NASPI North American SynchroPhasor Initiative

National infrastructure for Artificial Intelligence on the Grid (NI4AI)

Panel Presentation & Discussion

April 15, 2020

ni4ai.org

Goals for this Webinar:

- Update on ARPA-E funded Project
- Share ideas and inspiration for new methods in data analytics
- Recruit participation in ni4ai
- Cultivate a collaborative research community

Logistics

- All participants are muted, camera off.
- If you would like to speak, click on the "Raise Hand" button.
  The host will give you access to speak, and the moderator may call on you (time permitting), but you still need to unmute yourself.
- The chat window can be used and seen by all participants.
- We will use a separate site, Pollev.com/ni4ai245 , to solicit questions and feedback during this webinar. You can also post comments to the same site by text message from your phone: **Text "ni4ai245" to 37607** (not case sensitive). Your comments will be visible to all participants.

Today's Panelists

Sean P. Murphy
CEO, PingThings
sean@pingthings.io

ARPA-E Project
Objectives and Status

Alexandra (Sascha) von Meier
Adjunct Professor, EECS
UC Berkeley
vonmeier@berkeley.edu

Introduction
Motivation for NI4AI
Discussion/Conclusion

Mohini Bariya
PhD Candidate, EECS
UC Berkeley
mohini@berkeley.edu

Illustrative Use Case:
Voltage Sag Exploration

Luigi Vanfretti
Professor, Rensselaer
Polytechnic Institute
vanfrl@rpi.edu

RT ML-Based Forced
Oscillation Detection

Tetiana Bogodorova
Research Scientist
Rensselaer Polytechnic Institute
tetiana.bogodorova@gmail.com

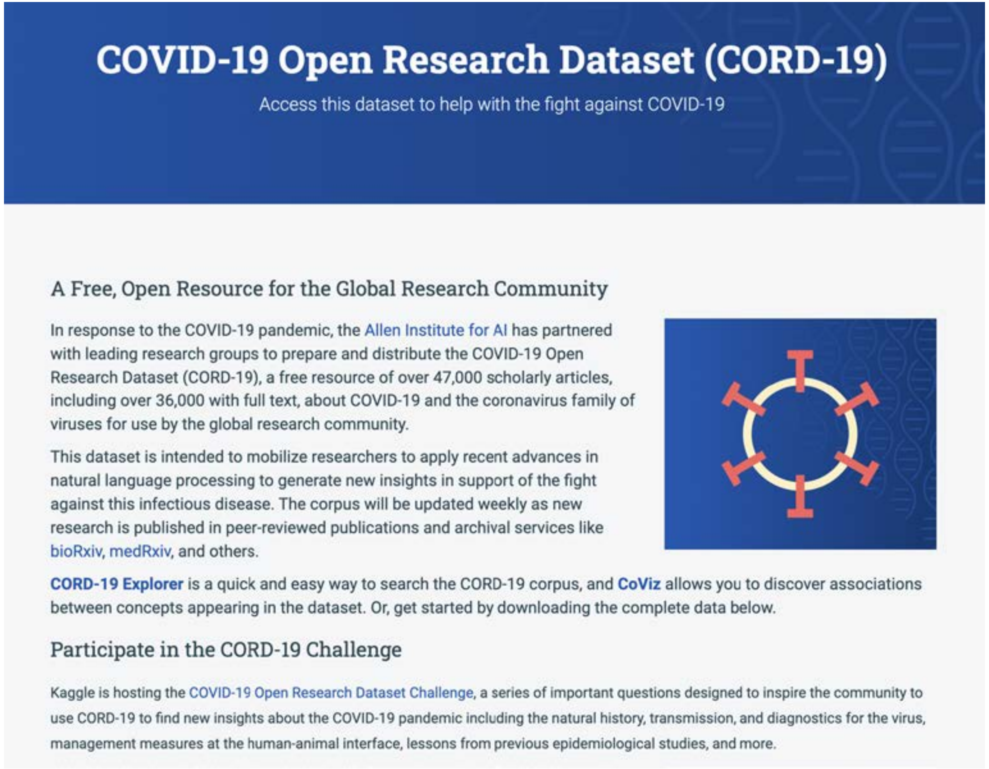
Synthetic Data for ML-Based
Decision Making Applications

Kevin D. Jones
Engineering Analytics
Dominion Energy - Power Delivery
kevin.d.jones@outlook.com

Utility Perspective on
Collaborative Big Data Analytics

NI4AI Motivation

Collaborative science with shared, anonymized data is vastly faster and more powerful than what any one team can accomplish alone.



COVID-19 Open Research Dataset (CORD-19)

Access this dataset to help with the fight against COVID-19

A Free, Open Resource for the Global Research Community

In response to the COVID-19 pandemic, the Allen Institute for AI has partnered with leading research groups to prepare and distribute the COVID-19 Open Research Dataset (CORD-19), a free resource of over 47,000 scholarly articles, including over 36,000 with full text, about COVID-19 and the coronavirus family of viruses for use by the global research community.

This dataset is intended to mobilize researchers to apply recent advances in natural language processing to generate new insights in support of the fight against this infectious disease. The corpus will be updated weekly as new research is published in peer-reviewed publications and archival services like bioRxiv, medRxiv, and others.

CORD-19 Explorer is a quick and easy way to search the CORD-19 corpus, and **CoViz** allows you to discover associations between concepts appearing in the dataset. Or, get started by downloading the complete data below.

Participate in the CORD-19 Challenge

Kaggle is hosting the COVID-19 Open Research Dataset Challenge, a series of important questions designed to inspire the community to use CORD-19 to find new insights about the COVID-19 pandemic including the natural history, transmission, and diagnostics for the virus, management measures at the human-animal interface, lessons from previous epidemiological studies, and more.

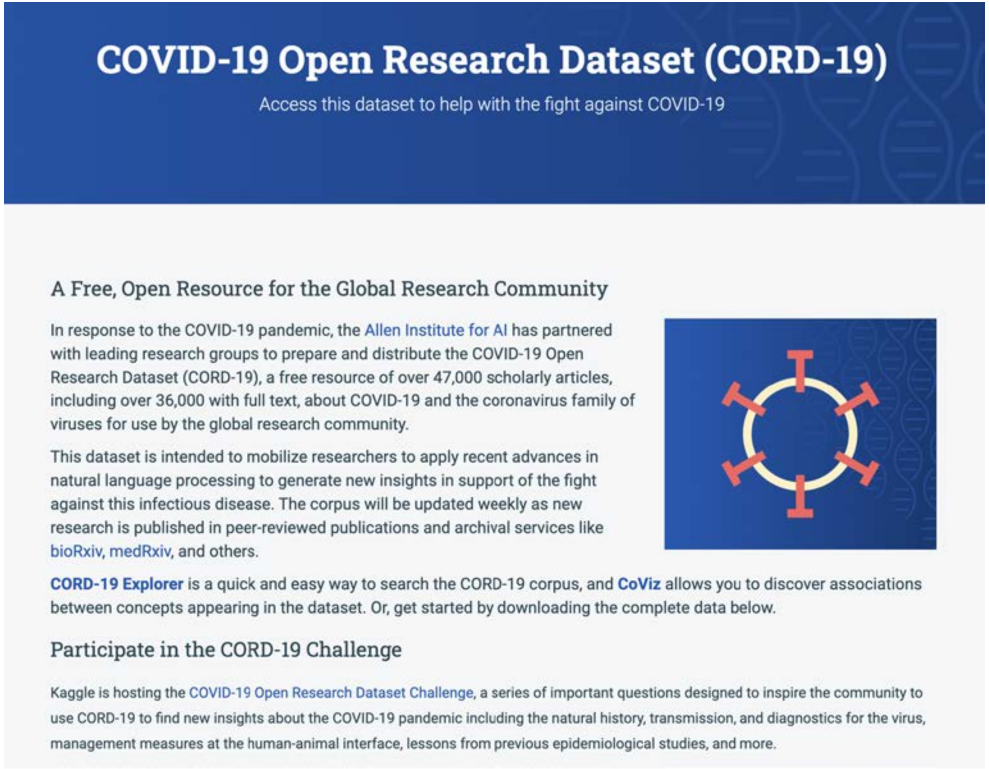
https://pages.semanticscholar.org/coronavirus-research

NI4AI Motivation

Collaborative science with shared, anonymized data is vastly faster and more powerful than what any one team can accomplish alone.

Requirements:

- Common goals
- Well-defined objectives
- Accessible data
- Suitable tools



COVID-19 Open Research Dataset (CORD-19)

Access this dataset to help with the fight against COVID-19

A Free, Open Resource for the Global Research Community

In response to the COVID-19 pandemic, the Allen Institute for AI has partnered with leading research groups to prepare and distribute the COVID-19 Open Research Dataset (CORD-19), a free resource of over 47,000 scholarly articles, including over 36,000 with full text, about COVID-19 and the coronavirus family of viruses for use by the global research community.

This dataset is intended to mobilize researchers to apply recent advances in natural language processing to generate new insights in support of the fight against this infectious disease. The corpus will be updated weekly as new research is published in peer-reviewed publications and archival services like bioRxiv, medRxiv, and others.

CORD-19 Explorer is a quick and easy way to search the CORD-19 corpus, and **CoViz** allows you to discover associations between concepts appearing in the dataset. Or, get started by downloading the complete data below.

Participate in the CORD-19 Challenge

Kaggle is hosting the COVID-19 Open Research Dataset Challenge, a series of important questions designed to inspire the community to use CORD-19 to find new insights about the COVID-19 pandemic including the natural history, transmission, and diagnostics for the virus, management measures at the human-animal interface, lessons from previous epidemiological studies, and more.

https://pages.semanticscholar.org/coronavirus-research

NI4AI Motivation

Collaborative science with shared, anonymized data is vastly faster and more powerful than what any one team can accomplish alone.

Requirements:

- Common goals ✓
- Well-defined objectives
- Accessible data
- Suitable tools

NI4AI Motivation

Collaborative science with shared, anonymized data is vastly faster and more powerful than what any one team can accomplish alone.

Requirements:

- Common goals ✓
- Well-defined objectives
- Accessible data
- Suitable tools

Barriers to AI on the grid:

- Accessing data within different silos
- Data quality assurance and maintenance
- Matching skills and tools between AI/ML experts and grid community
- Transitioning research into production

Online Poll

Participate by visiting

https://pollev.com/ni4ai245

(link posted in zoom chat)

1. Enter your name (if you wish)
2. Type your response
3. Upvote others' responses

What is your favorite grid component?

You have responded 4 times

Type your response here

1	Sensors!
0	Capacitor bank
0	Transformers
0	Wood poles

Click to upvote answers

What challenges have you experienced in working with grid data?

Top

62	Large dataset that needs cleaning, preparation and sometimes change of data from one software platform to another.
31	Accessing the data we have
30	Retrieving data fast enough
29	As a student researcher, access to grid data is very scarce. Another main problem is the missing data issue and noise in data.
26	Running analytics
16	Getting access to data
15	Data Quality, Data availability, label of historical events.
10	Moving data from one database into another
8	Data quality issues mainly.

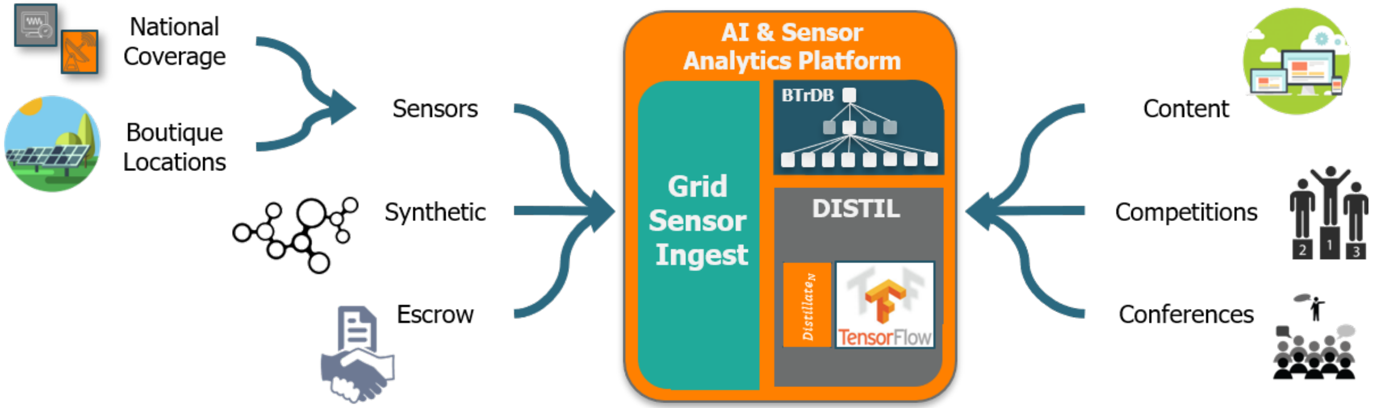
A National Infrastructure for Artificial Intelligence on the Grid (NI4AI)

The overarching objective is to remove any and all obstacles to the rapid development, adoption, and deployment of new use cases based on analytics, machine learning (ML), and artificial intelligence (AI) for sensor data measuring the electric grid.

Data

Platform

Community



National Coverage

Boutique Locations

Sensors

Synthetic

Escrow

AI & Sensor Analytics Platform

Grid Sensor Ingest

BTrDB

DISTIL

Distillate

TensorFlow

Content

Competitions

Conferences

(1) Data - Motivation

(1) Historical Data Sets

Objective: to provide a second life for projects that (1) have already collected data and (2) are willing to make it broadly available.

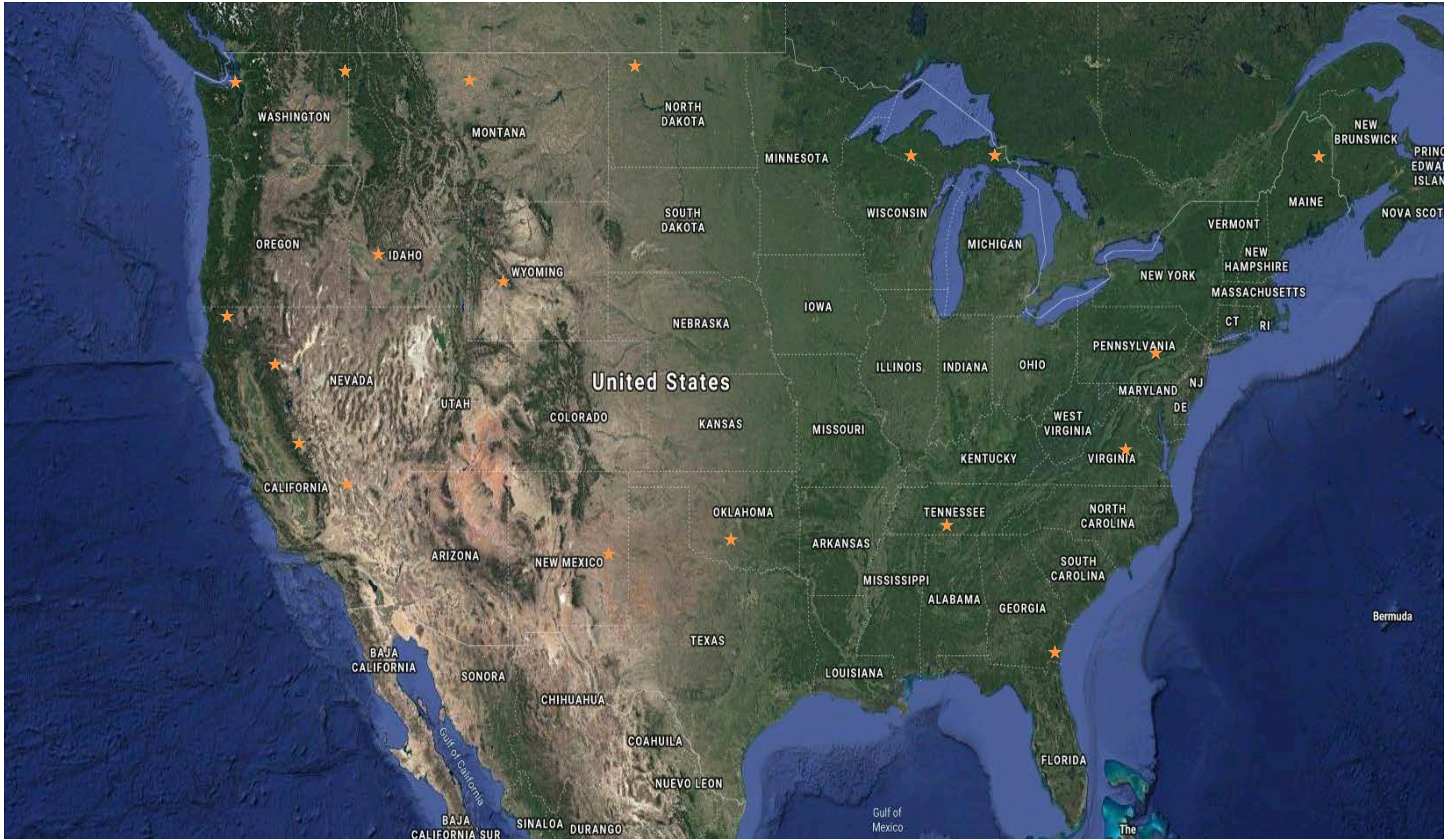
- 90 TB of continuous Point on Wave
  - 9 distribution sensors
  - 4 years of continuous data
  - 100 kHz
  - Geographically distributed

- 10 TB of uPMU data
  - Multiple distribution PMUs
  - Collected over a year
  - 60 Hz
  - California

- More Coming…
- Is there anyone that we should talk to?

