

**Enhancing PV Hosting Capacity of Distribution Feeders using Voltage Profile
Design**

Akshay Kumar Jain

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Electrical Engineering

Robert P. Broadwater
Virgilio Centeno
Jamie De La Ree

February 14, 2018
Blacksburg, Virginia

Keywords: Voltage Profile Design, PV Hosting Capacity, Clustering Analysis,
Representative Distribution Feeders

Copyright© 2018 Akshay Kumar Jain

Enhancing PV Hosting Capacity of Distribution Feeders using Voltage Profile Design

Akshay Kumar Jain

ABSTRACT (ACADEMIC)

Distribution feeders form the last leg of the bulk power system, and have the responsibility of providing reliable power to the customers. These feeders experience voltage drops due to a combination of feeder length, load distribution, and other factors. Traditionally, voltage drop was the major concern. Now, due to an ever increasing PV penetration, overvoltage has also become a major concern. This limits the amount of solar PV that may be integrated.

Few solutions exist to improve the voltage profile, where the most common is the use of voltage control devices like shunt capacitors and voltage regulators. Due to a large number of design parameters to be considered, the determination of the numbers and locations of these devices is a challenging problem. Significant research has been done to address this problem, utilizing a wide array of optimization techniques. However, many utilities still determine these locations and numbers manually. This is because most algorithms have not been adequately validated. The validation of a voltage profile design (VPD) algorithm has been presented here. The validation of this algorithm was carried out on a set of statistically relevant feeders. These feeders were chosen based on the results obtained from a feeder taxonomy study using clustering analysis. The algorithm was found to be effective in enhancing the amount of solar PV a feeder may host, while still maintaining all the voltages within the ANSI standard limits. Furthermore, the methodology adopted here may also be used for the validation of other algorithms.

Enhancing PV Hosting Capacity of Distribution Feeders using Voltage Profile Design

Akshay Kumar Jain

ABSTRACT (GENERAL AUDIENCE)

Utilities have the responsibility of providing reliable power supply to their customers. Traditionally, bulk power was generated and transmitted over long distances incurring losses and voltage drops along the way. Now, with the integration of distributed energy resources, particularly solar photovoltaic (PV) generators at the customer locations, overvoltage has also become a problem. This requires adoption of measures which can help in maintaining the voltages within standard limits.

Several options exist to compensate for these voltage issues, the most commonly used is voltage control devices like capacitor banks and voltage regulators. However, determining the required numbers of these devices and their appropriate locations is a challenging problem. Even though a number of algorithms have been proposed to give automated solutions to this problem, most utilities still use a manual approach. This is because these algorithms have not been validated on a statistically relevant set of feeders. To solve this issue, the validation of a voltage profile design (VPD) algorithm is presented in this thesis. The ability of this algorithm to enhance the amount of PV that may be connected to a distribution network has been validated on a set of feeders. The feeders were chosen based on the results obtained from clustering analysis, a machine learning concept. The cost effectiveness of this algorithm has also been investigated and significant savings were observed. Furthermore, the methodology adopted here can be easily extended for the validation of other algorithms as well.

DEDICATION

This thesis is dedicated to my parents. Without their love, affection and blessings I would not have been able to complete this work.

ACKNOWLEDGEMENT

Words are not enough to express my gratitude and appreciation to my advisor, Dr. Robert Broadwater, for his constant guidance and support all through my Masters. Dr. Broadwater was a constant source of encouragement and made the Masters a great learning experience. I would also like to thank Dr. Virgilio Centeno, Dr. Jamie De La Ree and Dr. Vassilis Kekatos for promptly providing any guidance and help that I needed.

I want to acknowledge the constant support of my parents, Mr. Hemant Kumar Jain and Mrs. Sunita Jain, who ensured that I had everything that I needed to do well in my studies. I also want to thank my elder brother, Himanshu Jain, and my sister-in-law, Surabhi Garhawal, who were always there to help me.

Finally, I want to say a big thank you to my friend, Surabhi Thapliyal, who stood like a wall behind me and supported me. I cannot thank her enough for the support she provided during the Masters, and for being incredibly patient with me.

CONTENTS

1. Introduction	1
2. Literature Review	4
2.1. Placement of Voltage Control Devices.....	4
2.2. Representative Feeders for evaluating VPD algorithm.....	5
3. Voltage Profile Design (VPD) Algorithm	8
3.1. Challenges in manual voltage profile design.....	8
3.2. Reactive Compensation from fixed shunt capacitors.....	9
3.3. Voltage Regulators to improve feeder voltages.....	11
3.4. Voltage profile design (VPD) algorithm.....	14
4. Clustering Analysis	19
4.1. Parameter Selection.....	19
4.2. Feeder Selection and data preprocessing.....	21
4.3. Data Scaling.....	23
4.4. Outlier Rejection.....	25
4.5. Principal Component Analysis.....	27
4.6. Variable Reduction.....	29
4.7. Stopping Criterion or Optimal number of clusters.....	34
4.8. Clustering.....	37
5. PV Hosting Capacity Analysis and Cluster Validation	41
5.1. PV Hosting capacity Analysis.....	41
5.2. Minimum and maximum PV hosting capacity.....	45

5.3. Determining characteristics responsible for differences in PV hosting capacity of feeders.....	46
5.4. Comparison of cluster average properties and medoids.....	51
6. Validation of VPD Algorithm.....	54
6.1. Selection of feeders.....	54
6.2. Minimum PV hosting capacity of selected feeders.....	56
6.3. Implementation of VPD algorithm on selected feeders.....	58
6.3.1. Feeder 5022.....	60
6.3.2. Feeder 5021.....	63
6.3.3. Feeder 5012.....	66
6.3.4. Feeder 4021.....	69
6.3.5. Feeder 3021.....	71
6.3.6. Feeder 5024.....	73
6.4. PV hosting capacity of feeders after implementation of VPD algorithm.....	75
6.5. Smart inverter functions.....	77
7. Conclusions.....	80
References.....	82

LIST OF FIGURES

Figure 1. IEEE modified 4-bus system without voltage control devices.....	8
Figure 2.1. Phasor diagrams for a feeder with lagging power factor (a) without and (b) with shunt capacitor.....	9
Figure 2.2. Possible overcompensation during low load conditions.....	10
Figure 3. Effect of adding capacitive compensation in the 4-bus system.....	11
Figure 4. Control circuit of a voltage regulator.....	12
Figure 5. Effect of addition of a voltage regulator.....	13
Figure 6. Box-plot of parameters after removal of correlated parameters.....	23
Figure 7. Comparison of data scaling techniques.....	25
Figure 8. Outlier rejection using distance from 12 closest neighbors.....	26
Figure 9. Feeders and all parameters on first two principal components.....	28
Figure 10. Covariance heat map.....	30
Figure 11. Optimal number of clusters using Calinski-Harbasz criterion.....	35
Figure 12. Optimal number of clusters using average silhouette width criterion.....	37
Figure 13. Final clusters of feeders with respective medoids.....	40
Figure 14. PV hosting capacity analysis flow chart.....	44
Figure 15. Minimum and maximum PV hosting capacity.....	45
Figure 16. Residential/Commercial-scale PV hosting capacity comparison.....	49
Figure 17. Utility-scale PV hosting capacity comparison.....	50
Figure 18. Comparison of medoid parameter values with mean values of all feeders in the respective cluster.....	53

Figure 19. Minimum PV hosting capacity of selected feeders.....	57
Figure 20. Feeder 5022; Base case (top) and New Design (bottom).....	61
Figure 21. Voltage profile of feeder 5022; base case (top) and New Design (bottom).....	62
Figure 22. Feeder 5021; Base case (left) and New Design (right).....	64
Figure 23. Voltage profile of feeder 5021; base case (top) and New Design (bottom).....	65
Figure 24. PV hosting capacity improvement in feeder 5021; base case (top) and new design (bottom).....	66
Figure 25. Feeder 5012; Base case (left) and New Design 1 (right).....	67
Figure 26. Voltage profile of feeder 5012; base case (top) and New Design 1 (middle) and New Design 2 (bottom).....	68
Figure 27. Feeder 4021; Base case (top) and New Design (bottom).....	69
Figure 28. Voltage profile of feeder 4021; base case (top) and New Design (bottom).....	70
Figure 29. Feeder 3021; Base case (left) and New Design (right).....	71
Figure 30. Voltage profile of feeder 3021; base case (top) and New Design (bottom).....	72
Figure 31. Feeder 5024; Base case (top) and New Design (bottom).....	73
Figure 32. Voltage profile of feeder 5024; base case (top) and New Design (bottom).....	74
Figure 33. Improved PV hosting capacity of the New Designs using VPD algorithm.....	76
Figure 34. Maximum voltages observed in feeder 3021 using volt-watt control.....	77
Figure 35. Volt-Var curve (top); Volt-Watt curve (bottom).....	78
Figure 36. Maximum voltages observed in feeder 3021 using volt-var control	79

LIST OF TABLES

Table 1. Common Parameters used for creation of a feeder taxonomy.....	21
Table 2. Covariance between parameters.....	32
Table 3. Kendall rank correlation coefficients between parameters.....	33
Table 4. List of violation criteria analyzed in detailed PV hosting capacity analysis.....	43
Table 5. Distance of chosen feeders from cluster medoid, showing unbiased selection.....	47
Table 6. Parameters critical for PV hosting capacity analysis.....	48
Table 7. Selected feeders and their characteristics.....	55
Table 8. Actual minimum PV hosting capacity of distribution feeders.....	58
Table 9. Comparison of operational parameters of base case and new design of feeder 5022.....	63
Table 10. Cost benefits analysis of base case and new design of feeder 5022.....	63
Table 11. Comparison of operational parameters of base case and new design of feeder 5021.....	65
Table 12. Cost benefits analysis of base case and new design of feeder 5021.....	65
Table 13. Comparison of operational parameters of base case and new design of feeder 5012.....	67
Table 14. Cost benefits analysis of base case and new design 1 and 2 of feeder 5012.....	69
Table 15. Comparison of operational parameters of base case and new design of feeder 4021.....	70
Table 16. Cost benefits analysis of base case and new design of feeder 4021.....	71
Table 17. Comparison of operational parameters of base case and new design of feeder	

3021.....	72
Table 18. Cost benefits analysis of base case and new design of feeder 3021.....	73
Table 19. Comparison of operational parameters of base case and new design of feeder 5024.....	74
Table 20. Cost benefits analysis of base case and new design of feeder 5024.....	75
Table 21. Comparison of the base case and the new design using the VPD algorithm on all feeders.....	77

1. Introduction

In distribution systems voltage must be maintained within the limits set by the American National Standards Institute (ANSI) at customer loads [1]. Feeders experience voltage drops due to a combination of feeder length, load distribution, and other factors. Traditionally, voltage drop was the major concern for distribution system engineers, however, now due to an ever increasing PV penetration, overvoltage has also become a major concern.

One solution to voltage problems is line reinforcement. This may require replacing the existing conductors or adding parallel conductors. Although this solution can effectively reduce feeder losses while mitigating over voltage problems introduced by photovoltaic (PV) generators, it is very expensive [2]. In order to maintain customer voltages within acceptable limits, voltage control solutions, like load tap changing transformers, voltage regulators, switched capacitor banks, and/or fixed capacitor banks, are traditionally used. Some utilities employ mostly capacitor banks, while other utilities employ voltage regulators. Volt/VAR control systems use either local or centralized control.

However, determining the location for the placement of these devices remains a challenging task. A lot of research has been done in this area and many optimization based techniques have been proposed [3]–[9]. However, still most of the utilities determine the location using a manual approach. The most likely reason for the adoption of a manual design approach by the utilities is non-availability of built-in applications in distribution systems analysis software for determining these locations. Even if they are available, the algorithm used in such an application may not be applicable to all the different types of feeders under a utility's purview.

The manual placement of voltage control devices is usually done at the time point of peak load on a feeder, using a distribution systems analysis software. This time point is chosen because the maximum voltage drops will occur at the peak load. This single time point design may lead to violations at other time points. For instance it may lead to overvoltage violations at the minimum load time point. Secondly, overcompensation from a large number of fixed capacitor banks may lead to increased losses due to excessive current flow.

The above discussion suggests that manually designing a distribution feeder's voltage profile, by suitably placing voltage control devices, while considering all of the design aspects, is nearly impossible, and necessitates the use of automated tools and algorithms. These algorithms should be flexible to allow the placement of devices with different ratings, and generic enough to be applicable on different feeder types. This was the motivation behind the development of a Voltage Profile Design (VPD) algorithm. The VPD algorithm takes the desired voltage limits and the number of voltage control devices to be placed as inputs from the user. The suitable locations for the devices are then determined by the algorithm while considering the time varying load profiles, minimum and maximum load time points, and required compensation. The algorithm also determines the numbers of each type of device required to bring the feeder parameters within the user defined limits starting with the most inexpensive device, to give a cost-effective design.

A major challenge for evaluating any smart grid technology, including the VPD algorithm, is choosing the test feeder models. Distribution feeders vary widely in their characteristics, like line length, existing PV penetration, topology, protective devices used, and existing PV penetration to name a few. This implies that a random selection of feeders will not give statistically relevant results, and testing the algorithm on every feeder model can be time consuming. Thus such evaluation often requires selecting a reasonably small set of representative feeders that can capture the diversity of feeder designs in the region of interest. Cluster analysis, also called clustering, is a well-recognized technique for grouping a set of objects based on their similarities, such that they are more similar to objects within the group and distinct from others. Thus, grouping distribution feeders using clustering methods can provide a statistically valid and reasonably small set of feeders that can then be modeled and used for testing of new technologies.

Clustering is simply a statistical approach of grouping similar objects. The main objective of this research is determining the efficacy of the VPD algorithm in increasing the PV hosting capacity of distribution feeders, where PV hosting capacity of a feeder is defined as the maximum PV penetration which does not lead to any parameter violation. So, to make the clustering based approach relevant to the objective of this research, a dataset of feeder characteristics considered to be relevant for determining the PV hosting capacity will have to be used. This will enable grouping

of feeders with similar characteristics. Finally, the representative feeders obtained from clustering will have to be tested for their PV hosting capacity to determine if they actually differ in the amount of PV they can host. If these feeders do vary in this aspect, then the VPD algorithm may be tested on these feeders to get statistically relevant results.

Determining the PV hosting capacity of these representative feeders is itself a challenging task. A detailed PV hosting capacity analysis will require a number of stochastic PV deployments for every PV penetration level. This time intensive analysis can be simplified using an automated application. Since the VPD algorithm seeks to improve a feeder's voltage profile, it would be worthwhile to determine the impact of these improved point of common coupling (PCC) voltages provided by the VPD algorithm on the reactive power support provided by the smart PV inverters.

Even if the VPD algorithm is found to be significant in enhancing the PV hosting capacity, it will be practical only if it is cost effective. This requires a cost-benefits comparison between the existing feeder and the new design obtained from the VPD algorithm. This can be accomplished using a time series analysis over an entire year. Thus, if the VPD algorithm is found to be applicable to different feeder types, can enhance their PV hosting capacity, and be cost effective, while maintaining the voltages within the ANSI standard limits, then it can be used by utilities for achieving better voltage profiles. Also, the testing methodology adopted here can be extended for the evaluation of other smart grid technologies.

The main objectives of this thesis will be presenting the voltage profile design algorithm, developing a clustering methodology, and validating it using a detailed PV hosting capacity analysis. The clustering results will then be used for determining statistically relevant sets of feeders on which the VPD algorithm will be evaluated. The implementation of VPD algorithm on this feeder dataset will help in determining the algorithm's cost effectiveness and efficacy in enhancing the PV hosting capacity of feeders.

2. Literature Review

2.1. Placement of voltage control devices

In distribution systems voltage must be maintained within the ANSI standard limits at customer load [1]. Feeders experience voltage drops due to a combination of feeder length, load distribution, and other factors. In order to maintain customer voltages within acceptable limits, voltage control solutions like load tap changing transformers, voltage regulators, switched capacitor banks, and/or fixed capacitor banks are traditionally used. Some utilities employ mostly capacitor banks, while other utilities employ voltage regulators. Volt/VAR control systems use either local or centralized control. One solution to voltage problems is line reinforcement. Although this solution can effectively reduce feeder losses while mitigating over voltage problems introduced by photovoltaic (PV) generators, it is very expensive [2].

Capacitor banks are generally considered the least expensive upfront investment for providing voltage control. Much research has focused on the optimal sizing and placement of capacitors. The objective functions mostly seek to bring node voltages closer to a reference, minimize feeder losses, and/or minimize the number of tap changes (for switched capacitor banks) [3]–[5]. Particle Swarm Optimization is used in [6] and also in [7], [8]. Other optimization algorithms employed include differential evolution and pattern searches [9].

Applying capacitor banks alone does not always completely solve voltage problems. In some cases, voltage regulators may be more effective than pure reactive power control in solving voltage problems. Optimal capacitor and voltage regulator placement and sizing is considered in [10]–[12]. However, an ‘approximate power flow method’ is used in these studies. The power flow is termed ‘approximate’ because of a few assumptions used in this research. The voltage regulator placement problem was decoupled from the capacitor placement problem by assuming capacitive compensation exactly balances out the reactive power requirements of the grid, and consequently only real power flow needs to be considered for regulator placement. However this assumption is incorrect. Secondly, the changes in voltage angles were ignored, assuming these changes to be relatively small in distribution systems. These approximations limited the accuracy of the results. This was improved upon in [13], minimizing the number of required regulators to optimize a cost

function, but the philosophy for initial voltage regulator selection, placement, and tap setting is similar to that of [10]–[12].

With the large, ongoing increase in distributed generation interconnections, especially PV generation, more and more research is devoted to smart inverters [14]–[16]. Smart inverters can be set to operate at desired phase angles between voltage and current, and thus the inverters may either source or sink reactive power. Until recently, restrictions established by IEEE 1547 prohibited active voltage control by inverters. The recently adopted amendment, IEEE 1547a, has permitted active voltage regulation, but only after approvals from the distribution grid operators, but there are still concerns with using smart inverters [17]. One concern is the large number of smart meters and associated controllers required. Based upon current technology, it has been demonstrated that covering as few as 25% of renewable generation points of common coupling with smart meters would not be economical, even over 10 years [18]. A second concern involves issues of islanding, where UL 1741 listed inverters operate alongside non-certified equipment or synchronous generators [19].

None of the voltage design solutions mentioned above takes into account the time varying load patterns of the feeder and its individual loads. The ability to control the feeder voltage profile within a narrow range over time effects both conservation voltage reduction and PV hosting capacity benefits. Here the enhancement of PV hosting capacity of distribution feeders and economic advantages derived from maintaining a feeder’s voltage profile within a narrow range over time are investigated [20]–[24]. A new design algorithm for accomplishing this is evaluated. The design algorithm is referred to as Voltage Profile Design (VPD). Evaluating the VPD algorithm posed a challenge due to the large characteristic differences that exist among feeders. So, the results from a feeder taxonomy study, which highlighted the characteristics important for determining the PV hosting capacity, were used to pick representative feeders on which the VPD algorithm was evaluated.

2.2. Representative feeders for evaluating VPD algorithm

Evaluation of smart grid technologies, like the VPD algorithm, at a utility service area scale can be difficult due to the need for a complete set of accurate feeder models and often the

application of considerable computational capability. Thus, such evaluation often requires selecting a reasonably small set of representative feeders that can capture these differences in feeder construction present in a given service territory. Cluster analysis, also called clustering, is a well-recognized technique for grouping a set of objects based on their similarities such that they are more similar to objects within the group and distinct from others. Thus, grouping distribution feeders using clustering methods can provide a statistically valid and reasonably small set of feeders that can then be modeled and used for validation of technologies like VPD.

Willis et al. first attempted to cluster feeders [25]. They also modeled the feeders and presented results based on electrical analysis, not only statistical measures. Research by the Pacific Northwest National Laboratory used feeder data from the entire United States [26]. This led to including feeder parameters, such as climatic region, while also using graph theory concepts to help narrow the number of metrics needed without eliminating feeder characteristics.

Although most research on clustering feeders has focused on primary distribution systems, research in Germany has focused on low-voltage or secondary distribution systems. This is because significant amounts of distributed generation are being interconnected on the low-voltage side. In Germany clustering was used to develop “benchmark” grids to represent low-voltage networks [27]. Nationwide, data-based clustering has also been done in Australia [28]. This detailed report includes a valuable discussion on the initial data review, which is critical for optimal clustering.

Representative feeders have also been used to test smart grid technologies. A best approach for conducting PV hosting capacity studies was determined in [29] using the benchmark grids developed in [27]. The benefits of clustering are not limited to technologies for enhancing PV integration alone. Clustering was used to increase the 6.5% voltage band available to medium-voltage grids by freeing up the unused part of the 13.5% voltage band of the low-voltage grids [30]. The authors also compared the representative models to detailed models to validate their effectiveness; however, the data used were from a single medium-voltage grid, making them less representative, although it was sufficient for the objectives of the particular project.

So as a starting point for this research, a thorough review of the clustering practices used internationally was completed to determine an exhaustive sequence of steps that might be used to develop representative feeders, and is presented in [31]. It was found that one step or another was missing in most of the research discussed above. Further, the techniques used at each step varied considerably.

To address these issues the most relevant of these techniques and steps were implemented, and a feeder taxonomy was created. The most significant contribution, however, is the validation of clustering results, which is absent from any research considering a sizeable data set. This research validates the final results by comparing the PV hosting capacities within and among clusters of feeders. Similar hosting capacity within a cluster, and different hosting capacities among clusters, would signify that a valid feeder taxonomy has been created. It can also help in highlighting the most relevant feeder characteristics for determining PV hosting capacity of feeders. These can be used in choosing representative feeders from other utility databases without performing a detailed taxonomy study again.

The parameter values of representative feeders were also compared to the average values of the other feeders in the respective clusters, which helped in determining the ability of representative feeders to effectively represent their respective clusters. All of the methods were implemented on a data set of more than 3,000 distribution feeders with more than 30 parameters, chosen by utility experts, where the parameters were considered important for determining PV hosting capacity from a utility in North America [32].

3. Voltage Profile Design (VPD) Algorithm

3.1. Challenges in manual voltage profile design

The analysis of distribution systems can be more involved than transmission systems as many of the simplifying assumptions cannot be used. The distribution systems cannot be assumed to be balanced or consist of transposed lines. Due to mutual couplings among phases, symmetrical components cannot decouple the distributions system under consideration. This necessitates the analysis of the three phases separately.

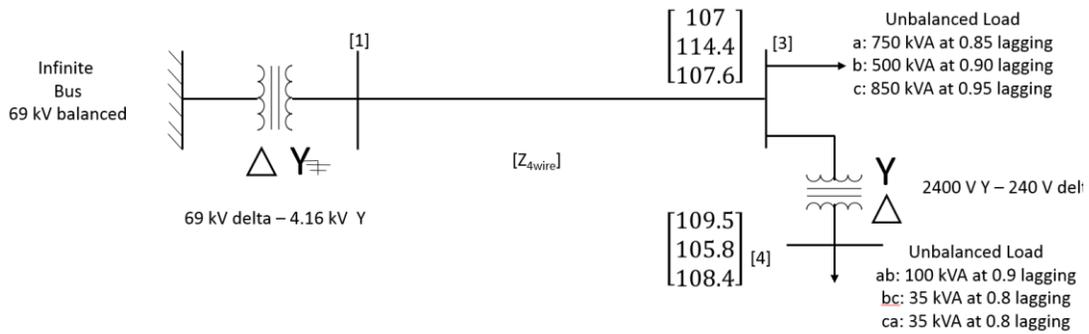


Figure. 1 IEEE modified 4-bus system without voltage control devices

The modified IEEE 4-bus system, shown in figure 1, has been used to explain the modeling of voltage control devices and the problems involved in re-designing the voltage profile manually [33], [34]. The substation is connected with the infinite bus, and the substation primary voltage is kept constant for all the power flow iterations. The substation transformer is delta-wye grounded and is connected to a distribution line consisting of 3 phase conductors and a neutral conductor. The series impedance of the line is represented through a 3x3 phase impedance matrix Z_{4wire} , where the effect of a neutral conductor, grounded at multiple points, has been included using Kron reduction [35]. Since the loads are unbalanced, the neutral current is non-zero, and so the effect of the return path of current through ground has been included using Carson's equations [36]. Both delta and wye connected loads are considered. The 3-phase ungrounded wye-delta distribution transformer itself consists of three single phase transformers of different ratings. So, none of the simplifying assumptions can be used, even in this simple, 4 bus system, and the problem gets compounded when real world distribution systems composed of thousands of nodes are considered.

The voltages at the load buses 3 and 4 are shown in figure 1. As per Range A of ANSI standard limits, the service voltages should be maintained within 126-114 V at all times [1]. However, the voltages are much below this minimum voltage threshold. Also, the load values considered in this example are for a single time-point. Considering the large voltage drop, this time point will most likely represent the peak load condition. Traditionally the peak load time point has been used for determining the position for placement of voltage control devices, like capacitor banks and voltage regulators to maintain desired service voltages. However, this single time point design may lead to overvoltage conditions during the low load conditions.

3.2. Reactive Compensation from fixed shunt Capacitors

A capacitor bank is the cheapest option available to a distribution engineer for improving the voltage on a feeder with predominantly inductive loads. These may be connected in series or in parallel with the lines. Parallel, or shunt connected capacitor banks, are most commonly used for providing reactive power support to counteract the out-of-phase component of current required by an inductive load [37]. By deploying a shunt capacitor the magnitude of source current is reduced,

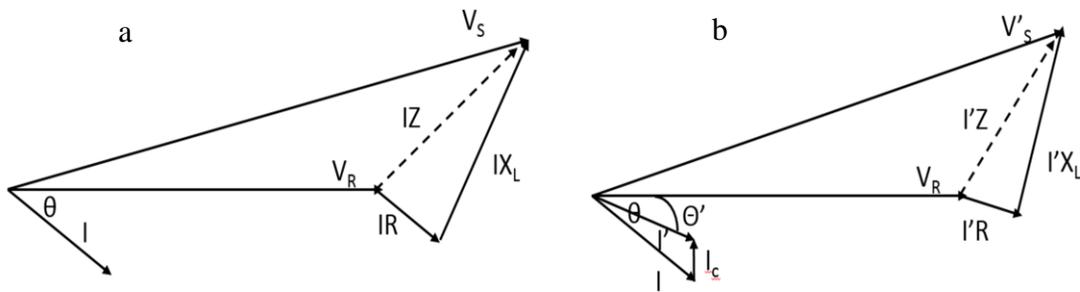


Figure 2.1: Phasor diagrams for a feeder with lagging power factor (a) without and (b) with shunt capacitor

and the power factor gets improved, which leads to a reduction in voltage drop from the source to the load. This can be seen in figure 2.1, where the phasor diagrams show the effect of the addition of a shunt capacitor [37]. The compensating current I_c is out of phase with the inductive current

and shifts the overall system current I with a lagging angle θ with the voltage to I' which now lags by $\theta' < \theta$.

The fixed shunt capacitor as discussed above cannot be switched automatically. So it may lead to overvoltage during low load conditions. In such a scenario the compensating current I_c will make the system current I' lead the system voltage. The load voltage will become larger than the source voltage and may even exceed the ANSI limits. This can be seen in the phasor diagram shown in figure 2.2. This problem is mitigated by the use of switched shunt capacitors, which have a controller. The controller controls the switching ON or OFF of the capacitor based on the reactive power flow it senses at its point of connection with the grid. However, these are more expensive than the fixed capacitors. So, to ensure overvoltage does not occur, fixed shunt capacitor placement was done at the minimum load condition in the VPD algorithm.

Capacitive compensation can thus be used to remedy the voltage drops in the 4-bus system. As can be seen in figure 3, a 3 phase fixed capacitor bank has been connected to bus 3. The capacitor bank consists of 3 single phase capacitor banks of 300 kVar/phase connected to the three phases. Capacitive compensation is usually connected equally in all of the three phases, as a standard utility practice. A shunt capacitor bank is modeled as a constant susceptance (B_i) load. Based on the voltage and kVar ratings, a susceptance value is determined, which is kept constant in all the iterations.

