# The Grid**Visibility** Platform: Enabling Artificial Intelligence and Machine Learning in the Distribution Grid

**Ralph W Brown** | brown@gridvisibility.com

# Introduction

It is well understood that the quality of data used in any Artificial Intelligence (AI) or Machine Learning (ML) application is critical to the accuracy and effectiveness of the results. This is true for the data that is used to train the AI/ML model as well as for the data that is fed to the model once it is deployed (serving data). Once a model has been deployed, serving data is often used to refine and improve the model continuously. Since the results generated from deployed AI/ML models may be used in planning and operational decision making, getting the data right (both in training and in deployment) is essential for making the best planning and operational decisions. This is particularly true when considering the use of AI/ML applications in the distribution grid that are necessary to "keep the lights on".

While the use of big data analytics and AI/ML for the grid has been a topic of research since at least 2015 [1], the primary challenge preventing wide scale adoption of AI/ML methods in planning and operations has been consistent access to high-quality data across the distribution grid. The current instrumentation available to distribution grid operators, e.g., Distribution-level Phasor Measurement Units (D-PMUs, a.k.a. micro-PMUs or μPMUs), Harmonic Phasor Measurement Units (H-PMUs), Waveform Measurement Units (WMUs), and Advanced Metering Infrastructure (AMI) are not able to provide consistent, high-quality data required for training and deployment of AI/ML methods at scale in planning and operations of the distribution grid. A primary challenge in making use of this instrumentation is the lack of a secure and reliable communications infrastructure with sufficient bandwidth to carry the serving data from the instrumentation to the centralized platforms that execute the deployed AI/ML models.

This paper provides a review of the literature on AI/ML applications in the power grid, a review of the factors affecting data quality, an evaluation of how the limitations of current instrumentation impact data quality, and a novel approach to efficiently obtain high quality, synchronous waveform or Continuous Point On Wave (CPOW) data from the distribution grid at scale.

# Summary of Literature Review

The application of big data analytics or AI/ML to detection of faults in the distribution grid has been the topic of numerous technical papers beginning as early 2010. Russell and Benner [1] early on recognized the benefits of waveform measurements, data analytics, and automation in detecting and isolating incipient and other faults in the distribution grid. They also identified challenges the industry faced in leveraging the insights that these waveform measurements could provide, including limited communications capacity, management of large volumes of data, and the need for automation to extract the relevant information necessary for distribution operators to remediate faults in the distribution grid in a timely manner. These early insights have been reinforced in technical papers ([2] through [19]) over the intervening years. In reviewing these technical papers several themes emerge (a detailed review of these papers follows the conclusions):

1. The value of waveform data in identifying different types of faults in the distribution grid, as more data is analyzed more types of faults emerge as being uniquely identifiable
2. The value of precise time synchronization of this waveform data across multiple sensors in the distribution grid, the ability to correlate events across sensors provides unique insights
3. The challenges of managing (e.g., transporting, processing, and storing) the volume of data represented by this waveform data, the classic big data problem

With the emergence of AI/ML algorithms to deal with very large volumes of data (theme 3 above), a number of researchers have explored applying AI/ML algorithms to identifying, classifying, and isolating faults in the power grid (papers [20] through [26]). In addition to the three themes identified above, three additional themes emerge from a review of these seven AI/ML focused papers (a detailed review of these papers also follows the conclusions):

4. The dearth of existing training data available (limited data sets or simulations were used) and even what is available represents sparse data

5. The challenges in obtaining serving data for use in the trained AI/ML models (ditto 4 above)

6. The challenges in obtaining quality data for either training or serving applications (limitations in the data available to the researchers)

These themes identify the need for consistent access to high quality data, both for training and for serving data. It is important to understand how training and serving are used to appreciate the importance of data quality to the results of AI/ML algorithms, particularly in the context of the distribution grid.

# Data Quality for AI/ML in the Context of the Distribution Grid

As described by M. T. Jones [27] there are generally three models of learning used in AI/ML algorithms: supervised learning, unsupervised learning, and reinforcement learning, as shown in Figure 1. Supervised learning involves the use of labeled training data that identifies the desired outputs for given inputs. Unsupervised learning does not require labeled training data and does not produce specific desired outputs. Reinforcement learning enables learning from feedback, for example refinements generated from the incorporation of serving data back into the ML model.

