

University of Tennessee, Knoxville
Trace: Tennessee Research and Creative Exchange

Masters Theses | Graduate School

8-2014

Statistical Analysis of Disturbances in Power Transmission Systems

Liu Liu
University of Tennessee - Knoxville, lliu25@vols.utk.edu

This Thesis is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Masters Theses by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a thesis written by Liu Liu entitled "Statistical Analysis of Disturbances in Power Transmission Systems." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Computer Engineering.

Hairong Qi, Major Professor

We have read this thesis and recommend its acceptance:

Yilu Liu, Leon M. Tolbert

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Statistical Analysis of Disturbances in Power Transmission Systems

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Liu Liu

August 2014

© by Liu Liu, 2014
All Rights Reserved.

This thesis is dedicated to my always hardworking mother, and to the memory of my father who always believed in me.

Acknowledgements

First of all, I would like to express my enormous gratitude towards my advisor Dr. Hairong Qi for her continuous support and constant inspirations for my research and study, for the immense knowledge, incredible patience, motivation and encourage she gives me all the time.

Also I want to thank the rest of my committees: Dr. Leon M. Tolbert, Dr. Yilu Liu for their insightful comments, continuous support and patience.

My sincere thanks also goes to my fellow labmates in Advanced Imaging and Collaborative Information Processing (AICIP) group: Wei Wang, Li He, Rui Guo, Shuangjiang Li, Jiajia Luo, Zhibo Wang, Yang Song, Cristian Capdevila, Alirea Rahimpour. I cannot finish my work without your generous help and surprisingly great master of knowledge. Meanwhile, I want to thank Jidong Chai, Zhuohong Pan, Dr. Changgang Li, Yin Lei, Jiahui Guo from Dr. Yilu Liu's group, for their patience and generous help as well.

Last but not least, I want to express my appreciation to my family, far away in China, for their enormous support and completely selfless love.

Per aspera ad astra

Abstract

Disturbance analysis is essential to the study of the power transmission systems. Traditionally, disturbances are described as megawatt (MW) events, but the access to data is inefficient due to the slow installation and authorization process of the monitoring device. In this paper, we propose a novel approach to disturbance analysis conducted at the distribution level by exploiting the frequency recordings from Frequency Disturbance Recorders (FDRs) of the Frequency Monitoring Network (FNET/GridEye), based on the relationship between frequency change and the power loss of disturbances - linearly associated by the Frequency Response. We first analyze the real disturbance records of North America (1992 to 2009) and confirm the power law distribution; we discover that small disturbances are log-normal distributed. Then based on the real records from 2011 to 2013 (EI), the disturbances in megawatt and the corresponding frequency change records are studied in parallel. We prove that the frequency change of disturbances and its megawatt records share similar power law distribution when the disturbances are large; the frequency change can be delineated by a log-normal distribution with its numerically approximated coefficient when the disturbances are small.

Meanwhile, activities like FIDVR in the power systems reflected as voltage signature patterns recorded at the transmission level are worth studying since each pattern corresponds to a certain type of behavior. Pattern recognition is used in this problem. Initially the records are preprocessed through eliminating ineligible records and rescaling. Feature extraction is applied to obtain a better

representation of signature dataset by statistics of amplitude, wavelet transform and Fourier transform. With the extracted features, k-means, an unsupervised clustering algorithm is exploited to generate root patterns; furthermore we use heuristic selection to remove the mis-classified patterns. The extracted root patterns then serve as training dataset to train a support vector machine (SVM). After the parameters of kernel function in SVM is optimized, a subset of voltage signature records is generated as testing dataset, based on which the performance of SVM is evaluated. With all patterns we achieve an accuracy of 80.12% of multi-label classification; and if only considering dominant patterns, the accuracy reaches 86.20%.

Table of Contents

1 Introduction **1**

2 Distribution-Level Disturbance Analysis Using Frequency Information **6**

2.1 Analysis of Real Power Grid System Data 6

2.1.1 Extended Analysis of Megawatt Records of Disturbances . . . 6

2.1.2 Disturbances in Megawatts vs. Disturbances as Frequency Change 12

2.2 Theoretic Proof 16

2.2.1 Large Disturbances Modeling 16

2.2.2 Small Disturbance Modeling 18

2.3 Stochastic Process of Frequency Response 20

3 Voltage Signature Pattern Recognition in Power Grid **22**

3.1 Feature Extraction 23

3.1.1 Preprocessing 24

3.1.2 Amplitude Statistics 24

3.1.3 Fourier Transform Statistics 25

3.1.4 Wavelet Statistics 26

3.2 Classification 28

3.2.1 Root Patterns Extraction 28

3.3 Supervised Classification 31

3.3.1 Support Vector Machine 31
3.3.2 Parameter Selection for SVM 35
3.3.3 Multi-label Classification 37
3.4 Performance Evaluation 38

4 Conclusions and Future Works **41**

Bibliography **43**

Appendix **47**

Vita **52**

Chapter 1

Introduction

The power grid system is a complex system which includes numerous generators, consumers and connections. Taking the Eastern Interconnection (EI) as an example, it contains more than 15,000 generators and hundreds of thousands of miles of transmission and distribution lines Wang et al. (2013). Such complexity makes the grid prone to faults due to either operations, natural disasters or malicious attacks, to name a few; leading to potentially cascading disturbances or even blackouts. One of the most devastating disturbances is the August 2003 massive power outage took place throughout parts of the Northeastern and Midwestern United States and part of Canada Andersson et al. (2005), affecting an estimated population of over 50 million totally.

Traditionally the disturbances are recorded as power loss in power grid and we usually focus on the megawatt (MW) sized disturbances, corresponding to the frequency change with the magnitude of about 10^{-2} Hz. Recent studies on the statistical modeling of the disturbance size (in megawatt) share similar findings that the disturbance size follows a power-law tail distribution Carreras et al. (2004); Chen et al. (2001); Holmgren and Molin (2006); Sachtjen et al. (2000). In fact, Power law distribution can be found in a wide variety of physical, biological, and man-made phenomena Newman (2005), like the sizes of earthquakes, the foraging pattern

of various species. Similar to the disturbance size in the power grid, few empirical distributions fit a power law throughout whole data set, but rather follow a power law in the tail. That is, the data set only follows the power law distribution when its magnitude is large enough. We will show that for the smaller disturbance, its size is actually log-normal distributed.

Although power loss is a good indicator of power grid stability, it is difficult to estimate and usually takes a long time to obtain the data, making it nearly impossible to perform real-time analysis for disturbances. On the other hand, frequency dynamics, as another important parameter in the power system, usually can reflect power grid dynamics accurately. Frequency information can be collected by wide area measurement system (WAMS) using PMU (Phasor Monitoring Unit) or other sensors placed at the transmission level across North America. Although they monitor the grid more stably, the major drawbacks are the high cost and delay in installation and difficulty in data access. The former two lead to coarse resolution in monitoring and the latter causes the inefficiency of data analysis. Consequently they cannot provide an efficient surveillance in large scale. Therefore it is desirable to deploy a measurement network that is economical such that dense deployment would be possible. As a member of the PMU family, the frequency disturbance recorder (FDR) was developed at Virginia Tech in 2003 Qiu et al. (2001). At the distribution level, FDR collects instantaneous phasor and frequency information taken at low voltage level from the 120 V wall outlet and transmits the measured data remotely via the Ethernet Tsai et al. (2004); Gardner and Liu (2007); Zhong et al. (2005). The uniqueness of FDR is that it can measure essential transmission level information at the distribution level using widely distributed and low-cost sensors. Based on the FDRs, a US-wide Frequency Network (FNET/GridEye) has been implemented and some power system monitoring applications are being developed by taking full advantage of the FDRs. Thus far there are more than 200 FDRs installed in the United States and another 39 installed worldwide, as shown in Figure 1.1. It is worth noting that since the FNET/GridEye is deployed at the distribution level, it

potentially can collect more disturbances at local scale in the future extension of functionality, making it especially useful in the analysis of microgrid.

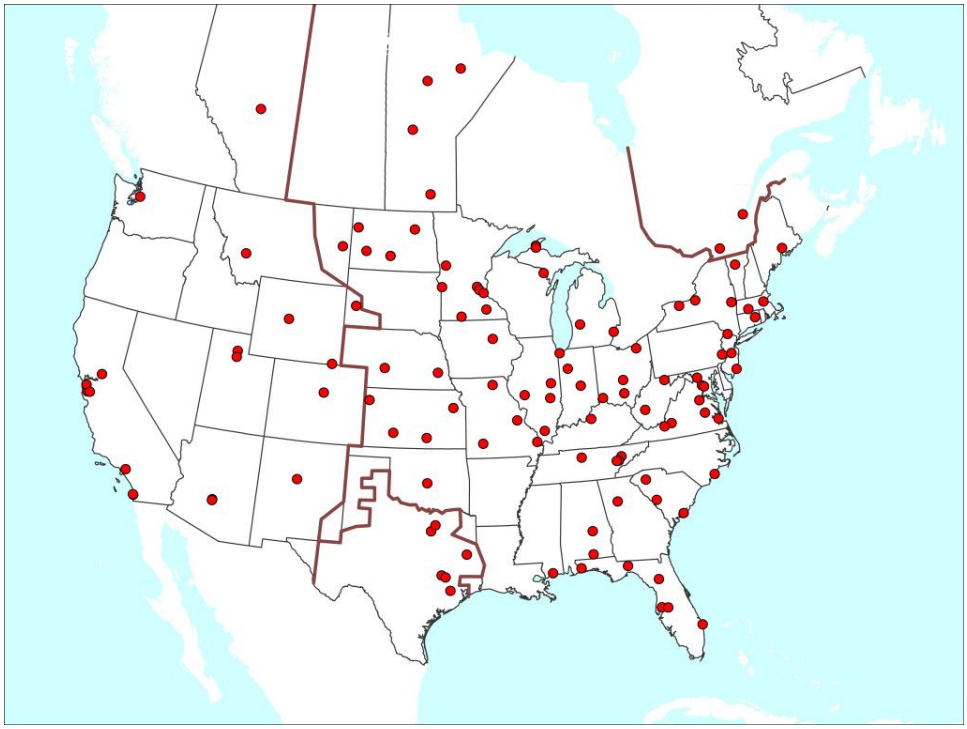


Figure 1.1: Deployment map of FDR in FNET/GridEye

The frequency change and the power loss of a certain disturbance are connected by introducing a quantity, Frequency Response, denoted as β. Generally β is treated as the ability of an Interconnection to react or respond to a change in frequency domain, expressed in megawatts per 0.1 Hz (MW/0.1 Hz) Lauby et al. (2013). Mathematically the frequency change and the power loss are linearly related, $P = \beta\Delta f$, and β itself is a stochastic process. Under certain assumptions, β can be approximated as a random variable, which will be addressed later. Note that originally $P = -\beta\Delta f$ since Δf is usually a negative quantity. We use the absolute value of P, β and Δf throughout the paper for the purpose of making analysis more convenient and straightforward. Both P and Δf can represent the size of a certain disturbance.

