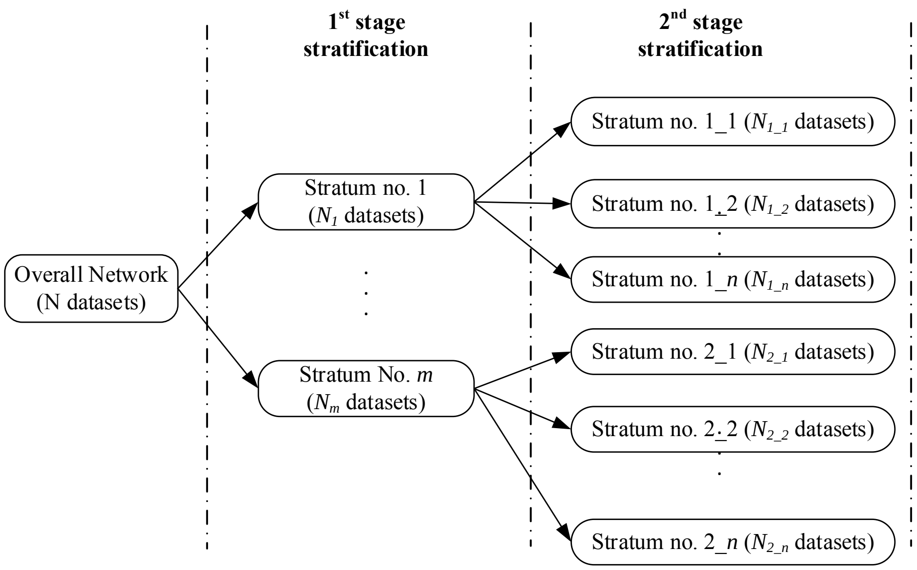

Figure 8 Example of two (2) stage data stratification

For example, consider two (2) level of stratification, as shown in Figure 8 and described in Equation 3 and Equation 4. The 1st stage stratification will generate m stratum, and then, for each 1st stage stratum, it is again stratified to generate n more stratum. The total/final no of stratum will be $m \times n$.

$$N = N_i + N_{i+1} + \cdots + N_m, \quad i = 1, \dots, m \qquad \text{(Equation3)}$$

$$N_i = N_{i\_k} + N_{i\_(k+1)} + \cdots + N_{i_n}, i = 1, \dots, m, k = 1, \dots, n \qquad \text{(Equation 4)}$$

Where:

N = total no of MV network datasets
i = index for stratum number for 1st stage stratification
k = index for stratum number for 2nd stage stratification
m = total number of stratum for 1st stage stratification
n = total number of stratum for 2nd stage stratification

In many existing research works, it was found that, voltage levels and location of areas served (i.e. rural, urban and sub-urban) which generally serves as defining and design factors of distribution feeder, are commonly used to characterize distribution feeders [9],[17].

2. OBJECTIVES OF RESEARCH

1. To develop a set of new Malaysian electricity reference distribution network which is representative of various network types and demographic location.
2. To validate the developed reference network against the real network in terms of network characteristic and network performance.
3. To test the effectiveness of the developed reference network against the real network with the connection of emerging technologies, such as electric vehicle, photovoltaic and energy storage.

3. RESEARCH METHODOLOGY

Development of MV representative feeder model

Distribution feeders and its load characteristics are never the same for every feeder. Thus, making any type of assessment of power distribution network oftentimes involves rigorous, time-consuming and complex analysis due to the large amount of network datasets and associated parameters required. One solution is by using a set of generic and representative feeder (RF) model. This simplified and representative feeder model could help to reduce the amount of data to be handled, and thus simplify the complexities of analysis, which includes estimating TL. Therefore, it is the aim of this section to establish a set of RF and feeders, such that, will be used to effectively estimate the TL of the entire MV distribution system.

