How to Generate Realistic Synthetic PMU dataset with Deep Generative Model?

Dr. Nanpeng Yu
Full Professor and Vice Chair
Department of Electrical & Computer engineering
Director of Energy, Economics, and Environment Research Center
Email: nyu@ece.ucr.edu

Outline

- Motivation
- Overview of pmuBAGE
- Sample Synthetic PMU Data for System Events
- Background of Event Participation Decomposition
- Overall Framework and Algorithms for Generating Synthetic PMU Data
- Numerical Results
- Conclusion

The Need for Synthetic PMU Data

- Why do we need synthetic PMU dataset?
  - Researchers/developers of machine learning algorithms for transmission system always identify the lack of large-scale and realistic PMU data set as a bottleneck for innovation
    - Security concerns, common problem for both academia and industry
  - Benchmarking across algorithms is hard when they're all tested on different data
- Is PMU data generated from dynamic simulation sufficient?
  - Advantages
    - PMU data generated is consistent with simulated dynamic system
    - Simulation model can be configured to answer any hypothetical research questions
  - Disadvantages
    - IEEE dynamic test cases can not match the complexity of real-world transmission systems
    - Parameterization of generic models (e.g. renewables) are extremely difficult to match observed dataset
    - Lack realistic details (PMU data in response to real-world events often can not be easily emulated by dynamic models, noise, missing values, outliers)

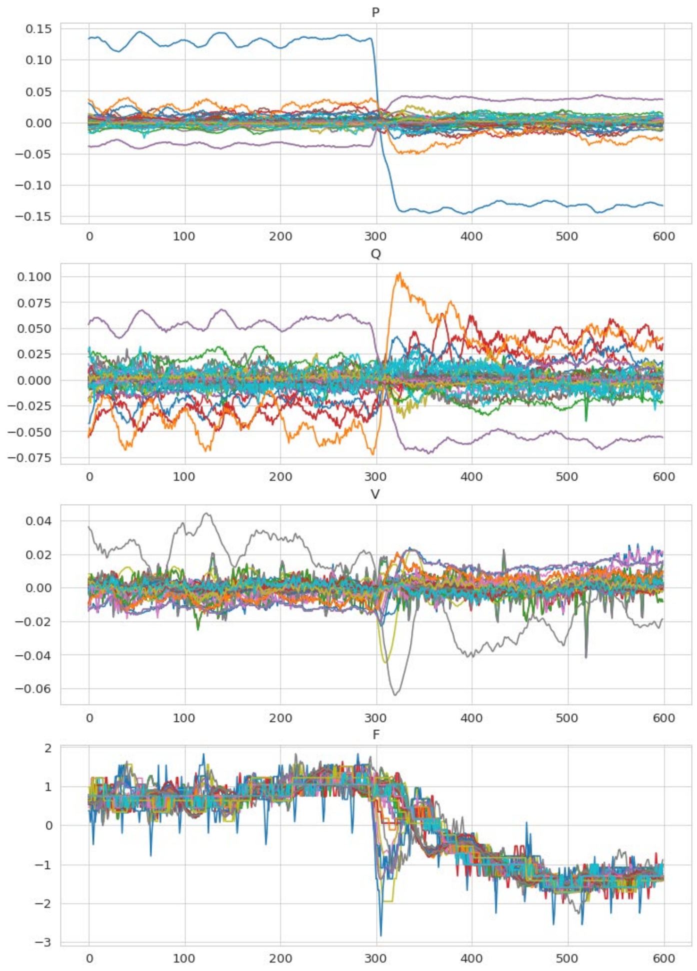
Overview of pmuBAGE: The Benchmarking Assortment of Generated PMU Events*

- pmuBAGE: the result of training a generative model on ~1,000 real-world power system events in the Eastern Interconnection.
- Publicly available at https://github.com/NanpengYu/pmuBAGE
- Advantages: accessibility, homogeneity of results & unprecedented level of realism
- Contains 84 synthetic frequency events and 620 synthetic voltage events
- 4 channels (PQ|V|F), 20 seconds event window length, 100 PMUs
- Key Ideas
  - Decompose PMU data during an event into: *Event Signatures* and *Participation Factors*
  - Event signatures can be separated into two types: inter-event and intra-event
  - Physical event signatures are PMU private and are used directly
  - Statistical participation factors are synthesized with generative model

* B. Foggo, K. Yamashita and N. Yu, "pmuBAGE: The Benchmarking Assortment of Generated PMU Data for Power System Events," in *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 3485-3496, March 2024, doi: 10.1109/TPWRS.2023.3280430.