Meanwhile every now and then there are thousands or even more activities (not necessarily disturbances or faults), like switching, or fluctuations. Similarly they can

be picked by sensors like PMU at the transmission level or FDR at the distribution level. Among them certain voltage signature caused by the Fault-Induced Delay Voltage Recovery (FIDVR) attracts the attentions because it is important to the power system. The FIDVR is a phenomenon, whereby system voltage remains at significantly low levels, after a fault clearing in transmission, subtransmission or even distribution system, for several seconds, in which high load currents and large reactive power demands caused the delays up to tens of seconds for voltage to recover after a fault clearing (**?**). They are of concern because they show a temporary loss of voltage control in an area, and they pose a risk of cascading to a larger area, especially if another unexpected event occurs while the voltage is depressed. FIDVR is poorly recorded by PMU at the transmission level and FDR at the distribution level tends to record more.

Although there are countless voltage signature records sensed by PMU, many of which share similar patterns, which corresponds to certain types of disturbances. Consequently it is intriguing to recognize those patterns and then furthermore identify certain disturbances. There have been several studies on the topic of pattern recognition for voltage signature or disturbances in power system. Abdullah et al. (2007); Kang et al. (2010); Safty et al. (2004) all had discussed the pattern recognition for power-quality disturbances. Basically they performed feature extraction using either wavelet or frequency domain information. Then artificial neural network was used as the classifier to classifier the voltage (sinusoidal) waveform. One main drawback was that their experiment is established on the small scale simulation with no consideration of real-world adaptation. Furthermore they lack systematic evaluation of the performance of their classifier.

Based on the real-world data collected from Eastern Interconnection (EI), we are able to perform pattern recognition for the voltage signature with taking real-world scenario into account. And ultimately we are able to perform the on-line detection of activities occur in the power system. Furthermore, with more FDR records, it is possible to use the same approach to recognize FIDVR at the distribution level.

The contributions of the thesis are five-fold:

1. we extend the existing study of disturbance from the years 1984 through 1999 to the span of 1992 - 2009 and confirm the power-law tail characteristics of megawatt size disturbances; Furthermore the behavior of small disturbances is studied and characterized as a log-normal distribution.

2. we analyze the frequency change of disturbance, and discover the similar power-law tail distribution with the Frequency Response following a Gaussian distribution. The distribution of small disturbances also obeys a log-normal distribution with numerically approximated coefficient.

3. we theoretically prove the relationship between power loss and frequency change, and the possibility of studying disturbance at the distribution level, making the analysis more prompt and convenient.

4. we successfully extract 15 recognizable patterns of voltage signature from PMU records

5. we implement the multi-label SVM based on binary SVM and achieve high accuracy regarding multi-label classification of voltage signature patterns.

The thesis is organized as following: Chapter 2 discusses the statistical modeling for both megawatt size of disturbances and the corresponding frequency change; Chapter 3 focuses on the pattern recognition of voltage signature by implementing multi-label support vector machine (SVM); Chapter 4 draws the conclusion and discuss the future works.

Chapter 2

Distribution-Level Disturbance Analysis Using Frequency Information

2.1 Analysis of Real Power Grid System Data

Statistical analysis of disturbances in power grid is based on real-world records, and consists of two cases. The first case is the extended study on the megawatt (power loss) records of disturbances from the year 1992 to 2009. The frequency change, Frequency Response and power loss of disturbances are studied in the second case.

2.1.1 Extended Analysis of Megawatt Records of Disturbances

Although there have been some studies on the power-law tail distribution fitting for large disturbances, it is worthwhile validating such statistical characteristics of disturbances on recent years' disturbance records. As previous study Chen et al. (2001) pointed out, power-law distribution has been found in the disturbances in the power system as well as other natural or artificial phenomena like the

sizes of earthquakes, craters on the moon, etc, as illustrated in Figure 2.1. The complementary cumulative distribution function (complementary CDF, or CCDF) is used to study the statistical characteristics of the size of disturbance because that only limited data can be obtained and that studying the histogram or the probability density function (PDF) tends to include great discontinuity and fluctuations. When the disturbance size is large enough, i.e., greater than a certain boundary, denoted as x_0, its distribution follows the power-law tail, denote X as the disturbance size and the unit is megawatt (MW), $\exists x_0 \in \mathbb{R}^+, s.t., X > x_0$, we have

$$F_X(x) = P(X \leq x),\ x \geq x_0. \tag{2.1}$$

The corresponding probability density function is

$$p(x) = C_0 x^{-\alpha}, \quad x \geq x_0,\ \alpha > 1. \tag{2.2}$$

where C_0 and α are the coefficients for the power law distribution. Then the cumulative distribution function is

$$\begin{aligned} F_X(x) &= \int_{x_0}^{x} C_0 x^{-\alpha}\,\mathrm{d}x \\ &= C_0 \left(\frac{1}{-\alpha+1} x^{-\alpha+1} \right) \Bigg|_{x_0}^{x} \\ &= \frac{C_0}{\alpha-1} x_0^{-\alpha+1} - \frac{C_0}{\alpha-1} x^{-\alpha+1} \end{aligned} \tag{2.3}$$

Since $F_X(x)$ is CDF, it must follow $\lim_{x\to\infty} F_X(x) = 1$. Therefore,

$$\frac{C_0}{\alpha-1} x_0^{-\alpha+1} = 1 \Rightarrow C_0 = \frac{\alpha-1}{x_0^{-\alpha+1}} \tag{2.4}$$

As a result, there is only one free coefficient to be determined. Still, we keep C_0 and x_0 in further analysis for simplicity in format.

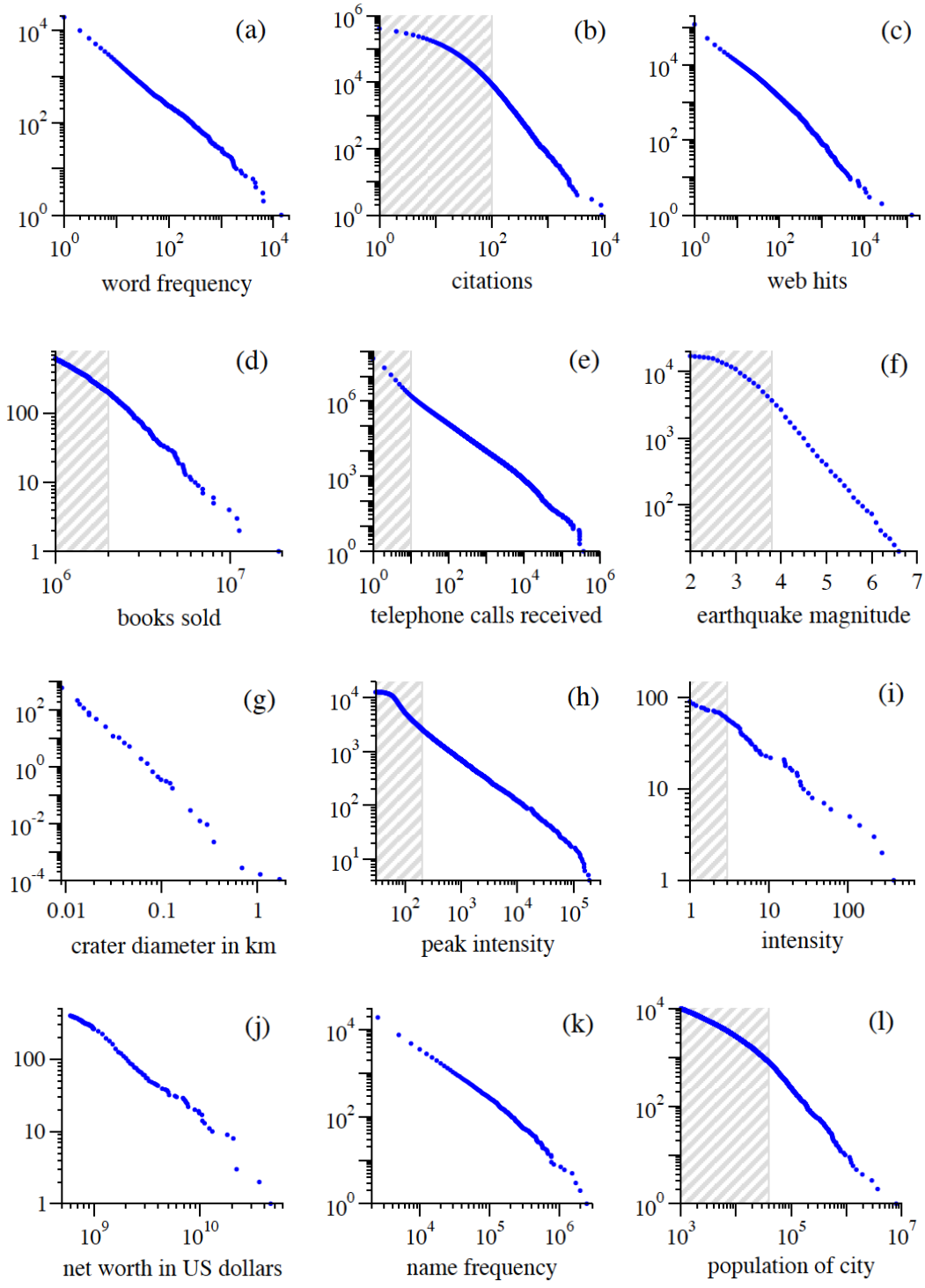


Figure 2.1: Power law in natural or man-made phenomena

The complementary CDF can be represented as

$$\begin{aligned} F_X^C(x) &= P(X > x) = 1 - P(X \leq x) \\ &= 1 - F_X(x) \\ &= \frac{C_0}{\alpha - 1} x^{-\alpha+1} \\ &= \lambda_0 x^{-\gamma} \end{aligned} \tag{2.5}$$

where $\lambda_0 = \frac{C_0}{\alpha-1}$ and $\gamma = \alpha - 1$

Taking logarithm of both sides of Eq. 2.5,

$$\log(F_X^C(x)) = -\gamma \log x + \log \lambda_0 \tag{2.6}$$

Eq. 2.6 indicates that the log-log plot for power-law tail distribution should be a linear function, which is demonstrated in Figure 2.2.

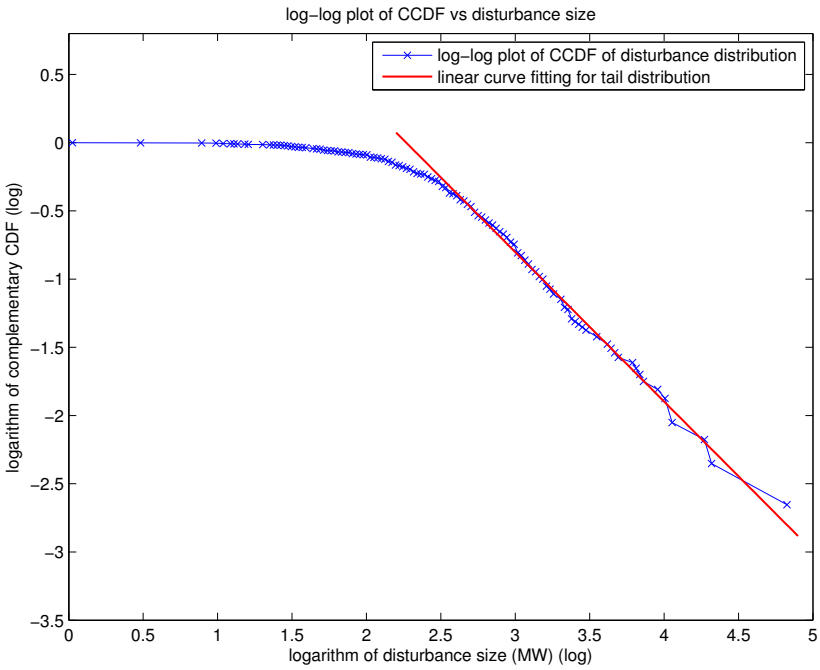


Figure 2.2: Log-log plot of complementary CDF of disturbance and the linear tail curve fitting, 1992 - 2009

We observe that when the size of disturbance is large enough, it obeys power-law distribution. In order to more systematically analyze the disturbance in the power grid system, it is desired to derive a statistical model for small disturbances. It is found that the small disturbances actually obeys a *log-normal* distribution, that is, the disturbance size is a random variable whose logarithm is normally distributed. Denote X_s as the random variable for the megawatt size of disturbances. When the size is small enough, i.e., smaller than the aforementioned boundary x_0, the PDF is

$$f_{X_s}(x;\mu_s,\sigma_s) = \frac{1}{x\sigma_s\sqrt{2\pi}} e^{-\frac{(\ln x-\mu_s)^2}{2\sigma_s^2}}, \quad 0 < x \leq x_0 \tag{2.7}$$

where μ_s and σ_s are the mean and variance, respectively, of the variable x on a logarithmic scale.