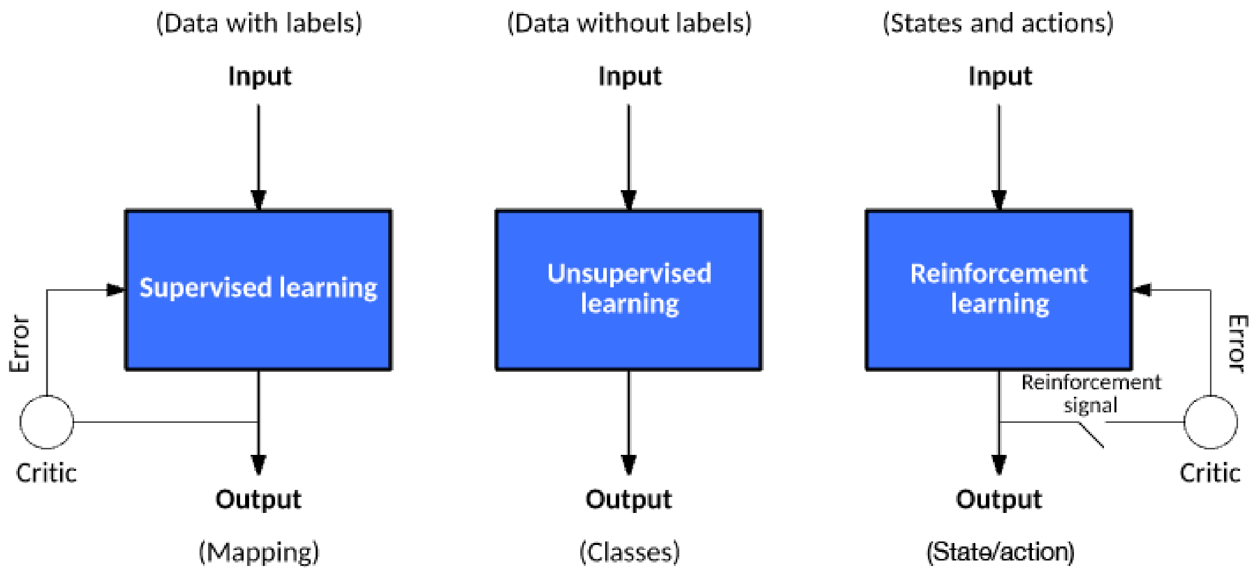


**Figure 1 - Three learning models for algorithms [27]**

While there is an intuitive understanding of the impact of data quality (or the lack thereof, a.k.a. garbage in, garbage out) on the results of AI/ML algorithms, there are several aspects of data quality that are generally recognized by the AI/ML community as being paramount. The impact of data quality on the performance of AI/ML applications has been studied in other contexts (e.g., banking credit, telecommunication provider customer churn, home sales, automobile sales price, etc. [28]) and focus on these attributes of data quality.

- Accuracy – Errors in training or serving data can result in erroneous or misleading results from AI/ML models
- Completeness – Incomplete, or sparse, training or serving data can either create incomplete AI/ML models or miss critical inputs, also resulting in erroneous or misleading results
- Consistency – Training and serving data should be uniform across different data sources, inconsistencies can result in errors in analysis

- Timeliness – Training and serving data must be current and up to date to ensure relevance, outdated information can skew results and diminish the model's effectiveness
- Bias – Selection of training data can skew AI/ML models, making it crucial to ensure representativeness in training data

These aspects of data quality apply to AI/ML generally and often apply to unstructured datasets, such as images, sound, or text. They can also be interpreted in the context of signal processing applications such as those related to synchronous waveform analysis in the distribution grid. S. R. Salkuti, et.al., [29] provide a survey of ML algorithms in the context of signal processing and identify five challenges of big data in its application to AI/ML algorithms in conjunction to signal processing: volume, variety, velocity, veracity, and value. These challenges translate directly to signal processing the context of the distribution grid and highlight the factors that impact the data quality in this context:

- Continuous availability – While the training data can be snippets of measurements that are relevant to the types of events of interest (such as those from triggered event captures), not all types of events, e.g., incipient events, are captured leading to missing training data. These snippets of data represent sparse data and relying sparse data in deployment introduces greater risk, as critical events could be missed, negatively impacting operations. As identified in [20], continuous monitoring is required to provide high quality training data and high-quality serving data for the deployed AI/ML models.
- Location – Obtaining data from only a subset of the distribution grid (e.g., from a few feeders at a few substations) does not provide an accurate view of the entire distribution grid. Faults on an un-instrumented portion of the distribution grid may not be accurately reflected on those portions that are instrumented, which can lead to poor planning and operational decision making. This again is missing data that can negatively impact the results of the training and deployed models.
- Feature accuracy – There are several factors affecting the feature accuracy. First, the sampling frequency and resolution of the data must be sufficient to capture the events of interest. For example, if the sampling frequency is less than the Nyquist sampling frequency for the events of interest these will not be reflected accurately in the data. Similarly, digitization of the power signal introduces quantization error and lower resolution (number of bits per sample) samples will introduce greater quantization error reducing data quality. Second, the representation of the data, e.g., waveform versus phasor data. The superior quality of waveform data versus phasor data is well documented [18][20].
- Latency – While latency is not a factor in training AI/ML models, it can be a factor in the efficacy of the deployed models. The delay in the serving data from measurement to ingest by the deployed AI/ML model and the generation of subsequent results can be critical to operations. Each step in that process contributes to the overall latency of operations.
- Accurate time synchronization – The data used in training and deployment of AI/ML models for the distribution grid do not come from a single sensor, rather data from multiple sensors are evaluated by these models. Errors in the time synchronization of data from multiple sensors represent one of the most significant factors in data quality, since phase relationships are critical to the operation of the distribution grid. For this reason, most of the currently available instrumentation rely on GPS time-stamped data to provide microsecond timing accuracy.
- Absence of Personally Identifiable Information (PII) – Privacy is a concern in the application of AI/ML models with the potential for compromising PII. For example, AMI data, by definition, include PII data as it relates to customer billing information. Care should be taken that the data used in training and deployed models is free from PII.

This raises the question, how does one instrument the distribution grid to ensure the highest quality data for use in AI/ML algorithms to enhance and improve its operation?