pmuBAGE – Sample Frequency Event

pmuBAGE frequency event

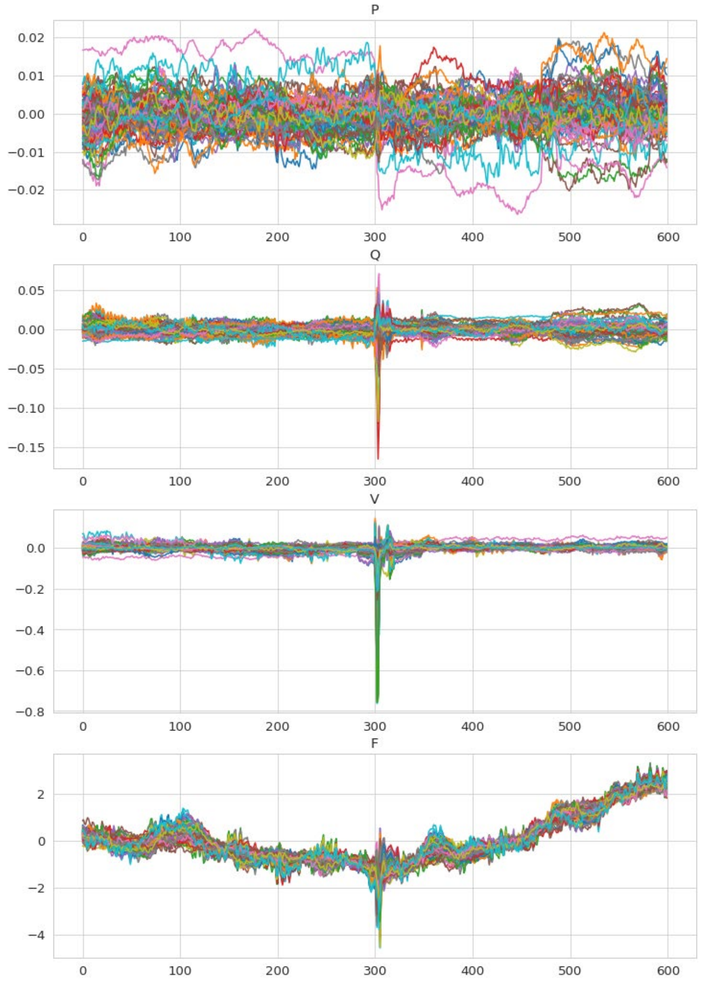


An actual frequency event

The interval between two time indices is 1 / 30 seconds. The presented data is scaled to per unit values.

pmuBAGE – Sample Voltage Event

An actual voltage event



pmuBAGE voltage event

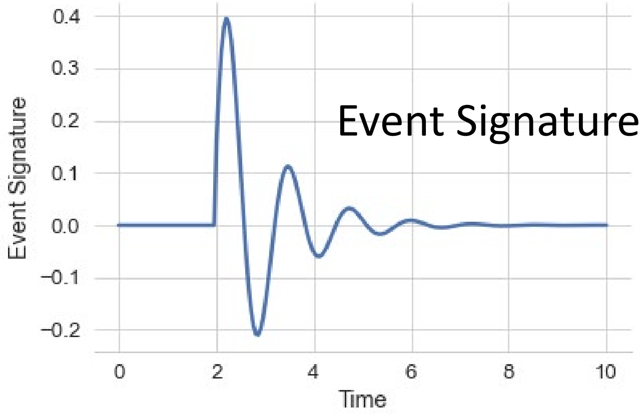
The interval between two time indices is 1 / 30 seconds. The presented data is scaled to per unit values.

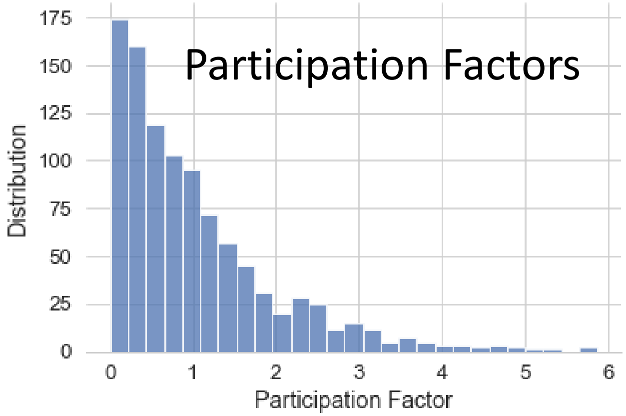
Background for Event-Participation Decomposition*

