# Machine learning Approach to Event Detection for Load Monitoring

Priyabrata Sundaray

A Thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Science

(Electrical Engineering)

at the

University of Wisconsin - Madison

2019

# Machine Learning Approach to Event Detection for Load Monitoring

## Abstract

Today's Electricity market is highly interconnected. The actions of customers, transmission and distribution systems, generation systems, markets, and service providers are correlated. The dynamic stability of the grid depends on the dynamic characteristics of all these components and systems. The least understood component is the consumer load. While average power is metered for billing, the voltage or current relationships and dynamic characteristics of loads are not monitored.

It is the purpose of this work to analyze high-bandwidth voltage and current measurements from an office building to characterize certain load features. The focus is on transient events under the premise that some features will be revealed when loads switch on and off, or change mode of operation. This approach involves certain tasks. First an event needs to be identified, and the duration of transient needs to be estimated. Second, features that might be useful for classifying transients must be determined. Third, classification is performed using machine learning Techniques.

Signal Processing tools based on changes in current waveforms are used to detect events. Many features are examined in this work including voltage and current harmonics, power factor, active, reactive, and distortion power. Machine Learning techniques guide the classification task. These include self-organising maps and neural network.

Analysis and results are presented based on actual load measurements.

# Acknowledgments

I am grateful to Prof. Bernard Lesieutre, an erudite, for his precious time and for providing me this opportunity to work on an interesting problem of Load Monitoring in the broad field of power system reliability. I thank him for his confidence on me and his constant guidance throughout the period of my research work.

Thanks to Dr. Varun Jog for his kind feedback on the model for clustering the events using k-means and and his opinion on whether to use of Fuzzy Inference systems with regards to the data used for the work. I thank Dr. William Sethares for his precious mentoring provided throughout the work of using Machine Learning Models and training Neural Networks. I am grateful to Dr. William Sethares for his kind feedback on the performance metric used to compare the performance of the models. I thank Dr. Line Roald for her feedback about the features to be used in the work and other possibility to use optimization on the features selection technique.

Thanks to Bonneville Power Administration (BPA) an American federal agency operating in the Pacific Northwest, for providing the Data and research funding.

I am thankful to the University of Wisconsin-Madison as well as the faculty and staff of the Department of Electrical and Computer Engineering for their constant support and encouragement, to enable me to work on this research.

Lastly, I am grateful to my Parents, my brother and my friend Lane Meddaugh, for their patience and support throughout the time of my research and the supreme personality of Godhead for providing me this opportunity.

# Abbreviations

**AMPD** . . . . Automatic Multi-Scale Peak Detection

**BMU** . . . . . Best Matching Unit

**BPA** . . . . . Bonneville Power Administration

**DTW** . . . . . Dynamic Time Warping

**FFT** . . . . . Fast Fourier Transform

**ILM** . . . . . Intrusive Load Monitoring

**LMS** . . . . . Local Maxima Scalogram

**MMPD** . . . Modified Multi-Scale Peak Detection

**MSE** . . . . . Mean Square Error

**NILM** . . . . Non-Intrusive Load Monitoring

**PCA** . . . . . Principal Component Analysis

**PF** . . . . . Power Factor

**RMS** . . . . . Root Mean Square

**SOM** . . . . . Self Organising Map

**THD** . . . . . Total Harmonic Distortion

**VAR** . . . . . Volt Ampere Reactive

# Contents

# List of Figures

# List of Tables

# Chapter 1

# INTRODUCTION

## 1.1 OVERVIEW OF RESEARCH - TRACKING LOADS

The composition of loads affects the stability in an electric power grid. Further knowledge about the composition of the loads can help both customers and grid operators. The information can provide customer with greater control over their own equipment. It may help them reduce their excess consumption or simply provide a degree of freedom in order to do further analysis on their use of electricity. Moreover, the information that load monitoring provides to the utilities may guide their future actions in load management. The correlation between the load types and the peak load level, over a period of time can help to predict the load trend [6] [15]. Further, the trends in the use of electricity by the consumers or of a community is a very valuable information for policy makers.

The literature is abundant with methods to decompose loads based on load sig-

natures. Due to the diverse definitions of event and even a more diverse set of signatures associated with every event, the literature is full of various methods to detect events. However, the majority of these methods have used abnormalities in the current signal and some have used voltage signal discrepancies to detect disturbances [22].

Recently there has been a surge in the use of Machine Learning techniques to understand the behavior of events [23]. With the rapid development of measuring instruments at higher sampling frequencies, the large amount of data being received from energy meters can be coupled with Machine Learning tools to perform analysis and make predictions. This work focuses on performing detailed energy sensing, of loads. Moreover this study focuses on selecting the optimum features for Machine Learning Algorithms.

Load Monitoring can be performed in two ways.

**i.** Intrusive Load Monitoring (ILM) or Distributed sensing methods, and
**ii.**Non-Intrusive Load Monitoring (NILM) or Single point sensing methods

Intrusive Load Monitoring or Distributed sensing methods requires one or more than one sensor per appliance for monitoring [23]. But this method has obvious practical disadvantages, which includes high costs, multiple sensor configuration, and installation complexity. Non-Intrusive Load Monitoring (NILM) estimates the energy consumed by individual appliances and their states utilizing aggregate measurements from a small number of power meters [7]. Among the two methods,

NILM doesn't require intrusion into the customers' premises for monitoring energy consumption.

This work is non-intrusive, as it relies on data gathered at the service entry of a building. In this work, 60Hz Current and Voltage signals from a building are analyzed. The waveforms have been sampled at a rate of 5kHz. The high frequency sampling helps determine Electric Load Signatures from the real-time monitored data.

In order to detect the events, the characteristic features of the individual loads are ideal for signatures. Because the measurements are aggregated, individual load characteristics may be revealed during turn-on events and turn-off events. The unique characteristics in the electrical signals are attributed to the load signatures and form the basis for event detection. Other disturbances may also be present, including events such as disturbances with equipment, arcing events, and equipment failures etc. These disturbances in the signal can also be found and by analyzing the signatures based on the observations and may lead to preventative actions for eliminating the problems.

Out of the multiple components connected to the system, some include power electronic components. Knowledge of the level of power electronics load can be useful for certain analysis, especially stability. In order to distinguish a power electronic load, the features of those loads in terms of both steady state and dynamic state are used in this work to help in the process of load classification.

The features if properly selected, can help to determine the starting turn-on instance of a component and also help to determine when the component was turned off, based on the similarity in the feature space of the component. As some loads may dominate the characteristics, and it may be difficult to isolate the smaller components from the larger ones. Also, if the feature space is too large, it becomes difficult for measured features to converge to unique component classification. Hence, there is a trade off between the number of features and classifying both larger and smaller loads. Although it is difficult to classify all load components connected to the system, it may be possible to cluster select load classes and classify those into electrical or power electronic loads based on the feature space of the individual classes.

## 1.2 BACKGROUND AND LITERATURE REVIEW

In literature, authors have achieved load segregation by decomposing the total load into individual appliance loads. Almost three decades back, George Hart [7] first formulated the technique of separating total power into a sum of constituent powers:

$$P(t) = \sum_{i=1}^{n} P_i\ (t) \tag{1.1}$$

**where,**

n: The total number of active appliances

P(t): Total power

$P_i$ (t): Power consumption of individual appliances

In contrast to this traditional methods for load monitoring, our work considers multiple features of the given signal. More detailed descriptions about the features and the feature selection methods are explained in Chapter 3. These features are used to cluster the total load into independent loads using Machine Learning Clustering techniques as explained in Chapter 5. Note that events with 'n' number of unique features can be associated with an n-dimensional feature space. Then, the clustering technique clusters the events into subgroups as described in Chapter 5.

### 1.2.1 EVENT DETECTION

The majority of literature work have focused on using abnormalities in the current signal in order to detect events. Authors in [21] and [9] have used the threshold technique on superimposed fault current component; [18] has used a simple threshold technique; [14] has presented a composite method combined with Threshold technique to detect a disturbance. The technique used in this work for detecting events is explained in a detailed manner in Section 2.3.2.

## 1.3 CONTRIBUTION OF THIS RESEARCH

This work provides insight into the use of both time and frequency domain analysis of a signal to enhance the capability of a Load Monitoring system.

The contributions are in the following broad categories:

**i**: Comparison for Fast vs Slow Transients

**ii**: Selection of optimal features for the segregation of Loads

**iii**: Clustering the loads using Machine Learning Techniques

## DATA SETUP

The analysis relies on data obtained from BPA Headquarters, Portland, Oregon.

The data constitutes the building current and voltage measurement from mains supply and 5 feeders, sampled at 5kHz. The analysis has been performed on a total of 21 data sequences, in this work.

Every 1 sec worth of a signal comprises 5000 sample points. A current waveform for a period of 1 sec, is as shown in Figure 1.1.

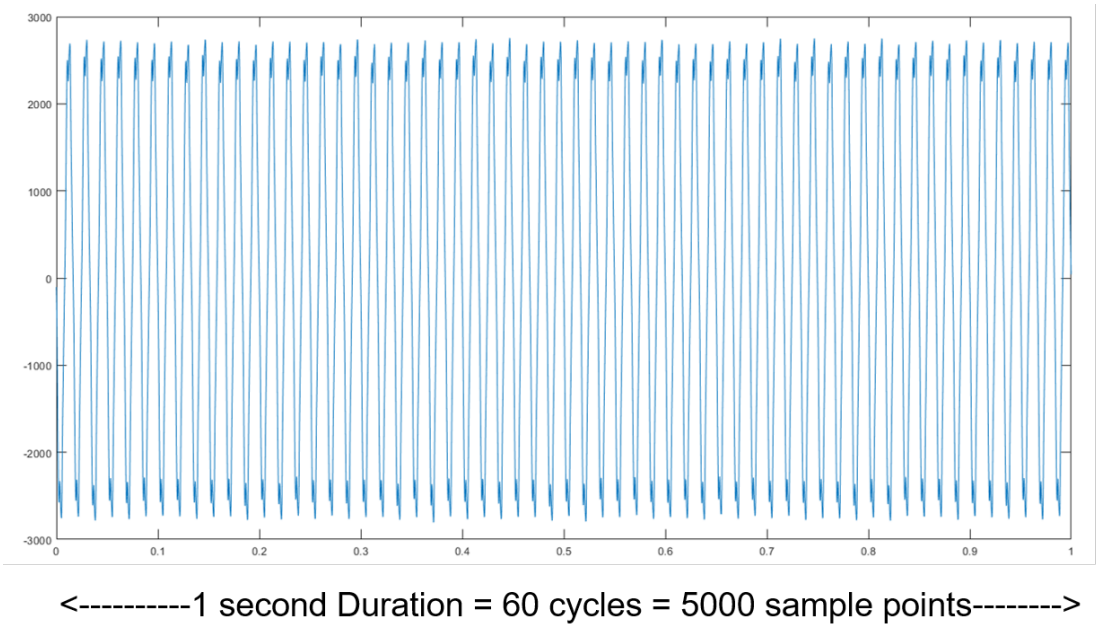



Figure 1.1: Current Waveform for a period of 1 sec

## DETECTION OF EVENTS

Every time, a component is turned on or turned off, an event is triggered. A detection signal helps to determine the time at which the triggering of the event started

as well as the end of the transient response.

For detection of events, a detection waveform is used based on the characteristics of the given input signal. The segment RMS value of the differential current waveform is used as the detection signal in this study. Then a peak detection algorithm is used to detect the event occurrences. The starting and ending of the event is detected using the inverse peaks of the detection signals. The duration of the event is determined and the event is characterized by peak location as well as the duration of the event.

**LOAD ANALYSIS**

The events are classified into clusters, formed on the basis of features associated with every event. In this work, the various electrical parameters which are used to analyze the load signatures are : mains current and voltage signals, power of differential current signal, electrical power signals such as active power, reactive power and apparent power for the fundamental component as well as the harmonic components up to the $15^{th}$ harmonic, and power factor.

## 1.4 OUTLINE OF THESIS

The objective of the research is to perform detailed energy sensing and to provide information on the energy consumed. Hence, the overall research is divided into the following steps. First, obtain the load data. After obtaining the load data, the data is passed through a pre-processing phase, to clean the data. Then the signal undergoes a reduction phase. The reduction of the signal is performed in order to obtain the

detection function signal. This form the basis for further analysis. Next, based on the detection function signal, electrical signatures are obtained. These serve as the features. Third, after analyzing the features obtained, segregation of loads is performed via supervised training models and using a unsupervised classification model. Finally, the performance is evaluated in terms of segregation of electrical and electronic loads.

# Chapter 2

# DETECTING EVENTS

## 2.1 METHODS OF DETECTING WAVEFORM ABNORMALITY

The literature has an abundant set of methods to detect electrical events. The majority of these use abnormalities in the current signal and some use voltage signal discrepancies as well [22]. This section presents some research in this area using reference [22] and a literature review. Finally a method to detect events by combining the advantages of the methods in the literature is presented.

### 2.1.1 ABNORMAL CURRENT METHOD

The superimposed fault current component which is the current signal less the normal load component, is used in references [9] and [21] to detect disturbances. The abnormal current component can be derived as follows.

$$i_{F(k)} = i_{(k)} - i_{(k-N_M)} \tag{2.1}$$

where,

$i_{F(k)}$: Superimposed abnormal component current values for phase A, B and C

$i_{(k)}$: Instantaneous current values for phase A, B and C

k: Sample index

$N_M$ : Multiple of Number of samples per power cycles

Both the references [9] and [21] have used $N_M = 2^*N_1$, where $N_1$ represents the number of samples per power cycle. Abnormal components are very small under steady state, normal operating conditions [2]. However during faults or other switching events the abnormal components will be higher than certain threshold value. There are two ways to determine the threshold value for a disturbance to occur.

**a.** Using threshold on magnitude of fundamental frequency abnormal current

**b.** Using threshold on instantaneous abnormal current

**a. Using threshold on magnitude of fundamental frequency abnormal current**

The fundamental component of the superimposed abnormal current waveform is calculated using a Fourier transform. If the component exceeds a certain threshold as established or determined by analyzing current cycles, a disturbance is detected.

$$i_{F(MAG)} > i_{THRESHOLD} \tag{2.2}$$

where,

$i_{F_(MAG)}$: Magnitude of Fundamental Frequency Abnormal current values for phase A, B and C

$i_{(THRESHOLD)}$: Threshold values for phase A, B and C

Both references [21] and [9] have used a similar methods. The difference being, [21] has performed a full cycle Fourier transform, whereas [9] has performed a half cycle Fourier transform to obtain the fundamental component of abnormal current.

**b. Using threshold on instantaneous abnormal current**

This is presented in [14] for a threshold technique on instantaneous values of abnormal components to detect arcing events.

$$|i_{F(k)}| > i_{THRESHOLD} \tag{2.3}$$

where,

$i_{F_(k)}$: Absolute values of Abnormal current values for phase A, B and C

$i_{(THRESHOLD)}$: Threshold values for phase A, B and C

k: Sample index

### 2.1.2 WAVELET CURRENT ANALYSIS

It is proposed in reference [21] to use a wavelet analysis technique to detect a fault with a short duration $\leq 3$ cycles. Using this technique the waveform can be analyzed using low frequency approximation coefficients and high frequency detail coefficients. The low frequency approximation coefficients can represent the fundamental frequency component, while the high frequency detail coefficients can represent the transient state [21]. There are two triggering methods, by which a decision about a disturbance can be made:

**a.** Using approximate coefficients

**b.** Using detailed coefficients

#### a. Using approximate coefficients

Using low frequency (0 - 240 Hz) approximation coefficients, the following detection rule is formulated.

$$RMSCR = \frac{RMS_{latest-half-cycle} - RMS_{one-cycle-before}}{RMS_{one-cycle-=before}} > THRESHOLD \quad (2.4)$$

where,

$RMS$: Root mean Square values for phase A, B and C

RMSCR: A derived parameter for phase A, B and C

This rule is insensitive to noise because it is not related to the high frequency components [22].

**b. Using detailed coefficients**

Using high frequency (240 - 960 Hz) detailed coefficients, a disturbance is detected when the following rule is satisfied.

$$ENGR = \frac{Energy_{latest} - Mean(Energy_{past})}{STD(Energy_{past})} > THRESHOLD \tag{2.5}$$

where,

$Energy_{latest}$: Energy of the latest detail coefficients

$Energy_{past}$: Array of energy of the past detail coefficients

MEAN: Average function

STD : Standard deviation function.

This rule has good performance in a low noise environment. Since it does not consider the low frequency component, it is insensitive to the slow change of fundamental frequency component [22].

### 2.1.3 FUNDAMENTAL FREQUENCY CURRENT METHOD

In reference [13], a technique has been presented to detect faults in medium voltage circuits.

$$i_{(MAG)} > i_{THRESHOLD} \tag{2.6}$$

where,

$i_{(MAG)}$: Magnitude of Fundamental Frequency current values for phase A, B and C

$i_{(THRESHOLD)}$: Threshold values for phase A, B and C

In order to obtain the fundamental component of the waveform, a DFT on every one eighth of a power cycle is performed and the magnitude of the fundamental component of the waveform is compared with a threshold value. If the fundamental component of the current waveform exceeds certain threshold a disturbance is detected.

### 2.1.4 VOLTAGE WAVEFORM METHOD

To detect a disturbance by comparing consecutive cycles of a voltage waveform, the following methods have been adapted in literature [22].

(1) A voltage waveform is superimposed by shifting the waveform by one power frequency cycles. If the resulting voltage abnormal component exceeds a certain threshold for a certain amount of duration, a disturbance is considered to be detected.

(2) A difference of the consecutive cycles squared is calculated for the voltage waveform . The absolute values of the difference forms MAVSA (Mean Absolute Variation in Squared Amplitude). The MAVSA value is compared with a certain threshold value, and if it exceeds the threshold, a disturbance is detected.

(3) The RMS value of the voltage differential waveform is calculated after the voltage waveform is superimposed by shifting the waveform by one power frequency cycle. A percentage value is calculated as the ratio of the RMS value of the differential cycle to that of a healthy cycle. If the percentage value exceeds a certain

threshold, a disturbance is detected.

The advantages and disadvantages of voltage waveform methods are discussed in [22]. The first method is prone to detect a lot of inconsequential disturbances or miss important ones, if the threshold value is not selected carefully. The third method calculates a percentage value based on an entire cycle, it may not detect those disturbances which have a duration of less than a cycle.

