Toward an Ideology of Waveform Analytics: How to Automate Transient Point-on-Wave Analysis for Actionable Information

Jeffrey A. Wischkaemper, *Senior Member, IEEE*, B. Don Russell, *Life Fellow, IEEE*, Carl L. Benner, *Fellow, IEEE,* and Karthick Manivannan, *Senior Member, IEEE*

***Abstract*—The creation and analysis of large datasets has transformed multiple aspects of commerce and technology over the past decade. Applications of so-called "big data" range from sports teams seeking an edge against opponents to real-time language translation, and of course marketing companies trying to better target consumers with advertising. It is not surprising, then, that there are no shortage of products in the power system marketplace which claim to turn data into information which can transform day-to-day utility operations.**

One source of data commonly seen as "low hanging fruit" for improving power system operations comes from phasor measurement units (PMUs). PMUs typically produce a current and voltage phasor for each power system cycle along with a precise timestamp, allowing for comparisons of voltage angles across an entire power system. PMUs are broadly deployed across transmission nodes both in North America and Europe, and a substantial body of literature exists regarding their use and operation.

As applications developed for PMU data, practitioners began to recognize the need for higher speed, higher quality data than PMUs natively provide, particularly during transient conditions. This data, branded as "point-on-wave", is in effect traditional power quality transient data with precise timestamping. That said, while the data itself may look familiar, the scale at which it is generated requires new approaches that go beyond the practice of traditional power quality.

This paper explores many of the issues with leveraging so-called "point-on-wave" data to produce actionable intelligence and proposes a framework for developers seeking to capitalize on this emerging data source.

***Index Terms*— high impedance faults, downed conductors**

I. Introduction

IN overly simplistic terms, traditional power quality can be separated into concerns about harmonics on one hand, and concerns about the effects of power system transients on voltage (e.g. voltage sags, swells, rapid voltage change, etc.) on the other. This dichotomy is perhaps most clearly seen in the division of IEEE standards 1159 and 519. [1, 2] Monitoring power system transients through oscillography has a long history going back to at least the late 1930's. [3] Until the 1980's, the majority of oscillography collected was ad-hoc and used for research purposes, usually to analyze characteristics of fault behavior. [4]

Through the 1980's and 1990's, changing load profiles and customer needs led power system operators to recognize that *reliability* alone was not a sufficient metric to capture conditions which led to adverse impacts for customers, particularly as measured by standard indices. [5, 6] Reliability indices might capture when customers were without power for some length of time, but sustained interruptions were not the only conditions that created problems for customers. As the head of power quality for a large North American transmission operator once quipped, "It's entirely possible to deliver perfectly reliable, but crappy power." This is particularly true for customers with sensitive loads or those running industrial processes. As defined in IEEE 1366, the IEEE Guide for Electric Power Distribution Reliability Indices, a sustained interruption (colloquially understood as an "outage") does not begin until a customer has been without power for five minutes. As a result, if a utility can manually or automatically switch and reconfigure circuits to restore power to a customer in four minutes and fifty-nine seconds, that customer has technically not experienced an event which would contribute to traditional reliability indices like SAIDI, SAIFI, CAIDI, etc. Five minutes was chosen as a boundary primarily because it was believed that any manual (i.e., SCADA-enabled) switching to restore customers could likely be completed within that time window. IEEE 1159, the IEEE Recommended Practice for Monitoring Power Quality, uses one minute as its demarcation for a sustained interruption, based on the assumption that most automated reclose sequences should be complete within one minute. As of the writing of this paper, both IEEE 1159 and IEEE 1366 are under revision, and it is anticipated that IEEE 1159 will change to adopt a five-minute interval to align with IEEE 1366.

In any event, for sensitive industrial customers, a one-minute or four-minute interruption can have the same impact as a one-hour or four-hour interruption, in that it requires a full restart of processes and potentially the loss of whatever product was in production at the time the interruption occurred. For extremely sensitive customers (e.g., chip fabs), even a deep voltage sag can cause substantial problems, resulting in customers defensively designing their systems to mitigate the negative

All authors are with the Department of Electrical and Computer Engineering at Texas A&M University. Correspondence should be directed to jeffw@tamu.edu.

effects of power quality at the point of common coupling. [7] The emergence of large datacenters which are both sensitive to and a cause of power quality issues will likely create additional challenges over the coming decades. That said, the detection and analysis of power system transients will continue to be an important feature of power quality for the foreseeable future.