# INSTRUMENTING THE DISTRIBUTION GRID

The idea of instrumenting the distribution grid is not a new one. AMI SmartMeters, microPMUs, WMUs, and other types of instrumentation have been deployed by utilities in an attempt to instrument the distribution grid, yet the amount and quality of the data provided by them is insufficient to enable the deployment of AI/ML algorithms at scale. The common challenge to these approaches is the lack of network connectivity that is necessary to backhaul data from the instrument to the centralized point of aggregation and analysis. This lack of network connectivity also drives design decisions for these devices that inherently restrict the amount and quality of data they can capture. Why build the capability into a device that has no way to report its data? One approach to overcoming this challenge is to leverage the adjacent broadband infrastructure that runs parallel to the distribution grid as described in IEEE PES TR-127 [31], T. Peck and S. Caruso [32], and D. Kopin, et.al.[33].

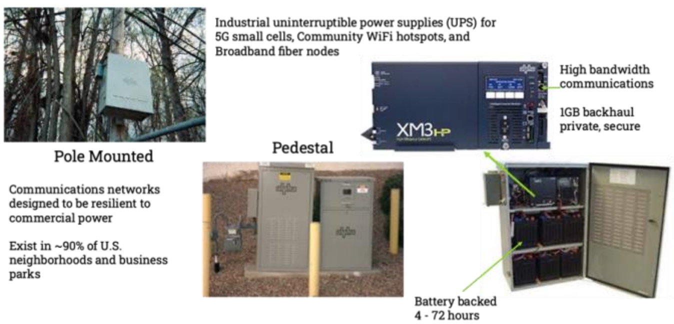


**Figure 2. Broadband Network Industrial UPS**

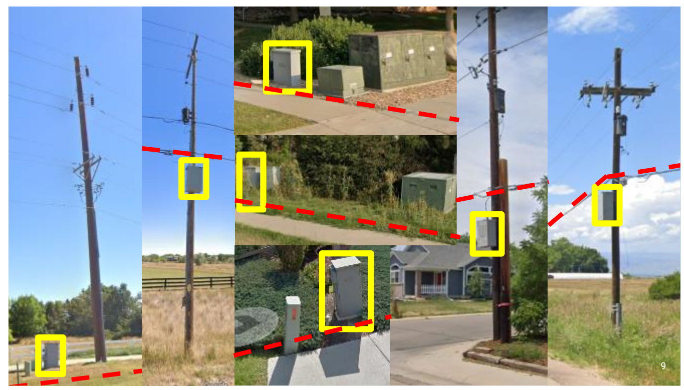

Figure 3. Broadband Network Parallels Distribution Grid

In Section 2.5 "Synchro-waveform Measurements from Adjacent Infrastructure", TR-127 explains how the existing broadband communication networks run parallel to the existing distribution grid and draws power from it. These broadband networks also deploy hundreds of thousands of industrial UPS that are monitoring the grid for power outages in order to provide backup power to the broadband network; see Figures 1 and 2. Consequently, broadband communication networks can be utilized for the purpose of instrumenting the distribution grid by placement of simple instrumentation in locations where the utility may not have access to high-speed communication. This removes the constraint of limited communication capacity to address the problem of instrumenting the distribution grid. To explore this idea, CableLabs® (see: https://www.cablelabs.com) initiated an innovation project in that led to the development of the GridVisibility Platform (GVP) that has been licensed and commercialized through Grid**Visibility**, inc.

# DEVELOPMENT OF THE GRID VISIBILITY PLATFORM (GVP)

The history of the development of the foundational intellectual property incorporated in the GVP began in 2017 CableLabs when established an innovation project to explore an idea sparked by the recognition that broadband operators often know when and where there is an outage in the distribution grid before the distribution grid operators themselves. This is a direct result of the ability for cable operators to monitor the state of the UPSs that power their broadband networks utilizing the cable modems that are embedded in them. Cable operators monitor this data via the Simple Network Management Protocol (SNMP) Management Information Base (MIBs) implemented in the UPS cable modems. In the following year, this type of data was aggregated from multiple broadband operators providing the first holistic, independent collection of voltage and outage data at the neighborhood level.

Beginning in 2019 CableLabs, in collaboration with the Department of Energy's (DOE) National Renewable Energy Laboratory's (NREL), explored advanced power distribution sensing and communications through cable broadband networks under NREL's Situational Awareness of Grid Anomalies

 

(SAGA) project. The output of this collaboration is documented in two NREL reports [34][35] evaluating use of aggregated broadband power data in grid cyber security applications.

As part of this collaboration, CableLabs began the development the a sensor package to measure the distribution grid power signal at high-resolution. An early outcome from the SAGA project was the realization that the broadband infrastructure's density and distribution is an ideal platform measuring the "last mile of the power grid". At the encouragement of NREL, the CableLabs innovation team developed a high-fidelity sensor that could be deployed into the UPS.

Another outcome of this collaboration was the development of the ANSI SCTE 271 standard [36] for sensing the power grid. The standard defined how the broadband industry could measure power signal with 12 bits of resolution, at 10,000 samples/second, and timestamped within 0.5 microseconds accuracy. The sensor implements this standard with the added capability of streaming via the Raw Data Transport Protocol (RDTP) enabling streaming Continuous Point-On-Wave (CPOW) data measurements.

In 2022 the first pilot test was deployed. After several iterations and rigorous testing, the technology platform received certification for pilot deployments into the broadband operator's production access network equipment. Early results demonstrate an immediate level of visibility previously impossible.

In 2023 the DOE Oak Ridge National Laboratory (ORNL) launched the Fault Location, Isolation, and Service Restoration (FLISR) project inclusive of the sensor deployments. This program supported the deployment of the sensors concentrated around Distributed Energy Resources (DERs) locations. The insights on grid behaviors in and around DERs identified the opportunity to support many applications beyond FLISR, including measuring power quality, frequency monitoring (islanding), voltage variations, and transients (flicker).

