Detecting Bad Data and Managing Synchrophasor Data Quality

Panel – Event Characterization Using Synchrophasor Big Data
IEEE PES GM 2022

Yi Hu
Quanta Technology

We Need High Quality Data

The very foundation for getting credible data analytics results

- Low quality data either is **not usable** or **require substantial time and effort to preprocess** the data
- Preprocessing won't fix all data quality issues!

A MUST for real-time applications in control rooms and protection & control (P&C) systems

- Bad data ➔ Wrong decisions/ actions ➔ major adverse consequences
- Time constraints may limit what type of and how much preprocessing could be done in real-time

Project and dataset information

BDSMART (Big Data Synchrophasor Monitoring and Analytics for Resiliency Tracking) Project

- A US Department of Energy (DOE) co-funded project of FOA 1861 program “Big Data Analysis of Synchrophasor Data”

DOE provided field recorded synchrophasor data through Pacific Northwest National Lab (PNNL) to all 1861 projects

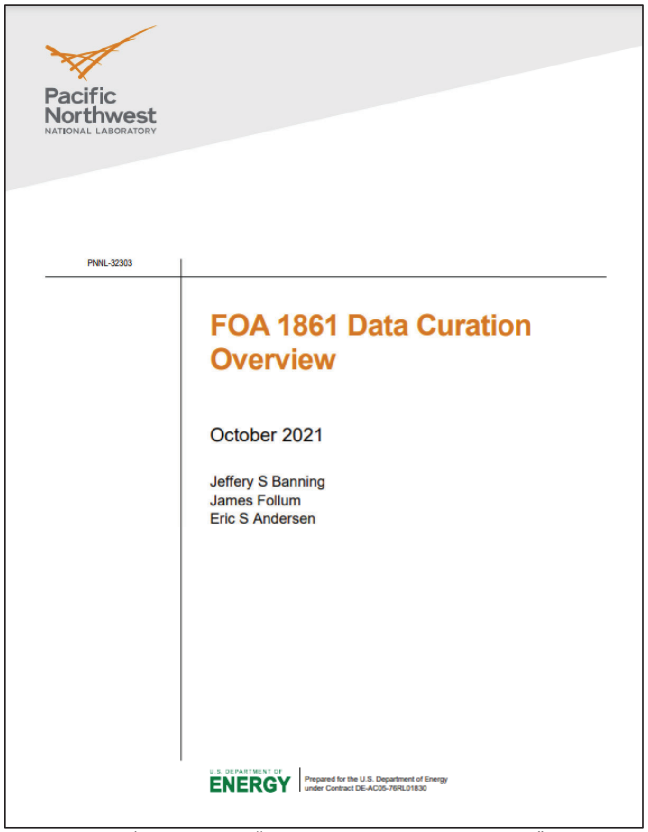
- Collected from multiple data providers from all three US interconnections
- Two years of data

Curated data provided by PNNL*

- Anonymized
- Adjusted to uniform PMU format
- Converted to Parquet file format
- Broke into two datasets: training and test
- Event logs were provided for training dataset
- Total datasets size: ~27 TB

