A Quantitative Analysis of Wide-Area Phasor Measurement Unit Data

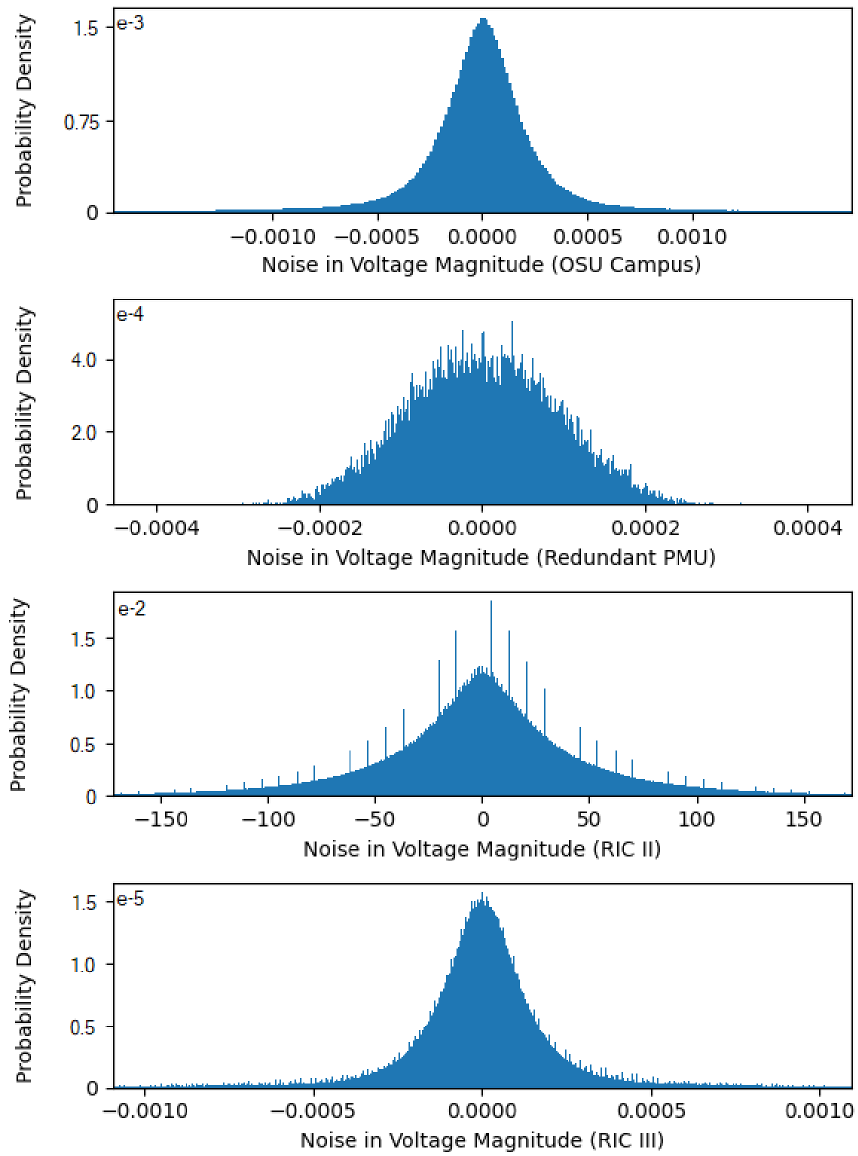
Travis Hagan[1], Eduardo Cotilla-Sanchez[1], Jinsub Kim[1], Shashini De Silva[1], and Dilan Senaratne[1]

[1]Oregon State University College of Engineering

September 16, 2022

Abstract

This paper contributes to the growing body of work that aims to characterize similarities and differences between synchrophasor data from real-power systems and those from synthetic power systems with emulated Phasor Measurement Units (PMUs). In particular, we survey previous works that characterize PMU noise and analyze the impacts on applications of these time-series data into machine learning algorithms in the power systems domain. We benchmark these methodologies with three datasets: data from an Oregon State University local PMU network, from two PMUs using the same set of sensors, and from multiple-utility interconnect-wide data. We found that it is important to consider each signal individually when synthesizing PMU data with noise, and that the noise needs to be adjusted by key statistical metrics.