II. Typical Power Quality Transient Analysis

In many respects, the analysis of power quality transients by utilities – particularly in North America – is in part customer service and in part pre-emptive defense against litigation. Power quality engineers are frequently tasked with investigating customer complaints to determine whether the underlying cause of a power quality problem is on the utility system or within the customer facility. Often, this involves temporarily installing a power quality meter at a customer's location, recording data for some period of time, generally days or weeks, and then retrieving the device and data for analysis back in the office. Even when a utility installs permanent power quality monitors, the data is frequently analyzed only on exception, in response to a customer complaint, and the analysis is almost always performed manually by an engineer, usually looking at individual waveforms.

An arrangement like described above does have value in certain situations but, as one can imagine, also has obvious shortcomings. First and most importantly, too much of the process relies on humans – both the need for customers to complain and for engineers to analyze. Power quality problems only become "problems" if they are severe enough for a customer to notice and complain to a utility. Data is only collected in limited circumstances, in response to a known problem. Only the most severe disturbances are typically recorded, partially because power quality monitors traditionally have limited amounts of memory, but also because all data needs to be analyzed offline by a person. This results in relatively insensitive triggering thresholds. In practice, these limitations mean that many – perhaps most – power quality problems go unnoticed and undiagnosed.

While this type of approach can be effective to diagnose and remediate major problems that lead to customer complaints, it is not sufficient to proactively diagnose problems that have not yet risen to the notice of customers. Retroactive, ad hoc analysis of data only when a problem is detected through other means is quite different from the sort of online analysis required for advanced applications like incipient fault detection or condition-based maintenance.

III. Ideology for waveform analytics systems

Multiple groups in recent years have advocated continuous waveform recording, or gapless data, as a solution to the problems listed in the previous section. While continuous recording offers the illusion of "not missing anything," the reality is more complex. Proponents of continuous recording assert that one of the most complex aspects of recording power quality transients – triggering – is rendered unnecessary if you simply record all the data. But this is not quite true. The problem of triggering hasn't been *eliminated* by recording all the data, it has simply been deferred to a different time and location. While we would certainly grant that the opportunities to mine the data post hoc allow for more creative and speculative triggering, this is a different thing than saying that triggering is irrelevant for continuous waveform recordings. Even if one possesses all the needles in a haystack, one must know where in the haystack to look to find said needles. Claiming, "I could examine each individual piece of the combined hay and needlestack and determine which pieces are hay or needles," may be theoretically true, but is not convincing as a practical matter.

Additionally, we would note that continuous recording provides the illusion that "nothing was missed," but fails to consider events which occur below the noise floor of sensors or instrumentation. In other words, if an incipient signature occurs that falls below the noise floor of the measurement system – for example, the initial stages of a vegetation fault spanning two conductors, or a conductor resting on a wooden crossarm – one might be tempted to say that no electrical activity was present because it was not observed in recorded waveform signatures. In reality, however, the monitoring system was simply not sensitive enough to distinguish the electrical activity in question from electrical noise, which is not the same thing.

As many utilities push for IEC 61850 enabled digital substations with the availability of continuous sample value streaming, multiple IEEE and CIGRE working groups have formed to explore using continuous high-speed waveform samples for advanced applications like condition-based maintenance, asset management, incipient fault detection, and the like. It should be noted that while IEC 61850 may enable broad dissemination of sampled waveform values, data rates and resolution may or may not be sufficient to enable certain functionality. As an example, the most common rate for 61850 SV is 80 points per cycle, or 4,000/4,800 Hz (on 50/60 Hz systems, respectively). This sample rate may be sufficient for metering and protection systems, but is not sufficient to meet IEEE 519, EN 50160 or IEC 61000-4-7 requirements for harmonic measurements.