### 2.1.5 WAVELET VOLTAGE ANALYSIS

To detect voltage transients, fundamental frequency and high frequency components are analyzed in reference [16]. Peaks in the high frequency components are compared with a part of the signal which is assumed to be disturbance-free [22].

### 2.1.6 COMPOSITE METHOD

A composite method using voltage and current signals to detect disturbances is presented in [2]. The running average of the following parameters are considered for the purpose of disturbance detection [22].

**a.** Components below $16^{th}$ harmonic are computed for each current input, for every period of two cycles.

**b.** Current and voltage RMS values are computed for each input, for every period of two cycles.

**c.** The real, reactive, and apparent power are computed for every period of two cycles.

**d.** Energy is computed for the high-frequency current channels.

Then,

$$X_{(i)} > \alpha * X_{MEAN}$$
$$X_{(i)} < \beta * X_{MEAN} \tag{2.7}$$

where,

$X_{(i)}$: The $i^{th}$ value of any of the above parameters

$X_{(MEAN)}$: The running average value

$\alpha$ : The coefficient (larger than 1) corresponding to the upper threshold

$\beta$ : The coefficient (less than 1) corresponding to the lower threshold

Post calculation of the above parameters, each parameter is compared with an upper and a lower threshold value, every two cycles interval. If, any of the parameters exceeds the thresholds, a disturbance is detected.

## 2.2 METHOD OF DETECTING EVENTS

An abnormality detection method has been proposed in [22] by combining the advantages of the methods in literature. It follows the principle of using a differential waveform between two consecutive cycles. Then the RMS of each segment is calculated after dividing the difference waveform into M segments. This method detects the disturbance based on a threshold technique. Moreover, another detection method has also been proposed in [22], by analysing the RMS value of the current

and voltage signal space.The RMS value of the signal is evaluated in every time period and if the value exceeds a certain threshold, there is an event detected.

Figure 2.1 shows the comparison of current and voltage waveform and the RMS techniques discussed above, along with the ability of the techniques to capture abnormalities or disturbances [22].

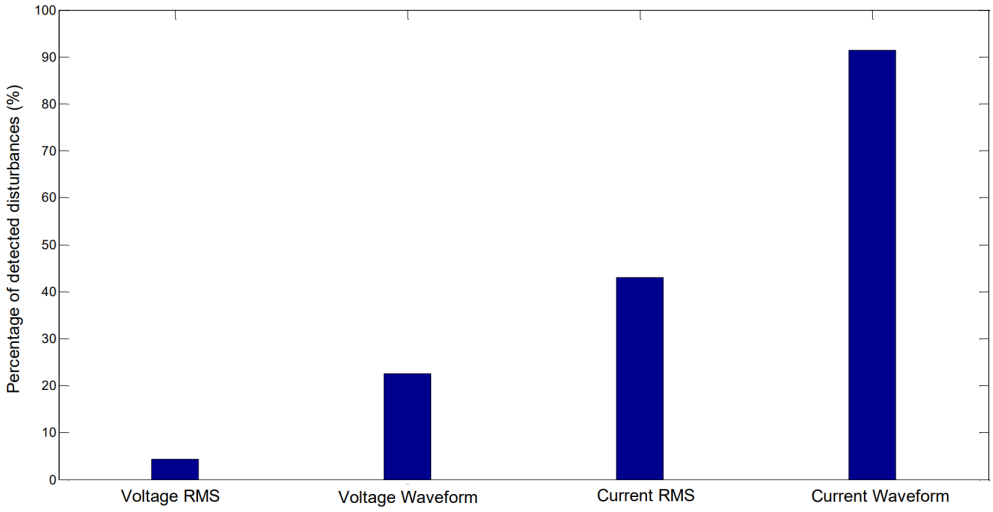



Figure 2.1: Plot for comparison of various techniques in terms of percentages of events detected [22]

The bar graph in Figure 2.1 shows that the current waveform technique has the highest capability in capturing disturbances compared to other techniques. The approach proposed in this research work uses current signal abnormalities as the basis for detection of events. In order to address to the diverse set of signatures associated with every event, this study uses voltage signal abnormalities, fundamental component powers as well as harmonic component powers, fundamental power factor, and the current and voltage harmonics at and below the $15^{th}$ harmonic. Instead of using the two cycle difference used in [14], or a consecutive cycle difference as used in [22], this study analyses multiple cycle difference starting from 3 cycle to 60 cycle, which

is one complete power frequency cycle. After calculating the 'X' cycle difference of the current signal, where 'X' indicates the number of cycles of difference for the waveform; the segment RMS value of the differential waveform gets determined, as described in the Section 2.3.1.

Moreover, instead of using the threshold technique to detect disturbances, this method first obtains a detection signal by using a moving window over 3 cycles for the 'X' cycle differential waveform as well as using a 3 cycle moving window for calculating the frequency domain components of the signal. Then, this study uses an Automatic Multi-Scale based Peak Detection algorithm (discussed in Section 4.1.1) in order to detect the peaks which specify the disturbance and has encapsulated the events, in the sense of deriving the starting time and ending time of the events.

The procedure employed in the majority of peak detection algorithms is as follows, the original signal is pre-processed to improve the performance of the subsequent stages, and then a detection function is derived at a lower sampling rate, to which a peak-picking algorithm is applied to locate the peaks [1]. In order to detect the events, this work has transformed the input waveform into a detection function which provides the positions of the events in the time domain. This is described in Chapter 3 and Chapter 4.

## 2.3 DETAILS OF ALGORITHM

In order to detect a Waveform abnormality in the current waveform, a modified waveform detection method is implemented. In this method, the differential current

signal between '$n_c$' consecutive cycles is calculated. The segment RMS value is calculated by segmenting the resulting waveform as described in the following steps.

**Step 1**: The original signal and the signal after moving '$n_s$' sample points are superimposed to eliminate normal load component.

where,

$$n_s = \left(\frac{f_s}{f_0}\right) * n_c \tag{2.8}$$

and,

$n_s$: Number of sample points of difference

$f_s$: Sampling frequency value of the input signal

$f_0$: Fundamental frequency value of the input signal

$n_c$: Number of cycles of difference

**Step 2**: The resulting signal is divided into M segments.

**Step 3**: The RMS value for each segment of the signal is calculated to obtain the detection function.

**Step 4**: An event is detected using the Modified Multi-Scale Based Peak detection method as explained in Section 4.1.1 on the detection function obtained in Step 3.

The use of '$n_c$ ' cycles is empirical based on it's performance of smoothing the signal as well as retaining enough information to detect events at a later stage.

### 2.3.1 EQUATIONS

In order to calculate the detection function by calculating the differential current signal between '$n_c$' consecutive cycles, the original signal is being superimposed by the same signal, shifted by $n_s$ samples, as described below:

$$I_{\mathrm{D}(k)} = I_{(k)} - I_{(k-n_s)} \tag{2.9}$$

where,

I : Instantaneous current values for phase A, B and C

$I_D$ : Instantaneous differential current values for phase A, B and C

k: Sample index

$n_s$: Number of sample points of difference

The above method is applied to the mains as well as to each of the 3 phase output signals from the 5 feeders. The segment RMS value of the differential current signal is calculated as shown in 2.10 and 2.11, in order to determine the precise locations easily.

$$V_{RMS}\ \ (k) = \sqrt{\frac{\sum_{n=k-n_{rms}+1}^{k} V_D\,(k)^2}{(n_{rms})}},\ \forall\ k \in [1,\ length(V_D)] \tag{2.10}$$

$$I_{RMS}\ \ (k) = \sqrt{\frac{\sum_{n=k-n_{rms}+1}^{k} I_D\,(k)^2}{(n_{rms})}},\ \forall\ k \in [1,\ length(I_D)] \tag{2.11}$$

$$n_{rms} = \frac{n_s}{M} \tag{2.12}$$

where,

$V_D$: Instantaneous differential voltage values for phase A, B and C

$V_{RMS}$ : Segment RMS value of voltage signals for phase A, B and C

$I_D$: Instantaneous Differential current values for phase A, B and C

$I_{RMS}$ : Segment RMS value content of current signals for phase A, B and C

k: Sample index

$n_s$: Number of sample points of difference

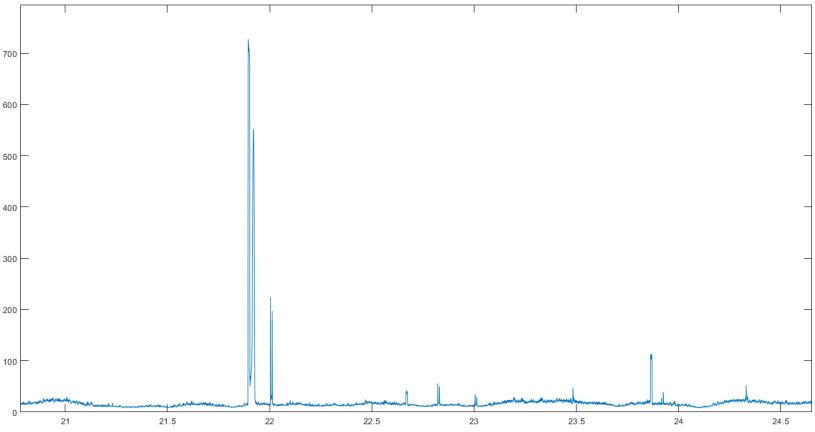
M : The number of segments, the differential signal is filtered over, per $n_c$ number of cycles

$n_{rms}$: The number of sample points in each segment of the signal

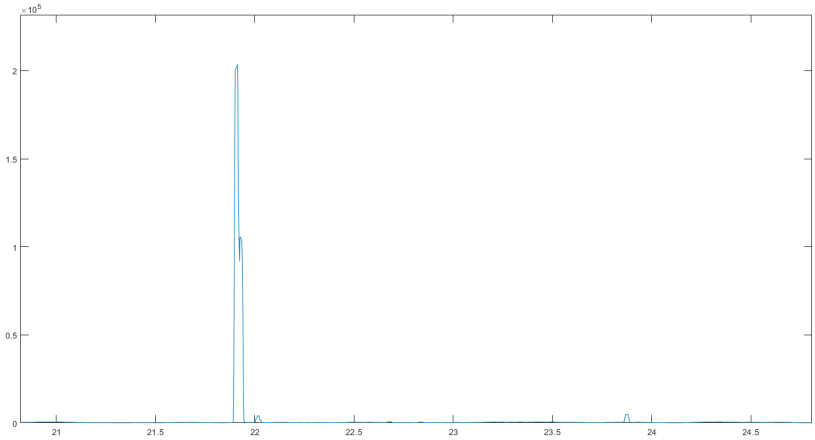
In order to calculate the segment RMS value of the signal, as performed above, the superimposed and shifted signal is squared. The resulting signal is then divided into M segments. Then, the mean value of each segment is calculated over $n_{rms}$ number of sample points in each segment of the signal. Lastly the square root of the signal hence obtained, providing the segment RMS value of the signal. It is repeated for the complete length of the signal for all the sample points.

The plot showing the various steps of the segment RMS value calculation is shown in Figures 2.2a, 2.2b and 2.2c.

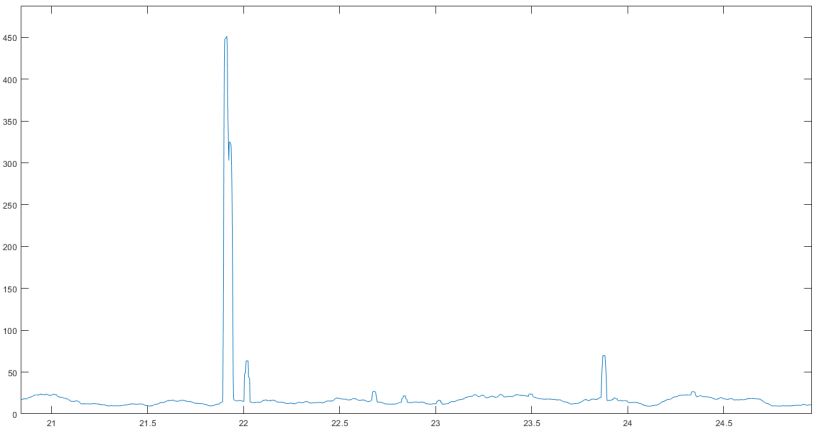
Comparing Figures 2.2c and 2.2b with the differential current signal shown in Figure 2.2a, it is evident that peaks are smooth and distinguishable from non peaks.




(a) Differential current waveform in Amps




(b) Mean value of squared differential current waveform in square Amps




(c) Detection function waveform in Amps

Figure 2.2: Plots for waveform lasting for a period from 21.0 min to 24.5 min. (a) Plot for differential current waveform in Amperes. (b) Plot for the mean value of squared differential current waveform in squared Amperes. (c) Plot for detection function waveform in Amperes.

### 2.3.2 EXAMPLES

This section demonstrates the concepts expressed above using data for a duration of one day. The value of '$n_c$' is taken to be '30' and, the value of 'M' is chosen to be 100, for this study. For the purpose of illustration, Figure 2.3 shows the superposition of the original signal with another signal after moving it by $n_c = 3$ cycles.

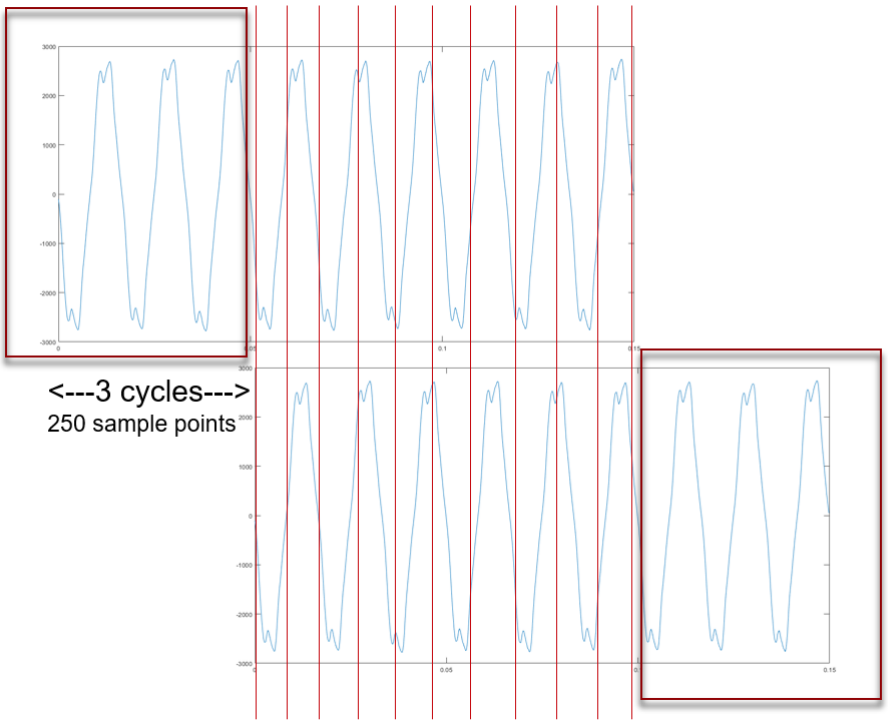
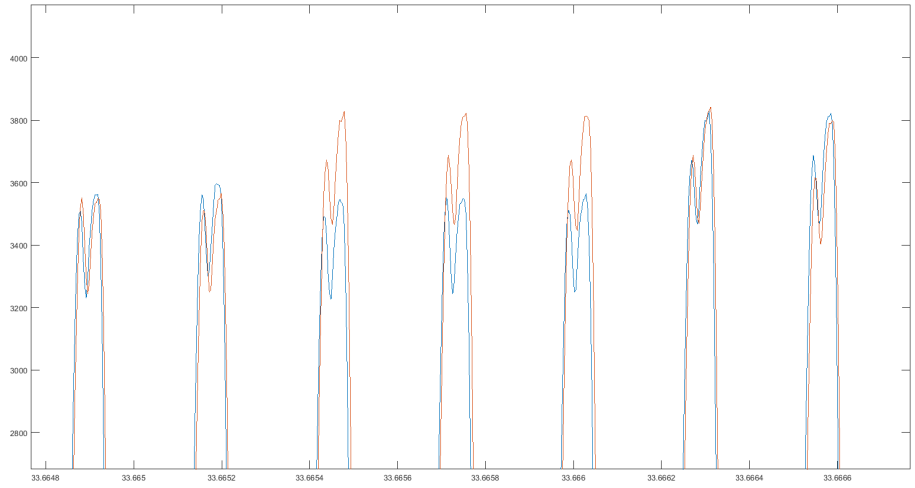
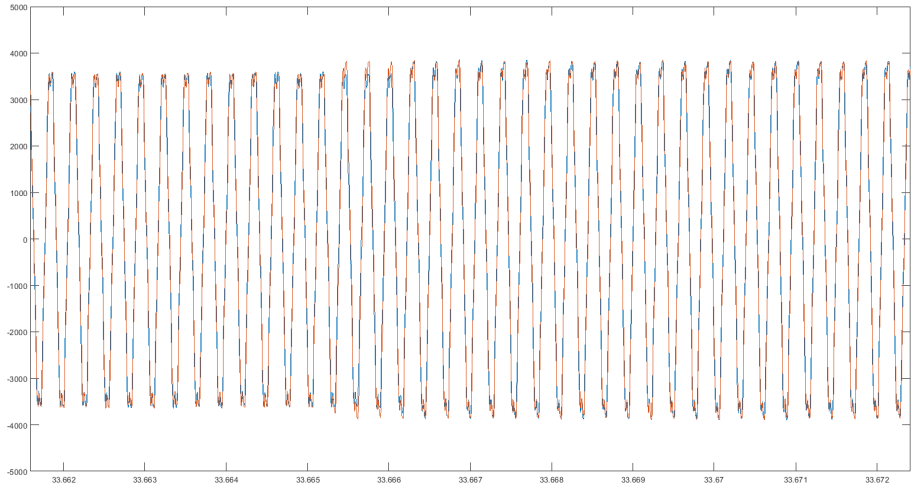



Figure 2.3: Illustration of the method, showing the original and shifted waveforms ($n_c = 3$ cycles).

The method explained above has been implemented on a 1 sec portion of the signal data set. The resulting waveform is shown in Figures 2.4a and 2.4b.




