TECHNISCHE UNIVERSITÄT
ILMENAU

Fusion of PMU and SCADA data to develop novel applications for dynamic control centres

Masterarbeit zur Erlangung des akademischen Grades
Master of Science
an der Technischen Universität Ilmenau

Submitted by: Arun Lal

Date Of Birth: 28.04.1990

Course of Study: Media Technology

Matriculation Number: 54566

Department: Fakultät für Elektrotechnik und Informationstechnik

Fachgebiet Elektrische Energieversorgung

Professor: Univ.-Prof. Dr.-Ing. Dirk Westermann

Supervised by: M.Sc. André Kummerow

Dipl.-Ing. (FH) Christoph, Brosinsky

Submitted on: 14.07.2019

Acknowledgement

First of all, I would like to express my deepest thanks to my supervisors M.Sc. André Kummerow and Dipl.-Ing. (FH) Christoph Brosinsky for all the discussions, guidance, valuable suggestions, explanation, and advice. Many thanks to other thesis committee member, Univ.-Prof. Dr.-Ing. Dirk Westermann for taking effort to evaluate my thesis.

Secondly, I would like to thank my friends in Ilmenau for their support, encouragement. Especially, Ms. Duygu Pelin for her constant moral support.

Finally, I would like to thank my beloved family in Pakistan, especially my parents, for their never-ending prayers, and for my brother and sister for their never-ending supports. And last but not least, I would like to thank God, that it would have not been possible to accomplish my study at TU Ilmenau without His blessings.

Kurzfassung

In modernen Energiesystemen gewinnt der Einsatz der PMU-Technologie neben SCADA-Systemen aufgrund ihres verbesserter Beobachtungsfähigkeit zunehmend an Bedeutung. Daher ist es unerlässlich, ein neuartiges und effizientes Modell für die Fusion von PMU- und SCADA-Daten zu entwickeln. Im Rahmen dieser Arbeit wird eine neue Methodik für die Fusion von PMU- und SCADA-Daten für neue Leitstellenanwendungen entwickelt. Es besteht aus einem Klassifikator auf von Entscheidungsbäumen zur Identifizierung und Lokalisierung von Generator- und Leitungsausfällen unter Verwendung von SCADA-Daten und einem zusätzlichen Fusionsalgorithmus unter Verwendung logistischer Regression, um diese Ergebnisse mit einem PMU-Klassifikator für eine erhöhte Robustheit und Zuverlässigkeit zu kombinieren. Dieser Ansatz wird auf simulierte SCADA- und PMU-Daten trainiert und in verschiedenen Testszenarien bewertet, u.a. hinsichtlich Veränderungen in der Spärlichkeit der SCADA-Daten sowie der Verfügbarkeit von PMU-Messungen im Stromnetz.

Abstract

In modern power systems, the utilization of PMU technology is gaining more attention beside SCADA systems due to their increased situational awareness. Hence it is essential to develop a novel and efficient model for the fusion of PMU and SCADA data. Within this thesis, a new methodology is developed for the fusion of PMU and SCADA data for new control room applications. It is comprised of a decision tree-based classifier to identify and locate generator and line outages with the use of SCADA data and an additional logistic regression based fusion algorithm to combine those results with a PMU classifier for an increased robustness and reliability. This approach is trained on simulated SCADA and PMU data and evaluated within different test scenarios, including changes in sparseness of the SCADA data as well as availability of PMU measurements in the power system.

Table of Contents

1 Introduction

1.1 Background and Motivation

The electrical power system is comprised of many interconnected nodes, sources and loads. Such a complex system requires a lot of maintenance and constant monitoring of the dynamic behaviour of the system. Nowadays, most of the substations are equipped with Intelligent Electronic Devices (IEDs) which can collect huge amounts of data in addition to perform their own intended functions. These IEDs include phasor measurement units (PMUs), remote terminal units (RTUs), etc [1], [2]. Traditionally, SCADA systems constitute the main infrastructure for nowadays monitoring and operating power systems, whereas PMUs are special measurement devices that collect synchronized phasors at a much higher rate and in a continuous manner. Despite the high potential of synchro phasor online and offline applications PMUs are not yet used universally in SCADA based control rooms.

The integration of PMU and SCADA data remains a challenge for several reasons. The difference of sampling rates is one major problem. PMU measurements are collected with high sampling rates (typically between 10 and 50 times a second), whereas SCADA measurements are spontaneous with low sampling rates (typically 1 or 2 times a second). To minimize installation costs and limit transmission rates, PMUs are not placed all over the power system and typically don’t collect discrete measurements like circuit breaker status values or tap changer information. This makes the system monitoring a tedious task and requires the merging of different data sources before deploying them in applications. It has been proven through some recent field demonstrations by utilities and vendors that integration of synchro phasor measurements enhances awareness of system operators (see literature review). However, it remains an issue how different data sources shall be merged to aid operators in making decisions. In this context, this thesis focus on the fusion of PMU and SCADA data to support control room applications.

1.2 Goal and Objectives

Based on a state-of-the-art analysis of existing PMU and SCADA applications for power system control rooms, suitable methods are to be developed to aggregate and fuse PMU and SCADA data to support control room applications.

The information gained from the fusion of PMU and SCADA data will be used to develop an application-oriented visualisation in order to support the operation in transmission network control rooms.

- State of the art of applications for network control rooms based on PMU and SCADA data.
- Development of novel algorithms for the combined online evaluation of PMU and SCADA data.
- Development of visualization techniques to support operational operations in dynamic control rooms.

1.3 Thesis Layout

This thesis is organized in the following way:

Chapter 1: Introduction

This chapter encloses the background consideration that motivates this research area. Then it continues with research objectives and ends up with outlining the structure of the thesis.

Chapter 2: State of the Art

Mainly, this chapter illustrates literature review on filed devices (PMU and RTU), machine learning, data fusion and state of the art applications for control room applications.

Chapter 3: Enhanced detection of power system outages with limited PMU measurements

This chapter describes proposed application derived from the state of the art within this thesis.

Chapter 4: Methodology

This chapter describes the methodology to aggregate PMU and SCADA data and the development of novel algorithm for the online evaluation of PMU and SCADA data.

Chapter 5: Evaluation

This chapter describes the evaluation of the proposed methodology on different test scenarios. Performance of the classification and fusion model is evaluated and assessed.

Chapter 6: Conclusion and Future Work

In this chapter, main conclusion, concepts/suggestions for the development and improvement of the presented methodology are discussed. Additionally, further data fusion applications for control room identified within this thesis are presented as well as general concept for integrating and visualizing the presented methodology in existing control rooms.

2 State of the Art

This chapter reviews the literature on existing concepts related to this thesis. It starts by presenting basic concepts and addressing main features regarding field devices (PMU and RTU) which serves as data sources for the developed algorithms within this research work. Then, the important subjects like data fusion and machine learning are described along state-of-the-art applications for control rooms (PMU based, SCADA based and hybrid). The chapter ends with the research questions that are derived from the research analysis.

2.1 Phasor Measurement Unit

Wide Area Measurement Systems (WAMS) are used world-wide in transmission and distribution power systems [3], [4], [5]. WAMS are composed of multiple Phasor Measurement Units (PMUs), which are interconnected by high-speed communication channels with one or more Phasor Data Concentrators (PDCs) [6]. PMUs provide valuable state information of electrical power systems under both static and dynamic conditions. Therefore, measurements are performed at selected nodes of the electrical power grid [7]. They provide phasor estimations synchronized to within a microsecond. This has been made possible by the utilization of Global Positioning System (GPS) and the sampled data processing techniques developed for computer relaying applications [8], [9].

The development of PMU technology can be traced back near the end of 1980s and the first products arrived in market in the early 1990s [10], [11], [12], [13]. With the propagation of PMUs, the study on its application has become an attractive area since the middle of 1990s [14], [15], [16].

In order to understand how PMUs enhance grid operations and planning, it is useful to understand phasor technology. PMUs estimate voltage and current phasors as well as frequency and frequency changes in the electricity grid by processing and analysing the raw instantaneous measurements from voltage and current transformers. A phasor is a complex number that represents both the magnitude and phase angle of sinusoidal waveforms at a specific point in time [6],[17],[18].

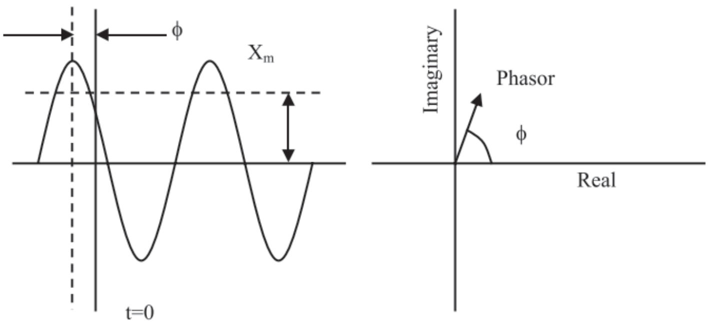


Fig. 2. 1: Phasor Representation [6]

Each phasor measurement is time-stamped against Global Positioning System universal time. Then these time-synchronized measurements are combined to provide a precise and comprehensive view of the entire region [17], [19], [20].

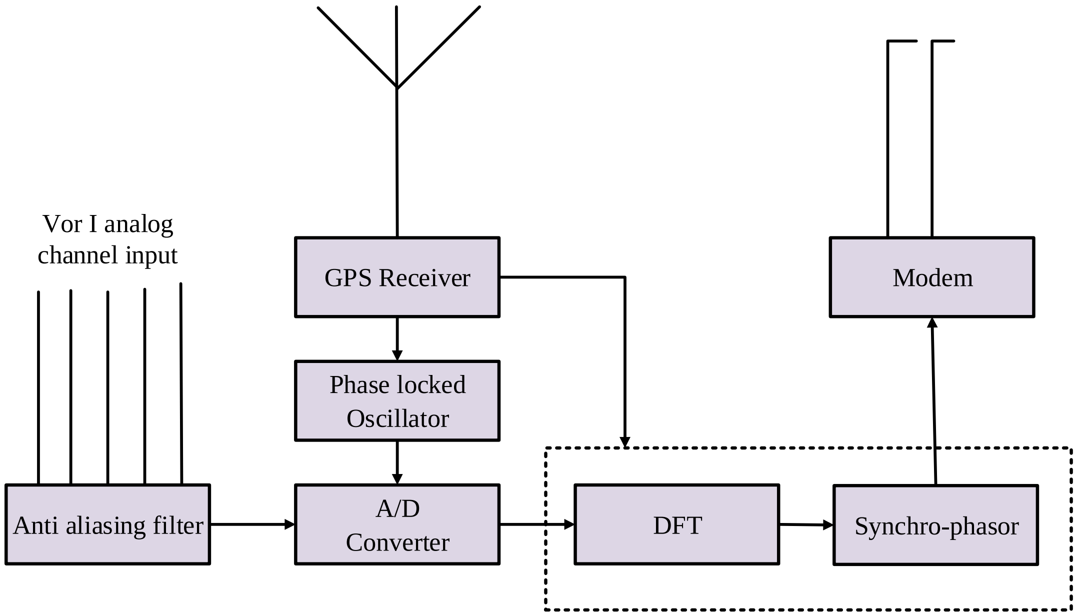


Fig. 2. 2: Block diagram of the Phasor Measurement Unit [14], [21]

Fig. 2.2 shows a simple block diagram explaining the processing of a measured voltage or current analogue signals. The instantaneous measurements are sampled at high frequencies and passed through a hardware low-pass filter for antialiasing and converted into digital data by the analog-to-digital converter. Then the fundamental frequency component is calculated using discrete Fourier transform (DFT) and converted into a complex number which represents the phasor of the sampled waveform. Phasors of the three phases can be combined to produce the positive, negative and/or zero sequence measurements. [7], [21].

2.2 Remote Terminal Unit

SCADA system consists of a number of Remote Terminal Units (or RTUs) collecting field data and sending that data back to a master station via a communication system. The master station displays the acquired data and also allows the operator to perform remote control tasks. Typically, an RTU converts the electrical signals from the equipment to digital values such as the open/closed status from a switch or a valve, or measurements such as pressure, flow, voltage or current [22], [23], [24].

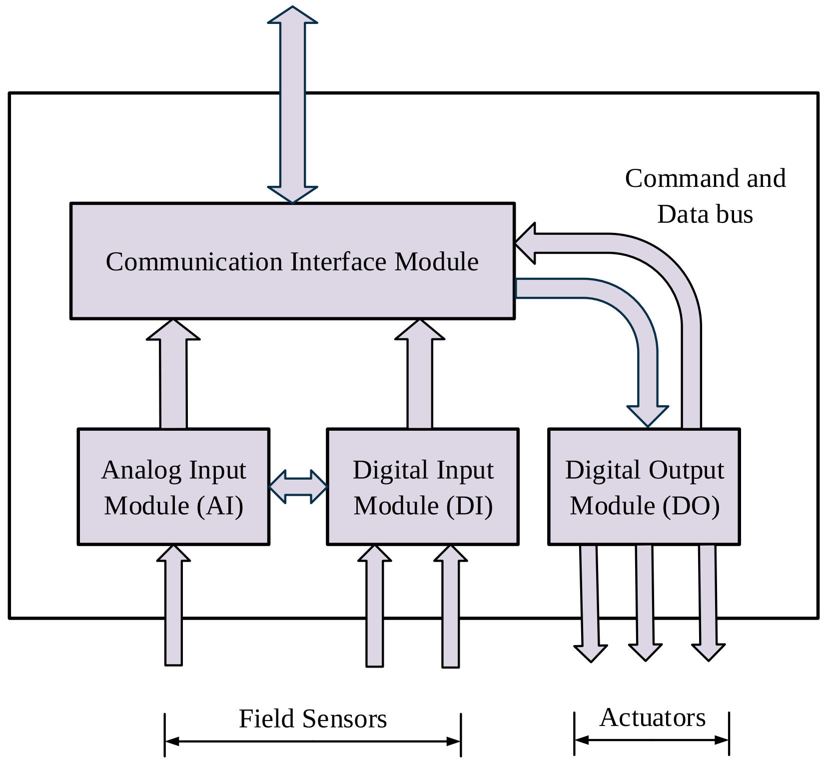


Fig. 2. 3: Structural design of a Remote Terminal Unit [25]

Reference to Fig. 2.3, a generalized structure of a RTU is described in [25], [26], [27]. It includes digital and analog input modules, digital output modules and a communication interface module. The basic functionality of an RTU can be implemented using above mentioned modules. Field sensors can be interfaced with input modules depending upon their type whereas actuators can be operated by interfacing them with digital output modules. The communication interface module is capable of sending data from the RTU to the SCADA center utilizing several means of telemetry.