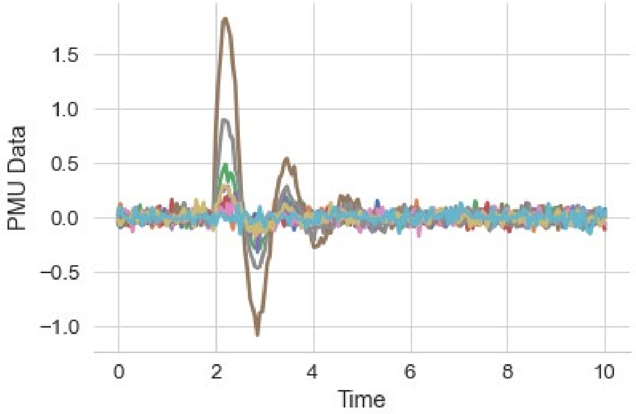
- Decomposes PMU data in an event window into:
  - A dynamic component shared by all PMUs – the <u>Event Signature</u>
  - A static component which varies by PMU – the <u>Participation Factor</u>

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{E}^{\boldsymbol{T}} + \boldsymbol{\eta}$$

$\boldsymbol{X} \in \mathcal{R}^{\boldsymbol{N}\times\boldsymbol{T}}$: Event Tensor, $\boldsymbol{P} \in \mathcal{R}^{\boldsymbol{N}\times\boldsymbol{c}}$: Participation Factors, $\boldsymbol{E} \in \mathcal{R}^{\boldsymbol{T}\times\boldsymbol{c}}$







- Properties of Physical Event Signatures
  - Depend on all PMUs, but don't depend much on any *single PMU*.
  - Event signatures are PMU private and can be used directly to generate synthetic PMU data.
- Properties of Statistical Participation Factor
  - Participation factors are not PMU private by definition.
  - They must be synthesized.

* B. Foggo and N. Yu, "Online PMU Missing Value Replacement via Event-Participation Decomposition," *IEEE TPS*, vol. 37, no. 1, pp. 488-496, Jan. 2022.

Mathematical Foundation for Event-Participation Decomposition

- Baseline Generative Model for PMU Data

R: PMU sampling rate

F_S: Nominal frequency

$$X_{t+1} = base(X_t) + \varepsilon_t \qquad base(f_t) = f_t$$

$$base(X_t) = \begin{cases} X_t & \text{Magnitude data} \\ X_t + \frac{360}{R}(f_t - F_S 1_N) & \text{Angle data} \end{cases} \qquad \varepsilon_t \sim N(0; \sigma_1 I_N)$$

- The Event-Participation Model

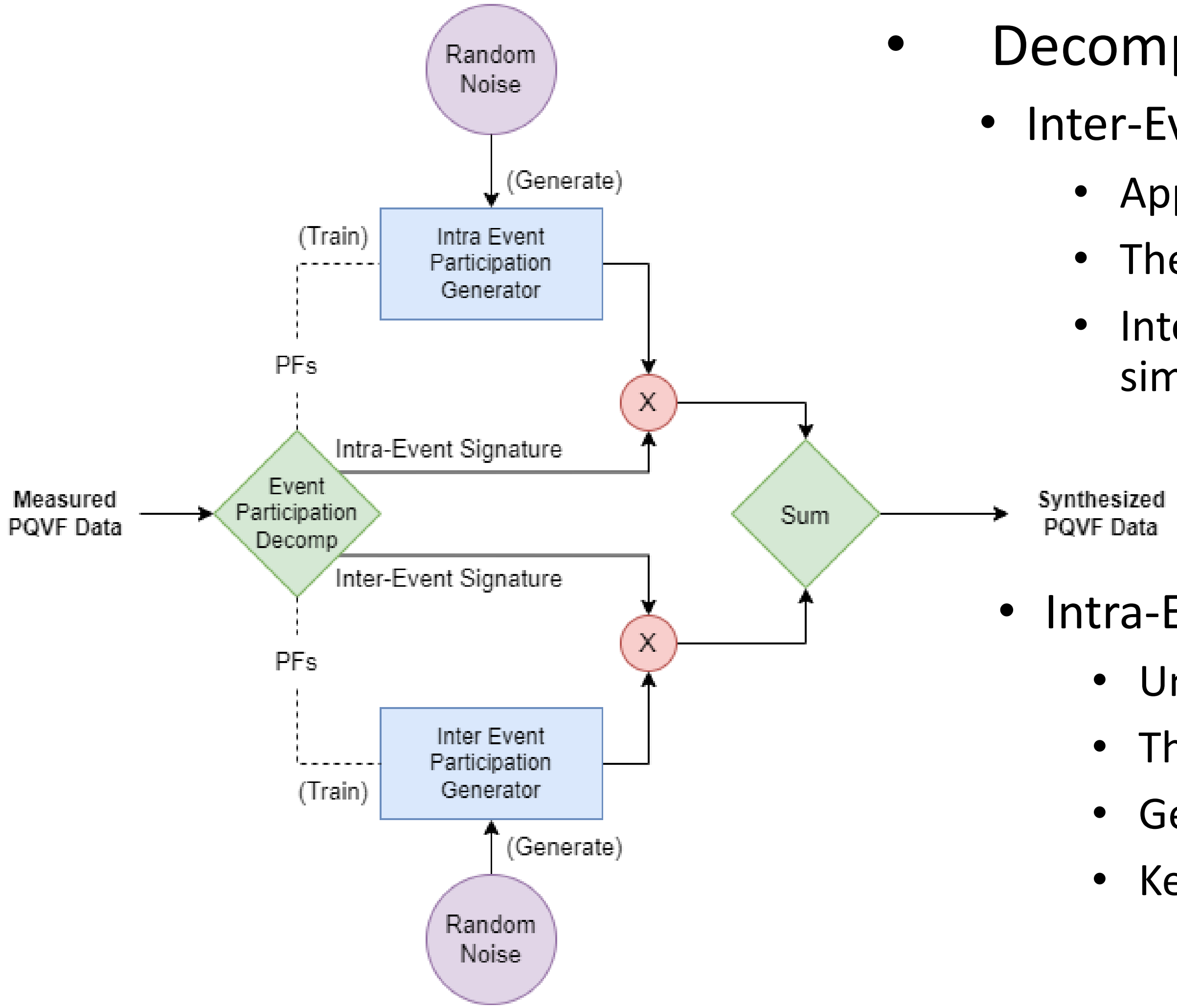
$$x_{t+1}^j = base\left(X_t^j\right) + v^j d_t + \eta_t^j,\ \eta \sim N(0; \sigma_2 I),\ v \sim \prod_j Laplace(0; \tfrac{1}{\theta})$$