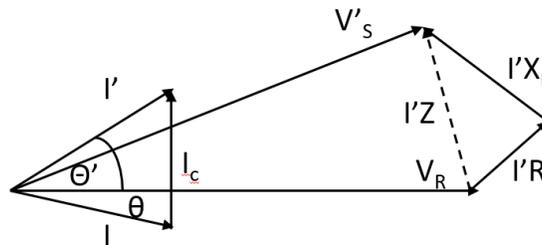


Figure 2.2: Possible overcompensation during low load conditions

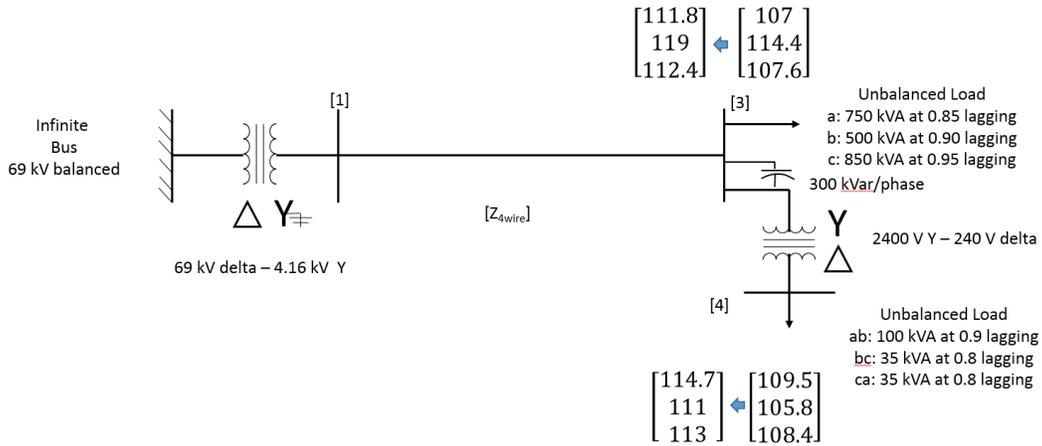


Figure 3: Effect of adding capacitive compensation in the 4-bus system

$$B_i = \frac{kvar}{kV_{LN}^2 * 1000} S, \text{ where } i = a, b, c \quad (1)$$

The current injection in every iteration is given by,

$$IC_i = jB_i * V_{in} \text{ where } i = a, b, c \quad (2)$$

Here the voltage V_{in} , will keep changing in every iteration, and consequently the compensating current injected per phase will also vary. Figure 3 shows that busses 3 and 4 voltages have improved after addition of the fixed shunt capacitor, but some of the phase voltages are still below the limit. Thus, capacitive compensation alone cannot reduce the voltage imbalance. Voltage imbalance can be improved using voltage regulators.

3.3. Voltage Regulators to improve feeder voltages

Voltage regulator, or step-voltage regulator, is basically an autotransformer with a load tap changing (LTC) mechanism [33]. The tap position on the series winding determines the voltage at the output of the voltage regulator. The most commonly used voltage regulators have a regulation range of $\pm 10\%$ with 32 steps. This corresponds to a 0.75 V change per step on a base of 120 V. The tap position is determined by the line drop compensator (LDC) as shown in figure 4. The LDC has a fixed impedance setting Z_{ldc} , which gives the estimated impedance from the voltage regulator

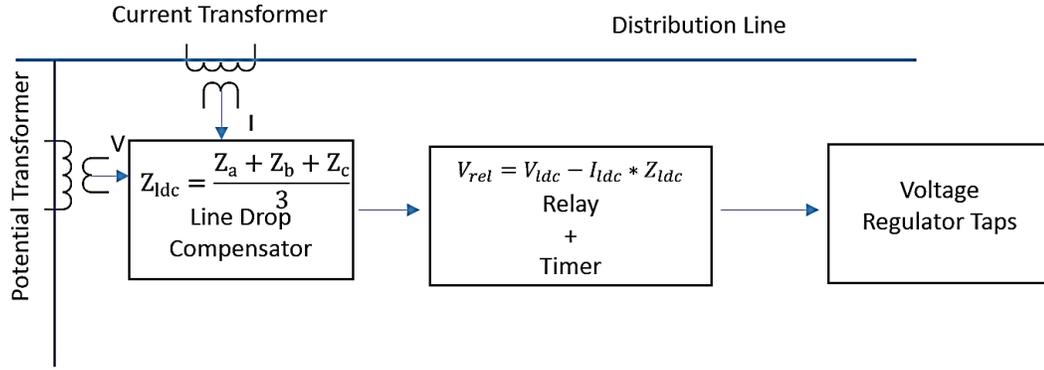


Figure 4: Control circuit of a voltage regulator

to the load center (i.e. the feeder location where voltage has to be controlled [33]). This impedance setting may also be set to zero, in which case the voltage at the output of the voltage regulator will be controlled. Z_{ldc} is the average of the ratio of the difference of the regulator node voltage and the load center voltage, and the line current of the three phases. These voltage and current values are determined by locking the regulator in the neutral position and running power flow to convergence. Figure 5 shows the effect of adding a voltage regulator after bus 1. The regulator node is node 2, and load center node is node 3. So Z_{ldc} is,

$$Z_{ldc} = \frac{Z_a + Z_b + Z_c}{3} \quad (3)$$

where, $Z_{a,b,c} = \frac{V_{2(a,b,c)} - V_{3(a,b,c)}}{I_{23(a,b,c)}}$

Once Z_{ldc} is set, the tap position is determined by estimating the load center voltage, which is given by the relay voltage. The real time line current (I_{ldc}) and regulator node voltage (V_{ldc}) is sent to the LDC through current and potential transformers, respectively, as shown in figure 4. The relay voltage (V_{rel}) is given by,

$$V_{rel} = V_{ldc} - I_{ldc} * Z_{ldc} \quad (4)$$

If V_{rel} is within the set bandwidth, then the taps will not change, otherwise if the voltage stays outside the bandwidth, then taps will change to bring the voltage within limits. The regulator shown in figure 5 has a $\pm 10\%$ regulation range. The tap positions for the three phases are different. The regulators taps in heavily loaded phases 'a' and 'c' had to change more from the neutral position to bring the voltages within the bandwidth. The voltages at bus 4 are still outside the ANSI

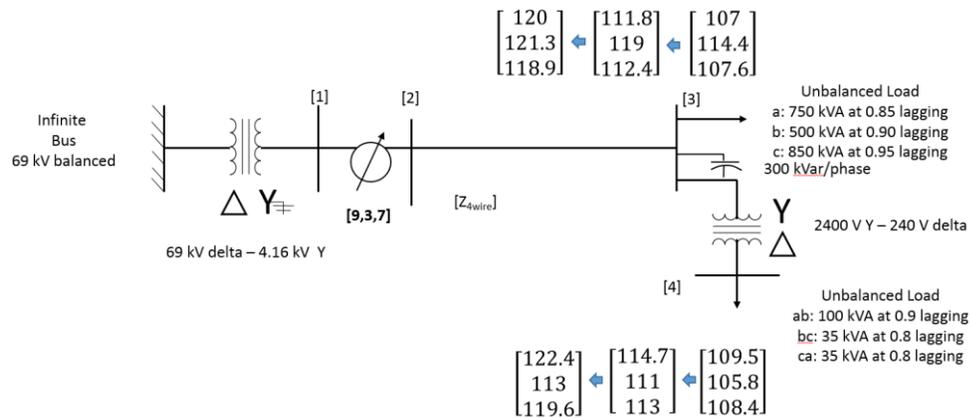


Figure 5: Effect of addition of a voltage regulator

limits. Thus, even for a simple 4 bus system, maintaining the voltages within ANSI limits is a challenging task. Actual feeders have thousands of nodes and large variations in the topology.

This example has helped in explaining the modeling and effects of addition of shunt capacitors and voltage regulators, and the difficulties associated with voltage regulation. The key problems highlighted are:

- Voltage profile design using loading conditions at a single time-point may not work effectively at other time points. This calls for the use of time varying load profiles available through smart meters and advanced metering infrastructure (AMI).
- Fixed shunt capacitors may lead to overvoltage at low load conditions. This may be mitigated through more expensive switched shunt capacitors. However, a more cost effective approach would be to use fixed capacitors, as long as they do not cause violations before moving to switched shunt capacitors.
- Capacitive compensation alone may not be sufficient to achieve desired voltage profiles. Voltage regulators may have to be used, especially to reduce large voltage imbalances.
- Determining the location and number of each type of voltage control device to be placed in a distribution feeder, and to get a cost effective solution, is a challenging task.

Considering all of these points, manual determination of voltage control device locations may not be feasible. However, all of these points have been considered in the VPD algorithm implemented in the Distributed Engineering Workstation (DEW) platform [38].

3.4. Voltage Profile Design (VPD) algorithm

The VPD algorithm works for radial distribution feeders with renewable generators located on either the primary or secondary of the feeder. Here the focus is on PV renewable generation. The algorithm seeks to determine appropriate locations to place voltage control devices and the necessary number of devices. Major designer inputs to the VPD algorithm are:

- **Beginning of Feeder Voltage** is the target voltage set point, usually the LTC set point, in volts.
- **Voltage Range** consists of an **Upper Limit** and a **Lower Limit**, both in volts. This specifies the desired range within which voltage should be controlled at all customer meters for all time.
- **Period for analysis** is typically an annual evaluation involving 8760 load points, one for each hour of the year. In case only the peak load data is available, the algorithm can run for just the peak load condition, but typically the results are very different for the single time point design, and the performance is less desirable as to energy savings and PV hosting capacity.
- **Time Step** is the time interval used in the design and evaluation of benefits. It should be no shorter than the load data time interval, and for the analysis considered herein is one hour.
- **Fixed Capacitor**
 - The **Maximum Number of Fixed Capacitors** is the largest number of fixed capacitor banks per feeder to be placed. May be set to zero.
 - The **Fixed Cap Size (C_F)** is the fixed capacitor size per phase in kVar.
- **Switched Capacitor**

- The **Maximum Number of Switched Capacitors** is the largest number of switched capacitor banks per feeder to be placed. May be set to zero.
 - The **Switched Cap Size (Cs)** is the capacitor size per phase in kVar.
- **Voltage Regulator**
 - The **Maximum Number of Regulators** is the largest number of voltage regulators per feeder to be placed. May be zero.
 - The **Control Bandwidth** must be no greater than half of the **Voltage Range** considered above.
 - **Control Percentage** is either 5% or 10%.
 - The **Regulator Control Zone Voltage Drop (V_{zone})** determines the target voltage difference from the new regulator to the lowest voltage location within the control zone of the regulator.

The VPD algorithm will only use the number of control devices of each type needed to achieve the specified voltage range. However, the algorithm will not place more than the specified maximum number of devices. Following are the algorithm steps:

- 1) All of the existing voltage control devices are turned off, and a feeder evaluation is performed. Feeder evaluation involves running power flow at the specified **Time Step** over the entire **Period for analysis**, which is usually the entire year at an hourly interval. In this case power flow would be run 8760 times based on the different load conditions for each hour of the year. This analysis identifies the minimum and maximum load times for the specified duration.
- 2) The voltage set point of the substation transformer LTC is set as the input **Beginning of Feeder Voltage**.
- 3) The most economical option, fixed shunt capacitors, are placed first. This helps in providing a cost effective solution. To avoid fixed capacitors creating overvoltage violations, the design is performed for the minimum load condition determined in step 1. The lowest customer voltage point in the feeder is located, and a feeder path trace [39] to a conductor that satisfies the conditions defined in (5) and (6) is performed.

Equations (5) and (6) help to avoid overcompensation at the beginning of the feeder and downstream of the location where the capacitor is installed, while compensating for the reactive losses and improving the voltage profile.

$$\min(Q_t) - C_F * m > C_F \quad (5)$$

$$\sum_{i=1}^N Q_i + \sum_{j=1}^K Q_{LOSSj} - \sum_{k=1}^Y C_k - C_F > 0 \quad (6)$$

, where

Q_t = average reactive flow per phase at the beginning of a feeder at time point t ($t \in [0, T]$). 0 and T represent the time indices for the ‘From’ time and ‘To’ time of the *Period for analysis*, respectively;

C_F = fixed capacitor size per phase (*Fixed Cap Size*);

m = number of fixed capacitors previously installed by VPD;

Q_i = average capacitive load per phase at customer i ;

N = total number of customers downstream from a given location;

Q_{LOSSj} = kVar loss of conductor j ;

K = total number of conductors downstream from a given location;

C_k = kVar output per phase of the capacitor k ;

Y = total number of capacitors downstream from a given location already installed by the program.

In the feeder path trace if a conductor is found that satisfies (5) and (6), then a fixed capacitor is inserted after the conductor. Power Flow is then run again to check for low voltage problems. If no low voltage violations exist, VPD exits; if there is any low voltage violation, and m is smaller than the *Maximum Number of Fixed Capacitors* requirement, step 3 is repeated before proceeding to step 4.

- 4) Switched shunt capacitors are placed next. At the maximum load condition, the lowest customer voltage point in the feeder is identified, and feeder path traces to the first conductor that satisfies the conditions defined in (7) and (8) are performed.

$$\max(Q_t) - C_F * m - C_S * n > C_S \quad (7)$$

$$\sum_{i=1}^N Q_i + \sum_{j=1}^K Q_{LOSSj} - \sum_{k=1}^Y C_k - C_S > 0 \quad (8)$$

, where

C_S = switched capacitor size per phase (*Switched Cap Size*);

n = number of the switched capacitors previously installed by VPD.

A switched capacitor is then inserted after the conductor that satisfies (7) and (8). If any low voltage violations remain, step 4 is repeated until n reaches the **Maximum Number of Switched Capacitors**, or condition (7) is violated. If in either case low voltage violations still exist, VPD proceeds to step 5; otherwise VPD exits.

- 5) Lastly, voltage regulators are placed. At the maximum load condition, a feeder path trace is performed starting from the lowest customer voltage location in the feeder until a conductor that satisfies the conditions defined in (9) is found.

$$V_x - V_{lowest} \geq V_{zone} \quad (9)$$

, where

V_x = per phase average customer level voltage at a given location;

V_{lowest} = customer level voltage at the lowest voltage location within a regulator control zone;

V_{zone} = Regulator Control Zone Voltage Drop.

A voltage regulator is inserted after the conductor identified by (9). If any low voltage violations remain, step 5 is repeated until the number of regulators reaches the **Maximum Number of Voltage Regulators**. If the maximum number of regulators is

reached and there are still low voltage violations, VPD proceeds to step 6; otherwise VPD exits.

- 6) Voltage set points on all voltage regulators, including the LTC set point, are raised by a 0.5 volt increment until there are no low voltage violations or until the *Upper Limit* of the input *Voltage Range* is reached. This is where VPD ends.

Thus using this approach, the VPD algorithm not only gives the appropriate location but also the required number of devices to be placed. It can help in maintaining a flat voltage profile in a distribution feeder. Also, by avoiding overcompensation, good power factors can be maintained at all times.

The next task was the validation of this algorithm. Distribution feeders vary widely in their characteristics. The line length, number of existing voltage control devices, kW and kVA ratings, etc., vary widely. So, testing the algorithm on random feeders cannot validate it. Clustering analysis was therefore used to choose representative feeders which contained the characteristics considered important for PV hosting capacity. The VPD algorithm was then used to redesign the voltage profile of these feeders and a PV hosting capacity study was performed to determine the change in the feeder's PV hosting capacity.

4. Clustering Analysis

The main objective of this research was to see if the PV hosting capacity of distribution feeders can be increased by redesigning the voltage profile using the VPD algorithm. So clustering analysis was used to select the feeders for evaluating the VPD algorithm. Clustering, followed by PV hosting capacity studies on some randomly chosen feeders from the final clusters helped in highlighting the characteristics most important for the variation in PV hosting capacity of feeders. These characteristics were then used for choosing detailed feeder models from a North American utility on which the VPD algorithm was evaluated.

The analysis presented here has been conducted on a dataset of 3000+ distribution feeders, with about 30 different characteristics of each feeder, from the state of California, which has already experienced a large amount of PV penetration. An exhaustive sequence of steps was developed based on a thorough review of similar feeder taxonomy studies done worldwide, and the most important technique(s) for accomplishing those steps were implemented on the feeder dataset.

4.1. Parameter Selection

The first step is the creation of a database. This step was not performed in this research, as the database was already available, however, any future database creation could certainly benefit from this discussion.

Engineering experience was used for making the initial parameter selection in almost all of the studies. Furthermore, the selection of parameters should be such that it covers all the parameters which are considered even remotely important to the objectives of the project.

The parameters to be included can vary widely based on the geographical area considered. For instance, a study consisting of feeder data at the national level may require inclusion of geographical parameters. The feeder taxonomy study at Pacific Northwest National Laboratory (PNNL) aimed at creating feeder taxonomy for the entire nation [26]. Considering the vast geographical area, the study categorized the feeders in terms of climatic regions as well. The

original 89 parameters considered were narrowed down to 35 variables by grouping them into voltage and region classes. The feeders belonging to slightly different voltage classes like 12.00kV, 12.47 kV, 12.5 kV and 13.8 kV, were grouped into a single voltage class of 12.47 kV. A similar approach was also followed by Broderick et al as well to narrow down the clustering variables [40]. Furthermore, the PNNL study used certain graph theory concepts, like degree, diameter and closeness, which can aid in capturing a wide variety of feeder parameters in fewer metrics.

Once the data is available for these parameters from the participating utilities for a large number of feeders, it might be worthwhile to ensure that this data is actually representative of the feeders at the national level. This might be a bit difficult to do if a national survey of feeder characteristics does not exist. But in the presence of one it will definitely be a strong validating point for the final clustering. This was done in the research by Commonwealth Scientific and Industrial Research Organization (CSIRO) and Ausgrid for creating national feeder taxonomy for Australian feeders [28]. The set of feeders being considered for the study were divided into groups based on voltage class and states and voltage class and reliability (short rural, urban etc.). The percentage of feeders in each group was compared against the data from a national feeder survey. A similar percentage of feeders in each group gave credence to the fact that the feeders being used for clustering are indeed representative of the feeders surveyed at the national level.

Finally, going into the actual parameters to be used for the taxonomy, it is a gray area with a lot of uncertainty and variability involved. It will depend heavily on the objectives of the project, and is also restrained by the availability of the data. However, specifically with the objective of creating feeder taxonomy for higher PV penetration, several parameters were common to most of the studies. The feeder kVA rating, primary voltage, feeder length and the percentage of overhead

and underground cables, number of regulators and capacitors, number of delivery points, installed PV capacity, feeder impedance and type of customers were mostly used. For low voltage (LV)

Table 1: Common parameters used for creation of a feeder taxonomy

Studies	NREL and APS	Sandia Labs	PNNL	CSIRO & Ausgrid	German Benchmark feeders	German Voltage band study
Parameters						
Feeder kVA Rating	✓	✓	✓	✓	*	*
Primary Voltage	*	✓	✓	✓	*	*
Feeder length	✓	✓	✓	✓	✓	✓
Customer Count or type	✓	✓	✓	✓	✓	✓
Voltage Regulators	✓	✓		✓		
Distinction between Underground and overhead cabling	✓	✓	✓	✓	✓	✓

*The studies in Germany used LV grids which are almost at the same voltage level and variance due to kVA rating was negligible. Similarly the NREL and APS study considered primary voltage to be an important criterion but was not included probably because the feeders considered were at the same voltage level.

feeders, branching and distance between delivery points were also considered to be very relevant [27]. Again these are only some of the common aspects and are by no means exhaustive. Table 1 shows some of the parameters commonly used in most of the studies.

4.2. Feeder selection and data preprocessing

The data for the selected parameters has to be obtained from participating utilities. There are several important points to consider while selecting the feeders for which the data will be used. First of all the more the number of feeders considered the better. However, this will be constrained by a number of practical limitations. Gathering data is a time consuming and tedious task. There could certainly be a number of feeders amongst the thousands of feeders under a utility's purview for which data is not available, or is insufficient. Non-availability of data, even for a few parameters, may necessitate elimination of those feeders. Similarly selecting the feeders for which data will be used should be unbiased. Feeders should be selected from across the area under a utility, and not simply a bunch of feeders from the same substation. This will ensure true representation as feeder data for multiple feeders from a single substation tends to be similar.

A very important consideration, especially while dealing with multiple utilities, is to ensure that a common method has been followed for determining the value of a parameter. The same problem may very well exist within a single utility as well. For instance, kVA rating of the feeder is a parameter used in almost all the studies. There may be a case where one utility has used SCADA or smart meter data for determining the kVA and another might have simply used the distribution transformers rating as a metric to determine the same. Most likely the latter would tend to overestimate the kVA rating. It might then be necessary to introduce a scaling factor so that the two are comparable.

One key point to make the research results more holistic would be the inclusion of meshed networks as well. Most of the research considers only radial distribution feeders. The feeder taxonomy research by CSIRO and Ausgrid does include meshed networks, as they consider Central Business Districts (CBDs), which usually tend to be meshed [28]. Also, it might be prudent to group them as a separate cluster altogether because their characteristics will be significantly different than other networks in terms of kVA/m, cable length, etc.

Having considered these points, and once all the data is available, a final data check for missing or abnormal data will be helpful in formation of meaningful clusters. Then the variables should be transformed in such a way that they foster clustering. This can be illustrated by again considering the CSIRO and Ausgrid research [28]. The kVA rating for the short rural, long rural and urban feeders was very close to one another for the feeders considered in the New South Wales/Australian Capital Territory (NSW/ACT) area. They however represent very different types of feeders and should be clustered separately. Thus, simply using kVA rating is not a very effective method. So, instead of kVA rating, kVA/m was used, as this parameter will vary vastly among these types of feeders, as urban feeder length is considerably less. Having completed these two steps, the statistical analysis process of forming the clusters may be started.

To summarize, the gist of these initial steps is to use an unbiased selection so that the feeder data set is representative of the entire region, and to use a similar methodology to determine the parameter values. Also, the initial data set should include all parameters that are considered important to determining PV hosting capacity. A simple box plot of the parameters can help in detecting bad data. This is evident in figure 6, which shows negative values of summer and winter kW demand. These were rejected from the cluster analysis as obvious bad data.

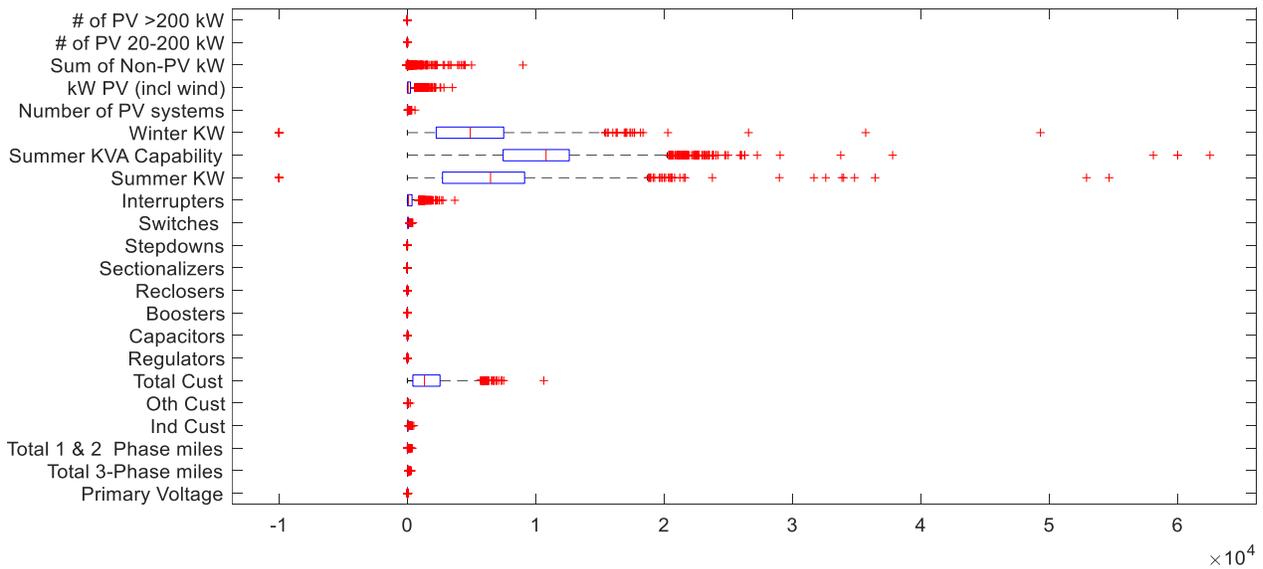


Figure 6: Box-plot of parameters after removal of correlated parameters

4.3. Data Scaling

This is the first step to be followed in the clustering process and is unavoidable. The clustering methods rely heavily on the variance and co-variance between the parameters considered. These values will be highly biased if the parameters vary vastly in their range. For example, the range for feeder length could easily be several hundred kilometers, whereas the number of regulators may be between 0 and 5. Thus, to give equal weightage to each variable considered, they have to be scaled. This step has been used globally, but tends to vary in the scaling methodology followed. Overall three different scaling methods could be found in the feeder taxonomies considered in this report as follows:

- Unit Variance Scaling and Mean Centering Transformation
- Probability Integral Transformation

- Natural log

- ***Unit Variance Scaling and Mean Centering Transformation (SDM)***

This technique has been used in [40], [41]. This involves two steps. The first step, unit variance scaling, involves dividing each value of a variable by its standard deviation. This ensures that each variable has unit variance. The second step is mean centering transformation where the mean for each variable is calculated and subtracted from each entry. This further improves the interpretability of the model.

- Dividing each of the variable column vectors \mathbf{v}_i by its standard deviation

$$\mathbf{v}_i \rightarrow \frac{\mathbf{v}_i}{\sigma(\mathbf{v}_i)} \quad (10)$$

- Remove the mean from the column vectors:

$$\mathbf{v}_i \rightarrow \mathbf{v}_i - \mu(\mathbf{v}_i) \quad (11)$$

- ***Probability Integral Transformation (PIT)***

This technique has been used in the research by PNNL [26]. This involves subtracting the minimum value of each variable from all of its entries followed by dividing each entry by the range of the data.

$$PIT = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (12)$$

This transformation ensures that all the variables lie in the range from 0 to 1.

- ***Natural log***

This has been used in the research by CSIRO and Ausgrid [28]. In this research several parameters like number of customers, feeder kVA rating etc., which could range in the thousands, were transformed by taking their natural log. The rest of the parameters were left as is.

To choose the data scaling technique, a comparison of the ability of the techniques to maintain the structure of the data was done, as shown in figure 7. The data set is a synthetic dataset with a large range. SDM refers to the first technique, PIT to the second, and log to the third. The actual parameter values are plotted on the secondary vertical axis. SDM is not only most widely used but also maintains the data structure most accurately amongst these three. So SDM was chosen as the data scaling technique and all the entries of the dataset were normalized.

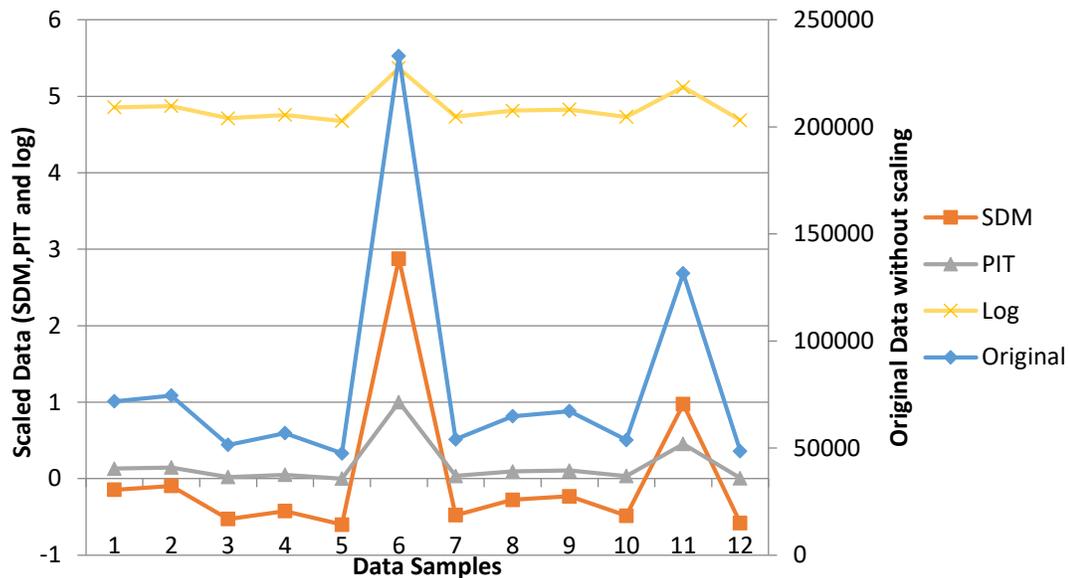


Figure 7: Comparison of data scaling techniques

4.4. Outlier Rejection

This step is necessary to ensure that the members of a cluster indeed belong there. Each feeder belonging to a cluster should be very similar to the others. If there is a feeder which is not similar to any other feeder and has been randomly assigned to a cluster, then it might skew the overall properties of that cluster. Also the clustering methods like k-means and k-medoids are susceptible to outliers [41]. So before proceeding to the other clustering steps the outliers were removed.

