Automating Detection and Diagnosis of Faults, Failures, and Underperformance in PV Plants

Scott Sheppard[1], Daniel Fregosi[2], Wayne Li[2], Kamran Paynabar[3]

[1]Turbine Logic, Atlanta, Georgia, 30308, USA

[2]Electric Power Research Institute, Charlotte, North Carolina, 28262, USA

[3]Georgia Tech, Atlanta, Georgia, 30332, USA

Introduction

- Large-scale PV plants have real-time instrumentation throughout the DC collector field, often recording data as granular as combiner box data.
- Regardless, hardware fault detection at PV plants is commonly performed by aerial imaging and is typically performed only on an annual or bi-annual basis.
- Utility Monitoring and Diagnostic centers have a strong need to utilize this data to better detect faults without flooding detection systems with false alarms and to more accurately identify where within the plant the fault occurred

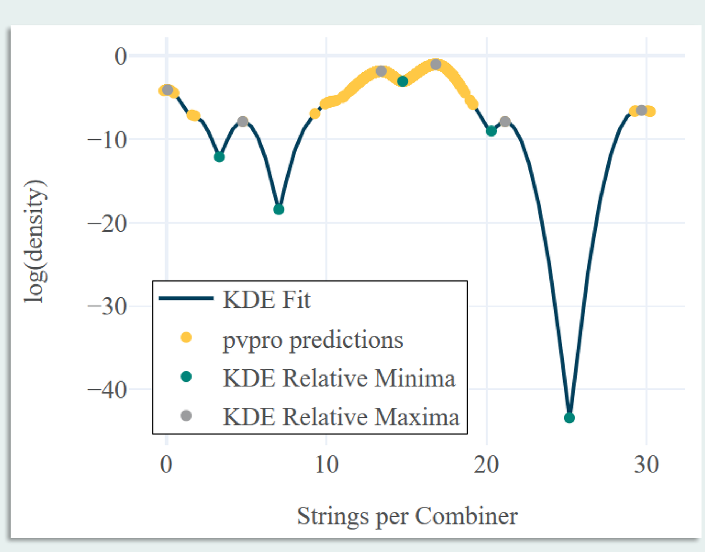
Objectives

- Improve existing fault detection algorithm and quantify detection limits
- Develop and automate fault diagnosis algorithm
- Make "golden data set", then make public for benchmarking
- Benchmark against commercial algorithms
- Develop software tool to link sensors and failures, and make public

How is AI/ML Used in this Project?

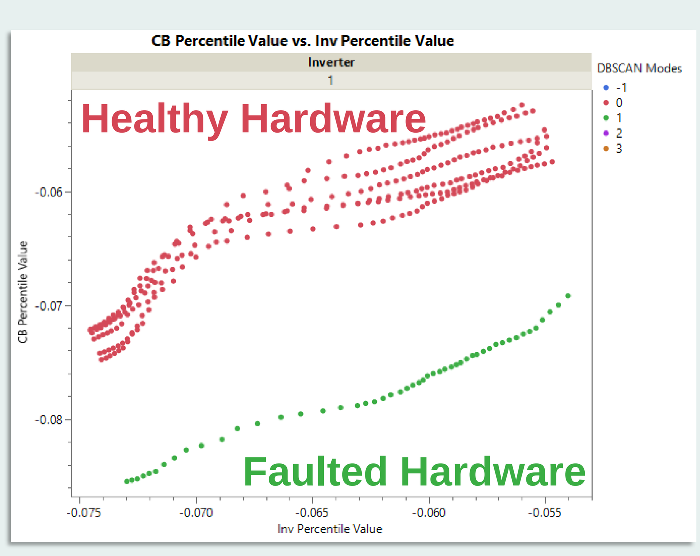
Clustering

- Clustering was used in two applications: automated setup of physics-based models and identifying faulted hardware



Combiner box architecture was estimates using PVPRO – estimates were grouped via clustering to mimic actual plant layouts

Data features extracted from the model were clustered to identify faulted hardware