2.3 Comparison of PMU and SCADA Measurements

PMUs are measurement devices capable of providing 30-60 measurement samples per second whereas SCADA system provides only 1-2 measurement samples per second [28], [29]. SCADA provides different measurement information than PMUs and PMUs are normally located only at specific stations whereas SCADA is available at all stations.

IEEE Standard 1344 defines the formats for output files provided by the Phasor Measurement Units [30], [31]. In SCADA system, the standard IEC 61850 “Communication Networks and Systems in Substations” supports all communication for substation tasks like control, protection and monitoring. The IEC 61850 client and server functionality of the RTU allows the combination of traditional protocols, parallel wiring and the IEC 61850 station bus [20], [32], [33]. Traditionally IEC 60870-5-101/104 is used as communication protocol or DNP3 in U.S. IEC 61850 is used for substation communication and can be used also for communication between substation and control center.

Tab. 2. 1: General comparison between SCADA and PMU measurements, as outlined in [28], [29]:

Attribute	SCADA	PMU
Resolution	1-2 sample per second	30-60 samples per second
Measured Quantities	Magnitude only	Magnitude and Phase angle
Time Synchronization	No	Yes
Focus	Local monitoring and control	Wide area monitoring and control

2.4 Data Fusion Techniques

Hall and Llinas [34] provide the following well-known definition of data fusion:

“Data fusion techniques combine data from multiple sensors and related information from associated databases to achieve improved accuracy and more specific inferences than could be achieved by the use of a single sensor alone.”

Data fusion techniques have been extensively employed on multisensory environments with the aim of fusing and aggregating data from different sensors [35]. Integrating multiple data from different sources into a framework makes a rich database which is more reliable than individual sources [36]. Recently, data fusion methods have been widely considered in different areas to improve decision making process using available sources of information [37]. Authors in [38] provide a short overview of data fusion techniques that can be used to interpret wearable sensor data in the context of health monitoring applications. Another application of data fusion can be observed in the field of image processing, where the goal of data fusion is to represent visual information contained in input images into a single fused image without introducing distortion or information loss [39].

In power system operation, traditional approaches to fault diagnosis based on information from the circuit breakers and relays is often ineffective [40]. Therefore, a combination of discrete data increases fault diagnosis capability in power system [41], [42]. A typical classifier in machine learning domain works on a given data source. In multi sensor system, not every sensor output is available all the time, moreover, system is based on different data sources with different time resolutions, information contents and different spatial distribution of the sensors which is hard to use in classifier algorithm. Therefore, an additional data processing is required [43]. This explains the motivation of using data fusion technique within this thesis to address both data sources (PMU and SCADA).

Data fusion models have been categorized in the literature in a number of distinctly different ways: in terms of architecture, in terms of information levels at which the fusion is accomplished, the objectives of the fusion process, the application domain and others [35], [44], [45], [46].

The characterization most commonly encountered in literature based on the level of detail in the information comprises of data level, feature level, and decision level [45], [47], [48], [49].

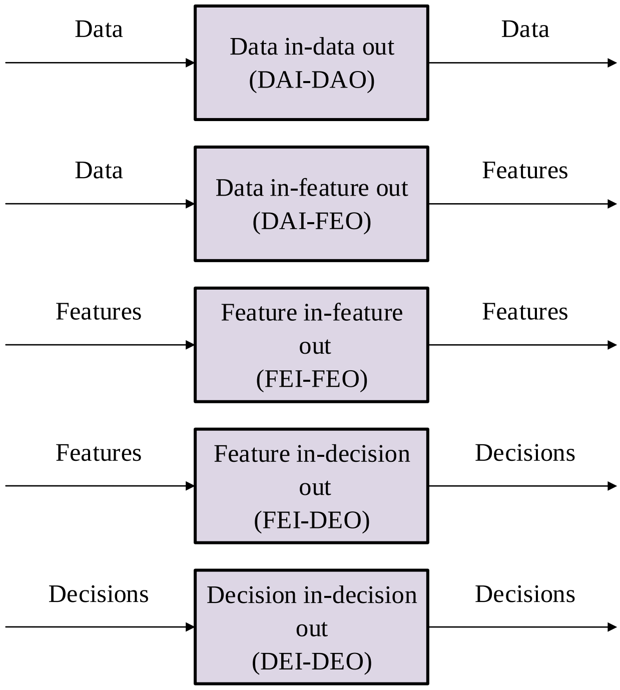


Fig. 2. 4: Classification Based on the input/output data types and their nature [35]

(1) Data in-data out (DAI-DAO)

This type is the most basic or elementary data fusion method that is considered in classification. This type of data fusion process inputs and outputs raw data; the results are typically more reliable or accurate than input raw data. Data fusion at this level is conducted immediately after the data are gathered from the sensors. The algorithms employed at this level are based on signal and image processing algorithms.

(2) Data in-feature out (DAI-FEO)

At this level, the data fusion process employs raw data from the sources to extract features or characteristics that describe an entity in the environment.

(3) ***Feature in-feature out (FEI-FEO)***

At this level, both the input and output of the data fusion process are features. Thus, the data fusion process addresses a set of features with to improve, refine or obtain new features. This process is also known as feature fusion, symbolic fusion, information fusion or intermediate level fusion.

(4) ***Feature in-decision out (FEI-DEO)***

This level obtains a set of features as input and provides a set of decisions as output. Most of the classification systems that perform a decision based on a sensor’s inputs fall into this category of classification.

(5) Decision In-Decision Out (DEI-DEO)

This type of classification is also known as decision fusion. It fuses input decisions to obtain better or new decisions.

2.5 Machine Learning

Statistical learning plays a key role in many areas of science, finance and industry. The science of learning plays a key role in the fields of statistics, data mining and artificial intelligence, intersecting with areas of engineering and other disciplines [50]. This section describes the literature concerning machine learning. After some initial considerations, the chapter includes description of the machine learning techniques relevant to this thesis.

"Machine learning is the body of research related to automated large-scale data analysis. The field also encompasses many of the traditional areas of statistics with, however, a strong focus on mathematical models and also prediction. Machine learning is now central to many areas of interest in computer science and related large-scale information processing domains." [51]

Bishop, C. M. in his book [52] states machine learning as a field of computer science that uses statistical techniques to provide computer systems the ability to learn with data, without being explicitly programmed.

Richert, W. and Coelho, L. in [53] refers machine learning to the changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. The "changes" might be either enhancements to already performing systems or synthesis of new systems [53].

In the context of this thesis, machine learning techniques with their pattern recognition, learning capabilities and identifying features seems to be the answer to overcome the challenges imposed by processing the huge volumes of raw data involved in power system operation. Operation of modern PMU-based power system involves massive amounts of data, in this regard machine learning methods are emerging for processing the information provided and extract valuable knowledge for system operators [17]. Instead of declaring complex analytical models, learning to recognize patterns and identifying features seems to be the answer to overcome the challenges imposed by processing the huge volumes of raw data involved in power system operation [52]. Moreover, numerical conventional methods are computationally expensive, which makes it difficult to use for the on-line security assessment. Machine learning techniques with their pattern recognition, learning capabilities and high speed of identifying the potential security boundaries can offer an alternative approach.

Supervised machine learning builds a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data [53].

Two types of machine learning techniques are outlined in [54], namely classification and regression. Authors refer classification as supervised learning with applications where the class level is precharacterized. However, prediction is used to predict future data sets. In this thesis, aforementioned techniques are used to develop a novel algorithm for control room application.

This section provides the description of the specific algorithms used in this thesis, including Decision Tree (DT) for classification, Logistic Regression for data fusion model and feature extraction techniques.

Decision Tree

Authors in [55] define *"A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node."* A typical decision tree is shown in Figure 8.2.

Literature in [55], [56], [57] describes decision tree as a model widely used to classify datasets into distinct classes. According to the literature, DT is composed of nodes and branches. The objective of DT is to partition the input space by learning simple decision rules on chosen features in a tree-wise manner until all classes are separated.

Mathematically, [55], [58] defines DT as follows:

$$T \in \{T_{1,} T_{2,} \, . \, . \, . \, , T_k, t_1\} \tag{2.1}$$

Where, T is a binary tree, k is an index number, t is a tree node and t_1, the root node.

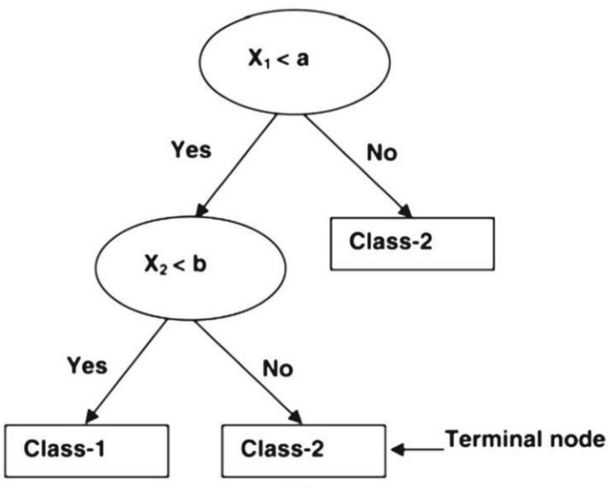


Fig. 2. 5: Decision Tree Classification [59]

Figure illustrates DT where each node has if-then rules and nodes are further partitioning into subsets until classification. One of the hyperparameter of a decision tree is the criteria of splitting feature at each step/node. Gini coefficient is used which provides a way to measure the quality of the split. Whereas, The minimum number of samples required to split an internal node.

Figueiredo [60] used decision trees for classification of electric energy consumer in distribution system. Ramos et. al. [61] have used decision trees for characterization and classification of consumers in distribution system. A Mosavi [62] used a multiple criteria Decision making pre-processing using Data mining tools.

Logistic Regression

"Logistic regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes)." [63]

The logistic regression model uses a logistic function to estimate probability outputs between 0 and 1, which indicate if the observation x belongs to a given class or not. As described in [51] and [64], logistic regression equation can be mathematically defined as:

$$\hat{y} = \frac{e^{(b_0+b_1x)}}{(1+e^{(b_0+b_1x)})} \tag{2.2}$$

Where $\hat{y}$ is the output, b_0 is the intercept and b_1 is the coefficient for the single input value x.

Feature extraction techniques

Feature extraction is generally used for all types of machine learning applications, as it reduces the dimensionality of the data [65], [66], [67]. This provides motivation to use feature extraction techniques within this thesis in order to obtain adequate classification results.

Within this thesis, a number of statistical based feature extraction technique are used as described in [68], [69] :

Arithmetic mean (AM)

The arithmetic mean is the average of the values $\{x_1, x_2, \ldots, x_m\}$. It is calculated by the following equation:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2.3}$$

Variance

Variance of the values $\{x_1, x_2, \ldots, x_m\}$ is calculated by the following equation:

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2 \tag{2.4}$$

Standard Deviation

The standard deviation of the values $\{x_1, x_2, \ldots, x_m\}$ is calculated by the following equation:

$$\sigma = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2} \tag{2.5}$$

However, feature selection techniques are applied to assess the relevance of the features for the classification step. This includes the computation of correlation matrices as well as the utilization of the feature importance given by the classification algorithm.

2.6 Control Room Application

In this section, the state-of-the-art applications based on PMU and SCADA data for network control rooms are briefly reviewed. Within this thesis, analysis of the state-of-the-art applications allows to develop novel algorithm for the online evaluation of PMU and SCADA data for dynamic control room.

2.6.1 PMU based Applications

In recent years, significant research effort was dedicated to develop applications using phasor measurement data. In particular, PMUs became an indispensable tool for post-mortem analysis [70], . Some examples can be found in Brazilian [71] and Colombian blackouts [72].

Research work in [73] mentions that collection of PMU data allows several possibilities for system studies and planning. Moreover, authors in [74] highlight the importance of PMU technology in power system operation. According to the paper, measurement rates of PMU between 6–60 Hz are ideal for observing system dynamics and and to develop applications for real time system monitoring and post disturbance analysis. Author in [75] explains how PMU provides new solutions for wide area monitoring, protection and control (WAMPAC). WAMPAC systems require the transmission of specific node information to a remote station and all information should be time synchronized. It provides a complete instantaneous snapshot of the power system. M. Parashar, A. Jampala and Jay Giri in [76] summarize some of the common experiences and early real-time PMU applications such as angular separation, oscillatory stability monitoring, and disturbance detection applications.

Previous studies [77], [78], [79], [80] have grouped PMU applications into online and offline applications. Online applications typically comprise of wide are monitoring and visualizing, oscillating detection, frequency stability monitoring, current stability monitoring and state estimation. In offline mode, PMUs can be used fro post event analysis, data visualization and model validation. Creation and developments in PMU technology offers wide area monitoring of electrical power grids [73], [81]. Wide area monitoring has provided a wide variety of opportunities to enhance the backup protection and system protection of modern power systems [6], [13], [82], [83].

2.6.2 SCADA based Applications

Main focus of the thesis is to fuse SCADA and PMU data. Although SCADA data is a widely used in several industries, requirements within the electric utility industry for remote control of substations and generation facilities has probably been the driving force for modern SCADA systems [74]. Authors in [84] mention some industrial applications like operation of chemical plants, oil and gas plants, electrical power systems, etc. Hayashi et. al. [85] presented an EMS (energy management system) that relied upon a SCADA system. The proposed approach comprised of data exchange with remote centres as well as the web-based real-time electricity demand metering, made for the first time within one Electricity Company in the South East Europe.

A survey in [74] reports SCADA applications used mainly in power system. Moreover, it provides an overview of SCADA functions in power grids!, state estimation, load forecasting, power flow control, data maintenance and voltage/reliability monitoring. Qiu and Gooi [86] proposed to apply the Intranet technology to SCADA for power system and presented the result of development of the Intranet-based SCADA trial system. The Intranet-based SCADA was concerned with real-time performance and reliability of supervisory control. They discussed various measurements at the trial system including picture display time, supervisory control response time and failover time. Real time performance, system cost and maintainability of the Intranet-based SCADA were evaluated based upon the trial system.