There are various methods which can be used for rejecting outliers. However, caution should be exercised before completely eliminating them. A feeder with properties unlike any other may be a candidate for further detailed examination. So they may be removed from the normal clustering process, but not eliminated completely from the examination process. It may be advisable to form entirely separate clusters for these outliers. This was done in the research by

CSIRO and Ausgrid [28]. Two different types of feeders, which might have acted as outliers, were removed from the clustering process and assigned to their own clusters. The first one was a very long rural feeder with an overhead length of 1828 km, which would have otherwise been grouped in the clusters for long rural feeders with a maximum overhead length of just 580 km. The other was the Central Business Districts (CBDs) which would have belonged to the cluster of urban feeders which have less than half the kVA/km than the CBDs. However as is evident, complete elimination of these feeders would have made the research less representative of the actual feeders that exist.

The method used in this research is the ‘distance from the 12 closest neighbors’. This method is discussed in the research by Broderick et al and can be used before starting the clustering process [40], unlike some other methods as mentioned in section 4.5. The outliers were determined by defining a threshold based on the average distance of a feeder from its 12 closest neighbors. The threshold was determined by visual inspection, and a line was drawn where the separation was most apparent, as shown in Figure 8. All the feeders above the threshold value of 8 were rejected as outliers.

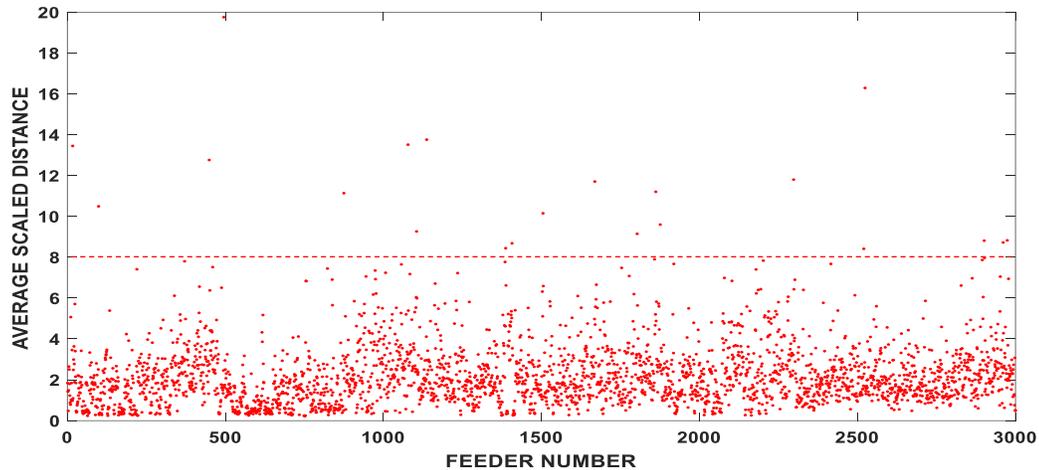


Figure 8: Outlier rejection using distance from 12 closest neighbors

The distance function used for subsequent clustering steps is the ‘Cityblock’ function rather than Euclidean function, because it results in better clustering. For example, it led to a better cophenetic correlation coefficient when hierarchical clustering was used. It is also suggested to be more robust to outliers when used in conjunction with clustering algorithms [41]. The generalized distance function is:

$$\left[\sum_{i=1}^N |x_i - y_i|^p \right]^{\frac{1}{p}} \quad (13)$$

Substituting $p=1$ gives the ‘Cityblock’ function, and $p=2$ gives the commonly used Euclidean distance function.

4.5. Principal Component Analysis (PCA)

PCA is a widely used and established technique for statistical analysis. It uses orthogonal transformation to convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. It helps in representing the data along the axes of maximum variance. The first principal component corresponds to the largest eigenvalue of the covariance matrix of the variables. Also, it represents a major chunk of the variance in the data. Thus, by plotting the feeder data with the first two principal components as the plot axes, the data can be visualized in terms of the variance that exists between them. Each point on this plot will represent a feeder, and the feeders closer together will be similar to each other. Another advantage with this technique is that multiple clustering steps can be accomplished from a single analysis, provided the first two principal components contains a large percentage of the total variance ($> \frac{2^{rds}}{3}$).

PCA was applied to the scaled feeder dataset as shown in figure 9. However, the first two principal components had lesser than $\frac{2^{rds}}{3}$ of the total variance. So PCA alone could not be used for accomplishing clustering, but was used to verify the results obtained from other techniques. The length of the lines in figure 9 indicate the contribution of each parameter in the two principal components.

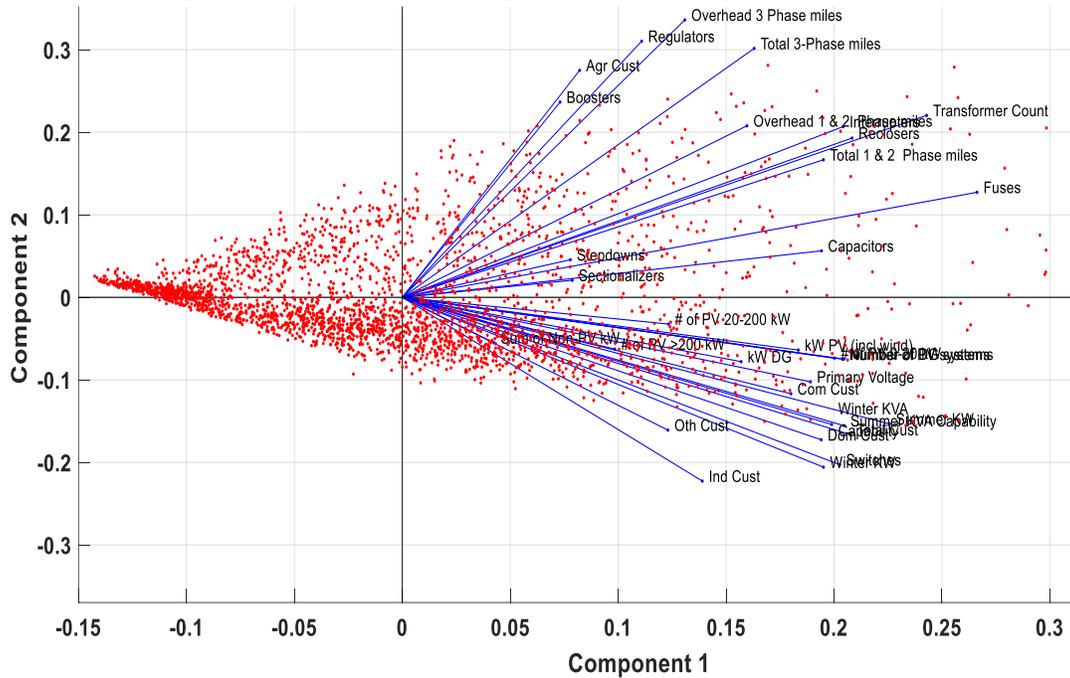


Figure 9: Feeders and all parameters on first two principal components

Having completed PCA, the final result can be used for several subsequent steps as follows:

- **Variable Reduction**

Due to lesser variance exhibited by the first two principal components, PCA results were only used for verifying this step. As can be seen in Figure 9, there are several variables which are almost overlapping. One of these overlapping variables can be eliminated to achieve better clustering, as suggested in [27], [41]. The reason is that clustering methods rely on variance in the data. Thus, two overlapping variables will essentially convey the same information.

- **Clustering**

The clustering methods will be discussed in a subsequent section. However assuming the method is known, it can directly be applied to the data as presented in terms of the first two principal components. This will only be true if the first two principal components have a larger share of the total variance, unlike figure 9. All the principal components can also be used, in which case there will not be any loss of information.

- ***Outlier Rejection***

This step, as discussed in section 4.4, was completed using the ‘average distance from the 12 closest neighbors’ method – however, PCA may also be used, but only after clusters have been formed. After clustering formation, several principal components based techniques can be used for outlier rejection, such as the error ellipse method using a 95% confidence interval [41]. Again, this could not be accomplished due to the lesser variance in the first two principal components.

Although each of these steps may be completed using other techniques, PCA is still a powerful technique, and can at least be used as a verification method.

4.6. Variable Reduction

Variable reduction techniques have been used in almost all the taxonomy studies. Optimal clustering can be obtained if dependent parameters are removed. This has been used in [40] and suggested in [41]. The reason being that the optimal number of clusters are determined with higher accuracy if the variables are uncorrelated and independent of each other. The effect of removing correlated variables on clustering has also been verified.

PCA, as discussed in the previous sub-section, can very well be the starting step, giving an initial idea regarding which variables may be rejected. The following two methods have been used in addition to PCA for outlier rejection.

- ***Covariance Heat maps***

This is just a graphical way of representing the covariance matrix of the feeder variables and have been used in [40], [41]. Figure 10 shows the covariance heat map between all variables considered initially. The regions in white represent higher correlation. One of the parameters with a higher correlation may be eliminated.

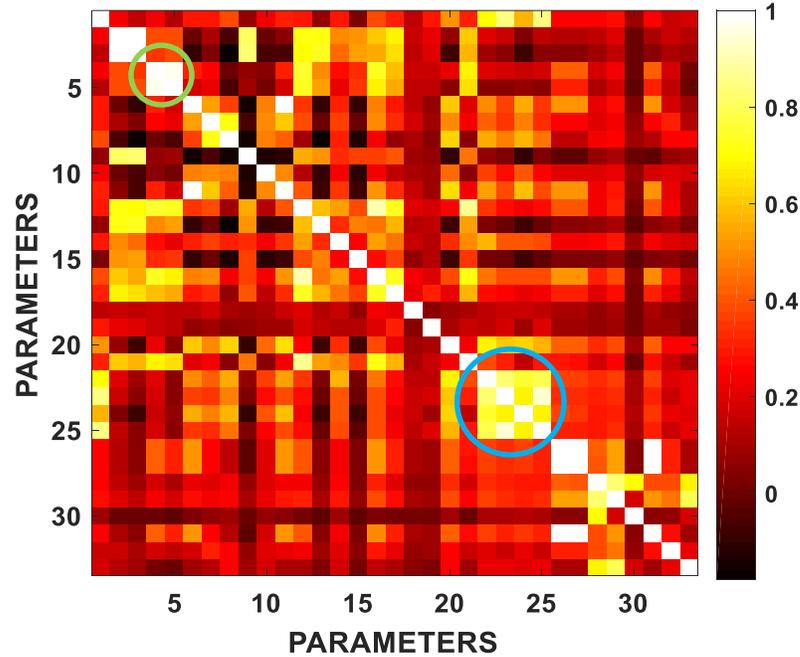


Figure 10: Covariance heat map

- ***Kendall Rank Correlation Coefficient***

This method was used for elimination of redundant input parameters by using their statistical dependence in [30]. This technique makes use of concordant and discordant set of pairs. Considering a set of observations $(x_1, y_1), \dots, (x_n, y_n)$; any pair of observations (x_i, y_i) and (x_j, y_j) such that $i \neq j$, is set to be concordant (or their ranks agree) if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. Otherwise if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$ then the pair is said to be discordant [42], [43]. The maximum number of pairs which may differ are $\frac{1}{2}(n)(n-1)$. So the Kendall rank correlation coefficient is defined as:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}(n)(n-1)}. \quad (14)$$

The denominator has the maximum possible number of pairs. Thus if all pairs are concordant then $\tau = 1$, else if all pairs are discordant then $\tau = -1$ and if equal number of concordant and discordant pairs exist then $\tau = 0$. $\tau = 1/-1$ represents perfect agreement/disagreement and $\tau = 0$

implies the variables are mutually independent. Thus a value close to 1 should indicate higher correlation and can serve as a metric for elimination.

Tables 2 and 3 show the numerical values of the covariance and the kendall rank correlation coefficients between the parameters, respectively. Based on similar results from the three methods, i.e. higher covariance, higher values of Kendall rank correlation coefficients and overlap of the same variables in the principal component axes of figure 9, 11 correlated, or dependent, variables were removed. For example, the green circle in figure 10 shows the high covariance among the ‘total 1 and 2 phase miles’ and ‘overhead 1 and 2 phase miles’. The same can be said for the winter and summer kVA capability, as shown in the blue circle. These parameters were also found to have higher τ values and overlap in Fig. 9 as well. After variable elimination, the original dataset of 33 variables was reduced to 22 independent variables. The parameters which were removed are:

- Winter kVA capability,
- Overhead 3 phase miles,
- Overhead 1 and 2 phase miles,
- transformer count,
- fuses,
- domestic customers,
- commercial customers,
- agricultural customers,
- kW DG,
- number of DG systems and
- number of PV systems between 0 and 20 kW

Table 2: Covariance between parameters. All covariances above 0.5 are highlighted

PARAMETERS	'ary	'Total Overht	'Total	'Overht	'Dom	'Com	'Ind C	'Agr C	'Oth C	'Total	'Tran	'Regul	'Capa	'Boost	'Fuses	'Reclo	'Secti	'Steps	'Switch	'Interri	'Sumi	'Sumn	'Winte	'Numb	'Numb	'kW D	'kW f	'Sum of	'# of P	'# of P				
'Primary Voltage'	1.00	0.23	0.14	0.23	0.12	0.28	0.28	0.39	0.07	0.24	0.31	0.28	0.07	0.29	0.04	0.38	0.30	0.14	0.29	0.51	0.31	0.69	0.84	0.57	0.84	0.26	0.26	0.25	0.27	0.09	0.25	0.15	0.18	
'Total 3-Phase miles'	0.23	1.00	0.98	0.38	0.38	0.00	0.11	-0.06	0.80	0.00	0.05	0.72	0.72	0.48	0.52	0.60	0.67	0.16	0.22	0.07	0.60	0.20	0.16	0.02	0.15	0.12	0.12	0.15	0.20	0.00	0.11	0.16	0.12	
'Overhead 3 Phase miles'	0.14	0.98	1.00	0.35	0.38	-0.10	0.04	-0.16	0.82	-0.06	-0.06	0.69	0.73	0.44	0.53	0.54	0.64	0.14	0.20	-0.08	0.58	0.09	0.05	-0.08	0.05	0.05	0.05	0.09	0.14	-0.01	0.04	0.11	0.08	
'Total 1 & 2 Phase miles'	0.23	0.38	0.35	1.00	0.97	0.30	0.26	0.00	0.06	0.06	0.31	0.76	0.49	0.26	0.32	0.74	0.58	0.15	0.16	0.21	0.67	0.19	0.18	0.16	0.17	0.43	0.43	0.17	0.24	-0.01	0.43	0.17	0.02	
'Overhead 1 & 2 Phase miles'	0.12	0.38	0.38	0.97	1.00	0.18	0.22	-0.05	0.08	0.01	0.19	0.72	0.53	0.22	0.34	0.67	0.56	0.15	0.14	0.05	0.64	0.06	0.05	0.07	0.05	0.30	0.30	0.11	0.17	-0.02	0.30	0.15	0.00	
'Dom Cust'	0.28	0.00	-0.10	0.30	0.18	1.00	0.53	0.36	-0.16	0.47	0.99	0.35	-0.10	0.30	-0.09	0.48	0.29	0.14	0.07	0.61	0.20	0.50	0.37	0.56	0.35	0.52	0.52	0.21	0.27	0.04	0.52	0.18	0.09	
'Com Cust'	0.28	0.11	0.04	0.26	0.22	0.53	1.00	0.71	-0.05	0.49	0.61	0.38	0.00	0.36	-0.01	0.49	0.32	0.18	0.09	0.49	0.23	0.43	0.34	0.47	0.33	0.26	0.26	0.20	0.23	0.06	0.25	0.30	0.13	
'Ind Cust'	0.39	-0.06	-0.16	0.00	-0.05	0.36	0.71	1.00	-0.16	0.48	0.42	0.09	-0.18	0.30	-0.13	0.24	0.07	0.10	0.06	0.62	0.05	0.55	0.48	0.57	0.46	0.13	0.13	0.24	0.26	0.09	0.12	0.26	0.26	0.26
'Agr Cust'	0.07	0.80	0.82	0.06	0.08	-0.16	-0.05	-0.16	1.00	-0.12	-0.11	0.55	0.52	0.32	0.37	0.36	0.39	0.09	0.07	-0.10	0.46	0.08	0.04	-0.09	0.03	-0.02	-0.02	0.08	0.11	-0.01	-0.02	0.06	0.09	0.09
'Oth Cust'	0.24	0.00	-0.06	0.06	0.01	0.47	0.49	0.48	-0.12	1.00	0.50	0.13	-0.09	0.23	-0.11	0.24	0.14	0.11	0.08	0.44	0.05	0.36	0.27	0.40	0.26	0.19	0.19	0.20	0.22	0.07	0.18	0.20	0.17	0.17
'Total Cust'	0.31	0.05	-0.06	0.31	0.19	0.99	0.61	0.42	-0.11	0.50	1.00	0.40	-0.07	0.34	-0.07	0.52	0.33	0.15	0.08	0.63	0.23	0.53	0.39	0.58	0.37	0.51	0.51	0.23	0.28	0.04	0.51	0.21	0.11	0.11
'Transformer Count'	0.28	0.72	0.69	0.76	0.72	0.35	0.38	0.09	0.55	0.13	0.40	1.00	0.58	0.52	0.39	0.90	0.73	0.21	0.18	0.25	0.86	0.35	0.27	0.24	0.26	0.40	0.40	0.23	0.31	0.01	0.40	0.23	0.10	0.10
'Regulators'	0.07	0.72	0.73	0.49	0.53	-0.10	0.00	-0.18	0.52	-0.09	-0.07	0.58	1.00	0.33	0.47	0.45	0.56	0.12	0.14	-0.08	0.53	0.03	0.00	-0.08	-0.01	0.06	0.06	0.06	0.10	-0.02	0.06	0.09	0.04	0.04
'Capacitors'	0.29	0.48	0.44	0.26	0.22	0.30	0.36	0.30	0.32	0.23	0.34	0.52	0.33	1.00	0.25	0.53	0.45	0.16	0.12	0.33	0.48	0.56	0.40	0.40	0.38	0.24	0.24	0.26	0.29	0.09	0.23	0.19	0.18	
'Boosters'	0.04	0.52	0.53	0.32	0.34	-0.09	-0.01	-0.13	0.37	-0.11	-0.07	0.39	0.47	0.25	1.00	0.28	0.36	0.08	0.12	-0.08	0.35	0.02	-0.01	-0.07	-0.02	0.03	0.03	0.03	0.06	-0.03	0.03	0.05	0.02	
'Fuses'	0.38	0.60	0.54	0.74	0.67	0.48	0.49	0.24	0.36	0.24	0.52	0.90	0.45	0.53	0.28	1.00	0.73	0.24	0.21	0.39	0.71	0.46	0.39	0.37	0.38	0.51	0.51	0.30	0.39	0.04	0.50	0.29	0.15	
'Reclosers'	0.30	0.67	0.64	0.58	0.56	0.29	0.32	0.07	0.39	0.14	0.33	0.73	0.56	0.45	0.36	0.73	1.00	0.22	0.32	0.22	0.57	0.27	0.25	0.21	0.23	0.31	0.31	0.20	0.27	0.02	0.30	0.21	0.11	
'Sectionalizers'	0.14	0.16	0.14	0.15	0.15	0.14	0.18	0.10	0.09	0.11	0.15	0.21	0.12	0.16	0.08	0.24	0.22	1.00	0.05	0.09	0.10	0.14	0.15	0.16	0.15	0.11	0.11	0.08	0.10	0.01	0.11	0.07	0.05	
'Stepdowns'	0.29	0.22	0.20	0.16	0.14	0.07	0.09	0.06	0.07	0.08	0.08	0.18	0.14	0.12	0.12	0.21	0.32	0.05	1.00	0.11	0.16	0.13	0.18	0.08	0.20	0.08	0.08	0.07	0.09	0.01	0.08	0.09	0.04	
'Switches'	0.51	0.07	-0.08	0.21	0.05	0.61	0.49	0.62	-0.10	0.44	0.63	0.25	-0.08	0.33	-0.08	0.39	0.22	0.09	0.11	1.00	0.23	0.68	0.60	0.61	0.57	0.43	0.43	0.33	0.39	0.09	0.42	0.26	0.25	
'Interrupters'	0.31	0.60	0.58	0.67	0.64	0.20	0.23	0.05	0.46	0.05	0.23	0.86	0.53	0.48	0.35	0.71	0.57	0.10	0.16	0.23	1.00	0.37	0.30	0.14	0.29	0.28	0.28	0.18	0.25	0.00	0.28	0.15	0.09	
'Summer KW'	0.69	0.20	0.09	0.19	0.06	0.50	0.43	0.55	0.08	0.36	0.53	0.35	0.03	0.56	0.02	0.46	0.27	0.14	0.13	0.68	0.37	1.00	0.83	0.75	0.79	0.37	0.37	0.33	0.36	0.11	0.36	0.20	0.22	
'Summer KVA Capability'	0.84	0.16	0.05	0.18	0.05	0.37	0.34	0.48	0.04	0.27	0.39	0.27	0.00	0.40	-0.01	0.39	0.25	0.15	0.18	0.60	0.30	0.83	1.00	0.68	0.94	0.32	0.32	0.31	0.33	0.11	0.31	0.16	0.22	
'Winter KW'	0.57	0.02	-0.08	0.16	0.07	0.56	0.47	0.57	-0.09	0.40	0.58	0.24	-0.08	0.40	-0.07	0.37	0.21	0.16	0.08	0.61	0.14	0.75	0.68	1.00	0.67	0.35	0.30	0.31	0.12	0.35	0.22	0.19		
'Winter KVA Capability'	0.84	0.15	0.05	0.17	0.05	0.35	0.33	0.46	0.03	0.26	0.37	0.26	-0.01	0.38	-0.02	0.38	0.23	0.15	0.20	0.57	0.29	0.79	0.94	0.67	1.00	0.31	0.31	0.29	0.32	0.10	0.30	0.17	0.21	
'Number of DG systems'	0.26	0.12	0.05	0.43	0.30	0.52	0.26	0.13	-0.02	0.19	0.51	0.40	0.06	0.24	0.03	0.51	0.31	0.11	0.08	0.43	0.28	0.37	0.32	0.35	0.31	1.00	1.00	0.41	0.53	0.05	1.00	0.30	0.11	
'Number of PV systems'	0.26	0.12	0.05	0.43	0.30	0.52	0.26	0.13	-0.02	0.19	0.51	0.40	0.06	0.24	0.03	0.51	0.31	0.11	0.08	0.43	0.28	0.37	0.32	0.35	0.31	1.00	1.00	0.41	0.53	0.05	1.00	0.30	0.11	
'kW DG'	0.25	0.15	0.09	0.17	0.11	0.21	0.20	0.24	0.08	0.20	0.23	0.23	0.06	0.26	0.03	0.30	0.27	0.10	0.09	0.39	0.25	0.36	0.33	0.31	0.30	0.29	0.41	1.00	0.82	0.70	0.39	0.38	0.67	
'kW PV (incl wind)'	0.27	0.20	0.14	0.24	0.17	0.27	0.23	0.26	0.11	0.22	0.28	0.31	0.10	0.29	0.06	0.39	0.27	0.10	0.09	0.39	0.25	0.36	0.33	0.31	0.32	0.53	0.53	0.82	1.00	0.17	0.50	0.47	0.79	
'Sum of Non-PV kW'	0.09	0.00	-0.01	-0.01	-0.02	0.04	0.06	0.09	-0.01	0.07	0.04	0.01	-0.02	0.09	-0.03	0.04	0.02	0.01	0.01	0.09	0.00	0.11	0.11	0.12	0.10	0.05	0.05	0.70	0.17	1.00	0.04	0.07	0.17	
'# of PV 0-20 kW'	0.25	0.11	0.04	0.43	0.30	0.52	0.25	0.12	-0.02	0.18	0.51	0.40	0.06	0.23	0.30	0.50	0.30	0.11	0.08	0.42	0.28	0.36	0.31	0.35	0.30	1.00	1.00	0.39	0.50	0.04	1.00	0.27	0.09	
'# of PV 20-200 kW'	0.15	0.16	0.11	0.17	0.15	0.18	0.30	0.26	0.06	0.20	0.21	0.23	0.09	0.19	0.05	0.29	0.21	0.07	0.09	0.26	0.15	0.20	0.16	0.22	0.17	0.30	0.30	0.38	0.47	0.07	0.27	1.00	0.21	
'# of PV >200 kW'	0.18	0.12	0.08	0.02	0.00	0.09	0.13	0.26	0.09	0.17	0.11	0.10	0.04	0.18	0.02	0.15	0.11	0.05	0.04	0.25	0.09	0.22	0.22	0.19	0.21	0.11	0.11	0.67	0.79	0.17	0.09	0.21	1.00	

Table 3: Kendall Rank Correlation Coefficients between parameters. All coefficients above 0.5 are highlighted