The general procedure to develop RF for the MV distribution network is summarized in Figure 9. The output of this process is a set of RF which has a generic characteristic of the real network based on the feeder type, voltage level, location of areas served, feeder length, load distribution and topology. The process starts with classifying the overall network datasets using stratification into two (2) main type, that is based on voltage level and location of areas served. Then, for each group, using dataset of matured urban and rural network, the datasets are again stratified based on location of areas served (i.e. urban, semi-urban and rural). Then, statistical analysis is carried out for each group to find the average feeder characteristics. Finally, a set of general representative feeders (RF) with average features of the Malaysian MV distribution feeder are also developed to further classify the feeders based on certain characteristics, such as voltage, feeder length, type and topology.

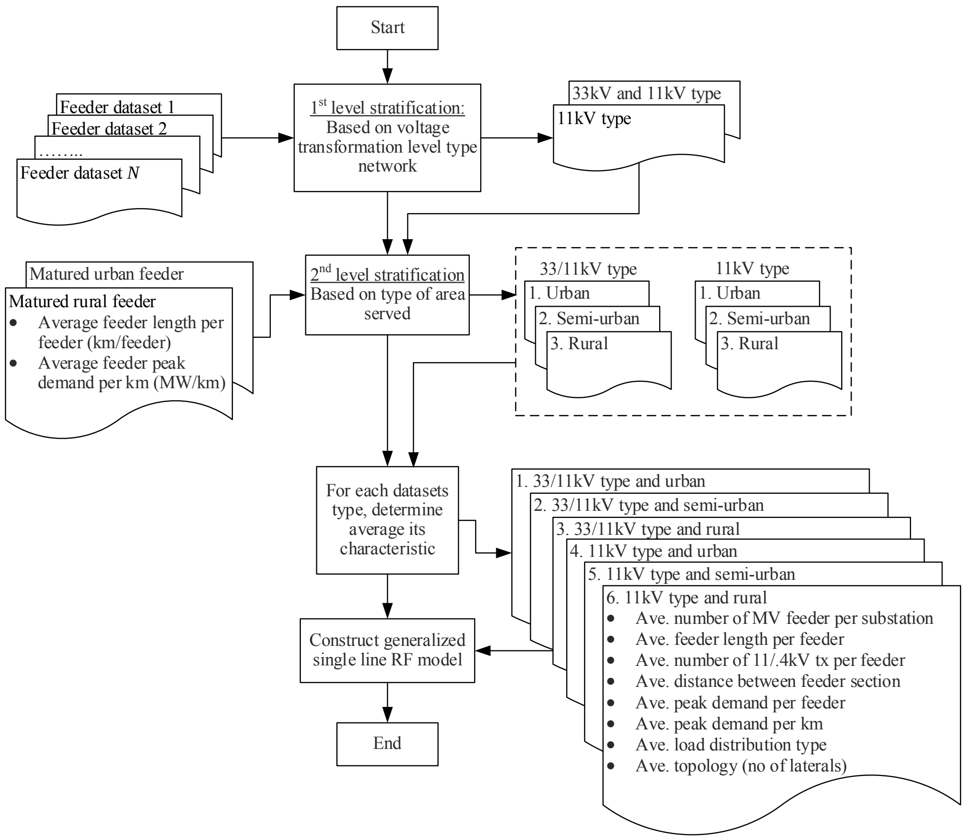

Figure 9 Flowchart to develop distribution MV RF model

These RF models then can be used for various applications and/or as reference for future research which requires understanding of the typical distribution network model available which are representative of the Malaysian distribution system. These set of RF can also be used to create a synthetic distribution MV network for various analysis. In this project, the RF model is used in combination with the energy flow model (EFM) in order to develop an effective method to accurately estimate the system wide TL of MV distribution network.