(a)

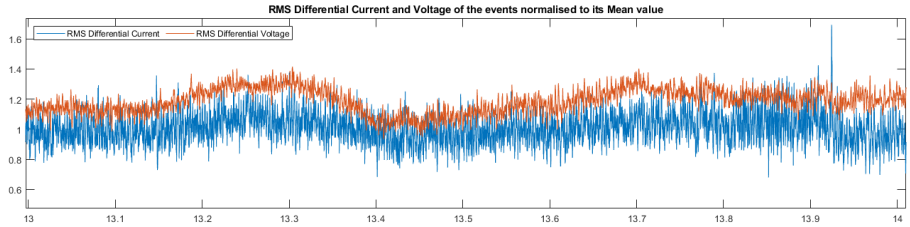



(b)

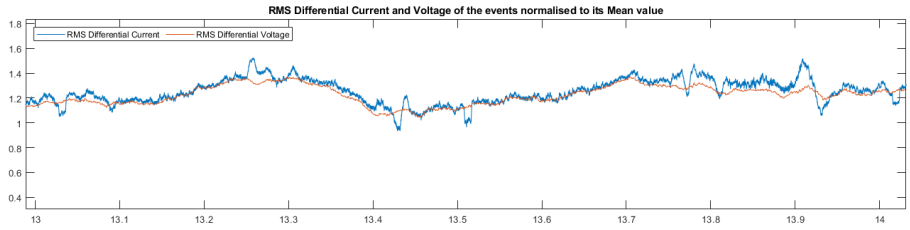
Figure 2.4: Plots for events. (a) Plot for an event in the current waveform in amperes lasting for a period of 33.6654 min to 33.6662 min. (b) Plot of current waveform in amperes against time in minutes, in presence of an event for the period of 33.662 min to 33.672 min.

In Figure 2.4a and Figure 2.4b, the blue waveform indicates the current signal at time 't' and red waveform on the background represents the waveform after t+(3 cycle). As is evident from Figure 2.4a and Figure 2.4b, the method isolates the events and the transient value detects the disturbance. In Figure 2.4a, after the transient ends at time of 33.6662 min, the steady state is regained.

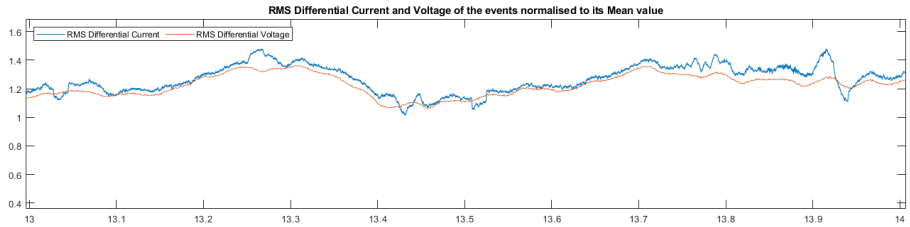
Using the real data from the mains circuit, the impact of the value of '$n_c$' is examined over a range of multiple cycles.




(a)



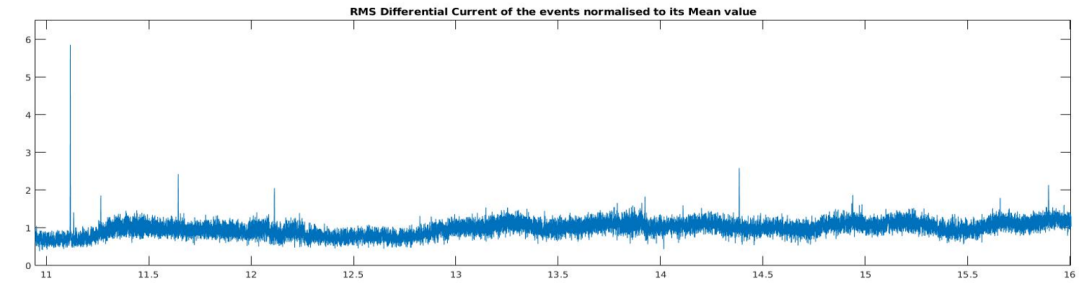

(b)



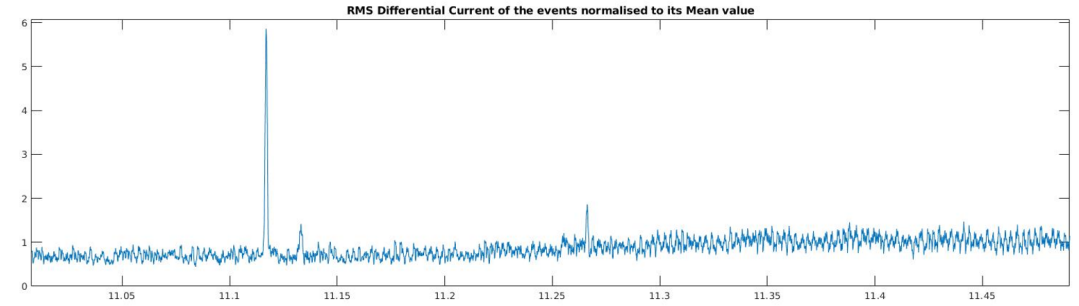

(c)

Figure 2.5: Comparison of cycle difference for the differential current waveform over time duration of 13 min to 14 min. (a) The top waveform represents the segment RMS value of the 3 cycle differential current waveform. (b) As for (a) but cycle difference over 30 cycles. (c) As for (a) but cycle difference over 60 cycles.

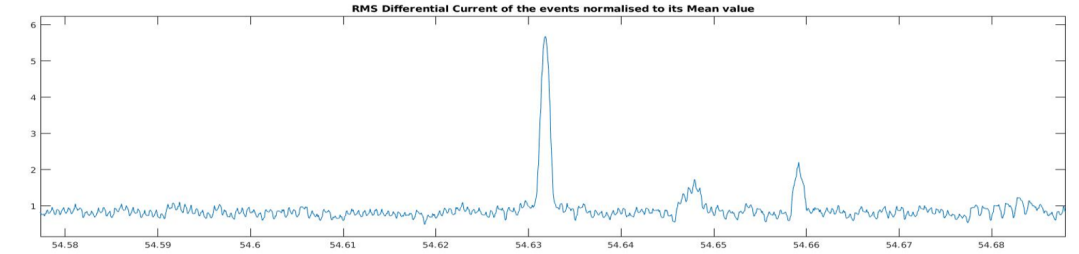
In Figure 2.5a, the 3 cycle differential waveform represents a noisy signal, and in Figure 2.5c the 60 cycle difference waveform is very smooth. Neither of these may be appropriate to be considered as a reference to detect events. But, the waveform shown in Figure 2.5b, represents an adequate combination of smoothed curve with spikes at the time of events.



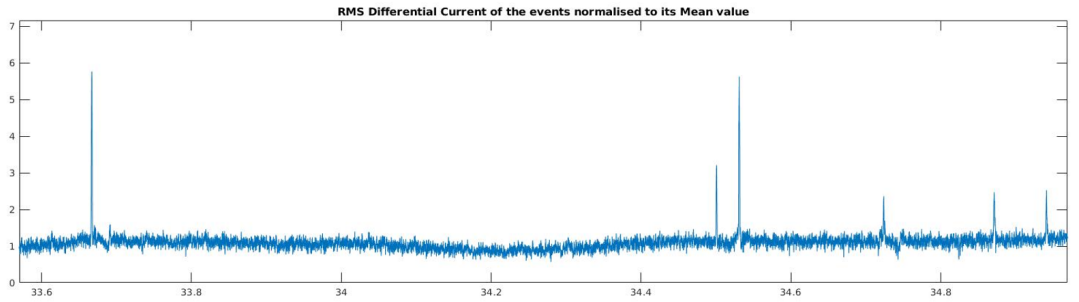

(a)




(b)




(c)




(d)

Figure 2.6: Segment RMS value of differential current waveform normalized to its mean value against time in minutes for multiple instances of events. (a) The plot illustrates the waveform from 11 min to 16 min. (b) The plot shows a narrower range of 11 min to 11.5 min. (c) As (a) but over different range of 54.58 min to 54.68 min and, (d) As (a) but over different range of 33.6 min to 35 min.

The reason for this can be attributed to the observation that, the 3 cycle difference captures only fast transients, whereas the 30 cycle difference waveform is capable of capturing both fast and slow transients. Hence, 30 Cycle waveform difference is chosen here for better event detection techniques. Moreover, the value of 'M' has been considered as 100 in this study. Hence, each segment comprises of 25 sample points

In Figure 2.6, it is easy to distinguish peaks visually. An important task is to program an algorithm to detect peaks automatically. This is not a trivial task as a simple threshold technique will fail. The topic of peak detection is addressed in the next chapter.

# Chapter 3

# FEATURE SELECTION

## 3.1 COMPARISON OF FEATURE SELECTION METHODS

In the literature, features for classification in power systems rely on observed power measurements. This is potentially limiting with more types of data and tools at hand to apply to the problem. For example with Machine Learning approaches unsupervised methods can find patterns for classification using non-traditional data sets and features.

### 3.1.1 FEATURE SELECTION

The purpose of this chapter is to define the features for the events detected, so that each event with 'n' number of unique features can be associated with an 'n' dimensional feature subspace. It becomes then easier to compare the events in terms of distance in the particular subspace and do further event analysis.

The selection of features is done based on the following factors:

**1**: Transient behaviors are primarily distinct.

**2**: Transient features do not overlap with steady state signatures

**3**: Avoid a major limitation of transient state features requiring a high sampling rate requirement in order to capture the transients

**4**: Allows Non-Traditional features to be used with the model, such as On and Off duration distribution and Frequency of Appliance Usage

Taking into consideration all the advantage and disadvantages of the features as discussed in Table 3.1, a set of potential features are shown in Figure 3.1. In this case, both the steady state and transient characteristics are considered as features.

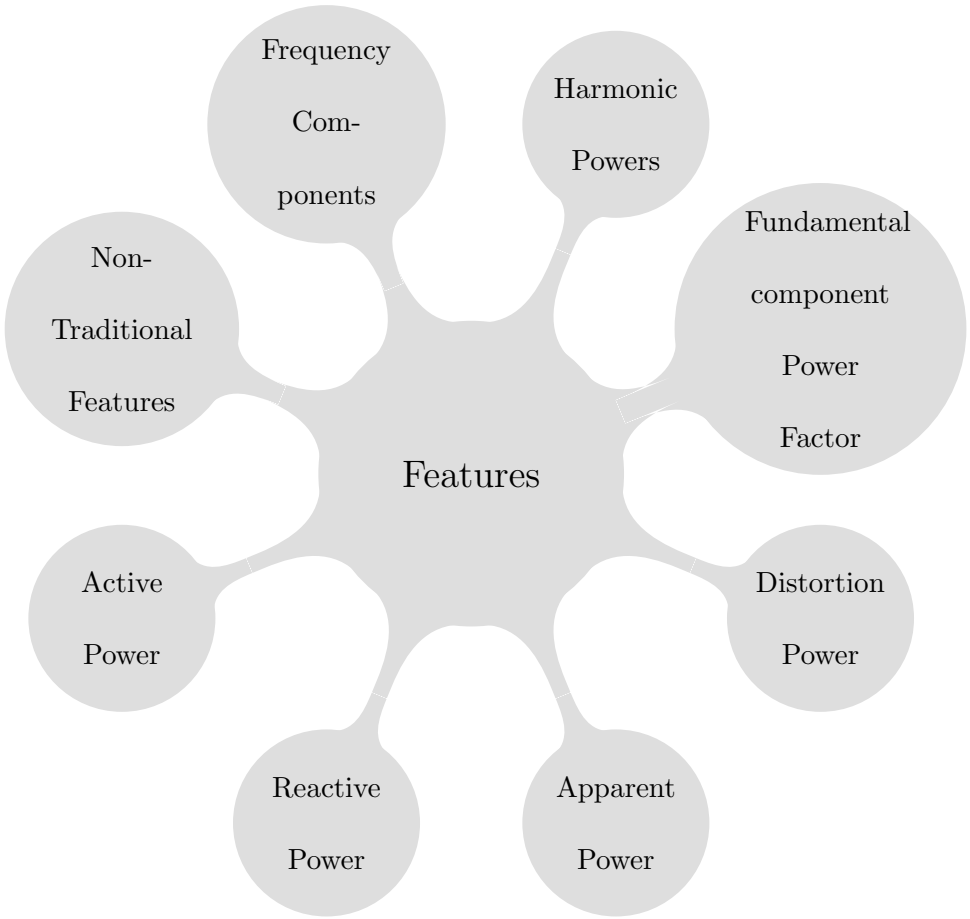
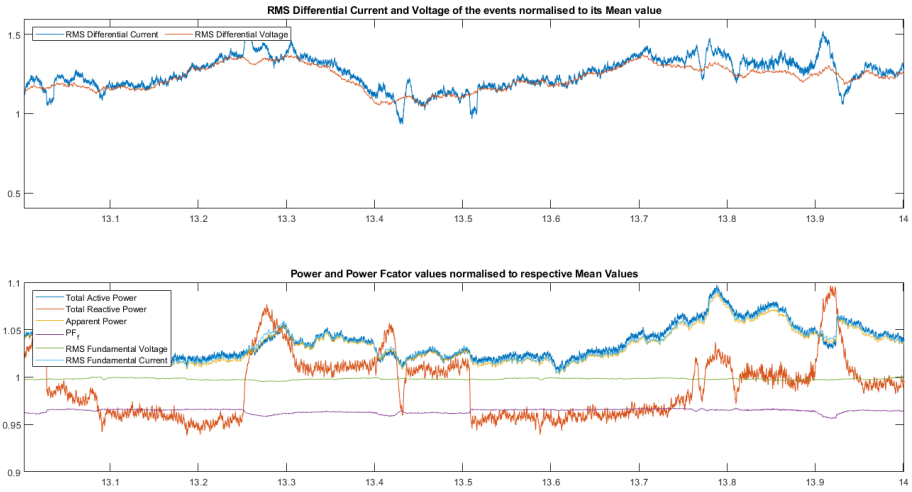
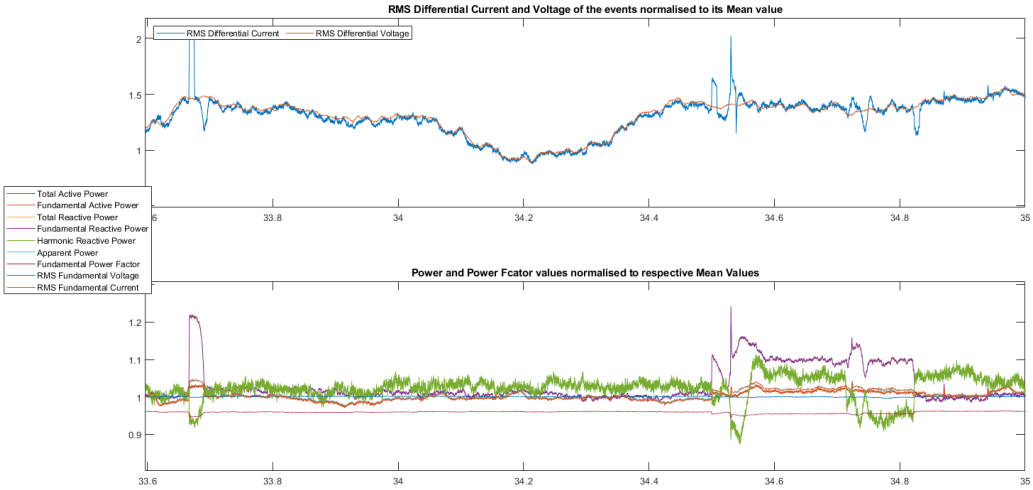



Figure 3.1: Useful features for classification

Figures 3.2a and Figure 3.2b illustrate that the features can be used as load signatures in the presence of an event.




(a)




(b)

Figure 3.2: Features representing two instances of events. (a) Plot for the events during the period of 13 min to 14 min. The top plot illustrates segment RMS value of differential current as the detection signal. The bottom plot shows transient features of a particular event. (b) As (a) but over a narrower range of 33.6 min to 35 min.

The primary reason for using Transient features is that they are primarily distinct compared to steady state signatures. Table 3.1 presents some of the advantages and disadvantages of transient feature selection.

| Transient Methods | Features | Advantages | Shortcomings |
|---|---|---|---|
| Transient Components | i. Power ii. Power Factor | Same Power draw characteristics can be easily differentiated | i. Continuous monitoring ii. High sampling rate Requirement |
| Start up current transient component | i. Current spikes, ii. Event duration iii. Event shape iv. Transient response Time | Distinct transient behavior in multiple load operation Scenario | Poor detection of simultaneous activation and deactivation of components |
| Transient Frequency Components | Frequency Analysis of transient event | Distinguish event behavior in multiple load operation scenario | i Variable definition of an Event ii.Expensive computation |

Table 3.1: Features, advantages and disadvantages of transient feature selection

The following section discusses the calculations used for obtaining the features discussed above. The analysis on steady state features made on the sample points using k-means technique, is discussed in Section 3.5.2. Further, in Chapter 5, the clustering of event sets using unsupervised Machine Learning techniques as well as

by training models using a Neural Network is presented.

## 3.2 DETAILS OF COMPUTATIONS

This section explains the calculations to obtain the multiple features. For the purpose of calculating the powers and power factor, IEEE 1459-2010 standard is used [10].

### 3.2.1 HARMONIC POWER CALCULATIONS

The voltage and current at any moment of time 't' can be defined as follows:

$$V(t) = \sqrt{2}\, V \cos\left(2\pi f\left(t_V + \frac{\theta}{2\pi f}\right)\right) \tag{3.1}$$

$$I_h(t) = \sqrt{2}\, I_h \cos\left(2\pi h f\left(t_{I(h)} + \frac{\Phi}{2\pi h f}\right)\right) \tag{3.2}$$

where,

$V$ : RMS value for power system frequency voltage (V)

$I_h$ : RMS value for current of harmonic order h (A)

$\theta$ : Phase angle of the phasor $V$ (rad)

$\Phi$ : Phase angle of power system frequency component of the phasor $I_h$ (rad)

$\beta_h$ : Phase angle difference between the phasors $V$ and $I_h$ (rad)

$t_{I(h)}$ : Time reference of zero crossing for phasor $I_h$ (s)

$t_V$ : Time reference of zero crossing for phasor $V$ (s)

f: Power system frequency (Hz)

h : Harmonic order

**Calculation of Phase Difference**

$$\begin{aligned}\beta_h &= [t_{I\ (h)} - \ t_V]\ (sec) \\ &= \left[\left(-\frac{\Phi}{2\pi hf}\right) - \left(-\frac{\theta}{2\pi f}\right)\right]\ (sec) \\ &= \left[\frac{(\theta - \frac{\Phi}{h})}{2\pi f}\right] 2\pi fh\ (rad) \\ &= [\theta\ *\ h\ - \Phi]\ (rad)\end{aligned} \tag{3.3}$$

where,

$t_{I(h)}$ : Time reference of zero crossing for phasor $I_h$ (s)

$t_V$ : Time reference of zero crossing for phasor $V$ (s)

In this work, we calculate the angle components and magnitudes for the first 15 harmonics and these are used as features.