	'Prim'	'Total'	'Over'	'Total'	'Over'	'Dom'	'Com'	'Ind C'	'Agr C'	'Oth C'	'Total'	'Transf'	'Regul'	'Capa'	'Boos'	'Fuses'	'Reclo'	'Sectid'	'Stept'	'Switch'	'Inter'	'Summ'	'Winte'	'Winte'	'Numt'	'Numt'	'kW F'	'Sum q'	'# of P'	'# of P'						
PARAMETERS (Rule>0.5)	1.00	0.33	0.19	0.28	0.11	0.20	0.26	0.34	0.16	0.22	0.23	0.31	0.08	0.27	0.04	0.38	0.31	0.12	0.33	0.45	0.33	0.55	0.63	0.46	0.63	0.18	0.17	0.25	0.24	0.21	0.16	0.17				
'Primary Voltage'		1.00	0.78	0.44	0.40	0.17	0.24	0.12	0.63	0.13	0.21	0.66	0.54	0.44	0.39	0.60	0.60	0.17	0.18	0.24	0.52	0.29	0.16	0.22	0.24	0.24	0.26	0.28	0.22	0.23	0.16	0.13				
'Total 3-Phase miles'			1.00	0.78	0.44	0.17	0.24	0.12	0.63	0.13	0.21	0.66	0.54	0.44	0.39	0.60	0.60	0.17	0.18	0.24	0.52	0.29	0.16	0.22	0.24	0.24	0.26	0.28	0.22	0.23	0.16	0.13				
'Overhead 3 Phase miles'				1.00	0.42	0.49	0.10	0.18	0.00	0.64	0.06	0.13	0.62	0.56	0.41	0.40	0.53	0.59	0.16	0.15	0.06	0.47	0.14	0.09	0.04	0.10	0.19	0.19	0.21	0.15	0.19	0.13	0.08			
'Total 1 & 2 Phase miles'					1.00	0.77	0.47	0.34	0.16	0.24	0.18	0.47	0.64	0.27	0.31	0.18	0.63	0.48	0.14	0.19	0.32	0.46	0.26	0.21	0.22	0.51	0.51	0.38	0.41	0.35	0.52	0.19	0.09			
'Overhead 1 & 2 Phase miles'						1.00	0.34	0.30	0.08	0.26	0.13	0.34	0.59	0.33	0.29	0.22	0.55	0.48	0.15	0.15	0.15	0.39	0.11	0.06	0.09	0.40	0.41	0.29	0.32	0.27	0.41	0.18	0.06			
'Dom Cust'							1.00	0.48	0.35	-0.06	0.41	0.93	0.38	-0.08	0.26	0.07	0.41	0.25	0.10	0.08	0.48	0.21	0.35	0.23	0.52	0.53	0.37	0.39	0.36	0.53	0.19	0.08				
'Com Cust'								1.00	0.61	0.03	0.44	0.54	0.41	0.00	0.34	0.00	0.45	0.29	0.12	0.11	0.48	0.26	0.35	0.25	0.38	0.26	0.34	0.34	0.32	0.33	0.28	0.15				
'Ind Cust'									1.00	-0.08	0.46	0.40	0.21	-0.15	0.31	-0.11	0.29	0.12	0.07	0.08	0.56	0.14	0.45	0.37	0.47	0.36	0.21	0.21	0.29	0.28	0.27	0.19	0.25	0.21		
'Agr Cust'										1.00	-0.09	-0.04	0.45	0.56	0.30	0.42	0.37	0.45	0.11	0.11	0.01	0.43	0.12	0.09	-0.04	0.09	0.08	0.08	0.11	0.13	0.07	0.08	0.07	0.08		
'Oth Cust'											1.00	0.43	0.21	-0.12	0.25	-0.10	0.27	0.14	0.08	0.06	0.43	0.11	0.31	0.22	0.35	0.21	0.24	0.24	0.27	0.27	0.25	0.22	0.23	0.19		
'Total Cust'												1.00	0.42	-0.06	0.30	-0.06	0.45	0.28	0.11	0.09	0.52	0.24	0.38	0.26	0.42	0.26	0.42	0.42	0.37	0.39	0.41	0.38	0.52	0.22	0.10	
'Transformer Count'													1.00	0.40	0.48	0.29	0.79	0.61	0.18	0.16	0.32	0.64	0.34	0.26	0.24	0.26	0.42	0.42	0.37	0.39	0.32	0.42	0.22	0.12		
'Regulators'														1.00	0.24	0.46	0.33	0.48	0.12	0.12	-0.05	0.35	0.02	0.00	-0.07	0.00	0.06	0.06	0.08	0.09	0.05	0.06	0.07	0.03		
'Capacitors'															1.00	0.21	0.47	0.40	0.14	0.12	0.32	0.42	0.46	0.33	0.32	0.32	0.24	0.24	0.28	0.28	0.24	0.23	0.19	0.17		
'Boosters'																1.00	0.23	0.33	0.12	0.10	-0.05	0.26	0.02	0.00	-0.05	0.00	0.04	0.04	0.06	0.07	0.03	0.04	0.04	0.01		
'Fuses'																	1.00	0.60	0.18	0.19	0.40	0.56	0.40	0.32	0.32	0.33	0.44	0.44	0.40	0.42	0.35	0.43	0.26	0.15		
'Reclosers'																		1.00	0.19	0.24	0.24	0.45	0.25	0.21	0.19	0.21	0.30	0.30	0.28	0.30	0.24	0.30	0.19	0.11		
'Sectionalizers'																			1.00	0.05	0.08	0.10	0.12	0.11	0.10	0.10	0.09	0.09	0.09	0.08	0.09	0.08	0.05	0.08		
'Stepdowns'																				1.00	0.13	0.15	0.14	0.47	0.09	0.19	0.10	0.10	0.10	0.10	0.10	0.10	0.07	0.03		
'Switches'																					1.00	0.24	0.54	0.46	0.53	0.44	0.37	0.37	0.39	0.39	0.36	0.35	0.24	0.20		
'Interrupters'																						1.00	0.32	0.26	0.13	0.25	0.27	0.27	0.25	0.27	0.22	0.26	0.14	0.11		
'Summer KW'																							1.00	0.62	0.59	0.59	0.27	0.27	0.32	0.32	0.28	0.25	0.19	0.19		
'Summer KVA Capability'																								1.00	0.49	0.77	0.20	0.19	0.26	0.26	0.23	0.18	0.14	0.18		
'Winter KW'																									1.00	0.49	0.27	0.27	0.30	0.29	0.28	0.25	0.20	0.17		
'Winter KVA Capability'																										1.00	0.49	0.27	0.27	0.30	0.29	0.28	0.25	0.20	0.17	
'Number of DG systems'																											1.00	0.99	0.63	0.66	0.59	0.96	0.29	0.12		
'Number of PV systems'																												1.00	0.20	0.20	0.26	0.25	0.23	0.19	0.15	0.17
'kW DG'																													1.00	0.62	0.66	0.58	0.97	0.29	0.12	
'kW PV (incl wind)'																														1.00	0.92	0.78	0.58	0.43	0.46	
'Sum of Non-PV kW'																														1.00	0.70	0.62	0.45	0.48		
'# of PV 0-20 kW'																															1.00	0.55	0.38	0.31		
'# of PV 20-200 kW'																																1.00	0.24	0.08		
'# of PV >200 kW'																																	1.00	0.20	0.20	

4.7. Stopping Criterion or Optimal number of clusters

Determining the optimal number of clusters is the most important step. There are two primary types of clustering methods, which are discussed in detail in the next section. They differ in the way they approach the problem, but they require a stopping point, which is the optimal number of clusters to be formed. All the steps discussed till now have been implemented using similar techniques and only a few variations exist. This particular step on the other hand is unique in almost every research. The selection of stopping criteria for this research were based on the following points,

- A good starting point for choosing the stopping criteria is the research carried out by Milligan and Cooper [44]. They have compared 30 different stopping criteria in their ability to form optimal number of clusters. The synthetic datasets used in this paper have been designed such that they will lead to the formation of a specific optimal number of clusters. Thus, a good stopping criterion should be able to find those specific numbers of clusters. Further, since the datasets are designed to have clearly demarcated cluster boundaries, a method which doesn't perform well here is unlikely to perform better in real world applications where clear boundaries may not exist.
- Calinski-Harbasz method was found to be the best stopping criterion in this study and formed optimal clusters in 390 of the 432 datasets [44].
- If k-medoids is used for clustering, then average silhouette width may be used, as it is said to be the standard criterion for this method in [28].
- Irrespective of the method used, the optimal number of clusters are mostly less than 25 and depends on the parameters and geographical extent considered. The key point to consider is that the number of clusters is such that each major classification has a chance to form a separate cluster. This was done in the research by CSIRO and Ausgrid where the major classifications were three reliability classes (urban, short rural and long rural) and 4 voltage classes. So, the optimal average silhouette width was chosen to be its maximum value after 12 clusters so as to ensure that each group corresponding to particular reliability and voltage class at least had a chance of forming a separate cluster [28].

Considering these points, the Calinski-Harbasz and average silhouette width criterion were implemented. The optimal number of clusters was chosen after 7 to allow the 7 different voltage classes in the dataset to get clustered in separate clusters as suggested in [28].

- **Calinski Method**

This particular stopping criterion is used in [41]. This method requires maximizing a cost function $C_d(N_c)$ over N_c , where N_c is the optimal number of clusters. So in the function for $C_d(N_c)$, different values of N_c are substituted and the value for which $C_d(N_c)$ is maximum determines the optimum number of groups.

$$C_d(N_c) = \frac{(n - N_c)}{(N_c - 1)} \cdot \frac{tr(B(N_c))}{tr(W(N_c))} \quad (15)$$

where n is number of feeders; N_c is the number of clusters which has to be varied; tr is the matrix trace function and B and W are the between and within clubbed sum of squares and cross-product matrices, respectively [44]. The Calinski-Harabasz criterion did not yield a clear result; however, using elbow criterion [26], it was observed that after 9 clusters the change in values was

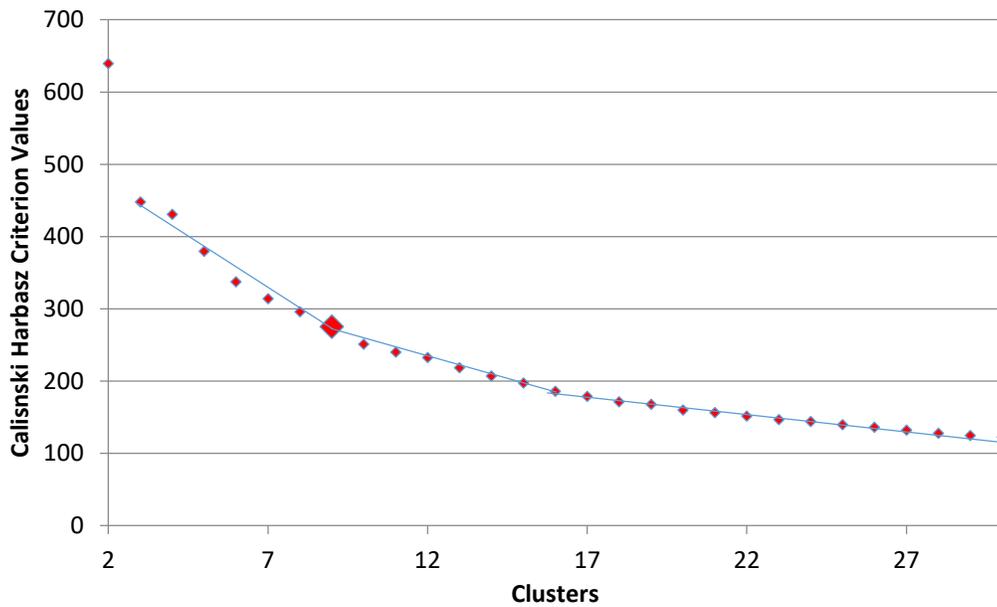


Figure 11: Optimal number of clusters using Calinski-Harbasz criterion

less gradual as shown in figure 11. This was further validated using the best possible sum of distances for varying cluster numbers, which again showed a similar trend.

- ***Average Silhouette Width***

This was used to determine optimal number of clusters in the research by CSIRO and Ausgrid [28] and is said to be the standard approach while using k-medoids method for clustering. Average Silhouette width,

$$s(e) = \frac{\text{intra_cluster_performance}(e) - \text{inter_cluster_performance}(e)}{\max(\text{inter_cluster_performance}(e), \text{intra_cluster_performance}(e))} \quad (16)$$

where,

- inter-cluster performance is simply the average distance from e to all other examples in the same cluster,
- intra-cluster performance is the maximum average-distance from example e to all examples in another cluster

The overall quality of a particular clustering is the average silhouette width. So, as average silhouette width gets closer to unity better clustering is obtained. The clusters will be closely knit when $s(e) = 1$.

- $s(e) = 1$: e is much more similar to its own cluster
- $s(e) = -1$: e is closer to another medoid
- $s(e) = 0$: e could belong to multiple medoids

So the maximum silhouette width after 7 was for 9 clusters as shown in figure 12. So based on similar results from the two criterion, 9 was chosen to be the optimal number of clusters for the case with independent parameters. A similar analysis was done when all parameters were considered, and that resulted in 11 as the optimal number of clusters. Further, because 10 replicates were used for each cluster number to avoid local minima, parallel computation was used to significantly reduce the computation time.

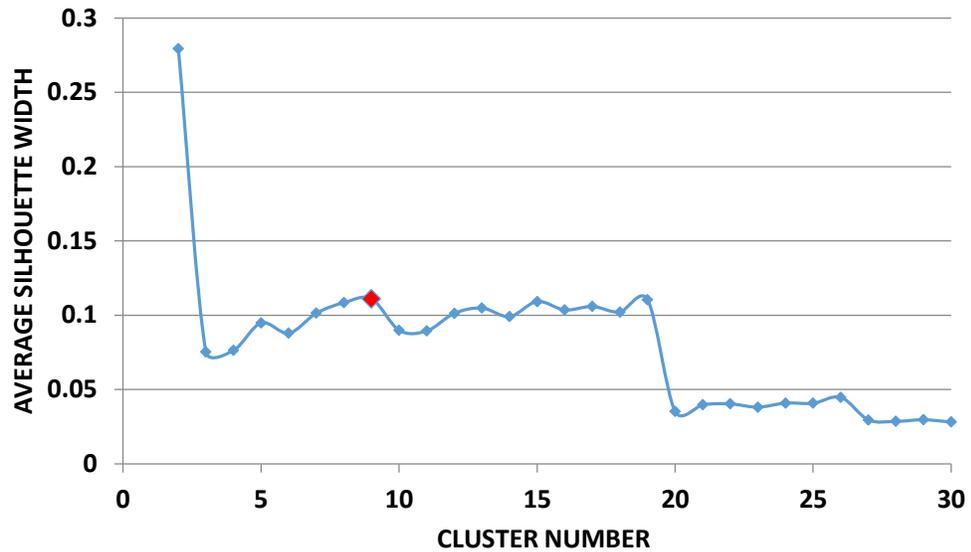


Figure 12: Optimal number of clusters using average silhouette width criterion

Now that the data is in a desirable format for cluster formation and the optimal number of clusters have been determined, the clusters can be formed.

4.8. Clustering

This is the final step in which the feeders are grouped, or clustered together, based on their similarities. The objective is to group the feeders in such a way that the feeders within a cluster are similar to each other and different from the feeders in other clusters. There are two main types of clustering methods:

- ***Hierarchical Clustering***

Here the clusters are not formed all at once. They follow either a bottom-up or top-down approach. They can be sub-divided into the following two categories:

- **Agglomerative methods:** In these methods initially all feeders are considered to be separate clusters. They are then fused together to form progressively larger clusters.
- **Divisive Methods:** These start with all feeders belonging to a single cluster initially. They are then divided into smaller clusters based on certain optimizing criterion.

The feeders are represented in the form of a dendrogram, with individual feeders at the bottom and single cluster at the top. Hierarchical agglomerative methods were used for clustering feeders in the PNNL study after pre-grouping them into clusters based on voltage class and climatic regions [26].

- ***Non-hierarchical Clustering***

These methods, unlike hierarchical methods, take the optimal number of clusters as an input and start forming all the clusters at the same time. Based on the optimizing criterion the clusters are updated in each iteration till the convergence criteria is met or no further improvement is possible. Two of the most commonly used methods are:

- k-means: This method has been used by Broderick et al in their research [40]. The objective of this approach is to minimize the within cluster distances.
 - It starts by selecting N_c (optimal number of clusters) different points called ‘means’ and groups the feeders closest to these means forming N_c different clusters.
 - In the next step these random initial means are replaced by the centroid of the clusters formed in the previous step.
 - The feeders are reassigned to those centroids from which their Euclidean distances are minimal. This and the previous step are repeated to convergence.

It has also been suggested that k-means converges faster than hierarchical clustering algorithms, but can be sensitive to outliers [45].

- k-medoids: This method is very similar to the k-means algorithm with some modifications. Here instead of initializing the algorithms with random ‘means’, actual feeders are chosen to be the initial medoids. Secondly, rather than using the calculated centroid to be the updated medoid in the future iterations, the feeder with the smallest cost function is used instead.

This method has been used in the research by Cale et al and CSIRO and Ausgrid [28], [41]. However, the implementation in these two studies differs slightly in the criteria it uses for clustering the variables together.

The method used by CSIRO involved optimization using the Euclidean distances, whereas the research by Cale et. al. uses the more generalized Minkowski distance function. The usage of this function in k-medoids makes the method more robust to noise and outliers. The algorithm for k-medoids described here is the Partitioning Around Medoids (PAM). The is used in [41] and involves the following steps:

- Clustering starts by selecting N_c (optimal number of clusters) different feeders as the initial medoids.
- In the next step individual feeders are assigned to the cluster of the closest medoid. Closeness is determined by the cityblock distance, which is obtained by substituting $p=1$ in the Minkowski distance function as shown. It can also be seen that substituting $p=2$ will give the Euclidean distance.

$$\left(\sum_{i=1}^N |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (17)$$

- The update of the medoids is done by evaluating the cost function, which basically calculates and stores the sum of Manhattan distances of the medoid and all the members of its cluster, while switching the medoid by each of the non-medoid members. The feeder which leads to the minimum cost function is assigned to be the new medoid. This and the previous step are repeated to convergence.

Thus, k-medoids seems to be the best method for clustering feeders amongst those presented here. It will have the convergence benefits of k-means, and though it is computationally more intensive than k-means, it is more robust to noise and outliers [46]. Further it can eliminate the final step of choosing representative feeders, as the post convergence medoid will be an actual

feeder. This is in contrast to k-means where the post convergence center is the calculated centroid, and we still have to choose the feeder which is closest to it.

Therefore, k-medoids using the partitioning around medoids (PAM) algorithm was used for clustering the feeders [47]. Figure 13 shows the 9 clusters with their respective medoids on the principal component axes. The clusters are not very distinct because of the lesser variance represented by the first two principal components. However, this is the only method of graphically representing the clusters in two dimensions while still including the contribution from all the parameters. So, the feeders are now grouped into clusters based on their similarities.

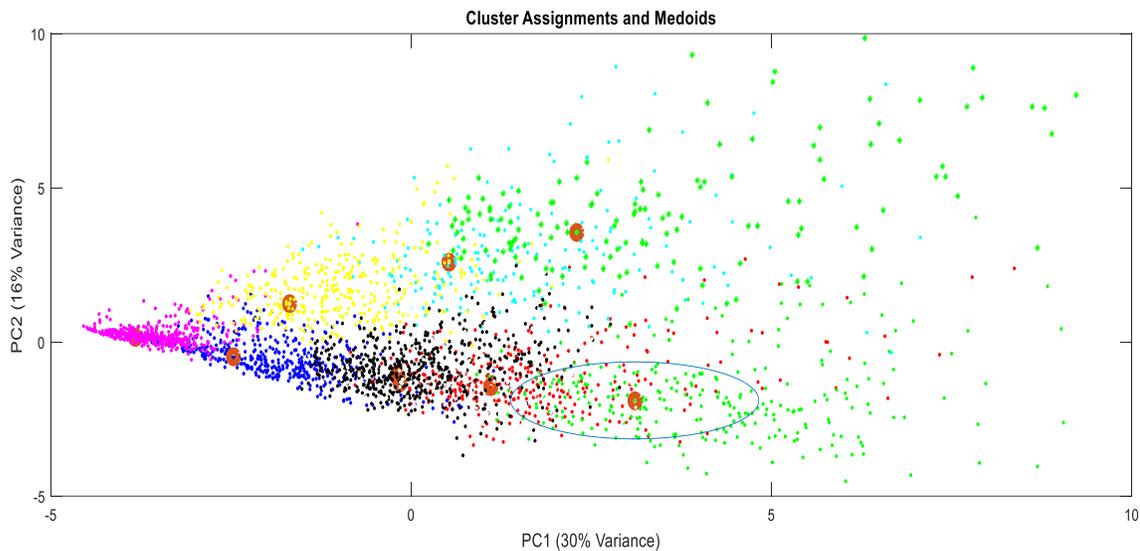


Figure 13: Final clusters of feeders with respective medoids

Now choosing some feeders from these clusters and testing their PV hosting capacities can help in the determination of the characteristics responsible for the differences in PV hosting capacities of feeders. Also, the results should be applicable to all the other feeders in their respective clusters because they have similar characteristics. The validation of clusters, determination of PV hosting capacity of a feeder and the determination of characteristics responsible for the variation in PV hosting capacities, is discussed in the next chapter.

5. PV Hosting Capacity Analysis and Cluster Validation

A significant percentage of PV interconnections are being done at the residential level in distribution systems. A lot of work has been done in the area of mitigating the adverse impact such interconnections may have on the grid. A number of standards and rules have been developed worldwide to aid in the process of determining whether an interconnection requires network upgrades or not [17], [48]–[50]. For example, Rule 21 by the California Public Utility Commission (CPUC) provides initial review screens and some shorthand equations to get a conservative estimate of a feeder’s PV hosting capacity [50]. PV hosting capacity is defined as the maximum amount of PV penetration, as a percentage of the peak load of a feeder, which will not cause any violations. These equations and review screens, though less time consuming, are only estimates. The true PV hosting capacity of a feeder can only be determined through a PV hosting capacity study on a detailed feeder model. The PV hosting capacity application on the DEW platform with a modified violation checker was used for the studies presented here [20], [38].

5.1. PV Hosting Capacity Analysis

For this research the objective was to determine the minimum hosting capacity of the feeders. Overvoltage is usually the first violation experienced as the PV penetration increases. So to determine the absolute minimum PV hosting capacity, the worst case time-point needs to be considered. The maximum chances of overvoltage will be when the ratio of the PV on a feeder to its load is maximum. However, determination of this time point is not easy as the power generated by the PV modules will depend on the insolation, and load conditions can only be known through time varying load profiles obtained from AMI and smart meters. The DEW platform helps in identifying this time point using the integrated system modeling approach, where the solar insolation measurements and the load profiles can both be compared [51]. Also, detailed load models are available which can help in accurately simulating the impact of the addition of solar PV generators on the distribution system.

The PV Hosting capacity analysis application uses a stochastic PV allocation approach to represent a wide variety of interconnection scenarios [52]. New PV sites are added to a study circuit at randomly selected locations in order to meet increasing PV penetration levels, and the

circuit is evaluated for any adverse effects. The effects of the added PV are tabulated, summarized and written out to a database for further analysis using a set of analysis functions within DEW.

The algorithm begins by running a base case analysis on the circuit to be analyzed. In this base case, the circuit can be started from an as-built state (with all existing PV, control devices, etc.), or it can be started from a case where all PV are removed from the system. The algorithm then checks and records violations on the circuit.

After the base case analysis is performed, a random allocation of PV is applied to the circuit. This allocation is based on three user-defined criteria: the size of the PV sites to be added, the maximum PV penetration level to test, and the penetration step size, defining the penetration levels to test. The PV connected to the circuit in the base case is included in the penetration calculation. The algorithm then makes a list of the customers on the circuit, randomizes that list and adds PV (of the size specified by the user) until the penetration level under test is reached. If there are not enough customers on the circuit to achieve the specified penetration step size, the algorithm increases the size of the individual PV sites until the desired step size is reached. The flow chart for this hosting capacity analysis algorithm is given figure 14.

After the new PV sites have been allocated, the application analyzes the circuit and records all violations resulting from the additional PV. The hosting capacity analysis procedure tests each random placement of PV for violations. These tests include steady-state over voltages and under voltages, transient over voltages and under voltages, flicker susceptibility, reverse flows at protective devices and voltage regulators, phase imbalances and overloads. A description of the violations checks is included below in Table 4. The circuit is then returned to the base state, and another random allocation of PV is applied using the methodology described above. The user may specify how many times this process is repeated at each penetration step. Once the number of iterations the user specifies has been reached, the application proceeds to the next penetration level and the process of adding a series of random PV allocations is repeated. The application continues in this fashion until the maximum penetration level as specified by the user is reached.

Table 4: List of Violation Criteria analyzed in detailed PV Hosting Capacity Analysis

Violation Variable	Comparison	Threshold	Units
Customer Level Overvoltage	>	126	Volts
Customer Level Undervoltage	<	114	Volts
Line Transformer Overvoltage	>	126	Volts
Line Transformer Undervoltage	<	114	Volts
Line Transformer Temporary Overvoltage	>	126	Volts
Line Transformer Temporary Undervoltage	<	114	Volts
Generator POI Overvoltage (Steady-State)	>	126	Volts
Generator POI Undervoltage (Steady-State)	<	114	Volts
Generator POI Temporary Overvoltage (During PV Output Change)	>	126	Volts
Generator POI Temporary Undervoltage (During PV Output Change)	<	114	Volts
Generator POI Flicker Sensitivity (PV Step Up)	>	2	Volts
Generator POI Flicker Sensitivity (PV Step Down)	>	2	Volts
Voltage Change at Voltage Controller	>	½ BW	Volts
Voltage Regulator Reverse Flow	<	-0.1	kW
Protective Device Reverse Flow	<	-0.1	kW
Feeder Reverse Flow	<	-0.1	kW
Feeder Current Imbalance	>	20	%
Component Voltage Imbalance	>	3	%
Component Overload	>	100	%

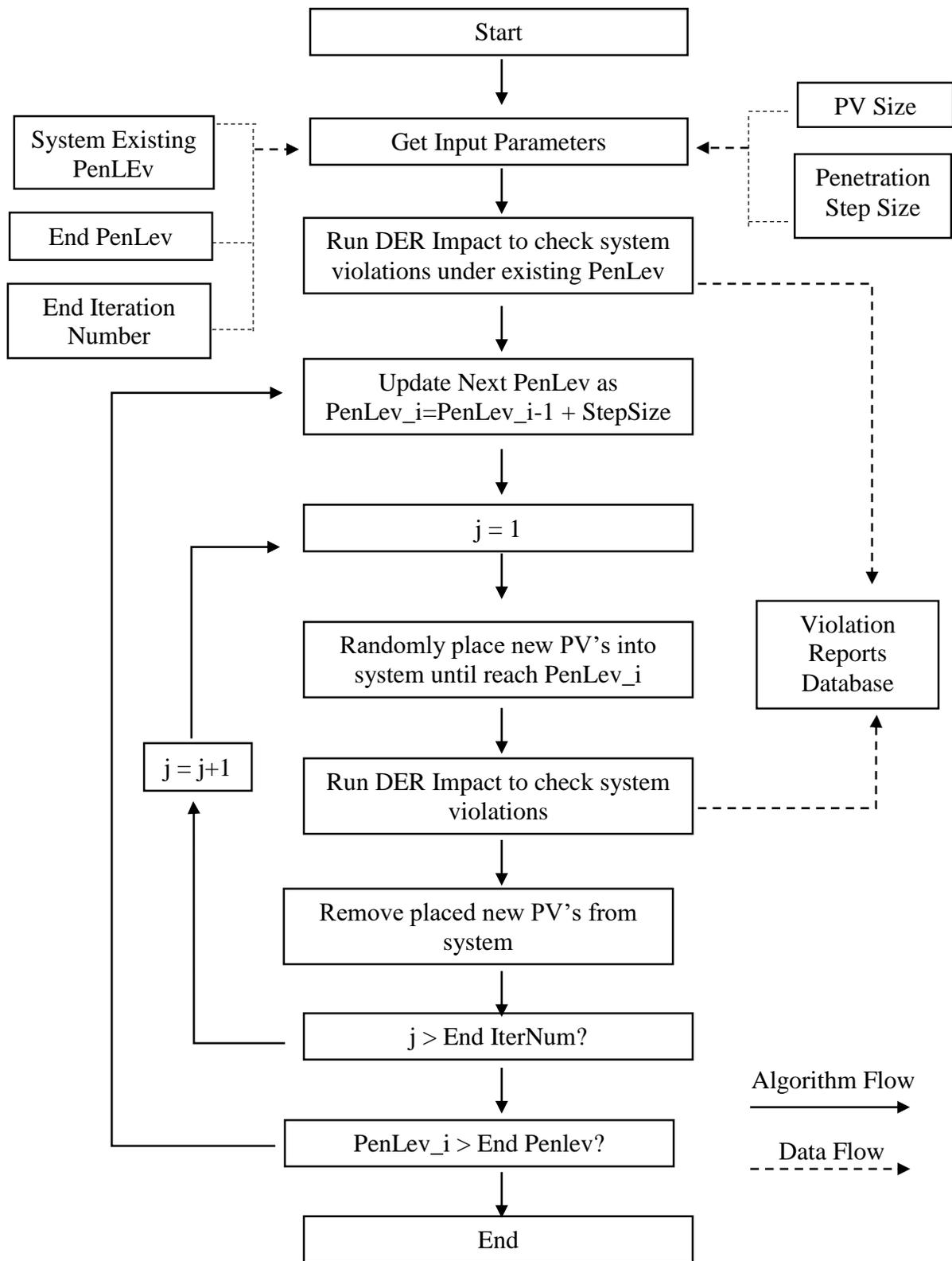


Figure 14: PV Hosting Capacity Analysis Flow Chart

5.2 Minimum and Maximum PV Hosting Capacity

The objective of the Hosting Capacity application is to define the limits at which a feeder or section of a circuit would be closed to further PV installations. Two penetration limits are defined in the results, the minimum penetration limit and the maximum penetration limit. The minimum penetration limit is the PV penetration at which a violation variable first crosses its failure threshold in the hosting capacity study. The maximum penetration limit is the point past which every tested random placement of PV results in threshold violations. These two penetration limits are depicted graphically in figure 15. This figure shows the maximum steady state voltage observed on a component in each deployment scenario. The minimum penetration limit is shown by the green line, and the maximum penetration limit by the red line. The PV penetration levels in between have some deployment scenarios, where there will not be violations if PV generators are placed at some specific locations, while other locations will lead to one or more violations.