(1) Wide Area Data Set

<u>Objective</u>: to explore the art of the possible for applications that transcend organizational boundaries.



(1) Boutique Data Sets

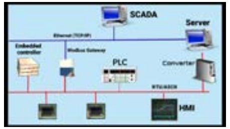
Objective: to instrument “interesting” portions of the grid to tackle specific problems.

Suggestions Welcome!

(1) Sensor Selection

- A Synchrophasor
- B Point on wave
- C Power quality
- D New sensor(s)?
- E All of the above

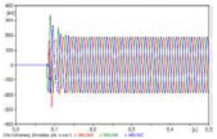
Smart Meters



SCADA

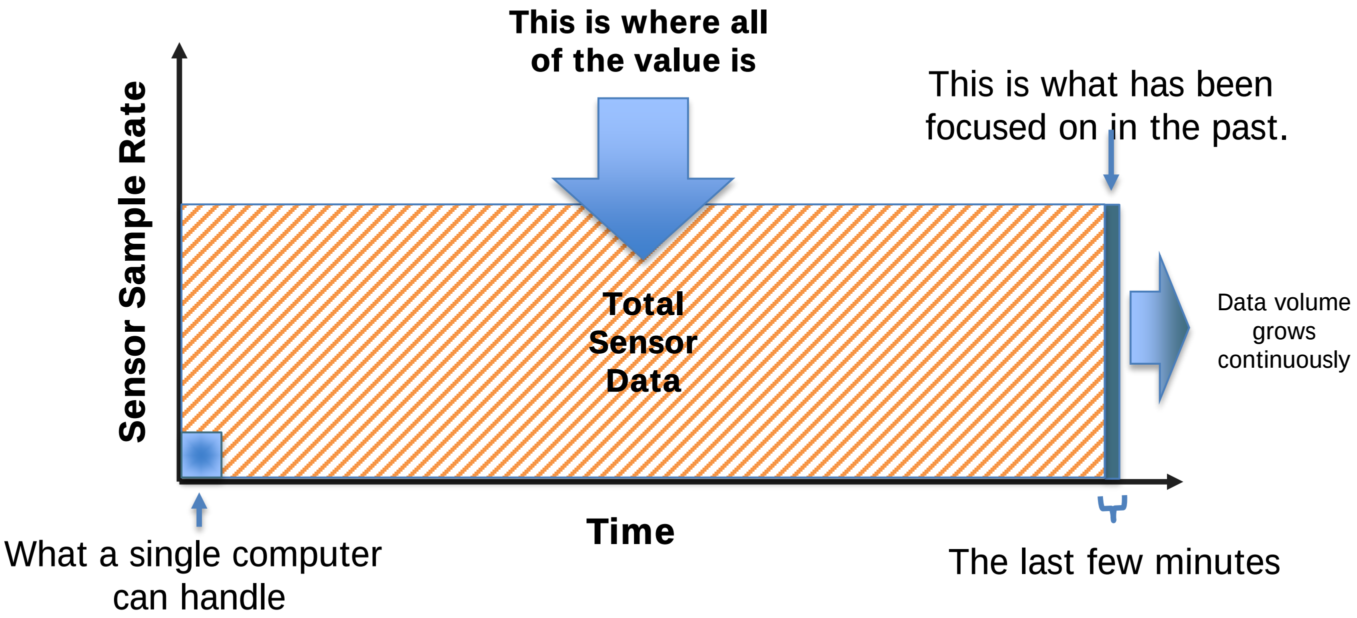
PMUs

Distribution PMUs



Point on Wave

(2) The Platform: Objectives



This is where all of the value is

This is what has been focused on in the past.

Total Sensor Data

Data volume grows continuously

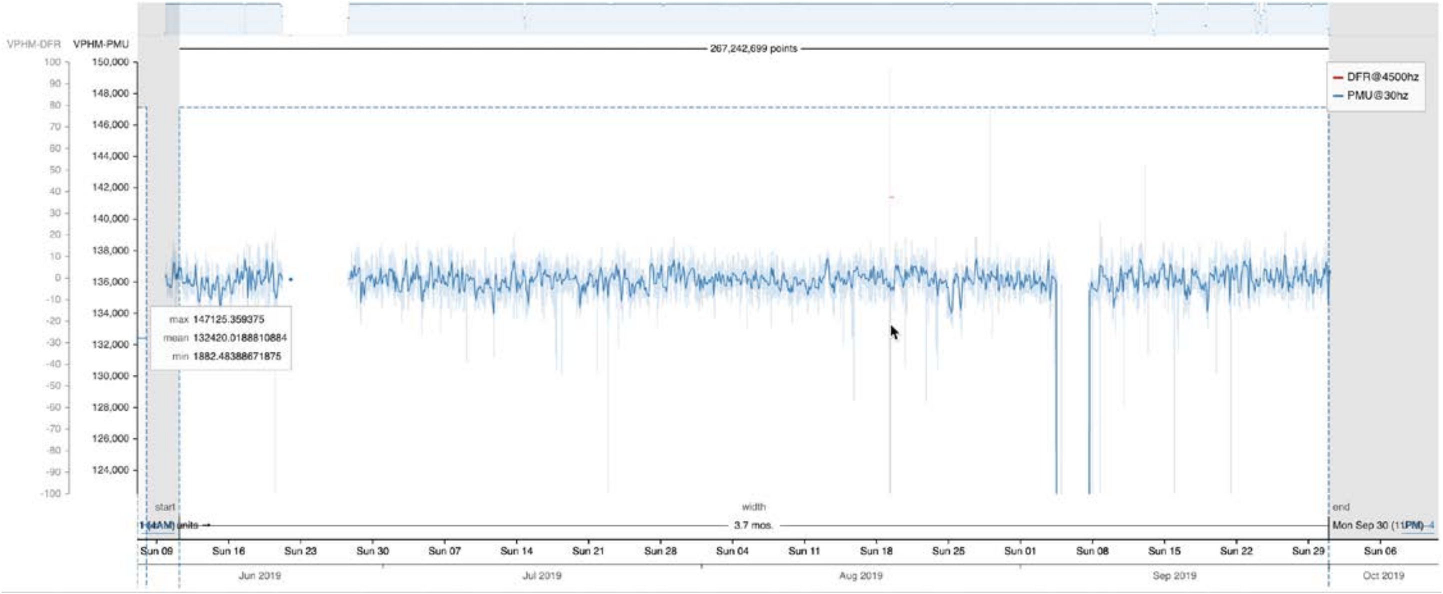
Sensor Sample Rate

Time

What a single computer can handle

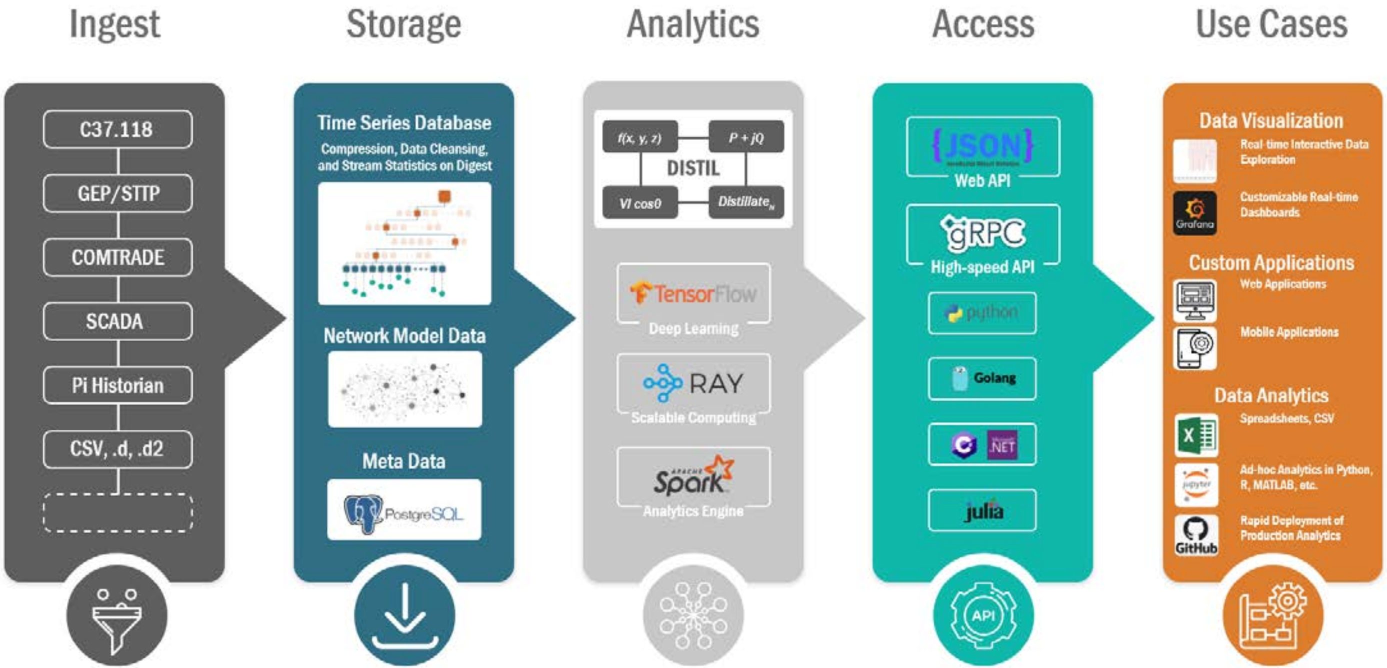
The last few minutes

(2) Visualization Demo



This unedited movie shows a user zooming in from months of data (267,242,699 data points) to microseconds, pinpointing an event lasting less than 1/10th of a second. The blue time series is sampled at 30Hz and the red at 4.5KHz

(2) The Platform

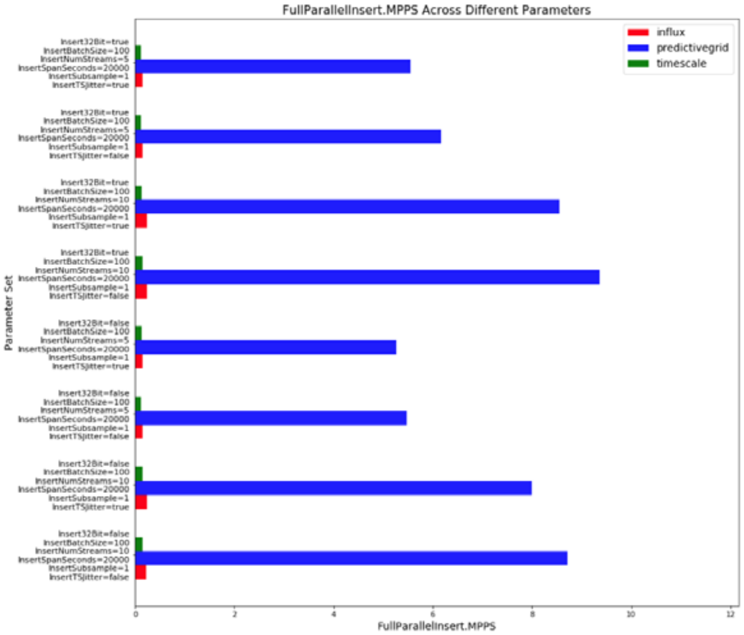


Demos are available upon request.

Contact sean@pingthings.io.

(2) The Platform: Benefits

- Exponentially accelerated queries **(10,000x faster than competitors)**
    - Visualization / Aggregation
    - Anomaly detection
    - Data quality analysis
- **10-1000x faster** for other operations
- Read/write 15M/30M points/sec/node
- Benchmarked to 100x larger deployments than Dominion Energy (1,000 pmus)



(3) The Community!

- Utilities
- Universities
- Colleges
- Industry
- Consulting groups
- Research orgs
- National Labs

Not just power engineering!

(3) The Community: Content

Conferences

Challenges

Content

Path to Commercialization

Team Composition

Sean Patrick Murphy
Principal Investigator
PingThings

Dr. Michael Andersen
Technology Architect
PingThings

Dr. Kevin Jones
Co-PI
Dominion Energy

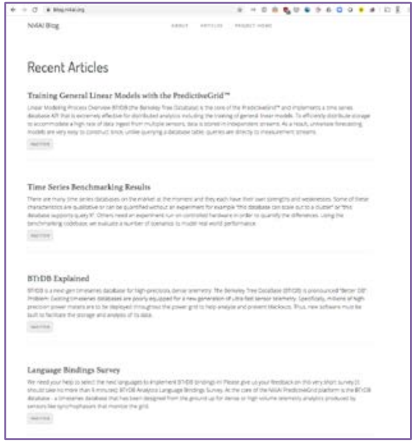
Dr. Sascha von Meier
UC Berkeley Lead
Berkeley UNIVERSITY OF CALIFORNIA

Next Steps

- What sensors should be used?
- What data sets should we create?
- What should be monitored?
- What questions should the data science competitions address?
- Do you have
  - simulated data,
  - Historical data,
  - Data to be collected
- that you would like to contribute?
- Do you want access to the platform and data?

Contact Information

Sean Patrick Murphy
PingThings, Inc
sean@pingthings.io



Recent Articles

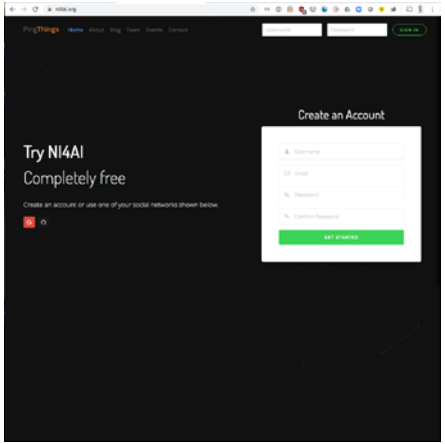
Training General Linear Models with the PredictiveGrid™

Time Series Benchmarking Results

BTrDB Explained

Language Bindings Survey

Blog
https://blog.ni4ai.org



Try NI4AI
Completely free

Create an account or use one of your social networks shown below.