The realization that the need for high fidelity visibility by transmission grid operators to understand the impact of Inverter Based Resources (IBRs) being rapidly deployed in the distribution grid occurred in early 2024. This resulted in the deployment of more sensors in time to observe the impact of the total solar eclipse on the distribution grid in the region. The deployment of 17 sensors went from concept to deployment in six weeks: an unprecedented speed for utilities. The sensor data captured, in real-time, the impact of the solar eclipse including voltage sags, voltage regulators responding, and the operation of tap changers, shown in Figures 3 and 4.

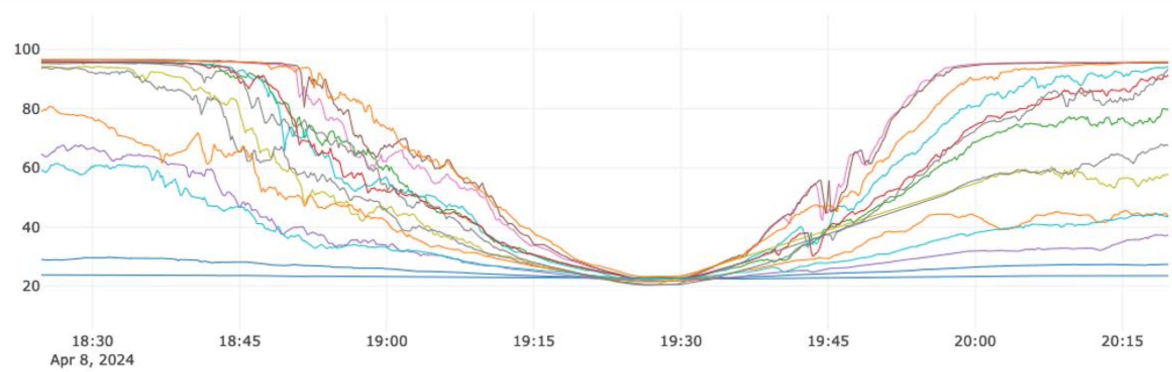


Figure 4. Solar Eclipse Light Sensor Data

 

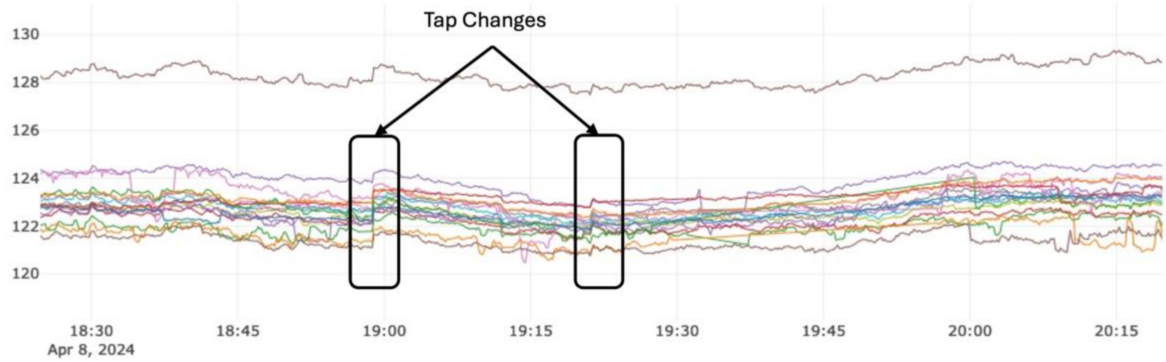


Figure 5. Solar Eclipse Voltage Sensor Data

Based on the successful deployment of these initial sensors the ORNL FLISR project deployed additional sensors in in the same geographic area in late 2024, bringing the total number of sensors in the area to 180. This deployment brought high-fidelity visibility to multiple substations and service territories. Table 1 describes the features and benefits of the current sensor and supporting software platform (GVP).

Table 1. GVP Features and Benefits

| Feature | Benefits |
|---|---|
| High fidelity: 12-bit samples at 10,000 samples per second | Continuous phase, magnitude, angle, & frequency with lossless aggregated data insights |
| GPS time synchronized to 0.5 microseconds, with reliable GPS lock | Holistic transmission & distribution GridVisibility with Continuous Point-on-Wave functionality |
| Resilient, continuous streaming data | Constant, state-of-the grid (24x7x365) |
| Battery-backed, resilient gigabit communications | Outage: 4 to 72-hour grid ride-through visibility & redundancy |
| Rapid time to data | Utility independent provisioning with no service interruptions |
| Grid visibility between substation and the meter | Secure data with no personally identifiable information (PII) |

# Conclusion

While the measurement data available to grid operators today from PMUs, DRUs, and WMUs (typically only available in snippets), is sufficient to provide training data for AI/ML fault models and can be used for post event analysis and identification, it is not sufficient to be useful in near real-time detection and mitigation. The limited availability of this data, both geographically and temporally, present real challenges to its near real-time application. To apply AI/ML models in near real-time would require placing these models in the measurement devices, increasing their cost, complexity, and configuration challenges. In addition, the detection of incipient events that are indicators of imminent failures or fault conditions are not typically detected or reported by the current measurement devices. This all points to the real need for accurate, high fidelity, time synchronized, secure CPOW data that is geographically dispersed for both training and near real-time fault identification and location. GVP provides this level of high fidelity CPOW data and can be rapidly deployed cost effectively.