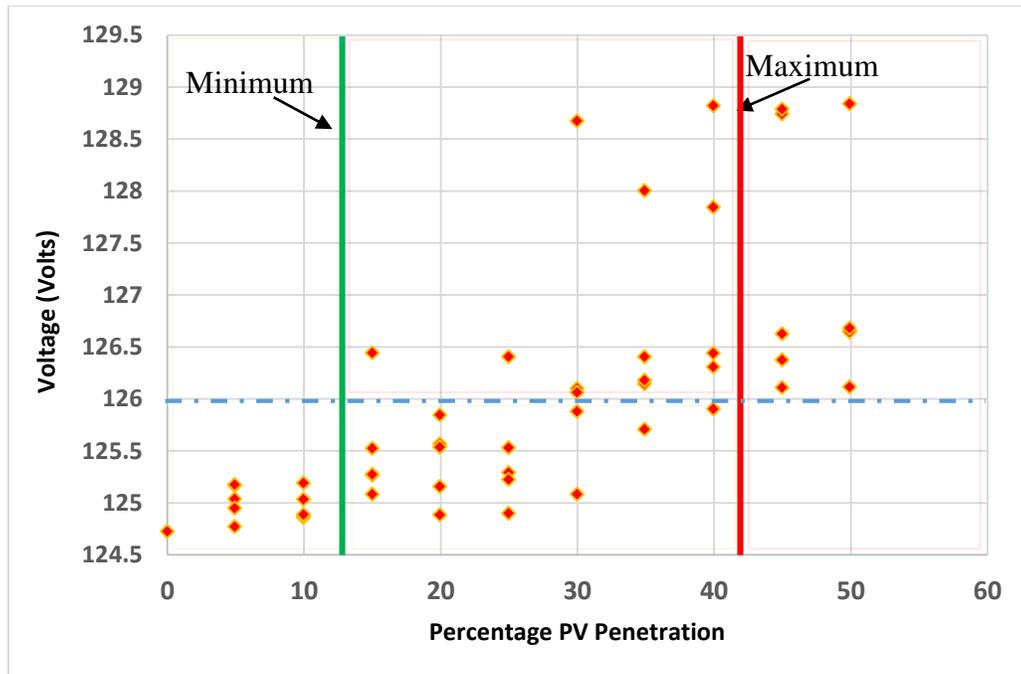


Figure 15: Minimum and Maximum PV Hosting Capacity

Each point on the plot of figure 15 represents one tested random placement of PV sites satisfying the PV penetration value on the horizontal axis. The location on the vertical axis gives the highest observed value of the violation variable for that random placement of PV. If the point

falls above the violation threshold, that placement of PV results in unacceptable conditions on the feeder. The minimum penetration limit occurs at the first PV penetration level which the violation threshold is exceeded. The maximum penetration limit occurs at the penetration level past which all tested random placements exceed the violation threshold. While the PV penetration is below the minimum penetration limit, small/medium PV can be interconnected anywhere on the feeder and no issues would be expected. As the penetration exceeds the maximum penetration limit, any subsequent PV interconnection would result in a violation, the circuit is effectively closed to additional PV deployments. In between these two penetration limits it is possible to add PV to the circuit, but it depends on the size and location of the new PV site.

5.3 Determining Characteristics Responsible for differences in PV hosting capacity of feeders

Clustering of a dataset of 3000+ distribution feeders was discussed in chapter 4. The feeders were clustered into 9 different clusters based on their characteristic similarities. The characteristics included in the dataset were chosen by experts from utilities and research organizations because they were considered to be relevant for the determination of PV hosting capacity of feeders. However, a better idea about the most important of these characteristics can be obtained by performing a detailed PV hosting capacity analysis on some of the feeders from these clusters. So, 7 feeders were chosen at random from these clusters and their PV hosting capacity analysis results were compared. Similar hosting capacities of feeders in a cluster and different hosting capacities among clusters will signify effective clustering, and also show that the parameters included in the dataset are indeed relevant for determining the PV hosting capacities of feeders. Then all 33 characteristics can be compared to determine the most relevant ones.

The dataset used in this research had previously been used in a similar clustering study [45], [53]–[55]. This previous study had conducted detailed PV hosting capacity analysis on the feeders it had determined to be the representative feeders. A comparison of these results provided a good validation opportunity for the clustering methodology followed here.

The representative feeders from the previous study are used here, so the selection is completely random and unbiased for this research. However, to be absolutely sure, the feeders were ranked based on their Euclidean distance from the cluster medoid. The parameters were given a ‘Nearness Rank’, as given in Table 5, which shows the number of cluster members closer to the medoid compared with the sampled feeder. The selected feeders can be seen to be fairly distant from the medoid and also from one other. Thus, the results obtained on these feeders can be assumed to be applicable for the other feeders as well.

Table 5: Distance of chosen feeders from cluster medoid, showing unbiased selection

		Feeder ID	Distance from medoid	Nearness rank		Average distance from medoid	Minimum distance from medoid	Maximum distance from medoid
Cluster 2	Modeled Feeders	1354	1.73	178/449	In general for all cluster members	2.32	0.34	12.76
Cluster 3		142	4.03	127/283		4.85	1.53	14.36
		281	3.21	66/283				
Cluster 5		888	2.40	423/456		1.22	0.23	6.82
Cluster 6		2885	1.84	55/341		3.23	0.90	16.22
Cluster 9		1140	5.35	72/142		5.76	2.33	13.56
		2093	2.96	12/142				

After selecting the feeders, their PV hosting capacities were compared. The residential/commercial scale and utility scale PV hosting capacities were compared. The former involves deployment of smaller PV units at residential/commercial customers, where the PV size is restricted to the maximum load of the customer. The utility scale PV consists of larger PV units of 500kW or more, to simulate the effect of addition of larger PV plants in the distribution feeders.

Figures 16 and 17 show the detailed PV hosting capacity results of these representative feeders [45]. The 6 bars are the hosting capacities for 6 feeder issues, namely primary node over-voltages (V_1) and voltage deviations (V_2), voltage regulation node voltage deviation (V_3), element fault current (P_1), sympathetic breaker tripping (P_2), and breaker reduction of reach (P_3). The areas in purple represent the PV penetration levels which do not cause violations at any location, yellow represents penetration levels which may lead to violations when PV generators are deployed at specific locations and orange represents penetration levels which will lead to violations

irrespective of the location [45]. The dotted line represents 15% of the peak load for the particular feeder.

It was found that two of the clusters (3 and 9) contained 2 each of these modeled feeders. It can be clearly seen that that the PV hosting capacities of the feeders in the same cluster are similar for both residential/commercial scale PV (figure 16) and utility scale PV (figure 17), for all the 6 parameters. This validates the effectiveness of the clustering. Another validation of the overall capability of the clustering methodology followed here to achieve better clustering is the grouping of feeders 1140 and 2093 (having similar parameters and hosting capacities) in the same cluster, as against different clusters in the previous study [45], which used the same data set. Having validated the clustering methodology by comparing the PV hosting capacity results, the characteristics of these feeders were compared to determine the ones most relevant for determination of a feeder’s PV hosting capacity. Table 6 shows these most relevant characteristics.

Table 6: Parameters Critical for PV Hosting Capacity

Cluster	Feeder ID	Primary Voltage	Total 3-Phase miles	Voltage Control Devices
3	142	21	26.89	0 regs ; 7 caps
	281	21	25.34	1 reg ; 7 caps
9	1140	12	70.55	9 regs ; 7 caps
	2093	12	128.79	9 regs ; 7 caps
2	1354	12	5.77	0 regs ; 2 caps
5	888	4	5.5	0 regs ; 0 caps
6	2885	12	50.41	4 regs ; 6 caps

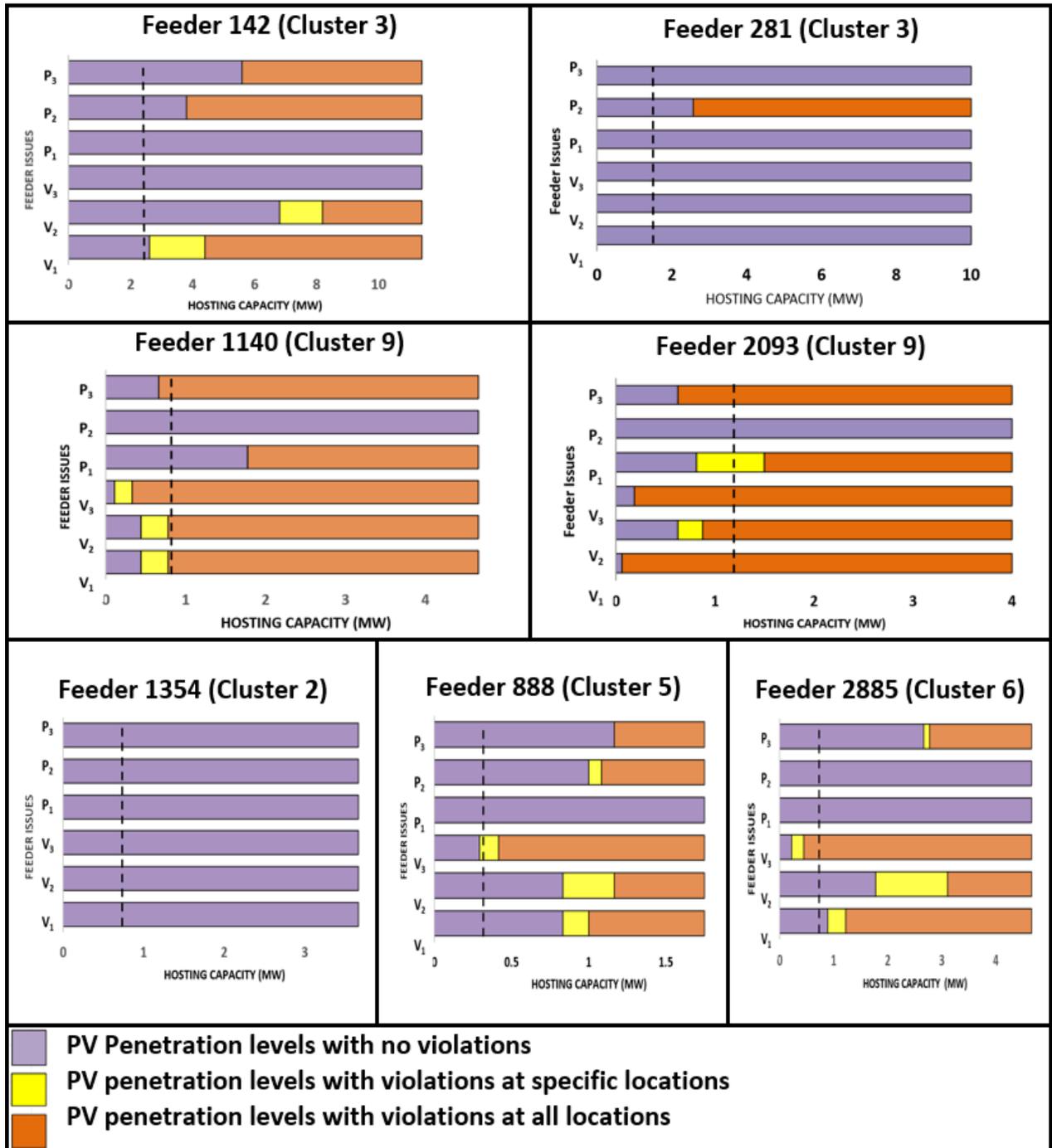


Figure 16: Residential/Commercial-scale PV hosting capacity comparison

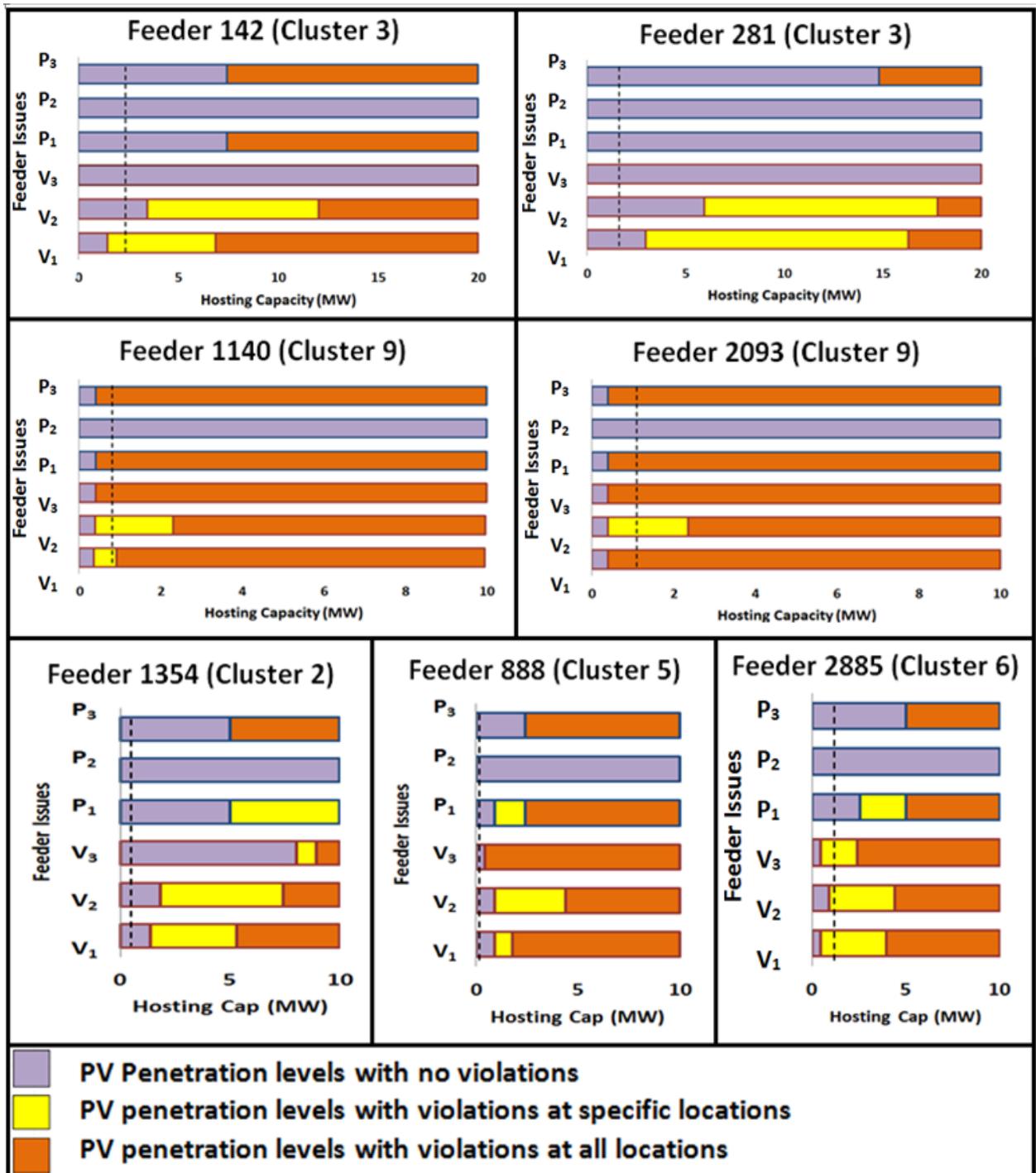


Figure 17: Utility-scale PV hosting capacity comparison

The characteristics shown in table 6 were the ones which differed the most among clusters and were similar amongst the members of a cluster. Similar results have been obtained in other studies as well. For instance, use of a large number of voltage control devices on a feeder shows the presence of existing voltage issues, and thus the feeder may exhibit lower hosting capacity. A similar result was obtained in [55], resulting in the addition of a new review screen in Rule 21 which states that any new proposed interconnection application will have to go through the supplemental review process if the feeder has an existing voltage regulator.

So, based on clustering and PV hosting capacity analysis on some feeders, the characteristics most essential for determining a feeder's PV hosting capacity were determined. A comparison of the hosting capacities of the feeders shown in figures 16 and 17 and their corresponding characteristics in table 6, shows that the feeders with a higher hosting capacity tend to be shorter in line length, have lesser number of voltage control devices and have a higher operational voltage level. Clusters 2, 3 and 5 exhibit these characteristics. On the other hand the feeders of clusters 6 and 9, having a lower hosting capacity, are longer in line length, have higher numbers of voltage control devices, and operate at a relatively lower voltage level. Thus, by selecting feeders which differ in these characteristics, feeders can be chosen from any dataset which should differ in their hosting capacities and an algorithm, such as the VPD algorithm, could be implemented on just those feeders for validation. If the algorithm can improve the hosting capacity on all of these feeders, then it can be said to be effective, or validated, and the results will be statistically relevant as well. Finally, these critical parameters determined from a few randomly chosen feeders were compared to the average properties of all the feeders in a cluster, and also to the properties of the cluster medoid, to ensure that the results are applicable to all the members.

5.4 Comparison of cluster average properties and medoids

One of the important steps for clustering was removal of dependent or correlated variables. So, figure 18 shows a comparison of the clusters formed when all the parameters, including the dependent ones, were used for clustering, and when only the independent parameters were used. The medoid values for all the parameters were compared to the average value of all the feeders within the cluster for the corresponding parameters. Figure 18 shows the former as a blue asterisk and the latter as a red circle for these three critical parameters. The first column of the figure shows

clustering results when all parameters have been used. Some of the medoids are farther away from the cluster average values. On the other hand, the second column, showing results for independent parameters, shows that the medoids are much closer to the means. A similar trend was seen in other parameters as well. Also, the cluster means, medoid values, and the values of these characteristics for the modeled feeders match very closely. Thus, the results obtained from the analysis on the randomly chosen feeders should also be applicable to all the other feeders in the respective clusters.

The results from the clustering analysis may be summarized as follows:

- The most relevant techniques used worldwide were used to accomplish clustering steps.
- Similar results from different techniques gave more credibility to the results.
- Cluster representatives had similar properties as the average properties of the cluster members. Also, feeders which were incorrectly grouped in separate clusters in a previous study were now grouped in the same cluster, further validating the methodology followed here.
- Removal of correlated variables led to better clustering, as all the feeder representatives are closer to the average properties of the cluster members.
- PV hosting capacity analysis of a few randomly chosen feeders showed that feeders in a cluster had similar hosting capacity, whereas the hosting capacity among clusters was different.
- Feeder length, number of existing voltage control devices and operational voltage level were found to be the most important parameters for determining the PV hosting of feeders amongst all the parameters in the dataset.
- These parameters may be used for choosing feeders to evaluate an algorithm for its effectiveness in enhancing the PV hosting capacity.

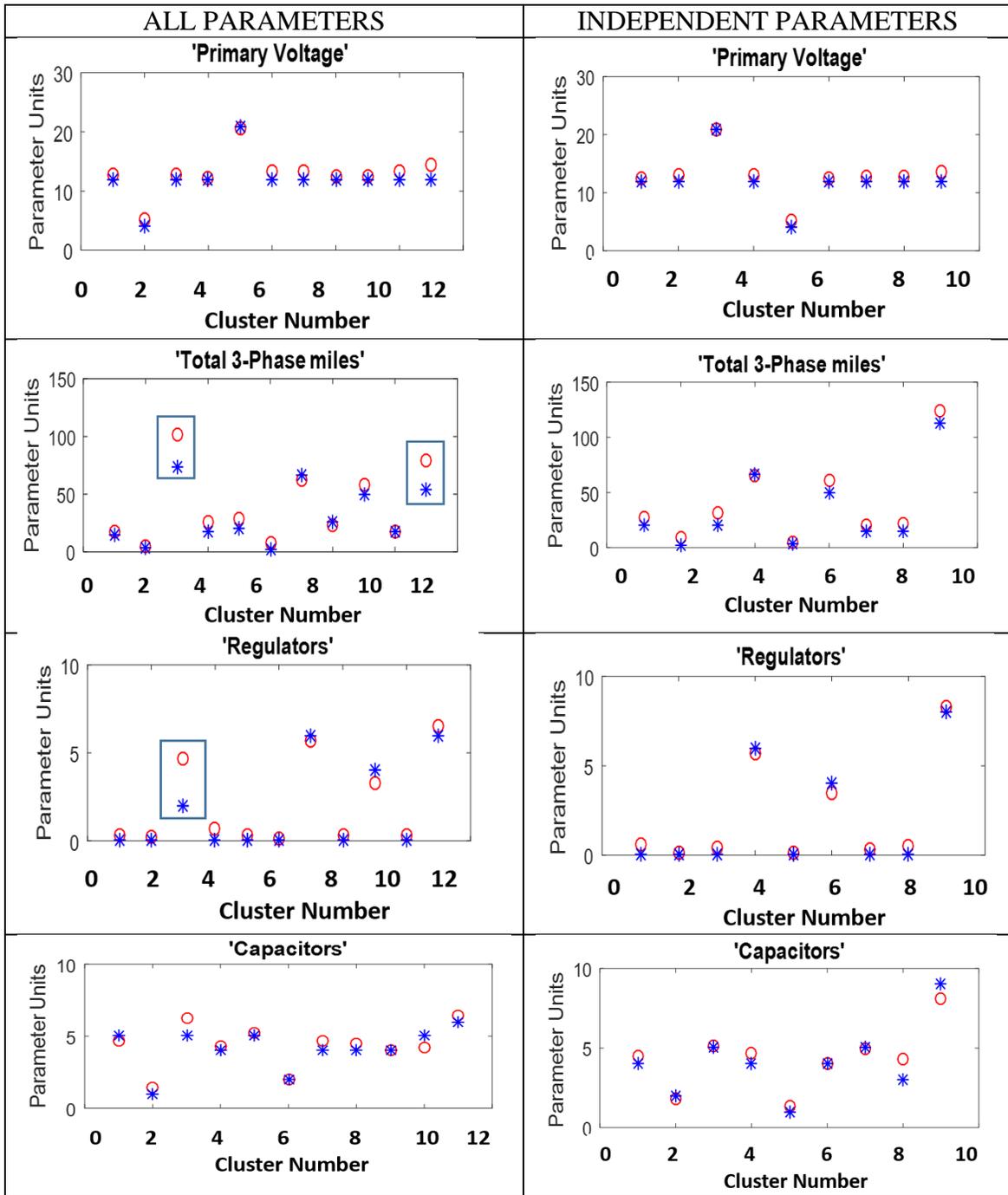


Figure 18: Comparison of medoid parameter values with mean values of all feeders in the respective cluster. Dependent variables have been removed while clustering in the second column (9 optimal clusters), whereas all the parameters are included in the first column (11 optimal clusters).

6. Validation of VPD Algorithm

In the previous chapter clustering was used to come up with the three most important characteristics for determining the PV hosting capacity of feeders. It was found that feeders with a shorter line length, lesser number of existing voltage control devices, and higher operating voltage tend to have a higher PV hosting capacity. The only objective for completing the clustering study followed by its validation using the PV hosting capacity analysis was to determine a statistically relevant set of feeders on which to validate the VPD algorithm. So in this chapter, the selection of a set of feeders and the results obtained after implementing the VPD algorithm are discussed.

6.1. Selection of feeders

On completion of the clustering analysis the 3000+ distribution feeders were grouped into 9 clusters and a representative feeder was determined for each of those clusters. Ideally the VPD algorithm should have been validated on these representative feeders which would exhibit the characteristic differences among the different feeder types. However, these feeders were not used because of the following reasons,

- Detailed feeder models, which should have included the feeder topology including the GPS coordinates and location data, and conductor information, both overhead lines and underground cables, was not available. Absence of this data made it impossible to accurately model the feeders, or conduct an effective PV hosting capacity analysis using solar measurements.
- Time varying load profiles were not available, which are essential for getting an effective design using the VPD algorithm. Also, the accuracy of a modeled feeder is determined by comparing the power flow results with the measurements.
- Detailed load models with time varying load profiles were available for distribution feeders from the New York region. Clustering was performed on the dataset of feeders from a single utility from the California region. Thus, choosing feeder models from New York region based on the clustering results provided an opportunity to further validate

the clustering results. Also, the VPD algorithm can now be validated on more accurate feeder models.

So, several feeder models were chosen from the New York region which varied in the characteristics determined from the clustering analysis. The selected feeders are shown in table 7. Based on the clustering results the chosen feeders should exhibit PV hosting capacities as shown in the last column of table 7.

- 5012: This feeder has a high number of voltage control devices, is the longest in line length amongst all the feeders, and has a significant section operating at a relatively lower voltage level. Thus, it can be expected to have a lower hosting capacity.
- 4021: This feeder also has a large number of voltage control devices and has a long line length. Thus, its hosting capacity should also be lower.
- 3021: This feeder has a large number of capacitors, but does not have any voltage regulators. It is also a long feeder and has a significant section operating at a lower 4.16 kV level. Thus, it is also expected to have a lower PV hosting capacity.
- 5024: This feeder has a lesser number of voltage devices, but is longer in line length, so it should have an intermediate hosting capacity.

Table 7: Selected feeders and their characteristics

Feeder ID	Fixed Shunt Caps	Switched Shunt Caps	Voltage Regulators	Total Miles	13.2 kV Miles	4.16 kV Miles	Expected Minimum PV Hosting Capacity
5012	5	1	1	47	42	5	Low
4021	5	0	2	39	39	0.06	Low
3021	6	1	0	46	35	10	Low
5024	3	0	0	37	37	0	Medium
5022	3	0	1	29	29	0.02	Medium-High
5021	2	0	0	17	15	2	High

- 5022: This feeder also has a lesser number of voltage control devices, including a single phase regulator, shorter in length and does not have any significant 4.16 kV section. Thus it can also be expected to have a medium to high PV hosting capacity.
- 5021: This feeder just has a couple of fixed shunt capacitor banks, is much shorter than the other feeders, and has a very small lower voltage section. Thus, it can be expected to have a higher PV hosting capacity.

So the selected feeders show a proper mix of these characteristics determined to be important for PV hosting capacity. Feeders 5012 and 5021 represent the worst and best case scenarios, respectively. The remaining feeders have a good combination of these characteristics and consequently are expected to have different hosting capacities. To validate these expected hosting capacities, a detailed PV hosting capacity analysis, as discussed in chapter 5, was carried out.

6.2. Minimum PV hosting capacity of selected feeders

As mentioned in chapter 5, in the discussion on PV hosting capacity analysis, the focus here is on determining the minimum PV hosting capacity of a feeder. The analysis time point chosen is when the PV to load ratio is maximum, and thus the chances of over-voltage violations are the most. Thus, the hosting capacities determined here are for the worst case scenario. This particular time point was of interest because if the VPD algorithm can increase the PV hosting capacity even in this worst case scenario, then it can be expected to work for all other time points as well.

A stochastic PV allocation approach is used here, covering a wide variety of interconnection scenarios. New PV sites are added to the study circuit at randomly selected locations to increase the PV penetration level. For each scenario the circuit is evaluated for adverse effects. Here the PV penetration level is increased in 5% step sizes, up to a penetration level of 50%, as the interest is only in determining the minimum hosting capacity. For each penetration level 5 random PV deployment scenarios are analyzed. The individual PV size used is 5kW. The time of analysis is when the PV to load ratio is maximum. It was found that overvoltage was the first violation encountered in almost all of the feeders. The same can also be seen in the PV hosting capacity results of figures 16 and 17 [45]. However as will be seen in the detailed discussion on individual feeder results, it is necessary to consider all the parameters.

The results in this and the next subsections are obtained by operating the PV inverters at unity power factor without real power curtailment, a commonly used setting for PV hosting capacity studies [14]. Figure 19 shows the maximum voltage of any component for all 5 deployment scenarios at a penetration level for all the 6 selected feeders. The minimum hosting capacity is determined assuming a linear transition between two consequent PV penetration levels and when the voltage crosses 126 volts [1]. Table 8 shows a comparison of the expected and actual hosting capacities of the 6 feeders.