**Instantaneous Power (W)**

According to the IEEE 1459-2010 standard [5], for steady-state conditions, a non sinusoidal periodical instantaneous power has two distinct components: the power system frequency components and the remaining terms.

$$p = p_a + p_q \tag{3.4}$$

where,

p : Instantaneous power (W)

$p_a$ : Instantaneous active power (W)

$p_q$ : Instantaneous reactive power (VAR)

**Active Power (W)**

In Equation (3.4), the component $p_a$ is the instantaneous active power, which is equal to the sum of harmonic active powers. It is produced by the active component of the current i.e., the harmonic active power of order h is caused by the harmonic voltage of order h and the component of the harmonic current of order h in phase with the harmonic voltage of order h.

$$p_a = V_0 I_0 \ + \ \sum_h V_h I_h \ cos \ \beta_h \ [1 + cos(2h\omega t + 2\Phi_h) \ ] \tag{3.5}$$

where,

$p_a$ : Instantaneous active power

$V_0$ : Direct voltage term

$I_0$ : Direct current term

$V_h$ : RMS value for voltage of harmonic order h (V)

$I_h$ : RMS value for current of harmonic order h (A)

$\beta_h$ : Phase angle difference between the phasors $V_h$ and $I_h$ (rad)

$\Phi_h$ : Phase angle of the phasor $I_h$ (rad)

$\omega$ : Angular frequency $2\pi$f (rad/s)

h : Harmonic Order

Each instantaneous active power of order h has two terms:

**i.** Active, or real, harmonic power, and

**ii.** Intrinsic harmonic power

The intrinsic power is always present when net energy is transferred to the load. This oscillating component does not contribute to net transfer of energy or to additional power loss in conductors.

The active power P, which is also called real power, is the average value of the instantaneous power during the measurement time interval $\tau$ to $\tau$ + kT, and it can be written as:

$$P\left(k\right) = \sum_{h=1}^{M} V_{\,h}\left(k\right)\ I_{\,h}\left(k\right) \cos\left(\beta_h\left(k\right)\right) \tag{3.6}$$

where,

T = $\frac{1}{f}$ : The cycle time (s)

k : A positive integer number

$\tau$ : The moment when the measurement starts

The above equation can be rewritten as:

$$P(k) = P_1(k) + P_H(k) \tag{3.7}$$

The components $P_1$ and $P_H$ are defined as follows:

$$P_1(k) = V_1(k)\ I_1(k)\cos(\beta_1(k)) \tag{3.8}$$

and,

$$P_H(k) = \sum_{h\neq 1} V_h(k)\ I_h(k)\cos(\beta_h(k)) \tag{3.9}$$

where,

P : Total average power or active power (W)

$P_1$: Power system frequency active power (W)

$P_H$ : Harmonic active power (W)

$V_h$ : RMS value for voltage of harmonic order h (V)

$I_h$ : RMS value for current of harmonic order h (A)

$\beta_h$ : Phase angle difference between the phasors $V$ and $I_h$ (rad)

k: Sample index

h : Harmonic order

M : Highest Order of harmonic

**Reactive Power (VAR)**

In Equation (3.4), the component $p_q$ is the instantaneous reactive power. It is produced by the reactive component of the current (i.e., the component that is in quadrature with the voltage), which is mathematically defined as follows:

$$
\begin{aligned}
p_q = & -\sum_{h} V_h I_h sin(\beta_h) sin(2h\omega t + 2\Phi_h) \\
& + 2 \sum_{\substack{p \\ p \neq q}} \sum_{q} V_q I_p cos(q\omega t + \theta_q)\ cos(p\omega t + \Phi_p) \\
& + \sqrt{2}\ V_0 \sum_{h} I_h cos(h\omega t + \Phi_h) \\
& + \sqrt{2}\ I_0 \sum_{h} V_h cos(h\omega t + \theta_h)
\end{aligned}
\tag{3.10}
$$

where,

$p_q$ : Instantaneous reactive power (VAR)

$V_h$ : RMS value harmonic voltage of order h (V)

$I_h$ : RMS value harmonic current of order h (A)

$\beta_h$ : Phase angle difference between the phasors $V$ and $I_h$ (rad)

$\theta_h$ : Phase angle of the phasor $V_h$ (rad)

$\Phi_h$ : Phase angle of the phasor $I_h$ (rad)

$\omega$ : Angular frequency $2\pi$f (rad/s)

h : Harmonic order

This energy component oscillates between the sources and the electromagnetic energy stored within the magnetic field of the inductors and electric field of the capacitors of electrical equipment. The average value of this rate of flow is zero, and the net transfer of energy to the load is nil; but, these power oscillations do lead indirectly to power loss (Joule and eddy-current) in the conductors.

The magnitude of the reactive power Q equals the amplitude of the oscillating

instantaneous reactive power $p_q$. According to Budeanu [3], the reactive power Q is given by:

$$Q\,(k) = \sum_{h=1}^{M} (V_{\,h}\,(k)\;\; I_{\,h}\,(k)\;\; sin(\beta_h\,(k)) \tag{3.11}$$

The above equation can be rewritten as:

$$Q(k) = Q_1(k) + Q_H(k) \tag{3.12}$$

The components $Q_1$ and $Q_H$ are defined as follows:

$$Q_1\,(k) = \;\; V_{RMS\;1}\,(k)\;\; I_{RMS\;1}\,(k) \sin\left(\beta_1\,(k)\right) \tag{3.13}$$

and,

$$Q_H\,(k) = \sum_{h\neq 1} \left(V_{RMS\;h}\,(k)\;\; I_{RMS\;h}\,(k) \sin\left(\beta_h\,(k)\right)\right. \tag{3.14}$$

where,

Q : Total Reactive power (VAR)

$Q_1$ : Power system frequency reactive power (VAR)

$Q_H$ : Harmonic reactive power (VAR)

$\beta_h$ : Phase angle difference between the phasors $V$ and $I_h$ (rad)

k: Sample index

h : Harmonic order

M : Highest order of harmonic

**Apparent Power**

The apparent power 'S' is the amount of active power that can be supplied to a load, or a cluster of loads, under ideal conditions. The ideal conditions may assume the load supplied with sinusoidal voltage and current. The loads are compensated by means of active or passive devices such that the line current is sinusoidal and in phase with the voltage that, ideally, is also adjusted to be sinusoidal.

Apparent Power is the product of the root mean square (RMS) voltage and the RMS current defined as follows:

$$S(k) = V(k) * I(k) \tag{3.15}$$

**Distortion Power**

The relationship of Active Power and Reactive Power with Apparent Power based on Equations (3.6), (3.11) and (3.15), is defined by the following inequality [4]:

$$S^2 \geq P^2 + Q^2 \tag{3.16}$$

The difference is due to the harmonic content of the signals. In 1927, Budeanu introduced the term distortion power [3], which is defined in terms of Active Power and Reactive Power as follows:

$$S(k)^2 = P(k)^2 + Q(k)^2 + D(k)^2 \tag{3.17}$$

where D is the distortion power. The distortion power can be expressed as

following:

$$
\begin{aligned}
D^2 &= S^2 - P^2 - Q^2 \\
&= \left[\sqrt{\sum_{h=1}^{M} V_h^2}\ \sqrt{\sum_{h=1}^{M} I_h^2}\right] \quad = \sum_{\substack{p \\ p\neq q}} \sum_{q} [V_q I_p [V_q I_p - 2\ V_p I_q\ cos\beta_q\ - \beta_p]\ ]
\end{aligned}
\tag{3.18}
$$

**Power Factor**

$$
\begin{aligned}
S^2 &= (VI)^2 \\
&= ({V_1}^2\ +\ {V_H}^2)({I_1}^2\ +\ {I_H}^2) \\
&= (V_1 I_1)^2\ +\ (V_1 I_H)^2\ + (V_H I_1)^2\ +\ (V_H I_H)^2 \\
&= (S_1)^2 + (D_I)^2 + (D_V)^2 + (S_H)^2 \\
&= (S_1)^2 + (S_N)^2
\end{aligned}
\tag{3.19}
$$

where,

$$
\begin{aligned}
S_1 &= (V_1\ I_1) \\
D_I &= (V_1\ I_H) \\
D_V &= (V_H\ I_1) \\
S_H &= (V_H\ I_H)
\end{aligned}
\tag{3.20}
$$

and,

S : Apparent Power (VA)

$S_1$ : Power system frequency apparent power (VA)

$S_H$ : Harmonic apparent power (VA)

$V_1$ : RMS value of power system frequency voltage (V)

$V_H$ : RMS value of non-power system frequency voltage (V)

$I_1$ : RMS values of power system frequency current (A)

$I_H$ : RMS values of non-power system frequency current (A)

$D_I$ : Current distortion power (VAR)

$D_V$ : Voltage distortion power (VAR)

Since, the harmonics in voltage is negligible, from the Equation (3.19), further reduction leads to following:

$$\begin{aligned}(S_N)^2 &= (D_I)^2 + (D_V)^2 + (S_H)^2 \\ &= (V_1\ I_H) \\ &= [(S_1)\ (THD_I)]^2 \\ [\frac{S_N}{S_1}]^2 &= (THD_I)^2\end{aligned} \tag{3.21}$$

Now the Power factor is defined as follows:

$$\begin{aligned}PF &= \frac{P}{S} \\ &= \frac{(P_1 + P_H)}{\sqrt{(S_1^2 + S_N^2)}} \\ &= \frac{\frac{P_1}{S_1}\ [1 + \frac{(P_1+P_H)}{\sqrt{(S_1^2+S_N^2)}}]}{\sqrt{1 + (\frac{S_N}{S_1})^2)}}\end{aligned} \tag{3.22}$$

$$\begin{aligned}PF &= \frac{P}{S} \\ &= \frac{(PF_1)}{\sqrt{1 + THD_I^2}}\end{aligned} \tag{3.23}$$

Where,:

$PF_1 = \frac{P_1}{S_1}$ : Power system frequency component of Power factor

$1 + \frac{P_H}{P_1} = 1$ : Assuming that Harmonic Active Power is negligible as compared

to Power system frequency Active Power $\frac{S_N}{S_1} = THD_I$: Total Harmonic Distortion of Current Signal

## 3.3 FREQUENCY ANALYSIS OF SIGNAL

Harmonic analysis of the signal provides an important characteristics for detecting and classifying electrical events [11]. Hence, the frequency components of the current signal are calculated from the $2^{nd}$ order harmonic to the $15^{th}$ order harmonic.

First, the original current signal is analyzed in the frequency domain using the Fast Fourier transform. The total harmonic distortion to the $15^{th}$ order harmonic is calculated. As explained in [15], the ratio of the power of the fundamental component to the power of all non-harmonic components, is a useful feature for each event comparing the original current signal with the multiple frequency components of the signal.

The frequency characteristics will later be used in Section 3.4, for the purpose of feature selection in order to classify the events.

### 3.3.1 CHANGES IN ACTIVE POWER, REACTIVE POWER, & POWER FACTOR OF THE FUNDAMENTAL COMPONENT OF THE WAVEFORM

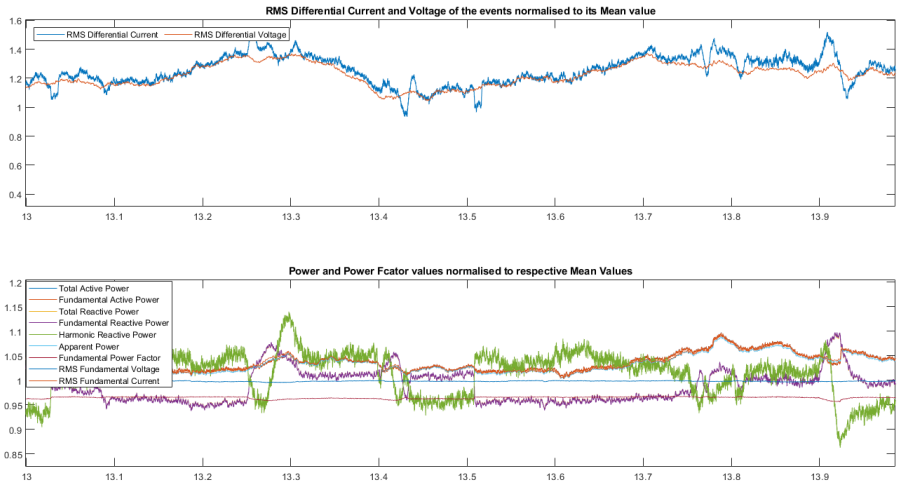



Figure 3.3: Illustration of the impact of features at the location of peaks in Detection function signal. The top Figure plots the detection function against time in minutes for a period of 13 minutes to 14 minutes; The bottom figure plots the features.

Figure 3.3 illustrates the plots of segments of the RMS value of the differential current along with multiple features at the instances of disturbances. The feature changes at the instances corroborate the events. This provides an opportunity to find the correlations between the features for characterizing the events, and then cluster them together into groups as will be discussed in Chapter 5.

## 3.4 METHOD OF FEATURE SELECTION

Selecting features which make greater distinction between events will help the clustering technique to perform better. This work has attempted to perform feature selection analysis on the set of features discussed above. For this purpose, a supervised machine learning model is used. For supervision, 25% of the sample points in a smaller portion of the data, taken out solely for the purpose of training and testing, is manually marked for events. Once a quarter of the sample points in the data is distinguished by events and non-events sample points, multiple set of features are used to train the classification models. In order to train the models based on the data selected, multiple features are used to identify the set of features that will give the highest performance with most of the models.

A 7-fold cross validation is used to reduce over fitting. A k-fold cross validation partitions the data into k disjoint sets. An operation performed on each set provides a performance metric in terms of the predictive accuracy of the trained model. For each set of data points, a model is trained using the data points in all of the sets except the selected set and the performance of the model is evaluated based on all the data points of the selected set.

The average error over all the sets is calculated and recorded for the purpose of comparison. The performances of the multiple models are analyzed and presented in the Table 3.2. It illustrates the performance of the models in terms of error while predicting the classification of the sample points that belong to one of either event or non-event classes.

| Sl. No. | Features | Best Classifier | Accuracy |
|---|---|---|---|
| 1 | a + b + c + d | Fine $KNN_{10}$ | 99.8 |
| 2 | a + b + c | Fine $KNN_{10}$ | 99.2 |
| 3 | a + b | Fine $KNN_{10}$ | 98.4 |
| 4 | a | Weighted $KNN_{100}$ | 98.9 |

Table 3.2: Model Performance comparison for feature selection

where,

**a.**

Feature 1: Differential current

Feature 2: Fundamental active power

Feature 3: Fundamental reactive power

Feature 4: Harmonic reactive power

Feature 5: Fundamental power factor

**b.**

Features 6-20: Current harmonics magnitude from the $2^{nd}$ order to the $15^{th}$ order and fundamental component magnitude.

**c.**

Feature 21-35: Current harmonics phase from the $2^{nd}$ order to the $15^{th}$ order and fundamental component phase.

**d.**

Feature 36-66: Voltage harmonics magnitude and phase from the $2^{nd}$ order to the $15^{th}$ order and fundamental component magnitude and phase.

**Trained Model**$_K$ indicates the model is trained with 'k' number of nearest neighbor points to classify each point for the purpose of prediction.

Through empirical considerations as shown in table 3.2, the following set of features are selected for analyzing the events in following sections.

**Feature 1.** Segment RMS value of Differential Current

**Feature 2.** Fundamental Active Power

**Feature 3.** Fundamental reactive power

**Feature 4.** Harmonic Reactive Power

**Feature 5.** Fundamental Power Factor

**Feature 6-20.** Current harmonics magnitude from the $2^{nd}$ order to the $15^{th}$ order and fundamental component magnitude.

**Feature 21-35.** Current harmonics phase from the $2^{nd}$ order to the $15^{th}$ order and fundamental component phase.

**Feature 36-66.** Voltage harmonics magnitude and phase from the $2^{nd}$ order to the $15^{th}$ order and fundamental component magnitude and phase.

## 3.5 K-MEANS CLUSTERING

After the preliminary task of feature selection is completed, are clustered, as discussed in detail in this section.

### 3.5.1 DEFINING THE NUMBER OF CLUSTERS

In a clustering method, the first challenge is determining the number of clusters. Generally, the supervisor provides the number of clusters, to which the Clustering technique finds the suitable fit for each input sample point or sample event.

In a clustering method, the sum of distances from each cluster centroid to the sample points within that cluster, for all the clusters, reduces as the number of cluster increases. This can be attributed partly to the fact that, as the number of clusters increase, fewer sample points are associated with each cluster. Ultimately the sum becomes zero when the total number of sample points equals the total number of clusters, as illustrated in Figure 2.4b. This can alternatively be visualized in terms of a trade off between the expected performance benefit, and the number of clusters.

