A Golden Dataset to Enable Automatic Photovoltaic Fault Detection Processes

Daniel Fregosi[1], Wayne Li[1], Steven Koskey[2], **Scott Sheppard[2]**, Chris Perullo[2]

[1]Electric Power Research Institute, Charlotte, NC
[2]Turbine Logic, Atlanta, GA

2023 PVPMC Workshop
May 9, 2023

The Problem…Using Available Data to Full Potential

- Large plant may have “10s” of inverters
- Limited sensor data available to detect DC side faults

- Each Inverter contains 10s to 100s of Combiner Boxes
- Each CB may have current and voltage measurements
- Can be used for diagnostics – not typically used today

Can we couple physics-based modeling and AI to better detect and localize string level faults?

Project Motivation

- **Goal:** Provide a better diagnostic solutions with fewer false alarms and actionable M&D insight
- SUBTLE FAILURES ACROSS THE DC COLLECTOR FIELD OFTEN GO UNDETECTED FOR LARGE AMOUNTS OF TIME
  - Determining the source of the failure is time consuming
  - Aerial inspections
    - Performed to detect small-scale faults across the DC collector field
    - Are typically performed infrequently
    - Current gold standard
- MODEL-DRIVEN APPROACH USES BIG DATA AVAILABLE AT PV SITES ENABLES REAL-TIME DETECTION
  - Improves ability to detect faults while reducing presence of false alarms
  - Improves ability to locate faults to more specific hardware components
  - Provides further aid in diagnosing cause of underperformance

M&D centers have abundant data available, how can it better be used for detection of subtle faults?

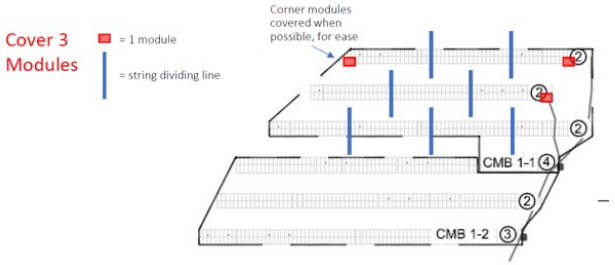
This work is funded in path by the U.S. Department of Energy Solar Energy Technologies Office, under award number DE-EE-0008976.

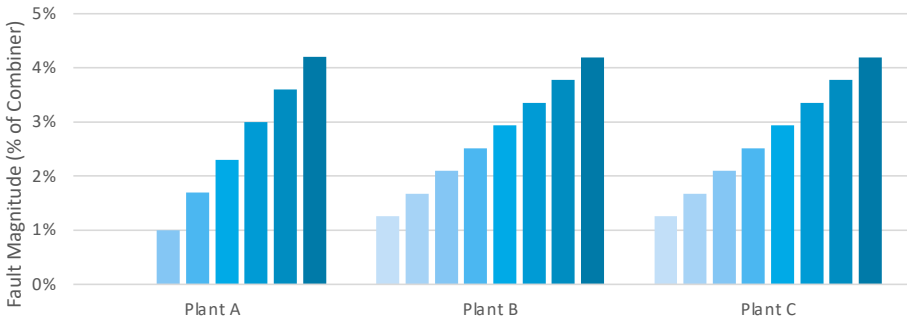
Creating "Clean" Validation Datasets

- **Purpose:** create "golden datasets" with known faults to use for benchmarking fault detection algorithms
- Introduce faults to site hardware where absolute impacts are known
- Intentional faults should be introduced in a **controlled** manner
  - Magnitude of each fault is known
    - How faulted does the hardware need to be to be detectable?
  - Specific hardware is known
    - Are we detecting faults on the correct hardware?
  - Exact time that each intentional fault was introduced is known
    - Do we detect the fault at onset?