Probability Density
e-3
1.5
0.75
0
−0.0010
−0.0005
0.0000
0.0005
0.0010
Noise in Voltage Magnitude (OSU Campus)
Probability Density
e-4
4.0
2.0
0
−0.0004
−0.0002
0.0000
0.0002
0.0004
Noise in Voltage Magnitude (Redundant PMU)
Probability Density
e-2
1.5
1.0
0.5
0
−150
−100
−50
0
50
100
150
Noise in Voltage Magnitude (RIC II)
Probability Density
e-5
1.5
1.0
0.5
0
−0.0010
−0.0005
0.0000
0.0005
0.0010
Noise in Voltage Magnitude (RIC III)

A Quantitative Analysis of Wide-Area Phasor Measurement Unit Data

Travis Hagan, Eduardo Cotilla-Sanchez, Jinsub Kim, Shashini De Silva, Dilan Senaratne
School of Electrical Engineering & Computer Science
Oregon State University
Corvallis, OR, United States of America
{hagantr, ecs, kimjinsu, desilvas, senaratg}@oregonstate.edu

***Abstract*—This paper contributes to the growing body of work that aims to characterize similarities and differences between synchrophasor data from real-power systems and those from synthetic power systems with emulated Phasor Measurement Units (PMUs). In particular, we survey previous works that characterize PMU noise and analyze the impacts on applications of these time-series data into machine learning algorithms in the power systems domain. We benchmark these methodologies with three datasets: data from an Oregon State University local PMU network, from two PMUs using the same set of sensors, and from multiple-utility interconnect-wide data. We found that it is important to consider each signal individually when synthesizing PMU data with noise, and that the noise needs to be adjusted by key statistical metrics.**

I. Introduction

Phasor Measurements Units (PMU) are a significant advancement in power systems technology. When combined with precise clocks, PMUs provide high-resolution insights into current operating conditions in the form of synchrophasors. However, the full potential of synchronized PMU data has yet to be realized. Large datasets exist but have few noted events with appropriate labels and are usually provided with no information about the underlying power system network and device characteristics. As such, the large datasets do not provide adequate detail for developing and tuning event detection, event classification, or other machine learning algorithms. In order to properly develop advanced detection algorithms that make use of frequency and phase angle differences, synthetic PMU data with known quantities is necessary.

This material is based upon work supported by the US Department of Energy under Award Number DE-OE0000780. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Accurate knowledge of the noise covariance matrix of PMU data is essential in many PMU data correction approaches [1], [2]. This information is frequently assumed to be known when tests are conducted with synthetic data. However, this is not always the case in real-world PMU data, and discrepancy between the actual and estimated noise covariance matrices can significantly impact the performance of such data correction approaches.

PMU noise impacts not only the time domain analysis, but also the frequency domain. Specifically, PMU noise results in a different variance in different frequency ranges in frequency domain measurements such as Short-time Fourier transforms (STFT). Thus, using an adequate range of noise characteristics in synthetic data when analyzing the frequency domain algorithm is important.

PMU measurements are widely used to estimate power system oscillation modes. Early work in the mode estimation algorithms assumed that the PMU measurements are noise-free and used a single PMU measurement stream. Thus different PMU measurements resulted in different modes due to the different noise characteristics. Although some techniques were introduced to effectively use multiple PMU measurements to estimate the mode [3], still the results can contain some additional modes estimates due to the effect of PMU noise. A typical technique is to fit the measurements to the higher-order system and remove the mode given by noise [4].

Work on synthetic PMU data generation has remained limited, with only a few research groups releasing published works on the subject. Furthermore, the papers do not generally agree on key pieces of the algorithms for synthetic PMU data generation or on what type of noise should be included. This paper compares previous works and adds new datasets to augment the analysis. The following sections will discuss sources of noise in PMU measurements, types and magnitudes of noise included in synthetic PMU data, and will quantify the type and magnitude of noise found in real world datasets.

A. Phasor Measurement Units

When generating synthetic PMU data it is important to consider all potential sources of error and noise. PMUs are analog-to-digital measuring devices that rely on an external array of equipment to produce phasor measurements. These external devices include Current Transformers (CT), Potential Transformers (PT), and the related wiring that convert current and potential to safe levels for input to the PMUs. Each external device may introduce noise and measurement error in addition to the PMU itself [5]. CTs and PTs are generally installed at the same time as the substation equipment, and can simultaneously be used as inputs to several measuring, monitoring, and protective devices.