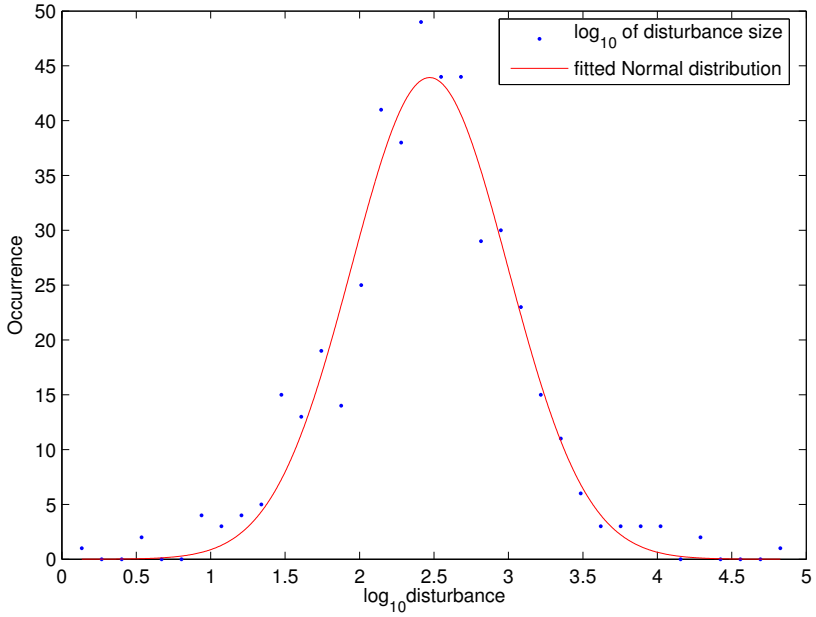


Figure 2.3: PDF of disturbance on a logarithmic scale - log-normal distribution, 1992 - 2009

The probability density function (PDF) is demonstrated in Figure 2.3. It can be observed that the disturbance size on a logarithmic scale follows a normal (or

Gaussian) distribution, indicating that the disturbance size follow a log-normal distribution.

In order to compare the difference between the distributions of small and large disturbances, i.e., log-normal distribution and power-law tail distribution, the complementary CDF of log-normal distribution is derived,

$$F_{X_s}^{C}(x) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\ln x - \mu_s}{\sigma_s\sqrt{2}}\right)\right] \tag{2.8}$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^{\infty} e^{-t^2}\text{d}t$ is the error function.

Figure 2.4 illustrates the log-log curve of disturbances, log-normal distribution for small disturbances and power law distribution for large disturbances. Furthermore, the performance of the fitting model is evaluated using the *normalized root mean square error* (NRMSE) is used,

$$NRMSE = \frac{RMSE}{x_{\max} - x_{\min}} = \frac{\sqrt{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2/n}}{x_{\max} - x_{\min}} \tag{2.9}$$

where x is the original curve or dataset, and $\hat{x}$ is the fitted curve. The NRMSE for log-normal distribution and power law distribution are $NRMSE_{power/log-normal} = 9.21\%$ and $NRMSE_{power/powerlaw} = 5.46\%$, respectively. Note that the log-normal distribution curve intersects with the linear log-log curve of power law at intersection A. It can be considered as the boundary x_0 of small disturbances and large disturbances, which roughly is $x_0 = 10^{2.7} \approx 500MW$. Therefore, the disturbance whose size in megawatt is smaller than intersection A $(500MW)$ is log-normal distributed, otherwise it is power law distributed.

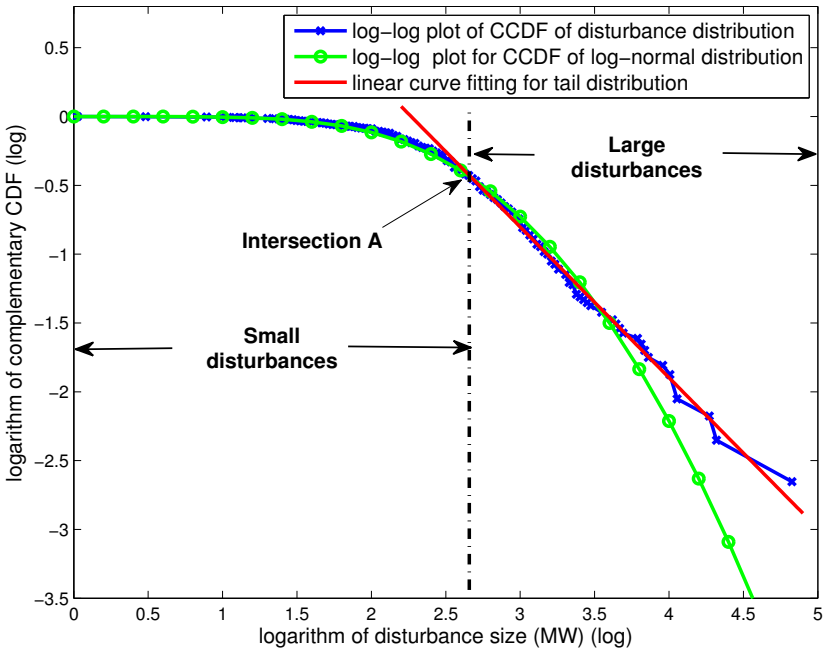


Figure 2.4: Complementary CDF of disturbance size, log-normal vs. power-law tail distribution, 1992 - 2009

We can conclude that the megawatt size of disturbances can be delineated with log-normal distribution for small disturbance and power law distribution for large disturbances. The boundary for small and large disturbances can be determined by the intersection of those two distributions. Furthermore the boundary is sample dependent, that is, if fewer smaller disturbances are included, the boundary x_0 will increase.

2.1.2 Disturbances in Megawatts vs. Disturbances as Frequency Change

Since megawatt size of disturbances and the frequency change are associated by Frequency Response, β, it is intriguing to study disturbance in megawatt and the frequency change in parallel. The real-world records of disturbances in both megawatt

and the corresponding frequency change records from Jan. 2011 to Jun. 2013 of EI are studied here.

Frequency Response

Frequency Response β is calculated by $\beta = \frac{P}{\Delta f}$, whose probability density distribution is plotted in Figure 2.5. This is consistent with Lauby et al. (2013) that Frequency Response is Gaussian/normal distributed. The mean for β based on this dataset is $2443MW/0.1Hz$, and the standard deviation is $550MW/0.1Hz$. The performance of fitting model measured by the aforementioned NRMSE is $NRMSE = 6.24\%$.

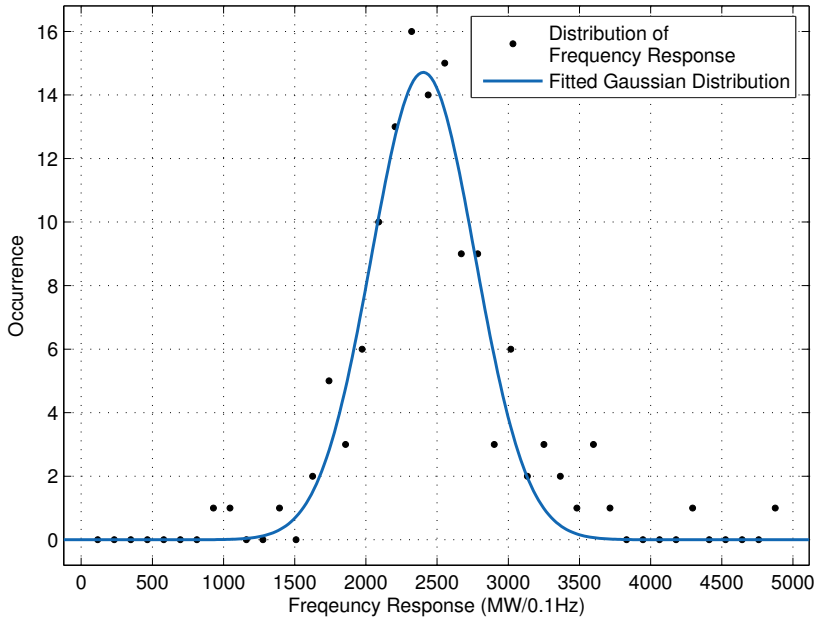
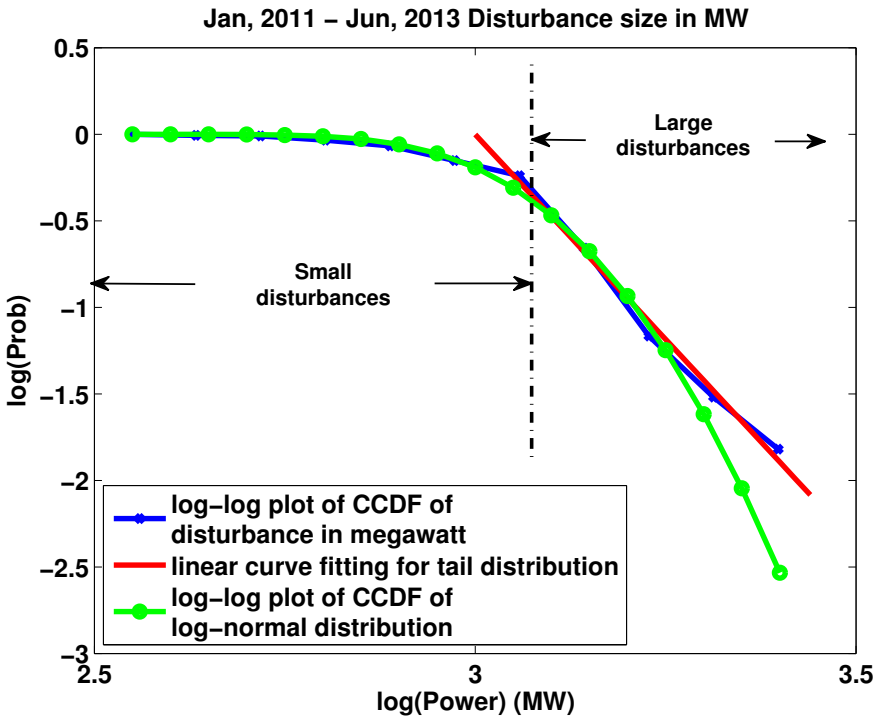


