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

www.epri.com

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
- <u>Can be used for diagnostics – not typically used today</u>

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?

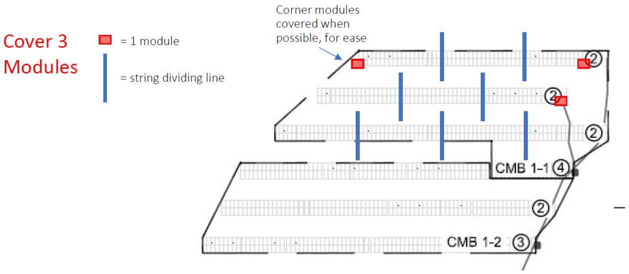
This work is funded in path by the U.S. Department of Energy Solar Energy Technologies Office, under award number DE-EE-0008976.

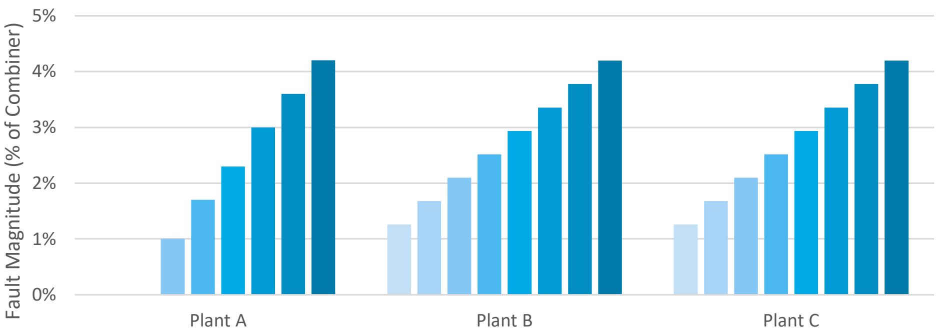
Creating “Clean” Validation Datasets

- **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

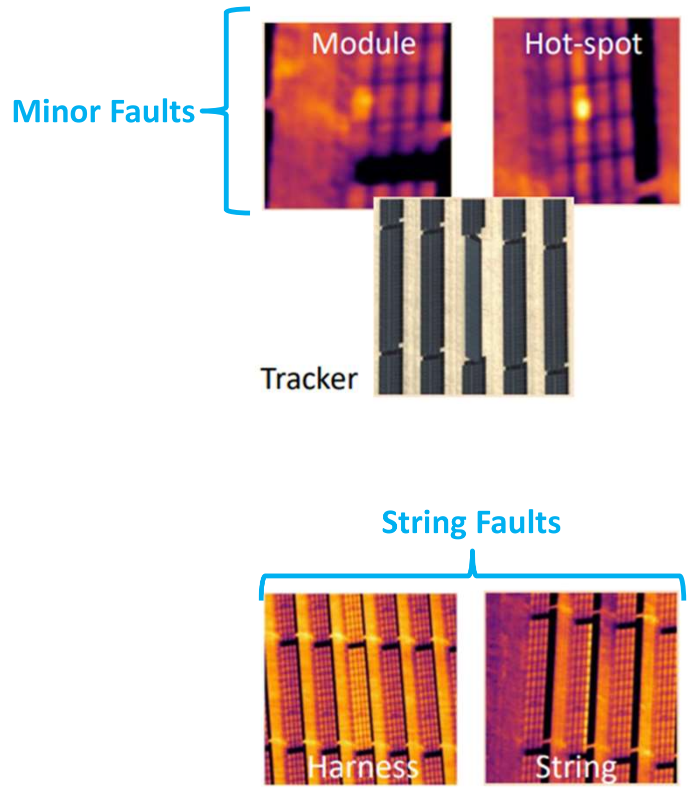
- Method: block sunlight from individual modules by covering them with opaque fabric
- Fabric effectively reduced sunlight reaching each covered panel, triggering bypass diodes
- Introduce faults introduced at three plants at different locations and site architectures