Create an Account

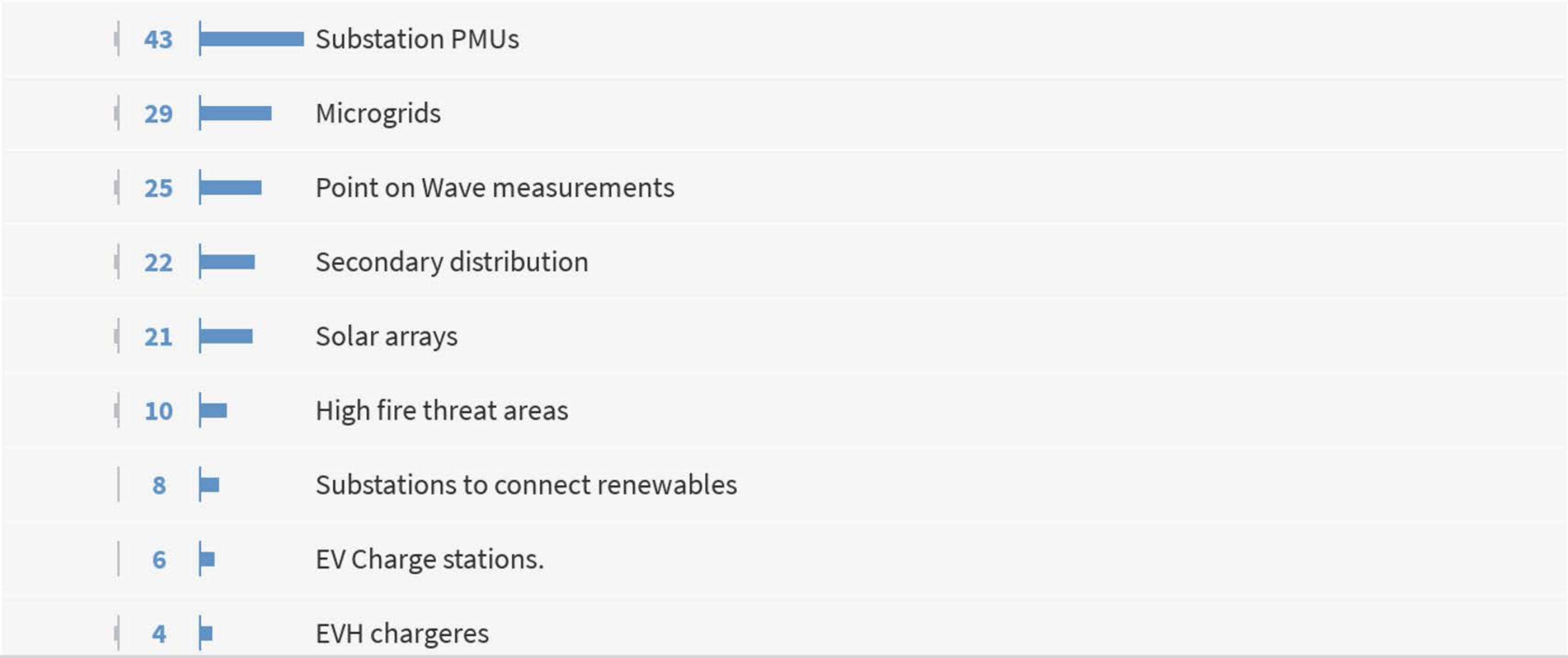
Website
https://ni4ai.org

Acknowledgement

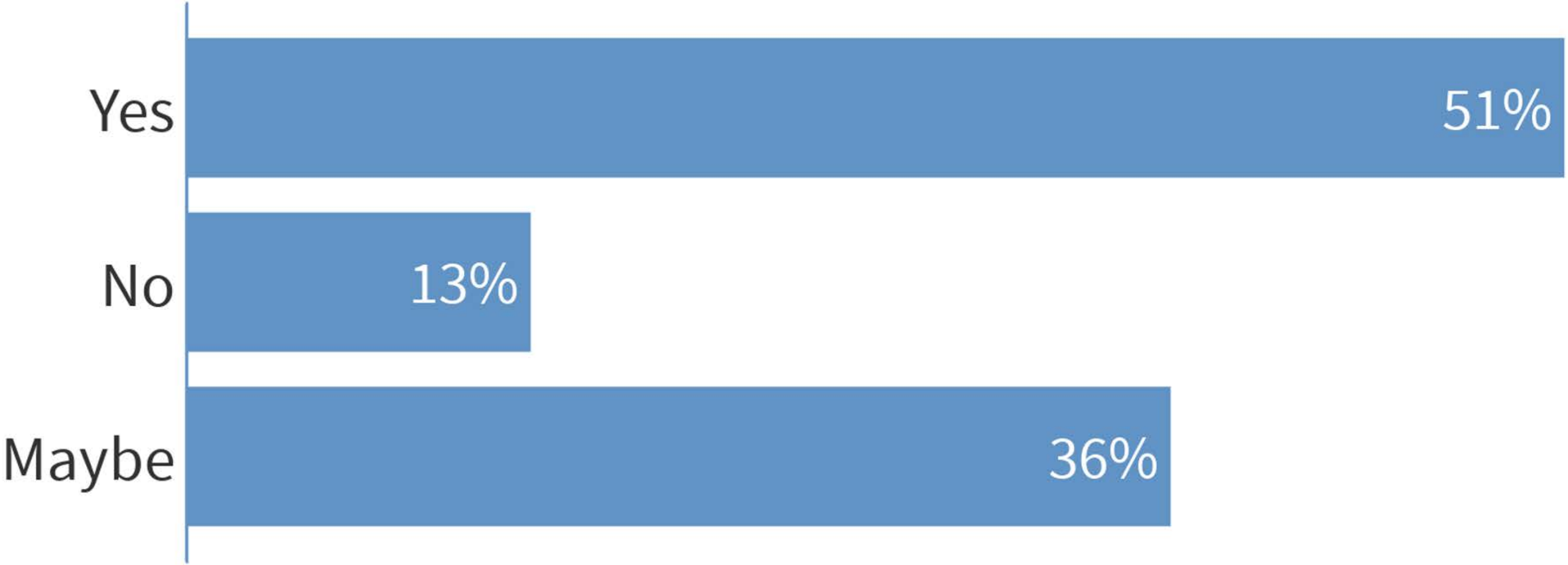
The information, data, or work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0001104. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Where would you want to see more sensors deployed, and what types of sensors?

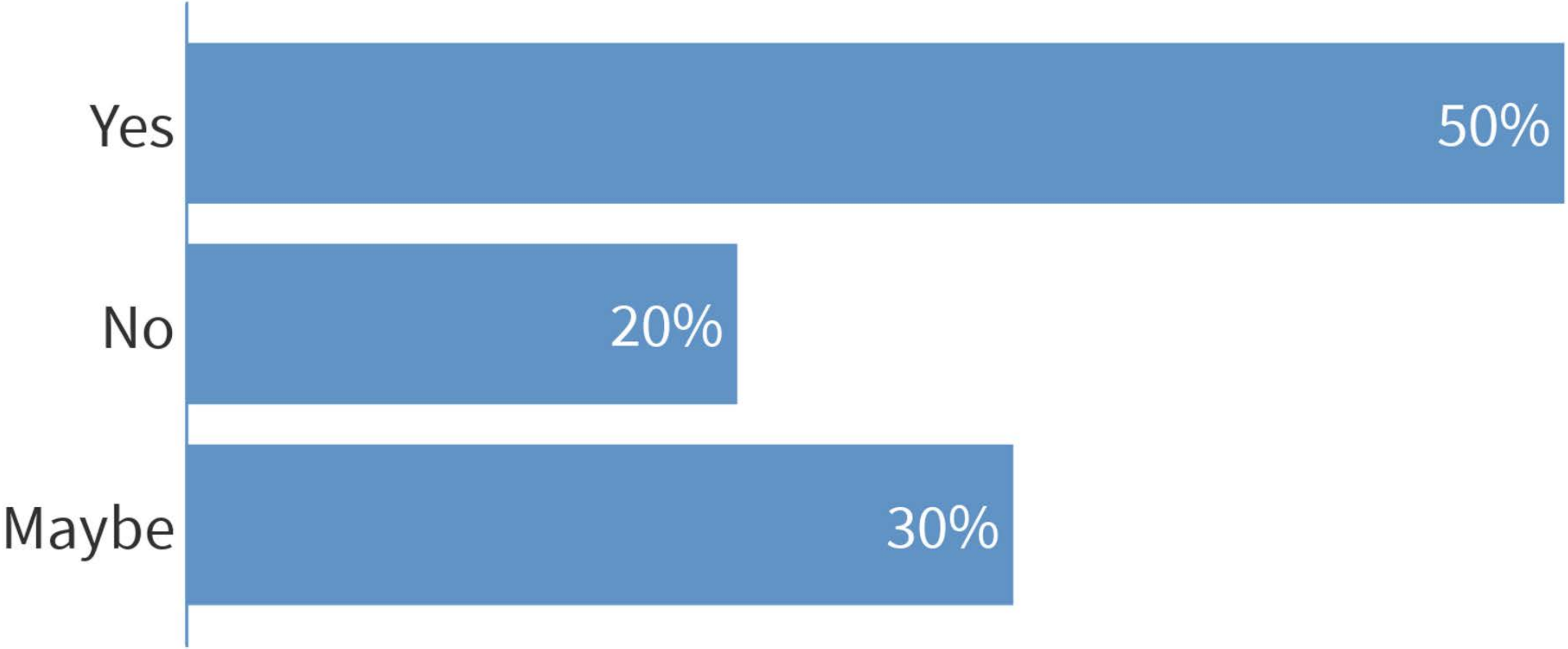
Top



Would your organization be interested in hosting NI4AI sensors, if you retain discretion over who can access the data?



Would your organization be interested in hosting NI4AI sensors, if all the data becomes public?



An example use case:

Voltage Sag Exploration

Mohini Bariya, UC Berkeley

The NI4AI platform makes it quick and easy to:

- Explore data
- Ask & answer questions
- Share results

I will demo the typical development process for one application: voltage sag detection and analysis.

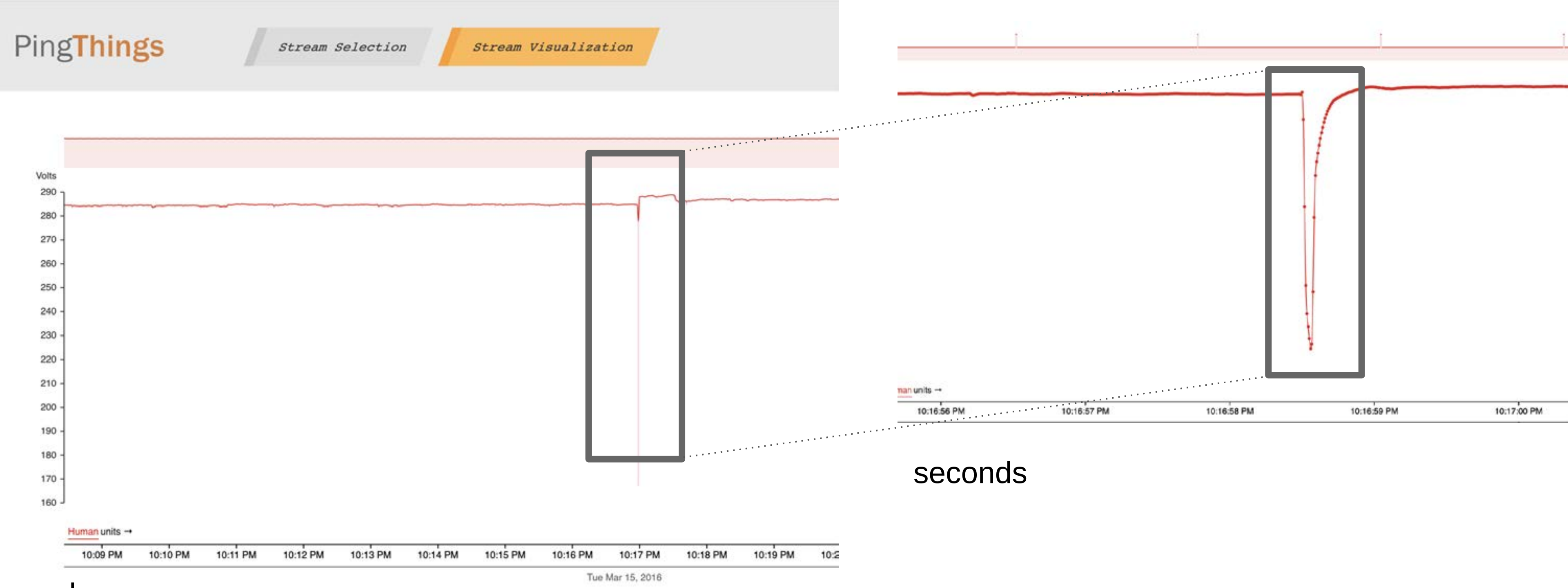
Step 1: Motivation & Questions

What are we interested in understanding? **Why** is this important?

- Voltage sags are brief, significant dips in system voltage.
- We want to understand the frequency & magnitude of sags.
- They can have a detrimental impact on sensitive equipment, causing inverters to trip offline, and can indicate safety issues such as faults.

Step 2: Explore the data

What do voltage sags look like? How can we automatically find them in the data?

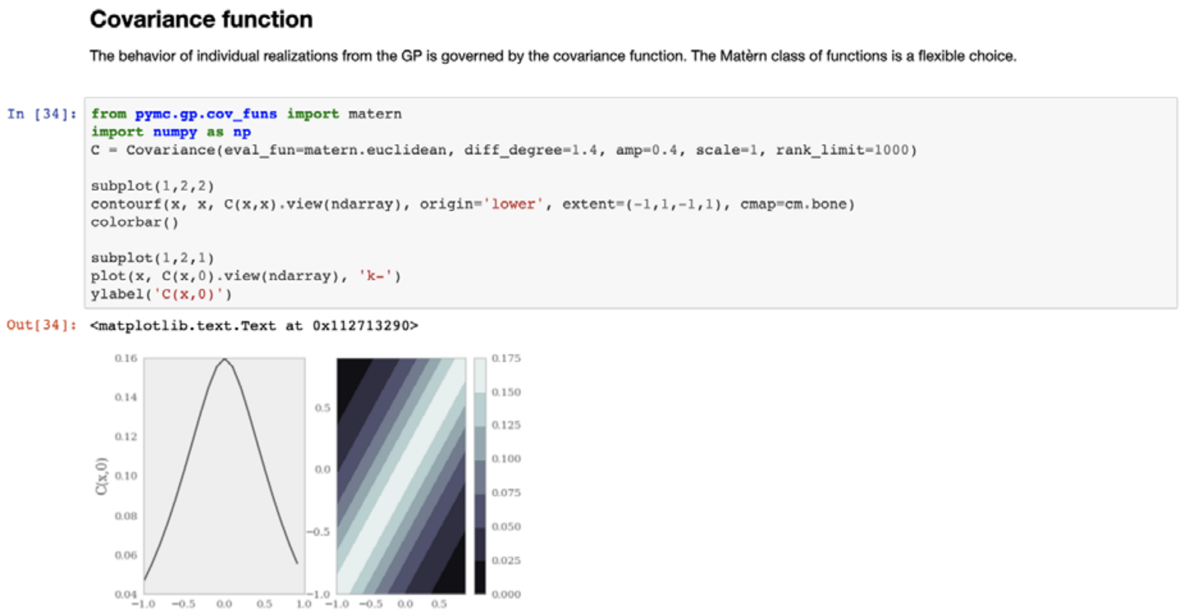


seconds

hours

Step 3: Automate sag detection

I can now write a script in a **Jupyter notebook** to automatically find voltage sags.



Covariance function

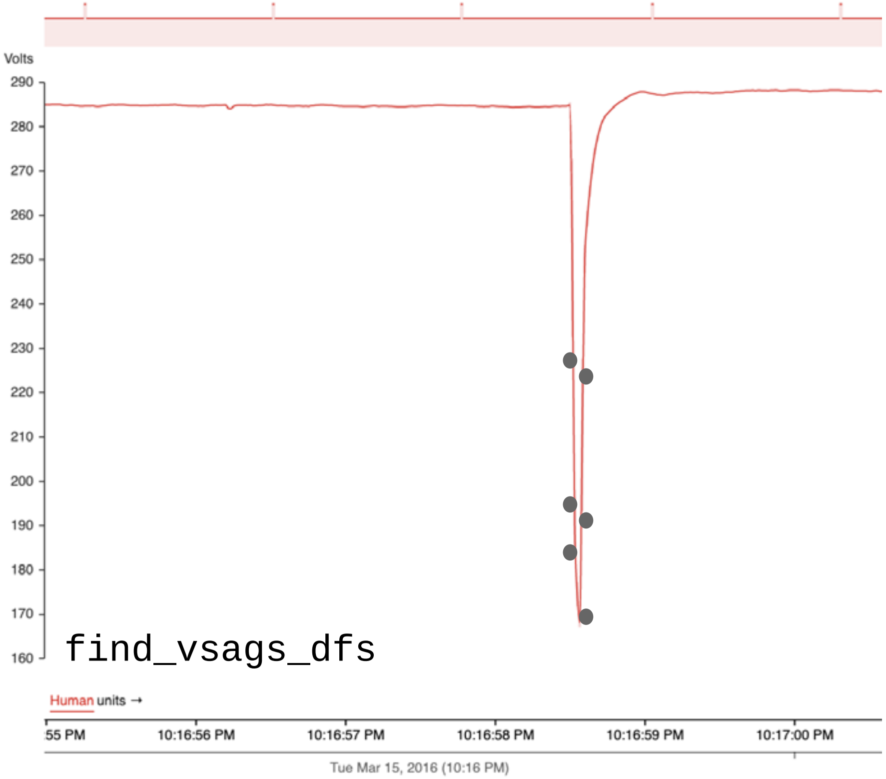
The behavior of individual realizations from the GP is governed by the covariance function. The Matèrn class of functions is a flexible choice.

```
In [34]: from pymc.gp.cov_funs import matern
         import numpy as np
         C = Covariance(eval_fun=matern.euclidean, diff_degree=1.4, amp=0.4, scale=1, rank_limit=1000)

         subplot(1,2,2)
         contourf(x, x, C(x,x).view(ndarray), origin='lower', extent=(-1,1,-1,1), cmap=cm.bone)
         colorbar()

         subplot(1,2,1)
         plot(x, C(x,0).view(ndarray), 'k-')
         ylabel('C(x,0)')

Out[34]: <matplotlib.text.Text at 0x112713290>
```

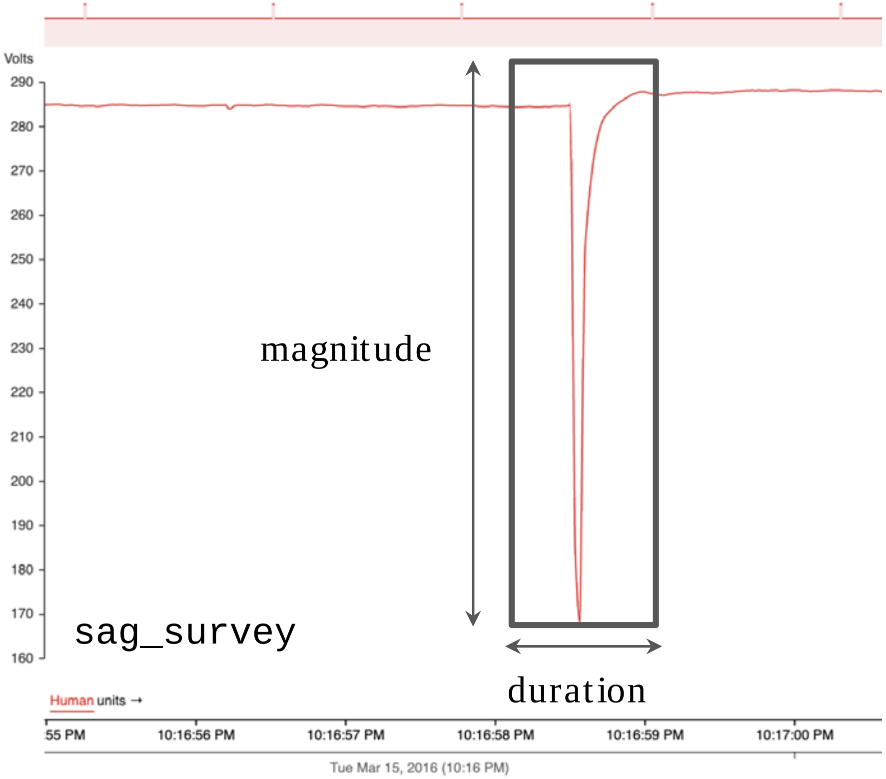
Jupyter notebooks allow for mixing of text, code, outputs including visualizations.

Enables exploration and sharing.

A sag is defined by voltage measurements below a threshold. I choose to split sag detection into two functions.