To understand how PMU quantities vary from synthetic measurements it is necessary to understand how a phasor is constructed. Ideal voltages and currents are sinusoidal signals fitting the form $x(t) = x_m * cos(\omega t + \theta)$ where x_m is the peak amplitude, ω is the frequency in radians per second, and θ is the phase angle in radians [6]. It is convenient to plot voltage and current as phasors, or vectors, where the magnitude is the vector length, and phase angle is the angle referenced from the x axis. PMUs measure the voltage and current directly, but must compute frequency and power, generally following the latest standard specifications available [7].

B. Noise

Sources of noise in PMU measurements lie internal and external to the PMU itself, and thus a formal noise characterization is not as simple as examining each type of PMU to determine what noise may be present. It is therefore necessary to consider the system as a whole instead of each component individually. There is little agreement in the literature about how much noise should be added to synthetic data to be able to benchmark and validate real-time and engineering applications. In most papers surveyed there was no detailed reasoning for choosing type and amplitude of noise.

A few trends are noticeable when surveying research papers using synthetic PMU data. First, the type of assumed noise is generally Gaussian white noise. Second, the variance used is usually less than 0.05 per unit (pu). In [8], additive Gaussian white noise is used with variance of 0.005 to 0.01 pu for quantifying uncertainty in PMUs for voltage stability assessment. Huang et al. similarly used Gaussian white noise but with a slightly higher variance of 0.15 pu [9]. Zhou discusses PMU noise while testing systems for detecting oscillations [10]. Zhou uses Gaussian white noise but with a fixed variance of 0.03 pu. Li et al. add Gaussian noise varying between 0 and 0.03 pu [11]. Gaussian white noise was also used with variance of 1e-6 to 0.01 pu for a state estimation application in [12]. Tripathy et al. used Gaussian noise, but with considerably lower variance of 0.0001 pu [13]. This noise level was found acceptable for estimating generator rotor angle, with a justification based on the IEEE standard C37.118 [7], which defines performance requirements of PMUs.

Random noise with variance ranging from 0.01 to 0.02 pu was used when testing a power system estimator [14]. Xie et al. assume white noise but with a Signal-to-Noise Ratio (SNR) of 92 dB when applying dimensionality reduction to PMU data [15]. Tate and Overbye state that impulse noise has been observed in real PMU data while also implying that Gaussian noise is also present [16]. In Overbye's other works on synthetic PMU data generation, noise is considered for the PMU and the rest of system separately [17]. White noise was introduced directly to the measurements with 80 dB SNR.

Shi et al. [18] reiterates that the PMU performance requirement is 1% total vector error (TVE) however many do not conform due to measurement noise and errors from instruments. This paper stands out not because it uses white noise with 0.002 pu variance, but instead due to the focus on adding bias and treating voltage and current differently. The focus of the paper was detection and correction of errors in PMU measurements. The 1% TVE PMU performance benchmark is tested in [19]. A PMU was tested and found to comply with most of the static requirements but passed fewer of the dynamic. The authors note the changes in amplitude, phase angle and frequency over an interval as the main contributor for PMUs not meeting the requirements, but do not discuss bias or other potential error sources.

Brown et al. was one of the first papers to experimentally characterize and quantify noise present in real PMU data [20]. The authors find that real-world data fit a Gaussian noise distribution with a SNR of 45 dB for frequency, voltage, and current. This was consistent for the three datasets analyzed. An important consideration in this work is how the noise was separated from the signal. The authors applied Tate and Overbye's method using a median filter on PMU data [16]. To use this approach, the data must be relatively free of events.

Frigo et al. followed up with a similar work to Brown et al. which also sought to determine noise distribution, but through different methodologies [21]. The authors used a methodology that did not require separation of static from dynamic PMU data before extracting noise from the signal, concluding that the noise for their dataset fits additive Gaussian white-noise with SNR close to 45 dB.

In summary, the relatively few published works that have analyzed real-world PMU data have agreed that noise is generally distributed as additive Gaussian white-noise with a SNR close to 45 dB. However, only a handful of datasets have been analyzed, of which many use a few hours of data for only a few PMUs. The following sections will expand on the current literature by analyzing a more comprehensive real-world PMU dataset.

II. Methodology

Two methodologies are used to analyze the noise in PMU data. The first is described in Brown et al. [20]. The process consists in taking a PMU signal, such as voltage magnitude, current magnitude or frequency, remove any signals with events, then extract the noise from the remaining signals, and finally compute noise characteristics. The following subsections will detail the process and parameters used to perform the computations.