A thorough data quality check was performed

- Provide important information for data analytics team

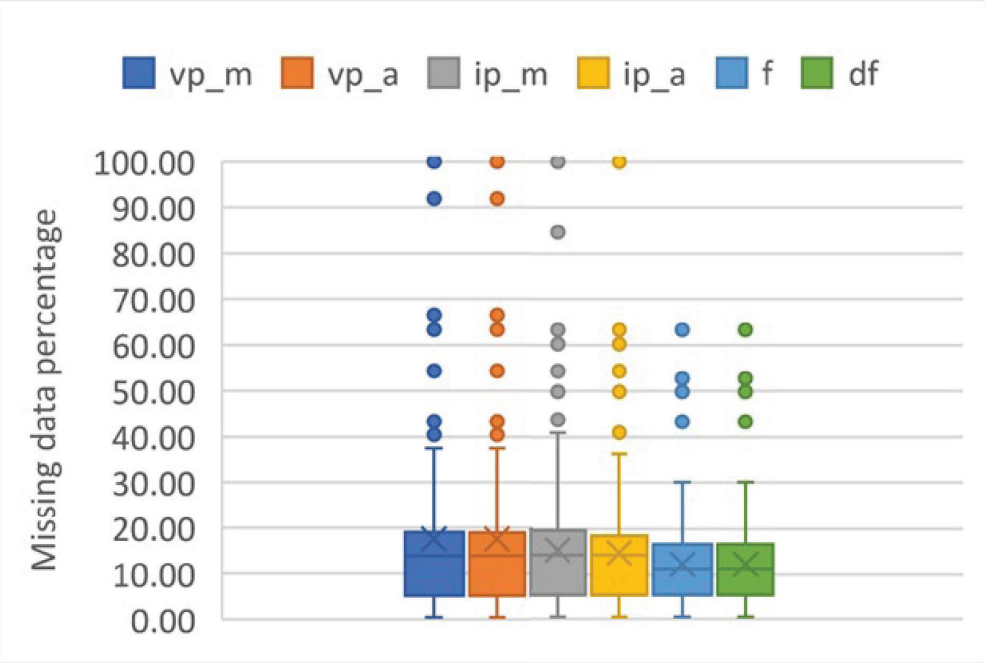


*: PNNL-32303 “FOA 1861 Data Curation Overview”
https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-32303.pdf

What have been found?

A lot of data are missing!

- Total number of PMUs: 188
- Recording period: 2016-2017
- Key measurement variables
  - Positive sequence voltage magnitude and angle: vp_m, vp_a
  - Positive sequence current magnitude and angle: ip_m, ip_a
  - Frequency: f
  - ROCOF: df (for df/dt)
- Most PMUs missing data between 5%-20%



What can be done?

Managing the data missing situation

Not much once the data has been received or archived

- Retrieve locally archived data – if possible/feasible
- Estimate – does not always work

Identify and fix the root causes

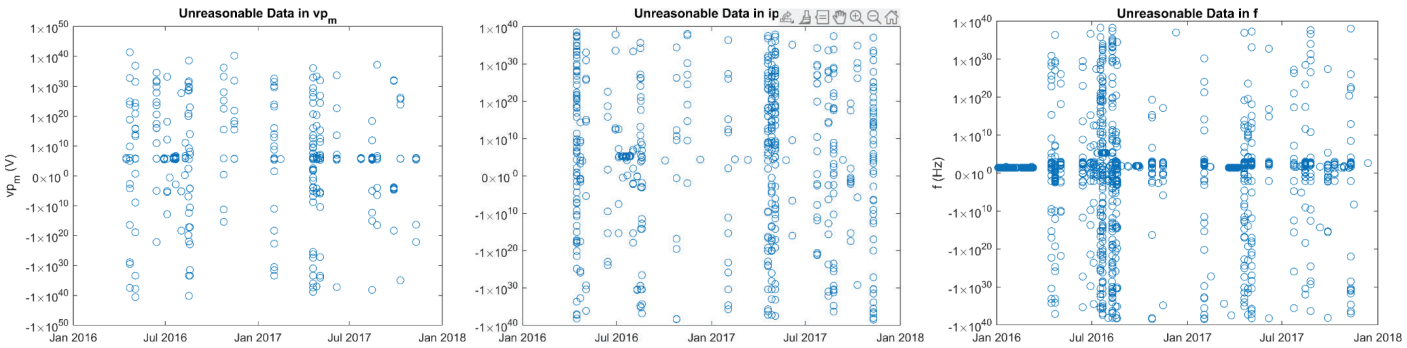
- Communication/network
  - Dedicated vs. shared, bandwidth, routers/switches, configurations, protocols, network resiliency, etc.
- Intermediate equipment
  - PDCs, gateways, etc.
- System architecture design and implementation
  - Redundancy, fallback, monitoring for fast issue reporting

What have been found?