- Parameter Inference: Maximize log probability → Regularized low-rank matrix approximation

$$\mathcal{L} = -\frac{\left\|y_{:t} - v d_{:t}^T\right\|_F^2}{2\sigma_2^2} - g(v),\ g(v) = \theta \sum_j \left\|v^j\right\|_1$$

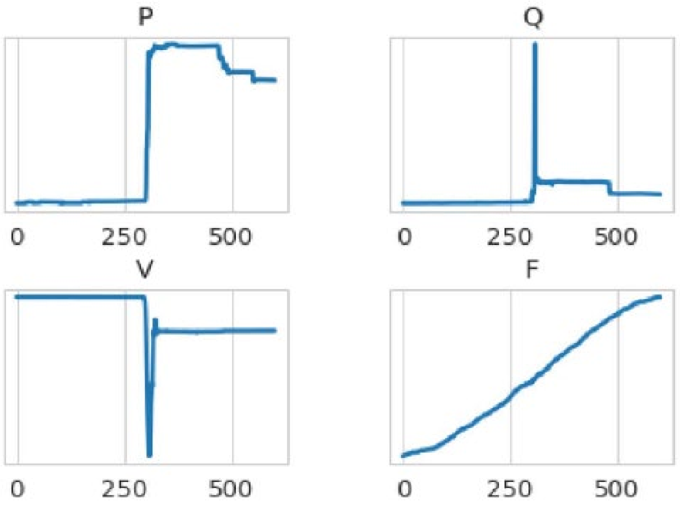
- Solve the optimization problem above with a proximal variant of the stochastic implicit Krasulina updates

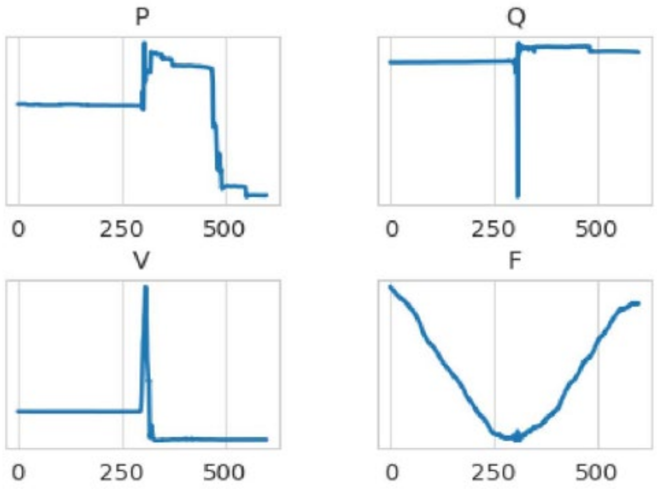
Overall Framework: Generating synthetic PMU data