First, power system events must be removed from the signal, S, with length n such that baseline noise is being analyzed and not power system variances. To identify power system events a variance based event detector is used. Sample variance of the input signal is computed over a sliding window of fixed length. Any variance above the threshold is considered to be a power system event and is then removed from the data to be used for noise analysis. In (1), $v(i)$ depicts computed variance from the PMU signal S received at time i, and with $2M$ being the sliding window size, see Table II. Several thresholds were tested and it was found that a wide range of values produced similar results. Ultimately, the mean variance was implemented as the threshold which generally resulted in less than 0.01% of the data being considered as event data and thus being removed.

$$v(i) = \frac{1}{2M} \sum_{j=i+M}^{n-M} (S(j) - \frac{1}{2M} \sum_{k=j-M}^{j+M} S(k))^2 \tag{1}$$

To separate the noise from the signal, we follow Tate and Overbye's median filter [16]. This filter has been used both for generating PMU data and for PMU data analysis. The median filter (2) removes the noise data without removing features by taking the median value inside of a sliding window with size $2N$, see Table II. Subtracting the filtered data S_f from the original signal S, leaves one with only the noise data.

$$S_f(i) = \text{median}[S(i-N), S(i-N+1), ..., S(i), ..., S(i+N-1), S(i+N)] \quad (2)$$

The second methodology consists of analyzing redundant PMUs to get the most accurate representation in practice, since there are some utilities that install PMUs devices in pairs for redundancy. In this experiment, two PMUs are connected through the same set of PT and CTs and thus should report from exactly the same measurements. However in practice the recorded measurements will vary by some amount regardless of whether the same or similar devices are used.

$$noise_1 = sig_1 - (sig_1 + sig_2)/2 \quad (3)$$

This methodology to analyze the data is relatively more straightforward. First, the two signals are normalized. Then the mean value of the two signals at the same point are subtracted from each of the signals, leaving only the noise and any difference in the PMUs as shown in equation 3. The redundant PMU data is also tested with the first methodology described before.

A. Datasets

The datasets represent a wide range of available PMU data. This includes two datasets from research devices [22], and three datasets from utility collected sources [23].

1) PMU Data from Research Devices: First, we use a portion of the Oregon State University (OSU) Corvallis campus dataset, composed by six research-grade PMUs [22]. The primary advantage of the campus data is that all necessary characteristics are known about the PMU location, manufacturer, and use. All the PMUs are Schweitzer Engineering Laboratories Inc. (SEL) relays. Three PMUs (SEL-351A, SEL-487E, SEL-751) are installed on campus, and three PMUs ((2) SEL-751, SEL-351A) are off campus.

The downside to placing PMUs in the University environment without a utility goal plan is that they are placed very close to noise sources. When placing a PMU at a university few options exist since installing the PMU requires a GPS clock, PT and four CTs to be a fully functional system. The wiring must also be de-energized and disconnected so the CTs can be installed.

The OSU campus system has some optimally placed and some poorly placed PMUs, but this is good for a smart grid research testbed. An optimally placed PMU is installed at a 115 kV utility substation located about 30 miles from campus. A PMU is installed near a set of pump, that may contribute additional noise. The PMUs in the University's WESRF (Wallace Energy Systems & Renewables Facility) lab are connected on the downstream side of a dedicated lab power transformer, and use only one set of PTs and CTs for all three of the PMUs. One of these PMUs a SEL-351A, is contained in the dataset the OSU Campus dataset. The other two comprise the redundant PMU dataset.

The second dataset uses redundant PMUs connected to the PTs and CTs in the lab [22]. By using two PMUs, the average of the two signals will be a much closer approximation to the true signal. A better test would be to have two PMUs on the same power source, only using different PTs and CTs, but due to practical reasons this is very difficult. The redundant PMU data uses two SEL-751 relays. The two relays were chosen such that they were as similar as possible. One is a newer unit with a touchscreen display. The other is an older unit without the touchscreen display. The relays were configured with identical settings. The data rate was set to 60 samples per second. 24 hours of data was used for analysis.