Many initiatives have focused on the mechanics necessary to enable solutions (for example, IEEE's Power System Relaying Committee group C55, or CIGRE working group B5.91). Application focused groups, like NASPI (formerly the North American SyncroPhasor Initiative, now rebranded as Novel Applications for Synchronized Power Instrumentation) have traditionally focused primarily on individual mathematical techniques or machine learning methods, rather than addressing system-level architecture concerns. While the development of standards-based approaches and competent analytics for waveform data is certainly a critical part of the problem, such development *must* be performed with system-level concerns in mind. To that end, we enumerate the following principles to serve as a guide for those hoping to develop practical analytics-based solutions:

A. Competent, automated data pipelines are essential for waveform analytics.

Data sources and data collection are critical components to any analytics system. The quality of output results is almost entirely dependent on the quality and reliability of available data. Real world data collection on operational circuits at scale is non-trivial. The level of effort to create systems which monitor, selectively capture, record, transmit, archive, and analyze data in a secure and automated manner should not be underestimated. The authors of this paper have maintained an online data collection platform for over 20 years which currently has over 3,000 circuit-years of exposure, and contains over 20 million distinct waveform transients, any of which can be graphed or analyzed on demand. The authors would freely admit that many mistakes were made / lessons were learned along the way, and that even today our system has room for improvement. At the same time, we are aware of very few systems which automate data production, collection, retrieval, and storage in a way that serves as a competent input to a hypothetical waveform analytics pipeline.

We also frequently consult with utilities who need to construct or reconstruct power system events from disparate data sources within their enterprise. Our experience helping utility companies analyze data obtained from third-party devices (relays, PQ monitors, line sensors, AMI and SCADA) suggests that it is not uncommon for data and metadata from major OEMs to contain multiple errors which substantially limit their value. These errors include but are not limited to issues like missing samples, inaccurate timestamps, and files supposedly originating from different devices at different times containing byte-wise identical data. Some of these errors can be blamed on poor data collection practices by customers or improper settings. Others are likely to be related to bugs (sometimes "features") in the hardware or firmware of the devices themselves. But the larger point is that standard business-as-usual practices at many utilities produce results that are unacceptable as inputs to a waveform analytics system.

Unfortunately, many discussions about "data quality" in the PQ space revert to the limited metrics of sample rate and/or bits of resolution. These are important factors, to be sure, but they are irrelevant if other aspects of the holistic data pipeline are ignored or overlooked. Factors like record length, triggering sensitivity, analog or digital anti-aliasing filters, and the quality of input sensors all contribute to the quality of data inputs for an analytics system.

An especially important caution regarding data quality involves combining waveform input from different types of devices into a common analytics framework. Many utility companies want to "use the data they already have" – which frequently means a mashup of digital fault recorders (DFRs), power quality monitors (PQMs), PMUs, relays, etc. Each of these types of devices were created for different purposes with different constraints and tradeoffs in terms of data quality. While there is nothing inherently wrong with any of these tradeoffs individually, combining all their outputs uncritically into a single unified data pipeline is a recipe for unexpected results.

Many years ago, the authors of this paper consulted with a utility experiencing unexpected behavior from a newly installed fault location, isolation, and service restoration (FLISR) system. It quickly became apparent that the utility was not only using different interrupting devices and controllers from multiple vendors, but also different types of low-energy sensing technology with different physical characteristics. The combination of different sensor technologies paired with different vendors made it impossible to debug the problem. While this seems like an extreme example, the nuances of inputs to waveform analytics systems are in many respects more sensitive and complicated than those for basic overcurrent protection.

Data quality issues are not limited to electrical waveform data. Information obtained from utility DMS/OMS systems – particularly outage cause information – should be viewed with suspicion, especially when the cause code is generic (e.g., "vegetation," "animal," "lightning"). Line crews are tasked with restoring power, and accurate record keeping is often given a lower priority. Attempts to accurately correlate waveform data to outage codes often require in-depth analysis from someone familiar with utility operations and apparatus, as key insights are buried in several layers of comments. Additionally, many utilities use multiple systems for tracking outages, with some information being available in one system but not the other.