# Detailed Literature Review

J. Wischkaemper, et. al., [2] describe the attributes of data quality necessary for waveform analytics. They identify several power system considerations for data quality, including: "1) other power system apparatus located between the event and the device (e.g. capacitors, power transformers, or even conductors), or 2) the power system sensors connected to the device itself (e.g. CTs, PTs, or alternative sensors)", sampling rate, bits of resolution, the implementation of the sensor, the dynamic range of the sensor, and duration of sensor recording.

X. Yu and Y. Xue [3] presents an overview of challenges for smart grids in the context of Cyber Physical Systems. It proposes that "Smart grids are electric networks that employ advanced monitoring, control, and communication technologies to deliver reliable and secure energy supply, enhance operation efficiency for generators and distributors, and provide flexible choices for prosumers." It identifies the importance of "ommunication Technologies, "Communication technologies are vital for efficient and effective interaction between the physical systems and the cyber systems. It is even more so for SGs as real-time distributed sensing and control (e.g., at the transient level) are critical for time-critical optimal performance. Two basic aspects of communication, namely, space and time, referring to the communication distance and time taken for transportation of information, should be considered when developing SG-tailored communication techniques at different levels, such as home area network, neighborhood area network, metropolitan area network, and wide area network. Key factors impacting real-time performance of SGs, especially in the transient layer, are time delays, packet errors and drops, and queuing delays. Some work has already been done in this direction, for example, […] an on-demand communication strategy was proposed to provide real-time tracking of dynamical systems and an embedded simulation environment created to synchronize with the dynamical system to inspect communication vulnerabilities. Given the trend of market-driven energy supply and demand in the future, competition and 'game playing' between various market participants may result in severe network congestions such as those occurred in India in 2012. The communication technologies as a whole need to be examined and improved upon in order for them to be used in real-time dynamic environments of SGs."

Brenner, et. al. [4] describe the Distribution Fault Anticipation (DFA) system developed at Texas A&M University. "DFA Devices continuously digitize current and voltage waveforms from current and potential transformers (5-amp circuit CTs and 120-volt bus PTs). Upon detection of anomalies, the record snapshots of the waveforms." They also note that, "DFA snapshots also are longer than would be typical for other technologies. More than a decade of DFA field research has shown that proper interpretation of certain events of interest requires analysis of these relative longer recordings."

H. Mohsenian-Rad, et. al. [5] identify challenges in processing data from µPMUs. "However, the main challenge is to go beyond manual methods based on the intuition and heuristics of human experts […]. Instead, it is crucial to develop the machine intelligence needed to automate and scale up the analytics on billions of µPMU measurements and terabytes of data on a daily basis and in real time. […] we make the case that big data analytics (BDA) is the key to addressing the challenges in working with µPMU measurements and so turn the data into actionable insights in a scalable fashion."

H. Akhavan-Hejazi and H. Mohsenian-Rad [6] identify the three Vs of big data, Volume, Variety, and Velocity and the barriers to adoption in power systems. The barriers they identify include:

- Discarded data - "the data in power systems should not be collected as need basis and the discarded data should be addressed."
- Siloed Data – "In power systems, siloed data poses as even a greater challenge."
- Real-time Analytics – "Utilities may need to upgrade their communication systems and to employ advanced network designs that support service differentiation, e.g., to distinguish delivering of critical protection-relay data from non-critical billing data. […] Finally, the network design must balance the overhead on the system with the speed needed for various signals."
- Coexistence of Centralized and Distributed Data Management – "supporting a coexistence and coordination among the existing centralized and the future distributed architectures is essential to enable BDA in power grids."
- Customized Data Management Systems to Cope with Fast Data – "The sampling rate of certain power sensor devices, such as PMUs, are so high, and the time window of some processes is so tight that the generic commercial database systems such as SQL or HDFS are not sufficient."

Bo Gao, et. al. [8] describe a method of detection subsynchronous resonance using time-stamped waveforms. "A new method for detecting a subsynchronous resonance at an early stage of the event has been proposed in this paper. With the use of time-stamped waveforms collected from both ends of a transmission line, the method can detect SSR current frequency, magnitude and damping in less one cycle of the SSR period."

M. Izadi and H. Mohsenian-Rad [9][12] argue that limitations on the data from distribution-level phasor measurement units (D-PMUs) favor “synchronized voltage and current waveform measurements to identify the location of events in power distribution systems.” and “The waveform measurements from WMUs are well-suited to study transient events in power distribution systems, in particular, when we compare them with the phasor measurements from distribution-level phasor measurement units (D-PMUs); a.k.a, micro-PMUs, which are another emerging class of sensors.”

W. Xu, et. al., [13] identify three industry trends driving the need for waveform data. “Firstly, the increased adoption of power electronic devices such as HVDC links and inverter-based resources has made it essential to add waveform monitoring capability at least for such devices since they work on waveforms. Secondly, modern power systems possess more complex dynamic responses such as inverter-related power oscillations and supersynchronous resonances. These phenomena can only be characterized and understood using waveform data. Thirdly, online condition monitoring of power apparatuses is gaining significant attention. The signs of emergent equipment failures are typically embedded in the waveforms. Therefore, waveform data are essential for the development of reliable condition monitoring tools.”