Research work in [87] presents a customized SCADA at customer side distribution automation system (DAS) for operating and controlling low voltage (LV) downstream of 415/240 V by using the Tenaga Nasional Berhad (TNB) distribution system. Proprietary software is used to develop algorithm for the controller and to develop HMI for monitoring and controlling functions for the operator. The SCADA system developed provides automatic fault isolation, monitoring and controlling functions for the operators and data collection for future analysis.

Authors in [88] introduces a paradigm for data integration where the substation field data recorded by monitoring and protection Intelligent Electronic Devices (IEDs) is used to supplement Remote Terminal Unit (RTU) data for Supervisory Control and Data Acquisition (SCADA) system, which improves SCADA and other applications. This data integration paradigm allows very detailed monitoring of the power system and subsequently a more comprehensive decision-making opportunities for the control applications.

The paper [89] describes and demonstrates a unique Internet-based application in a substation automation system that is implemented based on the existing system control and data acquisition (SCADA) system. The proposed application offers real-time data visualization and allows user to control the operation of the substation at the server site. Whereas, Creery and Byres [90] emphasizes on industrial cyber security aspect of power system and SCADA network. In this context, their work presents and evaluates methods to determine and reduce the vulnerability of networked control systems to unintended and malicious intrusions. Based on past assessments and incidents, security issues are identified, along with technical and procedural countermeasures to mitigate these risks. Moreover, study work in [91] investigates the nature of vulnerabilities to SCADA system and ways to protect SCADA system from ongoing cyber threats. Authors studies SCADA communication systems and focuses their attention to SCADA in the electric power grid. Based on a survey of common SCADA vendors, security policy to be put into practice is recommended to ensure the security of the system.

2.6.3 PMU and SCADA based Applications

A number of researchers explain the integration of time correlated information from PMU and SCADA data and their application in fault location, alarm processing and cascading post event analysis to help operators to better analyse and control the system [92], [93], [94], [95], [96]. Moreover, authors in [97], outline the problem of large amounts of data produced by WAMS on the timescale of milliseconds that further accounts for PMU as not a complete replacement for SCADA systems. In this regard, a PMU placement method is presented in their research work that complements both PMU and SCADA data. The proposed method applies PMU information radiation method to deduce the operating parameters of the buses (bus voltages and the current of the branches connecting to the buses) where PMUs are not installed and adopt state estimation to improve the precision of the handling data. Authors believe method can obtain the system’s operation states by using less PMUs, and is beneficial to the analysis and application of power system.

The cornerstone control room application observed in the literature reviewed, where PMU and SCADA measurements are fused is the power system state estimator [98], [99], [100], [101]. The state estimation is a powerful tool necessary for real time monitoring, control of power systems and is based on a static power flow model to estimate information such as voltage magnitudes and phase angles at each bus [29], [102], [103], [104]. There has been numerous studies that consider PMU and SCADA measurements, along with help of dynamical state estimation tool such as Kalman Filtering techniques has made possible to estimate the dynamic state of a power system. [105], [106], [107], [108], [109], [110]. Authors in [111] have extensively studied state estimation using both PMU and traditional SCADA measurements. The paper presents two ways to include PMU measurements in the state estimation process:

1) A single state estimator, where PMU measurements are mixed with the traditional power flow measurements;
2) A two-stage scheme, where the state estimate obtained from the traditional SCADA measurements in [112] is improved by using a second estimator that employs PMU measurements only.

M. Ghosal and V. Rao in [29], adopts a multi-rate multi–sensor data fusion algorithm and performs dynamic state estimation using standard Kalman filtering technique. The simulation results presented in this research work confirm that the fusion improves the final estimation. Kirinčić, Skok and Franković in [113], develops a recursive algorithm that estimates state estimation for the part of the system observable by PMUs (voltage and current phasors are measured), and the obtained state estimate is then merged with SCADA measurements to estimate the state of the entire power system. The proposed methodology was applied to the IEEE test systems with 30 and 57 buses as well as the model of the Croatian transmission system. Based on the comparison of results of the proposed state estimator with the results of the classical state estimator and other hybrid models, enhanced accuracy is observed for the developed model.

2.7 Research Questions

Based on the analysis of the literature reviewed, it can be noticed that the fusion of PMU and SCADA measurements to be used for new applications in control room is not yet explored adequately. However, previous works have specifically focused on the data fusion of PMU and SCADA measurements for the state estimation application. In this regard, this thesis can focus to investigate the following questions:

- Are there any other applications besides state estimation where data fusion techniques for PMU and SCADA data can be used?
- What improvement is possible against using single data source?
- What measurements/signals/channels are required from PMU and SCADA data for the application to be developed within this thesis?

3 Enhanced detection of power system outages with limited PMU measurements

As the analysis of the state of the art in previous chapter yields that most of the work accounts for state estimation hence no further investigation is required in this direction. Further, literature review highlighted the challenge faced by measurement and control systems to respond securely and reliably to various disturbances in power grid networks. Hence, research work within this thesis derives an application related to the enhanced detection of generator and line outages when only a reduced or limited number of PMU measurements is available along with the SCADA measurements. SCADA data could help to make a better detection of disturbances along with the PMU measurements.

As described in most of the PMU based state of the art applications (section 2.6.1), placement of PMU improves system observability. Few research works explored and proposed the possibility of placing PMUs on all nodes of the system. However, PMU placement has been limited so far, mainly due to the high costs of PMU.

In context of the proposed application, another reason for using limited PMU is to investigate whether the placement of limited PMU results in enhanced detection of disturbances along with the SCADA measurements, even if one of the measurement systems is not available.

As a first step a classifier is developed to classify disturbances in the power system based on the SCADA data only. These classification results are to be fused with the results from a PMU classifier to obtain a final decision in order to investigate whether the fused information is useful to enhance the detection of the power system outages. For this, the available PMU measurement are varied within different test scenarios.

This chapter proposes a machine learning based approach for the problem of disturbance classification in power systems using SCADA data and enhance the reliability by fusing it with the results of a PMU based classifier.

The proposed approach consists of the conventional steps of data pre-processing, feature extraction, and classification using SCADA data. The trained PMU classifier is a prerequisite to this thesis work; hence this thesis focuses on the fusion of the SCADA and PMU classification results. However, figure 4.1 illustrates the general workflow of the proposed model in the application phase the where input to the model is the field sensor-generated data from two different measurement systems, namely the SCADA data $\mathbf{X}^{S}$ and the PMU data $\mathbf{X}^{P}$. The input data consists of multiple time-series channels that represents important electrical parameters. Initially, data is pre-processed and prepared. After the stage of pre-processing, a set of features are calculated in order to feed the respective classification models for fault classification. Later, the outputs from both SCADA and PMU classifiers are fused to obtain final output in terms of a flag, which contains type of outage and location.

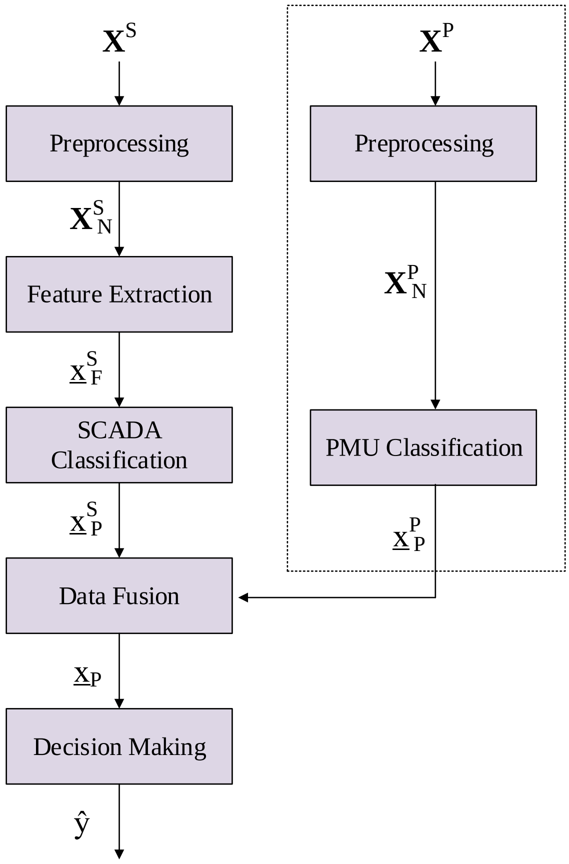


Fig. 3. 1: Steps of algorithm in application phase.

3.1 Pre-processing

At pre-processing stage, the input matrix $\mathbf{X}^{S}$ is normalized using min–max normalization procedure and provides a normalized input matrix ($\mathbf{X}_{N}^{S}$).

For feature extraction, $\boldsymbol{X}_{N}^{S}$ is an input matrix:

$$\boldsymbol{X}_{N}^{S} \in \boldsymbol{R}^{dxn} \tag{3.1}$$

Where, d= sample size, n= no. of stations*channels and S = SCADA.

Input matrix $\boldsymbol{X}_{N}^{S}$ is a normalized measurement matrix and hence consists of different samples from stations and channels. A channel $\underline{x}_{N}^{S}$ can be can be current, voltage etc.

3.2 Feature Extraction

After the data is preprocessed, feature extraction techniques are used to reduce the input dimension and a feature vector ($\underline{x}_{F}^{S}$) is created.

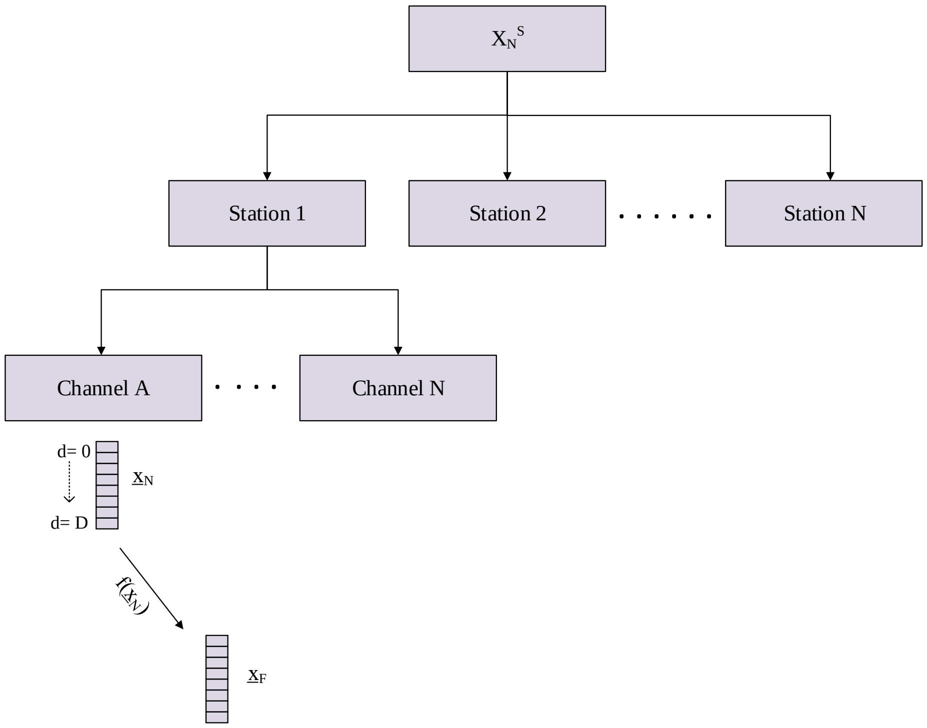


Fig. 3. 2: Steps of Feature Extraction.

A statistical feature is calculated for each channel in general according to:

$$x_F = f(\underline{x}_N) \tag{3.2}$$

Where, $f(x)$ = mean, variance, standard deviation as described in section 2.5

$\underline{x}_F$ is a feature vector that concatenates all computed single features.

$$\underline{x}_F = \{x_F\} \tag{3.3}$$

$\underline{x}_F$ is used as input to train the SCADA classification model.

In this thesis, a feature extraction technique based on the time characteristic of the signal is also designed. It basically calculates the ratio of time-span of the time signal to the time instances.

This technique is constructed on the basis of following definitions:

$$\boldsymbol{X}_N = \{\underline{x}_N^1, \; . \; . \; . \; , \underline{x}_N^k\} \tag{3.4}$$

$$X = \{\underline{S}_N\} \tag{3.5}$$

$$\underline{S}_N = \{\underline{S}_N^0, \; . \; . \; . \; , \underline{S}_N^n\} \tag{3.6}$$

$$\underline{t} = \{t^0, \; . \; . \; . \; , t^n\} \tag{3.7}$$

$$\underline{f} = \frac{\max(\underline{S}_N) - \min(\underline{S}_N)}{\max(\underline{t}_N) - \min(t_N)} \tag{3.8}$$

Where, $\boldsymbol{X}_N$ is a matrix/tensor of k number of normalized available observations, X is a time series signal, $\underline{S}_N$ is a time series signal with window of length k, $\underline{t}^n$ is time instances of time window.

3.3 SCADA Classification

The decision tree-based classification algorithm is proposed to identify and locate the disturbance events. As described in chapter 2, a decision tree is an effective supervised machine learning method that builds a classifier based on given known observations. As shown in Fig. 3.3, the input feature vector is passed to the decision tree which estimates the class affiliation of the current observation by a (pseudo) probability vector $\underline{x}_P^S$. The final class assignment $\hat{y}^S$ is done in the decision-making step based on the maximum probability value. Estimated and true class labels are compared to each other in a validation step and the empirical error e is used to optimize the classifier parameters $\underline{\theta}_C$ until all training instances are assigned correctly or a desired error rate is achieved.

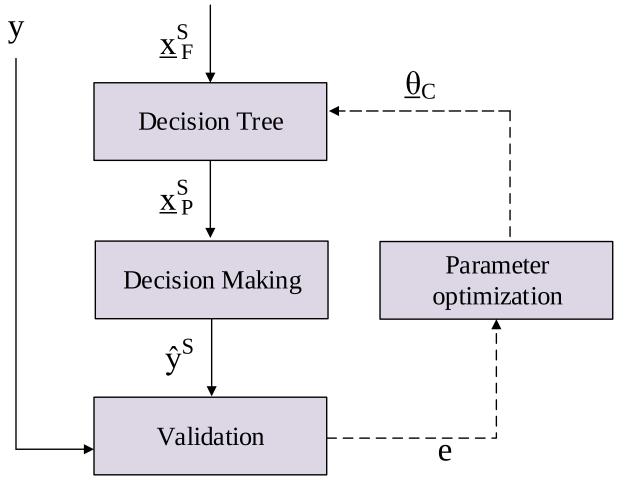


Fig. 3. 3: Training and design phase of SCADA classifier.