Data collection

There is no limit to the type and extent of data that can be incorporated into developing a RF. This is dependent upon the easy availability of data as well as the type of applications of the model. Therefore, the present study is only limited by the extent of adequate relevant data available for the Malaysian distribution network. A set of raw data obtained from

TNB containing network and load parameters of MV feeders and transformers are gathered using spreadsheet, as listed in Table 2.

The raw data consists of a total of approximately 524 datasets, of which, each datasets represent each network and consists of 46 parameters, containing all aggregated parameters for all MV and LV feeder and transformers. The list of these parameters are shown in Appendix F. Out of these 46 parameters, only 10 parameters are used for this study, as shown in Table 3. A significant effort of manual processing and verification was required to ensure that the information that was collected was relevant, consistent and free from any error, unrealistic, irrelevant or incomplete data. After data verification and cleansing process, this initial number of dataset was reduced from 524 to 444.

Table 2 TDIS network and energy raw data

Parameters	0	1	2	3	4	5	…	46
Dataset	Mnemonic	Tx[1] Ratio	Total Inflow Energy (MWh)	PPD[2] (MW)	33kv UG[3] feeder number	33kv UG[2] feeder length (km)	…	Distribution Tx capacity 33/0.4kv
1	ABCD	132/11	4088	7.24	0	0	…	0
2	FGHI	132/11	4303	7.62	0	0	…	0
3	WXYZ	132/33	25171	45.60	7	41.81	…	0
…	…	…	…	…	…	…	…	…
524	RSTU	132/11	6813	17.34	0	0	…	0

[1]Transformer
[2]Peak power demand
[3]Underground

Table 3Type of MV distribution network and load data

MV Feeder and transformer data	Load related data
1. Network pneumonic and location (i.e. name of city, region and state) 2. Transformer type at TDIS (132/11 kV OR 132/33kV) 3. Type and total number of 33 kV feeders 4. Total number of 33/11 kV transformers, type & capacity 5. Total number and type of 11 kV feeders 6. Total 33kV feeder length 7. Total 11kV feeder length	8. Monthly incoming energy units from GSS recorded at TDIS 9. Monthly recorded PPD at TDIS 10. Monthly recorded MV feeder maximum current

Feeder classification based on dataset stratification

In this work, a simple data stratification technique is employed to classify the feeder datasets into two (2) groups characterized based on voltage transformation type and location of areas served, as described in Table 4. Figure 10 describes in general how the total datasets are stratified into group of datasets characterized by voltage level and location of areas served.

Table 4 Stratification to classify Malaysian MV distribution network

Stratification stage	Stratification criteria
1^{st}	MV Network dataset with double voltage transformation (33kV and 11 kV) and single voltage transformation (11kV only).
2^{nd}	MV Network dataset which serves load located in the urban, semi urban and rural area.