`find_vsags_dfs`

Returns all measurement points below threshold

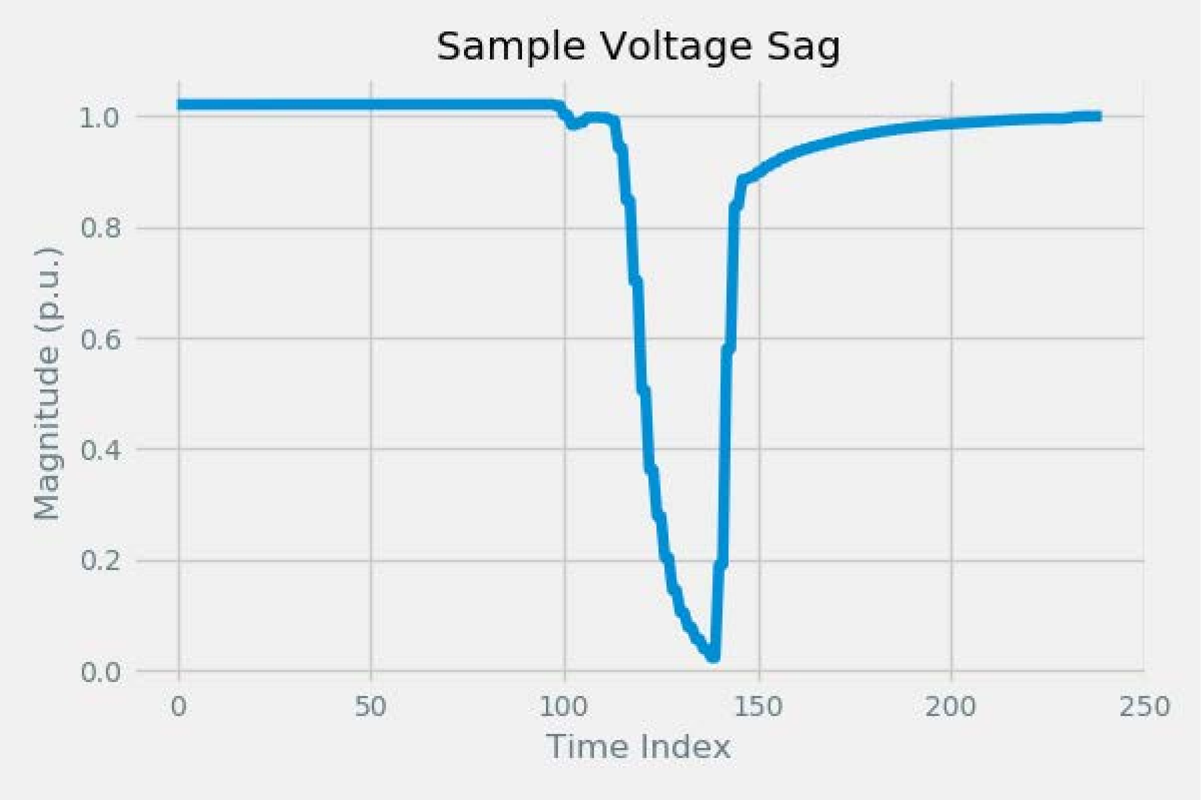

`sag_survey`

magnitude

duration

Consolidate points to get sag magnitude & duration

Now we can quickly find voltage sags and analyze their size and duration.

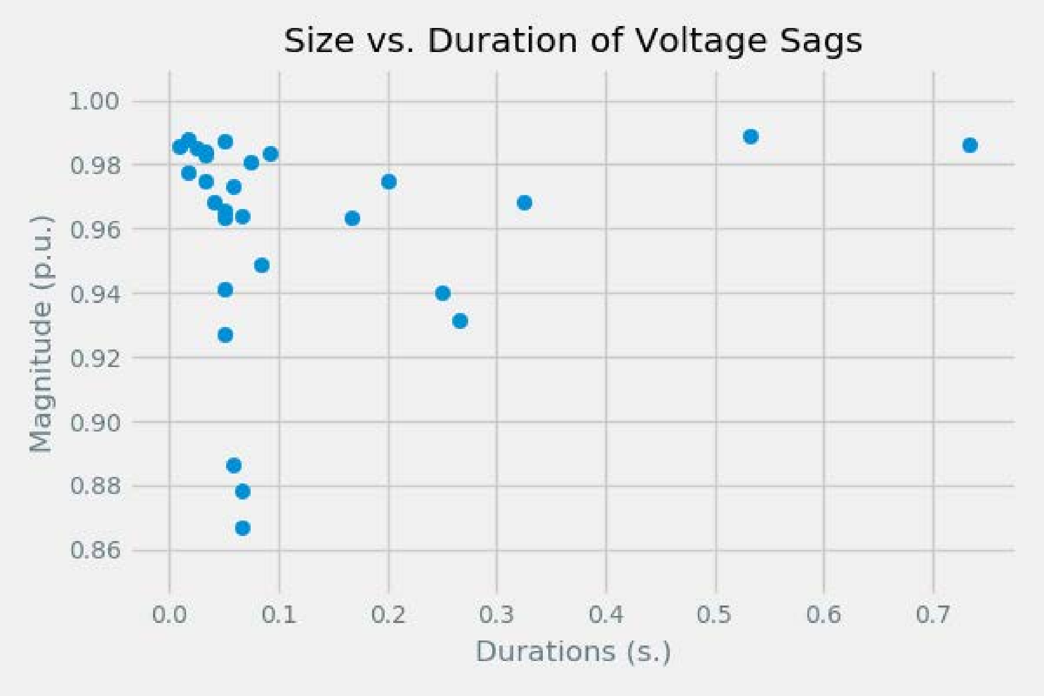


An example of a sag found by the algorithm.

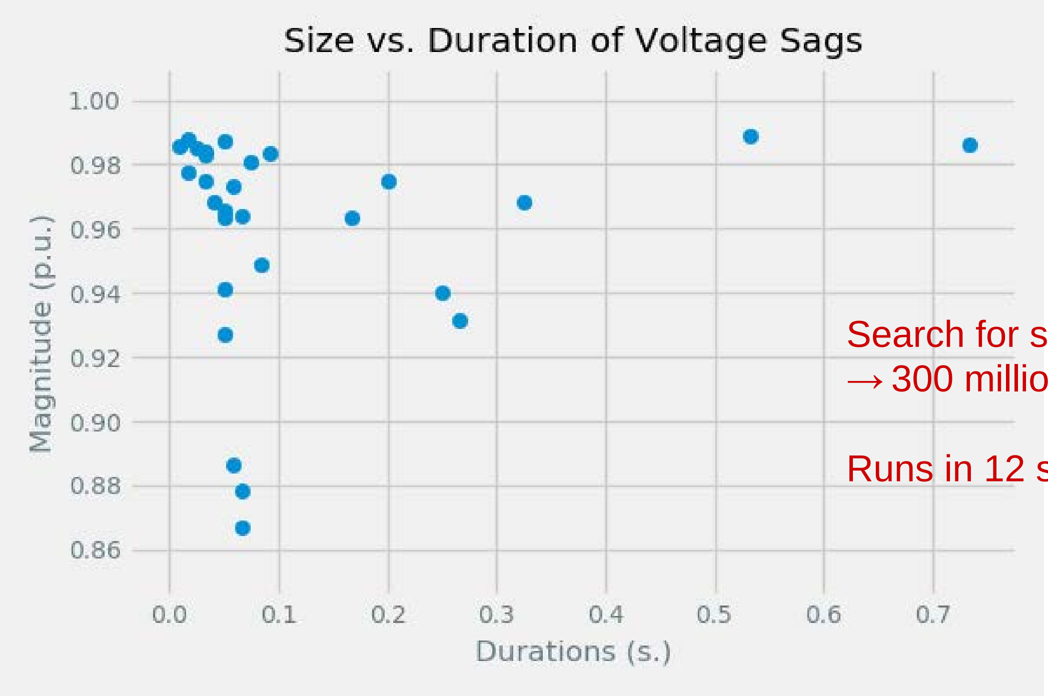
Step 4: Answer questions

- What are the magnitudes and durations of voltage sags at this location?
- How well does a model from the literature describe the frequency-vs-magnitude of the sags we observe?
- Do we see any impact of DG on sag magnitude?
- Are there weekly patterns in sag frequency?

- What are the magnitudes and durations of sags that occur at this measurement point?



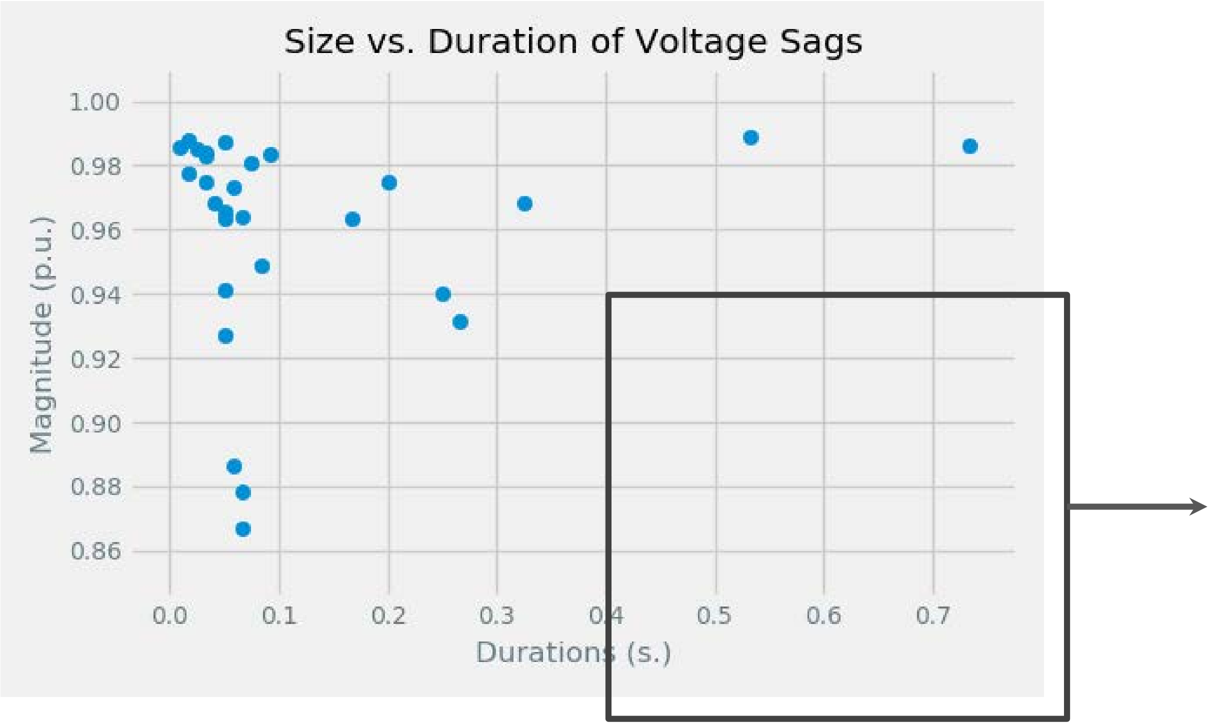
- What are the magnitudes and durations of sags that occur at this measurement point?



Search for sags across 1 month
→ 300 million data points.

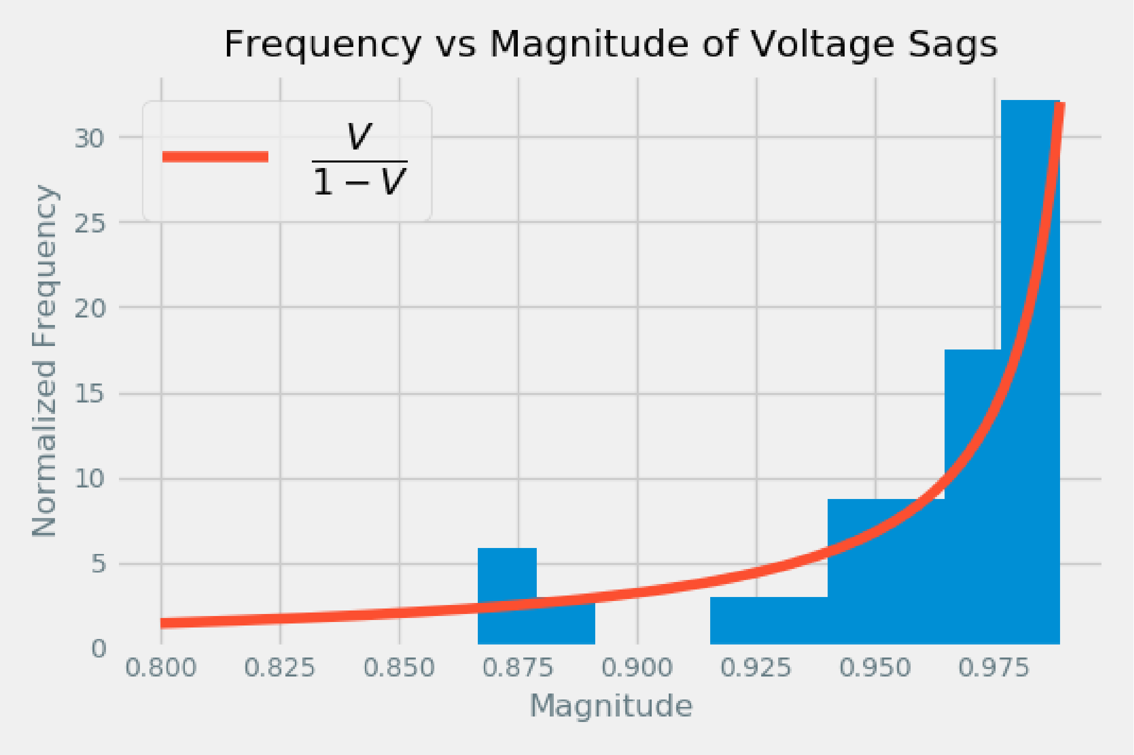
Runs in 12 seconds!

- What are the magnitudes and durations of sags that occur at this measurement point?



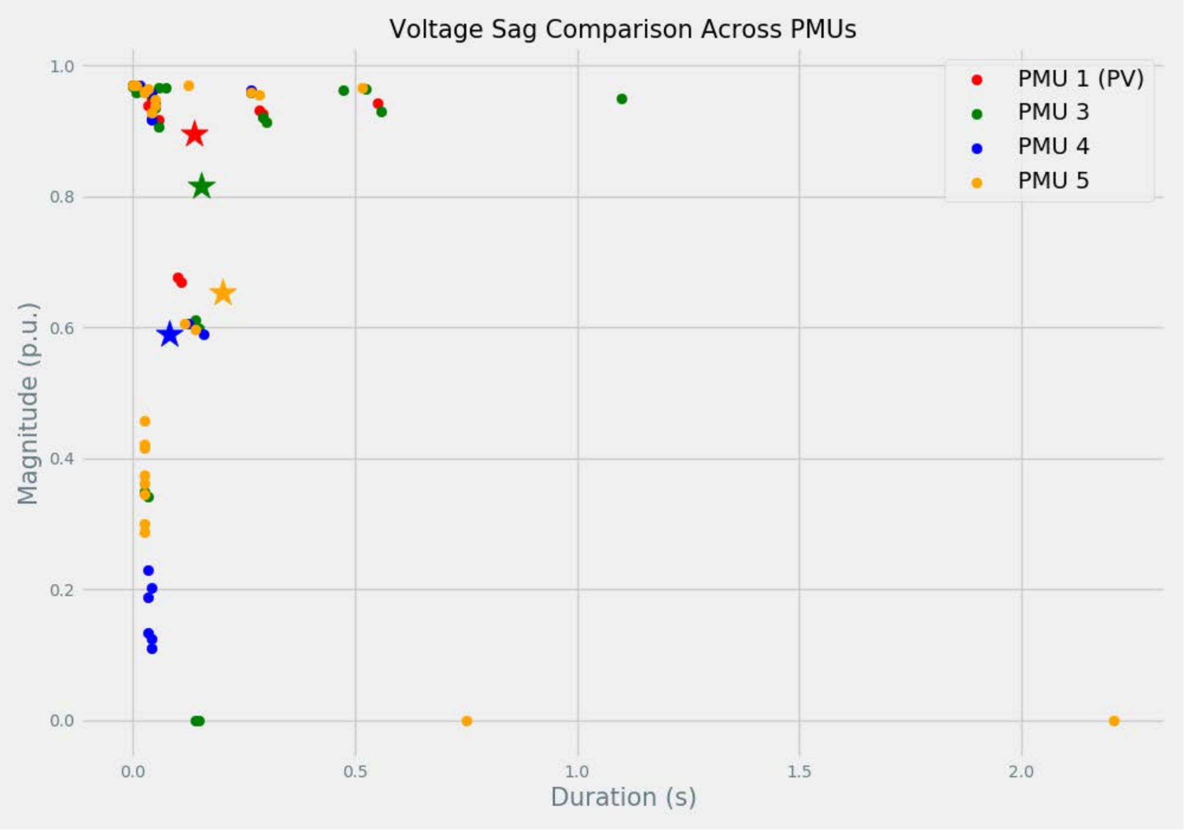
None in the danger zone!

- How well does a model from the literature describe the frequency-vs-magnitude of the sags we observe?