Determining Plant State via Aerial Scan

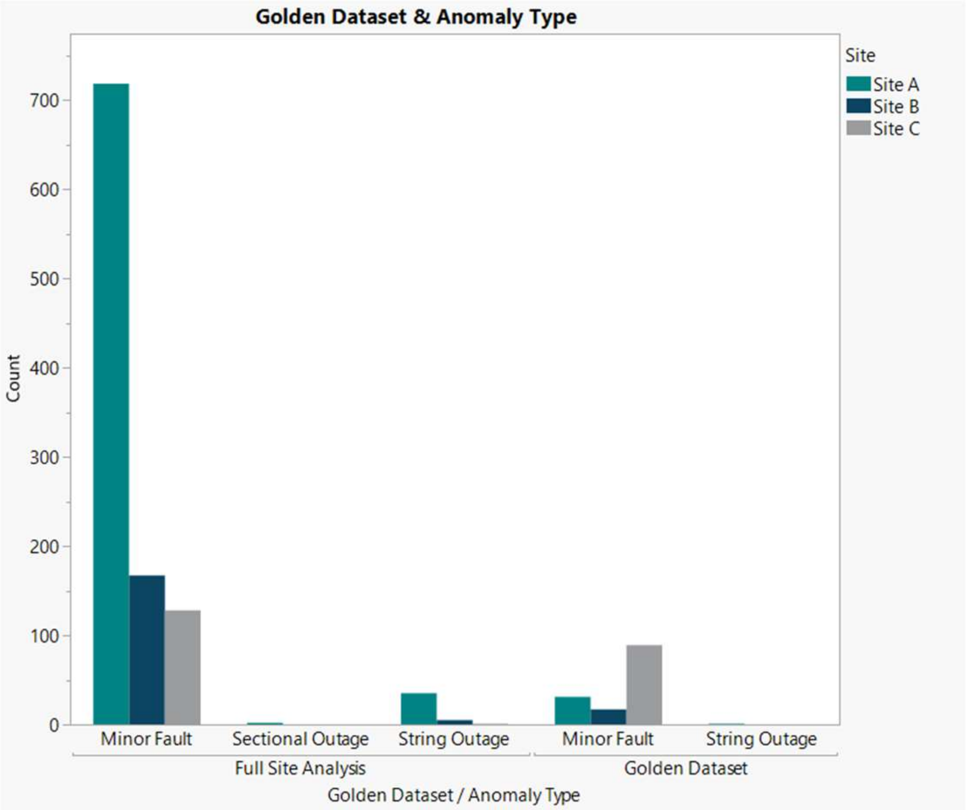
- Real site data often is **not** clean
  - Hardware may be faulted, sensors may not be reporting correctly
- IR and RGB aerial scans are common method to detect faults (down to specific modules)
- Aerial scans performed at each site prior the introduction of intentional faults



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

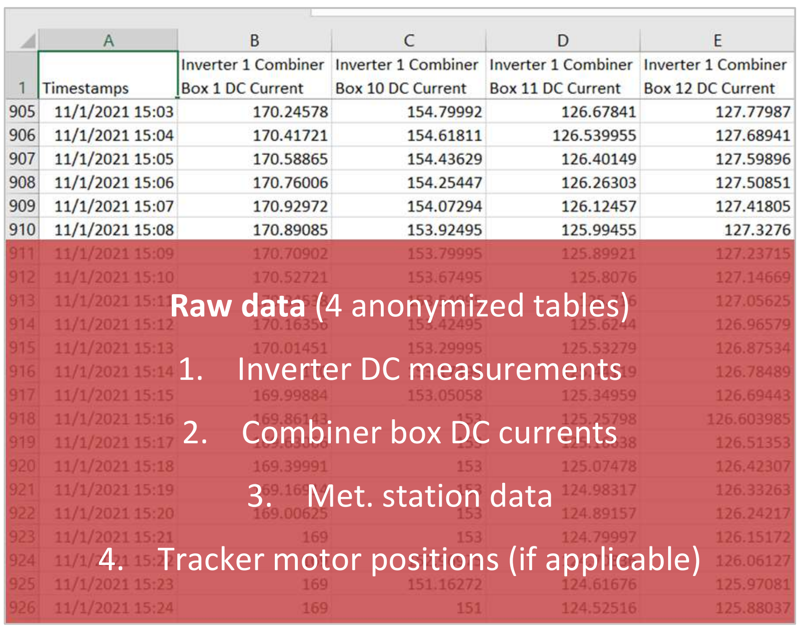
Aerial Scan Summary

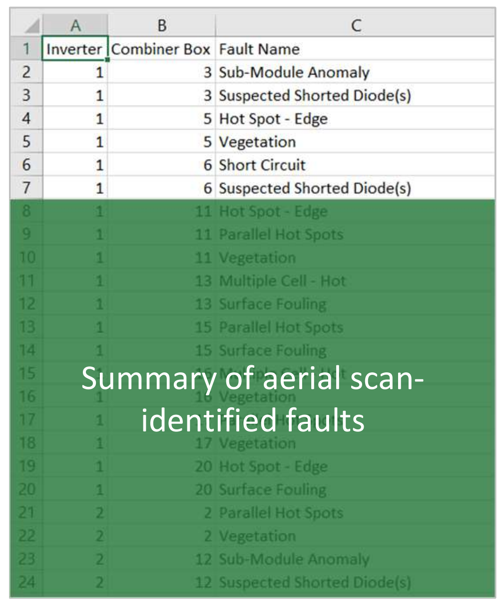
- PV Sites often have many faults of different types present at any given time
- Faults can be bucketed into different categories, based on fault magnitude
  - **Minor Faults**: Impact a single module
  - **String Outages**: Impact one or more strings
  - **Sectional Outages**: Impact several strings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