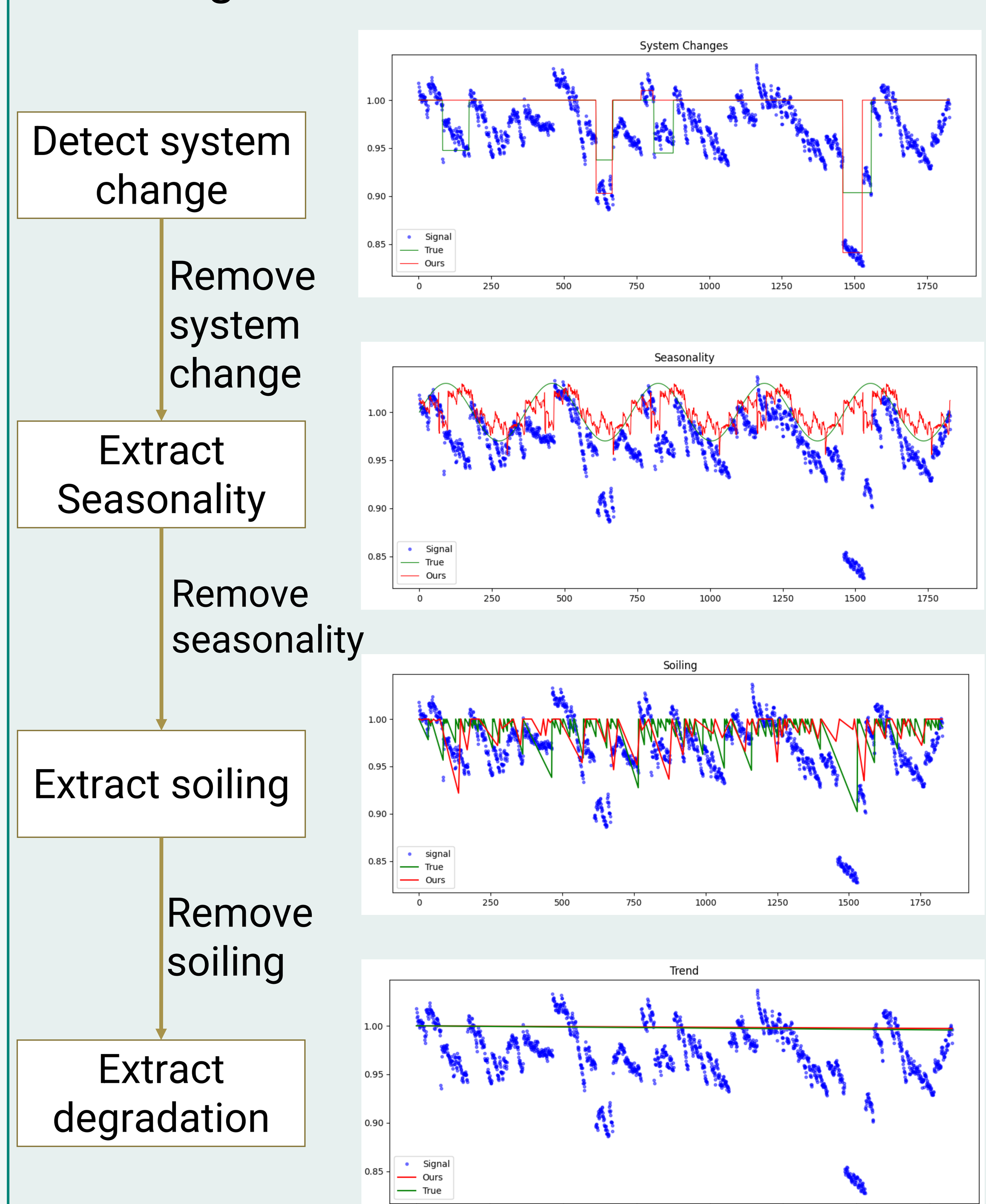


Sensitivity	TPR	FPR	F1
High	65%	19%	0.57
Med	49%	10%	0.54
Low	36%	7%	0.46

Detection results across nine utility sites

Signal Decomposition

- Signal decomposition enabled removal of seasonal and soiling effects to detect true degradation



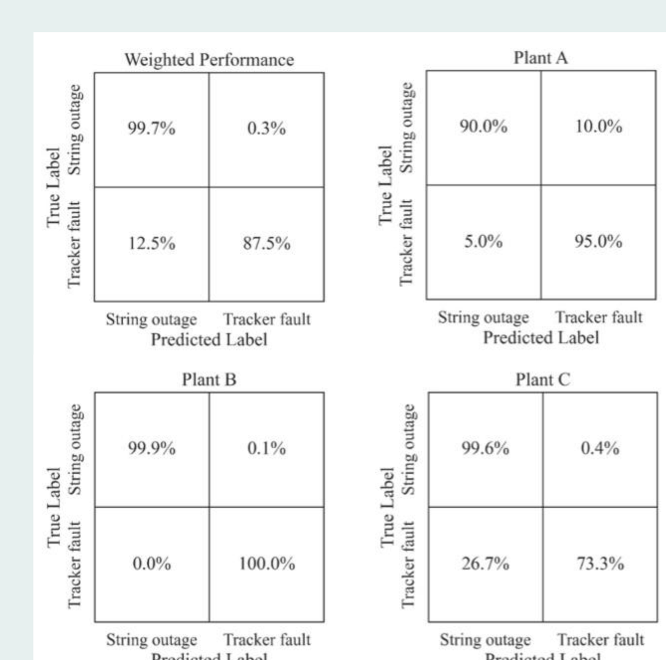
Classification

- Xgboost was used to identified if the fault was persistent or temporary
- Random forest classifiers identified faults in each category
- Classifier has been tested on both synthetic data and plant data

Table 1 Results on the Simulated Dataset

	Precision	Recall	F1-score	Support
Voltage Fault	0.99	0.96	0.97	20
Current Fault	0.99	0.98	0.98	20
Series Resistance Fault	0.90	0.87	0.88	20
Shunt Resistance Fault	0.91	0.96	0.93	20
Periodic Shading Fault	0.80	0.85	0.82	20
Snow Fault	0.92	0.89	0.90	20
Tracker Fault	0.91	0.97	0.93	20
Outage Fault	0.78	0.68	0.72	20

Classification results on synthetic data



Classification results on utility data

Challenges and Best Practices

- Due to the amount of hardware at PV plants, site drawings may vary from the actual layout and data channel labels. Automating model building can better match model parameters to actual site architecture.
- Real systems' data widely varies in quality. Choosing systems with full sensor suites and ample validation data aids in developing effective modeling solutions.
- However, industrial systems often have noisy, incomplete data. Model development must consider both high- and low-quality data together to better mimic "real world" deployment.
- Industrial applications have different constraints than research applications. Solutions that are simple to deploy and have very few false positives will see higher uptake.

Key Takeaways and Future Work

- AI/ML can be used to enable many steps throughout model building, fault detection, and fault diagnosis.
- AI/ML solutions should be flexible to varying hardware architectures and data availability. Adoptees will expect the solution to work on their unique system.
- Low inaccuracy is often a more important evaluation metric than high accuracy for industrial systems. False positives are seen as time wasters.
- These fault detection and diagnostic algorithms will be benchmarked against common commercial solutions as a next step.

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