Out-of-range values

Extremely large or negative values

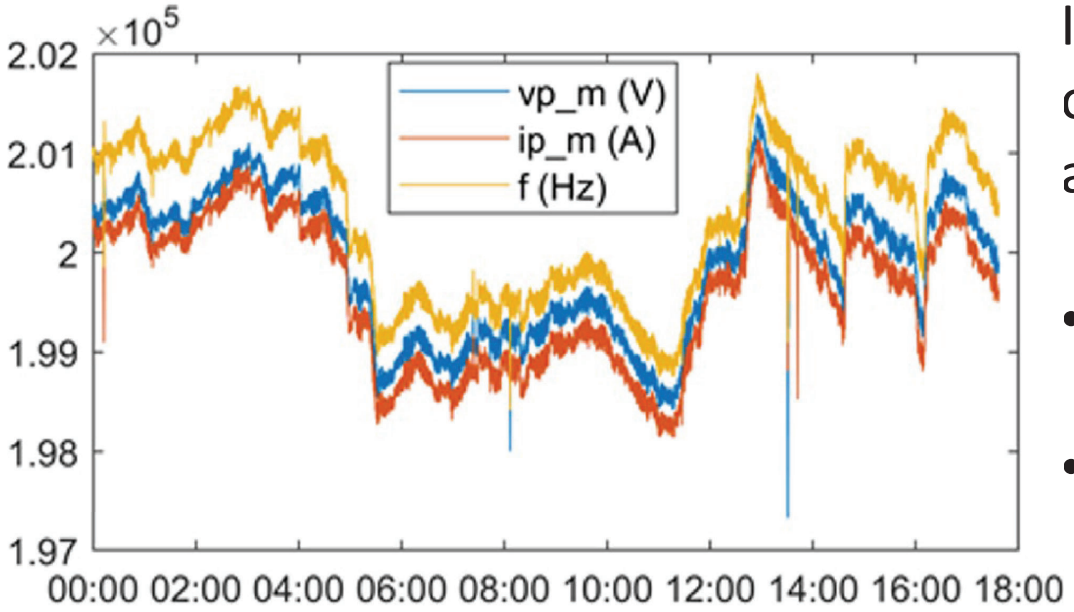
- Seen in all types of measurement signals
- 101 PMUs of out 188 PMUs had out-of-range values in one or more than one measurement signal.



- These out-of-range values are troublemakers for data analytics – must be excluded
- Root causes need to be identified and fixed
  - Rigorous PMU product testing
  - Issue tracing

What have been found?

Some out-of-range values caused by misplaced data



In one month, the 'ip_m' and 'f' of 57 PMUs had very similar value as 'vp_m'

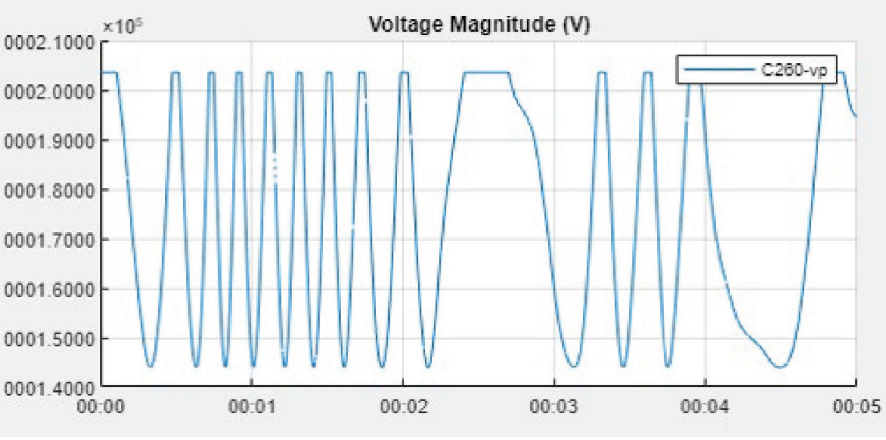
- Through visual inspection
- Automated approach needed for big dataset