R&D engineers should always adopt a hermeneutics of suspicion regarding both power system data and utility outage tickets. Unfortunately, our experience suggests that many analytics developers simply assume their data inputs are valid, leading to predictable garbage out results.

The overall quality of input data should be evaluated on multiple dimensions, including but not limited to relevance (the data can be used to make important decisions), trustworthiness (the data are consistent and don't contain unexplained discontinuities), and repeatability (data collection is fully automated and available for analysis when needed). If you do not have reliable, clean, consistent data to feed into your analytics framework, you will have a much harder time producing relevant, trustworthy and repeatable outputs. Unfortunately, many practitioners and utilities overlook this fact, wanting to use disparate waveform recorders with varying quality and sensing characteristics. Data collection at scale is not a trivial problem and should not be underestimated.

B. Any system that doesn't provide real-time, online information is of limited value.

Traditional power quality waveform monitoring often involves spot installation of equipment to record a set of data limited both in scope (at one customer) and time (generally a matter of days or weeks), for the purpose of gaining information about a specific problem that is either known or believed to exist. The equipment is then retrieved and data analyzed offline by engineers at the office to detect (generally) voltage related issues. This approach works fine for forensic PQ monitoring of load conditions and harmonics at a customer location but is wholly inappropriate for applications like incipient failure detection. Power system failures are dynamic, occurring in real time. At best, offline approaches allow you to retrospectively construct or determine what happened after the failure occurred. At worst, offline analysis gives you a false sense of what you *might* have done *if* you had analyzed the information in advance (which you *did not* and generally *cannot*).

Advanced waveform analytics applications require always-on, 24/7/365 data monitoring with near real-time access to data and alerts.

C. Think about scale at the beginning, not the end.

Many university or national lab research projects begin with a large pool of graduate students or staff scientists analyzing a relatively small amount of data – effectively a small problem scope and a huge labor scope. Frequently, little if any thought is given to how the approaches proposed in the research projects would work on a real-world system with hundreds or thousands of points. Or, if such thought is given to the matter, the assumption is that all users will have access to national lab or university-scale resources to run analytics. Solutions that appear successful on a handful of circuits often fail or encounter substantial challenges when scaled to real-world deployments.

Moving a proposed method from zero to ten circuits is usually a technology problem. In other words, this first stage is trying to answer the question, "Does my method work on a level of fundamental physics?" This is often the hardest stage, and the one where researchers spend much of their time. In truth, moving a research idea from the realm of working in a simulation to working on a single circuit in the real world is often the most difficult step. But even at this stage, lack of consideration for how the proposed method will scale can guarantee failure later in the process.

Assuming the underlying technology and physical principles are valid, moving from ten circuits to one hundred circuits is usually a platform problem. This stage often involves rearchitecting a research project and sanding off its rough edges to make it robust and presentable. It requires thinking like a user and not a developer. Questions that may pop up in this phase include, "Does my system have a steep learning curve? Does it require a lot of tinkering by trained personnel to keep it running? Is the user experience polished, or is it a command line experience with poorly documented, obscure option codes?"

The platform stage is also a time when breaking edge cases start to emerge as novel events are observed, sometimes requiring significant modifications to the core technology. Fundamental solutions which are brittle – for example, many machine learning approaches – can face profound problems at this stage as they try to incorporate new events while maintaining acceptable performance. Training and retraining costs and/or times can also balloon as the size of the dataset grows.

Finally, the platform stage also tends to identify software or hardware processes that scale as $O(n^2)$ (or worse, $O(c^n)$) – meaning that the complexity of your processes scale exponentially as your number of circuits scales linearly. This is particularly true for systems which rely on distributed sensors.

Moving from one hundred to one thousand circuits is usually an infrastructure problem. The question here is fundamentally, "Can I scale the hardware quickly enough in terms of compute, storage, and most importantly communications to keep the solution not only technically, but also economically feasible?" This is a particular concern for systems which use centralized architectures. Even with compression, high-speed waveform streams require megabits of bandwidth per circuit per second. This may be feasible within a single substation, but any attempt to stream all data for 1,000 circuits to a central location for processing and analysis will face serious technical challenges, and even more serious economic ones.