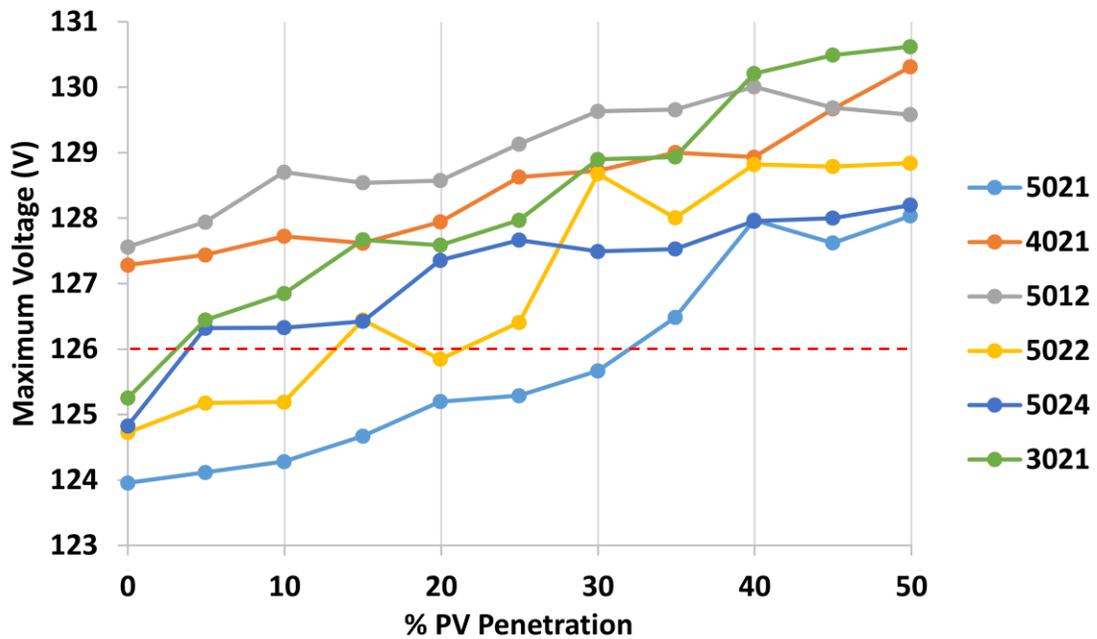


Figure 19: Minimum PV Hosting capacity of selected feeders

Table 8: Actual Minimum PV Hosting of Distribution Feeders

Feeder ID	Fixed Shunt Caps	Switched Shunt Caps	Voltage Regulators	Total Miles	13.2 kV Miles	4.16 kV Miles	Expected Minimum PV Hosting Capacity	Actual Minimum PV Hosting Capacity
5012	5	1	1	47	42	5	Low	0%*
4021	5	0	2	39	39	0.06	Low	0%*
3021	6	1	0	46	35	10	Low	4%
5024	3	0	0	37	37	0	Medium	5%
5022	3	0	1	29	29	0.02	Medium-High	14%
5021	2	0	0	17	15	2	High	33%

*0% implies overvoltage violations were in the base case itself without any additional PV units being placed

It can be clearly seen from the results shown in table 8 and figure 19 that the selected feeders exhibit a very similar trend in their PV hosting capacity as was expected. This not only provides further validation to the clustering results, but also provides a statistically relevant set of feeders on which to validate the VPD algorithm.

6.3. Implementation of VPD algorithm on selected feeders

The VPD algorithm has been implemented in the DEW platform, and is available as an interactive application. To implement the algorithm on any feeder model, the MV 90 measurements have to be loaded, using the measurement database, which has been set up as an ODBC database. This is followed by running a feeder evaluation. Feeder evaluation simply calls on another built in application ‘Feeder Performance’, which runs a time series analysis of user defined time step and duration. The time step chosen for this analysis was 1 hour and for the entire year, thus making it an annual evaluation. This application was used for getting the time points of maximum and minimum load conditions in terms of the real and reactive power flows in the circuit. This analysis being an annual evaluation is the only time consuming step in the VPD algorithm. However, this is only done once and the results are saved in a text file. A new design can be completed in just a few seconds using the results saved in this file.

All the other user-defined inputs, as discussed in chapter 2, were input to maximize the PV hosting capacity while using as few new voltage control devices as possible and still maintaining the voltages within ANSI limits. For example, the PV hosting capacity analysis on the feeders considered thus far have shown that overvoltage is usually the first condition to be violated. So the substation load tap changer (LTC) settings were kept relatively lower, around 123 volts, unlike the traditionally higher values of around 126 Volts. Also 'Upper Limit' and 'Lower Limit' were kept closer to the ANSI thresholds so as to maximize the usage of the available range.

Regulator control zone voltage was kept between 3-4 volts to get a flatter voltage profile. The switched shunt capacitor controller was programmed to switch off the capacitor when the reactive power flow at its point of connection was more than -100 kVar/phase and turn on when it was greater than 200 kVar/phase. New control devices were only inserted at overhead lines whose end node degrees equal to 2 (i.e. only one connection at the end), excluding the poles that have more than two lines connected to them. All new voltage regulators were installed at least 2000 feet from the substation, except for the regulator added at the beginning of the feeder when the LTC control is absent in the substation. However, all the feeders analyzed here already had LTCs.

To reduce the cost of redesign, the maximum allowable voltage control devices and their settings were kept exactly the same as were already installed in the base feeder. Since the VPD algorithm turns OFF all the existing devices before starting the redesign, it can suggest better locations using the same or lesser number of devices. Then just by relocation of a few devices, a better voltage profile and cost savings can be achieved. However since the voltages at all the feeder components have to be maintained within the ANSI limits, there were a few cases where a greater number of devices had to be used than the base case.

Since the substation LTC voltage set point is being reduced, and because the VPD algorithm is designed to give flatter voltage profile, a cost benefits analysis was also done to determine the cost savings achieved. Savings were expected because the conditions are similar to Conservation Voltage Reduction (CVR). CVR is a voltage management approach promoting energy conservation and peak demand reduction. The reason is that many loads exhibit a voltage dependency and their power demand is proportional to the applied voltage. So a reduction in

voltage can reduce their power demand and consequently reduce the total load on a feeder. To evaluate these economic benefits of the new design achieved using the VPD algorithm, a time series analysis, using hourly load data, was performed on both the base case feeder model and the new design for a whole year. Using this annual evaluation, the changes in load demand, changes in the annual feeder losses, and also the changes in the peak loading were evaluated. The voltage dependency factor used for these feeders was provided by the utility to be -0.15.

$$\text{Voltage Dependency Factor} = \frac{\Delta V}{\Delta I} \quad (18)$$

This implies that if the voltage is reduced by 1% ($\Delta V = -1\%$), then the current will increase by 0.15%. This will lead to a slight increase in the feeder losses, which should most likely be compensated by the large reduction in load demand. CVR factors, which are simply the ratio of percentage change in energy to the percentage change in voltage, were also determined to compare the achieved savings with previous CVR studies [22]–[24]. These studies have reported CVR factors ranging from 0.5-1.

$$\text{CVR Factor} = \frac{\% \Delta E}{\% \Delta V} \quad (19)$$

The VPD algorithm was implemented on all the selected feeders, considering all the aforementioned points, while giving the input parameters. This was followed by a cost benefits analysis. The results obtained from the 6 feeders are discussed here. The feeders where some particularly interesting results were obtained are discussed first.

6.3.1. Feeder 5022

The base case of this feeder, i.e. the actual model of the feeder as it has been implemented is shown in figure 20. This feeder has around 900 load buses, more than 300 distribution transformers and a peak load of 2292 kW. The feeder has experienced previous voltage issues

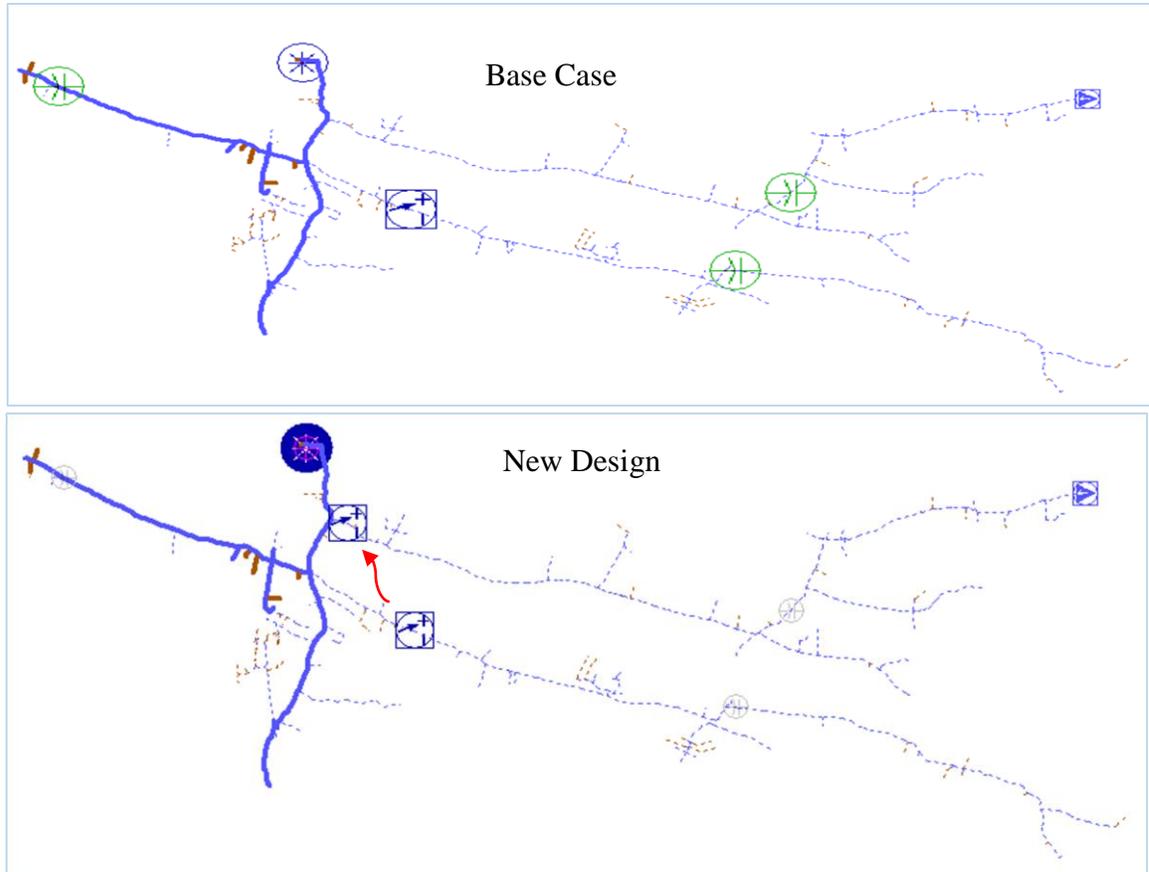


Figure 20: Feeder 5022; Base case (top), and New Design (bottom)

because voltage control devices have been used to maintain a desired voltage profile. The pre-existing voltage control devices (highlighted in figure 20) on this feeder are:

- 3 fixed shunt capacitor banks injecting 1100 kVar
- Single phase voltage regulator of 167 kVA

The voltage profile of the base case is shown in figure 21. The voltage profile is for the farthest point from the substation of the single phase lateral without the voltage regulator. The profile is drawn at the peak load condition to ensure no undervoltage violations because of the reduced substation LTC setpoint. There are no voltage violations in the base case.

The VPD algorithm was then used on the base case feeder. The maximum number of devices to be used were kept exactly the same as the base case, i.e. 3 fixed capacitor banks and a single phase voltage regulator. This new design is shown in figure 20. The original capacitor banks have been disabled (grayed out in the figure) but are still kept on the model for

comparison. The voltage regulator has been relocated to the other lateral, while the original regulator has been turned off, though still visible in the figure. The original regulator's impedance has been reduced to zero and its controller is not allowed to move, so effectively it does not exist in the circuit. This new design has only the single phase voltage regulator and none of the capacitor banks. The voltage profile of the new design is shown in figure 21. There are no voltage violations in the new design either. This was a particularly interesting result because even though 1100 kVars of the capacitor banks have been removed, there are still no voltage violations. The reason for this behavior can be seen in table 9. The power factors in the base case indicate that the excessive compensating kVars are leading to overcompensation. The VPD algorithm identified this problem and gave a design with better power factors.

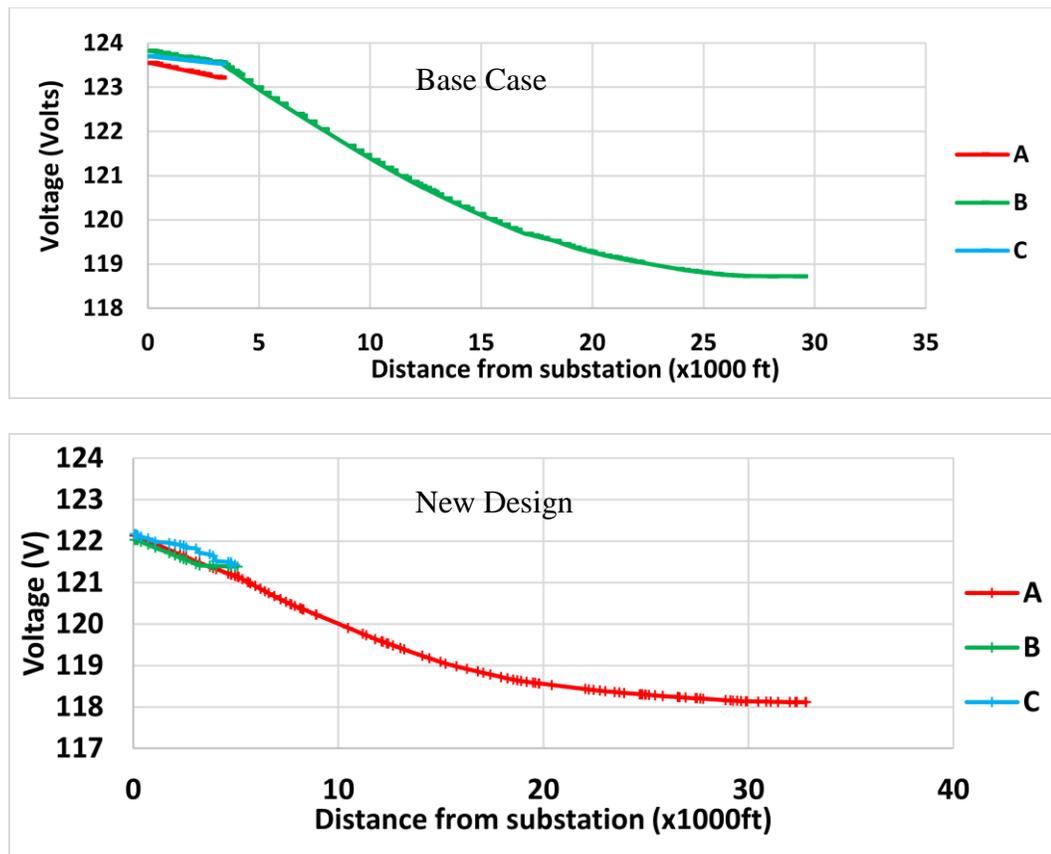


Figure 21: Voltage Profile of feeder 5022; base case (top), and New design (bottom)

Table 9: Comparison of operational parameters of Base case and New Design of feeder 5022

Design	Peak Load Time Point				Minimum Load Time Point			
	Max V	Min V	Avg V	P.F.	Max V	Min V	Avg V	P.F.
Base	124.7	118.7	122.1	-0.99	125.2	123.1	124.1	-0.63
VPD	122.7	117.3	119.7	0.91	122.7	120.5	121.8	0.92

The results from the cost benefits analysis are shown in table 10, where cost/unit is taken from [56]. This analysis consisted of a comparison between the base feeder and the new design for an entire year at an hourly time step (8760 time points). A significant peak reduction, load demand reduction and even loss reduction is observed. While the first two were expected, loss reduction was unexpected. A voltage dependency factor of -0.15 was used, which will increase the current as the voltage is reduced and consequently losses should have increased marginally. However there is a 23% reduction in losses. This can again be attributed to the overcompensation in the base case. Even though the losses increased due to the effect of the voltage dependency factor, the overall system current decreased due to the improved power factors. This led to an overall loss reduction.

Table 10: Cost Benefits Analysis of Base case and new design of Feeder 5022

	Annual MWh	Peak Load (kW)	Losses (MWh)
Base Feeder	12434	2292	237
New Design	12196	2255	182
% Change	-1.91%	-1.63%	-23%
Cost Savings due to load reduction			\$ 33,267*

* Cost per unit is taken to be 16 cents [56]

6.3.2. Feeder 5021

The base case of this feeder, i.e. the actual model of the feeder as it has been implemented, is shown in figure 22. This feeder has around 1790 load buses, more than 230 distribution transformers and a peak load of 5468 kW. The feeder has experienced previous voltage issues because voltage control devices have been used to maintain a desired voltage profile. The pre-existing voltage control devices (highlighted in figure 22) on this feeder are:

- 2 fixed shunt capacitor banks injecting 750 kVar

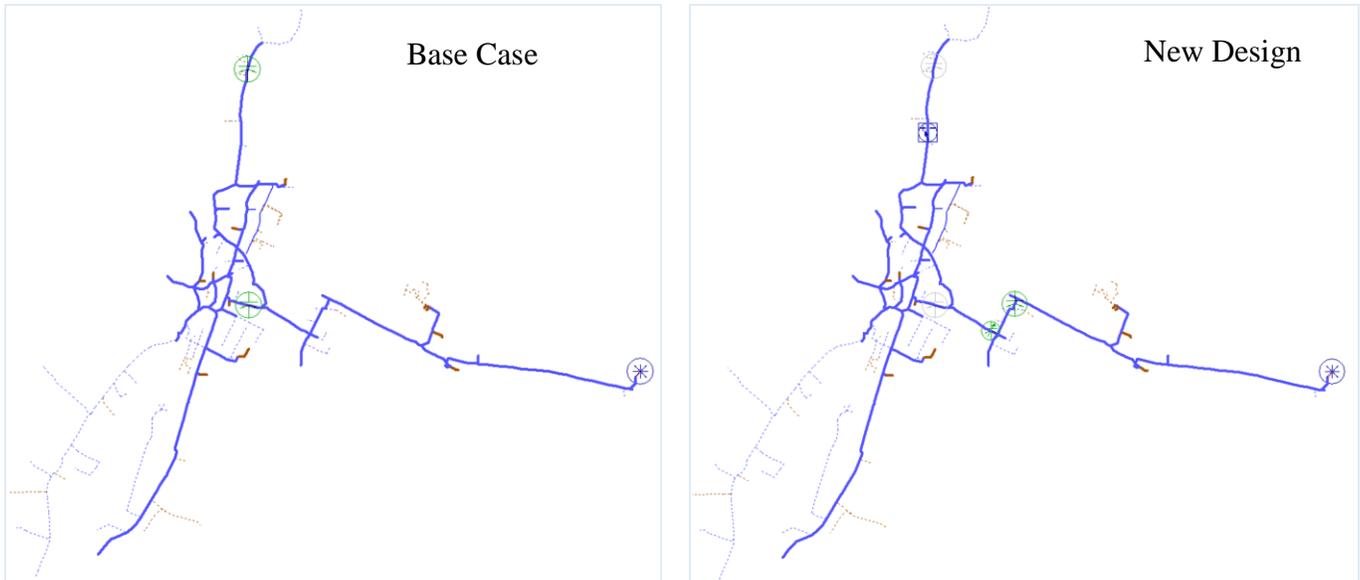


Figure 22: Feeder 5021; Base Case (left), and New Design (right)

Even though this feeder already had a higher PV hosting capacity, the VPD algorithm was implemented to determine if it can be further enhanced. The maximum number of devices allowed to be placed were the same as in the base case, i.e. 2 shunt capacitor banks. However, the new design showed minimum voltages of around 100 volts on a few components. So the maximum number of devices allowed were increased. The VPD algorithm placed a fixed shunt capacitor, a switched shunt capacitor and a voltage regulator to remove all the violations. This was again an interesting result because this feeder had a very high PV hosting capacity. On a closer examination it was found that this feeder was under-compensated and a few components had under voltage violations as shown in figure 23 and table 11. The voltage profile clearly shows that a few components at the extremities of the network experience voltages below the ANSI limits. The VPD algorithm identified these locations and placed a voltage regulator here to remove these violations, as can be seen in figure 22.

Further, a more thorough review of the results of the PV hosting capacity analysis showed that a few components in the base feeder were indeed experiencing very low voltages as shown in figure 24. This shows that it is essential to look at all the operational parameters and not just a few. So, just by looking at overvoltages, this feeder has a minimum hosting capacity of

Table 11: Comparison of operational parameters of Base case and New Design of feeder 5021

Design	Peak Load Time Point				Minimum Load Time Point			
	Max V	Min V	Avg V	P.F.	Max V	Min V	Avg V	P.F.
Base	123.5	113	118.6	0.92	128.23	121.3	123.2	0.97
VPD	124.1	116.6	119.5	0.96	124.2	122.7	123.6	1

around 33%, but effectively its hosting capacity is zero. On the other hand, the new design using the VPD algorithm does not experience any voltage violations, effectively increasing the hosting capacity from 0% to around 30%, as shown in figure 24.

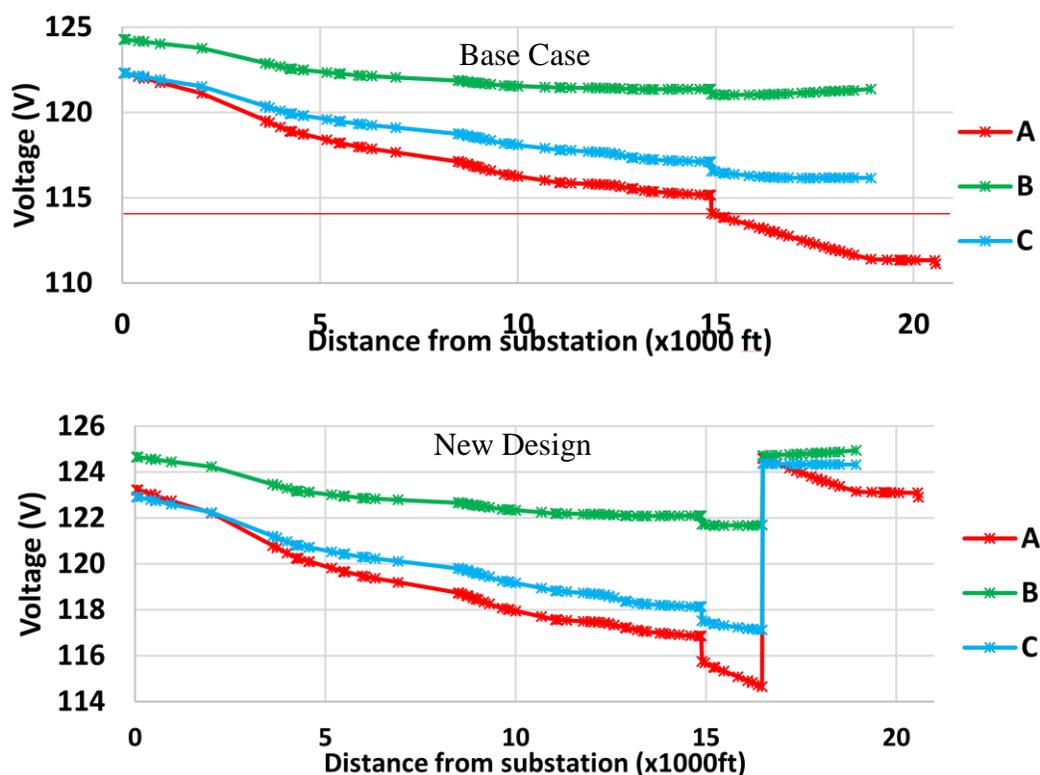


Figure 23: Voltage Profile of feeder 5021; Base case (top), and new design (bottom)

A cost benefits analysis was also done and the results are shown in table 12. As expected there is a marginal increase in the load demand, peak load and a reduction in losses due to the improved voltages across the feeder. The objective for the VPD algorithm is to

Table 12: Cost Benefits Analysis of base case and new design of feeder 5021

	Annual MWh	Peak Load (kW)	Losses (MWh)
Base Feeder	34990	5558	1084
New Design	35159	5587	856
% Change	0.48%	0.53%	-21%
Cost Savings due to load reduction			-\$ 23,657*

* Cost per unit is taken to be 16 cents [56]

provide a voltage profile which conforms to the ANSI standards. This may require addition of new devices as in this feeder. However, it led to an enhanced PV hosting capacity and much better voltage profiles.

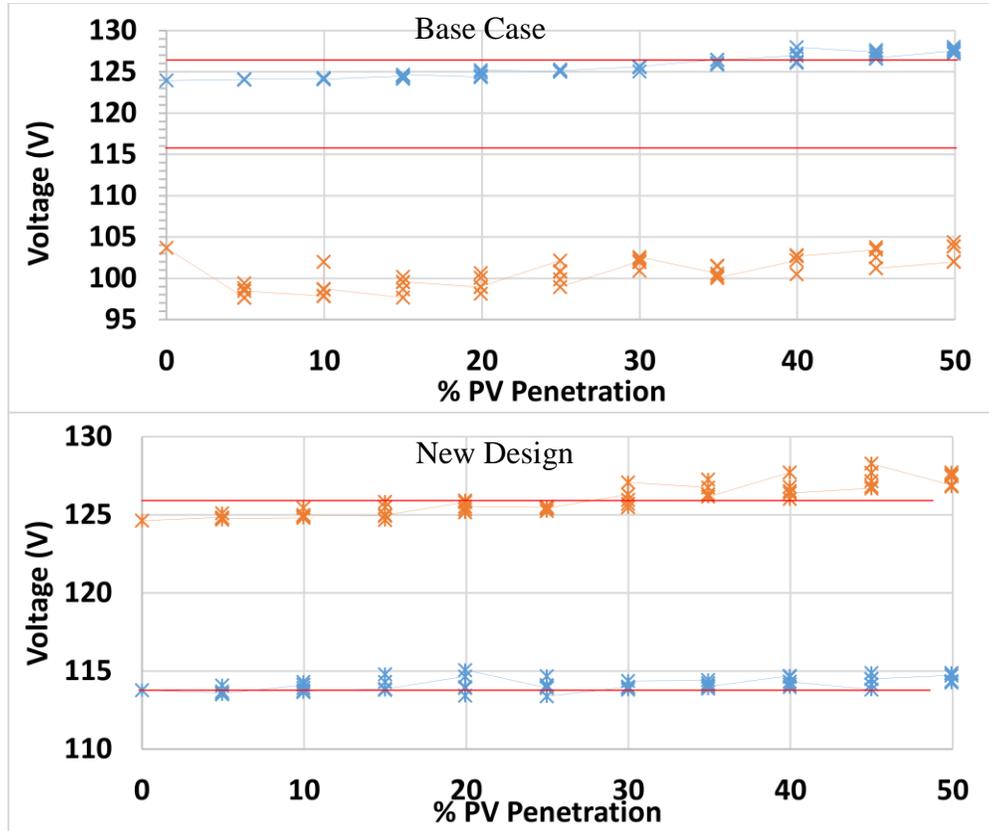


Figure 24: PV hosting capacity improvement in feeder 5021; base case (top), and new design (bottom)

6.3.3. Feeder 5012

The base case of this feeder, i.e. the actual model of the feeder as it has been implemented, is shown in figure 25. This feeder has around 2000 load buses, more than 500 distribution transformers and a peak load of 3601 kW. The feeder has experienced previous voltage issues because voltage control devices have been used to maintain a desired voltage profile. The pre-existing voltage control devices (highlighted in figure 25) on this feeder are:

- 5 fixed shunt capacitor banks injecting 2050 kVar
- 1 switched shunt capacitor bank injecting 600 kVar
- 1 3-phase voltage regulator rated 167kVA/phase

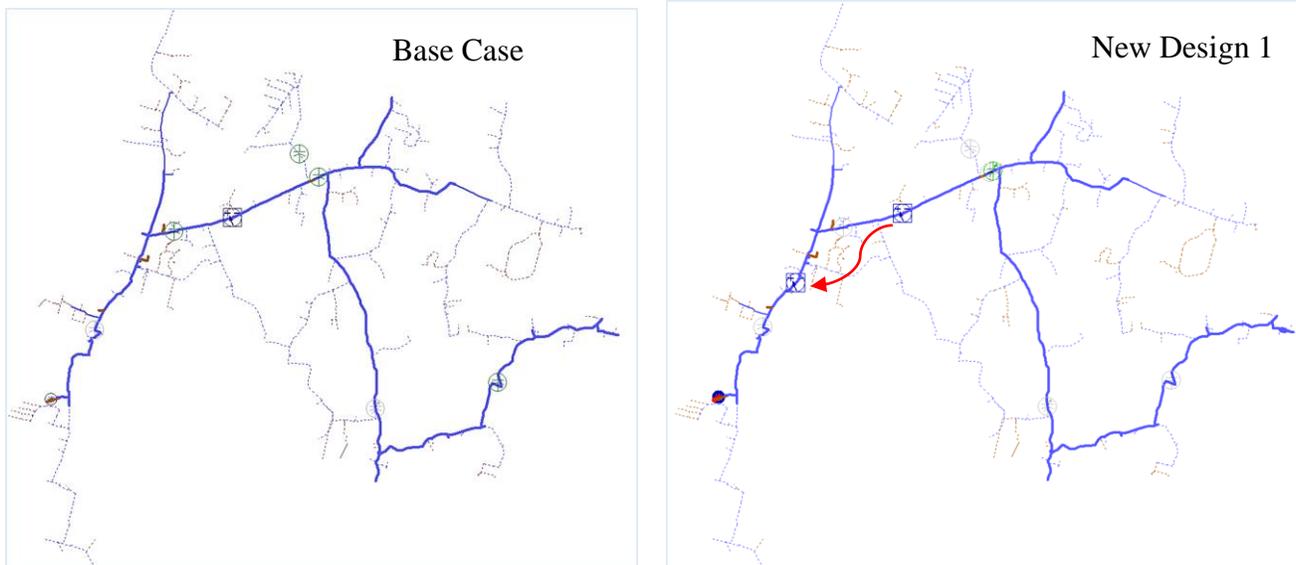


Figure 25: Feeder 5012; Base Case (left) and New Design 1 (right)

The voltage profile for this heavily compensated feeder during the peak loading conditions is shown in figure 26 and does not shown voltage violations. However, a comparison of the operational parameters in table 13 shows overvoltages. The VPD algorithm was implemented to see if a better design may be achieved. The input parameters allowed placement of 5 fixed shunt capacitors, 1 switched shunt capacitor and a voltage regulator. The algorithm repositioned the switched shunt capacitor and relocated the voltage regulator a bit closer to the substation. The repositioning of the voltage regulator was done based on the criterion of maintaining a 3-4 volt zone from the lowest voltage point to the regulator location. However, none of the fixed shunt capacitors were used. The reason is the same as for feeder 5022, overcompensation, as evident by the negative power factors seen in table 13. This new design can be seen in figure 25.