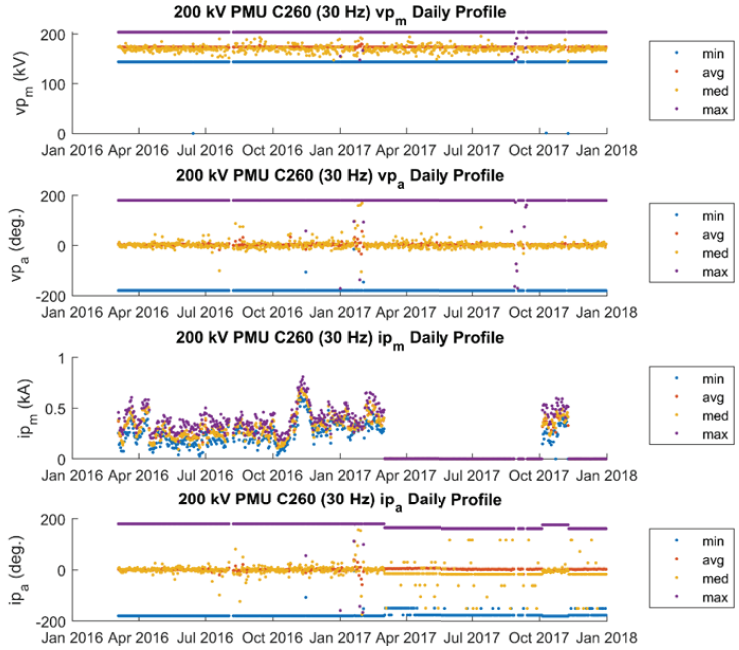
What have been found?

Constantly fluctuating voltage magnitude

One PMU's 'vp_m' values fluctuated between 145 kV and 205 kV throughout the two years.

- Reason unknown





What have been found?

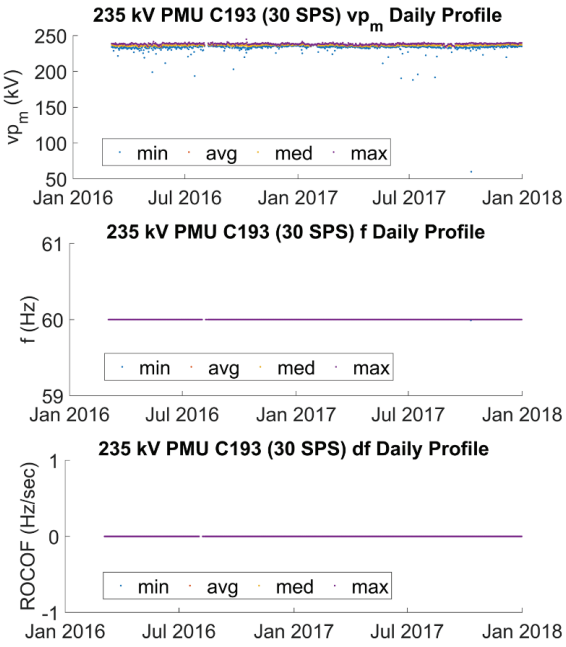
Constant frequency measurement values

18 PMUs had constant frequency measurement values

- 4 PMUs had constant 60 HZ value for entire two years
- 14 PMUs had constant “f” values during some time period(s) of the two years
  - 2 PMUs had constant 59.7 Hz values
  - 12 PMUs had constant 60 Hz values

Such values are of no use for data analytics – not real measurement values

- If no real measurement, assign “NaN” instead



What else have been found?