2) PMU Data from Utility Sources: The Real Interconnect (RIC) dataset was supplied from the US Department of Energy (DOE), was anonymized for this project, and spans the United States' three interconnections [23]. There are approximately 22 terabytes of data in the compressed form and it contains nearly 2 years of synchrophasor measurements. The data is collected from numerous independent system operators and was combined into three aggregated sets. No other information is known or given, and was previously anonymized in terms of utility owners, their locations, configurations, or manufacturers of PMUs. There are

TABLE I

Dataset	Signal	PMUs	Data Rate	M	N	Mean	STD	SNR	Unique (%)	SW	KS
OSU Campus	F	6	60	15	15	-9.77e-08	2.97e-05	45.3	0.232	0.0	0.0
	V_M	6	60	15	15	-5.64e-06	5.87e-04	32.3	22.5	0.0	1.23e-04
Redundant_1	V_M	2	60	30	30	-4.14e-06	2.71e-04	35.7	13.6	0.0	0.063
Redundant_2	V_M	2	60	-	-	4.23e-16	9.08e-05	40.4	98.7	2.35e-04	0.743
RIC I	F	9	30	15	30	3.24e-05	3.65e-03	24.4	0.22	0.0	0.0
RIC II	F	38	30, 60	15	30	1.02e-08	9.35e-06	50.3	0.016	0.0	7.72e-11
	V_M	38	30, 60	15	30	-1.66e-06	4.54e-04	33.4	0.316	0.0	0.311
	I_M	13	30	15	30	2.80e-06	6.47e-03	21.9	16.2	0.0	0.205
RIC III	F	183	30	15	30	-2.88e-08	1.44e-05	48.4	0.19	0.0	0.0
	V_M	92	30	15	30	-8.12e-06	3.65e-04	34.4	2.11	0.0	3.02e-03
	I_M	89	30	15	30	3.25e-04	5.32e-02	12.8	15.3	0.0	4.97e-09

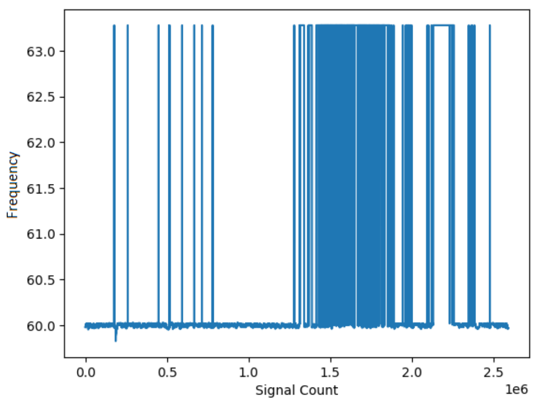


Fig. 1. Frequency graph for one PMU from RIC I dataset

many differences in quality of data, kV, and other attributes. This dataset provides a nearly complete spectrum of PMUs in use today. Taking into consideration what would be practical to process at a time for a real-time or engineering machine-learning-based application, one day worth of data was chosen from each dataset to analyze.

III. Results and Analysis

The noise extracted from all datasets is quantified through various methods, resulting in the summary Table II. The *mean* is computed to find any negative or positive bias. *Standard deviation* and Signal-to-Noise Ratio (*SNR*) are also used to characterize the noise data in range and strength. The Kolmogorov-Smirnov (*KS*) test and Shapiro-Wilk (*SW*) test are each used to determine if the histogram data fits a particular probability density function.

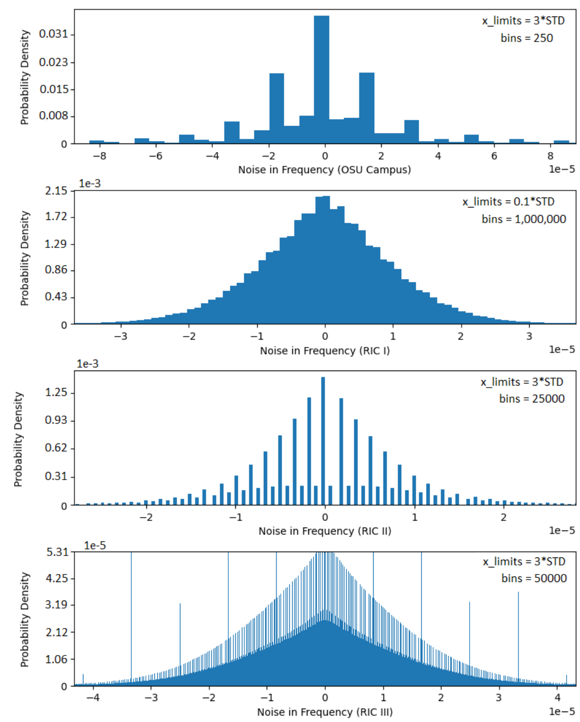


Fig. 2. Noise distributions for frequency data

The number of *unique* values is computed and reported as a percentage of the total values in table. This was added in response to some signals having just a few hundred different values in five