3.4 PMU Classification

It should be noted that PMU classifier is pretrained for this research work. However, PMU classification is based on the method proposed in [114] which uses linear classification model with logistic loss function. Within this approach time series shapelets are used to correctly assign class labels. Mathematical definition of the model is as follows:

$$\hat{y}_i = W_0 + \sum_k M_{i,k} W_k \tag{3.9}$$

Where M is the minimum distance between a time series and a shapelet, W is the weight, k is the number of the shapelets.

3.5 Data Fusion

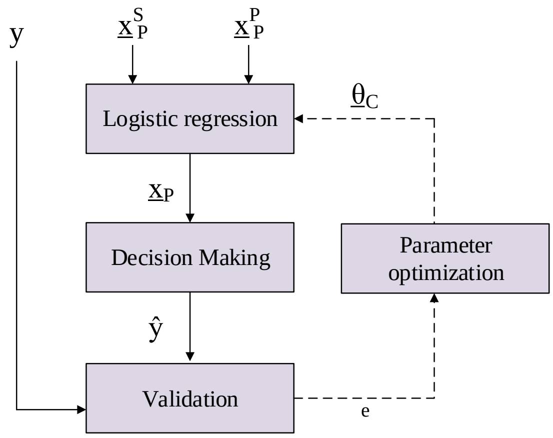


Fig. 3. 4: Training and design phase of data fusion model.

Based on a review of data fusion techniques conducted in the state-of-the-art chapter, a decision level data fusion is performed. Fusion model is based on the logistic regression algorithm described in section 2.5. It takes class probabilities from SCADA $\underline{x}_P^S$ and PMU $\underline{x}_P^P$ classifiers as inputs to obtain a better/final decision in the form of a SCADA flag $\hat{y}$ – see (3.10). It selects the maximum probability out of both classifier outputs. SCADA flag is solely the class label with contains information about the type y_{Type} and location y_{Loc} of the disturbance. Type of outage and location is important for an operator and both the information is contained in the class label. These steps are performed in loop to tune the parameters of the algorithm on training dataset - similar to the procedure described in the previous section. The logistic regression parameters are estimated via maximum likelihood estimation (e.g. Newton-Raphson method). Once, the parameters are tuned, performance of trained model is evaluated on validation dataset.

$$\hat{y} = [\hat{y}_{Type}, \hat{y}_{Loc}] \tag{3.10}$$

4 Evaluation

This chapter describes test environment and the application derived from the state of the art within this thesis. Proposed application defines which kind of outages and services are to be simulated and from which stations PMU and SCADA data are to be taken.

4.1 Power Grid and Generic Station Model

It is necessary to describe the system in order to know the origin of the data. Generally, PMU and SCADA data are received directly from the field sensors. However, in the absence of real measurements, the power grid model simulated in the Power Factory software used in this thesis to provide the data.

The power grid model simulates an electrical transmission system. Within this model, different kinds of outages from different stations are simulated and their effects on selected stations are evaluated.

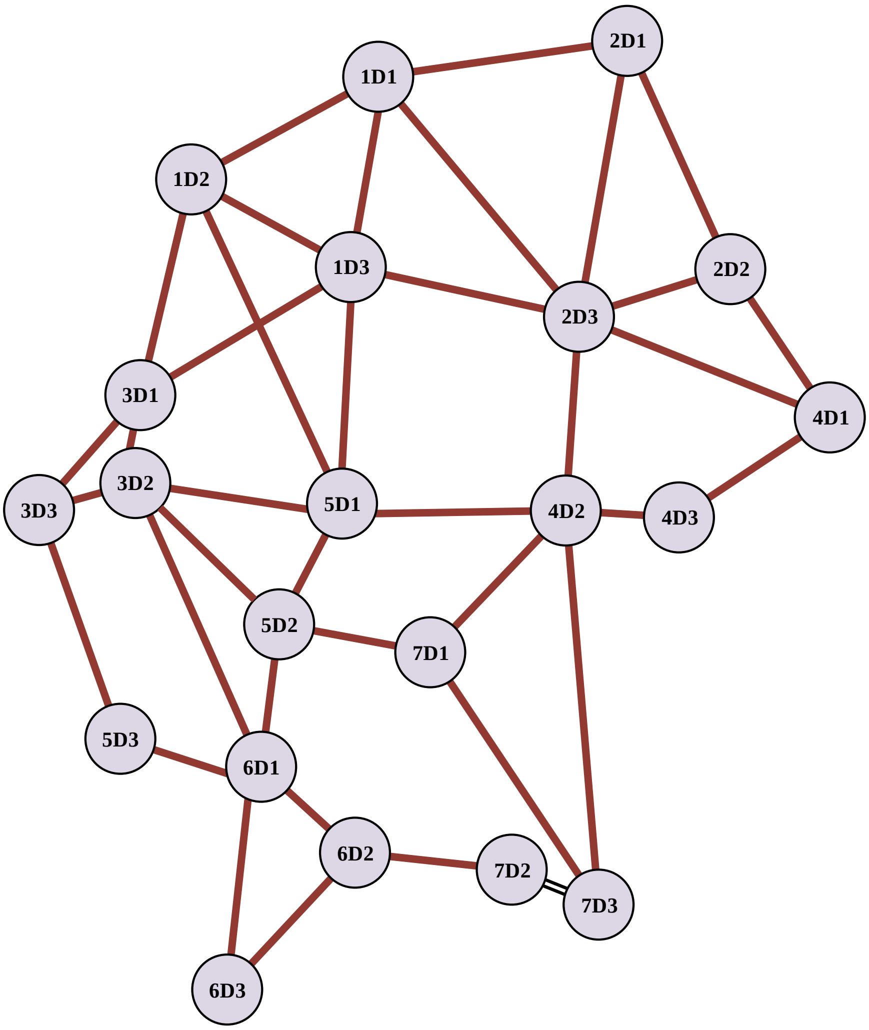


Fig. 4. 1: Topology of the power grid model, nodes representing substations interconnected through transmission lines (Single line can represent several transmission lines)

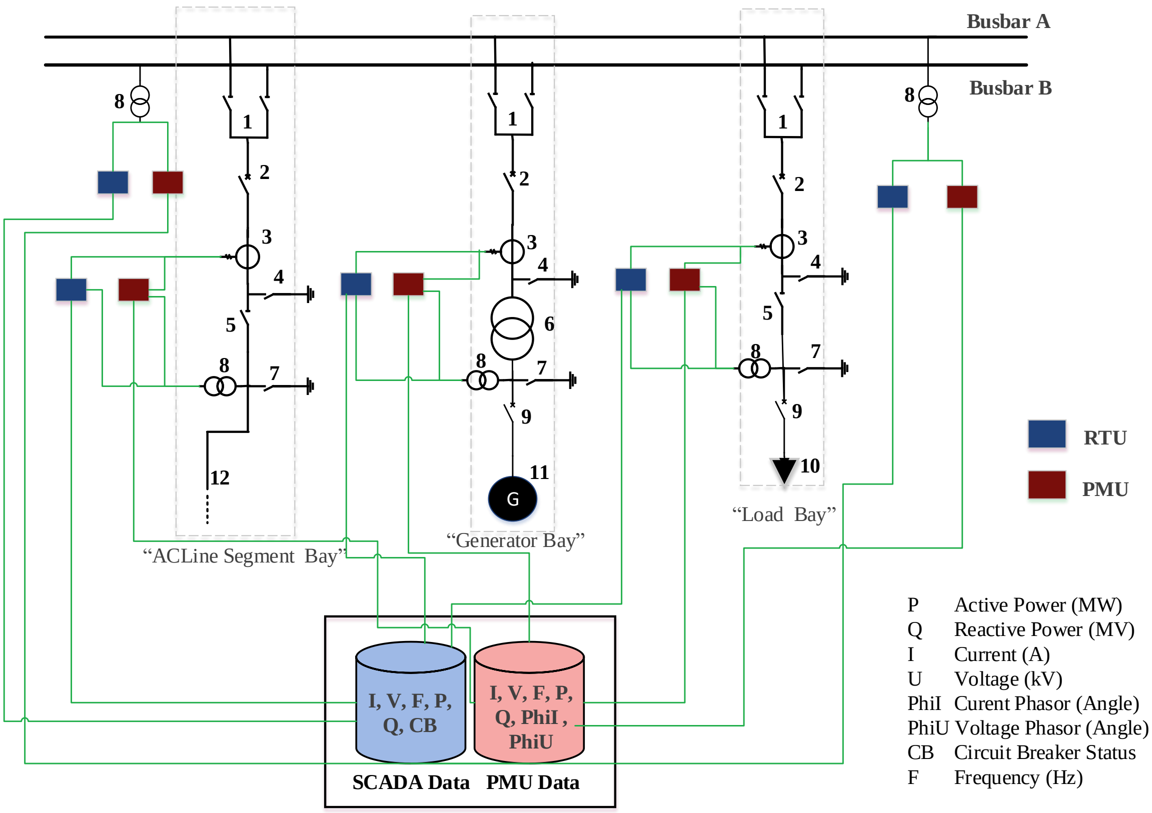


Fig. 4. 2: Generic station model. 1. Isolator of the busbars A and B 2. Circuit Breaker 3. Current Transformer 4. Earthing switch 5. Feeder disconnector. 6. Transformer 7. Earthing switch 8. Voltage Transformer 9. Circuit Breaker 10. Load 11. Generator 12. AC Line Segment

Fig 4.2 illustrates a generic station model to visualize data collection flow (shown with green lines) for both PMU and SCADA measurement system. It should be noted that PMU measures voltage phasor, frequency at bus bar and current phasor at bay/field. Whereas, SCADA measures voltage, frequency at bus bar and switch position, current, active and reactive power at bay/field. In case of SCADA, RTU receives measurements from the respective IED, which is further collected at SCADA database whereas PMUs receive measurements directly from respective voltage/current transformer and collected at PMU database.

In order to generate dynamic simulations from power factory model, certain steady state conditions were selected. For these steady state conditions, all possible line and generation outages were simulated by means of RMS Simulations with a time step of 1ms for all the simulated variables. PMU measurements are generated by taking average values every 100ms (corresponds to reporting rate of 10 f.p.s.).

Creation of SCADA data from dynamic simulations

SCADA data is generated from the PMU data by introducing sparseness and noise by a simple rule-based model. To make the SCADA data spontaneous the data change d_{ch} is calculated between two adjacent PMU observations x_t^P and x_{t-1}^P as well as the corresponding time span Δt between those two observations if a certain value of d_{ch} is exceeded. Additionally, white noise x_e is integrated into the model to account for the differences between PMU and SCADA measurement accuracies. The general equations are defined as follows:

$$x_S^t = \begin{cases} (x_t^P - x_{t-1}^P) > d_{ch} \; and \; \Delta t \geq d_{tmin}: x_t^P + x_e \\ \Delta t \geq d_{tmax}: x_t^P + x_e \\ No\ value \end{cases} \tag{4.1}$$

$$x_e = \mathcal{N}(0, \sigma_e) \tag{4.2}$$

The sparseness and noisiness of the SCADA data can be controlled by adjusting the parameters p_{ch} and p_e as follows:

$$d_{ch} = \frac{\sigma(\underline{x}_t^P)}{p_{ch}} \tag{4.3}$$

$$\sigma_e = \frac{\sigma(\underline{x}_t^P)}{p_e} \tag{4.4}$$

The sparseness model is applied for all measurement channels C independently to get a SCADA measurement vector $\underline{x}_t^S$:

$$\underline{x}_t^S = [x_t^S]_{C=1}^{C=N} \tag{4.5}$$

4.2 Performance Evaluation

This section shows the results obtained with methodology proposed. Firstly, some test scenarios are defined in order to implement the methodology on the test model. Then, the performance evaluation of the model is detailed. Input data is split into training and validation dataset (67% / 33%). The purpose of splitting data up into training and validation dataset is to evaluate the performance of the algorithm trained on training dataset with the data that has not been trained. Performance evaluation criterion is based on the classification accuracy of the model. Accuracy values account for the validation dataset.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{4.6}$$

Test Phase 1

Test phase 1 focuses on the training and performance evaluation of SCADA classification model, described in section 3.3. Firstly, type of measurements (channels), contingencies, stations are defined. Then, based on these assumptions, test scenarios are defined in order to train and evaluate the performance of the SCADA classifier. In the end, results of evaluation are discussed.

Definition of input measurements

In order to train the classifier, some known events are required. Generator and line disturbances were considered. The post-fault behaviour of the system is classified into four classes during the off-line supervised training process.

Generator and line current, voltage, frequency, active and reactive power are used as predictors whereas fault type and location are the targets. As described in section 2.5, mean, variance, standard deviation are the elements of the statistics-based feature vector, whereas a feature vector based on the time characteristic of the signal are collectively used as attributes to train the decision tree.

The dynamic simulations generated from the Power Factory model are used to construct data sequences with 850 samples of each channel i.e., current magnitude (I), voltage (U), frequency (F), active power (P) and reactive power (Q) corresponding to faults. This time series data is sampled to a window of 50 time steps. Sampling is performed since during application phase, the measurements are collected and processed within a sliding window fashion.

Test Scenarios

Test phase 1 comprise of test scenarios to evaluate the SCADA classifier. As described in literature review, feature extraction reduces dimensionality of the input data and provides a compact representation of the data. Hence, in order to train the SCADA classifier, a feature matrix needs to be extracted from the raw data. In this context, the goal of the following test scenarios is to find an optimal feature matrix which provides highest accuracy for a minimum number of features and measurement channels. Additionally, the impact of changed sparseness and noise of the SCADA data is evaluated.

Table 4.1 shows outages and stations considered within test phase 1. Furthermore, five test scenarios are listed in the table where scenarios (1.A-1.D) are based on different combinations of feature matrices. However, scenario (1.E) is based on the impact of noise/sparseness parameter of SCADA data on feature matrix. Feature matrix obtained from each test scenario is fed as input to SCADA classifier and performance of the model is evaluated on these combinations of features, which is described in the later section.