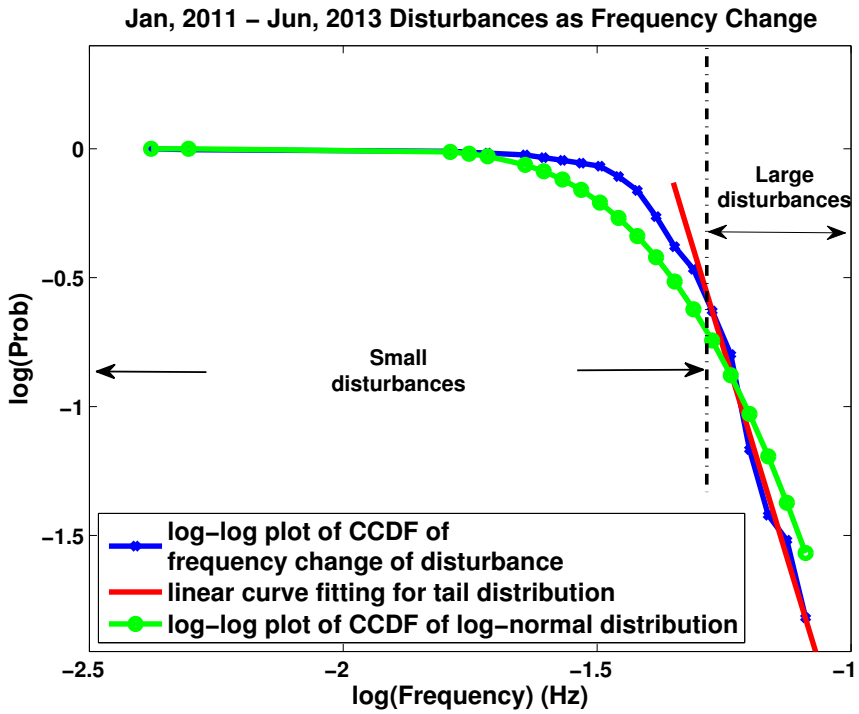
Figure 2.5: Distribution of frequency response β and fitted Gaussian distribution, Jan.2011 - Jun. 2013

Disturbances in megewatts & Frequency Change of disturbances

Following the same procedure, the complementary CDF is generated based on the EI records. Figure 3.2(m) demonstrates the fitted complementary CDF for disturbances in megawatt.



(a) Statistical fit for disturbance in MegaWatts



(b) Statistical fit for disturbance as frequency change

Figure 2.6: Statistical models for disturbance records: Jan. 2011 - Jun. 2013

It can be observed that the boundary between small and large disturbance in power domain is approximately $10^{-3.1} = 1,250MW$. Note the significant difference between the boundary derived using records from the year 1992 to 2009 and that from 2011 to 2013 of EI. It is because the threshold for disturbances that can be included in the former study is $1MW$ while the threshold for the latter study is $368MW$, indicating that there are fewer smaller disturbances included in the latter study. Therefore the consideration of small or large disturbances and thus the derivation of the statistical models are highly related to the boundary of disturbances size. The performance of the fitting model is also evaluated by $NRMSE$ with $NRMSE_{power/log-normal} = 15.4\%$ for small disturbances and $NRMSE_{power/powerlaw} = 7.65\%$ for large disturbances. The performance is generally worse than that of disturbances records between 1992 - 2009 mainly because of the limited records (only 132 records).

We also attempt to use the log-normal distribution to model small disturbances and power law distribution to model large disturbances for the frequency change of disturbances in EI. Figure 3.2(o) illustrates the statistical fitting for small disturbances ($\leq 0.045Hz$) and large disturbances ($> 0.045Hz$). And the performance of the fitting model $NRMSE_{freq/log-normal} = 76.75\%$ for small disturbances and $NRMSE_{freq/powerlaw} = 9.96\%$ for large disturbances.

It can be observed that the power law fits the frequency change of large disturbances well, while the log-normal distribution fails to provide a convincing model for the small disturbances. Besides the reason of limited records from FNET/GridEye, there is a more fundamental reason which will be addressed in the next section.

2.2 Theoretic Proof

In this section, we provide theoretic proof.

2.2.1 Large Disturbances Modeling

Since the megawatt size and the frequency change of disturbances can be associated with

$$P = \beta \Delta f, \quad P, \beta, \Delta f > 0 \tag{2.10}$$

we will prove that with β following the Gaussian distribution and P follows power-law tail distribution, the distribution of the ratio, i.e., distribution for frequency change Δf, also follows power-law tail distribution.

Assume that power P, frequency response β and frequency change Δf are all independent random variables, and

$$\begin{aligned} P \sim C_0 x^{-\alpha}, \quad \beta \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \\ \Delta f \sim C_f z^{-\tau}, \, \tau > 1 \end{aligned} \tag{2.11}$$

Denote X, Y and Z as the random variable for P, β and $Deltaf$, respectively. We have $X = YZ$, then $Z = \frac{X}{Y}$, and the cumulative distribution function (CDF) for Δf can be expressed as,

$$\begin{aligned} F_Z(z) &= P(Z < z) = P(\frac{X}{Y} < z) \\ &= P(Y > 0, X < Yz) P(Y > 0) \\ &\quad + P(Y < 0, X > Yz) P(Y < 0) \\ &= P(Y > 0, X < Yz) \\ &= \int_0^\infty p_Y(y) \left[\int_{-\infty}^{yz} p_X(x) \mathrm{d}x \right] \mathrm{d}y \end{aligned} \tag{2.12}$$

since $X, Y, Z > 0$ and are independent.

From Eq. 2.12, the PDF for Z can be derived which is the derivative of its CDF,

$$\begin{aligned} p_Z(z) &= \frac{\mathrm{d}F_Z(z)}{\mathrm{d}z} \\ &= \int_0^\infty p_Y(y) \frac{\mathrm{d}}{\mathrm{d}z} \left(\int_{-\infty}^{yz} p_X(x) \mathrm{d}x \right) \mathrm{d}y \\ &= \int_0^\infty y \cdot p_X(yz) p_Y(y) \, \mathrm{d}y \end{aligned} \tag{2.13}$$

Substituting Eq. 2.13 with Eq. 2.11,

$$\begin{aligned} p_Z(z) &= \int_0^\infty y C_0 (yz)^{-\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, \mathrm{d}y \\ &= z^{-\alpha} \cdot \frac{C_0}{\sqrt{2\pi}} \int_0^\infty y^{-\alpha+1} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, \mathrm{d}y \end{aligned} \tag{2.14}$$

Note that the integral in Eq. 2.14 is a definite integral. Now we prove the definite integral is integrable. Consider the Taylor series of $y^{-\alpha+1} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$,

$$\begin{aligned} \frac{e^{\frac{(y-\mu)^2}{2\sigma^2}}}{y^{\alpha-1}} &= \frac{1 + \frac{(y-\mu)^2}{2\sigma^2} + \frac{(y-\mu)^4}{2!(2\sigma^2)^2} + \cdots + \frac{(y-\mu)^{2k}}{k!(2\sigma^2)^k} + \cdots}{y^{\alpha-1}} \\ &= \frac{1}{y^{\alpha-1}} + \frac{(y-\mu)^2}{2\sigma^2 y^{\alpha-1}} + \cdots + \frac{(y-\mu)^{2k}}{k!(2\sigma^2)^k y^{\alpha-1}} + \cdots \end{aligned} \tag{2.15}$$

Then $\exists k_0 \in \mathbb{Z}^+$, s.t. $2k_0 - \alpha > 1$. Thus,

$$\begin{aligned} \frac{e^{\frac{(y-\mu)^2}{2\sigma^2}}}{y^{\alpha-1}} &= \frac{1}{y^{\alpha-1}} + \cdots + \frac{1}{k!(2\sigma^2)^k} \left(y^{2k_0-\alpha+2} + \cdots \right) + \cdots \\ &> \frac{1}{k_0!(2\sigma^2)^{k_0}} \left(y^{2k_0-\alpha+2} \right) \\ &> \frac{1}{k_0!(2\sigma^2)^{k_0}} y^2 \end{aligned}$$

Indicating

$$y^{1-\alpha} e^{-\frac{(y-\mu)^2}{2\sigma^2}} < \frac{1}{k!(2\sigma^2)^k y^2}$$

Therefore,

$$\int_0^\infty y^{\tau-1} e^{-\frac{(y-\mu)^2}{2\sigma^2}} < \frac{1}{k!(2\sigma^2)^k} \int_0^\infty \frac{1}{y^{-2}} = C < \infty$$

where C is a positive constant. As a result, Eq. 2.14 becomes,

$$p_Z(z) = C_f x^{-\alpha}, \tag{2.16}$$

if we allow

$$\begin{aligned} C_f &= \frac{C_0}{\sqrt{2\pi}\sigma} \int_0^\infty y^{\alpha-1} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \mathrm{d}y, \\ \tau &= \alpha \end{aligned} \tag{2.17}$$

we obtain exactly the same expression as Eq. 2.2, a power law distribution.

2.2.2 Small Disturbance Modeling

So far we have proved that for large disturbances, both the frequency change Δf and power P have the power-law tail distribution while the frequency response β follows Gaussian distribution, that is, theoretically the analysis of frequency change of disturbances is equivalent to the power of disturbances for large-size disturbances.

When the disturbance size is fairly small, we show that the ratio distribution, i.e., $\Delta f = P/\beta$ is not exactly log-normal distributed as P, which is in contrast to the scenario of large disturbances.

Recall the PDF for Δf is Eq 2.13. Substituting Eq 2.7 for small disturbance size and Eq 2.11 for β into Eq 2.13,

$$
\begin{aligned}
p_Z(z) &= \int_0^\infty \frac{C_0}{2\pi z\sigma_s\sigma} e^{-\frac{(\ln(yz)-\mu_s)^2}{2\sigma_s^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} \,\mathrm{d}y \\
&= \int_0^\infty \frac{C_0}{2\pi z\sigma_s\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\
&\quad \cdot e^{-\left[(\ln y-\mu_s)^2+(\ln z-\mu_s)^2+(2\ln y\ln z-\mu_s^2)/2\sigma_s^2\right]} \,\mathrm{d}y \\
&= \underbrace{\frac{1}{z\sigma_s\sqrt{2\pi}} e^{(\ln z-\mu_s)^2}}_{\text{log-normal distribution}} \\
&\quad \underbrace{\int_0^\infty \frac{C_0}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}-\frac{(\ln y-\mu_s)^2+(2\ln y\ln z-\mu_s^2)}{2\sigma_s^2}} \,\mathrm{d}y}_{\text{coefficient}}
\end{aligned}
\tag{2.18}
$$

As it can be observed in Eq.2.18 that the frequency change of small disturbance is supposed to be similarly distributed as is for megawatt records (Note that the coefficient is integrable since comparing to the integral in Eq 2.14, the exponential decays much faster than polynomial). Nevertheless the term $e^{-2\ln y\ln z/2\sigma_s^2}$ in the integral renders the coefficient associated with z, leading to the difference between the actual distribution and the log-normal distribution. This explains the deviation from the standard log-normal distribution observed in Figure 3.2(o). Although there is no closed form for the coefficient, it still can be approximated by numerical integration. Figure 2.7 shows the log-normal distribution with the numerical approximation of its coefficient, and the performance of the fitted model is $NRMSE_{freq/log-normal} = 16.12\%$, much lower than the original log-normal distribution fitting 76.75%.

