Chapter 1

A Prediction Model

1.1 Motivation

As discussed in Chapter **??**, the main objective of this work is to increase the reliability of power networks in terms of power quality within given financial budget/resources. In order to monitor the power quality, power quality meters are being deployed. Since power quality meters are expensive devices, we (in chapter **??**) proposed Max-Ent based algorithm for power quality estimation on unmonitored network segments. For meter placements, we proposed conditional entropy based efficient algorithms (in chapter chap:meterPlacement) that intelligently place power meters on selected network segments so that the power quality could be inferred as accurately as possible.

Since the power quality readings (exact readings from monitored links, and estimated PQ values from unmonitored links) are available now, we use these readings to estimate the state of the network and identify any potential malfunctioning device in the power network. Our objective in this chapter is to address the research challenge: *how to detect a potential malfunctioning device in the power network based on available power quality readings.* In the next section, we propose an algorithm that helps to detect the malfunctioning devices in the power network.

1.2 Problem Formulation

1.2.1 Assumptions

We make below assumptions about the problem:

1. *The transfer function or behavior $\left(f\left(d \right) \right)$ of each device d in the network is known.* As discussed earlier, a device-specific power quality transfer function could be estimated through physical modeling or through the assessment of historical PQ data. In Chapter **??**, using a real power quality dataset, we have shown how accurately an $f(d)$ could be estimated.

2. *All potential malfunction devices need to be on a monitored segment.* A potential malfunction device could effectively be detected when the device itself or any of its child device is monitored in real time using a PQ meter. This assumption is realistic in the sense that if the underline link is not monitored, we cannot get the real-time status of PQ values and would not be able to detect if a device is not behaving normally. The idea is to detect if a device is significantly deviating form its normal behavior.

3. *The power grid network is a tree-structured network where the electric current flows from root node to the child nodes.* Note that this is a reasonable assumption at any particular instance in time. While enterprise level power grids used in places such as hospitals and data centers often have two utility feeds available as well as an independent emergency power source, only one power source is typically used at one time. See the IEEE Gold Book [**?**] for further information on recommended practices in the design of critical power systems.

4. *The probability mass function $\left(f_x \left(d_0 \right) \right)$ of power quality values at the input link to the root node is known.* In other words, the distribution of power quality at the input to the network, usually the utility feed, is known. This is also a reasonable assumption, since electrical utilities typically report on indices such as System Average RMS Variation Frequency Index (SARFI) which is essentially a count of the number of times the magnitude and duration falls below a threshold. Furthermore, there are often independent bodies that gather statistics on power delivery service reliability that can also be incorporated into an estimate of power quality distribution [**?**].