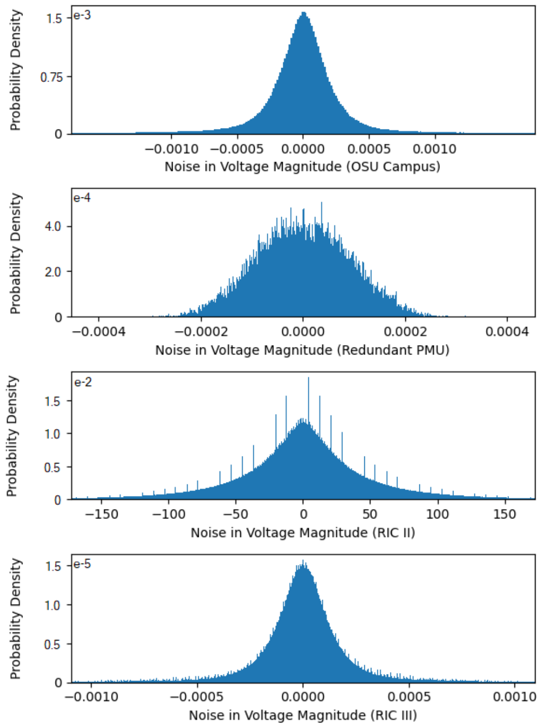


Fig. 3. Noise distributions for voltage magnitude data

TABLE II

β for Select Datasets		
Dataset	Signal	β
Redundant$_1$	V_M	1.12
Redundant$_2$	V_M	2.57
RIC II	V_M	0.83
RIC II	I_M	0.54

million samples as shown in Figure 5. This figure was generated with 1000 bins in the histogram for around five million samples. One-third of the bins remained empty. Plotting the full set of histograms presented mixed results, with some showing a well defined distribution, while other showed additional features in what would otherwise be a generalized Gaussian distribution.

The *unique* column shows the noise of current magnitude data has the highest range of unique values, followed by the voltage magnitude and finally frequency data. This is expected because frequency and voltage data is normalized and the mean value is always expected to be near one, per unit. This provides an advantage to the filters, whereas current magnitude data varies greatly over a day. Additionally, frequency data is computed as the first derivative of the synchrophasor phase angle according to the IEEE C37.118 standard for synchrophasor measurements [7], which reduces the number of unique values.

Mean noise values were least significant for frequency data and more significant for voltage magnitude and current magnitude. This is due to the

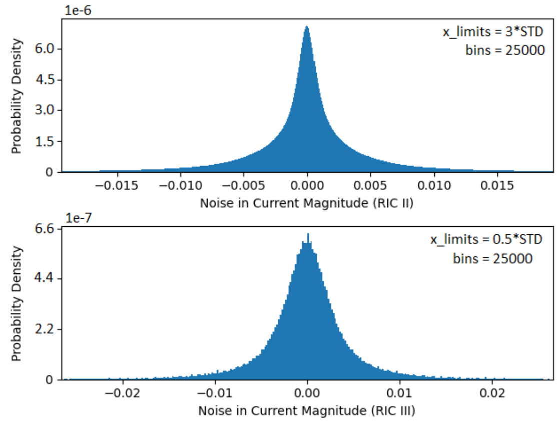


Fig. 4. Noise distributions for current magnitude data

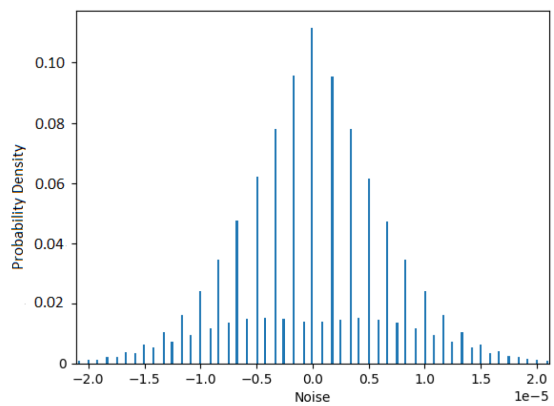


Fig. 5. RIC II frequency dataset has a finite number space for some PMUs

frequency being a quantity computed by the PMU and because the expected value is always close to the nominal frequency. Following a very similar pattern as the rest of the data, standard deviation was least for frequency and greatest for current magnitude. The outlier in this set is RIC I frequency standard deviation, which is one to two orders of magnitude greater, but given that only 9 PMUs of the 170 total had viable data, there is likely more issues in this dataset. The SNR for RIC I indicates this as well.