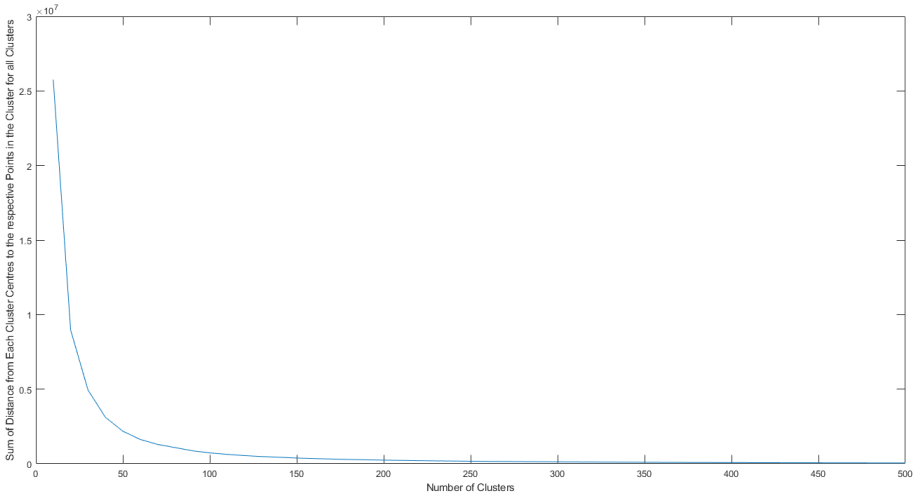



Figure 3.4: Plot for sum of distances from each cluster centroid to the sample points within that cluster, against the number of clusters

In order to define the number of clusters among the input data points, a method has been implemented based on a knee detection approach [19]. So, the curve on

which the analysis of point has been made is the curve that depicts the sum of distances from each cluster centroid to the sample points within that cluster, for all the clusters, against the number of clusters. In order to find the the knee points in the normalized curve, the point is calculated where the curve flattens out. This can be defined as the instance where the rate of increase/decrease of reduces significantly. Mathematically, it is based on calculating the Euclidean distance between each point on the curve and a vector $\vec{V}$ which connects the first and last point on the curve. The Knee Point is obtained by calculating the point on the curve with the maximum distance from the vector connecting the first and last point on the curve, as shown in the Equations (3.24) and (3.25).

$$\vec{d_k} = \vec{A_k} - \vec{A_V} \; \forall \; k \; \in \; 1, \; 2, \; \ldots, \; N \tag{3.24}$$

$$V_{KneePoint} = \; arg \; max \; (d_k) \; \forall \; k \; \in \; 1, \; 2, \; \ldots, \; N \tag{3.25}$$

where,

$\vec{d_k}$ : The perpendicular distance between $\vec{V}$ and the points on the curve

$\vec{V}$ : Vector connecting the first and last point on the curve shown in figure 3.4

$\vec{A_k}$ : Vector connecting the first point and the points on curve

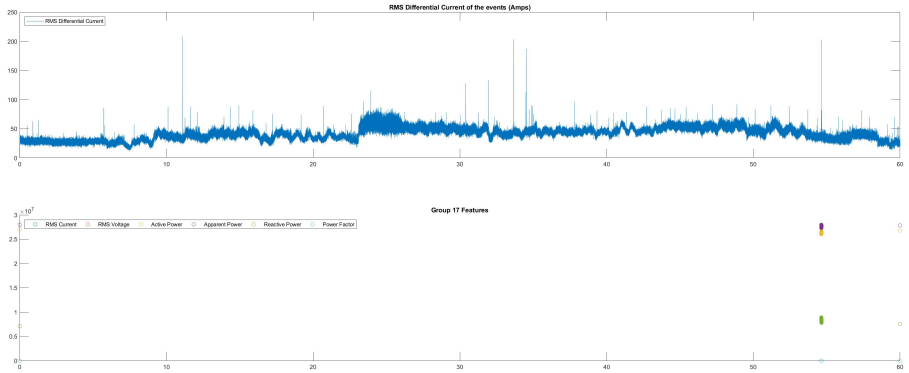
$\vec{A_V}$ : Projection Vector of $\vec{A}$ on $\vec{V}$

$V_{KneePoint}$ : The Knee point value

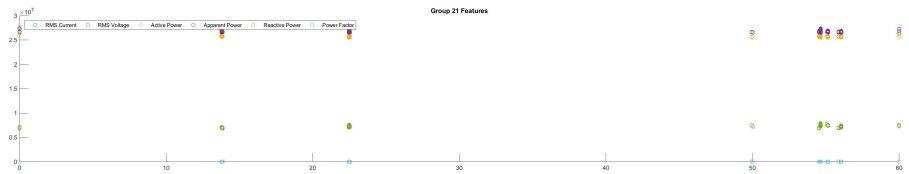
N : Maximum index of iteration in figure 3.4

### 3.5.2 CLUSTERING OF SAMPLE POINTS

After analyzing the performance of the models with multiple trial sets of features, a particular set is selected which provides the highest performance. In order to form the clusters of similar events in the event space, a k-means algorithm is used over the set of sample points which belong to the class of events. The knee/elbow point technique yields 38 groups for this particular data set. The Figure 3.5 illustrates three of these 38 groups.




(a)




(b)


(c)

Figure 3.5: Illustration of k-means clustering on the sample points of events. (a) The top waveform represents the segment RMS value of the 3 cycle differential current waveform and the bottom plot illustrates a selected set of features, for the group number 17. (b) As (a) but for group number 21. (c) As (a) but for group number 36.

## 3.6 DISCUSSION

The clustering analysis made on the sample points of the events using the k-means clustering technique resulted in 38 clusters. The result may not be efficient in terms of clustering, as the signatures in one event may be similar to other events and hence they may be clustered together. The objective is to cluster all the sample points in a particular event in one cluster. However, using k-means algorithm for clustering using sample points results in formation of partial clusters. Partial clusters are formed, when one fraction of sample points from one event form one cluster and another fraction of sample points from the same event form a second cluster.

To prevent the uncertainty in classification of events, each event is assumed to be represented by a multitude of Statistical measures for every feature set as discussed in Chapter 5. To calculate the features for a particular event, the objective is to identify the start and end of a particular event. The method to identify the peaks and the events has been discussed in Chapter 4.

# Chapter 4

# LOAD ANALYSIS

## 4.1 METHODS OF LOAD ANALYSIS

This research study proposes analysis of features during a transient to identify loads. These include power, power factor and frequency components. This will help in analyzing power demand in a better manner and improved performance of load operation can also be achieved.

In order to achieve this, the time duration of the event needs to be calculated. For this purpose the starting location of the event and end of the event needs to be detected. This was achieved by implementing the Automatic Multi-scale based Peak Detection (AMPD) algorithm for detection of peaks in noisy periodic and quasi-periodic signals [20], a method by Scholkmann Boss and Wolf. This is described in the following subsection. Further modifications have been made to the algorithm and a Modified Multi-scale Peak detection (MMPD) Algorithm is introduced in this work. A significant performance improvement is observed.

### 4.1.1 AUTOMATIC MULTI-SCALE BASED PEAK DETECTION ALGORITHM

According to the AMPD algorithm, for a given univariate uniformly sampled signal x = [$x_1$, $x_2$, ..., $x_i$, ..., $x_N$], in order to obtain the periodic or quasi-periodic peaks, the following steps are to be followed:

**Step 1**: The Local Maxima Scalogram matrix (LMS) 'M' of Signal 'x' is calculated

a. Signal x is linearly detrended.

b. Local maxima are determined using a moving window approach for every scale 'k', where k = 1, 2, ..., L | L = $\lceil \frac{N}{2} \rceil$ - 1 and $i = k+2, ..., N-k+1$, according to the following.

$$m_{k,\ i} = \begin{cases} 0, & (x_{i-1} > x_{i-k-1}) \cap (x_{i-1} > x_{i+k-1}) \\ \alpha + RV, & otherwise \end{cases} \tag{4.1}$$

where,

RV := Uniformly distributed random number

$\alpha$ : A constant Factor

**Step 2**: Row wise summation of the LMS matrix M

$$\gamma_k = \sum_{i=1}^{N} m_{\ k,\ i} \ \forall\ k \ \in \ 1,\ 2,\ \ldots,\ L \tag{4.2}$$

a. vector $\gamma$ has the information about local maxima (scale dependent distribution of zeros)

b. The global minimum value of $\gamma$ represents the scale with the most local maxima.

$$\lambda = \ arg\ min\ (\gamma_k) \tag{4.3}$$

**Step 3**: Reshape LMS Matrix 'M' by removing all elements belonging to scales greater than $\lambda$.

**Step 4**: Finally, the peaks are detected.

a. The column wise Standard Deviation $\sigma_i$ is calculated.

b. For all the indices for which the $\sigma_i = 0$ holds true $\forall$ i = 1 to N, there exists a peak.

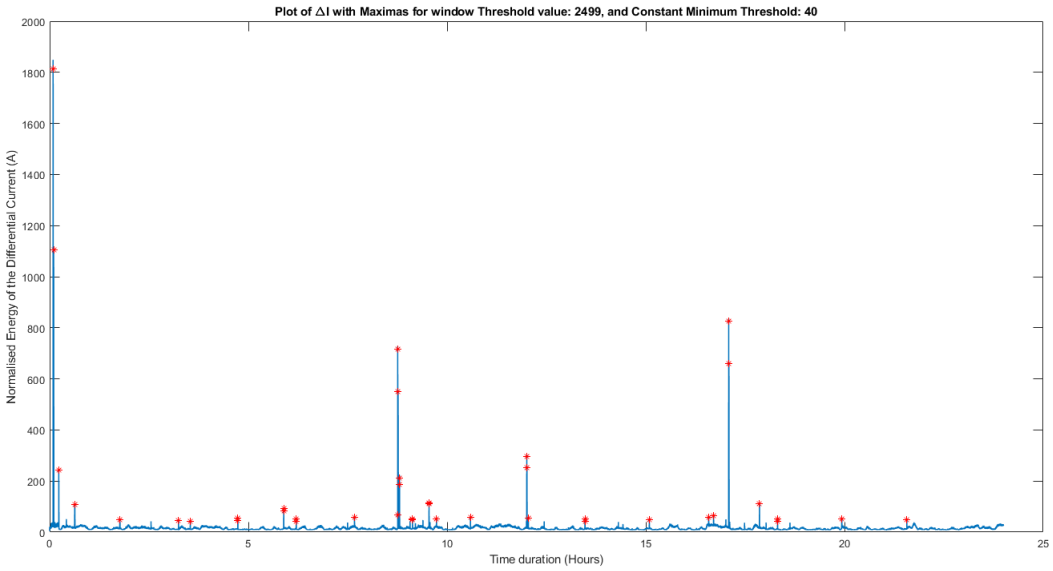



Figure 4.1: Plot for detection of maximas of differential current segment RMS value with a window threshold value of 2499, and a constant minimum threshold of 40 A, for a day

### 4.1.2 MODIFIED MULTI-SCALE BASED PEAK DETECTION ALGORITHM

In the AMPD algorithm, the multi scale operation has been performed in a moving window approach for every scale 'k', where k = 1, 2, ..., L | L = $\lceil \frac{N}{2} \rceil$ - 1, in order to obtain the Local Maxima Scalogram matrix. However, the calculation leads to longer jumps over far fetched values with distant indices. Hence, the performance is slowed down. A Modified Multi-Scale based Peak Detection (MMPD), splits the given univariate uniformly sampled signal x = [$x_1$, $x_2$, ..., $x_i$, ..., $x_N$] into multiple sections based on the window size of '$w_{LMS}$', in order to obtain a smaller sized LMS matrices. As a result the amount of jumps needed to perform the operation for LMS becomes smaller over values having low indices from the base index, improving the speed of MMPD significantly. The comparison in speeds of AMPD against MMPD algorithm is discussed and illustrated in Table 4.1.

The modifications in MMPD are described as followings:

**Modification 1: Windowing**: Before Step 1 operation of AMPD, the given univariate uniformly sampled signal 'x' is divided into multiple sections, with each section having a length equal to the window length of $w_{LMS}$.

**Modification 2: Complete element interaction**: For Step 2 operation of AMPD, in order to calculate the Local Maxima Scalogram matrix 'M' for each segment of the Signal 'x', the relationship between every element is determined for each scale, in order to ensure interaction among all the elements for calculating the scale dependent distribution of zeros, instead of the moving

window approach for every scale 'k' in case of AMPD.

**Modification 3: Reflection along edges**: Since multiscale operation is not possible on samples in the first and last window; reflection of the samples in the first and last window only, is performed along the first and last element. This is done in order to enable interaction between every single element. The reflection is performed by the addition of elements to the beginning and to the end of the vector, which is shown as the following:

$$\begin{aligned} & x_{(i-w_{LMS})} \;=\; x_{(i)} \\ & \forall\; i \;\in\; \{\; 1,\; 2,\; 3,\; \ldots,\; (w_{LMS}-2),\; (w_{LMS}-1),\; w_{LMS}\} \\ & \&, \\ & x_{(j+N-w_{LMS})} \;=\; x_{(j)} \\ & \forall\; j \;\in\; \{\; 1,\; 2,\; 3,\; \ldots,\; (w_{LMS}-2),\; (w_{LMS}-1),\; w_{LMS}\} \end{aligned} \tag{4.4}$$

where,

x: Given univariate uniformly sampled signal

N: Length of the signal 'x'

$w_{LMS}$: Window size along which the operation of MMPD is performed

**Modification 4: Constant value Assignment**: For assigning values to each element of the Local Maxima Scalogram matrix 'M', the Uniformly distributed Random Numbers get replaced with a constant value $\alpha$.

**Modification 5: Column wise Summation replace Standard Deviation calculation**: In Step 4 operation of AMPD, in order to obtain the

indices of the peaks, the sum of the columns of the reshaped LMS matrix after removing rows corresponding to values greater than the scale with the most local maxima, reveals important aspects about the peaks of the input signal 'x'. Moreover, as the column wise sum of the Reshaped LMS matrix, is a vector which produces zero values at the indices where the signal forms peaks, as shown in Figure 4.3, it is easier to obtain the peaks in the real signal 'x'.

Mathematically, the MMPD algorithm can be implemented to a given univariate uniformly sampled signal x = [$x_1$, $x_2$, ..., $x_i$, ..., $x_N$], as follows:

**Step 1**: The Local Maxima Scalogram matrix (LMS) 'M' of Signal 'x' is calculated

a. Signal x is linearly detrended.

b. Local maxima are determined using a moving window approach for every scale 'k', where k = 1, 2, ..., L | L = $\lceil \frac{N}{2} \rceil$ - 1 according to the following.

$$m_{k,\,i} = \begin{cases} 0, & (x_j > x_{w_{LMS}+i-k-1}) \cap (x_j > x_{w_{LMS}+i+k-1}) \\ \alpha, & otherwise \end{cases}$$

$$\forall\, j \in \{n_{LMS}(y)+1,\ n_{LMS}(y)+2,\ \ldots,\ n_{LMS}(y)+w_{LMS}-1,\ n_{LMS}(y)+w_{LMS}\} \cup \{n_{LMS}\left(\left\lfloor \frac{N}{w_{LMS}} \right\rfloor\right)+1, n_{LMS}\left(\left\lfloor \frac{N}{w_{LMS}} \right\rfloor\right)+2,\ \ldots,\ N\}$$

$$\&$$

$$\forall\, y \in \{1, 2, 3, ..., (\lfloor \frac{N}{w_{LMS}} \rfloor)\} \tag{4.5}$$

where,

$\alpha$ : A constant value

x: Given univariate uniformly sampled signal

N: Length of the signal 'x'

$w_{LMS}$: Window size along which the operation of MMPD is performed

$n_{LMS} : \{0, 1 * w_{LMS}, 2 * w_{LMS}, 3 * w_{LMS}, \ . \ . \ ., \ \lfloor \frac{N}{w_{LMS}} \rfloor * w_{LMS}\}$

**Step 2**: Row wise summation of the LMS matrix M

$$\gamma_k = \sum_{i=1}^{N} m_{\ k,\ i} \ \forall \ k \ \in \ 1, \ 2, \ . \ . \ ., \ L \tag{4.6}$$

a. vector $\gamma$ has the information about local maxima (scale dependent distribution of zeros)

b. The global minimum value of $\gamma$ represents the scale with the most local maxima.

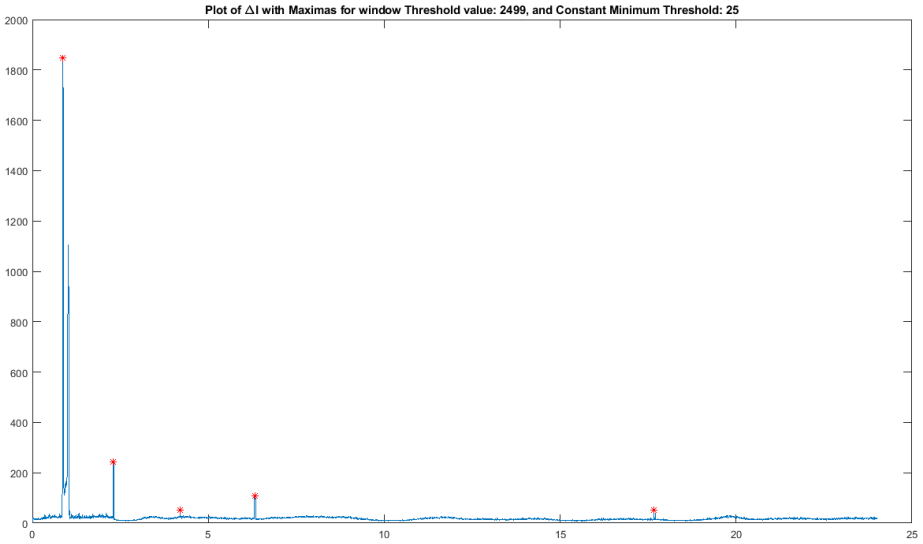
$$\lambda = \ arg \ min \ (\gamma_k) \tag{4.7}$$

**Step 3**: Reshape LMS Matrix 'M' by removing all elements belonging to scales greater than $\lambda$.

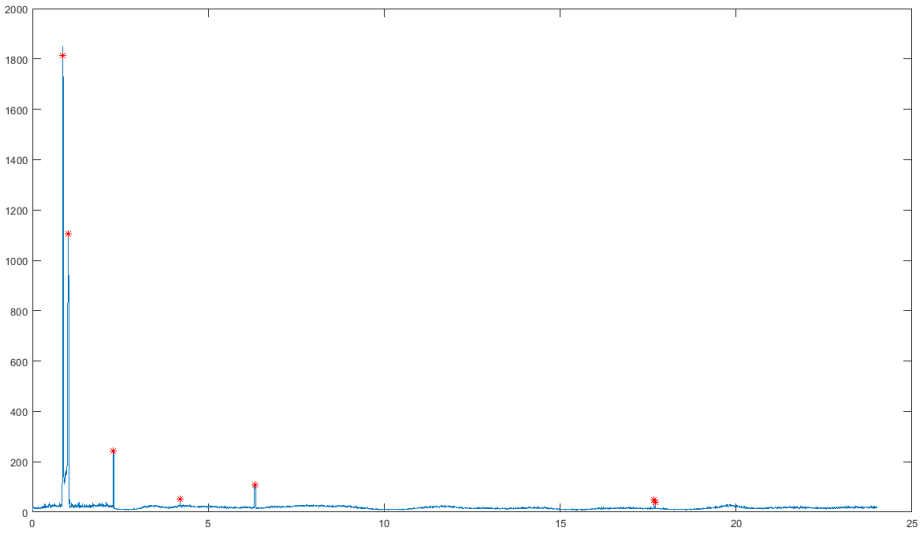
**Step 4**: Finally, the peaks are detected.

a. The column wise Summation vector $\sum_i$, $\forall$ i = 1 to N is calculated.

b. All the indices for which the $\sum_i = 0$ holds true $\forall$ i = 1 to N, there exists a peak.