1.2.2 The Problem

We now formalize the problem of detecting a malfunction device in the network. The objective is to detect a device d in the network as a malfunction device when its power
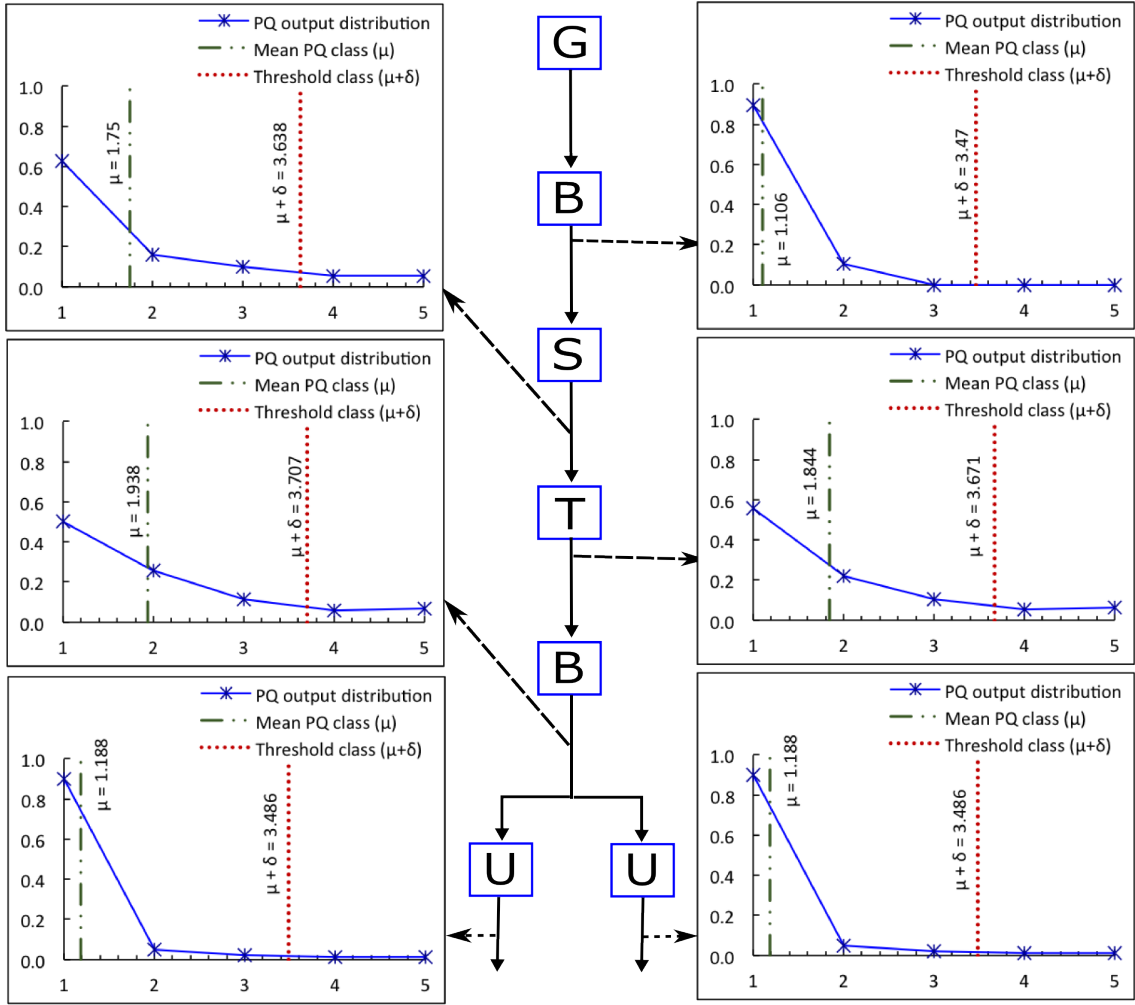
Figure 1.1: Power quality distributions at the output links of various devices. The average power quality class and computed threshold is shown as vertical lines in each distribution graph. The x-axis represents the power quality class c_i while the y-axis represents the probability of c_i. Further, the lower power quality class c_1 represents the best power quality while c_5 represents the worst power quality.

quality output degrades persistently from its normal behavior. Since the transition function $f(d)$ of each device d is known, we know the probability distribution of the output power quality $(f_x(d))$ at the output link of each device in the power grid network. Figure 1.1 shows the probability distributions for various devices in a sample network. The distributions are obtained/computed for the prior distribution of [0.9947 0.005 0.0002 0.00009 0.00001] at the utility feed/generator.

We mathematically analyze these distributions and based on their various statistical properties, we propose out detection algorithm in next section. The proposed

algorithm compares the normal behavior of a device (the $f_x(d)$) of each device with the observed behavior $(\widehat{f_x}(d))$ and compute the power quality degradation (the difference between the two distributions) as Δ_d. The objective function becomes:

$$\Delta_d = f_x(d) \stackrel{\text{def}}{=} \widehat{f_x}(d) \tag{1.1}$$

where the operation $\stackrel{\text{def}}{=}$ is unknown. In the next section, we propose algorithms to solve the equation 1.1.

1.3 Our Prediction Algorithms

1.3.1 A simple correlation measure

In order to compare two probability distributions, the correlation measure is usually used where its most familiar type is the *Pearson's correlation coefficient*, simply known as *the correlation coefficient*. In order to use the formula, we represent the actual PQ distribution as X while the obtained distribution as Y. For the random variables X and Y, it is defined as:

$$\rho_{X,Y} = corr(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

The above correlation measure could be used to compare the actual and observed PQ distributions. Generally speaking, this measure can accurately tell how much the observed distribution is similar (or different) from its actual one. When a device is perfectly behaving like its actual distribution, the measure returns a maximum value (+1). On the other hand, when a device is producing a very different (for instance producing very bad PQ) distribution, it returns a smaller value (the smallest possible is -1). The value 0 (or closer to 0) represents that the two distribution has no correlation, i.e., they are asymmetric. Although this technique is simple to use, we found various scenarios where it is not able to accurately detect the power quality degradation. We classify the identified scenarios in two classes: 1) a malfunction device is not detected; and 2) a better PQ producing device is classified as a malfunction. Both cases are explained as follows.
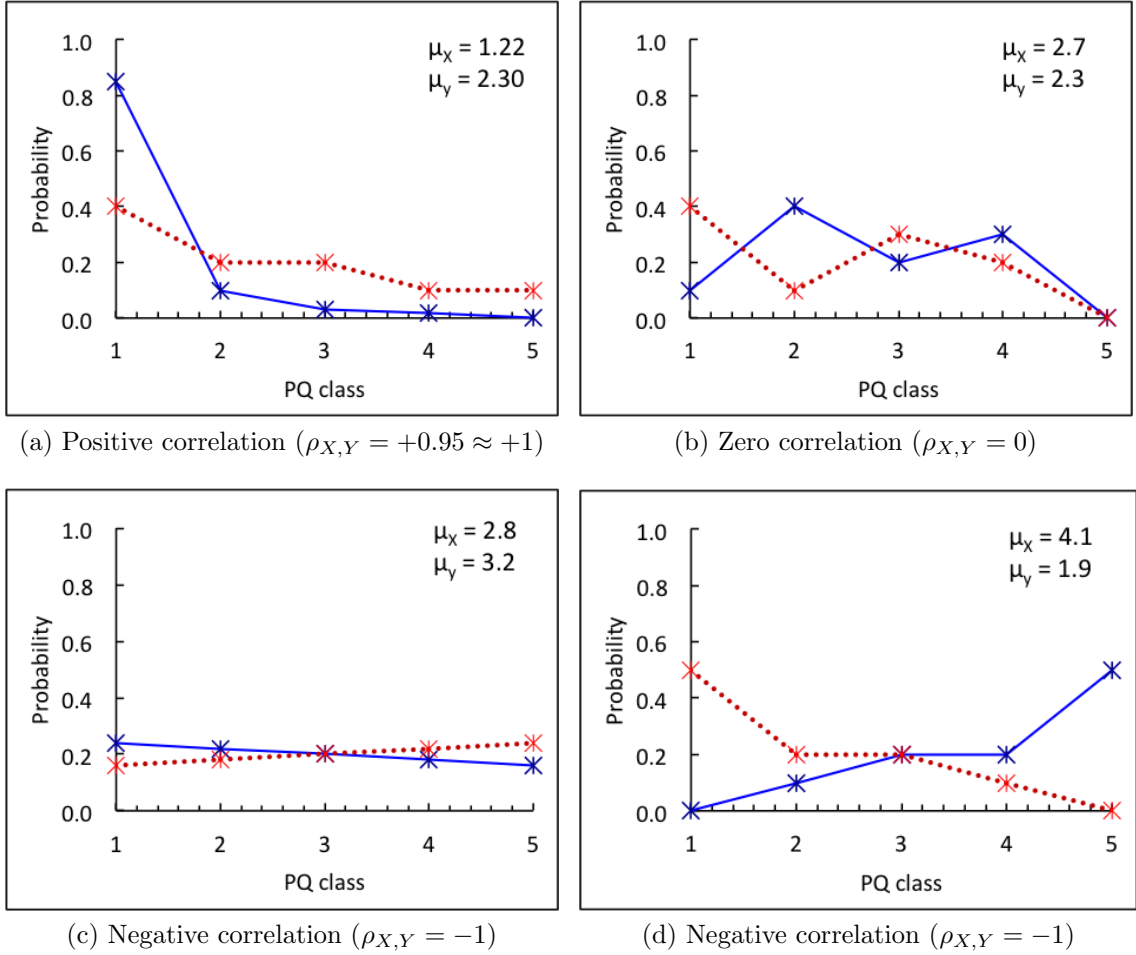
(a) Positive correlation ($\rho_{X,Y} = +0.95 \approx +1$)   (b) Zero correlation ($\rho_{X,Y} = 0$)

(c) Negative correlation ($\rho_{X,Y} = -1$)   (d) Negative correlation ($\rho_{X,Y} = -1$)

Figure 1.2: Various PQ distribution scenarios where the simple correlation technique is not useful.