Tab. 4. 1: General assumptions for test phase 1

	Parameter	**Value**
General Assumptio	Outages	Generator Outages: ▪ Outage_2D1.DKW, Outage_5D3.GKW Line Outages: ▪ Outage_L11, Outage_L28
	Stations observed by SCADA	1D1, 2D1, 5D1

Tab. 4. 2: Test scenarios for test phase 1

	Parameter	**Value**
1.A	Measurement channels	Currents, active power, reactive power, voltage, frequency
	Features per channel	Mean, variance, standard deviation, time-based feature
	Sparseness parameter p_{ch}	5
	Noise parameter p_e	10
1.B	Measurement channels	Currents, voltage, frequency
	Features per channel	Mean, variance, standard deviation, time-based feature
	Sparseness parameter p_{ch}	5
	Noise parameter p_e	10
1.C	Measurement channels	Currents, active power, reactive power, voltage, frequency
	Features per channel	Mean, variance, standard deviation
	Sparseness parameter p_{ch}	5
	Noise parameter p_e	10
1.D	Measurement channels	Currents, voltage, frequency
	Features per channel	Mean, variance, standard deviation
	Sparseness parameter p_{ch}	5
	Noise parameter p_e	10
1.E	Measurement channels	Currents, active power, reactive power, voltage, frequency
	Features per channel	Mean, variance, standard deviation, time-based feature
	Sparseness parameter p_{ch}	1
	Noise parameter p_e	1

Model Hyperparameters

Table 4.3 shows the hyperparameters of the decision tree which is used for SCADA classification.

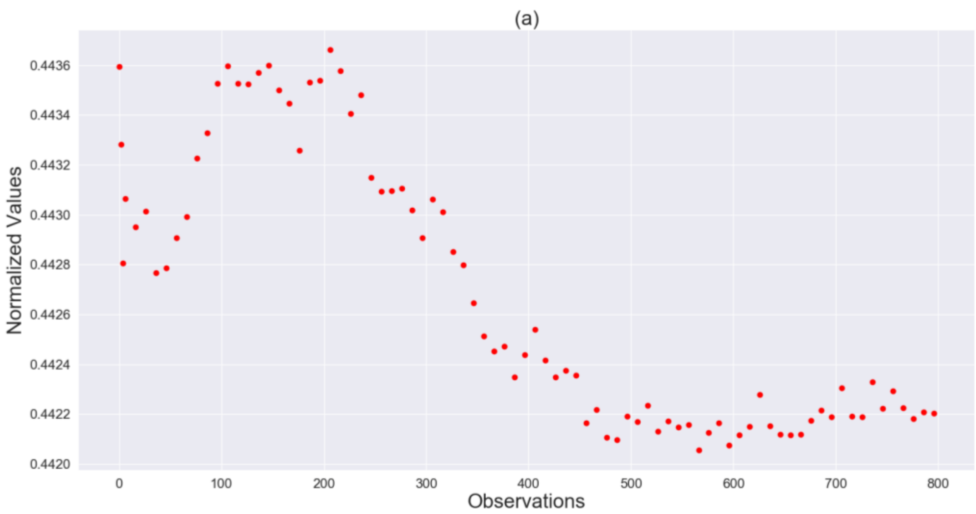
Tab. 4. 3: Hyperparameters of decision tree

Parameters	Description
Split criterion	Gini Coefficient
Minimum number of samples for splitting	2

General behavior of selected raw / normalized signals

As mentioned earlier, each time series data consists of different channels i.e., current magnitude, voltage, frequency, active and reactive power. It should be noted that frequency and voltage magnitudes are measured only at the busbar of each station. Whereas, current, active power and reactive power magnitudes are measured at the fields / bays of each station. Therefore, multiple current, active and reactive power measurements are obtained for each station. Based on this consideration, normalized current signals of different fields from station 1D1 in case of generator outage (2D1.DKW) are presented in figure 4.3. General behaviour of the raw signals from SCADA data can be observed in the figure.

Figure (a) and (b) refers to the current magnitudes of generators, whereas (c) and (d) refers to the current magnitudes of load and line respectively. From figures, a general behaviour of the signals can be noticed. After a disturbance in the system, variation in steady state is observed. However, system adapts dynamics and tends to achieve new steady state. Further, variation from steady state for some fields/bays found to be relatively small, such as line bays, load bays, hence current, active power and reactive power magnitudes of those fields/bays are not considered for feature extraction. Generator bays, photo voltaic source, wind energy source are considered.



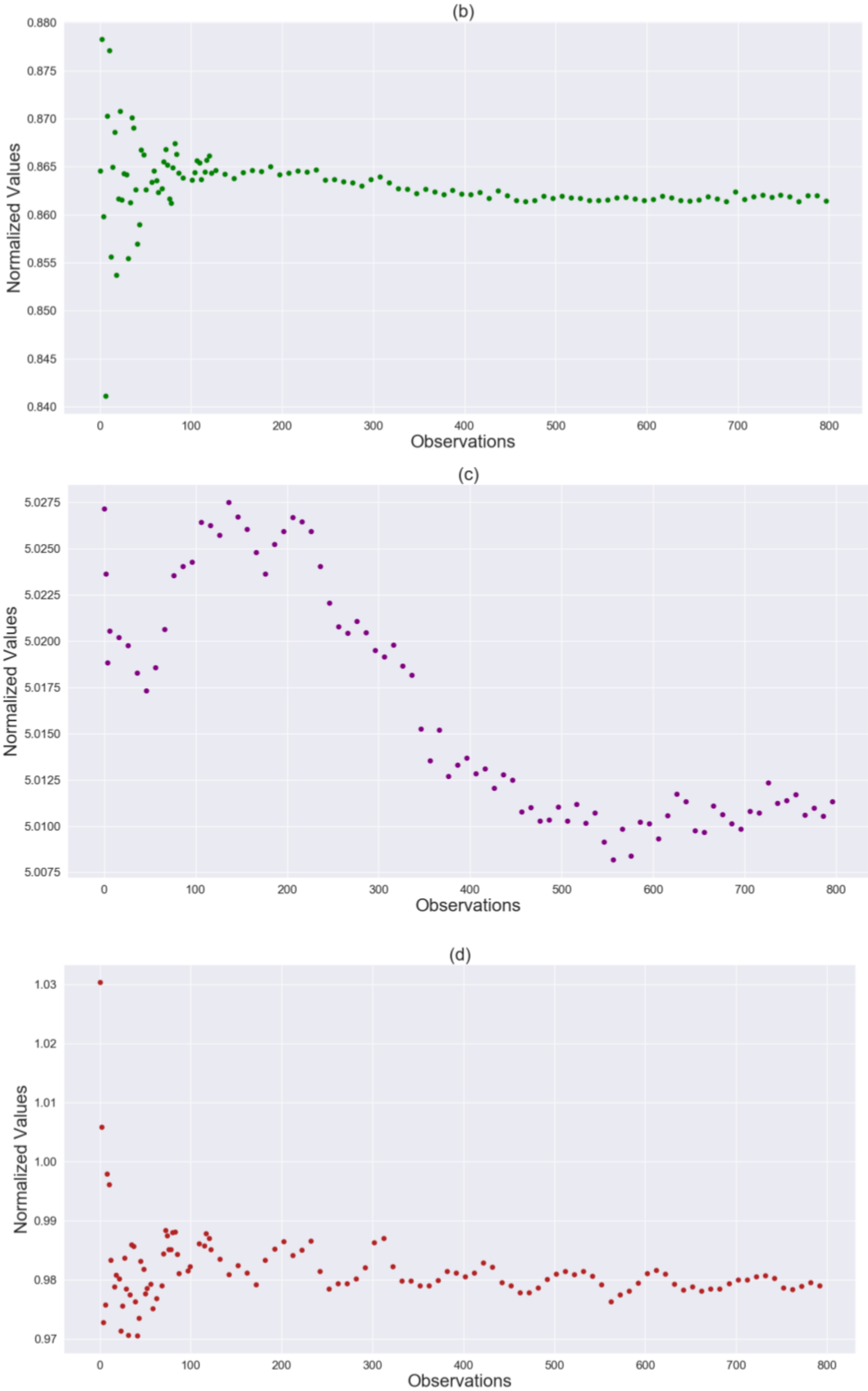


Fig. 4. 3: General behaviour of selected raw / normalized signals. (a) Normalised current magnitude of a bio mass generator (BMKW) at station 1D1 in case of 2D1.KW outage. (b) Normalised current magnitude of a diesel generator (DKW) at station 1D1 in case of 2D1.KW outage. (b) Normalised current magnitude of load (L1) at

station 1D1 in case of 2D1.KW outage. (d) Normalised current magnitude of line (L1_1) at station 1D1 in case of 2D1.KW outage.

Comparison of SCADA and PMU signal

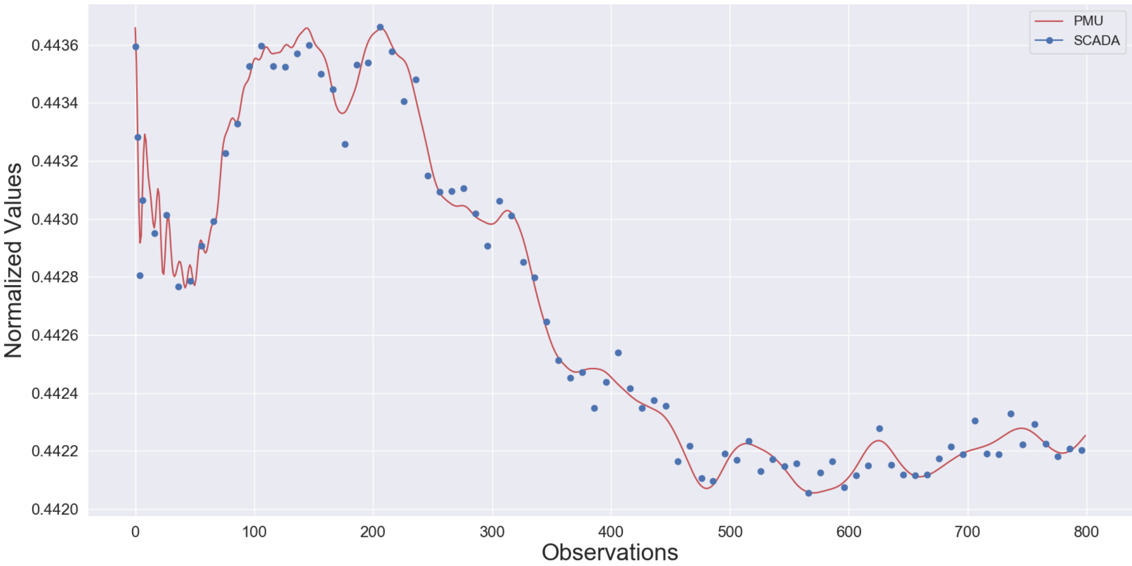


Fig. 4. 4: Comparison of SCADA vs. PMU signal

Figure 4.4 illustrates a current magnitude of a biomass generator (BMKW) located at station 1D1 measured by SCADA and PMU during a generator fault (2D1.DKW). It can be observed that SCADA signal is spontaneous as compared to PMU signal which is continuous. Moreover, a deviation is also observed for the same signal between the two measurement systems because of the introduced noise component – see equation (4.1). The increased sparseness and noise have also a negative impact on the classification accuracy when using SCADA data, which can be seen later.

Evaluation of features

After calculating statistical and time-based features using equations 3.2 and 3.8, a feature vector $\underline{x}_F$ (equation 4.3) with 240 features is created. This feature matrix contains computed features of all channels i.e., current, active power, reactive power, voltage and frequency magnitudes (arranged in the same order column wise).

In order to train the model effectively, an additional step of feature selection is performed after feature extraction (section 3.2) in training phase. However, in application phase features extracted from the input data will be directly fed to the classification model as described earlier in methodology chapter. Purpose of feature selection is to assess all features and remove correlated features. This results in further reduction of dimensionality of the data and increase in the model performance and the computational efficiency.

Feature Importance and Correlation (scenario 1.A)

Above mentioned feature matrix was further assessed by using feature importance scores and correlation analysis. The feature importance provide a score for each feature and are an additional result of the decision tree classifier. These score values represent the increase in the model's prediction error after permuting the feature. the rate of change of the classifier accuracy in case of small perturbations of the corresponding input features. High score values indicate a high relevance of the features. Figure 4.5 shows feature importance scores of all features. Features with low scores occur in the middle region of the plot, corresponds to active and reactive power channels. Moreover, correlation of all features visualized using a heat map also highlighted active and reactive power channels as highly correlated with current and voltage channels (see Appendix).

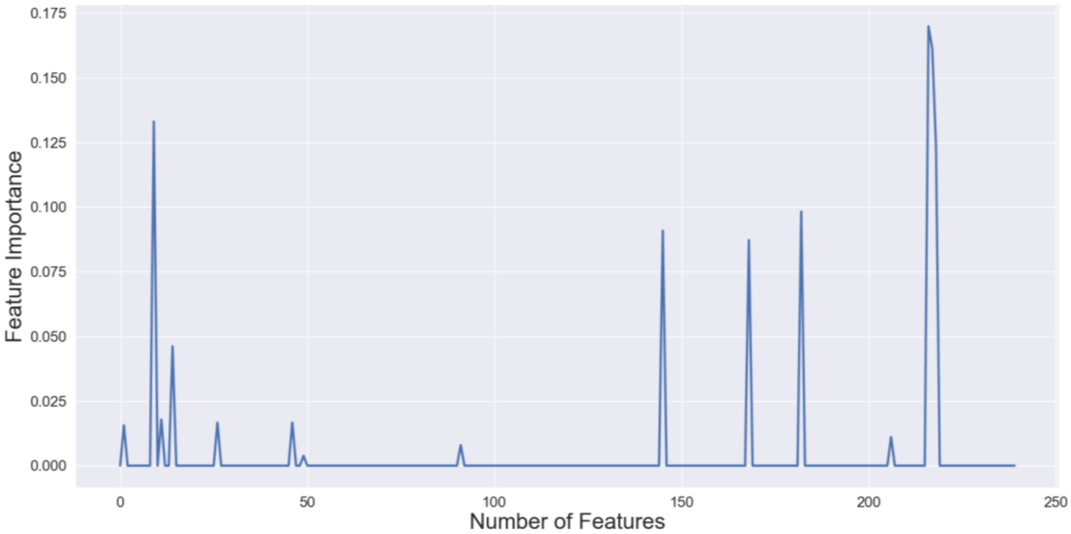


Fig. 4. 5: Feature Importance plot for Feature Selection.

As a result of feature selection, features of active and reactive power channels can be excluded from the feature matrix as it contains similar information from current and voltage channels. This reduced feature matrix refers to scenario 1.B, model performance based on this matrix is evaluated in next section.