The voltage profile for the new design is shown in figure 26. It can be seen that the voltage regulator is not stepping up the voltage a lot as it is very close to the substation and no substantial voltage drop has occurred till that location. Considering this fact, a new design was

Table 13: Comparison of operational parameters of Base case and New Design 1 of feeder 5012

Design	Peak Load Time Point				Minimum Load Time Point			
	Max V	Min V	Avg V	P.F.	Max V	Min V	Avg V	P.F.
Base	126.12	118.7	123.3	0.99	128.23	124.3	125.1	-0.79
VPD	122.21	115	120	0.96	124.2	121.1	121.9	0.93

created manually using the results of the VPD algorithm. The switched shunt capacitor was kept at the location determined by the VPD algorithm, but the voltage regulator was not moved. None of the fixed shunt capacitors were used as suggested by the VPD algorithm. This design is referred as New Design2. The voltage profile for this design is shown in figure 26. This design provides a more cost effective design as the expenses for moving the voltage regulator

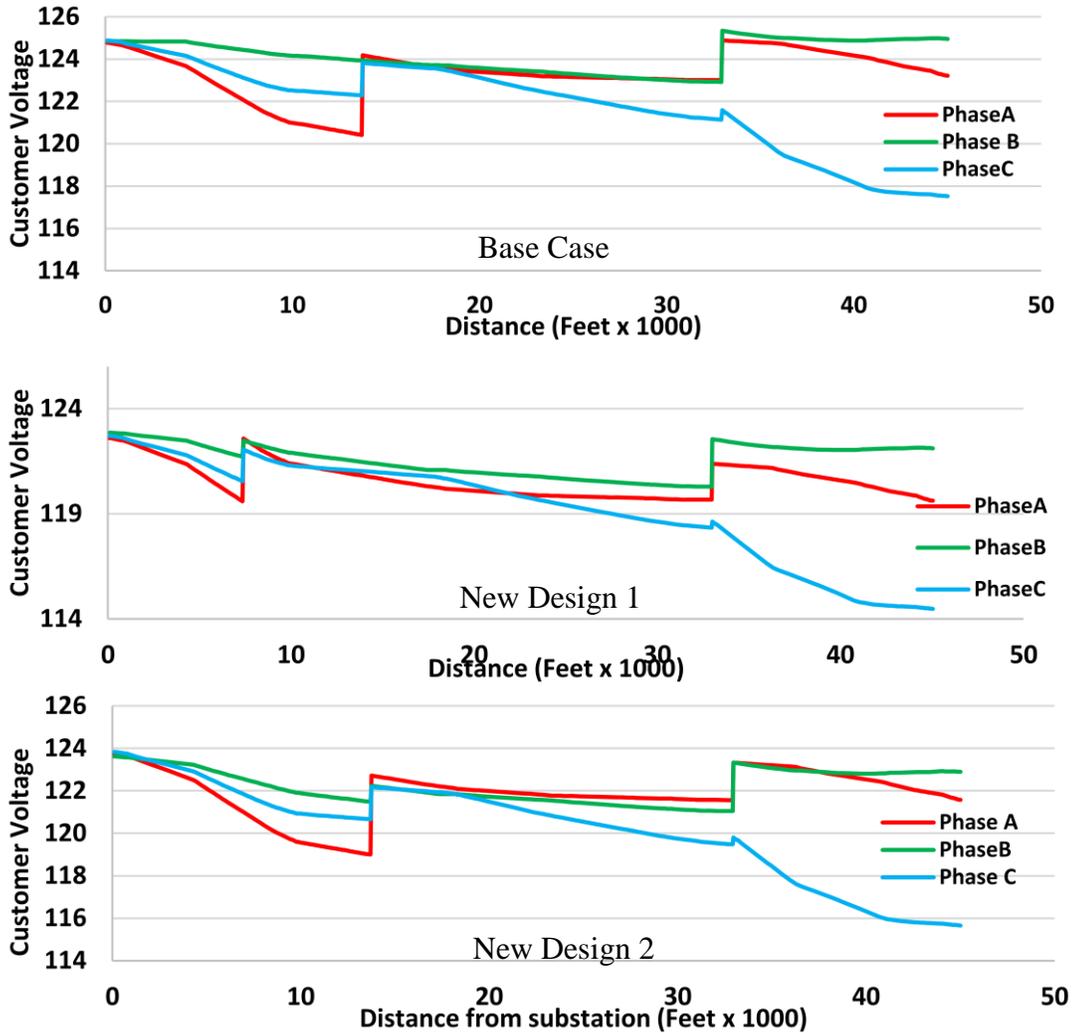


Figure 26: Voltage profile of feeder 5012; base case (top) and new design 1 (middle) and new design 2 (bottom)

and the downtime required for moving this series connected device can be avoided.

Finally, the cost benefits analysis was completed. As expected the VPD algorithm provided significant load and peak reduction with a marginal increase in the losses. A comparison of the base case and two new designs is shown in table 14.

Table 14: Cost Benefits Analysis of base case and new design 1 and 2 of feeder 5012

	Annual MWh	Peak Load (kW)	Losses (MWh)
Base Feeder	19875	3763	443
New Design1	19590	3713	445
% Change	-1.43%	-1.32%	0.45%
Cost Savings due to load reduction			\$ 45,460*
New Design2	19588	3714	448
% Change	-1.44%	-1.30%	1.12%
Cost Savings due to load reduction			\$ 45,853*

* Cost per unit is taken to be 16 cents [56]

Thus, New Design 2 is the cheapest and easiest to implement and leads to maximum cost savings.

6.3.4. Feeder 4021

The base case of this feeder, i.e. the actual model of the feeder as it has been implemented, is shown in figure 27. This feeder has around 1700 load buses, more than 400 distribution transformers, and a peak load of 4988 kW. The feeder has experienced previous voltage issues because voltage control devices have been used to maintain a desired voltage profile. The pre-existing voltage control devices (highlighted in figure 27) on this feeder are:

- 5 fixed shunt capacitor banks injecting 1800 kVar
- 2 voltage regulators, including a 2-phase and a 3-phase voltage regulator, rated 167 kVA/phase each

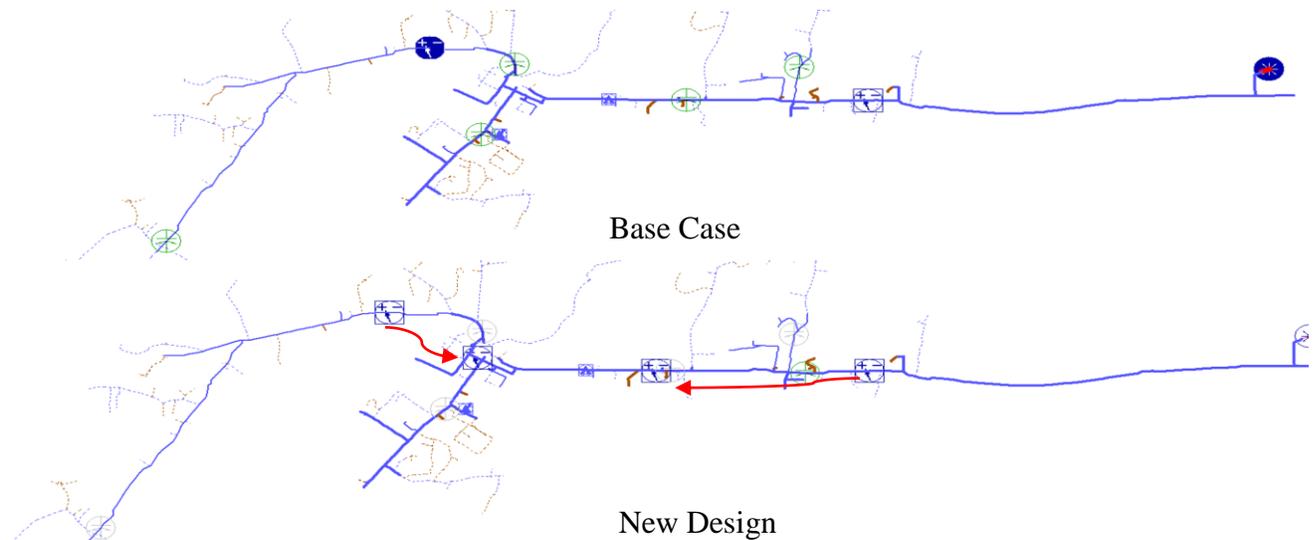


Figure 27: Feeder 4021; Base case (top) and new design (bottom)

The VPD algorithm was implemented restricting the maximum devices that can be placed to those already deployed on the feeder. The new design is shown in figure 26. The algorithm only used 1 of the 5 fixed shunt capacitor banks to avoid overcompensation, and repositioned the voltage regulators. The improved power factors can be observed in the operational parameters shown in table 15. The voltage profiles of the base case and this new design are shown in figure 28. No violations can be observed in the profiles because they have been plotted at the peak load condition. However, table 15 shows that the base case experiences overvoltage during minimum load conditions. The new design on the other hand keeps the voltage within limits while utilizing the entire available range.

Table 15: Comparison of operational parameters of Base case and New Design of feeder 4021

Design	Peak Load Time Point				Minimum Load Time Point			
	Max V	Min V	Avg V	P.F.	Max V	Min V	Avg V	P.F.
Base	125.9	121.5	123.9	0.99	126.7	125.1	125.9	-0.84
VPD	124.2	115.9	121.5	0.94	124.6	121.9	123.4	0.99

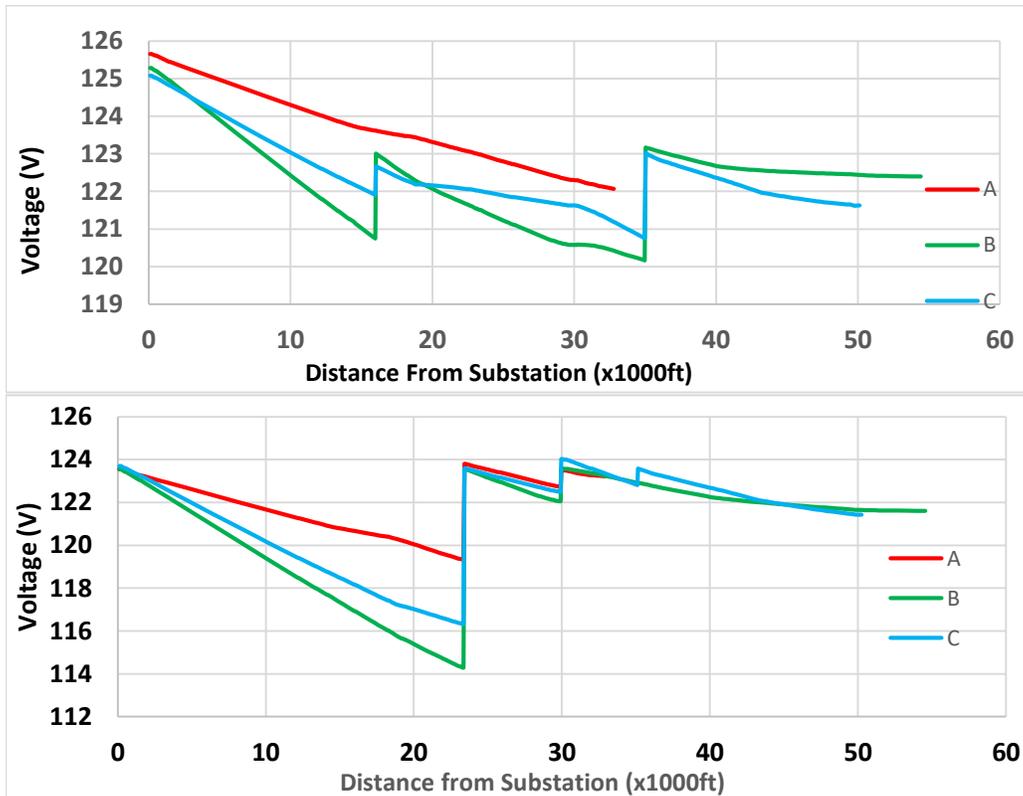


Figure 28: Voltage profiles of feeder 4021, base case (top) and new design (bottom)

Table 16: Cost Benefits Analysis of base case and new design of feeder 4021

	Annual MWh	Peak Load (kW)	Losses (MWh)
Base Feeder	27680	4950	918
New Design	27381	4945	940
% Change	-1.08%	-0.10%	2.33%
Cost Savings due to load reduction			\$ 41,809*

* Cost per unit is taken to be 16 cents [56]

Due to the improved voltages a significant load and peak reduction was also observed during the cost benefits analysis as shown in table 16. There was a marginal increase in losses due to the increase in the current because of the negative voltage dependency factor.

6.3.5. Feeder 3021

The base case of this feeder, i.e. the actual model of the feeder as it has been implemented, is shown in figure 29. This feeder has around 1800 load buses, more than 500 distribution transformers and a peak load of 4246 kW. The feeder has experienced previous voltage issues because voltage control devices have been used to maintain a desired voltage profile. The pre-existing voltage control devices (highlighted in figure 29) on this feeder are:

- 6 fixed shunt capacitor banks injecting 1850 kVar
- 1 switched shunt capacitor bank injecting 600 kVar



Figure 29: Feeder 3021; Base case (left) and new design (right)

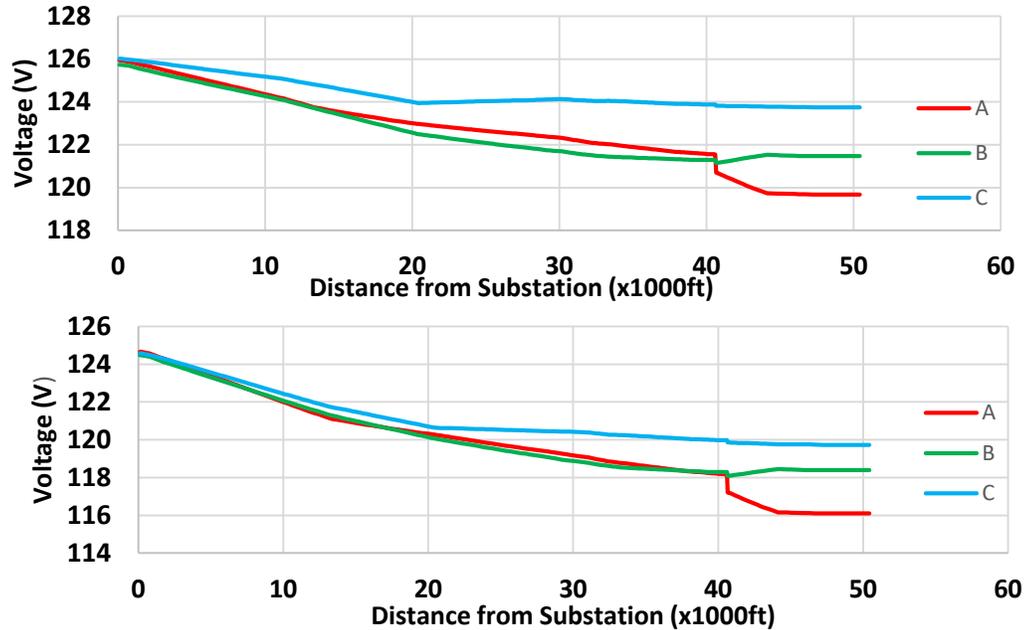


Figure 30: Voltage profile of feeder 3021; base case (top) and New Design (bottom)

The operational parameters for the base case are shown in table 17. This table shows that during the minimum load conditions there is an excessive overcompensation, causing overvoltage and leading power factors. Due to this reason an effective design could not be obtained just by repositioning the capacitor banks. So the VPD algorithm placed a voltage regulator on one of the phases to achieve a much better design. Only 1 of the 6 fixed shunt capacitor banks were used and the switched shunt capacitor bank was repositioned. The new design is shown in figure 29. The system voltages and power factors have improved significantly. These improved parameters and voltage profiles can be seen in table 17 and figure 30, respectively.

Due to the improved voltages a significant load and peak reduction was also observed during the cost benefits analysis, as shown in table 18. There was a marginal increase in losses due to the increase in the current because of the negative voltage dependency factor.

Table 17: Comparison of operational parameters of Base case and New Design of feeder 3021

Design	Peak Load Time Point				Minimum Load Time Point			
	Max V	Min V	Avg V	P.F.	Max V	Min V	Avg V	P.F.
Base	125.8	117.3	122.5	-0.99	126.6	123	124.3	-0.88
VPD	124.0	114	119.4	0.97	124.1	120.5	122.2	0.97

Table 18: Cost Benefits Analysis of base case and new design of feeder 3021

	Annual MWh	Peak Load (kW)	Losses (MWh)
Base Feeder	25397	4272	661
New Design	24997	4184	669
% Change	-1.58%	-2.06%	1.13%
Cost Savings due to load reduction			\$ 56,052*

* Cost per unit is taken to be 16 cents [56]

6.3.6. Feeder 5024

The base case of this feeder, i.e. the actual model of the feeder as it has been implemented, is shown in figure 31. This feeder has around 1350 load buses, more than 400 distribution transformers and a peak load of 3499 kW. The feeder has experienced previous voltage issues because voltage control devices have been used to maintain a desired voltage profile. The pre-existing voltage control devices (highlighted in figure 31) on this feeder are:

- 3 fixed shunt capacitor banks injecting 1800 kVar

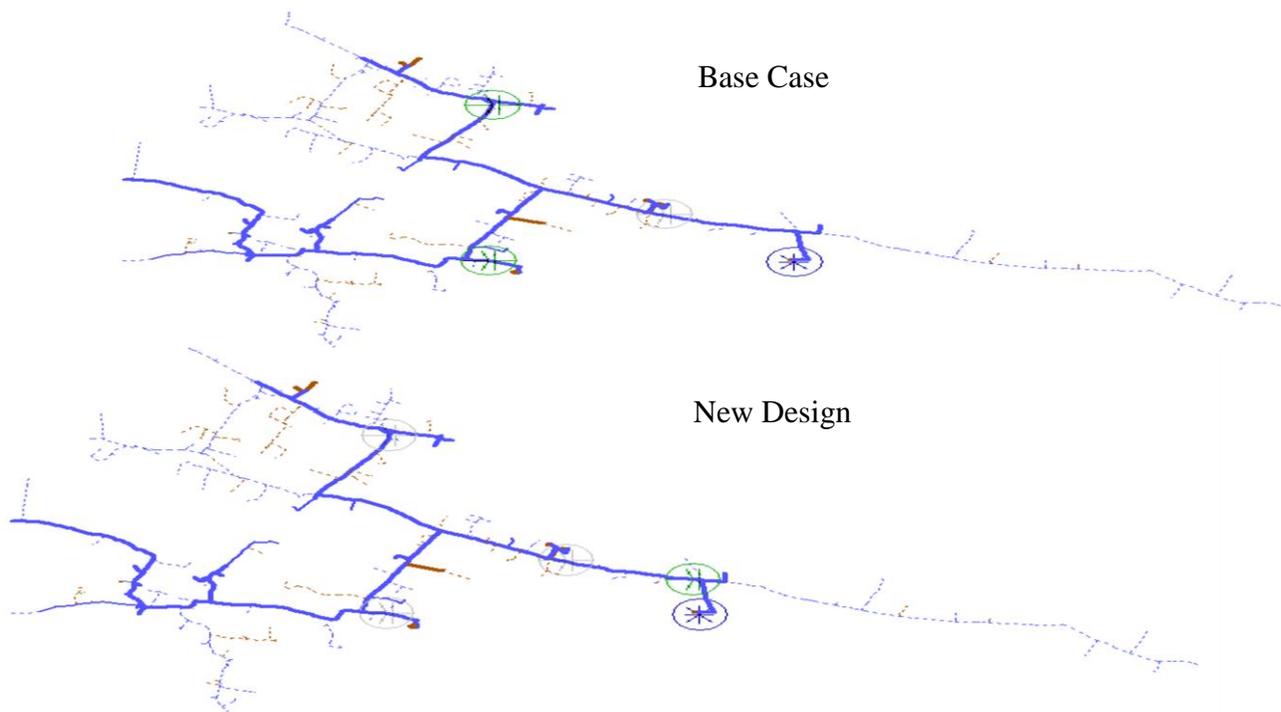


Figure 31: Feeder 5024; Base case (top) and new design (bottom)

Table 19: Comparison of operational parameters of Base case and New Design of feeder 5024

Design	Peak Load Time Point				Minimum Load Time Point			
	Max V	Min V	Avg V	P.F.	Max V	Min V	Avg V	P.F.
Base	124.3	119.8	122	0.98	125.3	124	124.5	-0.91
VPD	121.9	116	118.6	0.94	122.9	120.8	121.8	1

The operational parameters of the base case are shown in table 19. All voltages are within limits, however, there is some overcompensation as evident from the leading power factor during the minimum load conditions. To overcome this problem, the VPD algorithm was implemented and the new design is shown in figure 31. The algorithm ended up using just one fixed capacitor bank and no regulators were required. The improved operational parameters with the new design are shown in table 19. The voltage profiles are also well within ANSI limits as shown in figure 32.

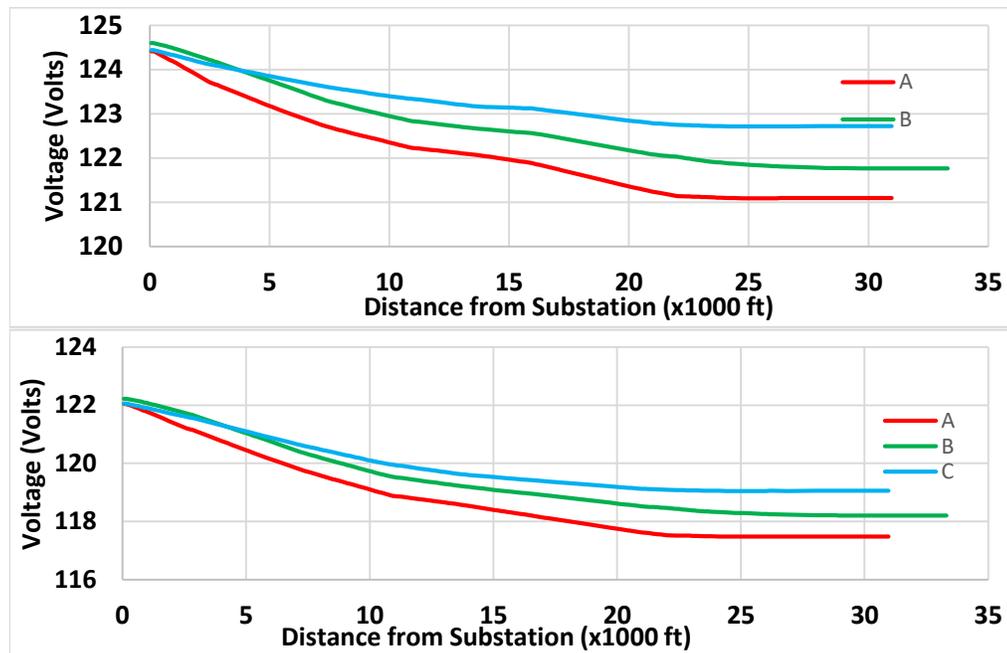


Figure 32: Voltage profile of feeder 5024; base case (top) new design (bottom)

Finally the cost benefits analysis was done to determine if the new design leads to any economic benefits. Due to the improved voltage profile of the new design, there is a significant load and peak reduction as can be seen in table 20. There is also an increase in the losses, which was expected due to the increase in current.

Table 20: Cost Benefits Analysis Feeder of base case and new design of feeder 5024

	Annual MWh	Peak Load (kW)	Losses (MWh)
Base Feeder	21744	3754	400
New Design	21324	3675	424
% Change	-1.93%	-2.10%	6.48%
Cost Savings due to load reduction			\$ 58,802*

* Cost per unit is taken to be 16 cents [56]

The main results from all the redesigns can be summarized as follows:

- The base cases of almost all the feeders had excess compensating kVars being injected. This led to overvoltage violations during the low load conditions. This was because the voltage profile design was done considering just the peak load conditions. The VPD algorithm was able to overcome this problem in all the cases by considering the time varying load profiles.
- Improved operational parameters could be obtained in all of the feeders just by relocating the existing voltage control devices. In some feeders fewer than existing devices were adequate to get the desired results.
- By making a better utilization of the ANSI limits, avoiding overcompensation, and by following a systematic approach where the cheapest available devices were positioned first before moving on to more expensive devices, a significant reduction in annual load demand and annual peak demand were observed. This demand reduction translated to significant cost savings due to the reduced energy consumption.

6.4. PV hosting capacity of feeders after implementation of VPD algorithm

The implementation of the VPD algorithm was found to lead to significant cost savings and load demand reduction. However, the main objective of this research was to determine if the PV hosting capacity of feeders can be enhanced using the VPD algorithm. So finally a detailed PV hosting capacity analysis was performed on all the new designs obtained using the VPD algorithm.