Intentional Fault Introduction

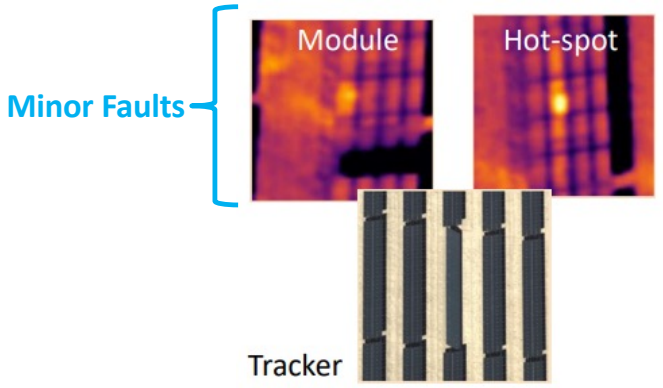
- Method: block sunlight from individual modules by covering them with opaque fabric
- Fabric effectively reduced sunlight reaching each covered panel, triggering bypass diodes
- Introduce faults introduced at three plants at different locations and site architectures

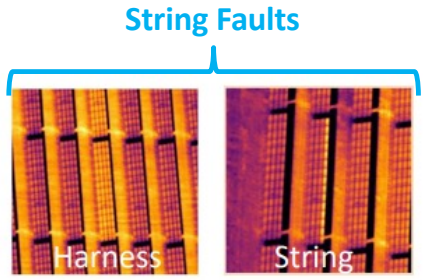




Determining Plant State via Aerial Scan

- Real site data often is **not** clean
  - Hardware may be faulted, sensors may not be reporting correctly
- IR and RGB aerial scans are common method to detect faults (down to specific modules)
- Aerial scans performed at each site prior the introduction of intentional faults

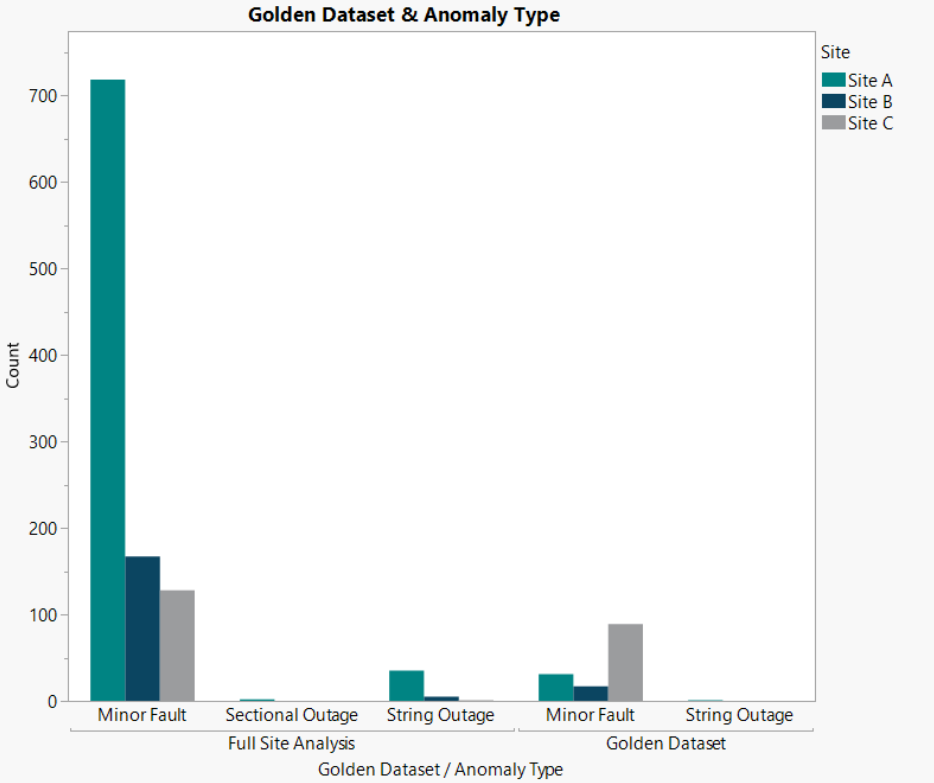




Additional faults present at each site have been catalogued and are shared as part of each dataset.