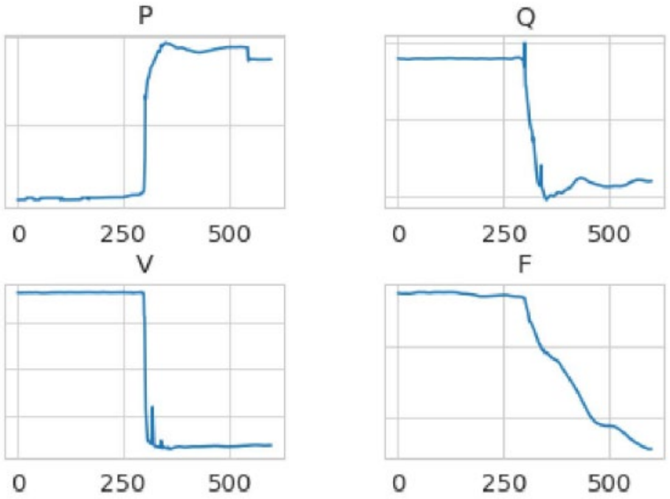
- Decompose event signatures into 2 types
  - Inter-Event Signature
    - Appear repeatedly across events with little variation
    - The corresponding participation factors are statistically simple
    - Inter-Event participation factors ~ Multivariate Gaussian after simple transformation
  - Intra-Event Signature
    - Unique components of an event
    - The corresponding participation factors are more complicated
    - Generated via a deep generative probabilistic program
    - Key architectural components
      - Feature extraction maps with cascaded convolutional network
      - Loss function with feature mapping and quantile loss

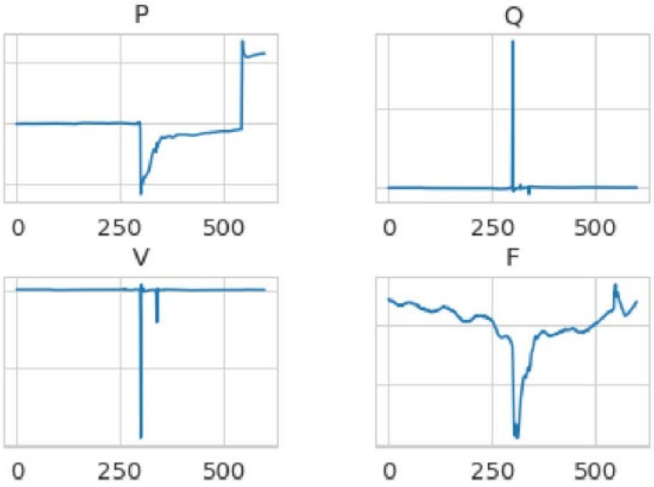
Inter-Event Signatures





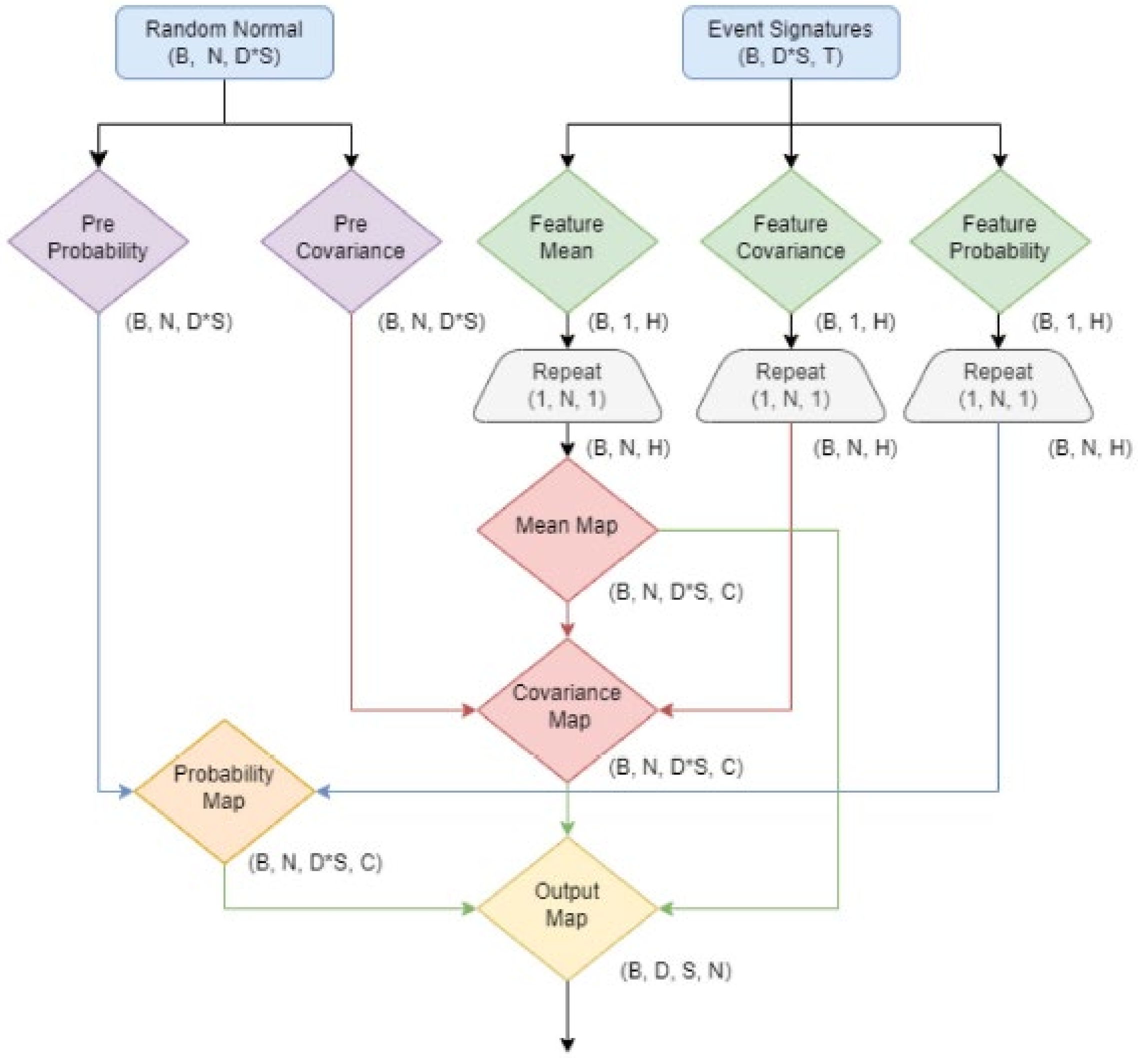
Top Two Inter-Event Signatures for Voltage Events