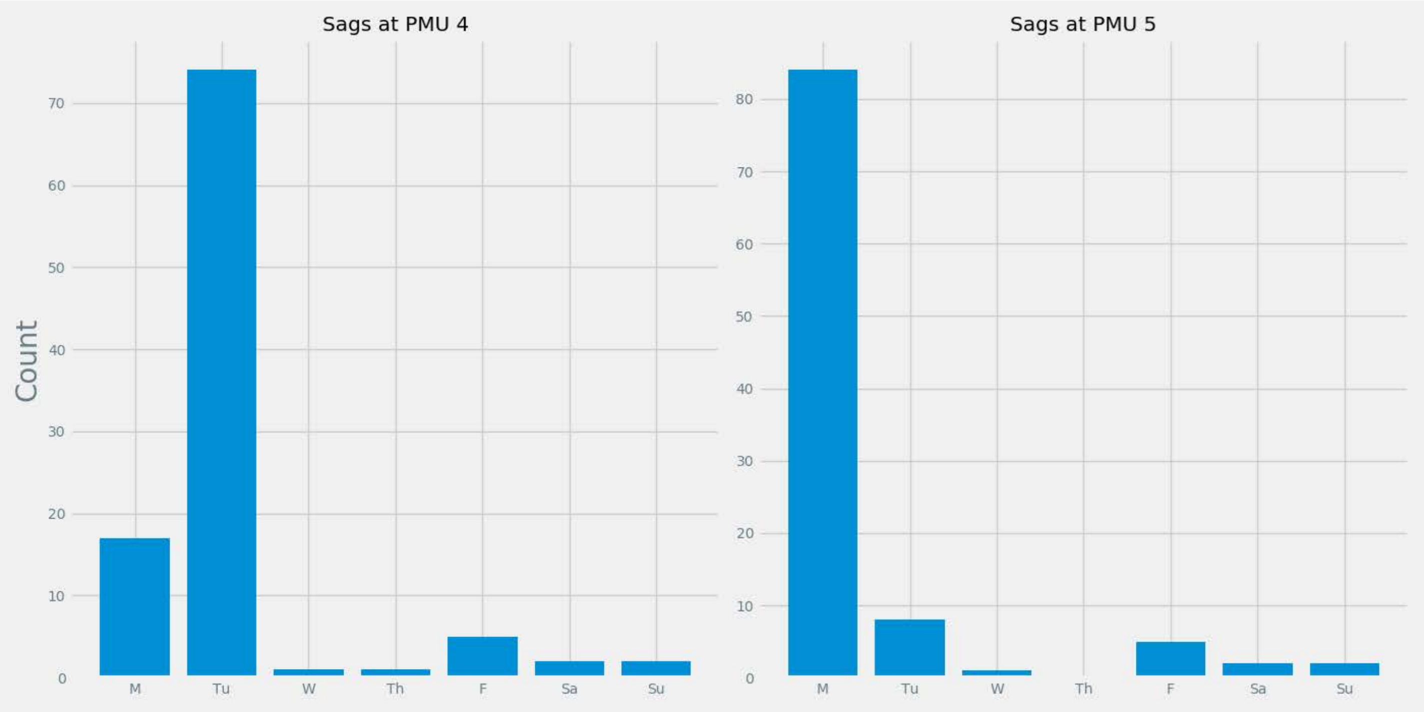
In the literature, a highly simplified model of fault-induced voltage sags predicts that the frequency of a sag with magnitude V will be proportional to V / (1-V)

- Do we see any impact of DG on sag magnitude?



We can compare sag magnitude & frequency at multiple PMUs including one with a PV injection.

- Are there weekly patterns in sag frequency?

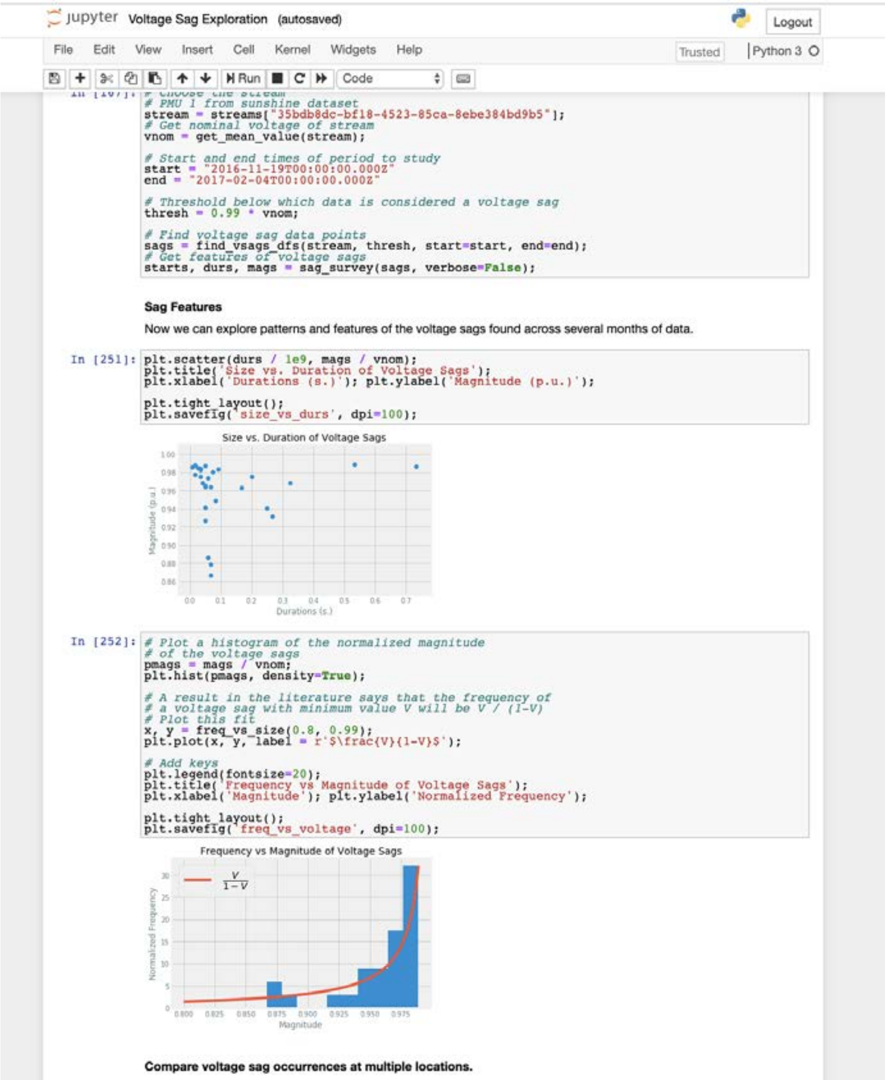


We can count sag occurrences per day to see if there are weekly patterns.

Step 5: Share Results

Creating use cases in a Jupyter notebooks allows for easy sharing of results.

- Visualizations
- Explanations
- Replicable code experiments



```
# PMU 1 from sunshine dataset
stream = streams["35bdb8dc-bf18-4523-85ca-8ebe384bd9b5"];
# Get nominal voltage of stream
vnom = get_mean_value(stream);

# Start and end times of period to study
start = "2016-11-19T00:00:00.000Z"
end = "2017-02-04T00:00:00.000Z"

# Threshold below which data is considered a voltage sag
thresh = 0.99 * vnom;

# Find voltage sag data points
sags = find_vsags_dfs(stream, thresh, start=start, end=end);
# Get features of voltage sags
starts, durs, mags = sag_survey(sags, verbose=False);
```

Sag Features

Now we can explore patterns and features of the voltage sags found across several months of data.

In [251]:
```
plt.scatter(durs / 1e9, mags / vnom);
plt.title('Size vs. Duration of Voltage Sags');
plt.xlabel('Durations (s.)'); plt.ylabel('Magnitude (p.u.)');

plt.tight_layout();
plt.savefig('size_vs_durs', dpi=100);
```

In [252]:
```
# Plot a histogram of the normalized magnitude
# of the voltage sags
pmags = mags / vnom;
plt.hist(pmags, density=True);

# A result in the literature says that the frequency of
# a voltage sag with minimum value V will be V / (1-V)
# Plot this fit
x, y = freq_vs_size(0.8, 0.99);
plt.plot(x, y, label = r'$\frac{V}{1-V}$');

# Add keys
plt.legend(fontsize=20);
plt.title('Frequency vs Magnitude of Voltage Sags');
plt.xlabel('Magnitude'); plt.ylabel('Normalized Frequency');

plt.tight_layout();
plt.savefig('freq_vs_voltage', dpi=100);
```

Compare voltage sag occurrences at multiple locations.

Mohini’s code is available here:

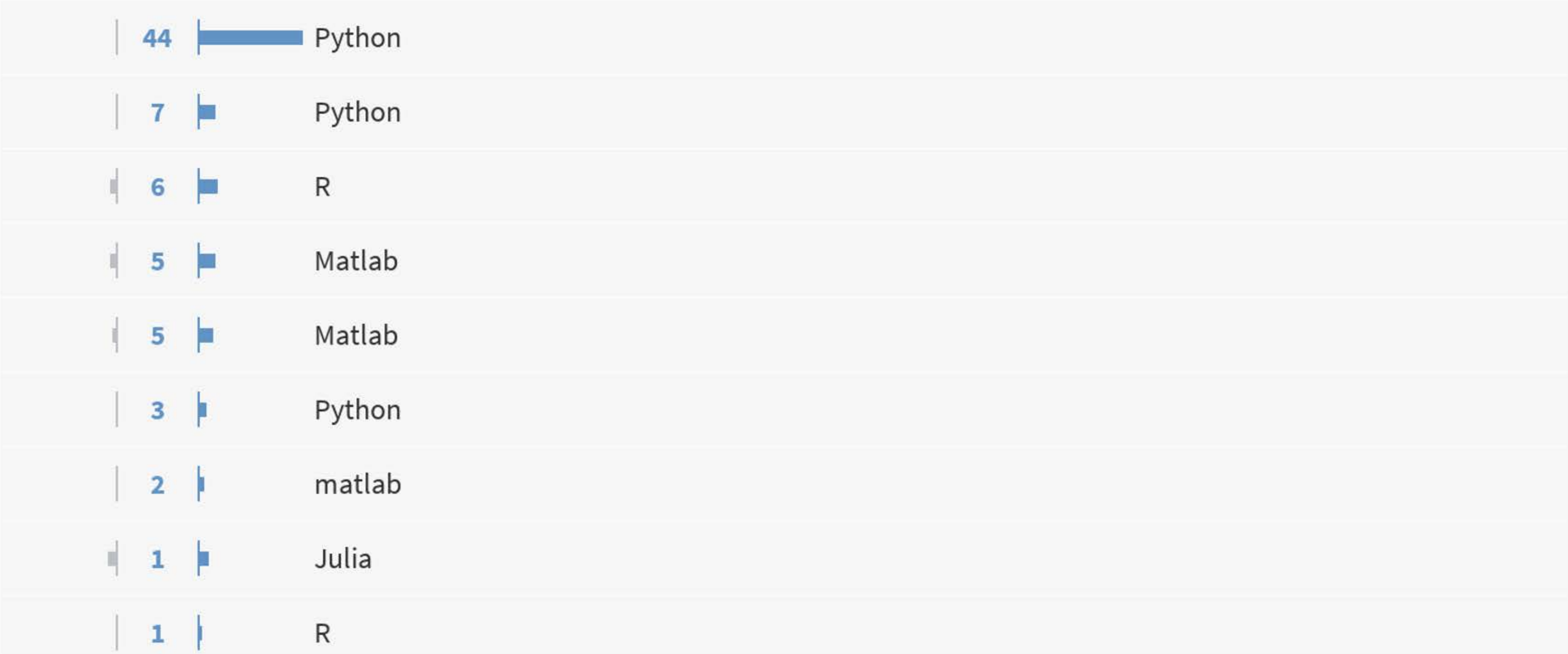
https://blog.ni4ai.org/post/2020-04-08-sunshine-data/

We will share code you can use to run similar analytics on your data. What analytics or engineering examples would you like to see illustrated?

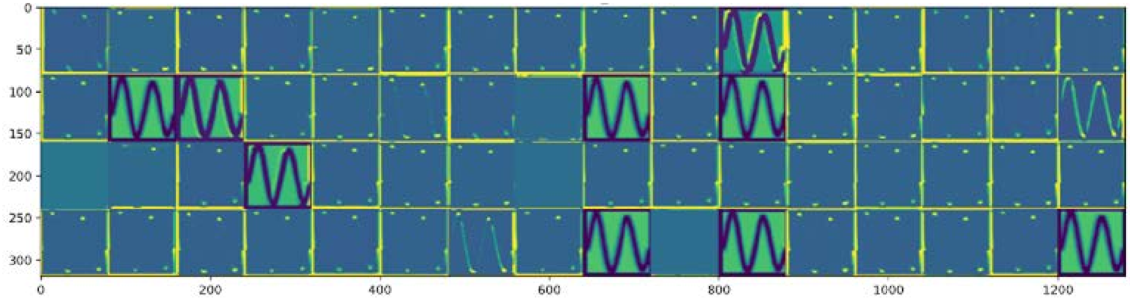
Top

What programming languages would you want us to support?

Top



44	Python
7	Python
6	R
5	Matlab
5	Matlab
3	Python
2	matlab
1	Julia
1	R



Synchrophasor Analytics using AI/ML

From inference at the edge to operator decision-support applications.

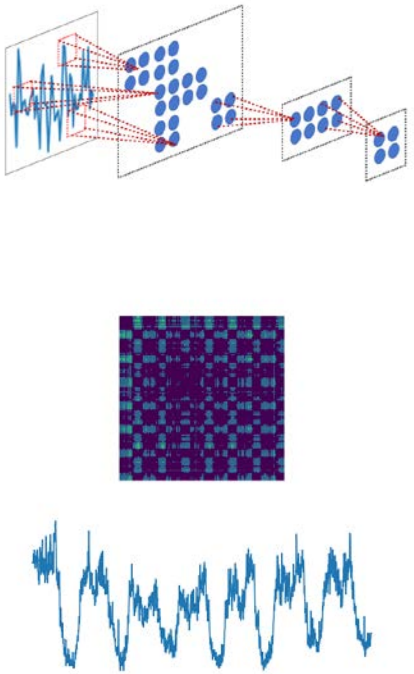
Luigi Vanfretti, Associate Prof.
RPI, ECSE
alsetlab.com,
luigi.vanfretti@gmail.com

Tetiana Bogodorova, Research Scientist
RPI, ECSE
alsetlab.com
tetiana.bogodorova@gmail.com

Overview

- This presentation is divided in two parts.

- **Part 1: (Luigi)**
  - Real-Time Forced Oscillation Detection at the Edge
  - Work presented at SASG 2019
  - Paper available at:
    - https://ecse.rpi.edu/~vanfrl/documents/publications/conference/2019/CP165_KA
      ST_ML_ForcedOscillations.pdf
- **Part 2: (Tetiana)**
  - Synthetic Data for ML Decision Making Applications
  - On-going work within DeepGrid project
  - Draft paper being prepared:
    - T. Bogodorova and L. Vanfretti, “Algorithms for Synthetic Data Generation for Power System Dynamics Simulation-Based Machine Learning Decision-Making Applications”.



Acknowledgement

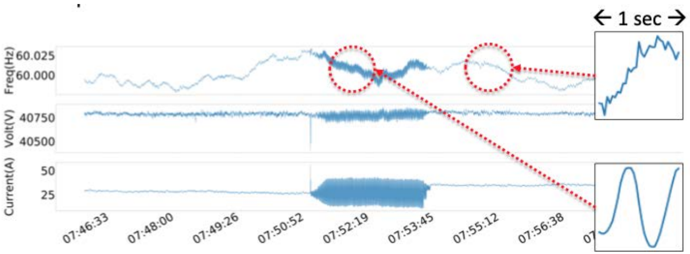
- The presenters would like the acknowledge the contributions of their colleagues:
  - ***KAUST:*** Mohammed-Ilies Ayachi and Prof. Shehab Ahmed
  - ***RPI:*** Sergio Dorado

  which are included in this presentation.

- **The work in the 1st part of this presentation is supported by**
  - The GridX project funded by the Center of Excellence for NEOM Research at the King Abdullah University of Science and Technology, Saudi Arabia.
- **The work in the 2nd part of this presentation is supported by:**
  - The DeepGrid project funded by the New York Research and Energy Agency (NYSERDA) and New York Power Authority (NYPA) through the Electric Power Transmission and Distribution (EPTD) High Performing Grid Program Program Opportunity Notice (PON) 3770, under Grant Agreement no. 137951.
- The funding support from the funding bodies above are gratefully acknowledged.

Part 1: Real-Time Force Oscillation Detection at the Edge (GridX Project)

- ***Problem:***
  Identify forced oscillations, (likely) product of sub-synchronous control interactions between WF and grid, and use it to take action (e.g. tripping/ramp-down) to stop it (not covered here)

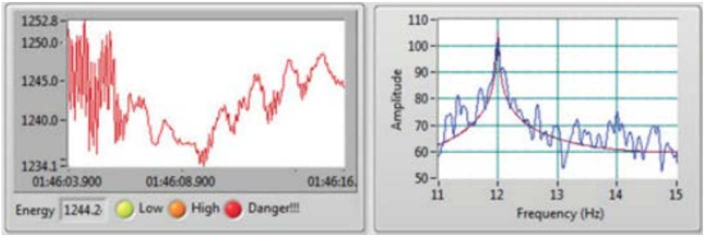


- ***Previous Work:***

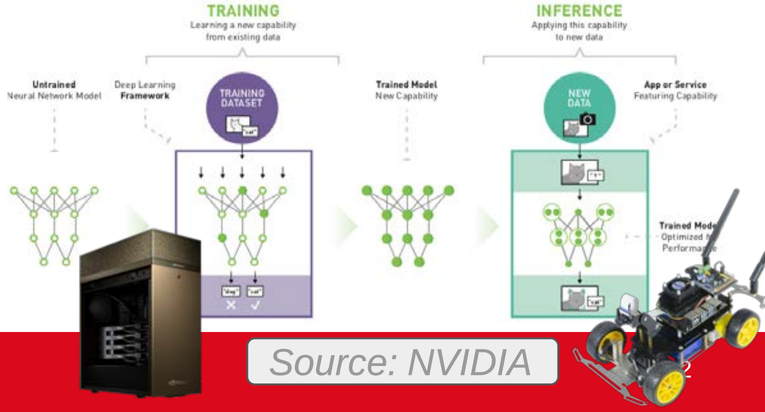
  Signal Processing-Based Methods (e.g. energy-based detection, link):

  Pros: experience with mode estimation → confidence from users has been gained.