Aerial Scan Summary

- PV Sites often have many faults of different types present at any given time
- Faults can be bucketed into different categories, based on fault magnitude
  - **Minor Faults**: Impact a single module
  - **String Outages**: Impact one or more strings
  - **Sectional Outages**: Impact several strings
- Overlapping faults presents a challenge for benchmarking, obscures how much impact each fault has



Golden Dataset Preparation

- Data has been anonymized to respect owner privacy
- Raw data from affected inverters and one on-site met station
- Generalized site metadata
- Findings from aerial scans performed prior to the introduction of intentional faults

	A	B	C	D	E
1	Timestamps	Inverter 1 Combiner Box 1 DC Current	Inverter 1 Combiner Box 10 DC Current	Inverter 1 Combiner Box 11 DC Current	Inverter 1 Combiner Box 12 DC Current
905	11/1/2021 15:03	170.24578	154.79992	126.67841	127.77987
906	11/1/2021 15:04	170.41721	154.61811	126.539955	127.68941
907	11/1/2021 15:05	170.58865	154.43629	126.40149	127.59896
908	11/1/2021 15:06	170.76006	154.25447	126.26303	127.50851
909	11/1/2021 15:07	170.92972	154.07294	126.12457	127.41805
910	11/1/2021 15:08	170.89085	153.92495	125.99455	127.3276
911	11/1/2021 15:09	170.70902	153.79995	125.89921	127.23715
912	11/1/2021 15:10	170.52721	153.67495	125.8076	127.14669
913	11/1/2021 15:11	170.34541	153.54996	125.71599	127.05625
914	11/1/2021 15:12	170.16358	153.42495	125.6244	126.96579
915	11/1/2021 15:13	170.01451	153.29995	125.53279	126.87534
916	11/1/2021 15:14	169.99884	153.17496	125.44119	126.78489
917	11/1/2021 15:15	169.99884	153.05058	125.34959	126.69443
918	11/1/2021 15:16	169.86113	152.9	125.25798	126.603985
919	11/1/2021 15:17	169.65826	153	125.16638	126.51353
920	11/1/2021 15:18	169.39991	153	125.07478	126.42307
921	11/1/2021 15:19	169.16	153	124.98317	126.33263
922	11/1/2021 15:20	169.00625	153	124.89157	126.24217
923	11/1/2021 15:21	169	153	124.79997	126.15172
924	11/1/2021 15:22	169	152	124.70837	126.06127
925	11/1/2021 15:23	169	151.16272	124.61676	125.97081
926	11/1/2021 15:24	169	151	124.52516	125.88037

Raw data (4 anonymized tables)

1. Inverter DC measurements
2. Combiner box DC currents
3. Met. station data
4. Tracker motor positions (if applicable)

	A	B	C
1	Inverter	Combiner Box	Fault Name
2	1	3	Sub-Module Anomaly
3	1	3	Suspected Shorted Diode(s)
4	1	5	Hot Spot - Edge
5	1	5	Vegetation
6	1	6	Short Circuit
7	1	6	Suspected Shorted Diode(s)
8	1	11	Hot Spot - Edge
9	1	11	Parallel Hot Spots
10	1	11	Vegetation
11	1	13	Multiple Cell - Hot
12	1	13	Surface Fouling
13	1	15	Parallel Hot Spots
14	1	15	Surface Fouling
15	1	15	Vegetation
16	1	17	Vegetation
17	1	17	Vegetation
18	1	17	Vegetation
19	1	20	Hot Spot - Edge
20	1	20	Surface Fouling
21	2	2	Parallel Hot Spots
22	2	2	Vegetation
23	2	12	Sub-Module Anomaly
24	2	12	Suspected Shorted Diode(s)