So far it can be concluded that the disturbances recorded as frequency change reveals the statistical characteristics as those recorded in megawatts. For large disturbances, the frequency change records of disturbances share the power-law tail distribution as the megawatt records; while for small disturbances, frequency change follows a log-normal distribution that can be approximated by numerical integration using Eq. 2.18.

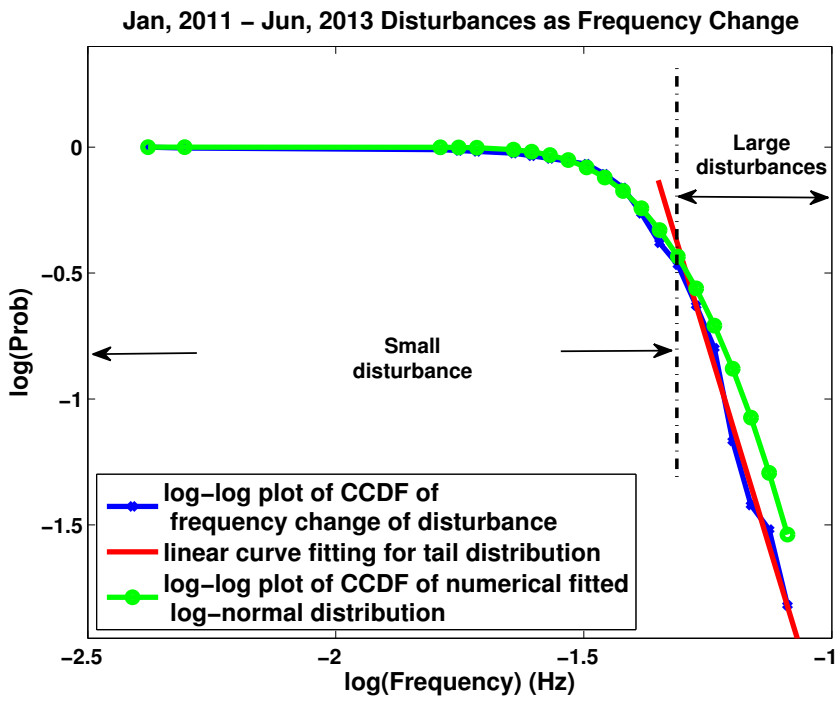


Figure 2.7: Numerical approximation for small disturbances: Jan. 2011 - Jun. 2013

2.3 Stochastic Process of Frequency Response

According to Lauby et al. (2013); Ingleson and Allen (2010), Frequency Response, or the calculated beta for a Balancing Authority is based on measuring a relatively small change in Net Actual Interchange coincident with a frequency excursion. Therefore frequency response is greatly influenced by contributing generators and motors, and changes by season and day of the week. There has been a long-term (30 years) decline of β in EI ner (2011). Furthermore, mixed results have been observed in other interconnections. In order to provide a more accurate model for the analysis of disturbance, frequency response should be considered as a stochastic process that varies along with time, making analysis of disturbance more difficult.

Although the Frequency Response is a non-stationary stochastic process, it can be approximated as a strict stationary process within relatively short duration. In our study, a period of 1.5 years is analyzed, which is considered short comparing

to the 30 years trend. This is because frequency Response is a slow time-varying random variable, and given a small time window, it can be approximated as a strictly stationary stochastic process, and furthermore as a random variable, whose statistical characteristics do not change over time and do not follow any trends, as is analyzed in Lauby et al. (2013). However the parameters of model for Frequency Response can be updated over the time so that it reflects the most current state of the power grid system.

Chapter 3

Voltage Signature Pattern Recognition in Power Grid

The pattern recognition of voltage signature mainly consists of the following steps:

1. Data preprocessing: since not all records are valid, problematic records must be eliminated. Then the signature records should be normalized so they are comparable and ready for feature extraction.

2. Feature extraction: original dataset are in high-dimensionality and containing high level of noise, thus unsuitable for pattern recognition directly. Consequently, statistics for amplitude, Fourier transform and wavelet transform are utilized to extract features as well as mapping the high dimensional dataset to relatively low dimensional dataset.

3. Unsupervised root pattern extraction: since the ultimate goal is to perform supervised pattern recognition, it is desired to obtain training dataset at the first place. This is realized by using unsupervised learning approach together with heuristic manual screening.

4. Supervised pattern recognition: With the training dataset, it is possible to perform the supervised learning technique. And the support vector machine

(SVM) is chosen as the classifier. Because this is a multi-label problem, multi-label SVM should be generated based on the binary SVM through binarization, which is implemented by One-vs-All (OvA) approach.

5. Performance Evaluation: the performance of multi-label SVM is evaluated by a subset of dataset as the testing dataset.

3.1 Feature Extraction

Voltage signature recorded by PMU at transmission level is represented as a 1-dimensional vector along time stamp, containing 1800 samples with sample rate of 1/60 s. The raw records also can be considered in a 1800-dimensional space, where each record is a point in the data cloud.

Intuitively the raw records can be utilized as the representation of different patterns. In practice however, this brings several drawbacks to the performance of pattern recognition algorithms:

1. raw data set usually resides in high dimensional space and can cause computational problems (curse of dimensionality)

2. raw data set contains relatively high level of noise, making classifiers less robust

Consequently it is preferred to extract features of the raw dataset that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in the feature space Jain et al. (2000), whose dimension generally is substantially lower than that of the raw data set. Naturally the statistical characteristics can be exploited to represent the patterns resides in the raw data set. And the effectiveness of the representation space or feature space is determined by the separation of patterns from different classes. Considering the variation of records corresponding to different patterns reflected in spatial (temporal) domain and frequency domain, three statistical approaches are adopted to generate features:

1. amplitude statistics, including a. mean; b. standard deviation; c. skewness; d. Kurtosis.

2. statistics of wavelet coefficient, including its mean and energy.

3. statistics of Fourier transform, including its mean and energy

3.1.1 Preprocessing

The raw records are not all usable as some of them are problematic. According to the settings (**more details**) of the transmission level, there are five voltage levels, and any records that fall out of these levels are treated invalid. Meanwhile, there exists sudden zero drops in some records, i.e., voltage signal abruptly drops from several 10k V to zero, which is impossible for actual voltage signature, but caused by recording errors. Furthermore, the length of some records is smaller than the normal 1800 samples, and becomes unusable because it is not clear which part is missing. Records that contain any of the aforementioned issues will be abandoned. And the eligible records are normalized by the corresponding voltage level.

Afterwards signature signals are rescaled to the range $[0, 1]$. This is because the voltage signature signals of the same pattern still have different scale, e.g., the voltage dip, etc. The rescaling process is defined as:

$$\hat{x}(n) = \frac{x(n) - \min x(n)}{\max x(n) - \min x(n)} \tag{3.1}$$

3.1.2 Amplitude Statistics

Different patterns reside in voltage signature are directly illustrated by the amplitude variations of signals. Therefore, the statistics of amplitude is an efficient tool to separate different patterns.

Mean of the signal represents its average level,

$$\mathrm{E}[x] = \mu = \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{3.2}$$

Standard Deviation shows how much variation or dispersion from the average exists, i.e., the fluctuation of the signal,

$$\sigma = \sqrt{\mathrm{E}[(X-\mu)]^2} = \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2} \tag{3.3}$$

Skewness is a measure of the asymmetry of the signal,

$$\gamma_1 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathrm{E}[(X-\mu)^3]}{\sigma^3} \tag{3.4}$$

and generally negative skew indicates the left tail is longer while positive skew means the right tail is longer.

Similar to the concept of skewness, **Kurtosis** is a descriptor of the shape, and is any measure of the "peakedness" of a signal,

$$\beta_2 = \frac{\mathrm{E}[(X-\mu)^4]}{(\mathrm{E}[(X-\mu)^2])^2} = \frac{\mathrm{E}[(x-\mu)^4]}{\sigma^4} \tag{3.5}$$

As a result, the amplitude statistics of the signal is equivalent to its first, second, third and forth moment.

3.1.3 Fourier Transform Statistics

The well-known Fourier transform is an expansion of a signal in terms of an infinite sum of periodic sines and cosines, which are mutually orthogonal,

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \mathrm{d}t \\ x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} \mathrm{d}\omega \end{aligned} \tag{3.6}$$

where $x(t)$ is the original signal in temporal domain and $X(\omega)$ is the correspondent frequency representation.

In practice discrete Fourier transform (DFT) is more used since in most scenarios the signals are discretely sampled,

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) e^{-jk\omega_0 n} \\ x(n) &= \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{jk\omega_0 n} \end{aligned} \tag{3.7}$$

Essentially DFT can be considered as sampling the signal in frequency domain with resolution ω_0, and $X[k]$ is the coefficients for each frequency component. Since DFT maps the original signal from temporal domain to frequency domain and the dimensionality remains the same, the statistics, i.e., the mean and energy (standard deviation) of the coefficients is used instead to generate a more compact representation.

$$\mathrm{E}[X(k)] = \frac{1}{N} \sum_{i=1}^{N} X(i) \tag{3.8}$$

$$energy[X(k)] = \sum_{i=1}^{N} X(i)^2 \tag{3.9}$$

3.1.4 Wavelet Statistics

Fourier transform reveals the frequency information satisfyingly for stationary signals, nevertheless it fails to take the temporal variant information into account, indicating that non-stationary signals like the voltage signature cannot be represented by Fourier transform completely.

Wavelet transform however gives a time-frequency representation of the signal, meaning that it provides the time-domain information that a specific spectral component occurs. Wavelets are derived from a signal generating function named

the *mother wavelet*, which should meet the following conditions,

$$\int \psi(t)\mathrm{d}t = 0,\ \psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{3.10}$$

Similar to DFT, usually discrete wavelet transform (DWT) is used. The DWT of a signal $x(n)$ is considered as passing it through a series of filters. For a signal level transform the signal is decomposed by using both a low pass filter g and a high-pass filter h simultaneously. And the output is

$$\begin{aligned} y_{low}(n) &= \sum_{k=-\infty}^{\infty} x(k)h(2n-k) \\ y_{high}(n) &= \sum_{k=-\infty}^{\infty} x(k)g(2n-k) \end{aligned} \tag{3.11}$$

The above decomposition is repeated to further increase the frequency resolution as well as the approximation coefficients decomposed with high and low pass filters and then down-sampled. The structure is demonstrated in Figure 3.1, which is the structure used for voltage signature signal analysis. Since DWT also maps original signal space to the coefficient space while keeping the dimensionality, Eq 3.11 is used to obtain the statistical feature of the DWT coefficient.