All the input parameters for the analysis were kept exactly the same as for the base feeders. The PV penetration level was increased in 5% step sizes, up to a penetration level of 50%, as the interest is only in determining the minimum hosting capacity. For each penetration level 5 random PV deployment scenarios were analyzed. The individual PV size used is 5kW. The time of analysis is when the PV to load ratio is maximum. The PV inverters were operated at unity power factor without real power curtailment, a commonly used setting for PV hosting capacity studies [14]. The results for the 6 feeders are shown in figure 33. The feeders show a significant improvement in their PV hosting capacities. A tabular comparison of the base cases and the new designs is shown in table 21. Figures in green represent an improvement, or that lesser than or the same number of devices were used. A few cases where an additional device had to be added are highlighted in red. The only exception is feeder 5021. However as discussed in details in section 6.3.2, this feeder

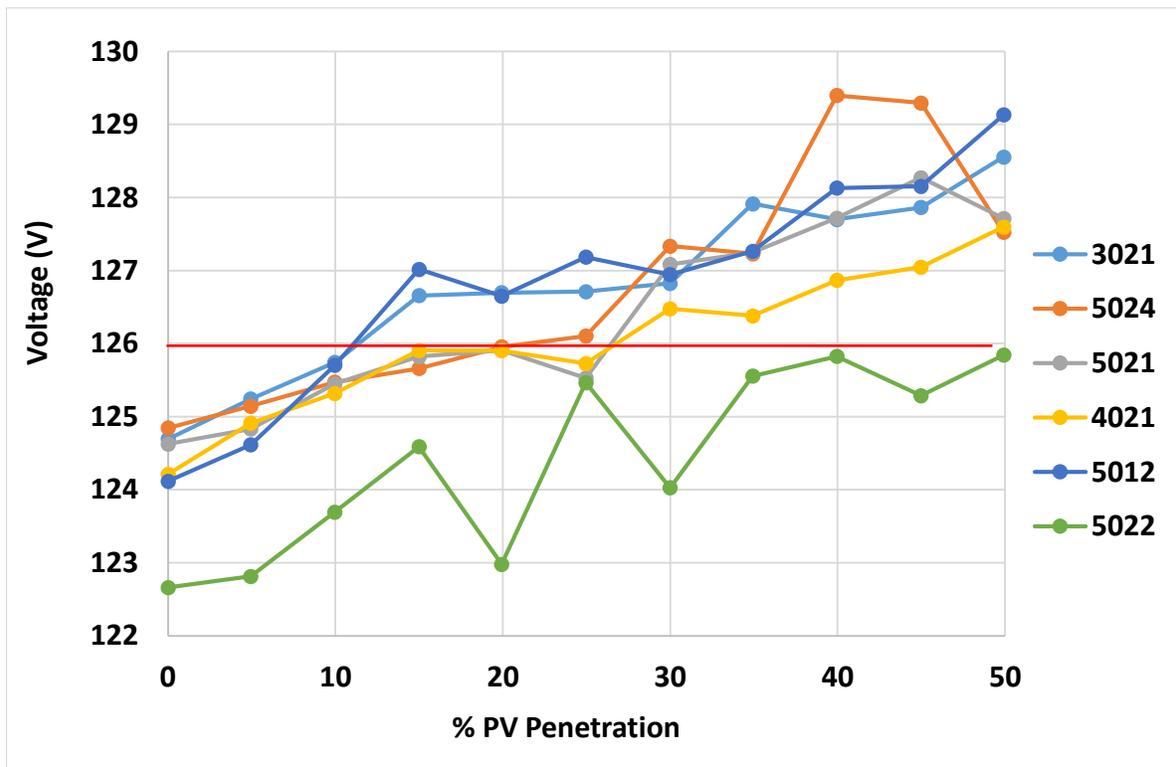


Figure 33: Improved PV hosting capacity of the new designs using VPD algorithm

Table 21: Comparison of the base case and the new design using the VPD algorithm on all feeders

Feeder #	Fixed Shunt Caps		Switched Shunt caps		Voltage Regulators		Actual Minimum PV Hosting Capacity		Load Reduction	Peak Reduction	CVR Factor $= \frac{\% \Delta E}{\% \Delta V}$	Min V
	Base	VPD	Base	VPD	Base	VPD	Base	VPD				
3021	6	1	1	1	0	1(1 phase)	4%	13%	1.58%	2.06%	0.7	114
5024	3	1	0	0	0	0	5%	22%	1.93%	2.1%	0.7	115
5021*	2	1	0	1	0	1	33%	27%	-0.48%	-0.53%	0.8	117
4021	5	1	0	0	2	2	0%	27%	1.08%	0.1%	0.5	116
5012	5	0	1	1	1	1	0%	12%	1.44%	1.31%	0.5	115
5022	3	0	0	0	1	1	14%	>50%	1.91%	1.63%	0.9	117

* Feeder 5021 had undervoltage violations in base case, but no overvoltage violations. Thus if the traditional approach of not considering parameters like undervoltage was used, then the PV hosting capacity would be 33%, as this is when the first overvoltage was observed. However the more detailed approach followed here showed that the effective hosting capacity of this feeder was 0%, which was increased to 27% by VPD algorithm as shown in section 6.3.2

was undercompensated and was experiencing undervoltages which necessitated the use of a few additional devices. Also the CVR factors are all within the range observed in previous CVR studies [22]–[24]. The minimum voltages at peak load conditions are also within ANSI limits.

6.5. Smart Inverter functions

Advanced inverters have the ability to dynamically control the local voltage profile by regulating their real and reactive power output [14]–[16]. In this study the effect of Volt-Var and

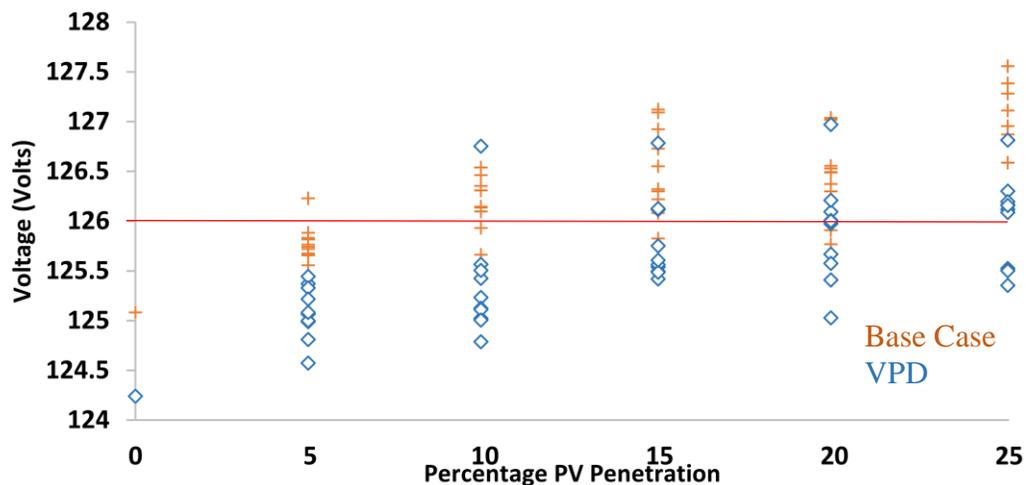


Figure 34: Maximum voltages observed in feeder 3021 using volt-watt control. Base case (orange) new design using VPD algorithm (blue)

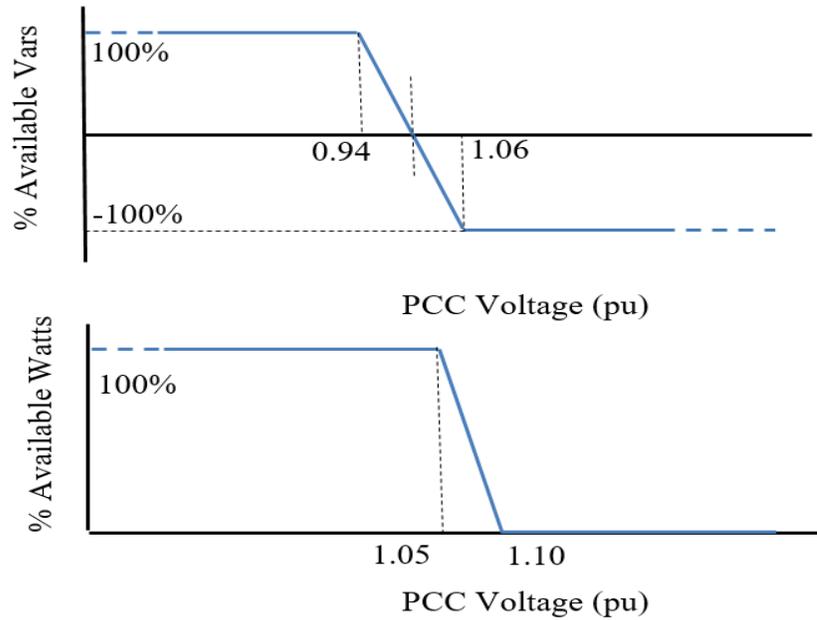


Figure 35. Volt-Var curve (Top); Volt-Watt curve (Bottom)

Volt-Watt control functions is analyzed, using the Point of Common Coupling (PCC) voltage. Smart inverters are being increasingly used to provide localized voltage support. So, it was considered important to see if the VPD algorithm can further enhance these benefits through the improved voltage profile it provides.

Figure 35 shows the Volt-Var and Volt-Watt control curves used for this analysis. The Volt-Watt control function is used when high PV penetration and low load conditions lead to occasional high voltages, or when multiple customers with PV installations being served from the same transformer experience higher voltages [16]. Under these situations curtailing the real power becomes necessary to keep the voltage within operational limits. As shown in the Volt-Watt curve in figure 35, the PV inverter will start curtailing real power once the voltage at the PCC reaches the maximum limit of 1.05 pu, and completely eliminates the power output if the PCC voltage reaches 1.10 pu. This feature is not being used a lot, because it involves real power curtailment to maintain grid parameters within limits, but there is no provision of compensation for this lost generation. This method was implemented on both the base case and the new design for feeder 3021. Figure 34 shows significant improvements in PV hosting capacity using the VPD algorithm over the base circuit design. However, this increase was lesser than that obtained using unity power factor control used in section 6.3. This was expected as this approach will curtail excess real power generation.

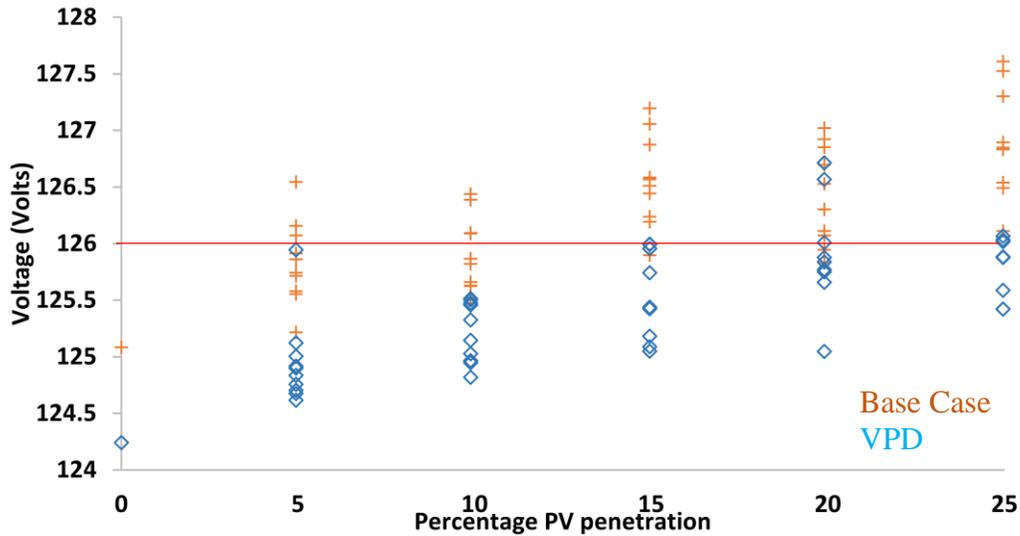


Figure 36: Maximum voltages observed in feeder 3021 using Volt-Var control; base case (orange) new design using VPD algorithm (blue)

The most commonly used smart inverter function is the Volt-Var control. This function can help in injecting as well as absorbing reactive power, thereby improving the local voltage profile. Thus, improved point of common coupling voltages using the VPD algorithm should be able to further enhance this voltage support.

The dead band for the Volt-Var curve, as shown in Figure 35, was set to zero to achieve continuous control. The entire available reactive power is injected when the PCC voltage drops to 0.94 pu, while the maximum possible reactive power is absorbed when the PCC voltage rises to 1.06 pu. The reactive power injection or absorption is limited by the inverter’s VA rating, while avoiding curtailment of real power. Fig. 36 shows that while the base case design had violations at about 5% PV penetration, even with Volt-Var control of the randomly placed PV inverters, the VPD circuit was able to accommodate more than 15% PV penetration without violations. This is due to the improved voltage profile obtained by the optimal placement of voltage regulation devices. So the most effective strategy for PV hosting capacity improvement seems to be the use of Volt-Var control of PV inverters in conjunction with the VPD algorithm.

7. Conclusions

The work presented in this thesis demonstrates the ability of a voltage profile design (VPD) algorithm to enhance the PV hosting capacity of feeders. The VPD algorithm follows a systematic and cost effective approach for the placement of voltage control devices, considering the effect that different types of devices have on the voltage profile. The most inexpensive devices, fixed shunt capacitors, are placed first to remove any voltage violations. If these alone are insufficient, then switched shunt capacitors and voltage regulators are used. Further, the maximum number and types of devices to be used, and their control parameters, are user defined. This makes the algorithm flexible enough to be used for different feeder types.

The main focus of this thesis was on the evaluation of the VPD algorithm. Distribution feeders vary widely in their characteristics, such as line length, existing PV penetration, voltage control devices used, operating voltage levels, and protective devices used to name a few. Due to these differences, testing the VPD algorithm on randomly chosen feeders would not have proved its efficacy. So a statistical approach, clustering, was used to group feeders based on their similarities. This was done using a dataset consisting of an exhaustive list of parameters considered to be important for determining the PV hosting capacity of feeders. Each step in the clustering analysis was accomplished using referenced approaches to clustering. The outcome of this analysis, was a small set of representative feeders, which vary in the characteristics responsible for determining the PV hosting capacity of feeders. The most significant contribution is the validation of clustering results using a detailed PV hosting capacity analysis, which is absent from any study using a sizeable dataset. The analysis showed that the representative feeders indeed vary in their PV hosting capacities. This helped in narrowing down the characteristics to the ones most essential for the determination of PV hosting capacity.

The representative feeders obtained from the clustering analysis should have ideally been used for evaluating the VPD algorithm. However, due to non-availability of feeder topology and time varying load profiles for these feeders, a new set of feeders were chosen using the clustering results. These new feeder sets, chosen from an entirely different geographical region than the ones used in the clustering analysis, provided a great opportunity for independently validating the

clustering results. A detailed PV hosting capacity analysis on these feeders showed that the clustering results hold for these feeders as well. These detailed feeder models thus provided a statistically relevant platform for evaluating the VPD algorithm.

The VPD algorithm was then implemented on all of these feeders. The maximum number of devices to be used by the VPD algorithm were restricted to those already used in the base models. This allowed for a direct comparison between the base case and the new design obtained from the VPD algorithm. It was found that the VPD algorithm was able to give better voltage profiles and power factors in all the feeders using fewer than the existing devices. In a couple of feeders, extra devices had to be used. However, the marginal expense on these new devices can be easily compensated by the significant savings achieved due to load demand reduction in the new designs. The savings were evaluated by comparing the base case and the new design using a time series analysis for an entire year (8760 load time points). Finally, a detailed PV hosting capacity analysis was done on the new designs for all the feeders. The PV hosting capacity was enhanced 2-3 times of the base case on average, while all the voltages were well within the ANSI limits.

Thus, the evaluation of the VPD algorithm proved its ability to give better voltage profiles, improved power factors, and enhanced PV hosting capacities at minimal costs. Also, the methodology adopted for the evaluation of this algorithm can be extended for the evaluation of other smart grid technologies.

The VPD algorithm and the approach followed for its evaluation can be merged to further improve the feeder voltage profile and enhance the PV hosting capacity of feeders. The cost-benefits and the PV hosting capacity enhancement achieved can be included as an objective function. This objective function can be evaluated while deploying every device. This can give an optimized result and is the future direction of work.

References

- [1] ‘ANSI C84.1-2011 - American National Standard for Electric Power Systems and Equipment— Voltage Ratings (60 Hertz)’, National Electrical Manufacturers Association, American National Standards Institute, Inc. 2011.
- [2] R. A. Shayani and M. A. G. de Oliveira, “Photovoltaic Generation Penetration Limits in Radial Distribution Systems,” *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1625–1631, Aug. 2011.
- [3] R.-H. Liang and C.-K. Cheng, “Dispatch of main transformer ULTC and capacitors in a distribution system,” *IEEE Trans. Power Deliv.*, vol. 16, no. 4, pp. 625–630, Oct. 2001.
- [4] T. Senjyu, Y. Miyazato, A. Yona, N. Urasaki, and T. Funabashi, “Optimal Distribution Voltage Control and Coordination With Distributed Generation,” *IEEE Trans. Power Deliv.*, vol. 23, no. 2, pp. 1236–1242, Apr. 2008.
- [5] Y. Agalgaonkar, B. C. Pal, and R. A. Jabr, “Distribution voltage control considering the impact of PV generation on tap changers and autonomous regulators,” in *2014 IEEE PES General Meeting | Conference Exposition, 2014*, pp. 1–1.
- [6] Masoud Ayoubi, Rahmat-Allah Hooshmand, Mehdi Torabian Esfahani, "Optimal capacitor placement in distorted distribution systems considering resonance constraint using multi-swarm particle swarm optimisation algorithm," *IET Generation Transmission and Distribution*, Volume 11, Issue 13, PP3210-3221, 2017
- [7] Hossein Dehghan Dehnavi, Yaghoob Azizi, Mohammad Shafiei, "A new method based on fuzzy system and gravitational optimal detector for capacitor placement, considering nonlinear loads," *2015 30th International Power System Conference (PSC)*
- [8] Shuheng Chen, Weihao Hu, Chi Su, Xiaoxu Zhang, Zhe Chen, “Optimal Reactive Power and Voltage Control in Distribution Networks With Distributed Generators by Fuzzy Adaptive Hybrid Particle Swarm Optimisation Method,” *IET Gener. Transm. Distrib.*, 2013, 9, (11), pp. 1096–1103
- [9] El-Fergany, A.A.: ‘Optimal capacitor allocations using integrated evolutionary algorithms’, *IET Gener. Transm. Distrib.*, 2013, 7, (6), pp. 593–601.
- [10] J. J. Grainger and S. Civanlar, “Volt/Var Control on Distribution Systems with Lateral Branches Using Shunt Capacitors and Voltage Regulators Part I: The Overall Problem,” *IEEE Trans. Power Appar. Syst.*, vol. PAS-104, no. 11, pp. 3278–3283, Nov. 1985.

- [11] S. Civanlar and J. J. Grainger, "Volt/Var Control on Distribution Systems with Lateral Branches Using Shunt Capacitors and Voltage Regulators Part II: The Solution Method," *IEEE Trans. Power Appar. Syst.*, vol. PAS-104, no. 11, pp. 3284–3290, Nov. 1985.
- [12] S. Civanlar and J. J. Grainger, "Volt/Var Control on Distribution Systems with Lateral Branches Using Shunt Capacitors and Voltage Regulators Part III: The Numerical Results," *IEEE Trans. Power Appar. Syst.*, vol. PAS-104, no. 11, pp. 3291–3297, Nov. 1985.
- [13] A. S. Safigianni and G. J. Salis, "Optimum voltage regulator placement in a radial power distribution network," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 879–886, May 2000.
- [14] Rich Seguin, Jeremy Woyak, David Costyk, Josh Hambrick, Barry Mather, (2016), 'High-Penetration PV Integration Handbook for Distribution Engineers,' Technical Report, NREL/TP-5D00-63114.
- [15] J. A. Nelson, 'Effects of cloud-induced photovoltaic power transients on power system protection,' Master of Science Thesis, Department of Electrical Engineering, California Polytechnic State University, San Luis Obispo, 2010. [Online]. Available: [http://digitalcommons.calpoly.edu/theses/430/.](http://digitalcommons.calpoly.edu/theses/430/)"
- [16] "Common Functions for Smart Inverters: 4th Edition. EPRI, Palo Alto, CA: 2016. 3002008217."
- [17] "IEEE 1547-2003 - IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems." [Online]. Available: <https://standards.ieee.org/findstds/standard/1547-2003.html>. [Accessed: 03-Jan-2018].
- [18] T. Stetz, K. Diwold, M. Kraiczy, D. Geibel, S. Schmidt, and M. Braun, "Techno-Economic Assessment of Voltage Control Strategies in Low Voltage Grids," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 2125–2132, Jul. 2014.
- [19] "Michael Coddington, Barry Mather, Benjamin Kroposki, Kevin Lynn, Alvin Razon, Abraham Ellis, Roger Hill, Tom Key, Kristen Nicole, Jeff Smith, 2012, 'Updating Interconnection Screens for PV System Integration.'"
- [20] D. Cheng, B. A. Mather, R. Seguin, J. Hambrick, and R. P. Broadwater, "Photovoltaic (PV) Impact Assessment for Very High Penetration Levels," *IEEE J. Photovolt.*, vol. 6, no. 1, pp. 295–300, Jan. 2016.
- [21] "Electric Power Research Institute, 'Stochastic Analysis to Determine Feeder Hosting Capacity for Distributed Solar PV,' 1026640."

- [22] R. F. Preiss and V. J. Warnock, "Impact of Voltage Reduction on Energy and Demand," IEEE Trans. Power Appar. Syst., vol. PAS-97, no. 5, pp. 1665–1671, Sep. 1978.
- [23] D. M. Lauria, "Conservation Voltage Reduction (CVR) at Northeast Utilities," IEEE Power Eng. Rev., vol. PER-7, no. 10, pp. 58–59, Oct. 1987.
- [24] A. Dwyer; R. E. Nielsen; J. Stangl; N. S. Markushevich , "Load to voltage dependency tests at B.C. Hydro," IEEE Transactions on Power Systems, 10, (2), pp 709-715.
- [25] H. L. Willis, H. N. Tram, and R. W. Powell (1985). 'A Computerized, Cluster Based Method of Building Representative Models of Distribution Systems.' IEEE Trans. Power Appar. Syst., vol. PAS-104, no. 12, pp. 3469–3474.
- [26] K. P. Schneider, Y. Chen, D. Engle, and D. Chassin, R. Pratt, S. Thompson (2008) 'Modern Grid Initiative - Distribution Taxonomy Final Report' PNNL.
- [27] J. Dickert, M. Domagk, and P. Schegner, "Benchmark Low Voltage Distribution Networks Based on Cluster Analysis of Actual Grid Properties," 2013 IEEE Grenoble Conference.
- [28] Berry, A.M., Moore, T., Ward, J.K., Lindsay, S.A. and Proctor, K. (2013) 'National Feeder Taxonomy – Describing a Representative Feeder Set for Australian Networks,' CSIRO, Australia.
- [29] W. Biener, G. Gunther, S. Killinger, B. Wille Haussmann, (2016) 'Evaluation of Methods for Estimating Distribution Grid Stress Due to Future Installations of Photovoltaic Units,' 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC).
- [30] G. Walker, H. Nägele, F. Kniehl, A. Probst, M. Brunner, S. Tenbohlen (2015), 'An application of cluster reference grids for an optimized grid simulation.' 23rd International Conference on Electricity Distribution (CIRED) paper 0753.
- [31] Akshay K. Jain, Barry Mather, Technical Report of the IEA PVPS Task 14 Subtask 2 – Activity 2.5 – Taxonomy of Distribution Feeders with High PV Penetration.
- [32] Matthew Rylander, Robert J. Broderick, Matthew J. Reno, Karina Munoz-Ramos, Jimmy E. Quiroz, Jeff Smith, Lindsey Rogers, Roger Dugan, Barry Mather, Michael Coddington, Peter Gotseff and Fei Ding, 'Alternatives to the 15% Rule,' technical report, Sandia National Laboratories.
- [33] William H. Kersting, "Distribution System Modeling and Analysis," Third Edition. .

- [34] Distribution Test Feeders - Distribution Test Feeder Working Group - IEEE PES Distribution System Analysis Subcommittee." [Online]. Available: <https://ewh.ieee.org/soc/pes/dsacom/testfeeders/>. [Accessed: 15-Jan-2018].
- [35] Kron, G., "Tensorial analysis of integrated transmission systems, part I, the six basic reference frames," AIEE Transactions, 71, 1952.
- [36] Carson, J.R., "Wave propagation in overhead wires with ground return," Bell System Technical Journal, 5, 539, 1926.
- [37] Turan Gonen, "Electric Power Distribution Engineering," Third Edition, CRC Press (2014). .
- [38] R. P. Broadwater, 'DEW Theory - Generic Analysis with Topology Iterators for Design, Operation, and Real-Time Control of Large Scale, Reconfigurable Systems,' in Course Note for Object Oriented Integrated System Analysis, Virginia Tech, 2007.
- [39] L. R. Feinauer, K. J. Russell, and R. P. Broadwater, "Graph Trace Analysis and Generic Algorithms for Interdependent Reconfigurable System Design and Control," Nav. Eng. J., vol. 120, no. 1, pp. 29–40, Mar. 2008.
- [40] R. J. Broderick, J.R. Williams, K. Munoz-Ramos (2014). 'Clustering Method and Representative Feeder Selection for the California Solar Initiative.' Sandia Report, SAND2014-1443, Sandia National Labs.
- [41] J. Cale, B. Palmintier, D. Narang, K. Carroll (2014). 'Clustering Distribution Feeders in the Arizona Public Service Territory.' in proc. of IEEE Photovolt. Spec. Conference.
- [42] https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient.
- [43] Kendall, M.G. (2009), 'Rank Correlation Methods.' New York: Hafner Publishing Co.
- [44] G. W. Milligan and M. C. Cooper, (1985) 'An examination of procedures for determining the number of clusters in a data set,' Psychometrika, vol. 50, no. 2, pp. 159–179, pp. 159–179.
- [45] Alternatives to the 15% Rule: Modified Screens and Validation. EPRI, Palo Alto, CA: 2015. 3002005791.
- [46] Aruna Bhat (2014), 'k-medoids clustering using partitioning around medoids for performing face recognition,' International Journal of Soft Computing, Mathematics and Control (IJSCMC), Vol. 3, No. 3,.

- [47] Akshay K. Jain, Barry Mather, 'Clustering Methods and Validation of Representative Feeders,' IEEE PES T&D Conference and Exposition, 2018.
- [48] German association for electrical, electronic & information technologies (VDE), "Guideline for the Connection and Parallel Operation of Generation Units at Low Voltage Level," VDE-AR-N 4105, 2011.
- [49] "Technical Guideline for the Connection and Parallel Operation of Generator Connected to the Medium Voltage Network," German association of energy and water industries (BDEW), 2008.
- [50] "Rule 21 Governing Generation Facilities Connection to the Distribution Grid." [Online]. Available: <http://www.cpuc.ca.gov/Rule21/>. [Accessed: 03-Jan-2018].
- [51] D. Cheng, Y. Liang, D. Zhu, and R. P. Broadwater, "Real-Time Power Electric System Modeling, Assessment and Reliability Prediction," in 2009 IEEE/PES Power Systems Conference and Exposition, 2009, pp. 1–6.
- [52] Jason Bank, 'PV Hosting Capacity Application User Guide and Instruction Manual,' DEW/ISM Modeling Software.
- [53] "Current Utility Screening Practices, Technical Tools, Impact Studies, and Mitigation Strategies for Interconnecting PV on the Electric Distribution Systems." EPRI, Palo Alto, CA: 2014. 3002003277.
- [54] "Clustering Methods and Feeder Selection for California Solar Initiative," EPRI, Palo Alto, CA: 2014. 3002002562."
- [55] "Alternatives to the 15% Rule: Modeling and Hosting Capacity Analysis of 16 Feeders. EPRI, Palo Alto, CA: 2015. 3002005812."
- [56] "EIA - Electricity Data." [Online]. Available: https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a. [Accessed: 10-Jan-2018].