Summary of aerial scan-identified faults

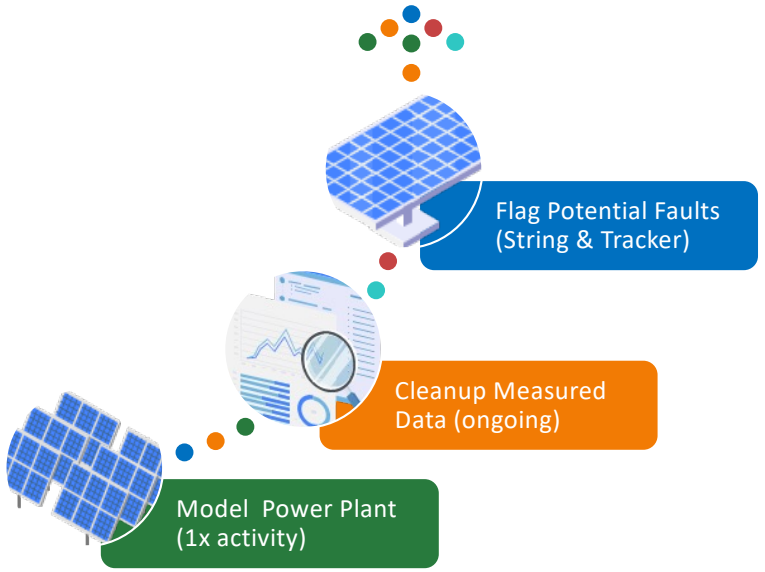
	A	B	C
1	Latitude	32.5	
2	Longitude	-84	
3	I_sc_ref	9.96	
4	V_oc_ref	48	
5	I_mp_ref	9.36	
6	V_mp_ref	38.4	
7			

Simple metadata

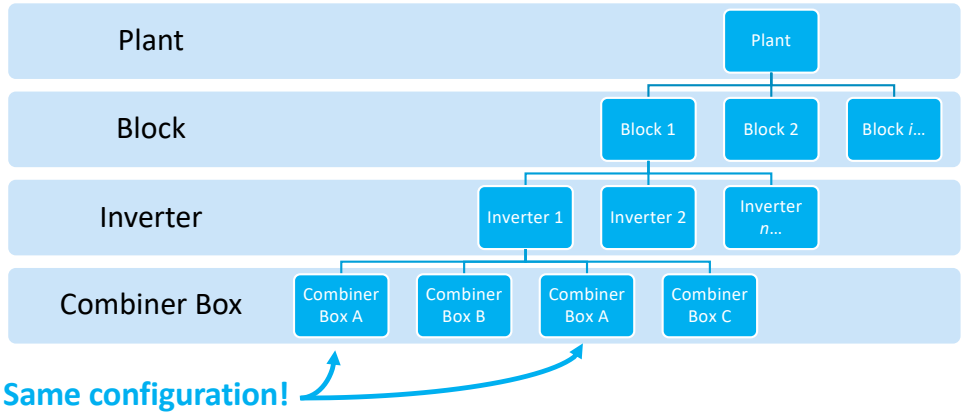
Modeling Approach

Physics-based models coupled with AI to identify failed string and tracker outages

TYPICALLY COMPRISE 3% OF LOST GENERATION



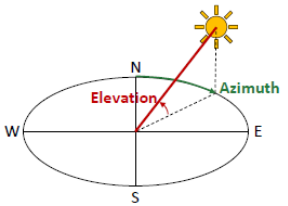
- Model leverages modularity of PV plant for quick customization to new sites
- Model can run in real-time or on historical data



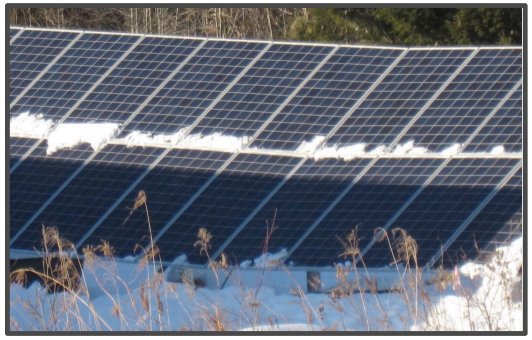
Data Filtering and Cleanup

- Removing noise and poor-quality data significantly improves detection capability
- IEC 61724-3 (right) provides guidance on filtering data from sensors
- Filtering against self-shading and clouded operation significantly reduces noise