Missed Detection Scenarios

This scenario arises when a device is producing a very bad power quality (malfunctioning) and the system does not detect it. The system will classify the observed distribution as perfectly normal (similar to the actual) when the correlation is $+1$ (or close to $+1$). For all normal behaving devices, the observed distribution will be very similar to the actual and a $+1$ correlation will perfectly classifying them as normal. Nevertheless, we identified cases where the PQ is significantly degraded and the distributions were highly correlated. In all those cases, we cannot use the simple correlation formula. Figure 1.2a shows one such example where the power quality is significantly degraded and the correlation coefficient is a $+1$.

False Detection Scenarios

We classify a detection as false when its power quality output is similar (or slightly degraded) to its actual distribution. Using the correlation measure, a device is detected as malfunction when the correlation coefficient is smaller (close to 0 or smaller than 0). Although we observed that this measure works in various cases, we have also been able to identify few scenarios where the simple correlation measure miss-classify devices as malfunction. These scenarios are explained in two categories as follows:

1. **Zero Correlation:** A zero correlation means the two distributions do not share any particular correlation. In our case, most probably, observing a distribution which do not share any correlation implies degradation in the power quality. But we cannot tell for sure if the observed distribution is good or bad. Figure 1.2b represents a scenario where the observed power quality is better and we can not classify the device as a malfunction simply based on a zero correlation.

2. **Negative Correlation:** In a negative correlation between two variables, the value of one variable increases while the other decreases, and vice versa. Although theoretically possible, it is rare to get a perfectly negative correlation in power quality measures. In a negative correlation, the observed distribution is either bad or good compared to the actual distribution. Although the probability of getting a bad PQ is higher than that of a good, it cannot always be guaranteed.

   Further, although the negative correlation can quantify the symmetry of the distribution, in cannot essentially quantify how much the distribution are different. For instance, Figure 1.2c shows a perfect -1 correlation but the actual PQ degradation is very minor. Finally, for the sack of theoretical proof, Figure 1.2d demonstrate a significant PQ improvement as a power quality degradation.

1.3.2   An average PQ measure

This measure is based on average power quality. A device is classified as a malfunction if its PQ output is significantly deviated from its normal average power quality. Clearly, if a device start malfunction by producing bad PQ output, its average output will deviate from its actual average output.
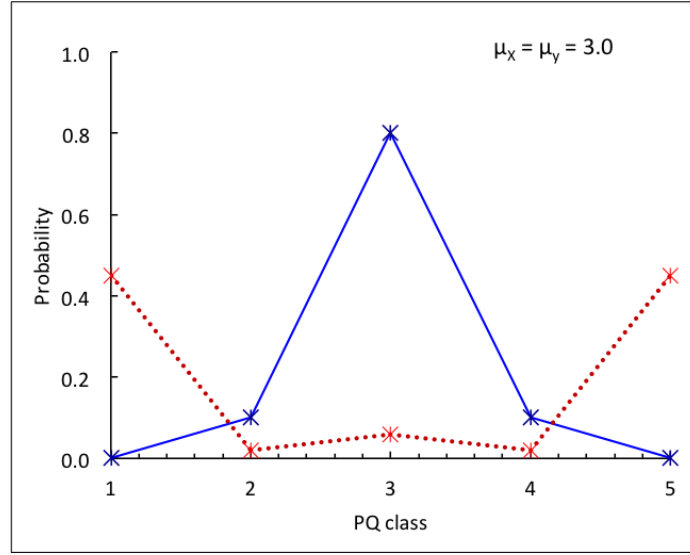
Figure 1.3: Two very different power quality distributions with same expected/average PQ output/class.

$$\Delta_d = E[f_x(d)] - E[\widehat{f}_x(d)] \leq \theta_d$$

where θ_d is the threshold/maximum value allowed at which a device is classified as malfunction. In other words, θ_d is the maximum allowed deviation between the expected PQ classes of the two distributions. The value of θ_d could be fixed for all devices, for instance, by using $\theta_d = 1$, any device whose average output is varied by 1 class from its actual average output will be classified as a malfunction device. We use θ_d equal to standard deviation of the observed output distribution, i.e., $\theta_d = \sigma_d$. Using standard deviation as threshold will allow us to use slightly larger/relaxed threshold for devices generating less uncertain outputs in normal conditions (compared to other devices in the same network). Figure 1.1 shows the expected value and corresponding threshold values form various devices in our sample power grid.