Top Two Inter-Event Signatures for Frequency Events

Overview of Generative Model to Simulate Event-Dependent Participation Factors



- Feature Extraction Maps
  - Feature Mean, Feature Covariance, Feature Probability
  - Encapsulates essential details of event signatures
- Pre-Maps
  - Pre Probability, Pre Covariance
  - Captures global non-Gaussian behavior amongst modes
- Mean Map
  - Represents the locations of these modes
- Covariance Map
  - Forces the average value over those PMUs to be zero
- Probability Map
  - Assign each of the generated PMUs to one of the modes
- Output Map
  - Combines probability map, covariance and mean maps

$$\mathrm{out}[i,j,s] = \sum_c p[i,j,s,c] \cdot (\mu[i,j,s,c] + \Sigma[i,j,s,c])$$

i – batch index, j – PMU index, s – PQVF index, c – mode index

Loss Function

- Evaluates how well the proposed algorithm models the given real-world PMU data
- Three categories of loss functions are used.

1. Standard Generative Adversarial Network (GAN) loss function

$$\mathcal{L}_{G_{disc}} = \mathbb{E}_{batch}\left[\left\|D\big(x, G(x)\big) - 1\right\|^2\right]$$ low error when generated points trick the discriminator into outputting a value of 1

$$\mathcal{L}_D = \mathbb{E}_{batch}\left[\left\|D\big(x, G(x)\big) - 0\right\|^2\right] + \mathbb{E}_{batch}\left[\left\|D\big(x, p(x)\big) - 1\right\|^2\right]$$ discriminator is trained to not be tricked by the generator

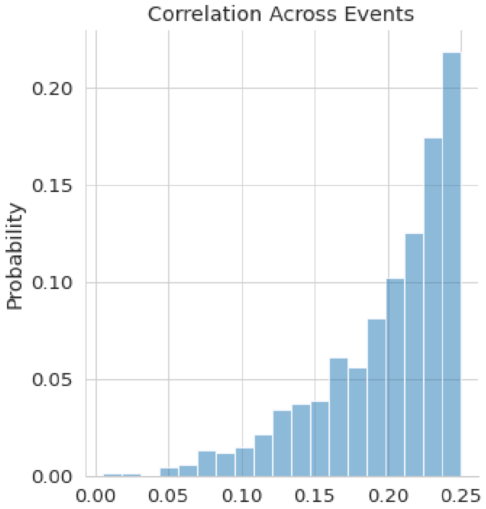
2. Statistics of generated participation factors

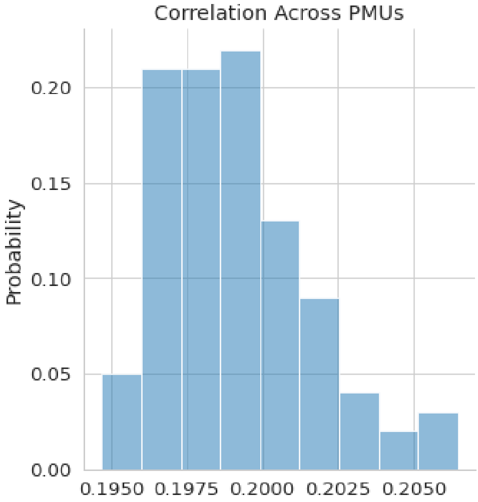
$$f_1(x) = x,\ \ f_2(x) = \min_{\mathrm{PMU-axis}}(x),\ \ f_3(x) = \max_{\mathrm{PMU-axis}} x, f_4(x) = \sigma(x * 2/(f_3(x) - \big(f_2(x)\big)),\ \ f_5(x) = f_4(-x)$$

$$f_6(x) = relu(x),\ \ f_7(x) = relu(-x),\ \ \ f_8(x) = relu(x)^2,\ \ \ f_9(x) = relu(-x)^2,\ \ \ f_{10}(x) = x^3$$