  Cons: inherited delays from filters, practically centralized - only operator decisions (slow), needs carefully parametrization/tuning from experts



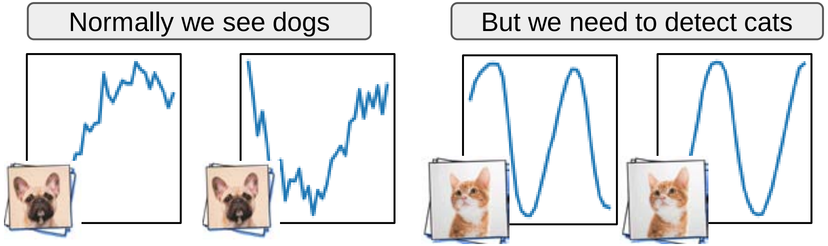
- ***Can AI/ML help meet evolving requirements/needs?***
  - Increase Detection Speed while maintaining (acceptable) accuracy
  - Train using:
    - Few (small) sets of ***useful*** data or no useful data at all!
    - Use simulation results for training
  - Possible to deploy in the field on low cost embedded system.

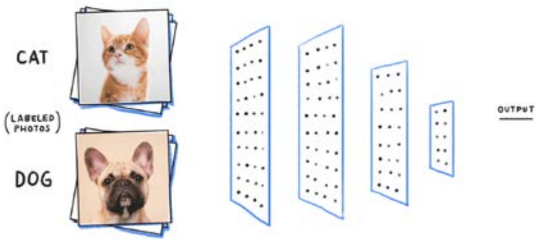


Source: NVIDIA

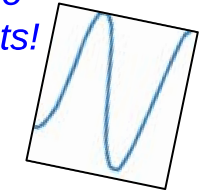
Detecting FO's through Classification

- ### *Detection and Classification*
  - We know what we are looking for! It's cats!
  - Event though our dog and cat pictures look different:

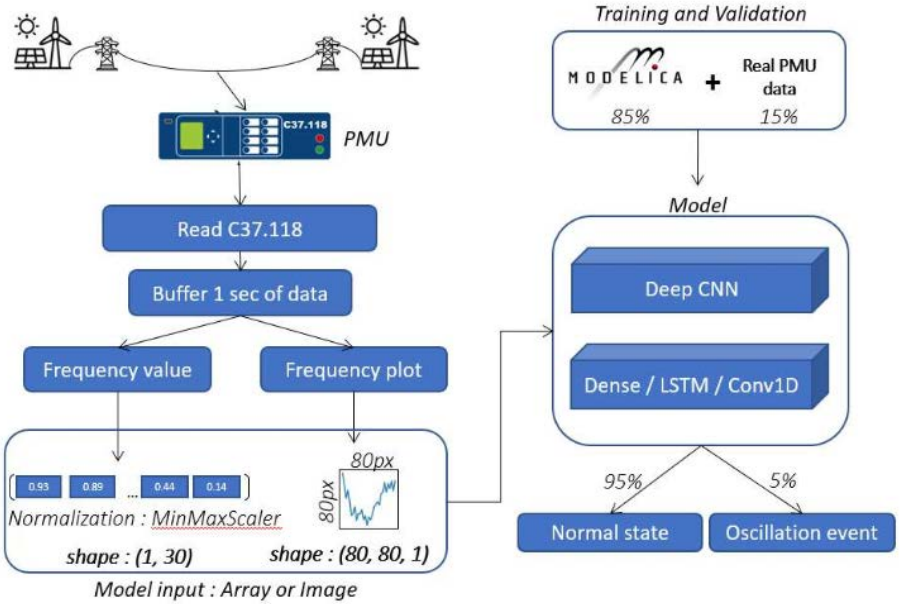




We can apply a similar approach to find our cats!



- ### *Proposed Approach*
  - Training: use ***labeled*** data (meas. and/or simulation)
  - Inference: Real-Time Detection every 1 sec. as a new input is passed through the trained ML/AI model.



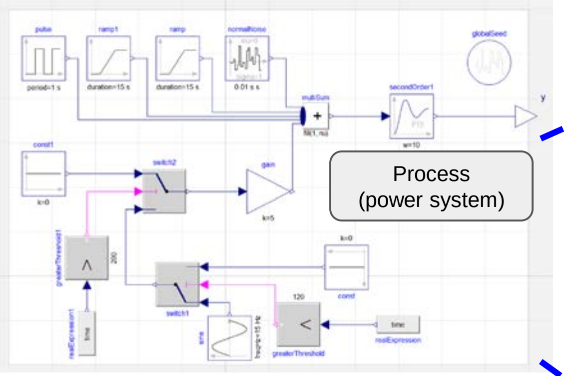
Labeling and Transfer Learning

- ***Labeling data***
  - PMU data is available: label as normal (🐶) or oscillation event (🐱) → was done manually, time consuming!
  - NO PMU available: generate it from a model (use Modelica/Dymola, 😊)
  - If you a have a few measurements, label them and combine them with simulation results.

- ***Transfer Learning:** Generating synthetic data*

This gives a library of 🐶s & 🐱s!

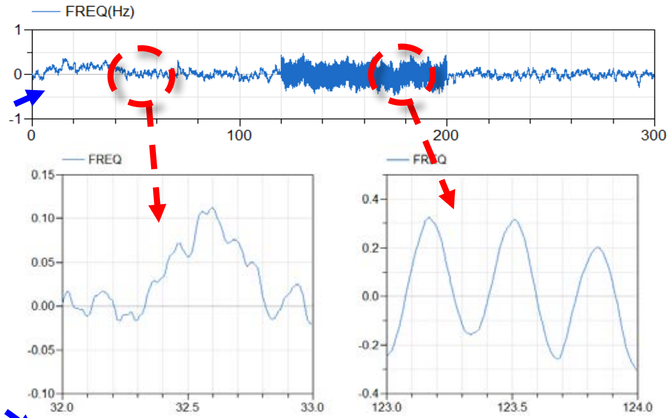
Model



Process
(power system)

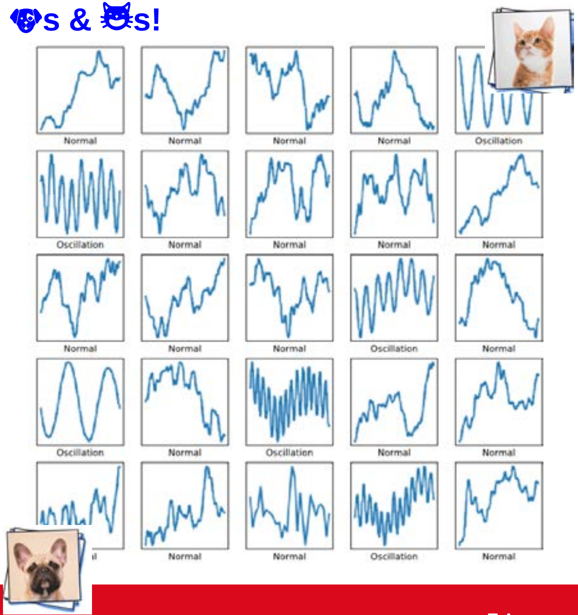
The bottom part is used to configure when the forced oscillation is injected, its amplitude and frequency!
Allows to automate labeling!

Simulation



Labeled - Normal

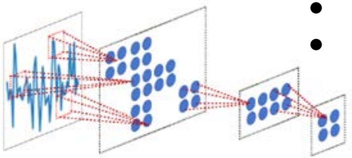
Labeled - Oscillation



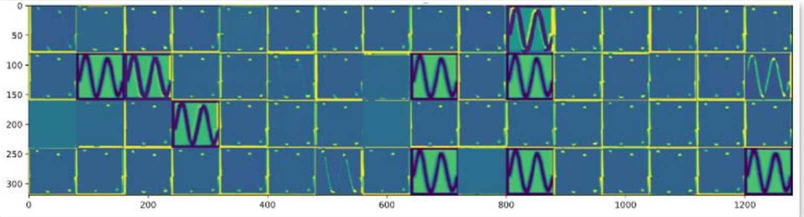
Training Neural Networks

- ***Training***
  - ○ Convolutional Neural Networks (CNN)

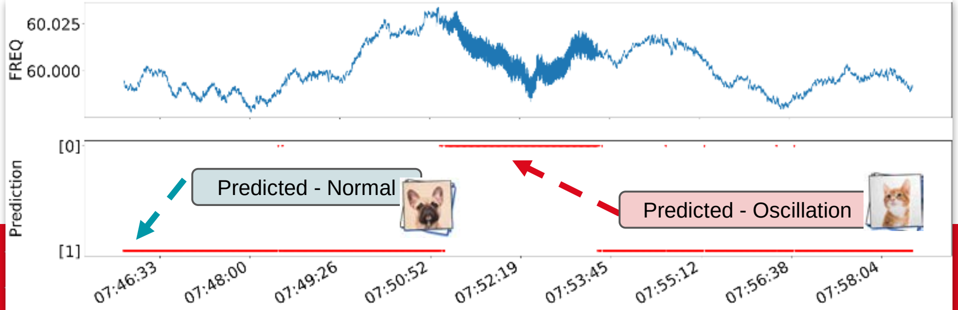
    Are classifiers that are known to have outstanding performances in the field of pattern recognition
  - ○ ***CNN patterns are learned through translation:***
    - ○ After learning a shape, a convnet can recognize it anywhere.
      - The first convolution layer learns small local patterns
      - A second convolution layer will learn larger patterns made of the features of the first layers, and so on.



Proposed CNN: To illustrate how an input decomposes into the different filters learned by the network, we have shown a feature map of our second convolution layer



- ***Trained CNNs***
  - ○ **Proposed 1-D CNN:** Conv1D model is composed of two Conv1D of 64 channels, one MaxPooling1D followed by a Dropout, then one Dense layer of 100 channels.
  - ○ **Proposed 2-D CNN:** similar to AlexNet's architectures with less depth.



Performance Assessment - Proposed and Existing ML Models

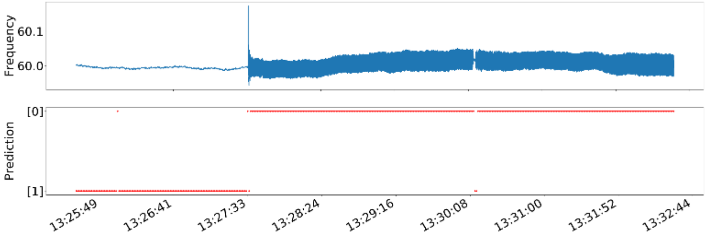
Model	Accuracy	False-positive	Missed event	Time for 1 prediction (sec)
Proposed CNN	97.41%	2	6	0.0047
Proposed Conv1D	98.06%	0	6	0.0027
MobileNet	97.74%	2	5	0.0074
MobileNet^2	98.71%	0	4	0.0074
AlexNet	94.51%	12	5	0.0098
ResNet-50	97.42%	4	4	0.0174
Dense	94.19%	6	12	0.0026
Stacked LSTM	94.19%	2	16	0.0054

Inference, Transfer Learning and ML Model Performance

- The transfer learning approach shows better results than using synthetic data only.
  - The simulation data consists of 11964 files for the training and 2991 files for validation (25% ratio).
  - The real data contains 967 samples (~15% of the synthetic dataset).
- The average prediction accuracy measured on 73 terminals (transmission lines):

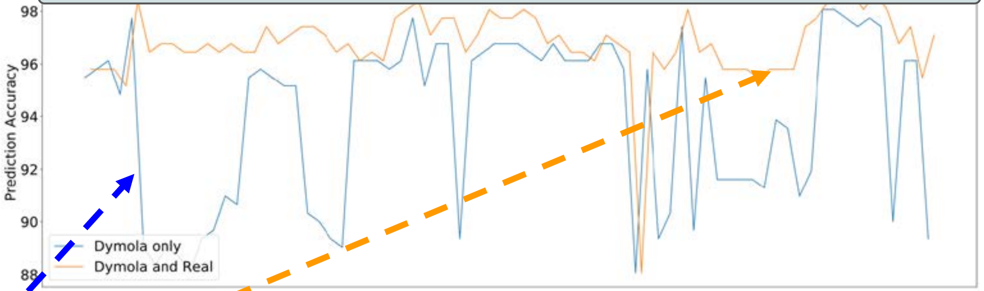
😑 Model trained with simulation data only is 93.94%

😁 Model trained on simulation + real data, it is 96.79%.



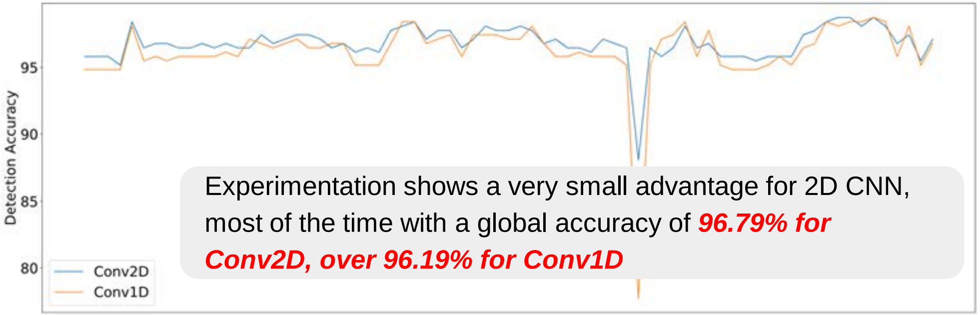
CNN Model trained on Modelica/Dymola's data only predicting with a sliding window of 1 seconds.

Transfer Learning



Each point on the x-axis is a different terminal (transmission line substation end).

Performance of 2D CNN vs 1D Conv1D Models



Experimentation shows a very small advantage for 2D CNN, most of the time with a global accuracy of ***96.79% for Conv2D, over 96.19% for Conv1D***

Each point on the x-axis is a different terminal (transmission line substation end).
2 minutes of normal events then 2 minutes of oscillation events)

Inference at the Edge

- The proposed CNN architecture shows a global accuracy of 97.41% and the proposed Conv1D shows an accuracy of 98.06% .
- The time to make a prediction is the average based on 1000 predictions made on different hardware.


Real-Time Forced Oscillations Detection from PMU Data
Using Machine Learning, deployed on Nvidia Jetson Xavier

Hardware	Cost	Time for 1 prediction with CNN	Time for 1 prediction with Conv1D
Windows PC Core i7 8700 – Nvidia 1080Ti	$2,900.00	0.0049 sec	0.0022 sec
Nvidia Jetson Xavier	$699.00	0.0357 sec	0.0170 sec
Raspberry Pi 3 (+sd card and accessories)	< $100.00	0.4698 sec	0.0114 sec

Average inference time based on 1000 predictions.

Part 1: Real-Time Force Oscillation Detection at the Edge (GridX Project)

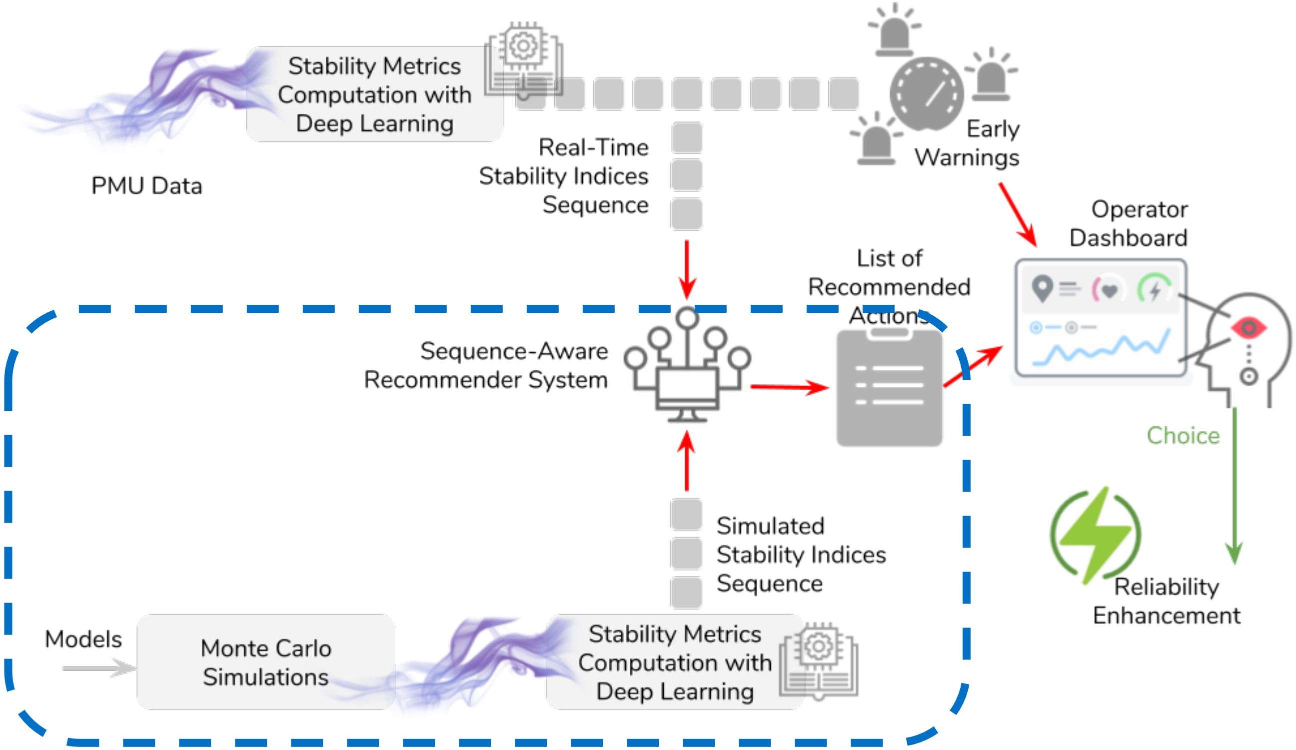
Key Takeaways

- **Real-Time F.O. Detection using ML:** Fast and accurate, can be deployed on the edge, but *requires good data* for training.
- **Not all data is created equal:** Unless the data has the behavior we want the AI/ML to recognize, even if you have Terabytes, it is of relatively of low value.
- **Labeling:** It is very time consuming and requires expert knowledge!
  - It is not easy to tell apart from a cat and a dog when you have only seen very few!
  - Common, curated data sets that can be shared are key for development.