To highlight the value of power waveform measurements over phasor measurements H. Mohsenian-Rad and W. Xu [16] explicitly state, “This example and other similar examples raise the following questions: Why should we tie our hands with phasor representation of the voltage and current waveforms, which are ‘processed’ data? and Why limit our imagination to one complex number as opposed to looking at the ultimate raw data in the time domain?”

W. Xu, et. al. [18] describes interharmonic power as a method to locate the source of power system oscillations and limitations on phasor measurements. “In summary, the above two examples have shown that dynamic fundamental frequency phasors have limitations to capture true oscillation behaviors. In fact, these limitations have been recognized by dynamic phasor researchers. Their proposed solution is to include dynamic ‘harmonic’ phasors. Since a window length greater than T1 is used in these works, the ‘harmonic’ phasors are actually interharmonics. This development confirms that interharmonics are indeed needed to analyze power system oscillation behaviors.”

H. Mohsenian-Rad, et. al. [19] explore the benefits of synchro-waveforms and their potential to enhance wide-area monitoring in distribution systems and focused on IBR waveform dynamics. Among the benefits identified are:

- High Sampling Rates and Time Synchronization – “One of the key advantages of synchro-waveforms is their ability to provide raw waveform measurement samples at high sampling rates.”
- Continuous Streaming of Synchro-Waveforms – “the main challenge in event-triggered waveform capture is that there is no guarantee that all of the informative cycles of the synchro-waveforms are captured at each WMU. This is due to the challenges in properly setting up the event-triggering functions. […] Ultimately, the main advantage of event-triggered waveform capture is to cope with the issues regarding the limitations of local data storage and communication. In the future, these issues will likely be addressed through information and communications technology advancements.”

J. Wischkaemper, et. al. [20] [21] describe how automated analysis identified various types of incipient faults including animal contact, vegetation, capacitor arcing, cable fitting failure, and failed transformer bushing. They also identified several criterial for successful operation of an incipient fault detection system, including always-on (7x24x365 monitoring), near real-time access to waveform data, extended record periods, and automated analysis of events.

F. Ahmadi-Gorjayi and H. Mohsenian-Rad [22] present two methods of data driven model development, one in the frequency domain and two in the time domain (a finite impulse response model and an auto-regressive exogenous model), to create a library of models. These models were used to demonstrate an improvement in selecting the appropriate model from a test disturbance. These experiments were conducted using 63 data sets of disturbances collected over a six-month period, 42 were used for training purposes and 21 were used to test the accuracy of the models created.

K. Sarita, et. al., [23] discuss the use of ML algorithms on wind turbine vibration data to proactively warn of imminent faults. They present the use of Principal Component Analysis (PCA) an unsupervised machine learning technique which can be useful to detect the fault condition but will not classify the type of fault that occurred. They discuss how this ML technique can be useful for predicting faults and estimating uptime gained by maintaining the equipment upon the receipt of the first alarm or warning. The dataset for this work was extracted from a transmission line model that was simulated in MATLAB/SIMULINK.

P. Onu, et. al., [24] review the application of ML algorithms for Fault Detection and Diagnosis (FDD) in smart grids (those "integrating advanced technologies such as sensors, communication networks, and intelligent algorithms"). They summarize the research on three types of ML algorithms, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Decision Trees (DT) and identify the following advantages of ML algorithms for FDD in smart grids: improved fault detection accuracy, rapid fault diagnosis, predictive maintenance, and operational efficiency. Several of the papers cited made use of a dataset obtained from a distribution network in China. They also identify the following limitations of the application of AI/ML algorithms for FDD: "the need for large amounts of reliable training data, potential algorithmic biases, and the interpretability of complex models such as deep learning networks."

M. Chingshom, et. al., [25] used multiple ML algorithms, including K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and XGBoost algorithms for fault detection and classification. All the training data was derived from a MATLAB/SIMULINK of a 300-kilometer transmission line with the simulation of various fault types, e.g., single-phase-to-ground (L-G), two-phase-to-ground (LL-G), three-phase-to-ground (LLL-G), phase-to-phase (LL), and three-phase (LLL), to access grid performance. The KNN algorithm was identified as suitable for the purpose of fault detection and classification.

NERC [26] provides a guide to questions that should be asked regarding AI/ML to thoroughly understand what they are capable of and what changes are needed to implement them properly across a broad range of near real-time applications. This paper considers the human factors considerations necessary, as well as the implications AI/ML at the technological, organizational, implementation, human interaction, and legal level. Examples of AI/ML applications identified in system operations include load forecasting, solar and wind forecasting, contingency/stability analysis, outage management, report generation and procedure drafting, EMS and planning model validations, system operator training, and anomaly detection. It also, identifies cybersecurity issues related to risk management and AI/ML threat modeling, e.g., threats, assets, and vulnerabilities. It discusses the importance of data quality in the context of data collection, data validation, and data cleaning. In the context of data validation, several types of validation are identified including data type validation, range and constraint validation, code and cross-reference validation, structured validation, consistency validation, and relevancy validation. Within data cleaning it identifies the following key tasks handling missing data, removing duplicates and errors, and normalization and standardization. This paper's conclusions are optimistic yet cautious. It identifies the potential benefits of the application of AI/ML to real-time system operations but cautions that the "electric power sector has no tolerance for significant 'trial and error' learning and needs to avoid the 'initial bumpy road' observed when new technologies are brought into real time".