By contrast, moving from one thousand to ten thousand circuits is usually an administrative problem, and of the challenges listed in this section probably the easiest to solve. This typically involves organizational decisions about subdividing the system into pieces and ensuring that appropriate personnel have access to the appropriate data.

D. Successful systems require a positive return on investment (ROI).

As noted in the previous section, the cost of a system frequently scales with the number of circuits monitored. This is true not only of the initial hardware and install cost to deploy the system, but also ongoing expenses to keep it running. Solutions that require large amounts of storage, compute, or communication to function can quickly incur costs that will draw the attention of executives or regulators, who will demand justification for the ongoing expenditures. This is particularly acute for solutions that rely on centralized storage and processing to function.

Designers would be wise to focus on and minimize the *investment* portion of ROI as soon as possible. Ultimately the reason why AMI deployments succeeded where many other "smart grid" technologies failed was almost entirely due to the ROI business case. [8] Even if waveform analytics systems "work" they must provide business value to see widespread deployment.

E. The user doesn't know what a waveform is (…and they shouldn't have to!)

Any system that relies primarily on the skill and training of the end user to derive value is destined to fail. In the modern utility landscape, results from waveform analytics systems can and will be used by users with varying levels of training and education, ranging from executives to chief engineers to linemen in a bucket truck. Some users become very proficient investigating and analyzing underlying waveform data; others have the skill to analyze the data but accept automated reports at face value; still others lack the skill to meaningfully interpret the underlying data at all. A user should never have to look at a waveform if they don't want to (or know how to).

F. "Throw it into TensorFlow/PyTorch and see what comes out" is not a solution.

Big data and machine learning tools provide opportunities, but only when supplied with high quality datasets, guided by appropriate domain expertise and applied to the correct aspects of the problem. This is a particular concern in early research, where data is often derived from simulation or a small number of circuits (which may or may not be representative of a "typical" case) and the work is often performed by personnel with expertise in machine learning or signal processing, but very little knowledge of how power systems operate in the real world.

Even at utility scale, the occurrence of high-value failure mechanisms is uncommon. Accumulating the quantities of distinct failures and signatures with accurate and appropriate labeling to serve as a competent input dataset for machine learning is a herculean task requiring years of monitoring on hundreds or thousands of circuits, along with substantial amounts of labor by trained personnel. Because operational systems are sensitive to both false positives and false negatives, it is critical to know not only that your system made a particular decision, but also to understand why that decision was made. Similarly, refining the system for better performance on edge cases is a critical requirement for any operational system at scale. Machine learning has value in parts of the waveform analytics problem but is poorly suited for many critical aspects of an overall solution.

G. The confusion matrix is not a flat space.

Classification problems are often represented as a confusion matrix, where "correct" classifications are represented along the diagonal, and "incorrect" classifications exist along off-diagonals. The unstated implication of this arrangement is that all correct and incorrect classifications carry more or less the same weight. In other words, if my routine is supposed to identify cats and I am presented with a photo of a fox, it doesn't matter whether I call it a tiger or a mountain lion – neither is correct, and I lose all points. While this has a certain appeal from a purist perspective, it is not how the waveform analytics classification problem works. Some classifications (and misclassifications!) matter far more than others.

Consider the following scenario: imagine a hypothetical waveform analytics engine which classifies disturbances into categories. Now suppose that I give the system 10,000 motor start events from different motors, and assume for the sake of argument that the analytics engine classifies 9,999 of them correctly as motor start events. From an absolute classification perspective, my system is 99.99% accurate! It's hard to get better than that. But, for the sake of argument, assume that for the one misclassification, the system declared that a motor start was a downstream overcurrent fault. Now, from the perspective of the utility, it doesn't matter that the system was 99.99% accurate – the only "important" event that the system reported on it was wrong about!

Evaluating the overall performance of a waveform analytics system is challenging, and fraught with complications. Regression testing can be even more difficult.