(a)

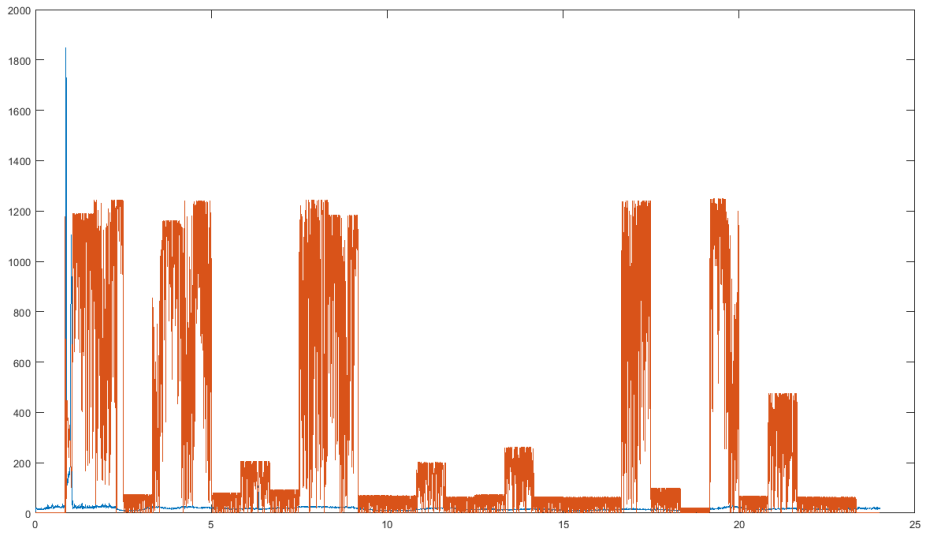



(b)

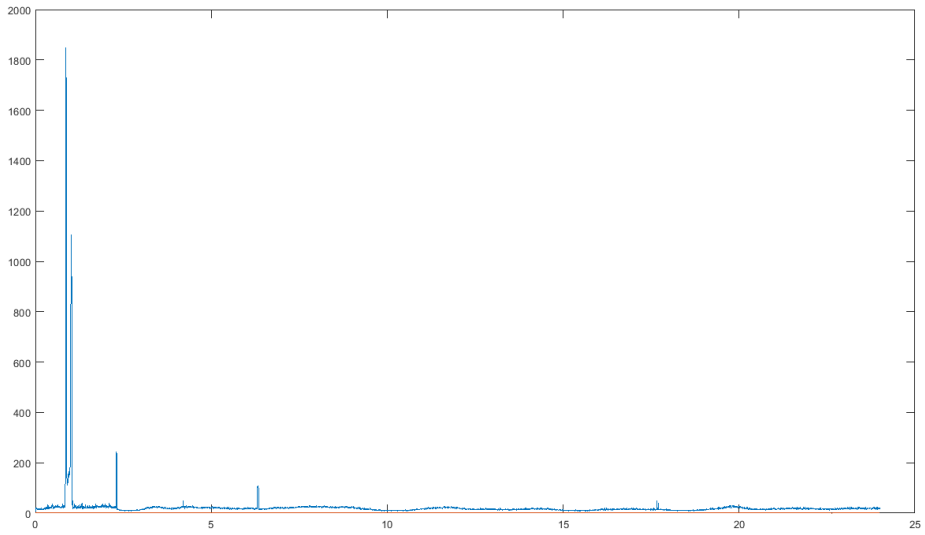
Figure 4.2: Comparison of Peak detection methods using MMPD and AMPD algorithms. (a) The plot illustrates the Input waveform along with detected peaks using MMPD algorithm for a period of 24 minutes. (b) As (a) but using AMPD.

Figure 4.3, illustrates the comparison in terms of Local Maxima Scalogram (LMS) matrix values, for MMPD as well as for AMPD. The crimson color waveform shows the LMS matrix values corresponding to the blue color detection function waveform. As can be observed from the waveform shown, LMS matrix values for

MMPD are very large, which can be attributed to the interaction of every single variable rather than a V-shape formation in the case of AMPD while forming the LMS matrix. However, in case of AMPD, the LMS matrix values are almost always less than 1 and hence cannot be observed easily in the figure.




(a)




(b)

Figure 4.3: Plot for comparison of Local Scalogram Matrices for both the methods of MMPD as well as AMPD. (a) The plot illustrates the original signal along with the sum along column of LMS for MMPD. (b) As (a) but using AMPD.

Further, in order to eliminate the peaks below certain threshold, determined for the purpose, a moving window threshold approach has been implemented after using MMPD to detect the peaks. The moving window approach makes the system to be robust to outliers. It is implemented as described in Section 4.1.3.

In order to compare the peak detection techniques of AMPD and MMPD methods, Table 4.1 illustrates the difference in the speed for peak detection on a data set of 3 minutes duration.

In this case, simulations are run 10 times, for each method, using the same data. The minimum value among the multiple run is selected as the choice for the final time. This process is repeated for different cases.

| Process Time (Seconds) | $\tau_{AMPD}$ | $\tau_{MMPD-250}$* | $\tau_{MMPD-2500}$* |
|---|---|---|---|
| Data 1 - Run 1 | 17.867576 | 0.209637 | 0.521135 |
| Data 1 - Run 2 | 16.064284 | 0.295011 | 0.396275 |
| Data 1 - Run 3 | 14.464602 | 0.112288 | 0.396813 |
| Mean | 16.132154 | 0.205645333 | 0.438074333 |

Table 4.1: Comparison of Performance in terms of simulation time in seconds, of AMPD against MMPD for the purpose of peak detection

*: MMPD-x indicates 'x' windowed operation of peak detection using MMPD algorithm as discussed in section 4.1.2.

In the case of the MMPD, the algorithm operates on smaller matrices as a result of windowed operation. Further, because the random numbers used in AMPD are

replaced with '1s' in MMPD, the performance is improved as well. The MMPD algorithm performs approximately two orders of magnitude faster than the previous AMPD algorithm as is evident from the Table 4.1.

### 4.1.3 THRESHOLDING

A drawback of fixed a threshold, is that it produces many false negatives, as a result of the wide range in the size of the peaks. But, in order to overcome the problem of large differences in peak sizes in determining the real peaks, dynamic thresholding can be used. Instead of using a fixed constant value to determine the peaks, the threshold value changes as we move along the signal. It varies based on the window size $w_{Threshold}$ and is determined by some statistical measure of the sample points within that particular window. It uses the following steps:

**Step 1**: A fixed cutoff value is assigned.

**Step 2**: A Window Length $w_{LMS}$ is assigned over which median value is calculated.

**Step 3**: For every window the median value of sample points within that window duration is calculated.

**Step 4**: The median is added with the fixed cut off value in order to obtain the new cutoff for each window.

In this work, a window length of 2499 sample points is considered. This corresponds to 12.5 secs. An odd value window length is chosen in order to obtain an integer median.

### 4.1.4 DETECTION OF EVENTS

A major portion of the study depends on the events and their features. In order to analyze the features, the starting point of the event as well as the end point for the same event needs to be defined.

Figure 4.4 illustrates the characteristics of an event before it occurred, during the transient, and after the transient settles. These characteristics are evident in Figure 4.5.

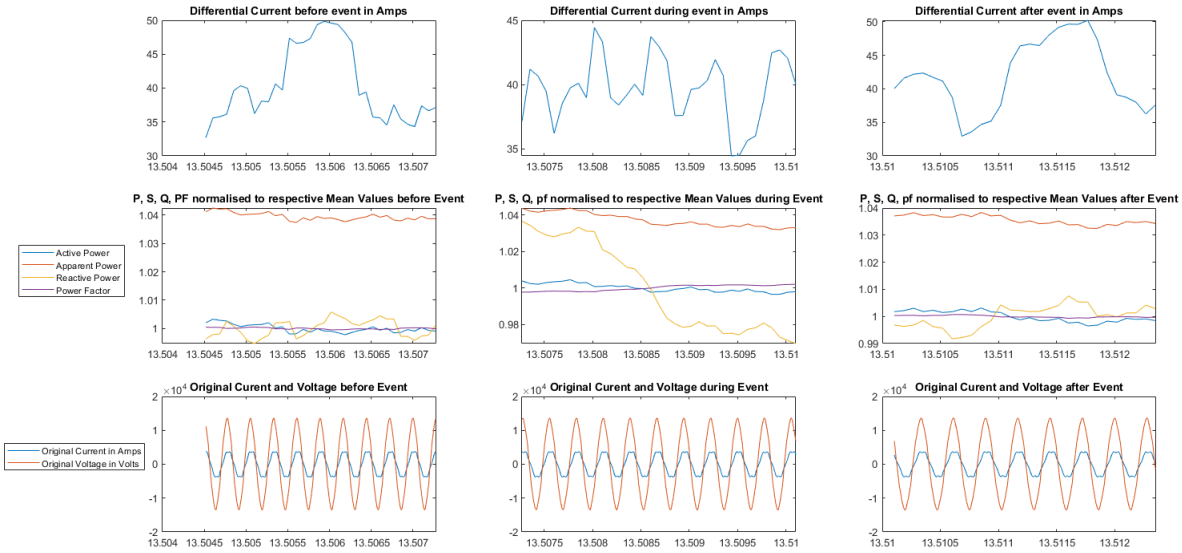



Figure 4.4: Plot for various features before, during, after a particular event

## 4.2 DISCUSSION

Once the start and end of a particular event is defined, the features for that particular event are calculated using the detailed calculation explained in Section 3.2. Using these features, clustering of events will be performed. The clustering analysis is discussed in Chapter 5. By using several statistical measures for each event, Chapter

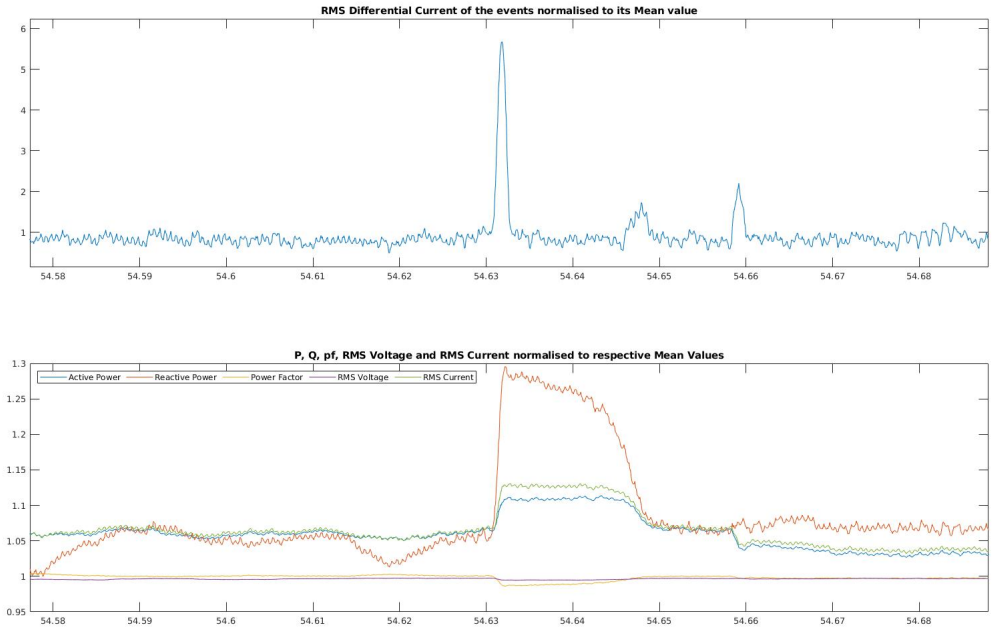



Figure 4.5: Plot for multiple features corresponding to an event, determined by the detection function signal

5 details the clustering algorithm and the methods for training models, to form event

clusters.

# Chapter 5

# CLUSTERING ANALYSIS

## 5.1 INTRODUCTION TO CLUSTERING

After the duration of the event is defined, and the unique features space associated with each event is calculated; the next step is to apply clustering techniques described in this section. The purpose of clustering the events is to be able to segregate the events into different categories and associate every single category with a representative value for the feature space. Ultimately this may enable us to design the definition for electronic loads.

A Self Organising Map (SOM) is used to cluster the events, detailed in the following section. In order to compare the performance of the clustering technique, the clustered data is then used to train a Neural Network model and the classification is done on the same data set using the Neural network model. This is discussed in the Section 5.2.2.

## 5.2 CLUSTERING OF DYNAMIC FEATURES USING SELF ORGANIZING MAPS AND NEURAL NETWORK

The sample points in this case are organized into clusters of events. Next, a statistical analyses of the the sample points provide the information to be used for clustering and analysis. The statistical variables that are used as the representative points are as follows:

**1.** Area under the curve

**2.** Mean of the sample points for each event

**3.** Mode of the sample points for each event

**4.** Variance of the sample points for each event

### 5.2.1 CLUSTERING OF EVENTS USING SELF ORGANIZING MAPS

The Self-Organizing Map (SOM) is an unsupervised clustering algorithm, which can cluster large data sets. This technique uses neurons, each of which is associated with a 'weight' vector in the input space. Each of the neurons has the same number of dimension as the input vector. Thus, the SOM has the capability to describe the mapping from higher-dimensional input space to a lower-dimensional space.

In training, all the neurons are initialized to some random weight. Through multiple iterations, the neurons gradually cluster around areas with a high density of data points.

The SOM technique can be used as a Sequential Algorithm or as a Batch Algorithm [12].

**Sequential Algorithm:**

The first sample point is introduced ⇒ Best Matching Unit (BMU) for the vector is chosen ⇒ Weight is updated ⇒ Next sample point is introduced ⇒ and so on.

**Batch Algorithm**:

All the sample points are introduced to the network ⇒ The BMU neuron for each input vector is chosen ⇒ Each weight vector is updated based on the average position of all of the input vectors for which it is a winner, or for which it is in the neighborhood of a winner based on neighborhood radial distance ($r_N$).

This work has used a **Batch Algorithm**, with neurons arranged in a 2-D Hexagonal grid. The radial neighbourhood distance for BMU is considered $r_N = 1$. The steps used in SOM sequential algorithm in order to achieve the desired result is explained below. The same can be extended to the Batch algorithm easily.

**Step 1. Allocation of Best Matching Unit (BMU)** : The Euclidean distance between the randomly chosen Input vector '$x$' and the weight vector of all the neurons is calculated and the neuron with the least distance is assigned

the term BMU.

$$\begin{aligned} \|d_i\| &= x - w_i \qquad \forall\ i\ \in\ 1,\ 2,\ \ldots,\ N \\ BMU &= \ arg\ min\ (d_i) \quad \forall\ i\ \in\ 1,\ 2,\ \ldots,\ N \end{aligned} \tag{5.1}$$

where,

BMU : Best Matching Unit

$d_i$ = Distance between the input vector $x$ and the neuron with weight $w_i$

$\forall\ i\ \in\ 1,\ 2,\ \ldots,\ N$

N = Maximum number of neurons defined by the supervisor

**Step 2. Update** : The weight vector '$w_i$' is updated according to Equation (5.2).

$$\begin{aligned} w_i(t+1) &= w_i(t) + \alpha(t)\ \theta_{BMU,i}(t)(x(t) - w_i(t)) \quad \forall\ i\ \in\ 1,\ 2,\ \ldots,\ N \\ \theta_{BMU,i} &= \{i\ \forall\ d_i\ \leq\ r_N\} \qquad\qquad\qquad\qquad\quad \forall\ i\ \in\ 1,\ 2,\ \ldots,\ N \end{aligned} \tag{5.2}$$

The weights for all the neurons within a certain radius of the BMU neuron are updated, using the Kohonen rule as shown in equation (5.2).

where,

t : Training Iteration index

$w_i(t)$ = Weight vector of neuron 'i' at time t

$\alpha(t)$: Learning rate at time 't'

$\theta_{BMU,i}$ : Neighborhood function which contains the indices of all the neurons less than or equal to the radial distance $r_N$ from the BMU at time t

N = Maximum number of neurons defined by the Supervisor

The training works in two phases:

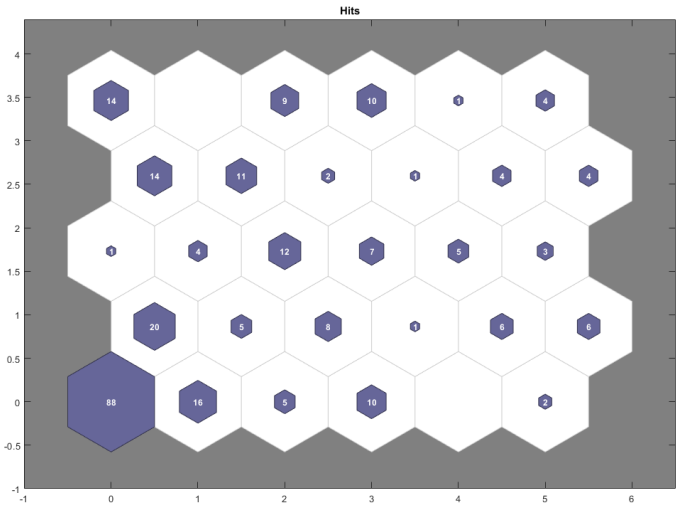
In the first phase, iterate using a large initial learning rate and large neighborhood radius is used.

In the second phase, a small initial learning rate and small neighborhood radius are used, to fine tune the SOM.