3. A New Loss Function Invented for This work, “Quantile Loss”
   - Similar to a feature match, but instead of matching the expectation of some value, we match the percentiles of the distribution to the original data at a coarse level

Correlation Analysis and Inception-Like Scoring





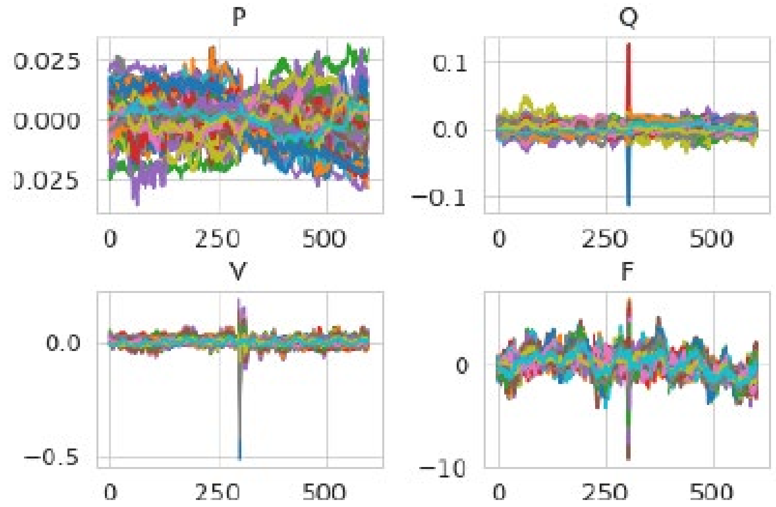
- Max correlation between synthetic and real events is 0.25
- No historical events used to train the model are compromised
- Max correlation between synthetic and real PMU measurements is 0.205.
- No PMUs used to train the model are compromised

- Quality of generated PMU data samples measured by "Inception-like score"
  - Train a standard ResNext model to classify event types of labels "frequency" and "voltage"
  - 200 epochs of training with a batch size of 50 with Binary Cross Entropy loss function

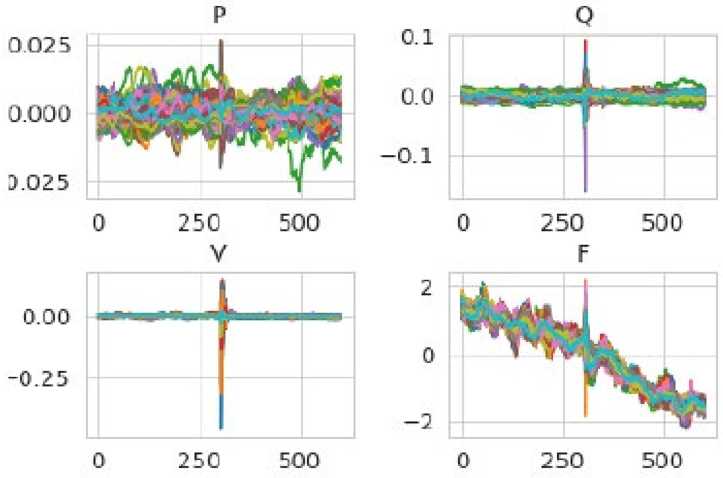
Training-Testing	Accuracy	F1	F2
Synthetic-Synthetic	99.9%	94.3%	93.3%
Synthetic-Measured	**94.3%**	**94.2%**	**92.8%**
Measured-Measured	99.8%	94.4%	91.2%
Measured-Synthetic	**93.2%**	**94.3%**	**92.7%**

- No significant degradation in F1 or F2 scores in cross-comparison compared to self comparisons.
- pmuBAGE could serve the community as a standard benchmarking tool for event detection and classification.