The data analyzed through the second methodology (*Redundant*$_2$), in which the signals are directly compared to determine noise instead of by filtering, shows some substantial differences to the noise extracted through filtering. First, the mean is less than half. Second, almost all the values are unique with 98.7% compared to 22.5% being the highest for the filtered noise.

A Shapiro-Wilk (*SW*) test and Kolmogorov-Smirnov (*KS*) test were performed in 100 batches of 5000 random samples. The choice in random number generator substantially influenced values computed in subsequent computations [24]. The mean value is reported in table II. The minimum and maximum p values were compared to the mean to look for substantial differences. The *SW* test compares the samples to determine if they come from a Gaussian distribution. The *KS* test is used similarly, but to test if the distribution fits a generalized normal distribution. The p values are reported in the table and for both tests, 0.05 is considered the threshold for if the samples come from the distribution. There was some further variation caused by the sampling. By using seed values for the random number generator, the p value increased in all tests.

Redundant$_2$ dataset has very high values for *KS* test. The minimum reported value for both PMUs data was 0.122, whereas the maximum 0.999. The mean value of 0.743. Comparing this to the *Redundant*$_1$ dataset that had a mean value of 0.064, minimum of 7.01e-10 and maximum value of 0.408 the different methodologies produce significant difference in distribution fitting. However, in both cases the test statistic is high enough that it is likely both come from a generalized normal distribution. The same cannot be said about the data fitting a standard Gaussian distribution as demonstrated in the *SW* column in Table II. A few of signals had non-zero p values but were in the range of 10^{-25} to 10^{-100} and thus were considered zero.

Considering the rest of the datasets, none of the frequency distributions had a p value for the *SW* and *KS* that was significant enough for the data to be from a Gaussian or generalized Gaussian. The voltage magnitude data only fit the generalized Gaussian in three of the five signals. The current magnitude only one of the two.

Table II reports the mean β values for the computed generalized Gaussian distribution in which the data fit the distribution. Note that *Redundant*$_2$, has the highest p value for the *KS* test along with the highest β value. β being over two indicates the distribution is wider in the center than a normal distribution. The rest of datasets have β much less than two indicating the distribution has a much sharper peak. This is evident in the figure 3. The redundant dataset has a much rounder central shape and heavier tails than the rest of the datasets indicating the Kurtosis of the distribution to be Platykurtic, while the rest would be Leptokurtic Kurtosis with sharper peaks and thinner tails.

The frequency distributions, figure 2, do not fit a normalized Gaussian. The hyper-geometric patterns shown in three of the four graphs is one reason the distribution does not fit well. The shape appears linear extending out from the peak. No distributions were found to fit the data well.

IV. Conclusion

This paper contributes to the broader analysis of PMU data in smart grids by quantifying signal and noise on real-world datasets spanning full power system interconnects; significantly expanding the amount of data analyzed than in previous works. The wide-area nature of the real interconnect data was used with few modifications and sought to capture the spectrum of installed devices. The results suggest that when generating synthetic PMU data it is necessary to consider each signal individually and to consider the means by which the PMU is reporting the values. Values measured directly by the PMU, (e.g.: current and voltage magnitude), as opposed to internally calculated have lower SNR, higher standard deviation, and higher noise mean. The extracted noise fit a generalized Gaussian distribution in about half of the signals sampled.

REFERENCES

[1] Shubhendra Vikram Singh Chauhan and Grace Xingxin Gao, "Synchrophasor data under gps spoofing: Attack detection and mitigation using residuals," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3415–3424, 2021.

[2] L. Vanfretti, J. H. Chow, S. Sarawgi, and B. Fardanesh, "A phasor-data-based state estimator incorporating phase bias correction," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 111–119, 2011.

[3] Daniel J Trudnowski, JM Johnson, and John F Hauer, "Making prony analysis more accurate using multiple signals," *IEEE Transactions on power systems*, vol. 14, no. 1, pp. 226–231, 1999.

[4] Guoping Liu, Jaime Quintero, and Vaithianathan Mani Venkatasubramanian, "Oscillation monitoring system based on wide area synchrophasors in power systems," in *2007 iREP symposium-bulk power system dynamics and control-VII. Revitalizing Operational Reliability*. IEEE, 2007, pp. 1–13.