Evaluation of DT classifier accuracies

The impact of different feature matrix combinations on classification accuracy is investigated in this section. The classification accuracy is given for the validation dataset after feeding the feature matrices defined in test scenarios into decision tree classifier. The classification results presented in Table 4.4 show that the decision tree classifier performs well on classifying given outages with accuracy values between 83% and 87%. However, no change in accuracy is observed after excluding features of active and reactive power channels. This confirms the previous investigations that active and reactive powers give no additional information to distinguish between the classes. A slight decrease in accuracy from 86.36% to 84% is observed, when time-based features of all channels are excluded. Consequently, the time-based feature has a small positive influence on the classification accuracy of about 2%. Further, overall influence of excluding time-based feature of all channels along with active and reactive powers is significantly low, around (1%). However, increasing sparseness and noise in SCADA data through respective parameters result in the lowest accuracy of the classifier with only 83%.

Tab. 4. 4: Classification accuracy for all test scenarios

	Feature Matrix	Accuracy (Validation)
1.A	Including all channels	86.36%
1.B	Excluding channels P and Q	86.36%
1.C	Excluding time-based feature	84%
1.D	Excluding P, Q and time-based feature	85%
1.E	Including all channels but sparseness and noise parameters are changed	83%

Conclusions for test phase 2 (scenario 1-4)

Comparing the accuracy and different combinations of feature matrix, classification results based on scenario 1.B where only P and Q channels are not included are proposed to fuse with the results of PMU classification to obtain final decision.

Test Phase 2

Test phase 2 focuses on the fusion of class probabilities $\underline{x}_P^S$ and $\underline{x}_N^P$ obtained from SCADA and PMU classifiers respectively. Fusion model is based on the logistic regression algorithm described in section 2.5. Initially, input parameters, contingencies, stations are defined. Then, in order to predict class label and evaluate model performance, test scenarios are defined. In the end, total accuracy of the model is compared with the accuracy of the individual models.

Definition of input parameters

For test phase 2, available PMU data is changed at the stations. In case of test phase 2, input to the fusion model are the class probabilities from SCADA and PMU classifiers instead of features. The fusion model selects the maximum probability out of both classifier outputs and provides final decision containing information about the type and location of the disturbance.

Test Scenarios

In order to evaluate the performance of the fusion model, test scenarios based on the availability of PMUs are defined within this test phase.

Table 4.5 shows general assumptions which include outages and stations considered within this test phase. Furthermore, four test scenarios are defined where scenario (2.A) is based on the minimum availability of the PMU and scenario (2.D) is based on maximum availability. Performance of the fusion model is evaluated for these scenarios, which is discussed later in the section.

Tab. 4. 5: General assumptions for test phase 2

	Parameter	Specification
General Assumptions	Outages	Generator Outages: ▪ 2D1.DKW

		▪ 5D3.GKW Two Line Outages: ▪ L_11 ▪ L_28
	Stations observed by SCAD	1D1, 2D1, 5D1

Tab. 4. 6: Test scenarios for test phase 2

Scenario	Parameter	Specification
2.A	Stations observed by PMU	1D1
2.B	Stations observed by PMU	1D1, 2D1
2.C	Stations observed by PMU	2D1, 5D1
2.D	Stations observed by PMU	1D1, 2D1 and 5D1

Model Hyperparameters

Table 4.7 shows the hyperparameter of logistic regression which is used for data fusion model. Logistic regression uses coordinate descent algorithm for optimization which supports in keeping the weights small, making the model simpler and avoiding overfitting.

Tab. 4. 7: Hyperparameter of Logistic Regression

Parameter	Description
Optimization	coordinate descent algorithm

Evaluation of SCADA classifier, PMU classifier and Fusion model accuracies

Table 4.8 shows accuracies of fusion model, SCADA and PMU classifier. For each test scenario, accuracy of the fusion model is compared with the accuracy of PMU classifier and SCADA classifier. For scenario 2.A, where PMU data is available at only one station, the accuracy of the PMU classifier is with 47.22 % significantly lower than the SCADA classifier with 86.36%. Hence, the fusion model focuses more on SCADA data and maintains a good total accuracy of about 90.62%. This can be observed in the model weights as well (See Appendix). The accuracy gain compared to the use of a single data source is about 4% in case of SCADA data and 43% in case of PMU data. Similarly, for scenarios 2.B and 2.C, fusion model maintains overall high accuracy values of about 93.75% taking into account both SCADA and PMU data. However, it should be noted that increase in the availability of PMU results in increase of accuracy of the model because the SCADA classifier configuration is kept constant over all scenarios. For scenario 2.D, where PMU is available at all three stations, no accuracy gain is achieved and the total accuracy is with

98.43% even lower compared to the accuracy of the PMU classifier with 100%. In this case a fusion of both classifiers has no advantage.

Tab. 4. 8: SCADA classifier, PMU classifier and Fusion model accuracies

Scenarios	SCADA	PMU	Fused
2.A	86.36%	47.22%	90.62%
2.B	86.36%	83.55%	93.75%
2.C	86.36%	79.27%	93.75%
2.D	86.36%	100%	98.43%

Figure 4.6 illustrates the accuracies of the individual models (SCADA and PMU) along with accuracy of the fusion model. From evaluation, it is cleared that fusion model provides improved system observability even when one measurement system is performing poorly. In that case, fusion model focuses on other measurement system and tends to maintain effective observability. Furthermore, it is cleared that fused information is useful to enhance the detection of the power system outages.

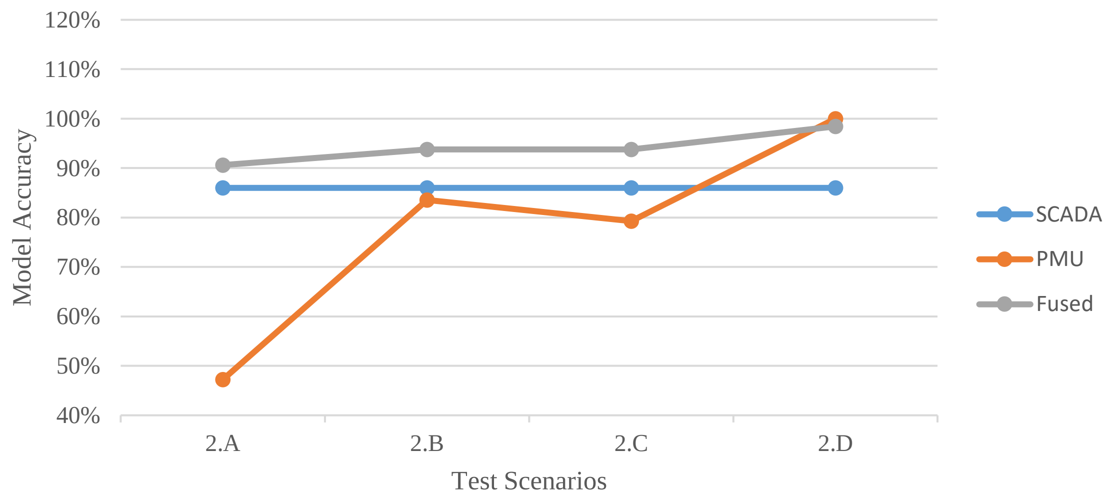


Fig. 4. 6: Accuracy plot of SCADA classifier, PMU classifier and Fusion model for all test scenarios.

5 Conclusion and Future Work

Summary and conclusions

This thesis has developed a novel algorithm which focuses on the fusion of PMU and SCADA data for control room application. Research of the state-of-the-art control room applications based on PMU and SCADA data revealed that a number of methodologies have been used for the fusion of data from PMU and SCADA. However, most of the research work have specifically focused on state estimation application. Therefore, this thesis focused to investigate other control room applications beside state estimation where data fusion techniques for PMU and SCADA data can be used. In this regard, literature review highlighted the need of enhanced security and reliability of measurement and control systems against a wide range of disturbances like outages of generators, transmission line, renewable energy plants or short circuits in power grid networks. Research work within this thesis presented an application of enhanced detection of power system outages with limited PMU data. Further, more applications identified within this thesis are listed in table (see Appendix).

The proposed methodology consists of two-stages based on machine learning techniques, where it classifies power system outages in the first stage and develops a novel method to fuse the classification results from SCADA and PMU. A final decision specifying type and location of the fault has been obtained. The trained PMU classifier was a prerequisite to this thesis work. Hence, the approach was to train a SCADA classifier using a database of offline contingencies, and fuse the classification results with the results of PMU classifier trained on the same database. Methodology has been implemented on a test power grid model simulated in power factory software. The database of contingencies has been generated from this power grid model with all possible generator and line contingencies. The training of SCADA classifier was performed using decision tree algorithm, with current, voltage, active power, reactive power, voltage and frequency magnitudes as measurement channels according to four types of outages in the system and their effects on three selected stations. An optimal set of features were extracted from the raw data. It was found that a feature matrix with the current, voltage, frequency as measurement channels provided a reasonable accuracy of 86.3 % for a minimum number of measurement channels.

In the second stage, class probabilities from both classifiers were fused using fusion model based on Logistic Regression algorithm. Evaluation of model was performed on scenarios based on the variation of PMU data. Results from the performance evaluation shown high accuracies gains achieved (up to 43%) compared to the case only using one data source (PMU), further more fusion model makes outage detection possible even when only few PMU sensors are available with satisfying accuracies over 93%.

From the evaluation of model, it was found that in case of SCADA data measurement channels, current, voltage and frequency are enough for detection. In case of PMU, frequency and voltage magnitudes are used for classification.

Visualization of results and integration into control rooms

Another important aspect based on the fusion of PMU and SCADA data arise from the completion of this study is the visualization of results to support operational operations in dynamic control rooms. The final decision obtained from the proposed methodology is a SCADA flag that is solely the class label with information about the type and location of the disturbance. These two information, probability values from PMU / SCADA classifier as well as from fusion model are important for an operator. Hence, visualization of these information will be useful for the operator.

Future work

However, possible future work identified within this thesis is as follow:

- Validate algorithms with real measurements
- Increase training set by integrating more OP points/contingencies/ stations
- As proposed application accounts for the enhanced detection of power system outages, therefore it can be evaluated for more scenarios to validate general enhancement.
- Try other classification models or data fusion algorithms. For example, majority voting algorithm can be used and validated in place of logistic regression to predict the maximum probability from SCADA and PMU classifiers.
- Cross validation can be used to minimize randomness and increase more consistency in results.

6 References

[1] M. Kezunovic and M. Knezev, “Temporal and Spatial Requirements for Optimized Fault Location,” in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, Waikoloa, HI, USA, Jan. 2008 - Jan. 2008, p. 169.

[2] M. Kezunovic and A. Abur, “Merging the Temporal and Spatial Aspects of Data and Information for Improved Power System Monitoring Applications,” *Proc. IEEE*, vol. 93, no. 11, pp. 1909–1919, 2005.

[3] I. Kamwa, R. Grondin, and Y. Hebert, “Wide-area measurement based stabilizing control of large power systems-a decentralized/hierarchical approach,” *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 136–153, 2001.

[4] I. Decker, D. Dotta, M. Agostini, S. Zimath, and A. Silva, “Performance of a Synchronized Phasor Measurements System in the Brazilian Power System,” in *2006 IEEE PES Power Systems Conference and Exposition, 2006: PSCE '06 ; Oct. 29 2006 - Nov. 1 2006, [Sheraton Atlanta, Atlanta, Georgia, USA*, Atlanta, Georgia, USA, 2006, p. 150.

[5] X. Xie, Y. Xin, J. Xiao, J. Wu, and Y. Han, “WAMS applications in Chinese power systems,” *IEEE Power and Energy Mag.*, vol. 4, no. 1, pp. 54–63, 2006.

[6] A. G. PHADKE, M. A. Pai, A. Stankovic, and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York, NY, Heidelberg: Springer, 2008.

[7] Priyanka Modi, “Modelling of PMU with various phasor estimation techniques in matlab,” Master of Engineering, Thapar University, Patiala, India, 2015.

[8] D. Dotta, J. H. Chow, L. Vanfretti, M. S. Almas, and M. N. Agostini, “A MATLAB-based PMU simulator,” in *IEEE Power and Energy Society general meeting (PES), 2013: 21 - 25 July 2013, Vancouver, BC, Canada*, Vancouver, BC, 2013, pp. 1–5.

[9] David A. Hart and David V. Uy and V. Vasudev Gharpure, Damir Novosel, Daniel Karlsson, and Mehmet Kaba, “PMUs - A new approach to power network monitoring,” ABB, 2001.

[10] S. Skok, I. Ivankovic, and Z. Cerina, “Applications Based on PMU Technology for Improved Power System Utilization,” in *IEEE Power Engineering Society general meeting, 2007: 24 - 28 June 2007, [Tampa, FL*, Tampa, FL, USA, 2007, pp. 1–8.

[11] J. Thorp, A. Phadke, and K. Karimi, “Real Time Voltage-Phasor Measurement For Static State Estimation,” *IEEE Trans. on Power Apparatus and Syst.*, vol. PAS-104, no. 11, pp. 3098–3106, 1985.

[12] J. S. Thorp, A. G. Phadke, S. H. Horowitz, and M. M. Begovic, “Some applications of phasor measurements to adaptive protection,” *IEEE Trans. Power Syst.*, vol. 3, no. 2, pp. 791–798, 1988.

[13] A. G. Phadke, “Synchronized phasor measurements-a historical overview,” in *IEEE/PES Transmission and Distribution Conference and Exhibition 2002, Asia Pacific: New wave of T & D technology from Asia Pacific : conference proceedings : Pacific Convention Plaza Yokohama, Yokohama, Japan, October 6-10, 2002*, Yokohama, Japan, 2002, pp. 476–479.

[14] A. G. Phadke *et al.*, “Synchronized sampling and phasor measurements for relaying and control,” *IEEE Trans. Power Delivery*, vol. 9, no. 1, pp. 442–452, 1994.

[15] R. O. Burnett *et al.*, “Synchronized phasor measurements of a power system event,” *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp. 1643–1650, 1994.

[16] C.-W. Liu, “Application of synchronised phasor measurements to real-time transient stability prediction,” *IEE Proc., Gener. Transm. Distrib.*, vol. 142, no. 4, p. 355, 1995.

[17] Pedro Emanuel Almeida Cardoso, “Deep Learning Applied to PMU Data in Power Systems,” FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO, 2017.

[18] Rohini Pradip Haridas, “Synchrophasor Measurement Technology in Electrical Power system,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 6, www.ijert.org, 2013.