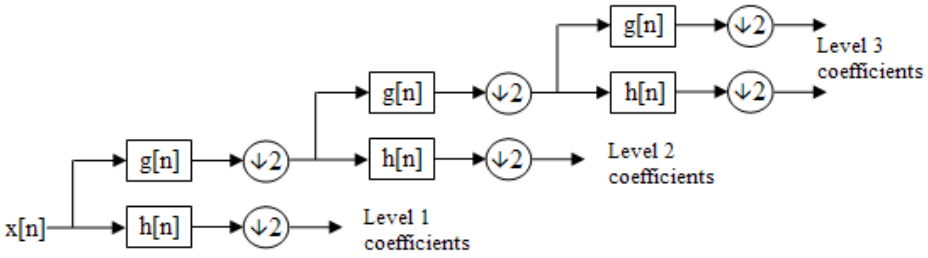


Figure 3.1: A three level filter bank

Once the statistical features of amplitude, Fourier transform and wavelet transform are generated, they are concatenated to form the feature space, which is a

24-dimension space and is much lower compared to the original 1800-dimension raw data space.

3.2 Classification

In order to successfully recognize different patterns of voltage signature, it is desired to use the root patterns to conduct supervised classification, i.e., the prior knowledge of the patterns should be obtained. Consequently root patterns have to be extracted first.

3.2.1 Root Patterns Extraction

Root Patterns can be considered as a set of well separated voltage signature signals in both original data space and feature space, whose labels are identified. They serve as the training data set for the supervised classifiers.

The data set of voltage signature contains around 10,000 records and it would be unfeasible to search for certain pattern of signature manually. A subset of original data set is used as training set and its index is uniformly scattered throughout the whole data set to ensure as many patterns being included as possible. K-means clustering, as a classic unsupervised classification is adopted to make the root patterns extraction more efficient.

K-means clustering tries to partition n observations or records into k clusters, and within each cluster the distance (or within-cluster sum of squares (WCSS)) between each observation are minimized,

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \mu_i\|^2 \tag{3.12}$$

where $\mathbf{x_j}$ is the jth element in a set of observations$(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$, and $\mathbf{S} = S_1, S_2, \cdots, S_k$ is the partitioned sets; μ_i is the mean of observations in S_i.

K-means problem is NP-hard, and there exist some efficient heuristic algorithms not computationally difficult and converge fast. The most common one, referred to as Lloyd's algorithm, uses an iterative refinement approach to reach a local minimum. It proceeds by alternating between two steps:

1. **Assignment**: assign each observation to the cluster whose mean yields the least WCSS:

$$S_i^{(t)} = \mathbf{x}_m : \|\mathbf{x}_m - \mu_i^{(t)}\|^2 \leq \|\mathbf{x}_m - \mu_j^{(t)}\|^2 \; \forall \, j, 1 \leq j \leq k, \tag{3.13}$$

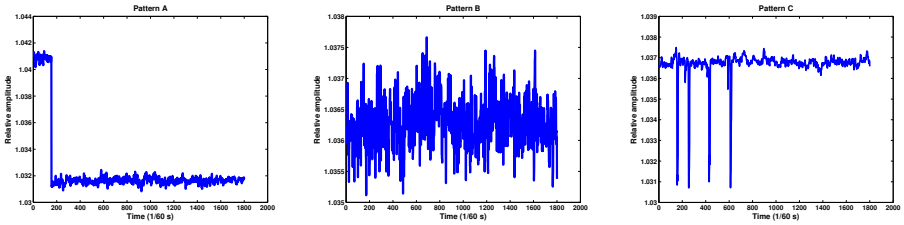
   where $\mathbf{x}_m$ is assigned to only one $S^{(t)}$.

2. **Update**: calculate the new means as the centroids of the new clusters:

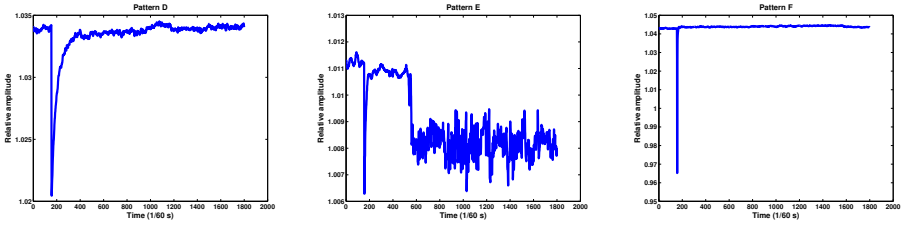
$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} \mathbf{x}_j \tag{3.14}$$

   where $|\cdot|$ is the cardinality of the set.

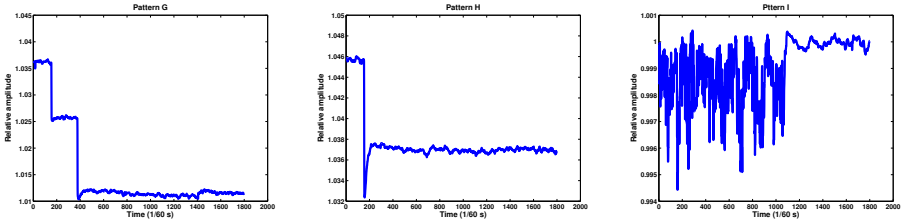
Because k-means is unsupervised, there is always one questions: how to choose k, i.e., how many clusters would be appropriate to identify all the different patterns. In practice we tried different k from 10 to 20. Once the clusters are generated, heuristic screening is used to furthermore eliminate the mis-classified records from the clusters and the root patterns are extracted. Finally there are 15 recognizable patterns extracted from the voltage signature records, displayed in Figure 3.2.



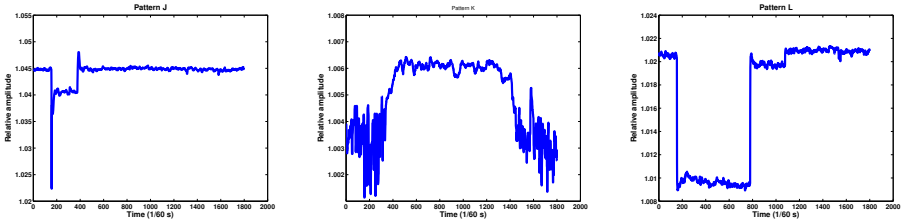
(a) Pattern A (b) Pattern B (c) Pattern C



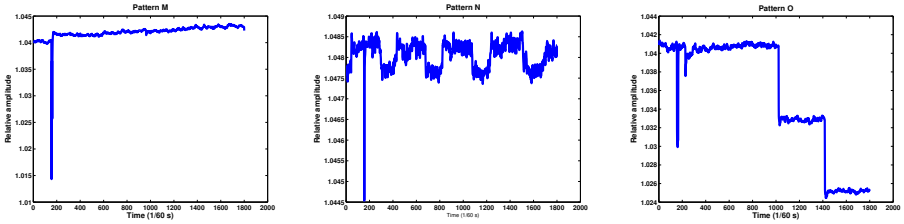
(d) Pattern D (e) Pattern E (f) Pattern F



(g) Pattern G (h) Pattern H (i) Pattern I



(j) Pattern J (k) Pattern K (l) Pattern L



(m) Pattern M (n) Pattern N (o) Pattern O

Figure 3.2: 15 recognizable patterns of voltage signature

3.3 Supervised Classification

Based on the root patterns extracted, it is possible to recognize patterns of voltage signature using multi-label classification techniques. And we use support vector machine (SVM) as the base binary classifier.

3.3.1 Support Vector Machine

Support vector machine (SVM) is a supervised learning methodology associated with various learning algorithms. SVM can be considered as a representation of data points in dataset space being mapped to another space in which they are separated by "gaps" as apart as possible. Usually SVM is used for classification and regression analysis.

There are many linear machines with margins already exist, and SVM shares their ideas of exploiting margins in order to provide a strong classification. In order to understand SVM, consider the simplest case: linear SVM with separable. However, SVM tends to represent patterns within the dataset in a high dimension - typically much higher than the original feature space of the dataset. This is realized through an appropriate (non)linear mapping $\phi(\cdot)$ to a sufficiently high dimension, such that data points from two categories can always be separated by a *hyperplane*.

The hyperplane (in the higher dimension) is defined by *support vectors* which essentially are a subset of (transformed) training patterns. These vectors can be treated as the most difficult patterns to recognize or classify, equivalent to the most informative subset in the training set. In order to understand SVM, consider the simplest case, linear SVM with separable dataset, as illustrated in Figure 3.3.

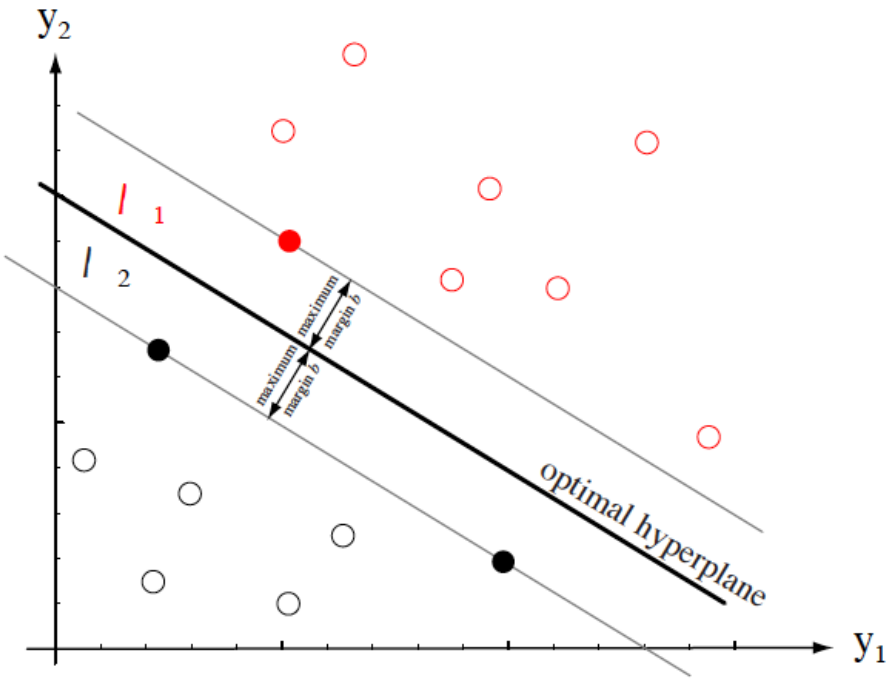


Figure 3.3: Support vectors in a 2-D space. The hyperplane is formed by the three support vectors (solid dots) with a maximum distance b

Assume that we have labeled data as training dataset $\{\mathbf{x}_i, y_i\}$, $i = 1, \cdots, l$, $y_i \in \{-1, 1\}, \mathbf{x}_i \in \mathbb{R}^d$. Suppose we have some hyperplane that separates the positive label from the negative label, and the hyperplane satisfy $\mathbf{w} \cdot \mathbf{x} + b = 0$ where $\mathbf{x}$ lie on the hyperplane, $\mathbf{w}$ is the normal to the plane. Then $|b|/\|\mathbf{w}\|$ is the perpendicular disturbance from the hyperplane to the origin. Define d_+ and d_- as the shortest distances from the hyperplane to the closest positive and negative examples respectively. Also define the "margin" as the sum of d_+ and d_-. It can be formulated as follows: if all the examples in the training dataset satisfy the following constraints:

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 \quad \text{for } y_i = +1 \tag{3.15}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 \quad \text{for } y_i = -1 \tag{3.16}$$

which can be combined into:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b \leq -1) \geq 0 \quad \forall i \tag{3.17}$$

The points lies on the hyperplane H_1: $\mathbf{x}_i \cdot \mathbf{w} + b = 1$ with normal $\mathbf{w}$ and perpendicular distance from the origin $|1-b|/\|\mathbf{w}\|$. For hyperplane H_2: $\mathbf{x}_i \cdot \mathbf{w} + b = -1$ with normal $\mathbf{w}$, and perpendicular distance from the origin $|-1-b|/\|\mathbf{w}\|$. Therefore the margin is $2/\|\mathbf{w}\|$ since $d_+ = d_- = 1/\|\mathbf{w}\|$. As a result, the problem is converted to the form:

$$\begin{aligned} &\min \|\mathbf{w}\|^2 \\ &\text{s.t.} \quad y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \end{aligned} \tag{3.18}$$

In order to solve the problem, positive Lagrange multipliers are introduced and the constraint equations are multiplied by positive Lagrange multipliers and subtracted from the objective function to form the Lagrangian:

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i(\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i \tag{3.19}$$

It is necessary to minimize L_p with respect to $\mathbf{w}$, b. And we can get:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \tag{3.20}$$

$$\sum_i \alpha_i y_i = 0 \tag{3.21}$$

Since these are equality constraints in the dual formulation, we can substitute them into Eq 3.19 to give

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{3.22}$$

So the solution can be found by minimizing L_P or maximizing L_D.