IV. Conclusions

Data is not a panacea. It is often difficult to collect, challenging to clean, and complicated to interpret, even by professionals. The transition from bits to action is seldom straightforward.

This is especially true when one envisions the move from bespoke analysis of events by trained professionals toward an automated system that functions at scale.

This paper is not intended to solve the waveform analytics problem, but rather to highlight both challenges and essential requirements for any system that hopes to leverage waveform signatures to help utilities operate their systems more effectively.

V. References

[1] "IEEE Recommended Practice for Monitoring Electric Power Quality," *IEEE Std 1159-2019 (Revision of IEEE Std 1159-2009),* pp. 1-98, 2019, doi: 10.1109/IEEESTD.2019.8796486.

[2] "IEEE Standard for Harmonic Control in Electric Power Systems," *IEEE Std 519-2022 (Revision of IEEE Std 519-2014),* pp. 1-31, 2022, doi: 10.1109/IEEESTD.2022.9848440.

[3] C. L. Gilkeson, P. A. Jeanne, and J. C. Davenport, "Performance of Ground-Relayed Distribution Circuits During Faults to Ground," *Transactions of the American Institute of Electrical Engineers,* vol. 61, no. 1, pp. 40-48, 1942, doi: 10.1109/T-AIEE.1942.5058429.

[4] J. J. D. Burke, D A; Lawrence, D J, "Distribution fault-current analysis. Final report.," EPRI, 1983.

[5] "IEEE Guide for Electric Power Distribution Reliability Indices," *IEEE Std 1366-1998,* pp. 1-21, 1999, doi: 10.1109/IEEESTD.1999.89432.

[6] M. McGranaghan and B. Roettger, "Economic Evaluation of Power Quality," *IEEE Power Engineering Review,* vol. 22, no. 2, pp. 8-12, 2002, doi: 10.1109/MPER.2002.981339.

[7] "IEEE Recommended Practice for Voltage Sag and Short Interruption Ride-Through Testing for End-Use Electrical Equipment Rated Less than 1000 V," *IEEE Std 1668-2017 (Revision of IEEE Std 1668-2014),* pp. 1-85, 2017, doi: 10.1109/IEEESTD.2017.8120227.

[8] O. Malik *et al.*, "Smart Grid Origins, Definitions, Technologies, and Emerging Trends: A Power Community Perspective," *Journal of Modern Power Systems and Clean Energy,* pp. 1-8, 2025, doi: 10.35833/MPCE.2025.000807.

VI. Biographies

Jeffrey Wischkaemper (SM'21) is a Research Associate Professor at Texas A&M University, and a Principal Research Engineer at Texas A&M Engineering. His work focuses on the characterization of power system transients, particularly the detection and remediation of incipient failure conditions on power systems, including those that can lead to wildfire ignition.

B. Don Russell (F'91) is a Distinguished Professor and Regents Professor in the Department of Electrical and Computer Engineering at Texas A&M University. He is director of the Power System Automation Laboratory and for over four decades has conducted research in the design, operation, and protection of electric distribution circuits. Dr. Russell Is past president of the Power and Energy Society of the Institute of Electrical and Electronic Engineers and is a member of the National Academy of Engineering. Dr. Russell is a fellow of five technical societies.

Carl Benner (F'14) serves as Research Professor in the Department of Electrical and Computer Engineering at Texas A&M University in College Station, Texas. He holds BS and MS degrees from Texas A&M and has been a researcher at Texas A&M for more than three decades. He is a registered professional engineer in the state of Texas, a Fellow of the IEEE (Institute of Electrical and Electronics Engineers), and a member of CIGRE (International Council on Large Electric Systems).

Karthick Manivannan (SM'21) received his M.S degree and Ph.D. in Electrical Engineering from Texas A&M University in 2002 and 2012. Dr. Manivannan is a Research Associate Professor at Power System Automation Laboratory. His work focuses on developing algorithms for automatic classification and reporting of power system events. His research interests include application of signal processing and pattern recognition techniques, data analytics, intelligent reporting and efficient algorithm design.