# APPENDIX A References


[1] "Intelligent Systems for Improved Reliability and Failure Diagnosis in Distribution Systems", B.D. Russell, C.L. Benner, IEEE Transactions On Smart Grid, June 2010, available online: https://ieeexplore.ieee.org/document/5446438.

[2] "Data Quality Considerations for Waveform Analytics", J. Wischkaemper, C.L. Benner, B.D. Russell, K. Manivannan, CIGRE US National Committee 2015 Grid of the Future Symposium, October 2015, available online: https://cigre-usnc.org/wp-content/uploads/2015/10/Wischkaemper.pdf.

[3] "Smart Grids: A Cyber–Physical Systems Perspective", X. Yu, Y. Xue, Proceedings of the IEEE, Vol. 104, No. 5, May 2016, available online: https://ieeexplore.ieee.org/document/7433937.

[4] "Application of DFA Technology for Improved Reliability and Operations", C. L. Benner, R. A. Peterson, B. D. Russell, 2017 IEEE Rural Electric Power Conference (REPC), June 2017, available online: https://ieeexplore.ieee.org/abstract/document/7967008.

[5] "Distribution Synchrophasors: Pairing Big Data with Analytics to Create Actionable Information", H. Mohsenian-Rad, E. Stewart, E. Cortez, IEEE Power & Energy Magazine, June 2018, available online: https://ieeexplore.ieee.org/document/8340896.

[6] "Power Systems Big Data Analytics: An Assessment of Paradigm Shift Barriers and Prospects," H. Akhavan-Hejazi and H. Mohsenian-Rad, Energy Reports, vol. 4, pp. 91-100, November 2018, available online: https://intra.ece.ucr.edu/~hamed/AHMRjEP2018.pdf.

[7] "SynchroWaveform Measurement Units and Applications", A. F. Bastos, S. Santoso, W. Freitas, W. Xu, Proceedings of the IEEE Power and Energy Society General Meeting (PESGM), July 2019, available online: https://ieeexplore.ieee.org/abstract/document/8973736.

[8] "Waveform-Based Method for Fast and Accurate Identification of Subsynchronous Resonance Events", B. Gao, R. Torquato, W. Xu, W. Freitas, IEEE Transactions on Power Systems, September 2019, available online: https://ieeexplore.ieee.org/abstract/document/8667317.

[9] "Event Location Identification in Distribution Networks Using Waveform Measurement Units", M. Izadi and H. Mohsenian-Rad, Proceedings of the IEEE PES Innovative Smart Grid Technologies Conference, October 2020, available online: https://intra.ece.ucr.edu/~hamed/IMRcISGT2020.pdf.

[10] "Event cause analysis in distribution networks using synchro waveform measurements," I. Niazazari, H. Livani, A. Ghasemkhani, Y. Liu, L. Yang, Proceedings of the IEEE North American Power Symposium (NAPS), April 2021, available online: https://ieeexplore.ieee.org/abstract/document/9449678.

[11] "Roadmap for Advanced Power System Measurements", J. Follum, H. Kirkham, A. Riepnieks, P. Etingov, L. Miller, X. Fan, E. Ellwein, Pacific Northwest National Laboratory (PNNL) Report, April 2021, available online: https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-31214.pdf.

[12] "Synchronous Waveform Measurements to Locate Transient Events and Incipient Faults in Power Distribution Networks," M. Izadi and H. Mohsenian-Rad, IEEE Transactions on Smart Grid, September 2021, available online: https://ieeexplore.ieee.org/document/9432388.

[13] "Synchronized Waveforms – A Frontier of Data-Based Power System and Apparatus Monitoring, Protection, and Control", W. Xu, Z. Huang, X. Xie and C. Li, IEEE Transactions on Power Delivery, February 2022, available online: https://ieeexplore.ieee.org/abstract/document/9403991.

[14] "Step Change Detection for Improved ROCOP Evaluation of Power System Waveforms", A. Karpilow, M. Paolone, A. Derviškadić, G. Frigo, Proceedings of International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA), May 2022, available online: https://ieeexplore.ieee.org/document/9806005.

[15] "1 ConvEDNet: A Convolutional Energy Disaggregation Network Using Continuous Point-On-Wave Measurements", A. Shirsat, H. Sun, K. J. Kim, J. Guo, D. Nikovski, In Proceedings of IEEE Power and Energy Society General Meeting (PESGM), July 2022, available online: https://ieeexplore.ieee.org/abstract/document/9916802.

[16] "Synchro-Waveforms: A Window to the Future of Power Systems Data Analytics", H. Mohsenian-Rad, W. Xu, IEEE Power and Energy Magazine, September 2023, available online: https://intra.ece.ucr.edu/~hamed/MRXjPEM2023.pdf.

[17] "Sub-cycle Event Detection and Characterization in Continuous Streaming of Synchro-waveforms: An Experiment Based on GridSweep Measurements", N. Ehsani, F. Ahmadi-Gorjayi, Z.-J. Ye, A. McEachern, H. Mohsenian-Rad, Proceedings of the IEEE North American Power Symposium, October 2023, available online: https://intra.ece.ucr.edu/~hamed/EAGYMMRcNAPS2023.pdf.

[18] "Interharmonic Power – A New Concept for Power System Oscillation Source Location", W. Xu, J. Yong, H. J. Marquez, C. Li, IEEE Transactions on Power Systems, January 2025, available online: https://ieeexplore.ieee.org/document/10857400.