Incorrect time tags and duplicated data

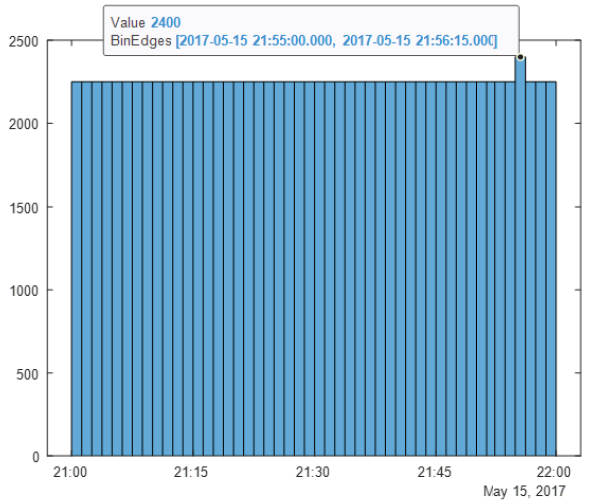
Three unique types of time tags

Type-I	Type-II	Type-III
'00:00:00.000000000'	'00:00:00.000000000'	'00:00:00.000000000'
'00:00:00.033340000'	'00:00:00.033000000'	'00:00:00.033000000'
'00:00:00.066680000'	'00:00:00.067000000'	'00:00:00.066000000'
'00:00:00.100000000'	'00:00:00.100000000'	'00:00:00.100000000'
'00:00:00.133340000'	'00:00:00.133000000'	'00:00:00.132000000'
'00:00:00.166670000'	'00:00:00.167000000'	'00:00:00.165000000'
'00:00:00.200010000'	'00:00:00.200000000'	'00:00:00.200000000'
10 PMUs	80 PMUs	98 PMUs

Two PMUs had incorrect time tags time to time

Time Tag
2016-02-29 13:15:17.7666700
2016-02-29 13:15:17.7667600

In one hour of a day, 79 PMUs had duplicated data



What about STAT bits?

Standard version unknown

- Embedded in data frame header
- Lost when pass through PDCs
- Generally unknown for archived data

Bits	C37.118-2005	C37.118.2-2011
15-14	Bit 15: Data valid, 0 when PMU data is valid, 1 when invalid or PMU is in test mode Bit 14: PMU error including configuration error, 0 when no error	00 = good measurement data, no errors. 01 = PMU error. No information about data 10 = PMU in test mode (do not use values) or absent data tags have been inserted (do not use values) 11 = PMU error (do not use values)
13	PMU sync, 0 when in sync	PMU sync, 0 when in sync with a UTC traceable time source
12	Data sorting, 0 by time stamp, 1 by arrival	Data sorting, 0 by time stamp, 1 by arriva
11	PMU trigger detected, 0 when no trigger	PMU trigger detected, 0 when no trigger
10	Configuration changed, set to 1 for 1 min when configuration changed	Configuration change, set to 1 for 1 min to advise configuration will change, and clear to 0 when change effected
09-06	Reserved for security, presently set to 0	Bit 09: Data modified, 1 if data modified by post-processing, 0 otherwise Bits 08-06: PMU Time Quality. Refer to codes in Table 7 in Standard
05-04	Unlocked time: 00 = sync locked, best quality 01 = Unlocked for 10 s 10 = Unlocked for 100 s 11 = Unlocked over 1000 s	Unlocked time: 00 = sync locked or unlocked < 10 s (best quality) 01 = 10 s ≤ unlocked time < 100 s 10 = 100 s < unlock time ≤ 1000 s 11 = unlocked time > 1000 s
03-00	Trigger reason: 1111–1000: Available for user definition 0111: Digital 0110: Reserved 0101: df/dt high 0100: Frequency high/low 0011: Phase-angle diff 0010: Magnitude high 0001: Magnitude low 0000: Manual	Trigger reason: 1111–1000: Available for user definition 0111: Digital 0110: Reserved 0101: df/dt high 0100: Frequency high/low 0011: Phase-angle diff 0010: Magnitude high 0001: Magnitude low 0000: Manual

What about STAT bits?

Overall situation

ERROR	SYNC	SORTING	TRIGGER	CONFIG.	MODIFY	PMUs
Error	Unsync	ByArrival	NoTrigger	NoChange	NoMod	97
Good	Unsync	ByTime	NoTrigger	NoChange	NoMod	111
Good	Sync	ByTime	Triggered	NoChange	NoMod	92
Good	Unsync	ByArrival	Triggered	NewCnfg	NoMod	97
Error	Unsync	ByTime	NoTrigger	NoChange	NoMod	11
Good	Unsync	ByArrival	NoTrigger	NoChange	NoMod	50
Error	Unsync	ByArrival	NoTrigger	NoChange	Modified	97
ErrNoInfo	Unsync	ByTime	NoTrigger	NoChange	NoMod	77
Good	Sync	ByTime	NoTrigger	NewCnfg	NoMod	106
ErrNoInfo	Unsync	ByArrival	NoTrigger	NoChange	NoMod	12