[19] G. Yang, M. Gordon, A. H. Nielsen, and J. Østergaard, “PMU Applications - From Situation Awareness to Blackout Prevention,” Technical University of Denmark, Denmark, 2009.

[20] S. Chakrabarti and E. Kyriakides, “Optimal Placement of Phasor Measurement Units for Power System Observability,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1433–1440, 2008.

[21] M. M. Eissa, M. E. Masoud, and M.M.M. Elanwar, “A Novel Back Up Wide Area Protection Technique for Power Transmission Grids Using Phasor Measurement Unit,” *IEEE Trans. Power Delivery*, vol. 25, no. 1, pp. 270–278, 2010.

[22] D.R.B. K.Vinay Kumar, “Supervisory control and data acquisition (SCADA) in subtransmission and distribution levels in power systems,” https://www.researchgate.net/publication/270159134_SUPERVISORY_CONTROL_AND_DATA_ACQUISTION_SCADA_IN_SUBTRANSMISSION_AND_DISTRIBUTION_LEVELS_IN_POWER_SYSTEMS, 2013.

[23] T. Yang, H. Sun, and A. Bose, “Transition to a Two-Level Linear State Estimator—Part I: Architecture,” *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 46–53, 2011.

[24] Douglas C. Osburn, III, “Remote Terminal Unit,” US 6,628,992 B2 09/826,988, Sep 30, 2003.

[25] M. Aamir, J. Poncela, M. A. Uqaili, B. S. Chowdhry, and N. A. Khan, “Optimal Design of Remote Terminal Unit (RTU) for Wireless SCADA System for Energy Management,” *Wireless Pers Commun*, vol. 69, no. 3, pp. 999–1012, 2013.

[26] F. Enejo and A. Ogunrinde, “Review of Remote Terminal Unit (RTU) and Gateways for Digital Oilfield delpoyments,” *IJACSA*, vol. 3, no. 8, 2012.

[27] G. R. Clarke, D. Reynders, and E. Wright, *Practical modern SCADA protocols: DNP3, 60870.5 and related systems*. Amsterdam: Elsevier, 2004.

[28] M. Zhao and J. Sha, *Communications and Information Processing: International Conference, ICCIP 2012 Aveiro, Portugal, March 7-11, 2012 Revised Selected Papers, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[29] M. Ghosal and V. Rao, “Fusion of PMU and SCADA Data for dynamic state estimation of power system,” in *2015 North American Power Symposium (NAPS): 4 - 6 Oct. 2015, The University of North Carolina at Charlotte, Charlotte, NC, USA*, Charlotte, NC, USA, 2015, pp. 1–6.

[30] K. E. Martin *et al.*, “IEEE Standard for Synchrophasors for Power Systems,” *IEEE Trans. Power Delivery*, vol. 13, no. 1, pp. 73–77, 1998.

[31] David G. Hart *et al.*, “PMUs- a new approach to power network monitoring.,” ABB, 2001.

[32] M. Mao, F. Mei, P. Jin, and L. Chang, “Application of IEC61850 in energy management system for microgrids,” in *IEEE 5th International Symposium on Power Electronics for Distributed Generation Systems (PEDG), 2014: 24 - 27 June 2014, Galway, Ireland*, Galway, Ireland, 2014, pp. 1–5.

[33] C. Wester, M. Adamiak, and J. Vico, “IEC61850 protocol - practical applications in industrial facilities,” in *IEEE Industry Applications Society annual meeting (IAS), 2011: 9 - 13 Oct. 2011, Orlando, FL, USA*, Orlando, FL, USA, 2011, pp. 1–7.

[34] D. L. Hall and J. Llinas, “An introduction to multisensor data fusion,” *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[35] F. Castanedo, “A review of data fusion techniques,” (eng), *TheScientificWorldJournal*, p. 704504, 2013.

[36] M. Kordestani and M. Saif, “Data fusion for fault diagnosis in smart grid power systems,” in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE): 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) took place 30 April-3 May 2017 in Windsor, ON, Canada*, Windsor, ON, 2017, pp. 1–6.

[37] K. Salahshoor, M. Kordestani, and M. S. Khoshro, “Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers,” *Energy*, vol. 35, no. 12, pp. 5472–5482, 2010.

[38] R. C. King *et al.*, “Application of data fusion techniques and technologies for wearable health monitoring,” (eng), *Medical engineering & physics*, vol. 42, pp. 1–12, 2017.

[39] J. H. Jang, Y. S. Kim, and J. B. Ra, “Image enhancement in multi-resolution multi-sensor fusion,” in *IEEE Conference on Advanced Video and Signal Based Surveillance, 2007: AVSS 2007 ; 5 - 7 Sept. 2007, London, United Kingdom ; proceedings*, London, UK, 2007, pp. 289–294.

[40] M. Hongxia and W. Honghua, “Research on fault diagnosis method of high-voltage circuit breaker based on fuzzy neural network data fusion,” in *International Conference on Computer Application and System Modeling (ICCASM), 2010: Taiyuan, China, 22 - 24 Oct. 2010*, Taiyuan, China, 2010, V11-231-V11-235.

[41] X. Yu and C. Singh, “A Practical Approach for Integrated Power System Vulnerability Analysis With Protection Failures,” *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1811–1820, 2004.

[42] Y. N. Wang *et al.*, “Novel hierarchical fault diagnosis approach for smart power grid with information fusion of multi-data resources based on fuzzy petri net,” in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2014: 6 - 11 July 2014, Beijing, China*, Beijing, China, 2014, pp. 1183–1189.

[43] R. Nowak, R. Biedrzycki, and J. Misiurewicz, “Machine learning methods in data fusion systems,” in *13th International Radar Symposium (IRS), 2012: 23 - 25 May 2012, Warsaw, Poland ; conference proceedings ; part of the 5th Microwave and Radar Week (MRW 2012)*, Warsaw, Poland, 2012, pp. 400–405.

[44] H. S. Carvalho, W. B. Heinzelman, A. L. Murphy, and C.J.N. Coelho, “A general data fusion architecture,” in *Proceedings of the Sixth International Conference on Information Fusion:*

FUSION 2003 : July 8-11, 2003, Radisson Hotel, Cairns, Queensland, Australia, Cairns, Queensland, Australia, 2003, pp. 1465–1472.

[45] B. V. Dasarathy, “Sensor fusion potential exploitation-innovative architectures and illustrative applications,” *Proc. IEEE*, vol. 85, no. 1, pp. 24–38, 1997.

[46] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[47] Y. Gu, M. Yang, Z. Shi, and Z. Wu, “The Applications of Decision-Level Data Fusion Techniques in the Field of Multiuser Detection for DS-UWB Systems,” (eng), *Sensors (Basel, Switzerland)*, vol. 15, no. 10, pp. 24771–24790, 2015.

[48] B. V. Dasarathy, “Decision fusion benefits assessment in a three-sensor suite framework,” *Opt. Eng*, vol. 37, no. 2, p. 354, 1998.

[49] M. M. Kokar, J. A. Tomasik, and J. Weyman, “Formalizing classes of information fusion systems,” *Information Fusion*, vol. 5, no. 3, pp. 189–202, 2004.

[50] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer, 2009.

[51] D. Barber, *Bayesian Reasoning and Machine Learning*, 2nd ed. Cambridge: Cambridge University Press, 2012.

[52] C. M. Bishop, *Pattern recognition and machine learning*. New York, NY: Springer, 2006.

[53] W. Richert and L. P. Coelho, *Building machine learning systems with Python: Master the art of machine learning with Python and build effective machine learning systems with this intensive hands-on guide*. Birmingham: Packt Publishing, 2013.

[54] S. Sumathi and S. N. Sivanandam, *Introduction to Data Mining and its Applications*. Berlin, Heidelberg: Springer, 2006.

[55] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining. Concepts and Techniques, 3rd Edition: The Morgan Kaufmann Series in Data Management Systems*, 3rd ed. 225 Wyman Street, Waltham, MA 02451, USA: TheMorgan Kaufmann, 2012.

[56] Teeuwsen *et al.*, *Eds.*, *Decision Tree Based Oscillatory Stability Assessment for Large Interconnected Power Systems*, 2004.

[57] A. Brunello, E. Marzano, A. Montanari, and G. Sciavicco, “J48SS: A Novel Decision Tree Approach for the Handling of Sequential and Time Series Data,” *Computers*, vol. 8, no. 1, p. 21, 2019.

[58] J.R. Quinlan, “Induction of Decision Trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[59] P. K. Ray, S. R. Mohanty, N. Kishor, and J. P. S. Catalao, “Optimal Feature and Decision Tree-Based Classification of Power Quality Disturbances in Distributed Generation Systems,” *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 200–208, 2014.

[60] v. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, “An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, 2005.

[61] S. Ramos and Z. Vale, “Data mining techniques application in power distribution utilities,” in *IEEE/PES Transmission and Distribution Conference and Exposition, 2008: T & D ; 21 - 24 April 2008, [Chicago, IL*, Chicago, IL, USA, 2008, pp. 1–8.

[62] A. Mosavi, “Multiple Criteria Decision-Making Preprocessing Using Data Mining Tools,” *IJCSI*, http://arxiv.org/pdf/1004.3258v1.

[63] David W. Hosmer Jr., Stanley Lemeshow, and Rodney X. Sturdivant, *Applied Logistic Regression, Third Edition*, 3rd ed.: John Wiley & Sons, Inc., 2013.

[64] Daryl Pregibon, “Logistic Regression Diagnostics,” *The Annals of Satistics*, vol. 9, no. 4, pp. 705–724, 1981.

[65] Robert T. Olszewski, “Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data,” Doctor ofPhilosophy, Carnegie Mellon University, Pittsburgh, PA 15213, February/2001.

[66] Asutosh Kar and Leena Das, Eds., *A Technical Review on Statistical Feature Extraction of ECG signal*: IJCA, 2011.

[67] O. P. Dahal and S. M. Brahma, “Preliminary work to classify the disturbance events recorded by phasor measurement units,” in *2012 IEEE Power and Energy Society General Meeting*, San Diego, CA, Jul. 2012 - Jul. 2012, pp. 1–8.

[68] M. Biswal *et al.*, “Signal features for classification of power system disturbances using PMU data,” in *19th Power Systems Computation Conference: PSCC 2016 Genova : June 20-24, 2016, Centro congressi Porto Antico, Magazzini del Cotone Conference Centre, Genova, Italy*, Genoa, Italy, 2016, pp. 1–7.

[69] Bilal Esmael1, Arghad Arnaout2, Rudolf K. Fruhwirth2, and Gerhard Thonhauser1, “A Statistical Feature-Based Approach for Operations Recognition in Drilling Time Series,” *International Journal of Computer Information Systems and Industrial Management Applications.*, vol. 5, no. 2150-7988, pp. 454–461, 2013.

[70] M. Akke and J. S. Thorp, “Sample Value Adjustment Improves Phasor Estimation at Off-Nominal Frequencies,” *IEEE Trans. Power Delivery*, vol. 25, no. 4, pp. 2255–2263, 2010.

[71] I. C. Decker, M. N. Agostini, A. S. e Silva, and D. Dotta, “Monitoring of a large scale event in the Brazilian Power System by WAMS,” in *IREP Symposium Bulk Power System Dynamics and Control - VIII (IREP), 2010: 1 - 6 Aug. 2010, Buzios, Rio de Janeiro, Brazil*, Rio de Janeiro, Brazil, 2010, pp. 1–8.

[72] C. A. Ruiz, N. J. Orrego, and J. F. Gutierrez, “The Colombian 2007 black out,” in *IEEE/PES Transmission and Distribution Conference and Exposition: Latin America, 2008: Bogota, Columbia, 13 - 15 August 2008*, Bogota, Colombia, 2008, pp. 1–5.

[73] Luis Fabiano dos Santos, Galina Antonova, Mats Larsson, Sergio Fujii, “The Use of Synchrophasors for Wide Area Monitoring of Electrical Power Grids,” 2015.

[74] N. Arghira, D. Hossu, I. Făgărăṣan, S. S. Iliescu, and D. R. Costianu, “MODERN SCADA PHILOSOPHY IN POWER SYSTEM OPERATION – A SURVEY,” *U.P.B. Sci. Bull*, vol. 73, no. 2, pp. 153–166, 2011.

[75] W. U. Rahman *et al.*, “Advancement in Wide Area Monitoring Protection and Control Using PMU’s Model in MATLAB/SIMULINK,” *SGRE*, vol. 03, no. 04, pp. 294–307, 2012.

[76] M. Parashar, A. Jampala, and J. Giri, “Practical Experiences in Operationalizing WAMS in the Control Room,” *IFAC-PapersOnLine*, vol. 49, no. 27, pp. 91–95, 2016.

[77] M. U. USMAN and M. O. FARUQUE, “Applications of synchrophasor technologies in power systems,” *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 2, pp. 211–226, 2019.

[78] G. Sanchez-Ayala, J. R. Aguerc, D. Elizondo, and M. Lelic, “Current trends on applications of PMUs in distribution systems,” in *IEEE PES Innovative Smart Grid Technologies Conference (ISGT), 2013: 24 - 27 Feb. 2013, Washington, DC*, Washington, DC, 2013, pp. 1–6.

[79] J. Bertsch, M. Zima, A. Suranyi, C. Carnal, and C. Rehtanz, “Experiences with and perspectives of the system for wide area monitoring of power systems,” in *Quality and security of electric power delivery systems: CIGRE/PES 2003 : 08-10 October 2003, Montréal, Canada*, Montreal, Quebec, Canada, 2004, pp. 5–9.

[80] E. Allen, “Real-Time Application of Synchrophasors for Improving Reliability,” North American Electric Reliability Corporation, Sep. 2010.

[81] A. G. PHADKE, P. WALL, L. DING, and V. TERZIJA, “Improving the performance of power system protection using wide area monitoring systems,” *J. Mod. Power Syst. Clean Energy*, vol. 4, no. 3, pp. 319–331, 2016.

[82] V. TERZIJA *et al.*, “Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks,” *Proc. IEEE*, vol. 99, no. 1, pp. 80–93, 2011.

[83] S. Chakrabarti, E. Kyriakides, T. Bi, D. Cai, and V. Terzija, “Measurements get together,” *IEEE Power and Energy Mag.*, vol. 7, no. 1, pp. 41–49, 2009.

[84] H. Saadat, *Power system analysis*, 3rd ed. [S.l.]: PSA Publishing, 2010.