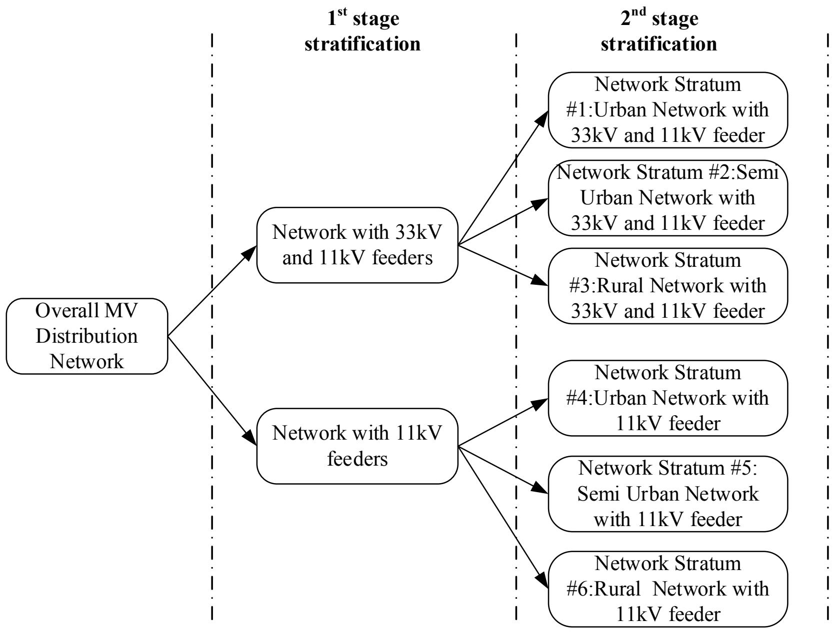

Figure 10 Stratification of Malaysian MV distribution network datasets

The stratification technique involves two most important steps: (i) determining the number of strata, L and (ii) determining the appropriate strata boundaries. To determine the strata boundary of the datasets for voltage level is relatively simple and straight forward since the data is available and is of categorical type (or discreet variable). Hence, there's no need to perform any statistical analysis to find the strata boundary. Thus, using simple data sorting commands in spreadsheet, the first step is to split the entire datasets into two (2) groups based on the two voltage transformation type described in Table 4. The 1^{st} level of stratification is now complete. Next, is the 2^{nd} stratification procedure.

The 2^{nd} stratification aim to further split each of the two (2) stratified dataset earlier into three (3) separate group based on locations of area served, i.e urban, semi-urban and rural. However, the main challenge here is that, there's no discreet data empirical available which classify each network dataset as urban, semi-urban or rural. Furthermore, there are no known literature which defines the criteria to differentiate these type of area served (urban, semi-urban or rural) based on the distribution network configuration. So, to solve this problem, the strata boundary need to be established by statistical analysis. Several works have suggest that the strata boundaries should be selected properly by approximation [56], principal component analysis [57], numerical optimization [58], dynamic programming [59]. However, these methods are useful if the distribution of the datasets are not known as well as it requires extensive datasets and complex mathematical manipulations.

Hence, to solve this problem, a simple approximation approach employed in this work is to use an empirical benchmark to identify segmentation and boundary recognition of the whole dataset population. Specifically, the idea is to select and analyze several benchmark datasets of which the feeders and substations are confirmed to serve loads which are located in a known location of area served. For example, in Malaysia, it can be confidently generalized that all distribution feeders located in major cities such as Kuala Lumpur, Ipoh and Johor Bharu can be a benchmark of a "typical urban type feeders". Similarly, for a selected rural area, such as in less populated such as the remote areas in mostly large states, such as Selangor, Terengganu, Kelantan and Pahang, where there are more rural areas than urban, is used as benchmark of "typical rural type feeders". Table 5 and Table 6 shows the selected network location which is identified as benchmark for urban and rural network respectively.

Then, the next challenge is to identify which of the feeder parameters are suitable to be used to stratify the datasets characterized by locations of area served. From observation of actual feeder layout, data obtained from TNB as well as feedback from local network planners/engineer, feeders located in a typical matured urban area are qualitatively identified and characterized as "*short length, highly meshed, with high load density with mixed residential, commercial and industrial load type*".

Table 5 Samples of selected distribution network located in KL

No.	Supply zone	No.	Supply zone	No.	Supply zone
1.	Ampang	11.	Imbi	21.	Pudu Ulu
2.	Batu Caves	12.	KL City Centre	22.	Sri Damansara
3.	Brickfield	13.	Kampong Lanjut	23.	Segambut
4.	Bukit Jalil	14.	KL Pavilion	24.	Sri Hartamas
5.	Bukit Mahkamah	15.	KL East	25.	Sentul Raya
6.	Bandar TunRazak	16.	KL North	26.	Bangsar
7.	Dang Wangi	17.	Manjalara	27.	Titiwangsa
8.	Damansara Height	18.	Mid Valley	28.	Vision City
9.	Galloway	19.	PandanMaju	29.	WangsaMaju
10.	HartaKemuncak	20.	Pantai		