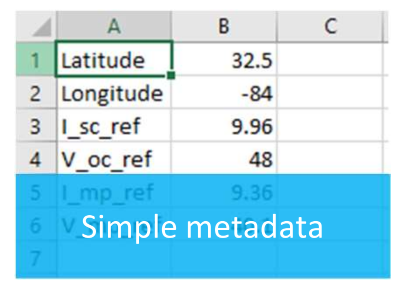


Golden Dataset Preparation

- 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



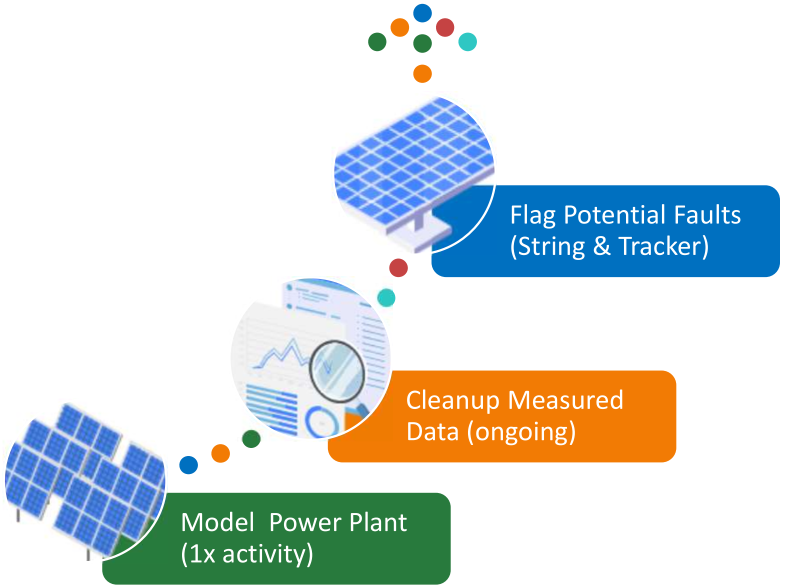




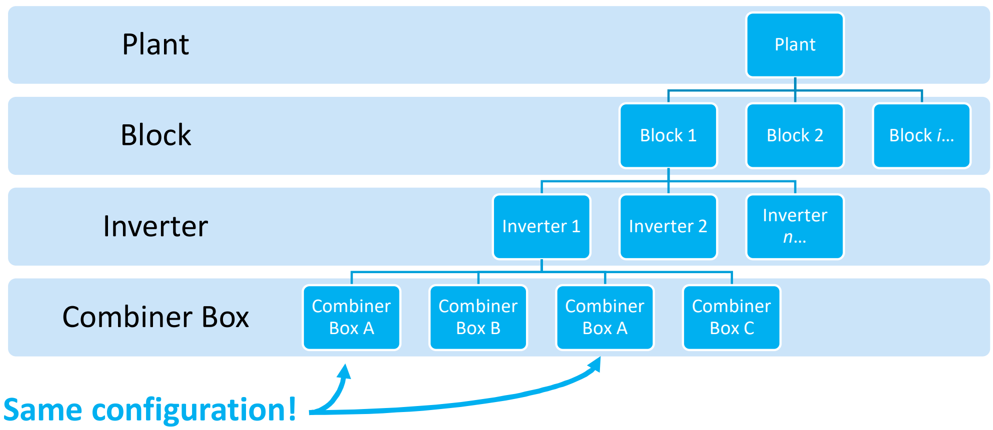
Modeling Approach

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

TYPICALLY COMPRISE 3% OF LOST GENERATION



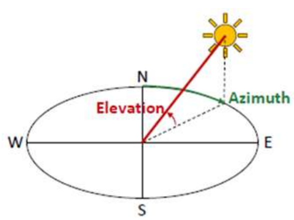
- Model leverages modularity of PV plant for quick customization to new sites
- Model can run in real-time or on historical data



Data Filtering and Cleanup

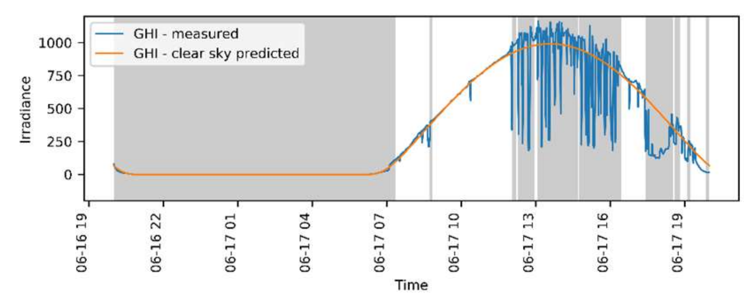
- Removing noise and poor-quality data significantly improves detection capability
- IEC 61724-3 (right) provides guidance on filtering data from sensors
- Filtering against self-shading and clouded operation significantly reduces noise

Flag type	Description	Suggested criteria for flag (15 min data)			
		Irradiance W/m^2	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

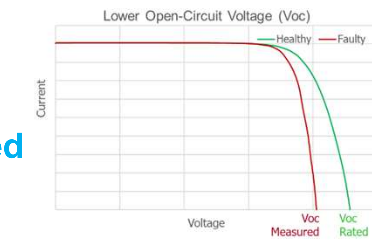


EPRI

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



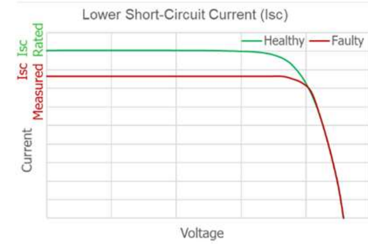
Bypass diode activation

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