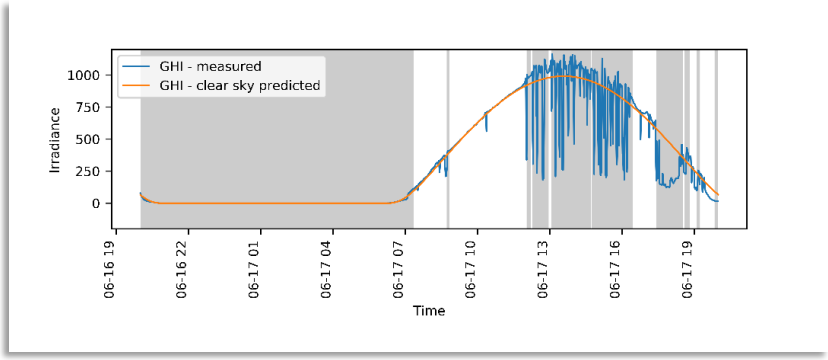
		Suggested criteria for flag (15 min data)			
Flag type	Description	Irradiance W/m²	Temperature °C	Wind speed m/s	Power (AC power rating)
Range	Value outside of reasonable bounds	< –6 or > 1 500	> 50 or < –30	>32 or < 0	> 1,02 × rating or < –0,01 × rating
Missing	Values are missing or duplicates	n/a	n/a	n/a	n/a
Dead	Values stuck at a single value over time. Detected using derivative.	< 0,0001 while value is > 5	< 0,0001	?	?
Abrupt change	Values change unreasonably between data points. Detected using derivative.	> 800	> 4	> 10	> 80 % rating
May be adjusted depending on the tilt of the system and the season of data acquisition.					



Self-shading from adjacent row depends on row spacing and is most prevalent in the winter

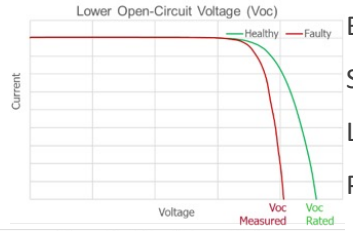


Filter data that is rapidly changing due to clouds, reduces false alarms and increases detection sensitivity



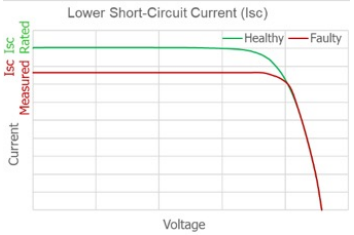
Fault Detection

- Various faults impact the current-voltage characteristic of the hardware
- **With online diagnostics, the full curve is not measured**
- Physical model allows for replication of the curve
- Feature extraction leads to identification of the presence of faults
  - Measured Data: I_{mpp}, V_{mpp}
  - Modeled Data: I_{sc}, V_{oc}



Voc Reduced

- Bypass diode activation
- Shorted module (fault)
- Light induced degradation
- Potential induced degradation



Isc Reduced

- Soiling, Shading, Snow
- Front glass corrosion, encapsulant yellowing
- Loss of string (fault or fuse)

Fault	Strings per CB	Panels per String	# Covered Panels	Fault Magnitude (@ Inv)	Detected
1	24	29	29	4.2%	✓
2	24	29	21	3%	✓
3	24	29	16	2.3%	✓
4	24	29	12	1.7%	✓
5	24	29	8	1.1%	

How to access

- https://github.com/epri-dev/PV-Golden-Datasets

- One dataset currently uploaded, the other two in the coming days
- Datasets will eventually be hosted on OSTI.gov