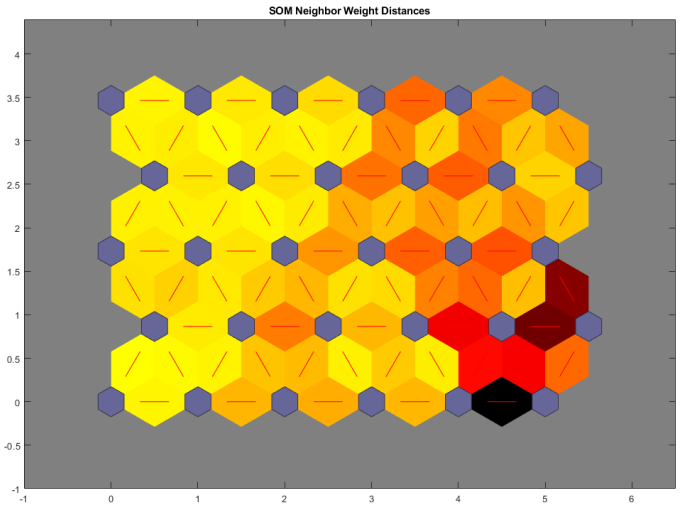
Areas with multiple neurons represent clusters in the data. As the weights of the neurons are updated, the neurons move in such a way that the overall shape of the data can be reflected by the positioning of the neurons in the grid.

Figure 5.1 represents the results obtained for the Self Organizing Map used to cluster the 283 events obtained y using the AMPD algorithm discussed in Section 4.1.1 on data worth 24 hours, using 82 features obtained as described in Section 3.1.1.

The clusters formed using the Self Organising Map technique are each associated with multiple features. In order to plot the multidimensional clusters, the study has used glyph plots with each axis representing a feature. Figure 5.2 shows the visualization of different loads in each cluster formed using the Self Organizing map technique. As is evident from the Figure 5.1(a), 3 Clusters have no events associated with them. This is because Self Organizing Maps differ from other artificial neural networks as they apply competitive learning as opposed to error-correction learning,




(a)




(b)

Figure 5.1: Results of Clustering by using Self Organizing Map (a) SOM Sample Hits: Plot of distribution of events in each neuron which it classifies, showing their number. The relative number of events is encoded using the size of neurons. (b) SOM Neighbor Weight Distance: Plot of neurons shown in gray-blue color with their direct neighbor relations indicated by red lines. The distance of the neuron's weight vector with their neighbors is color coded with black being farthest to yellow being closest.

and they use a neighborhood function to preserve the topological distribution of the input vectors they are trained on. Hence, the SOM network modeled here decided that out of the 35 neurons, only 33 neurons are sufficient to cluster the data, which can further be corroborated using the Neighbour weight distance as shown in Figure 5.1(b).

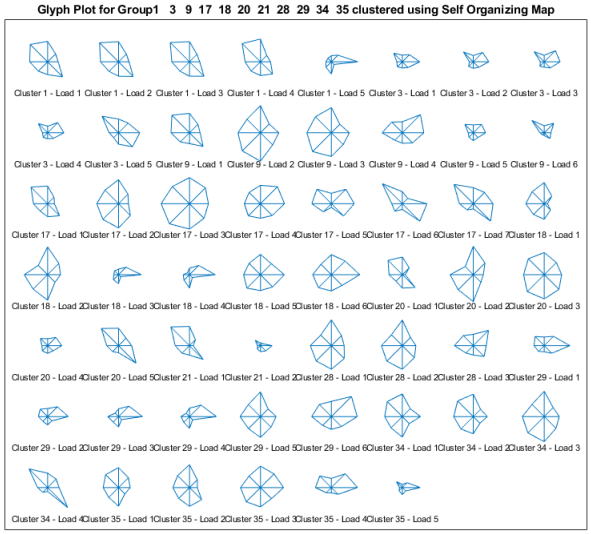



Figure 5.2: Visualization of multiple groups with limited set features, formed using SOM

An illustration of clusters formed using SOM, plotted against time in minutes, is shown in Figure 5.3.

### 5.2.2 PATTERN RECOGNITION AND CLASSIFICATION USING NEURAL NETWORK

Neural Networks solve the classification problem for nonlinear sets by employing hidden layers. The additional hidden layers can be interpreted geometrically as additional hyperplane, which enhance the separation capacity of the network [8]. The Input to the Neural Network as shown in the Figure 5.4.

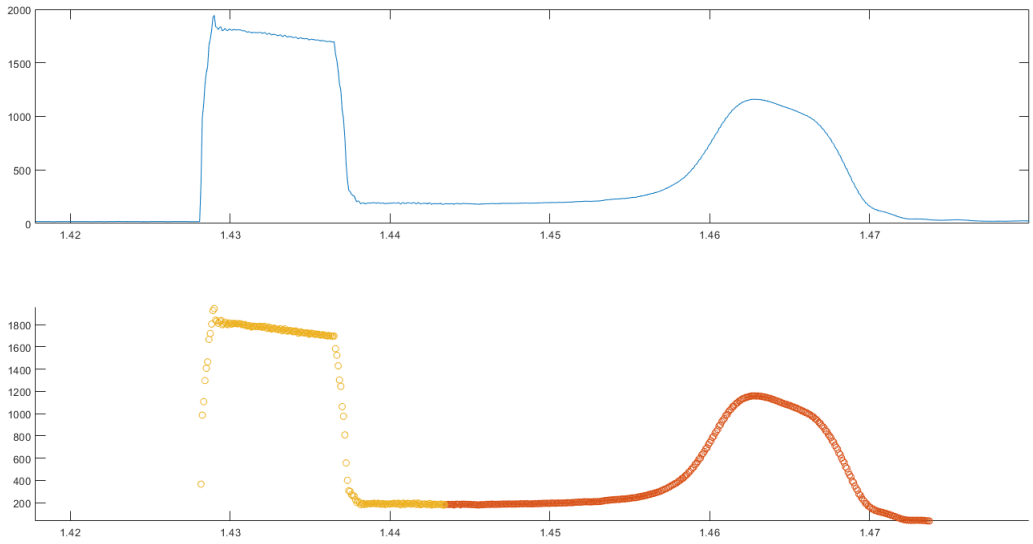



Figure 5.3: Visualization of detection function, with clustered groups distinguished by colors, formed using SOM, against time in minutes

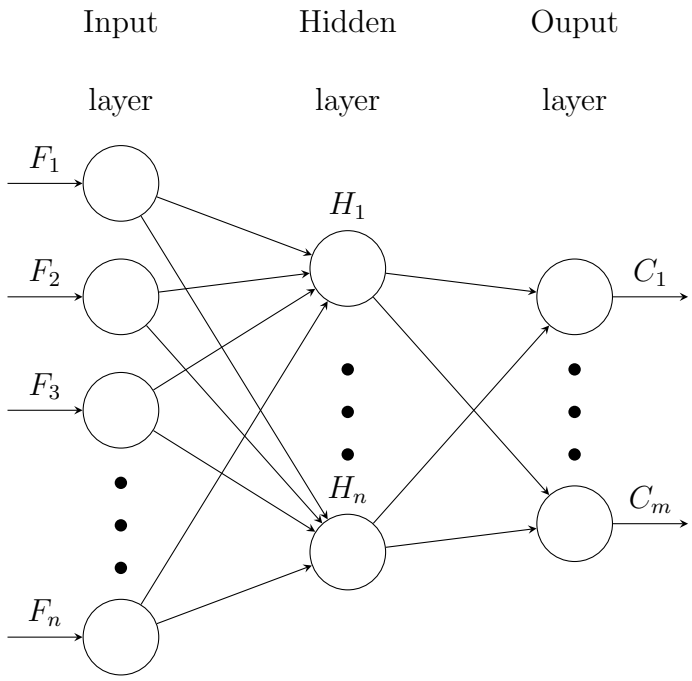



Figure 5.4: A Neural Network, with with 'n' features as input to the input layer obtained as discussed in chapter 3, for each event detected as discussed in the section 4.1.1, having 10 hidden neurons, results in 'm' output neurons with each neuron representing a cluster.

The hidden layer value of 10 has been obtained by empirical trials. Multiple Hidden layer values were compared against a fixed output layer value. In this case, the simulation is run 10 times for every change in the number of hidden layers, and the minimum value among the multiple results obtained is selected as the choice for that hidden layer. The process is repeated for different cases as mentioned in the Table 5.1.

| No. of Hidden Layers | No. of Output Layers | Cross Entropy | Mean Square Error |
|---|---|---|---|
| 1 | 35 | 0.0767 | 0.026 |
| 7 | 35 | 0.0437 | 0.0184 |
| 9 | 35 | 0.0378 | 0.0156 |
| 10 | 35 | 0.034 | 0.0145 |
| 30 | 35 | 0.0708 | 0.0233 |
| 100 | 35 | 0.0425 | 0.0143 |

Table 5.1: Comparison of performance in terms of errors, for the purpose of hidden layer selection

Using Neural Network Graphical User Interface [12] and the outputs of Self Organizing Map Clustering, the model is trained on 70% of the clustering data set, validation is performed on 15% of the data set, and testing is performed on the remaining 15% of the data set.

The clusters formed using the neural network model are each associated with multiple features. In order to plot the multidimensional clusters, the study has used glyph plots with each axis representing a feature. The Figure 5.5 shows the visu-

alization of different loads in each cluster formed using the neural network model trained using the data clusters obtained using SOM.

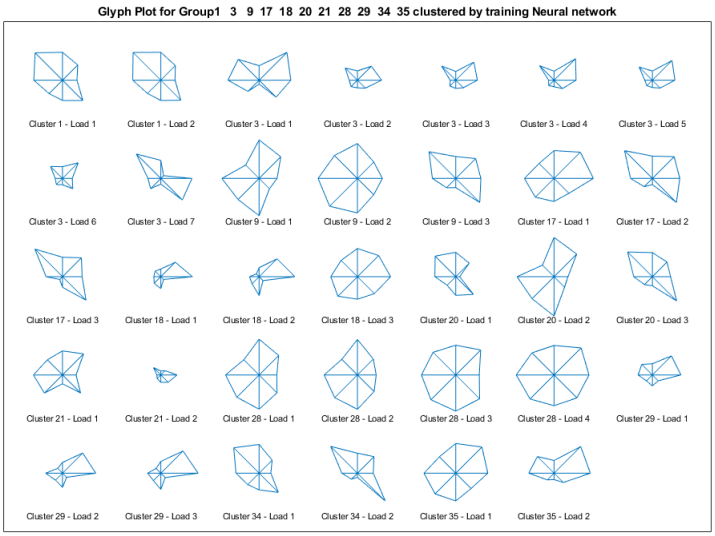



Figure 5.5: Visualization of multiple groups with limited set features, formed using Neural Network model

An illustration of clusters formed using neural network model trained using the clusters formed by SOM, plotted against time in minutes, is shown in Figure 5.6.

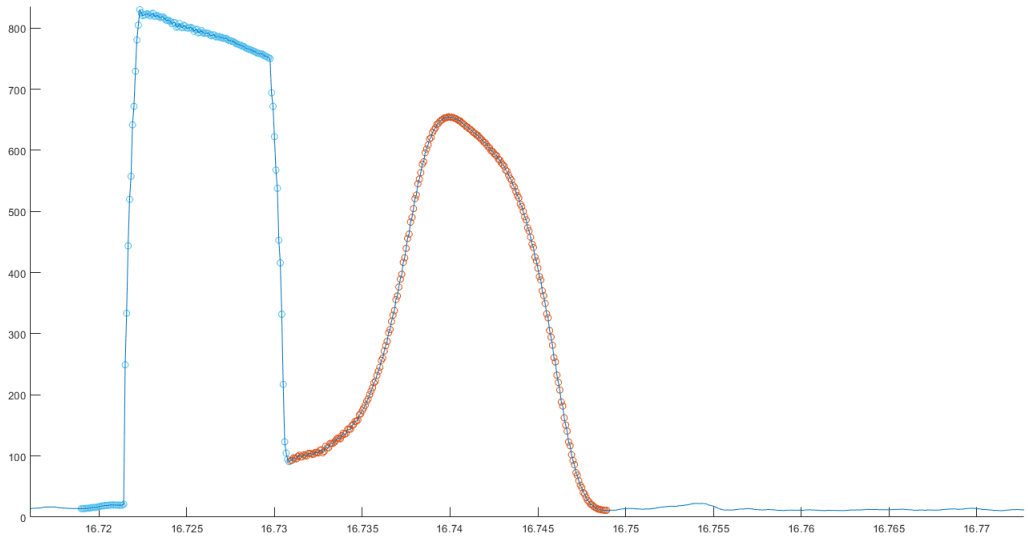



Figure 5.6: Visualization of detection function, against clustered groups distinguished by colors, formed using the Neural Network trained model

## 5.3 SIGNAL SIMILARITY TECHNIQUES

Once the clustering is performed, in order to compare the performance of the algorithm on the given data set, the correlation coefficient of events among each other for every cluster is calculated, to obtain a correlation matrix.

This work uses the Pearson product moment correlation coefficient as the measure of similarity between two signals. The Pearson product moment coefficient measures the linear correlation between two events $E1$ and $E2$, and is defined as follows:

$$\rho\left(E1, E2\right) = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{E1_i - \ \mu E1}{\sigma_{E1}} \right) \left( \frac{E2_i - \ \mu E2}{\sigma_{E2}} \right) \tag{5.3}$$

where,

$\rho$ : Pearson product moment correlation coefficient

E1, E2 : Two different events

$\mu$ : The sample mean $\sigma$ : The sample standard deviation

The challenge in finding the correlation between events is that the duration for events varies. The correlation coefficient methods described above considers both the signals to be of equal length. In order to achieve the objective of making the length of each pair of events to be equal, the work uses a technique called Dynamic Time Warping (DTW).

### 5.3.1 DYNAMIC TIME WARPING

In case of two time series signals, X = [$x_1$, $x_2$, . . . , $x_i$, . . . $x_{|X|}$], and Y = [$y_1$, $y_2$, . . . , $y_j$, . . . , $y_{|Y|}$], if one of the time series signals is shifted along the time axes, the Euclidean distance considers the two signals far apart from each other. Dynamic Time Warping (DTW) helps to overcome this limitation and provide an intuitive meaning to the similarity metric between the two time series signals.

Specifically, Dynamic Time Warping is defined as a method to obtain an optimized alignment between two time series signals, which can be obtained, by either stretching or shrinking the time axes of the signals [17] . Two time series signals x and y, are shown in Figure 5.7.

The objective is to create a path W = [$w_1$, $w_2$, . . . , $w_K$], where $K$ is the length of the warp path and the $k^{th}$ element of the warp path defined as $w_k = (i, j)$, such that $i$ is the index of time series signal $x$ and $j$ being that of the time series signal $y$.

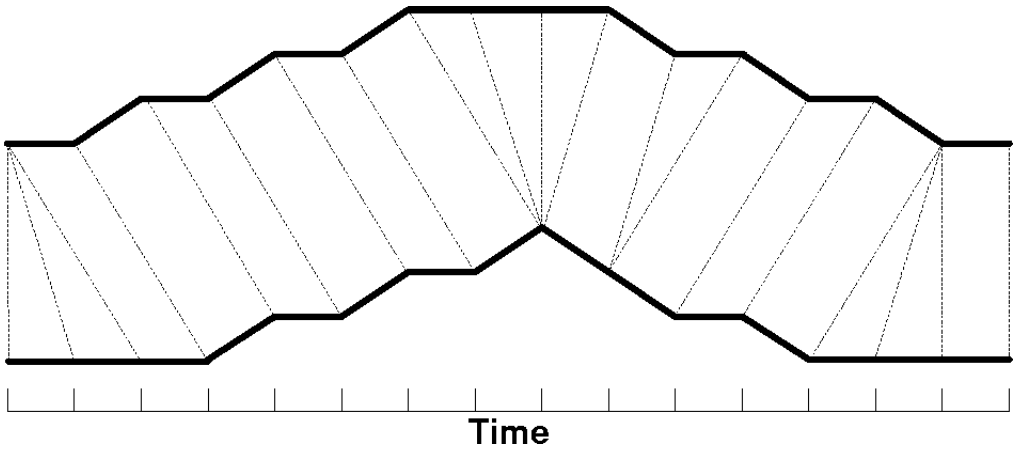



Figure 5.7: Plot of warping for two time series signals shifted along the time axis [17]

The objective is conditioned on the following constraints:

**1**: The warp path must start from the beginning of each of the time series signals and finish at the end of both the signals.

**2**: The warp path must either be incremented by one index or stay the same, to prevent overlapping of warp path.

**3**: Every index of the time series signals must be used

This can be formulated as an optimization problem:

$$\underset{W}{\text{minimize}} \quad \sum_{k=1}^{K} D(w_{\,k,\,i}\,, w_{\,k,\,j}\,) \tag{5.4a}$$

$$\text{subject to} \quad w_0 = (1,1)) \tag{5.4b}$$

$$w_K = (|x|, |y| \tag{5.4c}$$

$$i \leq i' \leq i+1 \tag{5.4d}$$

$$j \leq j' \leq j+1 \tag{5.4e}$$

where,

$D(w_{\,k,\,i}\,, w_{\,k,\,j}\,)$: The Euclidean distance between the indexes i and j of time series signals x and y, corresponding to the $k^{th}$ element of the warp path W.

If, $w_k = (i, j)$ then, $w_{k+1} = (i', j')$

$i$ is the index of time series signal $x$ and $j$ being that of the time series signal $y$

A programming approach has been used in this research for obtaining the optimised warp path. Figure 5.8 shows the plot of warp path in the Cost function matrix for

the two time series signals x and y, shown in Figure 5.7.

In Figure 5.7, the first index of time series Y is connected to first three indexes of time series X. Hence the first three indexes in the cost function matrix holds the values (1,1), (2,1), (3,1). The warp path in the cost function matrix in Figure 5.8 has been plotted in such a manner, taking into consideration the objective function and the constraints as explained.