- **ML Models/Methods:**
  - Wealth of developments needs to be carefully adapted for power applications.
  - No need to reinvent the wheel → using common tools enables portability and reuse; use TensorFlow, Keras, Pythorch!
- **Transfer Learning**
  - One can get very far using simulations data for training
  - But, accuracy improvements require ***real-world*** curated data.

- ***Inference at the Edge:***
  - *Low-cost platforms have great potential for automation: protection and control based on AI/ML.*
  - ***But the HW infrastructure to "train at scale" is cost-prohibitive → need of shared platform.***

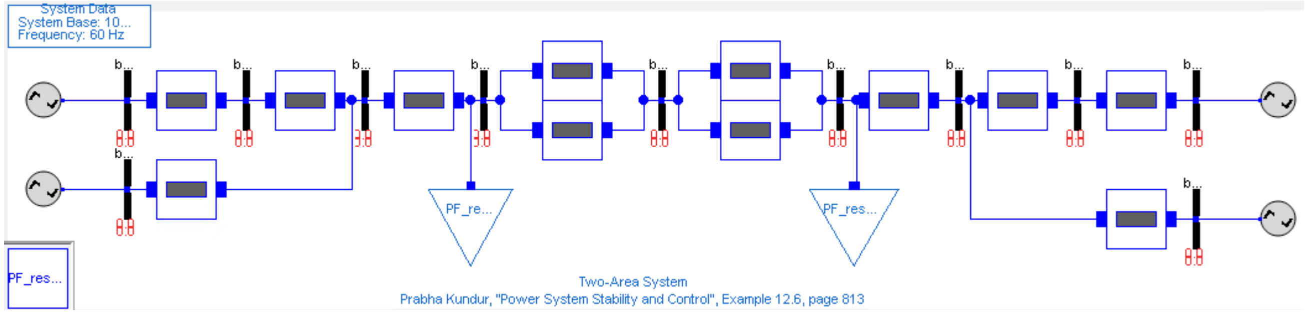
Part 2: DeepGrid Project (Overview)

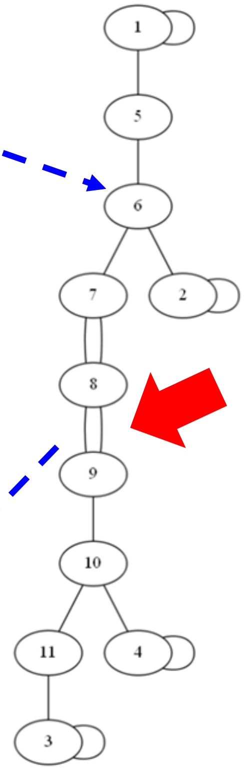


Focus of the remainder of this presentation

Realistic Contingency Selection for Synthetic Data Generation

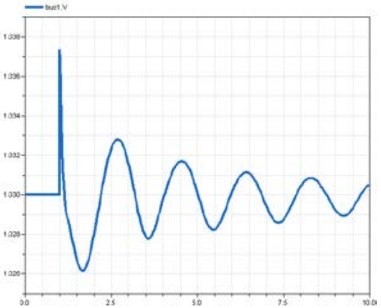
- This research proposes an algorithm for big data generation based on realistic selection of a set of contingencies for a system described by *an undirected graph*.



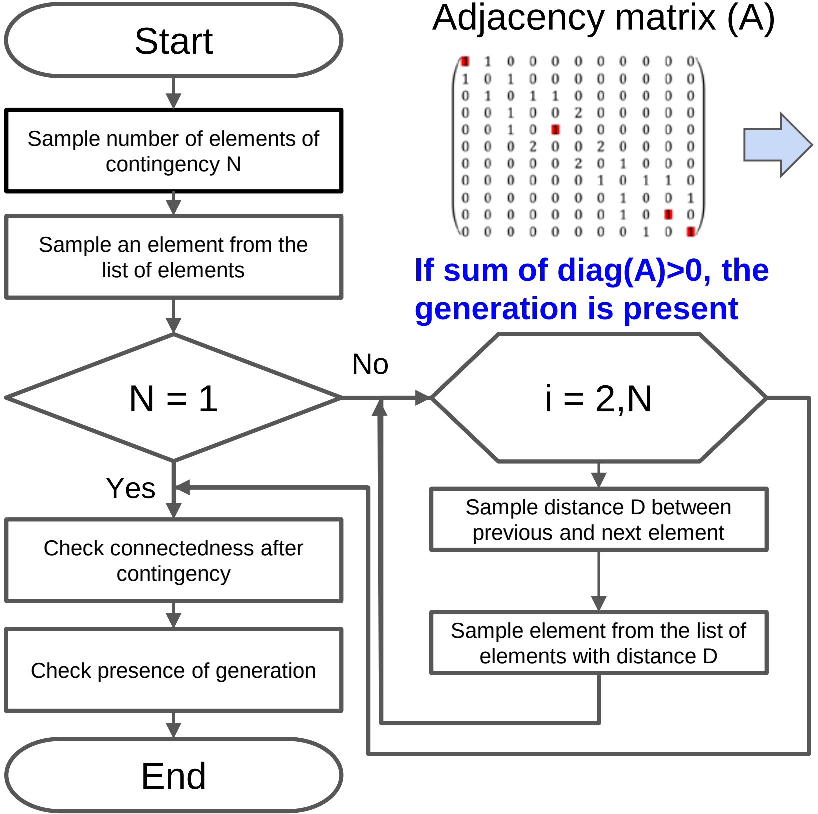


- The graph representation allows to detect change of topology of the system after a contingency.

- ***The idea*** is to ***generate trajectories of power system signals*** (magnitudes of I and V, angles of I and V) ***after a contingency***, but only for those cases where connectedness of the network is preserved.



Algorithm of Big Data Generation

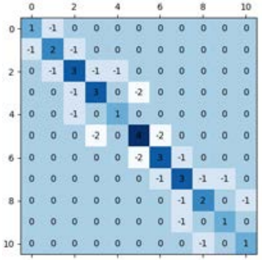


Adjacency matrix (A)

$$\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}$$

If sum of diag(A)>0, the generation is present

Laplacian matrix (L)



	0		2		4		6		8		10
0	1	-1	0	0	0	0	0	0	0	0	0
	-1	2	-1	0	0	0	0	0	0	0	0
2	0	-1	3	-1	-1	0	0	0	0	0	0
	0	0	-1	3	0	-2	0	0	0	0	0
4	0	0	-1	0	1	0	0	0	0	0	0
	0	0	0	-2	0	4	-2	0	0	0	0
6	0	0	0	0	0	-2	3	-1	0	0	0
	0	0	0	0	0	0	-1	3	-1	-1	0
8	0	0	0	0	0	0	0	-1	2	0	-1
	0	0	0	0	0	0	0	-1	0	1	0
10	0	0	0	0	0	0	0	0	-1	0	1

Eigenvalues of L

[6.51 4.45 3.96 2.80 2.25
2.00 **-5.01e-16** 1.28e-01 4.65e-
01 8.17e-01 6.30e-01]

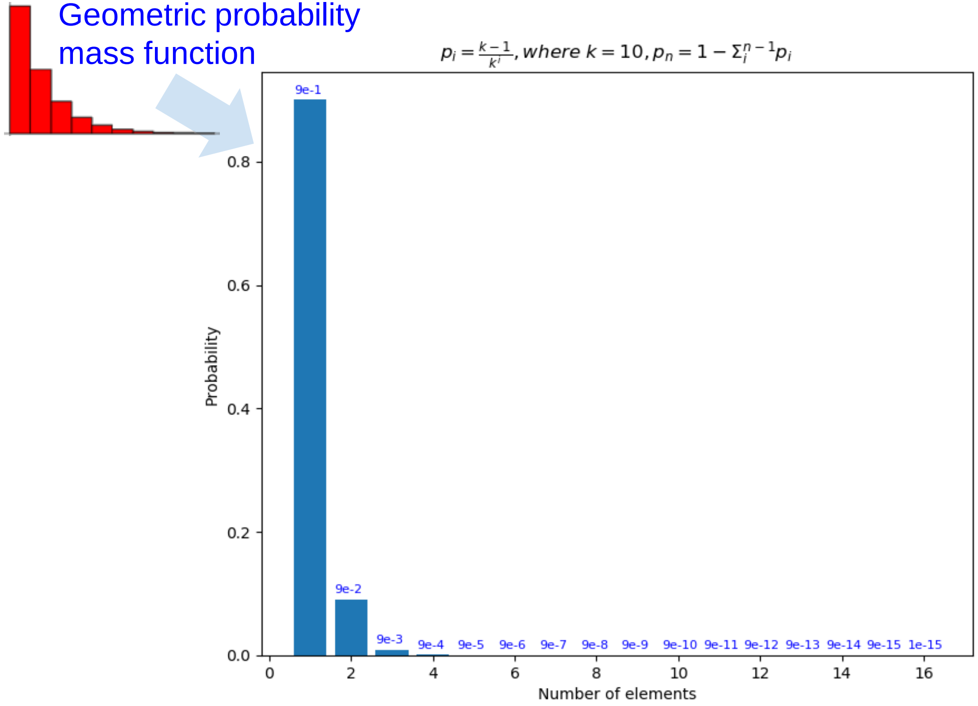
Number of eig(L)≅0 defines number of islands

- **Contingency - an outage of one or more elements** such as generators, transformers and lines that are presented as edges of undirected graph.
- After contingency **at least one generator has to be connected** to the network.
- After contingency **the cases of no islanding are of interest**

Algorithm = Sampling from pmf + graph theory to check validity

The ideas behind the algorithm: Sampling probability

Geometric probability mass function



1. Number of elements that are tripped follows a geometric distribution (GD) function that is *identified from historical data* on frequency of outages depending on number of elements [1].

1. *A distance between elements of a multi-event contingency follows* a geometric distribution function.

   GD is parameterized by identifying a ratio between probabilities of two consecutive values of distances between element of a contingency.

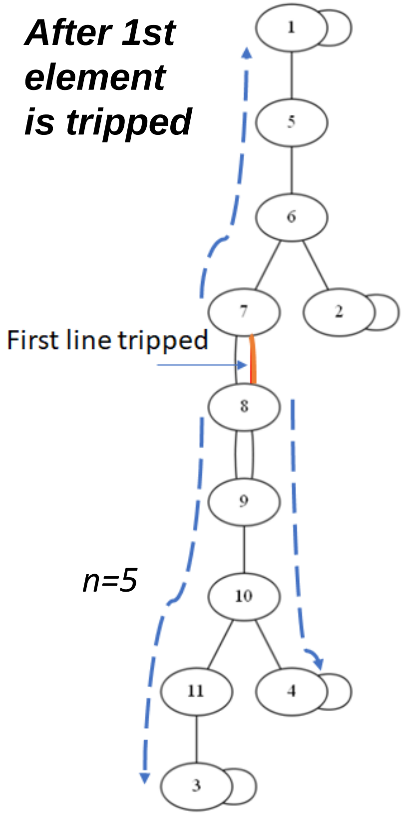
   This distribution function is *derived from a power flow change ratio as a function of a distance after contingency in [2].*

[1] The Western Electricity Coordinating Council (WECC), “Phase I Probabilistic Based Reliability Criteria Evaluation of Exceptions List Facilities,”Reliability Performance Evaluation Work Group, Tech. Rep., February 2001.

[2] S. Soltan, D. Mazauric, and G. Zussman, “Cascading failures in power grids: analysis and algorithms,” in Proceedings of the 5th international conference on Future energy systems, 2014, pp. 195–206.

Example of a contingency selection and probability calculation

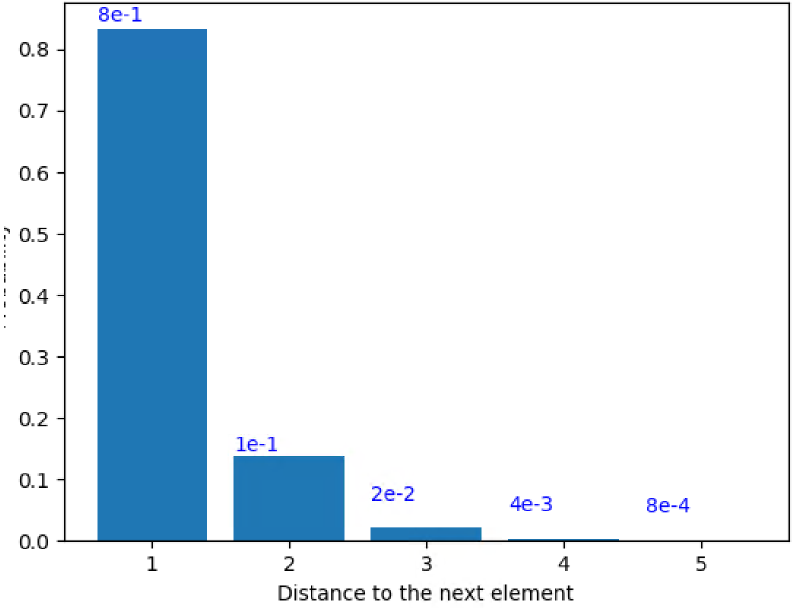
After 1st element is tripped

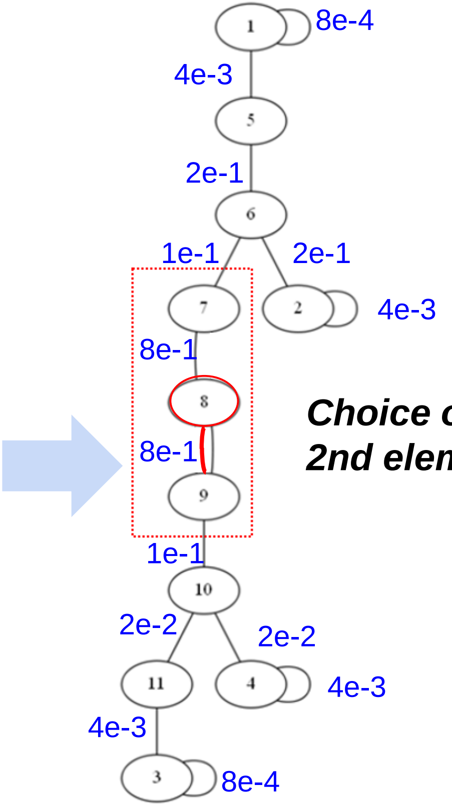


Probability how far the next element is located w.r.t. the 1st element

$$p_i = \frac{k-1}{k^i}, p_n = 1 - \Sigma_i^{n-1} p_i,$$

$$where\ k = 6, n = \ max.\ path\ from\ the\ tripped\ element$$

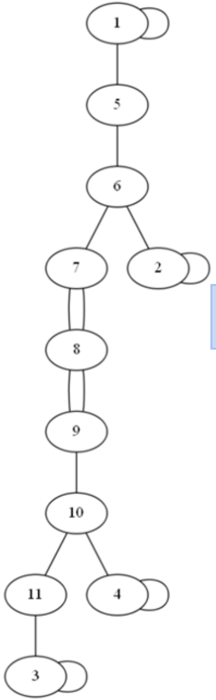




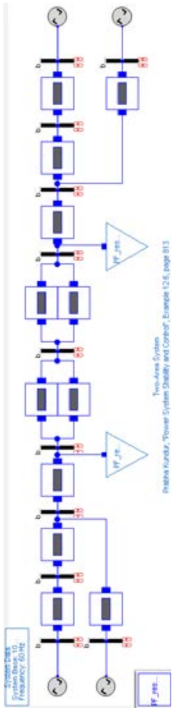
Choice of 2nd element

Data labeling for training ML model

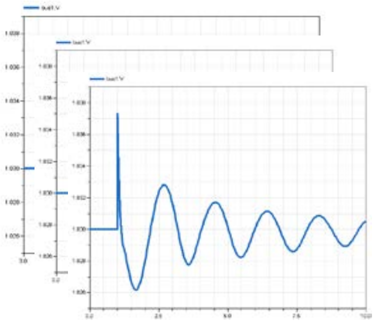
Choice from the list of scenarios:



Feed the scenarios to simulate in Dymola:



Label trajectories with SSS index:

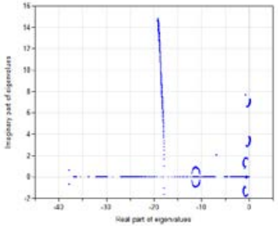


Labels:

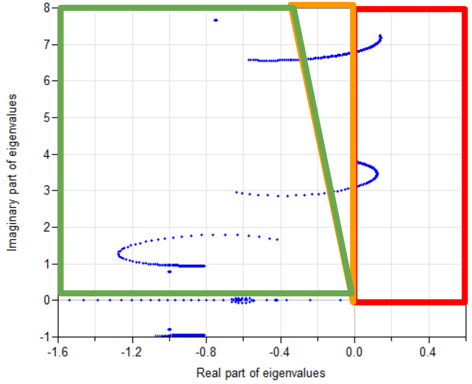
- Unstable ($\zeta < 0$)
- Marginally stable ($0 \leq \zeta < 0.05$)
- Stable ($\zeta \geq 0.05$)

SSS Index:

- Linearization after contingency is applied
- Calculate eigenvalues of A matrix
- Calculate damping



$$\zeta = \frac{-Re(eig)}{\sqrt{Re^2(eig)+Im^2(eig)}}$$



Part 2: Synthetic Data Generation

- **Key Takeaways:**
  - **The most time consuming** ML pipeline is **data extraction, generation** and **data preprocessing** (arguably takes 50-80%)
  - Be careful when generating data: ***What you give to the model to learn defines the output***
  - **Use real-data** directly or indirectly to design realistic data generation
  - **Tools are important** - define flexibility (Modelica gives full access to adjust parameters) and computational time (Multi-core computing for simulation, GPU computing for ML and Python - faster than Matlab and open source)

Result for 3.2 GHz 16-thread AMD Ryzen Processor

Cases to Simulate (Modelica/Dymola from Python)	Time (sec/min/**h**)
5	~ 86/1.43
100	~ 1178/19.63
9300	~ 63501/1058.35/**17.64**

We want to host content to make domain expertise accessible to a broader community. What would you like to learn more about?