[19] "Synchro-Waveforms in Wide-Area Monitoring, Control, and Protection", H. Mohsenian-Rad, M. Kezunovic, and F. Rahmatian, IEEE Power & Energy Magazine, February 2025, available online: https://magazines.ieee.org/pe/library/item/january_february_2025/4243445/.

[20] "Identification, Location, and Remediation of Incipient Fault and Failure Conditions Using Waveform Monitoring and Automated Analysis", J. Wischkaemper, B. D. Russell, C. L. Benner, K. Manivannan, 20th International Conference on Harmonics & Quality of Power (ICHQP), May 2022, available online: https://ieeexplore.ieee.org/document/9808824.

[21] "Online Automated System for Incipient Fault and Failure Detection of Distribution Apparatus Using Waveform Disturbances", J. Wischkaemper, C.L. Benner, B.D. Russell, K. Manivannan, 27th International Conference on Electricity Distribution, June 2023, available online: https://ieeexplore.ieee.org/document/10324360.

[22] Data-Driven Models for Sub-Cycle Dynamic Response of Inverter-Based Resources Using WMU Measurements", F. Ahmadi-Gorjayi, H. Mohsenian-Rad, IEEE Transactions on Smart Grid, September 2023, available online: https://intra.ece.ucr.edu/~hamed/AGMRjTSG2023.pdf.

[23] "Fault Detection of Smart Grid Equipment Using Machine Learning and Data Analytics", K. Sarita, S. Kumar, R. K. Saket, Advances in Smart Grid Automation and Industry 4.0, Select Proceedings of ICETSGAI4.0, April 2021, available online: https://www.researchgate.net/publication/351051756_Fault_Detection_of_Smart_Grid_Equipment_Using_Machine_Learning_and_Data_Analytics.

[24] "Machine Learning: A Comprehensive Exploration of Fault Detection and Diagnosis in Smart Grids", P. Onu, C. Mbohwa, A. Pradhan, III. International Conference on Electrical, Computer and Energy Technologies (ICECET 2023), 16-17 November 2023, available online: https://ieeexplore.ieee.org/abstract/document/10389596.

[25] "Fault Detection and Classification in Smart Grid Using Machine Learning Approach", M. Chingshom, B. Shakila, M. Prakash, 2024 International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS), December 5-6, 2024, available online: https://ieeexplore.ieee.org/document/10893627.

[26] "Artificial Intelligence and Machine Learning in Real-Time System Operations White Paper – Revision 1", North American Electric Reliability Corporation (NERC), November 2024, available online: https://www.nerc.com/pa/rrm/bpsa/Documents/Whitepaper-AI%20and%20ML%20in%20Real-Time%20System%20Operations.pdf.

[27] "Models for Machine Learning", M.T. Jones, December 2017, available online: https://developer.ibm.com/articles/cc-models-machine-learning/.

[28] "The Effects of Data Quality on ML-Model Performance", L. Budach, M. Feuerpfeil, N. Ihde, A. Nathansen, N. Noack, H. Patzlaff, H. Harmouch, F. Naumann, Proceedings of the VLDB Endowment, July 2022, available online: https://www.researchgate.net/publication/362386427_The_Effects_of_Data_Quality_on_ML-Model_Performance.

[29] "A survey of big data and machine learning", S. R. Salkuti, Q. Wu, G. Ding, Y. Xu , S. Feng, EURASIP Journal on Advances in Signal Processing, May 2016, available online: https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-016-0355-x.

[30] "Synchro-Waveform Measurements and Data Analytics in Power Systems", IEEE PES-TR127, December 2024, available online: https://resourcecenter.ieee-pes.org/publications/technical-reports/pes_tr_127_amps_1218724.

[31] “TCF-20-20213: Advanced Power Distribution Sensing and Communications through the Cable TV Broadband Network; Final Report. Period of Performance: June 2020 to June 2022”, M. Ingram, S. Caruso, R. Cruickshank, National Renewable Energy Laboratory (NREL), September 2022, available online: https://www.nrel.gov/publications.

[32] "Enabling the No-Touch HFC Power Network Through Next-Gen Instrumentation", T. Peck, S. Caruso, SCTE Cable-Tech Expo 2024, September 2024.

[33] “VELCO and VEC Pilot Project on Novel High-Resolution Point-On-Wave Monitoring for Transmission Event Analysis & Model Validation”, D. Kopin, C. Brunner, K. Thomas, A. Ospina, and R. Quint, 2025 Georgia Tech Fault Disturbance & Analysis Conference, May 2025, available online: https://www.ap-concepts.com/2025_FDA/modules/request.php?module=oc_proceedings&action=view.php&id=35&type=2&a=Accept.

[34] “Advanced Power Distribution Sensing and Communications through the Cable TV Broadband Network”, M. Ingram, S. Caruso, and R. Cruickshank, National Renewable Energy Laboratory (NREL), September 2022, available online: https://www.nrel.gov/publications.

[35] “Situational Awareness of Grid Anomalies (SAGA) for Visual Analytics-Near-Real-Time Cyber-Physical Resiliency Through Machine Learning”, M. Ingram, et.al., National Renewable Energy Laboratory (NREL), February 2023, available online: https://www.nrel.gov/publications.

[36] “Requirements for Power Sensing in Cable and Utility Networks”, ANSI/SCTE 271 2021, available online: https://account.scte.org/standards/library/catalog/.

## APPENDIX B Acknowledgements

The author gratefully acknowledges the support of DOE/ORNL, Comcast and VELCO in making this technology possible, as well as the review and feedback of <fill in names of reviewers>.