It is theoretically possible to produce two very different PQ distributions with same average output PQ. One such scenario is demonstrated in Figure 1.3. Although it is debatable whether the observed distribution in Figure 1.3 is acceptable as normal or not, we take a safer approach and address the problem by proposing a composite measure.

Algorithm 1: A malfunction device detection algorithm

Input: distribution function of input link to device 1 i.e., $f_x^{(0)}$, transfer function
$f(d)$, network topology, positions of the installed power meters M

Output: L (list of devices detected as malfunction)

begin

    foreach *(device d)* **do**

        $\hat{d} \leftarrow getParent(d);$

        /* Note that device 1 (root of the tree) has no parent i.e., *getParent(1) = 0*

        and $f_x(0)$ is given */

        $f_x(d) \leftarrow f_x(\hat{d}) \times f(d);$

        $H(d) \leftarrow -\sum_{i=1}^{n} p_{c_i}^{(d)} \log p_{c_i}^{(d)};$

        /* where $p_{c_i}^{(d)}$ is the ith component of $f_x(d)$ vector */

    end

    /* get N high entropy devices in vector H */

    $L \leftarrow getHighEntopyDevices(H, N)$

end

1.3.3 A composite measure

As discussed earlier, it is very clear the average power quality ($E[\widehat{f}_x(d)]$) will degrade (higher PQ class) when a device produce significantly bad power quality than its actual average power quality. Simply measuring the difference between expected values of actual, and observed PQ classes can detect if a device is a malfunction. Although this simple technique works well, theoretically it is possible that the actual and a very different observed PQ output distributions may produce similar expected value. To tackle this problem, we propose a composite expected value measure which guarantees the true detection of a malfunction device in terms of power quality degradation.

In the composite expected value measure, we calculate two separate expected values: 1) expected value of the complete distribution; and 2) expected value of the PQ classes greater than the average class computed on complete distribution. The second measure allow the algorithm to detect a potential malfunction device not detected by the first measure. In other words, for a malfunction device, if the observed expected value is similar to the actual expected value, the probability values for higher classes must be greater (or the device is not a malfunction). Our proposed algorithm is shown as Algorithm 1.

for every sample -¿ check if f(d) varies by f1(d)

for the detected segment, use maxent to estimate f(d1) to f(dn)

sort them by descending order of variation

the top one is high probable.

bad power quality events

unit time interval

window size

an interval is bad quality interval when it is classified as ¿= 4

threshold number of bad quality intervals

misbehaving device on the detected link

candidate links are

use maxent

using input pdf simulation dataset was generated

we use f(d) as in chapter 3

If no smart meter installed on that device, need to estimate from neighboring links

- The metered link detects deviation and estimate the deviations of all devices between itself and parent metered link

1.4   Evaluations

simulation setup dataset generation Performance Assume that one or more smart meters have continuously measured poor power quality. The power quality is indicated by some specific power quality index, for instance the System Average RMS Variation Frequency (SARFI) index . Based on the measured and estimated PQ values, we would like to know which device is most likely to be malfunctioning. Here, we assume that we know the transform function of each device if the device is working properly. We call these transfer functions the regular transform functions of the devices. Now, a malfunctioning device is the one whose estimated transform function deviates significantly from its regular transform function.

We plan to first give a comprehensive mathematical model of the problem. We then device a mechanism to calculate the deviation in transfer functions and finally we plan to propose an algorithm to efficiently identify any potential malfunctioning device in the power network. The proposed algorithm will be evaluated on various tree and bus structured network; specifically we will use the IEEE standard test networks for our simulations.

1.5   Conclusion

Power quality meters play an important role in the reliability of power networks. We plan to investigate the problem of identifying devices that degrade the power quality in the system. The proposed solution will be built on top of our already proposed work.