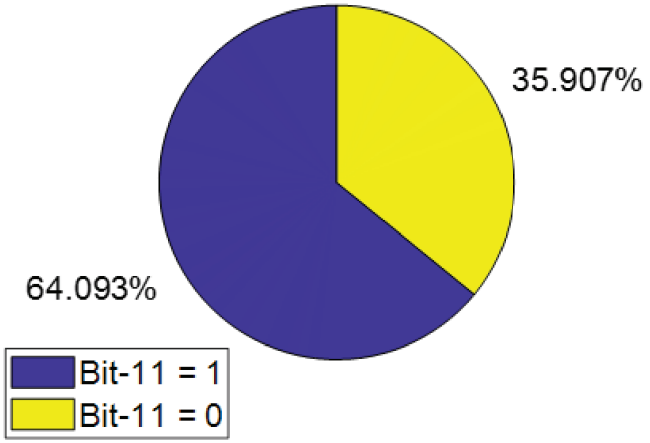
- Nonzero STAT bits occur in ~2% of the data points
- Top ten patterns cover 99.84% of them
- Most are related to time synchronization issues

What about STAT bits?

Inconsistencies in standard implementation

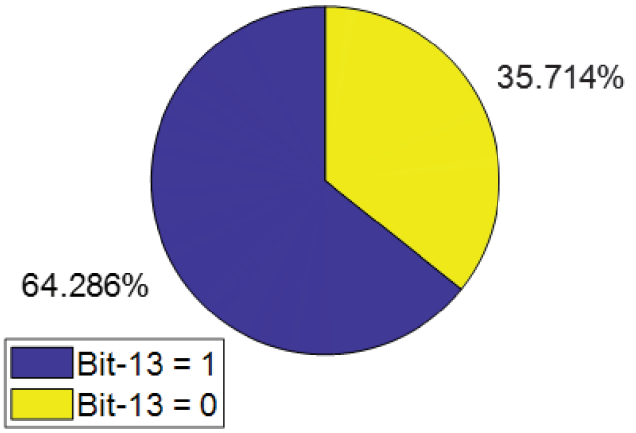
Trigger related bits

- Bit 11: PMU trigger detected
- Bits 3-0: Trigger reason



Out-of-sync related bits

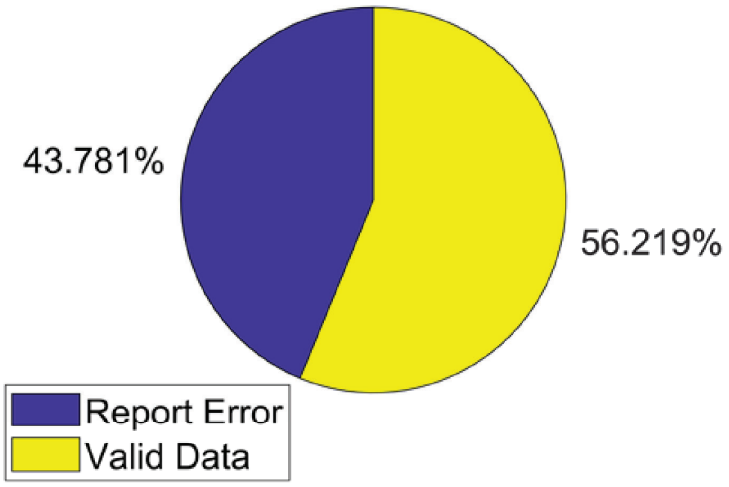
- Bit 13: PMU is out of sync
- Bits 05-04: PMU unlocked time



What about STAT bits?

Quality bits (bits 15-14) issues in implementation

Inconsistencies observed when out-of-sync bit 13 is set



Unreliable data quality indicator

- When there are data quality issues

Status Word	Partial Missing	All Missing	Unreasonable
Zero	84.75%	51.31%	99.76%
Nonzero	15.25%	48.69%	0.24%

Conclusions & recommendations

- Field recorded data shows that the data quality still has a lot of room to improve
- Most data quality issues can be effectively mitigated by identify the root causes of these issues
- Achieve high data quality should be an integral part of the synchrophasor system design, implementation, testing, and operation monitoring processes
- Standards should be more precise for uniform interpretation
- Products conformance to standards should be certified

Questions?

YHu@quanta-technology.com

QUANTA TECHNOLOGY