[5] A.P. Sakis Meliopoulos, G.J. Cokkinides, F. Galvan, and B. Fardanesh, "GPS-Synchronized Data Acquisition: Technology Assessment and Research Issues," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, Kauia, HI, USA, 2006, pp. 244c–244c, IEEE.

[6] Arun G. Phadke and James S. Thorp, *Synchronized Phasor Measurements and Their Applications*, Power Electronics and Power Systems. Springer International Publishing AG, Springer International Publishing, Cham, 2017.

[7] IEEE, *IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005)*, IEEE, 2011.

[8] Chen Chen, Jianhui Wang, Zhengshuo Li, Hongbin Sun, and Zhaoyu Wang, "PMU Uncertainty Quantification in Voltage Stability Analysis," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 2196–2197, July 2015.

[9] Qingqing Huang, Leilai Shao, and Na Li, "Dynamic Detection of Transmission Line Outages Using Hidden Markov Models," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2026–2033, May 2016.

[10] Ning Zhou, "A Cross-Coherence Method for Detecting Oscillations," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 623–631, Jan. 2016.

[11] Wen-Tai Li, Chao-Kai Wen, Jung-Chieh Chen, Kai-Kit Wong, Jen-Hao Teng, and Chau Yuen, "Location Identification of Power Line Outages Using PMU Measurements with Bad Data," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3624–3635, Sept. 2016, arXiv: 1502.05789.

[12] Esmaeil Ghahremani and Innocent Kamwa, "Local and Wide-Area PMU-Based Decentralized Dynamic State Estimation in Multi-Machine Power Systems," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 547–562, Jan. 2016.

[13] Praveen Tripathy, S C Srivastava, and S N Singh, "A Divide-by-Difference-Filter Based Algorithm for Estimation of Generator Rotor Angle Utilizing Synchrophasor Measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 6, pp. 1562–1570, June 2010.

[14] Jinghe Zhang, Greg Welch, Gary Bishop, and Zhenyu Huang, "A Two-Stage Kalman Filter Approach for Robust and Real-Time Power System State Estimation," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 629–636, Apr. 2014.

[15] Le Xie, Yang Chen, and P. R. Kumar, "Dimensionality Reduction of Synchrophasor Data for Early Event Detection: Linearized Analysis," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[16] Joseph Euzebe Tate and Thomas J. Overbye, "Extracting steady state values from phasor measurement unit data using FIR and median filters," in *2009 IEEE/PES Power Systems Conference and Exposition*, Seattle, WA, USA, Mar. 2009, pp. 1–8, IEEE.

[17] Ikponmwosa Idehen and Thomas J Overbye, "Large-Scale Generation and Validation of Synthetic PMU Data," *IEEE TRANSACTIONS ON SMART GRID*, vol. 11, no. 5, pp. 9, 2020.

[18] Di Shi, Daniel J. Tylavsky, and Naim Logic, "An Adaptive Method for Detection and Correction of Errors in PMU Measurements," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1575–1583, Dec. 2012.

[19] Dinesh Rangana Gurusinghe, Athula D. Rajapakse, and Krish Narendra, "Evaluation of steady-state and dynamic performance of a synchronized phasor measurement unit," in *2012 IEEE Electrical Power and Energy Conference*, London, ON, Canada, Oct. 2012, pp. 57–62, IEEE.

[20] Michael Brown, Milan Biswal, Sukumar Brahma, Satish J Ranade, and Huiping Cao, "Characterizing and quantifying noise in PMU data," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, USA, July 2016, pp. 1–5, IEEE.

[21] G. Frigo, A. Derviskadic, A. Bach, and M. Paolone, "Statistical Model of Measurement Noise in Real-World PMU-based Acquisitions," in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*, College Station, TX, USA, May 2019, pp. 1–8, IEEE.

[22] Oregon State University, "OSU campus PMU dataset," 2022, Accessed: 2022-08-15.

[23] US Department of Energy, "Big data analysis of synchrophasor data," 2019, Accessed: 2022-08-10.

[24] Timothy H. Click, Aibing Liu, and George A. Kaminski, "Quality of random number generators significantly affects results of Monte Carlo simulations for organic and biological systems," *Journal of Computational Chemistry*, vol. 32, no. 3, pp. 513–524, Feb. 2011.