In most occasions, we have to face non-separable dataset, and it is impractical to use a hyperplane that 100% separates dataset. Soft margin is found by allowing a small rate of misclassification.

SVM training is a process of finding the support vectors and margins then finally the hyperplane. Fundamentally the aforementioned mapping $\phi(\cdot)$ is a linear operator; however in many occasions hyperplane in lower dimension fails to provide a well-defined boundary to separate dataset. Therefore nonlinear kernel function, i.e., nonlinear $\phi(\cdot)$, is introduced to map the original feature space into space with higher dimension where there exists a linear hyperplane that better separates dataset. There are many kernel function that have been developed, among which polynomial kernel and radical basis function (Gaussian) kernel are most commonly used. They are the kernel functions used in the SVM for classification of voltage signature. Note that nonlinear functions do provide a better classification performance most times but the overhead (computational task) is higher than linear SVM, especially noticeable when dealing high-dimensional data (curse of dimensionality).

SVM generally is recognized as a superior classifier due to its capability of *generalization*. This is achieved by maximizing the margin area between two categories, which intuitively is correspondent to the tolerance of error of classification.

For the purpose of selecting the best parameters for kernel functions, balanced accuracy is utilized as the performance metric. Due to the severely unbalanced distribution of the patterns in voltage signature records, the traditional accuracy might be biased because it could favor the more weighted patterns (patterns appear more often than others), while balanced accuracy avoids inflated performance estimations over unbalanced datasets, defined as,

$$\text{balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2} \tag{3.23}$$

where *sensitivity* is the *true positive rate* and measures the correctly identified positives; *specificity* is the *true negative rate* and measures the correctly identified

negatives,

$$\begin{aligned} \text{sensitivity} &= \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \\ \text{specificity} &= \frac{\text{true negative}}{\text{true negative} + \text{false positive}} \end{aligned} \tag{3.24}$$

Although 15 patterns are extracted, they are severely unevenly distributed in the voltage signature records. While one dominant pattern takes around 50% of portion, some other patterns appear less than 5% of all records. Consequently we currently focus on the 7 dominant patterns, pattern A, B, C, E, F, G and H, as shown in Figure 3.2.

3.3.2 Parameter Selection for SVM

The Gaussian radical basis function (RBF) is used here as the kernel function. And *libsvm* Chang and Lin (2011) is used as the SVM toolbox. And the binary classification problem is described as the primal optimization problem (C-SVC),

$$\begin{aligned} \min_{\omega,b,\xi} \quad & \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{l}\xi_i \\ \text{s.t.} \quad & y_i(\omega^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \cdots, l, \end{aligned} \tag{3.25}$$

where training vectors $\mathbf{x_i} \in \mathbb{R}^n, i = 1, \cdots, l$, in two classes are given, together with label vector $\mathbf{y} \in \mathbb{R}^l$; $\phi(\mathbf{x_i})$ maps $\mathbf{x_i}$ into higher dimensional space, and $C > 0$ is the regularization parameter.

Furthermore, the RBF kernel is expressed as

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2} \tag{3.26}$$

where parameter γ defines the shape of the kernel function. As a result, (C, γ) is the set of parameters needs to be optimized for each base binary classifier. The approach

for parameter selection is "grid-searching", that is, search for a pair of $(\hat{C}, \hat{\gamma})$ that yields the highest balanced accuracy in the 2-D parameter space. Figure 3.4 illustrates the searching surface.

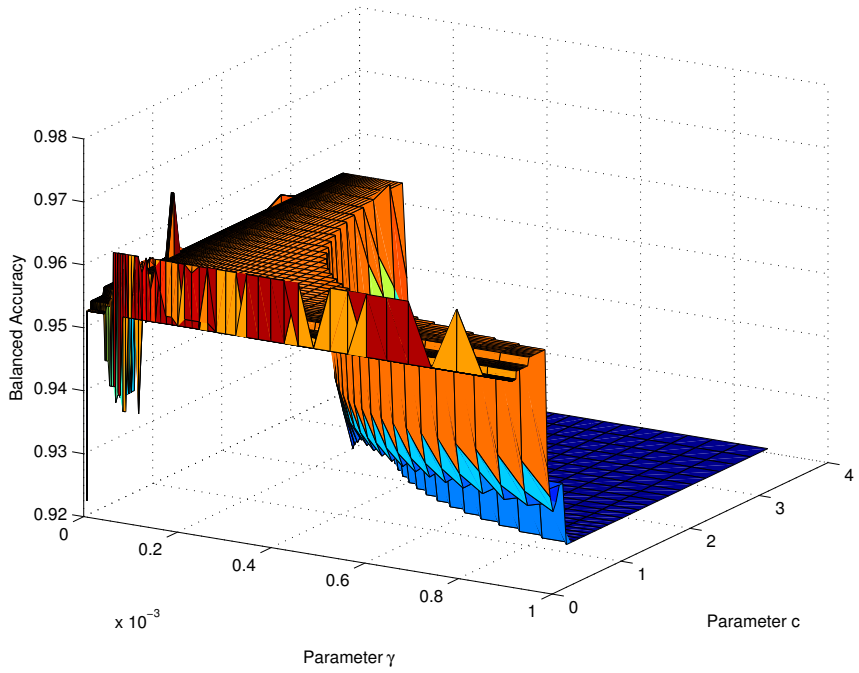


Figure 3.4: The surface of grid searching for the parameter selection of RBF kernel function in SVM

The parameter selection is conducted for the rest patterns, and table 3.1 gives the performance of SVM for seven dominant patterns.

Table 3.1: Balanced accuracy of seven binary SVM for dominant patterns

	sensitivity	specificity	Balanced Accuracy
Pattern A	98.00%	95.02%	96.51%
Pattern B	92.92%	90.71%	91.81%
Pattern C	51.85%	97.65%	74.75%
Pattern E	80.00%	93.44%	86.72%
Pattern F	51.72%	98.11%	74.92%
Pattern G	85.71%	100.00%	92.86%
Pattern H	91.67%	98.13%	94.90%

3.3.3 Multi-label Classification

Essentially the problem of pattern recognition for voltage signature is a multi-label classification problem, that is, each instance or sample should be assigned by one of multiple target label (different from "*multiclass classification*"). Usually the multi-label classification problem is divided into multiple binary classification (binarization) problems and can be solved by single-class classifiers. The most common strategies are:

1. "one-vs-one" (OvO) divides the problem into binary classification problems that considers all possible combinations of two classes, i.e., pair-wise. Then the outputs of these base classifiers are combined to produced the prediction of labels of each instance.

2. "one-vs-all" (OvA) or "one-vs-rest" builds a binary classifier for each class where the class is distinguished from all other (rest) classes. Then the results of all the classifiers are combined together.

There have several studies on the goodness of these two binarization of multi-label classification problem, and OvA is preferred here because,

1. the number of base binary classifiers for OvO is significantly more than that for OvA. Denote the number of labels are n_L, then the number of base binary classifiers for OvO is

$$n_{OvO} = \binom{n_L}{2} = \frac{n_L(n_L-1)}{2} \sim \mathcal{O}(n_L^2) \tag{3.27}$$

while the number of base binary classifiers for OvA is

$$n_{OvA} = n_L \sim \mathcal{O}(n_L), \tag{3.28}$$

2. OvA is easier to expand. New labels can be introduced when more records are studied and OvA does not have to change the classification as long as the patterns that corresponding to new labels are included in the training stage originally; while for OvO, new classification has to be developed to take the new labels into the consideration for multi-label classification purpose.

Multi-label classification strategy for OvA is straightforward: it assigns the label to a new instance with the largest probability, which is provided by base binary SVM. Nevertheless note that there is a special class in which the instances do not belong to any of the 7 dominant patterns, i.e., they belongs the rest insignificant patterns. They are voted as 8th pattern if the probabilities given by seven binary classifiers are all smaller than 50%.

3.4 Performance Evaluation

In order to evaluate the performance of SVM classifying voltage signature according to their patterns, a subset of 451 signature records is used as testing dataset, which are labeled manually as the ground truth. As mentioned before, we focus on the seven dominant patterns and consider the less significant patterns as another separate patterns.

According to the strategy in previous section, the results from seven binary SVM are fused to obtain the final labels for the testing dataset, and the confusion matrix 3.2 is generated based on the ground truth, where $Misc$ is the ensemble of less significant

Table 3.2: Confusion matrix for testing dataset with all categories of patterns

Predicted Label \ True Label	Misc	A	B	C	E	F	G	H
Misc	35	2	13	3	6	6	0	5
A	3	42	1	0	0	0	0	2
B	10	1	203	0	8	0	0	1
C	7	0	2	13	0	4	0	0
E	2	0	4	0	21	0	0	0
F	7	0	0	5	0	17	0	0
G	0	1	0	0	0	0	6	0
H	0	1	1	0	0	0	0	22

patterns, i.e., miscellaneous categories of patterns.

According to the confusion matrix table, the accuracy defined by confusion matrix M is defined as

$$\text{accuracy} = \frac{trace(M)}{\sum_{i=1,j=1}^{l} M_{i,j}} \tag{3.29}$$

where l is the number of labels and thus M is a $l \times l$ square matrix. The resulting accuracy is 80.12%, i.e., the error rate is 19.88%.

Furthermore the existence of less significant patterns interferes with the classification of the dominant patterns mainly because their features are not included in the training stage. Consequently it is hard to definitively delineate the boundary between the dominant patterns and less significant patterns. Removing the less significant patterns will lead to an unbiased evaluation for the performance of the multi-label classification, and the confusion matrix is illustrated in 3.3.

And the resulting accuracy for the reduced confusion matrix is 86.20%, higher than the accuracy for the complete dataset.