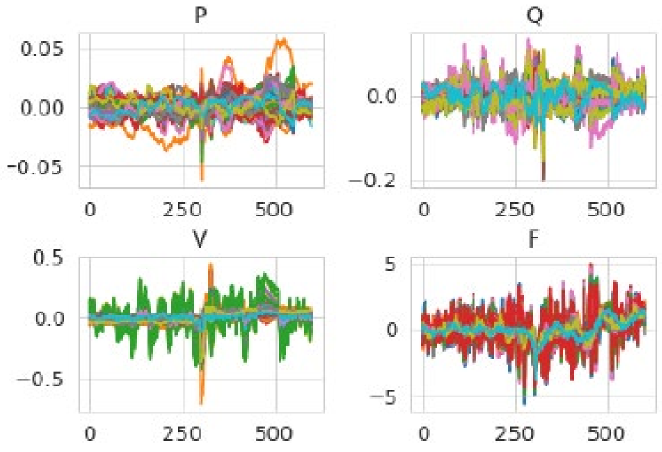
System Events with Different Causes



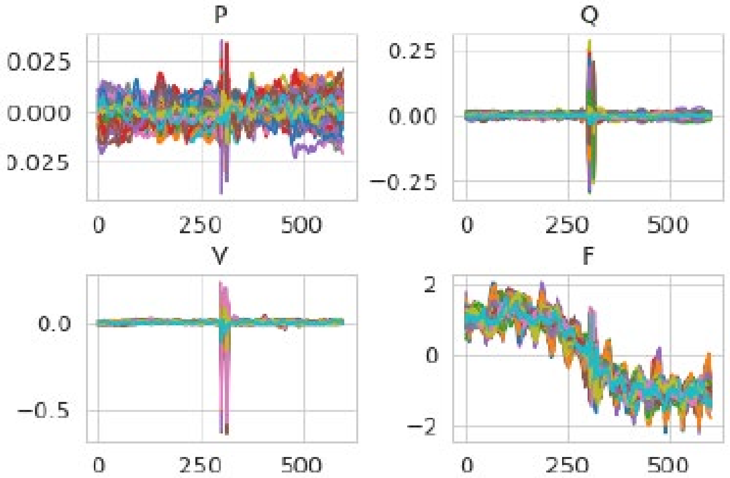
Voltage event causes by lightening strike



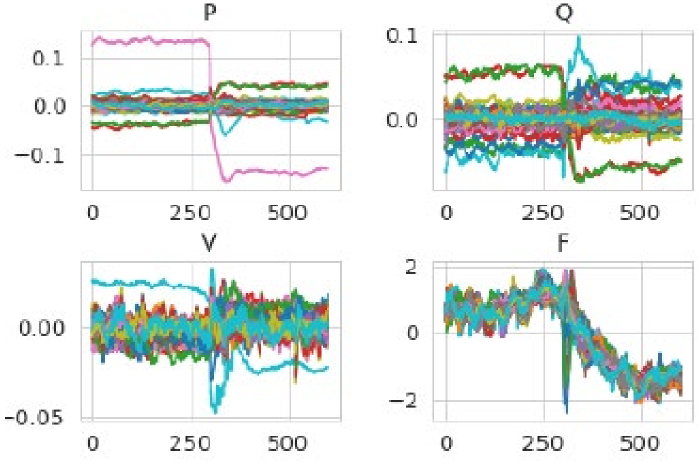
Voltage event causes by line tripping



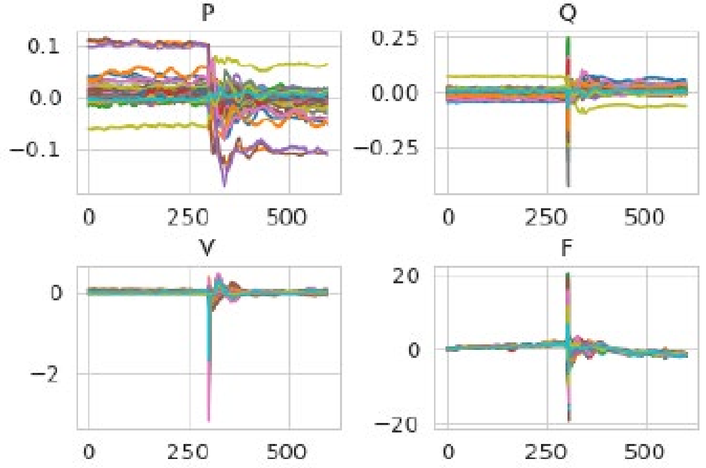
Voltage event causes by wind



Voltage event (equipment failure)



Frequency event (generator tripping)



Frequency event (generator equipment failure)

Conclusion and Future Work

- The synthesized PMU dataset created is highly realistic and does not significantly degrade important training evaluation metrics.
- pmuBAGE serves the community as a standard benchmarking tool for event detection and classification tasks.
- Improve realistic banding for frequency data – the tendency for several PMUs to have the same frequency behavior, especially during a frequency event.
- pmuBAGE occasionally displays inter-area oscillations that do not dampen as quickly as actual data.

Contact Information

Nanpeng Yu. Email: nyu@ece.ucr.edu

Thank You

Questions?

PhD Student and Postdoctoral Researcher: Brandon Foggo and Koji Yamashita

Funding Support for the Project: Department of Energy