Table 6 Samples of selected distribution network in rural area

No.	Supply zone	No.	Supply zone	No.	Supply zone
1.	Sungai Petani	12.	Kuala Lipis	23.	Taiping
2.	AlorSetar	13.	Kampar	24.	Gurun East
3.	Bentong	14.	Kota SarangSemut	25.	Kuala Krai
4.	Bidor	15.	Sri Manjung	26.	Kerayong

5.	Bukit KayuHitam	16.	Maran	27.	SeberangJerteh
6.	Benta	17.	TanjongPauh	28.	Sungai Siput
7.	Bintang	18.	Pekan	29.	TelukEwa
8.	Guar Chempedak	19.	Raub Bandar	30.	TanjungBatu
9.	GuaMusang	20.	Cameron Highland	31.	Kemaman
10.	Gurun	21.	TelukIntan	32.	Tanah Merah
11.	Temerloh	22.	Baling	33.	Ulu Melaka

On the other hand, the rural feeders is typically characterized as having "*long length, with sparsely distributed load, and low load density with majority residential type*". Subsequently, we can also approximate that, a semi urban is typically a "*medium length, moderately meshed and medium load density, with mixed residential, commercial and industrial load type*". Therefore, based on these qualitative measure and based on the availability of data, there's only two (2) possible quantitative parameters available which can be used to represent the feeder's location of areas served type, that is:

1. Average feeder length per feeder (km/feeder)
2. Average feeder peak demand per km (MW/km)

The calculation for the MV feeder length per feeder and linear load density performed on each dataset are shown in Equation 5 and Equation 6 respectively.

$$Average\ \ feeder\ length\ per\ feeder = \frac{total\ feeder\ length}{total\ number\ of\ feeder} \qquad \text{(Equation 5)}$$

$$Average\ \ feeder\ peak\ demand\ per\ km = \frac{total\ peak\ demand}{total\ feeder\ length} \qquad \text{(Equation 6)}$$

Then, the two parameters in Equation 5 and Equation 6 is calculated for both the benchmark urban and rural datasets (in Table 5 and Table 6) and use them as strata boundary to separate the remaining datasets into urban, semi-urban and rural network by applying the following estimation rules:

1. All datasets with average feeder length per feeder and peak demand per km which are equal or less than the values for the benchmarked urban feeders are stratified as urban network.
2. All datasets with average feeder length per feeder and peak demand per km which are equal or more than the ones for the benchmarked rural feeders are stratified as rural network.

All datasets with average feeder length per feeder and peak demand per km which falls in between the ones for the benchmarked urban and rural are stratified as semi-urban network.

4. RESULTS / FINDINGS

Results of MV RF model development

The aim of this subsection is to present the results of classifying the network datasets into groups based on location of areas served, i.e. urban, semi-urban and rural and voltage transformation type using stratification technique. In this work, feeder length per feeder and linear peak demand density are chosen as the parameters to set the boundary. To set the strata boundary, statistical analysis is conducted on samples datasets of known and matured urban and rural network, as benchmark network.

The statistical results from the stratified datasets of the benchmarked network are summarized in Table 7. The complete results are published and can be found in reference [60]. Based on the results from statistical analysis of the stratified datasets for the benchmark network, as expected, it can be seen that the network in the urban area has much shorter 33kV and 11kV feeders, compared to rural network. Furthermore, the linear load density of the 11kV feeders for the urban are much higher than the rural. Table 7 also compares the average values of the benchmark network. The next step is, the rest of the datasets are stratified into urban, semi-urban and rural using the results of average feeder length per feeder and linear load density values. Figure 11 shows the number of datasets in each stratification level. The semi-urban datasets are taken as the parameters, which are between the ones obtained for the urban and rural