It can be concluded that for both confusion matrix one of the misclassifications occurs misclassifying patter E, multiple voltage dips, as pattern C, single voltage

Table 3.3: Confusion matrix for testing dataset with dominant patterns

Predicted Label \ True Label	A	B	C	E	F	G	H
A	42	1	0	0	0	0	5
B	1	207	2	12	1	0	2
C	0	2	17	0	7	0	0
E	0	5	1	19	0	0	0
F	0	0	10	0	18	0	1
G	1	0	0	0	0	6	0
G	0	1	1	0	0	0	6

dip. This is closely related to the representation of the features extracted from the original voltage signature, suggesting the necessity of a better set of features that can distinguish different patterns but with more similarity.

Chapter 4

Conclusions and Future Works

Real-world disturbance records of power grid system in North America was studied in this paper and the power-law tail distribution was confirmed by extending the analysis to the data from year 1992 to 2009; Furthermore, the behavior of small-sized disturbance was found following a log-normal distribution. We also explored the disturbance records in frequency domain from FNET in EI, an innovative wide measurement network. With the relationship $P = \beta \Delta f$, the analysis of disturbances recorded as megawatt events and frequency change are equivalent, yielding the possibility of a faster and more convenient analysis of disturbances solely by frequency information at the distribution level. Finally we discussed the stochastic process of Frequency Response (β) and its influence on disturbance analysis. By approximating Frequency Response as a strict stationary process within a certain observation window, disturbance analysis follows the paradigm of aforementioned analysis.

For the pattern recognition of voltage waveform, real-world records are obtained from EI at the transmission level by PMUs. Dataset is preprocessed in order to be eligible for pattern recognition; features are extracted from the original dataset so that the dimensionality is reduced and the redundant information is removed at the same time; k-means is used to generate root patterns as the training dataset for the supervised classification. Finally multi-label SVM is used to perform the classification

and a subset of voltage waveform records is used as the testing dataset to evaluate the performance of the classification. In the multi-label SVM, one-vs-all (OvA) is used as the binarization approach due to its advantages on the computation complexity. Although the classification reaches over 80% accuracy, it still mis-classify certain types of patterns due to the similarity in their shape and features. This also indicates that during the feature extraction step, the features need to characterize more both local and global distinctive attributes.

For the future work, the stochastic characteristics of disturbance size in both power and frequency domain, i.e., the variation of P and Δf over a greater time scale, should be considered and more data are needed to provide an accurate estimation for the statistical model based on a longer period of records of power grid system. As for the pattern recognition of voltage waveform, more testing dataset should be used to include all the recognizable patterns, and thus provide a more general evaluation of performance. More real-world FIDVR data should be collected and covers as many different patterns as possible. It is expected to eventually detect FIDVR in an online fashion using the same paradigm of pattern recognition for general patterns in power grid system. Furthermore, it is intriguing to use other state-of-art techniques like *Deep Network* to learn a better representation of voltage waveform.

Bibliography

(2011). Balancing and frequency control. Technical report, NERC. 20

Abdullah, A., Sha'ameri, A., Sidek, A., and Shaari, M. (2007). Detection and classification of power quality disturbances using time-frequency analysis technique. In *Research and Development, 2007. SCOReD 2007. 5th Student Conference on*, pages 1–6. 4

Andersson, G., Donalek, P., Farmer, R., Hatziargyriou, N., Kamwa, I., Kundur, P., Martins, N., Paserba, J., Pourbeik, P., Sanchez-Gasca, J., Schulz, R., Stankovic, A., Taylor, C., and Vittal, V. (2005). Causes of the 2003 major grid blackouts in north america and europe, and recommended means to improve system dynamic performance. *Power Systems, IEEE Transactions on*, 20(4):1922–1928. 1

Carreras, B. A., Newman, D. E., Dobson, I., and Poole, A. B. (2004). Evidence for self-organized criticality in a time series of electric power system blackouts. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 51(9). 1

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 35

Chen, J., Thorp, J., and Parashar, M. (2001). Analysis of electric power system disturbance data. In *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, pages 738–744. 1, 6

Gardner, R. and Liu, Y. (2007). Fnet: A quickly deployable and economic system to monitor the electric grid. In *Technologies for Homeland Security, 2007 IEEE Conference on*, pages 209–214. 2

Holmgren, . and Molin, S. (2006). Using disturbance data to assess vulnerability of electric power delivery systems. *Journal of Infrastructure Systems*, 12(4):243–251. 1

Ingleson, J. and Allen, E. (2010). Tracking the eastern interconnection frequency governing characteristic. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–6. 20

Jain, A., Duin, R. P. W., and Mao, J. (2000). Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37. 23

Kang, S., Zhang, H., and Kang, Y. (2010). Application of signal processing and neural network for transient waveform recognition in power system. In *Control and Decision Conference (CCDC), 2010 Chinese*, pages 2481–2484. 4

Lauby, M., Bian, J., Ekisheva, S., and Varghese, M. (2013). Eastern interconnection frequency response trends. In *Power and Energy Society General Meeting (PES), 2013 IEEE*, pages 1–5. 3, 13, 20, 21

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323–351. 1

Qiu, B., Chen, L., Centeno, V., Dong, X., and Liu, Y. (2001). Internet based frequency monitoring network (fnet). In *Power Engineering Society Winter Meeting, 2001. IEEE*, volume 3, pages 1166–1171 vol.3. 2

Sachtjen, M. L., Carreras, B. A., and Lynch, V. E. (2000). Disturbances in a power transmission system. *Phys. Rev. E*, 61:4877–4882. 1

Safty, S., Ashour, H., Dessouki, H., and Sawaf, M. (2004). Online fault detection of transmission line using artificial neural network. In *Power System Technology, 2004. PowerCon 2004. 2004 International Conference on*, volume 2, pages 1629–1632 Vol.2. 4

Tsai, S., Zhang, L., Phadke, A., Liu, Y., Ingram, M., Bell, S., Grant, I., Bradshaw, D., Lubkeman, D., and Tang, L. (2004). Study of global frequency dynamic behavior of large power systems. In *Power Systems Conference and Exposition, 2004. IEEE PES*, pages 328–335 vol.1. 2

Wang, W., He, L., Markham, P., Qi, H., and Liu, Y. (2013). Detection, recognition, and localization of multiple attacks through event unmixing. In *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, pages 73–78. 1

Zhong, Z., Xu, C., Billian, B., Zhang, L., Tsai, S., Conners, R., Centeno, V., Phadke, A., and Liu, Y. (2005). Power system frequency monitoring network (fnet) implementation. *Power Systems, IEEE Transactions on*, 20(4):1914–1921. 2

Appendix

Appendix A
Summary of Equations

1. Complementary cumulative distribution function (complementary CDF or CCDF)

$$F_X^C(x) = P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$$

2. Power law distribution - PDF

$$p(x) = C_0 x^{-\alpha}, \quad x \geq x_0,\, \alpha > 1$$

3. Power law distribution - log-log plot for the CCDF

$$\log(F_X^C(x)) = -\gamma \log x + \log \lambda_0$$

where $\lambda_0 = \frac{C_0}{\alpha - 1}$ and $\gamma = \alpha - 1$

4. Log-normal distribution - PDF

$$f_{X_s}(x; \mu_s, \sigma_s) = \frac{1}{x\sigma_s\sqrt{2\pi}} e^{-\frac{(\ln x - \mu_s)^2}{2\sigma_s^2}}, \quad 0 < x \leq x_0$$

where μ_s and σ_s are the mean and variance, respectively, of the variable x on a logarithmic scale.

5. Log-normal distribution - CDF

$$F_{X_s}^C(x) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\ln x - \mu_s}{\sigma_s\sqrt{2}}\right)\right]$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} \mathrm{d}t$ is the error function.

6. Normalized root mean square error (NRMSE)

$$NRMSE = \frac{RMSE}{x_{\max} - x_{\min}} = \frac{\sqrt{\sum_{i=1}^{n} (\hat{x}_i - x_i)^2 / n}}{x_{\max} - x_{\min}}$$

where x is the original curve or dataset, and $\hat{x}$ is the fitted curve.

7. The power loss and frequency change of disturbances associated by Frequency Response

$$P = \beta \Delta f, \quad P, \beta, \Delta f > 0$$

8. PDF of ratio distribution, assume $Z = \frac{X}{Y}$

$$\begin{aligned} p_Z(z) &= \frac{\mathrm{d}F_Z(z)}{\mathrm{d}z} \\ &= \int_0^\infty p_Y(y) \frac{\mathrm{d}}{\mathrm{d}z} \left(\int_{-\infty}^{yz} p_X(x) \mathrm{d}x \right) \mathrm{d}y \\ &= \int_0^\infty y \cdot p_X(yz) p_Y(y) \, \mathrm{d}y \end{aligned}$$

9. PDF for frequency change of large disturbances

$$\begin{aligned} p_Z(z) &= \int_0^\infty y C_0 (yz)^{-\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, \mathrm{d}y \\ &= z^{-\alpha} \cdot \frac{C_0}{\sqrt{2\pi}} \int_0^\infty y^{-\alpha+1} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, \mathrm{d}y \end{aligned}$$

10. PDF for frequency change of small disturbances

$$\begin{aligned} p_Z(z) &= \int_0^\infty \frac{C_0}{2\pi z\sigma_s\sigma} e^{-\frac{(\ln(yz)-\mu_s)^2}{2\sigma_s^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} \,\mathrm{d}y \\ &= \int_0^\infty \frac{C_0}{2\pi z\sigma_s\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\ &\quad \cdot e^{-\left[(\ln y-\mu_s)^2+(\ln z-\mu_s)^2+(2\ln y\ln z-\mu_s^2)/2\sigma_s^2\right]} \,\mathrm{d}y \\ &= \underbrace{\frac{1}{z\sigma_s\sqrt{2\pi}} e^{(\ln z-\mu_s)^2}}_{\text{log-normal distribution}} \\ &\quad \underbrace{\int_0^\infty \frac{C_0}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}-\frac{(\ln y-\mu_s)^2+(2\ln y\ln z-\mu_s^2)}{2\sigma_s^2}} \,\mathrm{d}y}_{\text{coefficient}} \end{aligned}$$

11. Mean of a vector signal

$$\mathrm{E}[x] = \mu = \bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

12. Standard deviation of a vector signal

$$\sigma = \sqrt{\mathrm{E}[(X-\mu)]^2} = \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2}$$

13. Skewness of a vector signal

$$\gamma_1 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathrm{E}[(X-\mu)^3]}{\sigma^3}$$

14. Kurtosis of a vector signal

$$\beta_2 = \frac{\mathrm{E}[(X-\mu)^4]}{(\mathrm{E}[(X-\mu)^2])^2} = \frac{\mathrm{E}[(x-\mu)^4]}{\sigma^4}$$

15. Continuous Fourier transform

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\mathrm{d}t$$
$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\mathrm{d}\omega$$

16. Discrete Fourier transform

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-jk\omega_0 n}$$
$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{jk\omega_0 n}$$

17. The criteria of mother wavelet in wavelet transform

$$\int \psi(t)\mathrm{d}t = 0,\ \psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

18. The discrete wavelet transform

$$y_{low}(n) = \sum_{k=-\infty}^{\infty} x(k)h(2n-k)$$
$$y_{high}(n) = \sum_{k=-\infty}^{\infty} x(k)g(2n-k)$$

Vita

Liu Liu was born in Danyang, Jiangsu, China. He receives his B.S. degree in 2010 from Southeast University, Nanjing. He began his Ph.D. study at The University of Tennessee, Knoxville in 2012. Before that he has been studying at The Ohio State University for one year from 2011.