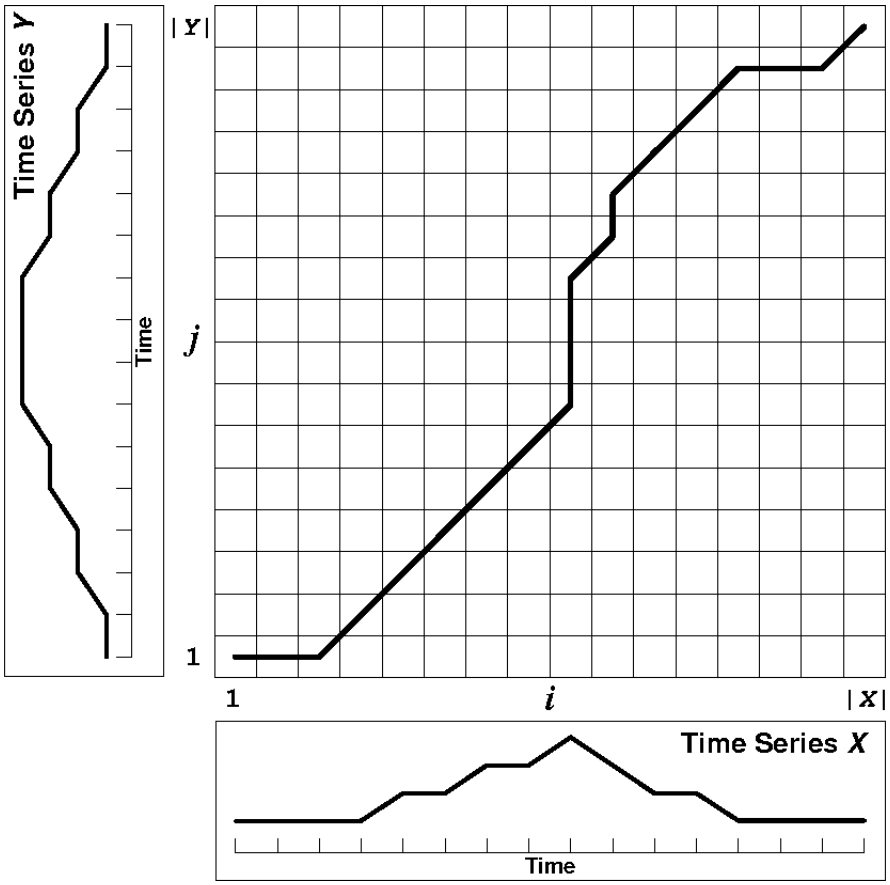



Figure 5.8: Plot of the cost function along with the warp path against time of time series $x$ and time of time series $y$ [17]

### 5.3.2 RESULTS OF SIGNAL SIMILARITY TECHNIQUES

The correlation matrix hence obtained for clusters by using the Self Organizing Map algorithm is plotted in the form of a Correlation Matrix, shown in Figure 5.9.

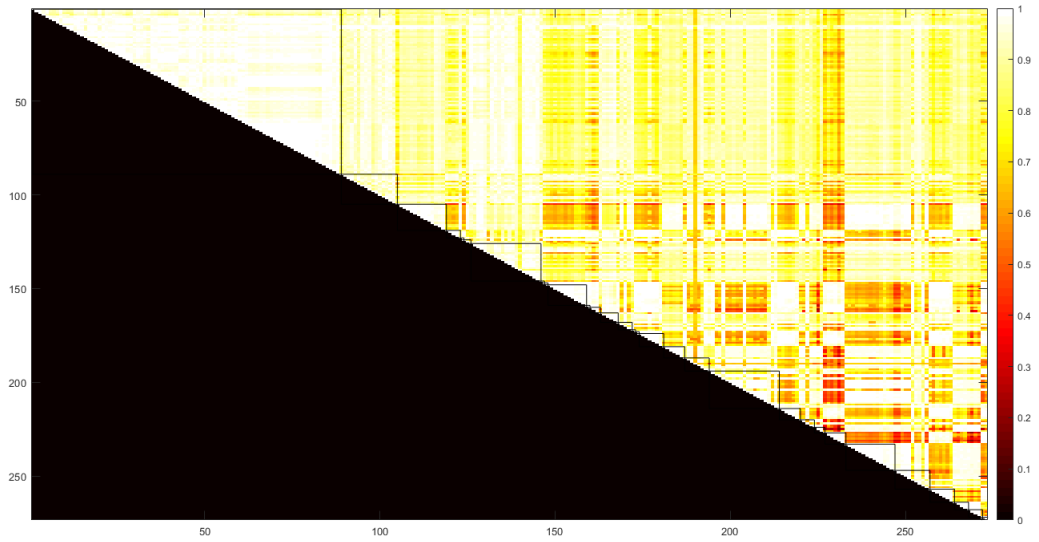



Figure 5.9: Correlation matrix for clusters formed using Self Organizing Map algorithm

After the training of the Neural Network model is done, clustering is performed using 10 hidden neurons. Then, in order to compare the performance of the algorithm on the given data set, the correlation coefficient of events among each other for every cluster is calculated and is plotted as Correlation Matrix, shown in Figure 5.10.

The plots of the correlation matrix represents the correlation between each event group with every other event groups obtained using the SOM method or by training the Neural Network model. The higher the value of correlation, the more correlated the event groups are, the highest value being 1. The lower correlation values move towards zero. The color encoding is used to visualize the correlation matrix among the event groups, where white represents the highest correlation among two events, and black represents zero correlation between the events. Only half of the Correlation matrix is shown in the figures, since it is a symmetrical matrix.

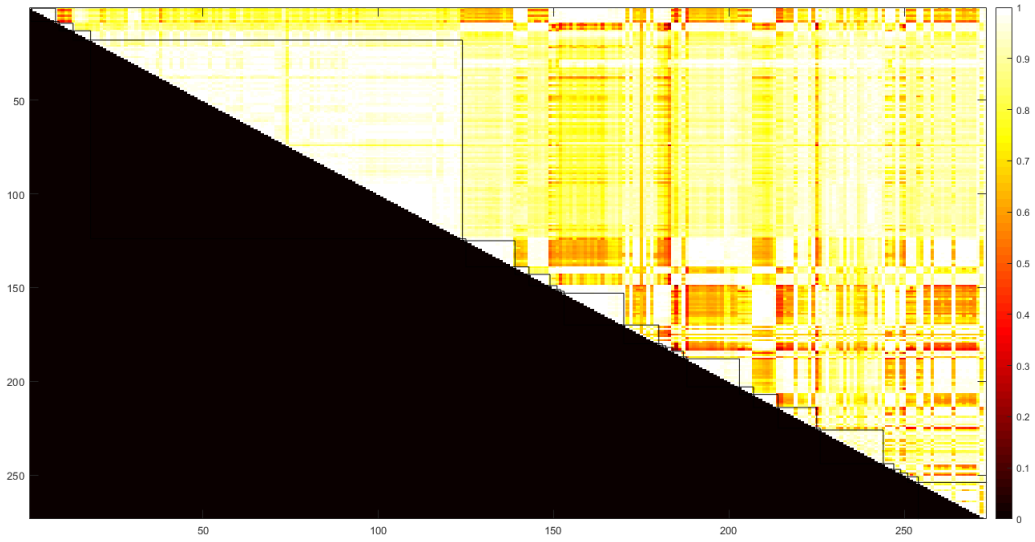



Figure 5.10: Correlation matrix for clusters formed using Neural Network training and clustering

The boundary shown by the lines indicates the self correlation values between pairs of events belonging to each group. Hence, the ideal value expected is a set of 1s in those triangles formed by enclosing of lines. However, practically we expect a much whiter space and less yellow colored regions in those triangular regions enclosed by lines. As is evident from the Figures 5.9 and 5.10, the self correlation values are higher in case of the pair of events in clusters, formed by using SOM as compared to the groups formed by training a Neural Network.

This may be attributed to the capability of SOM to be able to recognize the neighboring input space. Further, SOM learns the distribution of the sample points while clustering as well as the topology of the input training vector. The SOM performs well for clustering the events with higher fidelity. Moreover, the poor performance of Neural Network model as compared to SOM may be attributed to the fact that the

training data for formation of clustering is obtained by the output of SOM. Hence, the Neural Network considers the SOM to be the maximum performance and strives to do its best in order to match the performance of SOM, rather than matching the actual clusters. The performance of Neural Network is degraded, while striving for matching the SOM performance.

## 5.4 CONCLUSION ON CLUSTERING TECHNIQUES

From the analysis performed on the clusters and their correlation Matrix, the Self Organizing Map performs better than the Neural Network. The mean values of magnitude for each odd harmonic component till the $15^{th}$ harmonic value, for every cluster is represented in the following Table 5.2. In the Table 5.2, Clusters 4, 11, and 19 have no values, because the Self Organising Map decided that only 33 neurons are sufficient to cluster the data. Hence no events are grouped to the Clusters 4, 11, and 19. So the mean values of magnitude for each odd harmonic components is valued at zero for these clusters.

Table 5.2: Cluster-Harmonic Matrix

| Harmonic No. / Cluster No. | (3) | (5) | (7) | (9) | (11) | (13) | (15) |
|---|---|---|---|---|---|---|---|
| Cluster 1 | 18.9 | 133.5 | 66.4 | 26.1 | 48.7 | 15.5 | 10.5 |
| Cluster 2 | 19.7 | 119.4 | 55.6 | 23.1 | 44.8 | 12.4 | 8.3 |
| Cluster 3 | 19.6 | 104.4 | 54.1 | 21.3 | 41.6 | 8.2 | 8.1 |
| Cluster 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Table 5.2: Cluster-Harmonic Matrix

| Harmonic No. / Cluster No. | (3) | (5) | (7) | (9) | (11) | (13) | (15) |
|---|---|---|---|---|---|---|---|
| Cluster 5 | 19.7 | 106.4 | 51.0 | 21.7 | 43.4 | 11.3 | 7.8 |
| Cluster 6 | 20.2 | 133.4 | 58.1 | 26.0 | 51.8 | 17.5 | 9.3 |
| Cluster 7 | 18.4 | 133.8 | 62.9 | 24.8 | 45.8 | 11.7 | 9.8 |
| Cluster 8 | 20.4 | 130.0 | 59.2 | 25.4 | 49.2 | 16.0 | 10.2 |
| Cluster 9 | 20.3 | 140.4 | 59.5 | 27.4 | 55.6 | 19.0 | 9.6 |
| Cluster 10 | 19.1 | 110.1 | 60.4 | 21.6 | 42.7 | 9.6 | 9.7 |
| Cluster 11 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Cluster 12 | 19.0 | 110.6 | 62.5 | 22.0 | 42.9 | 8.1 | 10.7 |
| Cluster 13 | 22.4 | 131.3 | 59.5 | 27.5 | 51.9 | 16.1 | 9.5 |
| Cluster 14 | 19.7 | 118.2 | 55.0 | 23.6 | 45.3 | 12.7 | 8.5 |
| Cluster 15 | 20.7 | 118.3 | 52.5 | 24.4 | 48.3 | 14.3 | 8.1 |
| Cluster 16 | 18.3 | 128.9 | 61.7 | 24.3 | 46.5 | 14.4 | 9.6 |
| Cluster 17 | 21.2 | 125.2 | 62.1 | 24.5 | 44.7 | 11.8 | 10.6 |
| Cluster 18 | 19.8 | 110.5 | 45.9 | 22.2 | 46.0 | 14.2 | 7.1 |
| Cluster 19 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Cluster 20 | 19.5 | 135.9 | 57.3 | 25.3 | 48.1 | 16.6 | 9.7 |
| Cluster 21 | 18.2 | 108.8 | 49.2 | 22.8 | 45.8 | 9.1 | 8.4 |
| Cluster 22 | 19.1 | 125.5 | 59.5 | 24.1 | 47.9 | 14.5 | 9.5 |
| Cluster 23 | 20.7 | 131.8 | 54.6 | 26.4 | 54.5 | 15.9 | 9.3 |
| Cluster 24 | 21.4 | 137.5 | 58.4 | 27.7 | 56.3 | 19.7 | 9.6 |

Table 5.2: Cluster-Harmonic Matrix

| Harmonic No. / Cluster No. | (3) | (5) | (7) | (9) | (11) | (13) | (15) |
|---|---|---|---|---|---|---|---|
| Cluster 25 | 19.7 | 112.8 | 49.8 | 23.1 | 45.2 | 13.8 | 7.4 |
| Cluster 26 | 19.3 | 117.9 | 52.1 | 23.4 | 47.0 | 15.6 | 7.6 |
| Cluster 27 | 22.1 | 123.1 | 55.6 | 26.4 | 48.1 | 13.1 | 9.0 |
| Cluster 28 | 21.0 | 141.9 | 57.2 | 27.8 | 60.2 | 22.7 | 10.5 |
| Cluster 29 | 20.6 | 103.9 | 44.3 | 22.3 | 41.1 | 12.1 | 6.3 |
| Cluster 30 | 21.6 | 124.4 | 54.8 | 25.6 | 49.9 | 16.1 | 8.6 |
| Cluster 31 | 20.9 | 120.7 | 52.1 | 24.5 | 48.4 | 16.0 | 8.0 |
| Cluster 32 | 21.7 | 127.7 | 54.8 | 26.4 | 51.6 | 17.1 | 8.7 |
| Cluster 33 | 20.9 | 134.8 | 59.7 | 27.6 | 57.9 | 19.9 | 9.8 |
| Cluster 34 | 18.9 | 127.3 | 64.9 | 23.6 | 43.3 | 10.3 | 11.4 |
| Cluster 35 | 22.4 | 121.6 | 54.7 | 26.1 | 49.0 | 13.8 | 8.3 |

From Table 5.2, the clusters with relatively lower values can be classified as electrical loads and those with unusually higher values are classified as electronics loads.

# Chapter 6

# SUMMARY, CONCLUSIONS AND FURTHER WORK

A method for event detection has been proposed in this work. The work has used multiple load signatures based on both time and frequency domain analysis for detection of the loads. The harmonic content and the electrical characteristics of the waveform provide a unique way to achieve the objective of load monitoring.

This research focused on four major aspects of load monitoring. First, the comparison of Fast and Slow Transients. Multiple cycle difference values are compared for the purpose of calculating the detection function to detect events. The 30 cycle differential waveform is observed to have the perfect combinations of smoothed curve with spikes at the time of events. The reason for this can be attributed to the fact that, the 3 cycle difference captures only fast transients, whereas the 30 cycle difference waveform is capable of capturing both fast and slow transients.

Second, for the purpose of event analysis, the research used peak detection techniques. In order to determine the peaks, a Modified Multiscale Peak detection (MMPD) Algorithm has been proposed, which performs equally well to detect peaks and at the same time the MMPD algorithm performs approximately two orders of magnitude faster than the existing AMPD algorithm. This is due to operation on

smaller matrices as a result of windowed operation and the use of '1s' instead of random numbers.

Third, a feature selection method has been discussed. On testing multiple sets of features, with various supervised classification models, a final set of features are selected for analysing the events. Finally, the K-Nearest Neighbors model appears to have performed the best with a 99.8 % average performance.

Fourth, the features obtained using the methods, are used to cluster the events detected into groups. The Clustering of Events using a Self Organizing Map and a Neural Network Training Model are performed and the performance of the methods are compared. The Analysis of a similarity metric of both the Self Organizing Map and Neural Network Training Model reveals that SOM performs well for clustering the events with higher fidelity.

Further work can investigate improved ways to optimize the set of parameters for classification of events, and to find a novel measure of similarity/dissimilarity between events. Experimenting the techniques used in this study for analyzing larger systems with more data will be a step forward in this prospect. Future work opportunities include, optimizing the set of parameters for classification of events, which creates the largest distinction among the different load classes. Another task for future researchers involves comparing the different approaches to capture electrical events, such as the effectiveness of modelling events using Dynamic State Estimation approach against using Machine Learning approach.

Finally it would be useful to have a better method to identify the accurate duration of events, instead of relying on peaks using detection techniques to obtain the event duration.

# Bibliography


[1] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on speech and audio processing*, 13(5):1035–1047, 2005.

[2] C Benner, KB Purry, and BD Russell. Distribution fault anticipator". *EPRI, Palo Alto, CA*, 1001879, 2001.

[3] CI Budeanu. Reactive and fictitious powers. *Rumanian National Institute*, 2:127–138, 1927.

[4] Alexander Eigeles Emanuel. On the assessment of harmonic pollution [of power systems]. *IEEE Transactions on Power Delivery*, 10(3):1693–1698, 1995.

[5] Alexander Eigeles Emanuel. Summary of ieee standard 1459: definitions for the measurement of electric power quantities under sinusoidal, nonsinusoidal, balanced, or unbalanced conditions. *IEEE transactions on industry applications*, 40(3):869–876, 2004.

[6] Margaret F Fels et al. Prism: an introduction. *Energy and Buildings*, 9(1-2):5–18, 1986.

[7] George William Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

[8] Dimas Ruliandi Heneka Yoma Priyangga. Application of pattern recognition and classification using artificial neural network in geothermal operation. Technical report, Stanford University, 2018.

[9] Bogdan Kasztenny, Ilia Voloh, Christopher G Jones, and George Baroudi. Detection of incipient faults in underground medium voltage cables. In *2008 61st Annual Conference for Protective Relay Engineers*, pages 349–366. IEEE, 2008.

[10] Roberto Langella and Alfredo Testa. Ieee standard definitions for the measurement of electric power quantities under sinusoidal, nonsinusoidal, balanced, or unbalanced conditions. 2010.

[11] Christopher Laughman, Kwangduk Lee, Robert Cox, Steven Shaw, Steven Leeb, Les Norford, and Peter Armstrong. Power signature analysis. *IEEE power and energy magazine*, 1(2):56–63, 2003.

[12] MATLAB. *Classify Patterns with a Shallow Neural Network*, 2019.

[13] Mirrasoul J Mousavi, John J McGowan, James Stoupis, and Vaibhav D Donde. Apparatus and method for adaptive fault detection in mv distribution circuits, March 5 2013. US Patent 8,390,302.

[14] Karthick Muthu-Manivannan, Carl L Benner, Peng Xu, and B Don Russell. Arcing event detection, January 4 2011. US Patent 7,865,321.

[15] Leslie K Norford and Steven B Leeb. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. *Energy and Buildings*, 24(1):51–64, 1996.

[16] Olivier Poisson, Pascal Rioual, and Michel Meunier. Detection and measurement of power quality disturbances using wavelet transform. *IEEE transactions on Power Delivery*, 15(3):1039–1044, 2000.

[17] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007.

[18] Surya Santoso and DD Sabin. Power quality data analytics: Tracking, interpreting, and predicting performance. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–7. IEEE, 2012.

[19] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In

*2011 31st international conference on distributed computing systems workshops*, pages 166–171. IEEE, 2011.

[20] Felix Scholkmann, Jens Boss, and Martin Wolf. An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals. *Algorithms*, 5(4):588–603, 2012.

[21] Tarlochan S Sidhu and Zhihan Xu. Detection of incipient faults in distribution underground cables. *IEEE Transactions on Power Delivery*, 25(3):1363–1371, 2010.

[22] Transmission Working Group on Power Quality Data Analytics and Power Quality Subcommittee Distribution Committee. Electric signatures of power equipment failures. Technical report, IEEE, 2018.

[23] Ahmed Zoha, Alexander Gluhak, Muhammad Imran, and Sutharshan Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, 2012.