[85] Y. Ebata, H. Hayashi, Y. Hasegawa, S. Komatsu, and K. Suzuki, “Development of the Intranet-based SCADA (supervisory control and data acquisition system) for power system,” in *Power Engineering Society Winter Meeting, 2000 IEEE*, Singapore, Mar. 2000, pp. 1656–1661.

[86] B. Qiu and H. B. Gooi, “Web-based SCADA display systems (WSDS) for access via Internet,” *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 681–686, 2000.

[87] M. M. Ahmed and W. L. Soo, “Customized SCADA system for low voltage distribution automation system,” in *Transmission & Distribution Conference & Exposition: Asia and Pacific, 2009: IEEE T&D Asia 2009 ; 26 - 30 Oct. 2009, Seoul, Korea (South)*, Seoul, South Korea, 2009, pp. 1–4.

[88] M. Kezunovi, T. Djoki, and T. Kosti, “Automated Monitoring and Control Using New Data Integration Paradigm,” in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences: Abstracts and CD-ROM of full papers : 3-6 January, 2004, Big Island, Hawaii*, Big Island, HI, USA, 2005, 66a-66a.

[89] B. Qiu, H. B. Gooi, Y. Liu, and E. K. Chan, “Internet-based SCADA display system,” *IEEE Comput. Appl. Power*, vol. 15, no. 1, pp. 14–19, 2002.

[90] A. Creery and E. J. Byres, “Industrial Cybersecurity for Power System and Scada Networks,” in *Record of conference papers: Industry Applications Society 52nd Annual Petroleum and*

Chemical Industry Conference, Denver, Colorado, September 12-14, 2005, Denver, CO, USA, 2005, pp. 303–309.

[91] E. Chikuni and M. Dondo, "Investigating the security of electrical power systems SCADA," in *Africon 2007: 26 - 28 Sept. 2007, Windhoek, South Africa*, Windhoek, South Africa, 2007, pp. 1–7.

[92] M. Kezunovic, C. Zheng, and C. Pang, "Merging PMU, Operational, and Non-Operational Data for Interpreting Alarms, Locating Faults and Preventing Cascades," in *43rd Hawaii International Conference on System Sciences (HICSS), 2010 ; Honolulu, Hawaii, 5 - 8 Jan. 2010*, Honolulu, Hawaii, USA, 2010, pp. 1–9.

[93] W. R. Prince, B. F. Wollenberg, and D. B. Bertoignolli, "Survey on excessive alarms (power system control)," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 950–956, 1989.

[94] X. Luo and M. Kezunovic, "Implementing Fuzzy Reasoning Petri-Nets for Fault Section Estimation," *IEEE Trans. Power Delivery*, vol. 23, no. 2, pp. 676–685, 2008.

[95] C.-S. Chen, C.-W. Liu, and J.-A. Jiang, "A new adaptive PMU based protection scheme for transposed/untransposed parallel transmission lines," *IEEE Trans. Power Delivery*, vol. 17, no. 2, pp. 395–404, 2002.

[96] M. Kezunovic and B. Perunicic, "Automated transmission line fault analysis using synchronized sampling at two ends," *IEEE Trans. Power Syst.*, vol. 11, no. 1, pp. 441–447, 1996.

[97] A. Mao, J. Yu, and Z. Guo, "PMU Placement and Data Processing in WAMS that Complements SCADA," in *2005 IEEE Power Engineering Society General Meeting*, San Francisco, CA, USA, Jan. 2005, pp. 690–693.

[98] A. Simoes Costa, A. Albuquerque, and D. Bez, "An estimation fusion method for including phasor measurements into power system real-time modeling," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1910–1920, 2013.

[99] J. James and B. S, "Hybrid State Estimation including PMU measurements," in *2015 International Conference on Control, Communication & Computing India (ICCC): 19-21 November 2015, Trivandrum, Kerala, India*, Trivandrum, Kerala, India, 2015, pp. 309–313.

[100] S. A. Dowi and G. Li, "A new approach for including synchronized phasor measurements in dynamic state estimation," in *2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013: 23 - 24 Dec. 2013, Toronto, ON, Canada ; proceedings*, Toronto, ON, Canada, 2013, pp. 389–394.

[101] S. Chakrabarti, E. Kyriakides, G. Valverde, and V. TERZIJA, "State estimation including synchronized measurements," in *IEEE Bucharest PowerTech, 2009: June 28, 2009 - July 2, 2009, Bucharest, Romania*, Bucharest, Romania, 2009, pp. 1–5.

[102] R. Kumar and D. Prasad, "Effect of bad measurements on state estimation in power system," in *IEEE Fifth Power India Conference, 2012: 19 - 22 Dec. 2012, Murthal, Haryana, India*, Murthal, Haryana, India, 2012, pp. 1–5.

[103] W. Gao and S. Wang, "On-line dynamic state estimation of power systems," in *North American Power Symposium (NAPS), 2010: 26 - 28 Sept. 2010, Arlington, TX, USA*, Arlington, TX, USA, 2010, pp. 1–6.

[104] M. Meriem *et al.*, “Study of state estimation using weighted-least-squares method (WLS),” in *Proceedings of 2016 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, Marrakech & Bengrir, Morocco, 2016, pp. 1–5.

[105] L. U. Kumar and A. R. Devi, “Power System Dynamic State Estimation Using Kalman Filtering Technique,” in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, Jul. 2018 - Jul. 2018, pp. 1173–1178.

[106] R. Fan, Z. Huang, S. Wang, R. Diao, and D. Meng, “Dynamic state estimation and parameter calibration of a DFIG using the ensemble Kalman filter,” in *2015 IEEE Power & Energy Society general meeting: Denver, Colorado, USA, 26 - 30 July 2015*, Denver, CO, USA, 2015, pp. 1–5.

[107] G. A. Ortiz, D. G. Colome, and J. J. Q. Puma, “State estimation of power system based on SCADA and PMU measurements,” in *Proceedings of the 2016 IEEE ANDESCON: Arequipa, Peru, 19-21 October 2016*, Arequipa, Peru, 2016, pp. 1–4.

[108] Institute of Electrical and Electronics Engineers; IPEC; International Power Engineering Conference, *2007 conference proceedings / IPEC: Singapore, 3 - 6 December 2007*. Piscataway: IEEE Service Center, 2007.

[109] H. Saadabadi and M. Dehghani, “Large-scale power systems state estimation using PMU and SCADA data,” in *2016 24th Iranian Conference on Electrical Engineering (ICEE): Shiraz University, Iran, 10-12 May 2016*, Shiraz, Iran, 2016, pp. 906–911.

[110] C. Yang, Y. Li, B. Yang, M. Du, and S. Wang, “Dynamic variable-weight least squares for state estimation of distribution network based on data fusion,” in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2): Proceedings : Nov. 27-28, 2017, Beijing, China*, Beijing, 2017, pp. 1–5.

[111] M. Zhou, V. A. Centeno, J. S. Thorp, and A. G. Phadke, “An Alternative for Including Phasor Measurements in State Estimators,” *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1930–1937, 2006.

[112] A. Monticelli, “Electric power system state estimation,” *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[113] Vedran Kirinčić, Srđan Skok, and Dubravko Franković, “A State Estimator Using SCADA and Synchronized Phasor Measurements,” *International Journal of Electrical and Computer Engineering Systems*, vol. 7, no. 2, 2016.

[114] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, “Learning time-series shapelets,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, New York, New York, USA, 2014, pp. 392–401.

A. List of Abbreviations

PMU	Phasor Measurement Unit
SCADA	Supervisory Control and Data Acquisition
WAMS	Wide Area Monitoring Systems
IED	Intelligent Electronic Devices
DT	Decision Tree

B. List Of Figures and Tables

C. List of Symbols

Notation

The notation used in this work is described below using the letter "a" as an example. All variables are represented by italic letters with Cambria Math (a). A matrix/tensor is marked with a bold capital letter with italics ($\boldsymbol{A}$). Vectors are represented by a under bar ($\underline{a}$). Scalar variables are represented in physical quantities as italics (a).

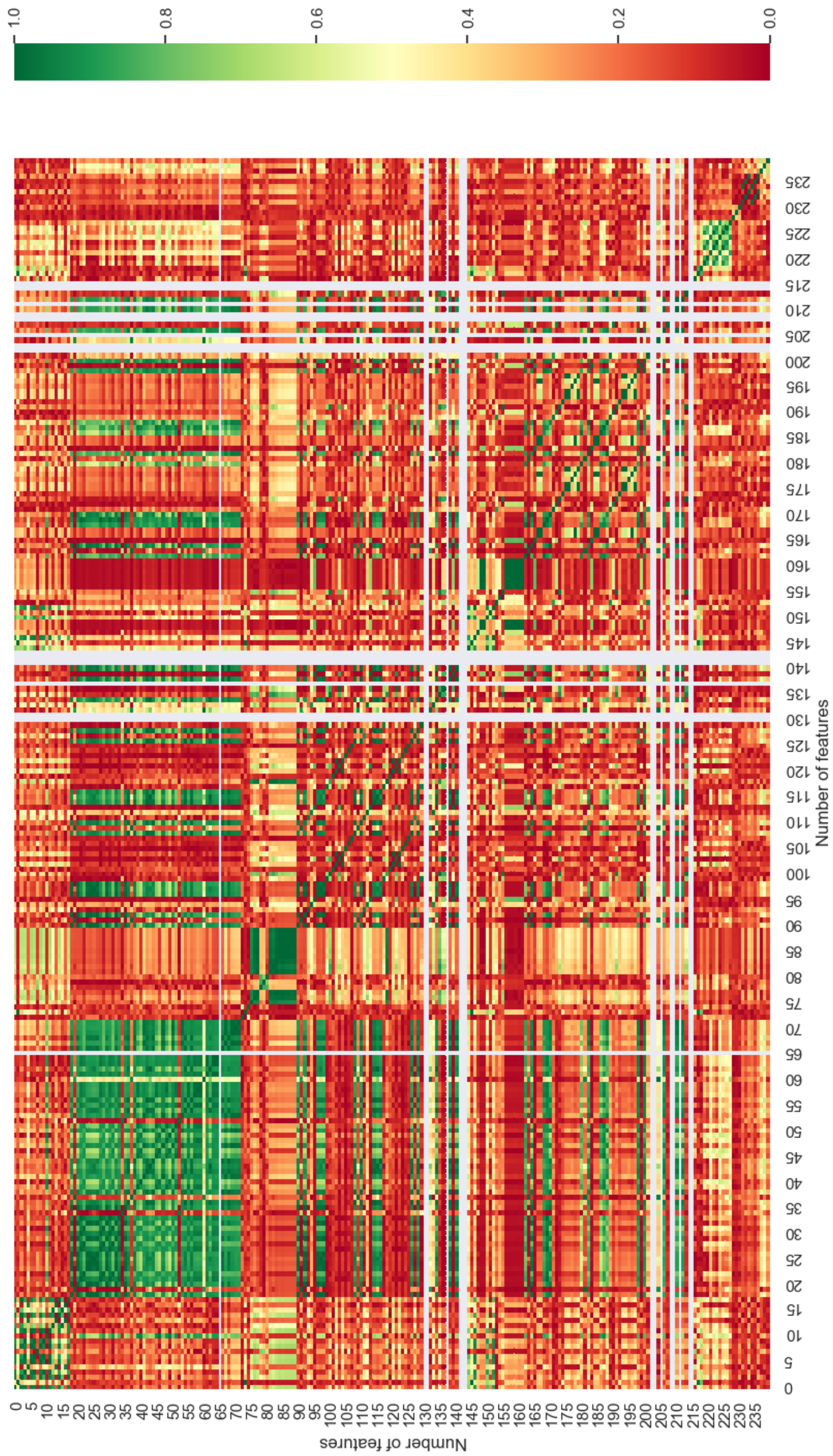
List of Symbols

T	binary tree
k	index number
t	tree node
t_1	root node
b_0	intercept
b_1	coefficient for the single input value
μ	arithmetic mean
σ^2	variance
σ	standard deviation
S	SCADA
P	PMU
$\mathbf{X}^{\mathbf{S}}$	SCADA measurement matrix
$\mathbf{X}^{\mathbf{P}}$	PMU measurement matrix
$\boldsymbol{X}_N^S$	normalized measurement matrix
d	sample size
n	no. of stations*channels
$\underline{x}_N^S$	single channel (SCADA)
$\underline{x}_F^S$	feature vector (SCADA)
x_F	single feature vector (SCADA)
$\underline{S}_N$	time series signal
$\underline{x}_P^S$	probability vector (SCADA)
$\underline{x}_P^P$	probability vector (PMU)
$\hat{y}^S$	final class assignment
$\underline{\theta}_C$	classifier parameters
M	minimum distance between a time series
W	weight
x_t^{tS}	SCADA measurement for specific channel C at time t
x_t^{tP}	phasor measurement for specific channel C at time t
$\underline{x}_t^P$	phasor measurement for all channels at time t

x_e	additive Gaussian error drawn from normal distribution
d_{ch}	minimum data change required for generating new scada value
d_{tmin}	minimum time range required for generating new scada value
Δt	time range between two subsequent scada values
σ_e	standard deviation of Gaussian error (white noise)
p_{ch}	parameter to control minimum data change threshold
p_e	parameter to control Gaussian error (white noise)

D. Appendix

Correlation matrix using heat map



Applications based on the fusion of PMU and SCADA data

No.	Application	Steps to be performed
1	Increase robustness against data manipulations when detecting generator and line outages	• Definition of contingencies • Train classifier on PMU and SCADA data basis • Manipulation of SCADA data to influence the Classification results. • Compare classifier accuracy with and without the usage of SCADA data.
2	Distinguish dynamic patterns which are created by outages or normal operations	• Introduction of SCADA flag • Visualization/Classification
3	Synthesis/estimation of PMU data at stations where no PMU measurements exists using SCADA data	• Definition of contingencies • Variation of available PMU measurements • Use SCADA data to estimate PMU signals at non-observed stations. • Compare estimated PMU signals with • Simulated ones.
4	Creation of a robust heterogenous PMU and SCADA data measurement database in presence of data manipulations and IT failures (data loss).	• Definition of contingencies • Variation of disturbed PMU and SCADA measurements • Generate different data manipulation signals • Robust estimation of PMU and SCADA values.

E. Ehrenwörtliche Erklärung

Ich versichere, dass ich diese wissenschaftliche Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen sind, wurden in jedem einzelnen Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht. Das gleiche gilt auch für die beigegebenen Skizzen und Darstellungen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Ilmenau, den 14.07.2019

Unterschrift

____

Vorname Nachname