Top

28	Grid applications and use cases
16	Real time Protection application for Grid based on Synchrophasor data.
14	Machine learning and AI tutorials
7	Success stories
7	Research highlights
5	Grid data & AI wiki
2	Creating containers to deploy applications on your platform
1	Yes

Utility Perspective

Kevin Jones, Dominion Energy

Three Important Questions

Why should a utility[†] care about NI4AI?

Why should a utility[†] participate in NI4AI?

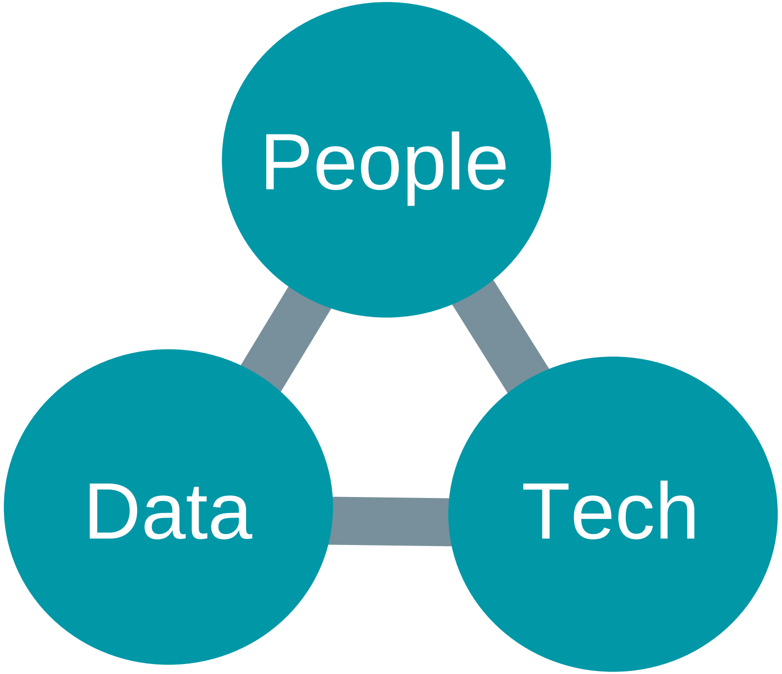
Why is Dominion participating in NI4AI?

[†] TO, RTO, ISO, etc.

Why Should a Utility Care About NI4AI?

NI4AI will answer key questions for utilities by solving the hardest problems...

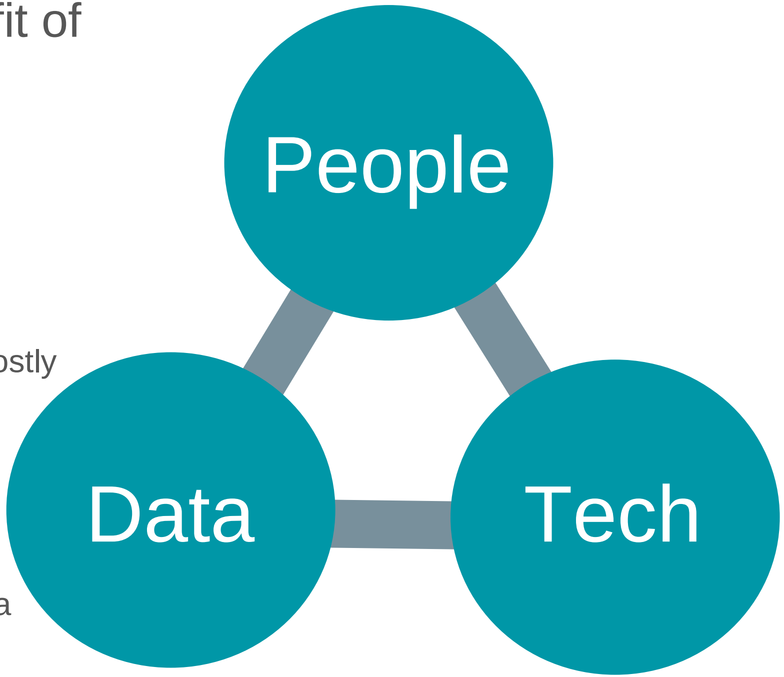
- Who to hire? Who to work with?
    - Employee recruitment
    - Vendor selection
- What to buy? - Tech proving ground for:
    - Sensors
    - Data platform & features
    - Use case & their implementations
- What to do?
    - Community driven use case selection
    - Ripe with content for business justification



Why Should a Utility Participate in NI4AI?

Use the full force of the community to the benefit of your organization with no additional W-2s

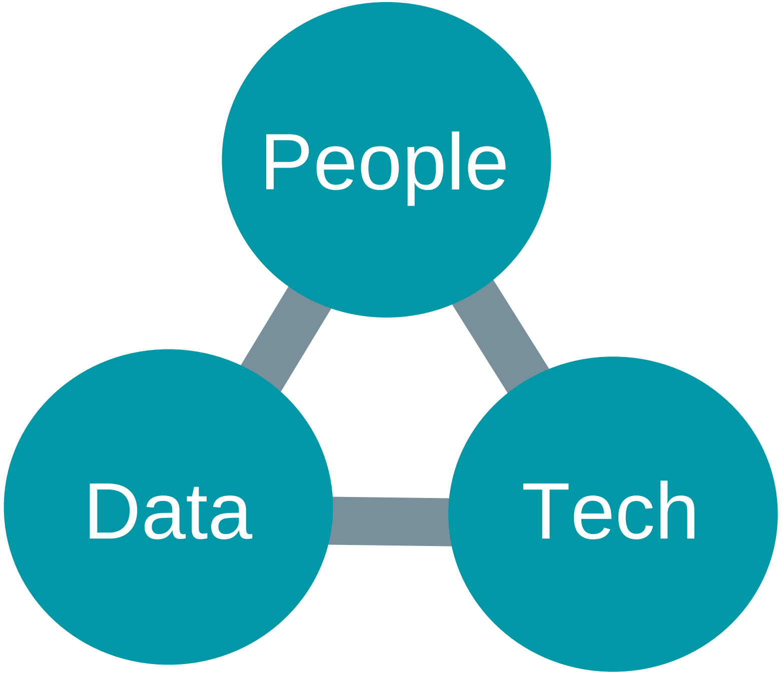
- A Virtual Team...
  - Have the best minds looking at your data
- Working on Your Data...
  - Sharing data with traditional means is labor intensive & costly
  - NI4AI makes it easy and secure to collaborate
  - Open sourcing data not required
- Solving Your Problems!
  - Technology most relevant to your problems will be easily deployable into your infrastructure to operate on your data



Why is Dominion Participating in NI4AI?

I Want to Unleash my Analytics Program!

- ~~Data~~ ← I have lots of it
- ~~Tech~~ ← I have the best
- People ?? ← I have a small, dedicated team but… is it enough?
- Ideas ?? ← So many ideas, but… which are the best?
- Cross-Validation & Insights ← I need this!

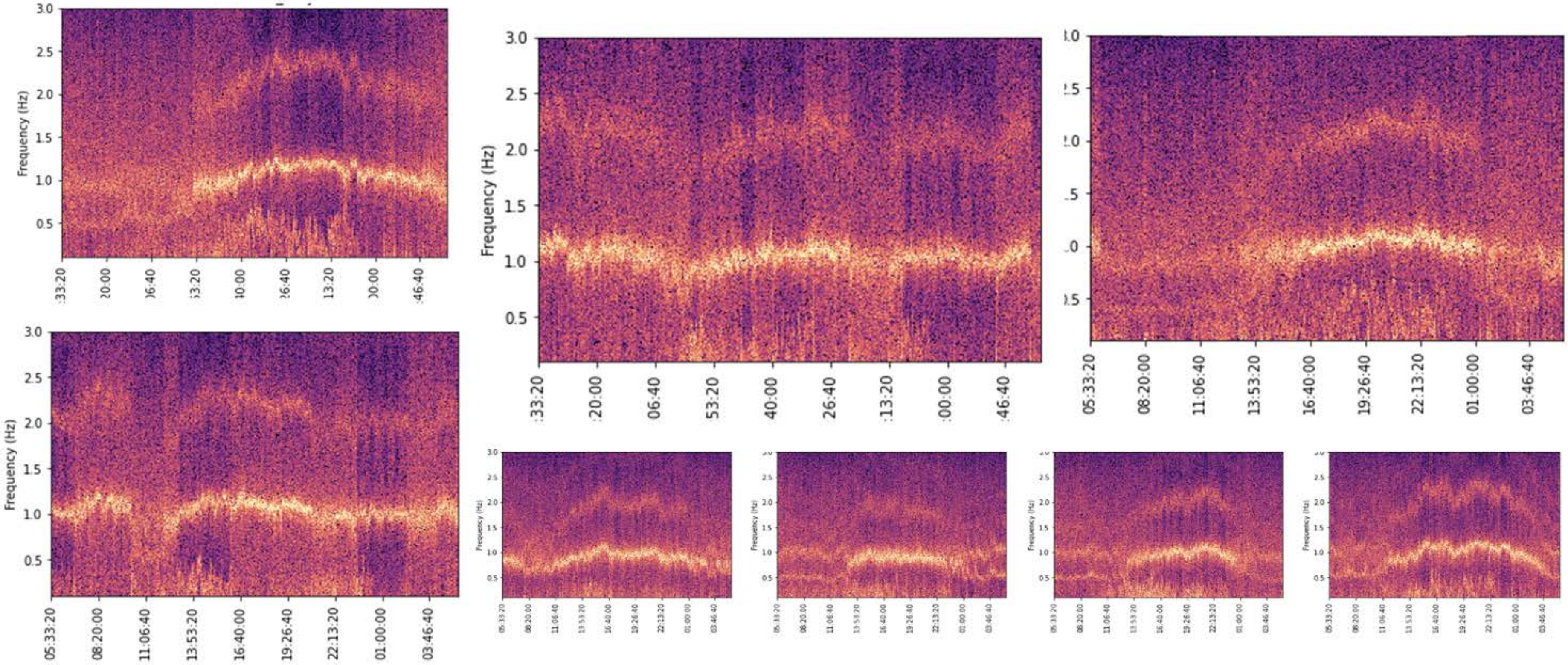


People

Data

Tech

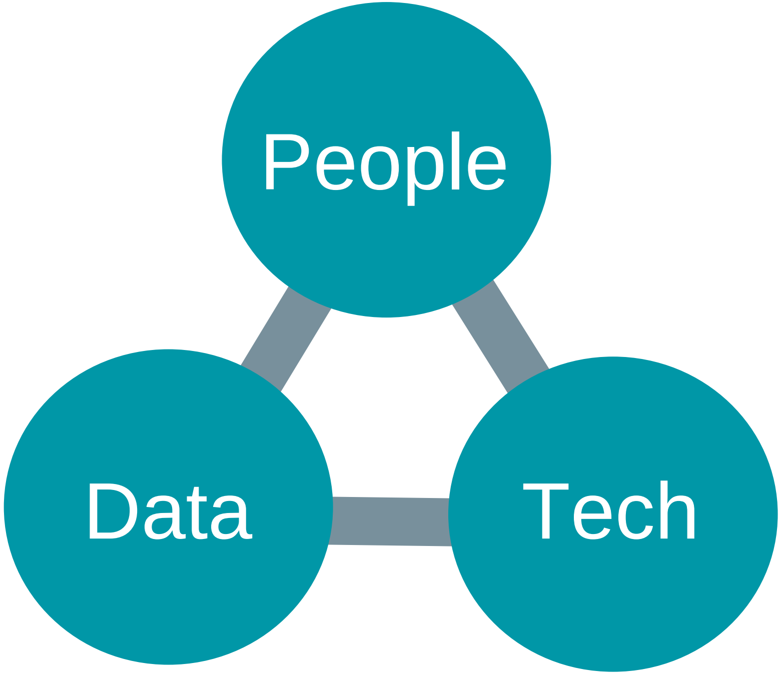
Some Inspiration



Why is Dominion Participating in NI4AI?

What can we offer to the community?

- Access to real system data at scale
- Real utility problems that matter
- Real pathways for Tech-to-Market
- Be the guinea pig

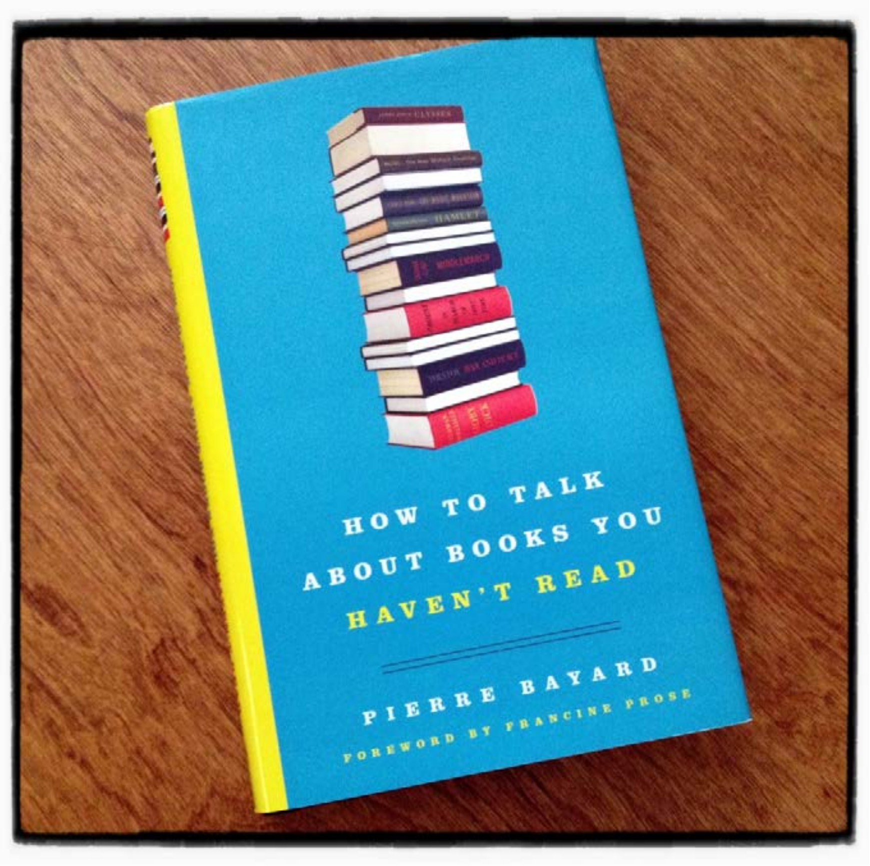


We Can’t Do It All Alone

The choice to read one book is a choice to not read many others.

The choice to take one action is a choice to not take many others.

The choice to explore one use case is a choice to not explore in many others.



What questions do you have about participating in NI4AI?

Top

18	It would be great if the tools and methodologies could become open source, as well as the data itself. But which license??
9	Will there be DOE funding support mechanisms to have students involved in this effort? Without direct funding support it is difficult to assign specific students to the effort for sufficient time to generate results that showcase value of the ideas and of the infrastructure.
8	Will the data be accessible from other geographies (outside US)?
5	Is it any 50Hz /60Hz or higher sampling rate public data available from distributed energy resources? This is a current obstacle to apply load monitoring methods in the academy focused on DER
4	Will there be an effort in developing teaching / training materials together with other US universities based on the platform? This should help developing the capabilities of US universities in bringing data proficiency to the classroom.
1	How is the particiaption fin NI4AI from countries outside US?
-8	We at I-EMS Group Ltd (www.iemsgroup.ca) provide SaaS to our customers for, 1) Predictive Analytics, 2) DERMS/OPF analysis, 3) Energy trading using our Blockchain based TE platform. Data is of essence to all of our SaaS services. My email: arkian@iemsgroup.ca, Cell: +1-647-606-3590

NASPI North American SynchroPhasor Initiative

Thank you!

Get Involved:

Join the community
Learn more. Get access.

Log in at ni4ai.org

Become a contributor
Have data? Resources? Ideas?

Contact sean@pingthings.io