Table 7 Strata boundary to stratify datasets based on locations of area served

Strata boundary parameter	Urban feeders	Rural feeders	Semi-urban feeders (estimated)
Average 33kV feeder length per feeder (km/feeder)	3.0 and below	18.9 and above	between 3.0 and 18.9
Average 11kV feeder length per feeder (km/feeder)	3.2 and below	16.6 and above	between 3.2 and 16.6
Average linear load density along 11kV (MW/km)	1.1 and below	0.2 and above	between 1.1 and 0.2

The extraction of dominant/average characteristics of feeders is based on the rule that each representative cluster or RF is created by the average value of the parameters associated with the real feeders allocated to the disaggregated group. Again, statistical analysis is used to find the average of each parameter relevant to the design. Mean values of relevant network parameters under each type of RF is shown in Table 8. Referring to the results in Table 8, the generic characterization of each RN for the Malaysian MV distribution network is then descriptively summarized in Table 9. Figure 12 and Figure 13 shows the single line diagram of the two (2) type of MV network.

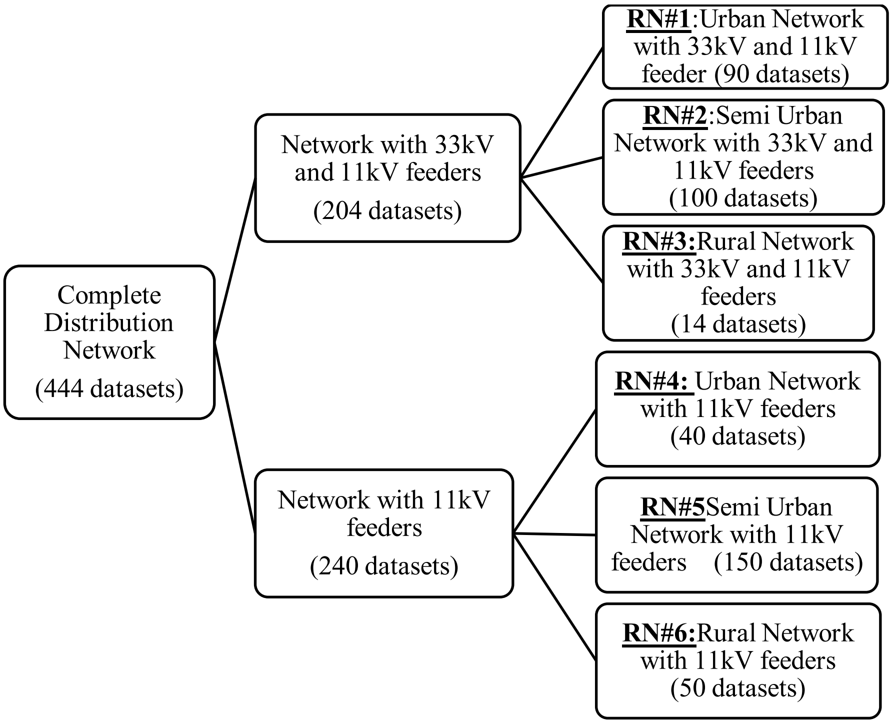

Figure 11 Data stratification for the Malaysian distribution MV feeders

Table 8Mean value for representative feeder characteristics

No	Representative feeder type	Mean value					
		RF#1	RF#2	RF#3	RF#4	RF#5	RF#6
1.	No of 11kV feeders per 33/11kV Transformer	5	4	3	-	-	-
2.	No of 11kV feeders (unit) for each main intake	-	-	-	11	10	9
3.	11kV Transformer number per 11kV feeder (number)	5	8	15	6	10	16
4.	LV feeder number per 11/.4 KV Transformer (number)	8	5	3	5	5	3
5.	Average distance between 11/0.4kV Transformer (per feeder in km)	0.6	1.2	2.1	1	1.3	2
6.	33/11kV Transformer MD (MW/ Transformer)	9.6	10	4.8	-	-	-
7.	11kV Feeder MD per feeder (MW/feeder)	2.5	2.4	2.8	1.6	1.97	2