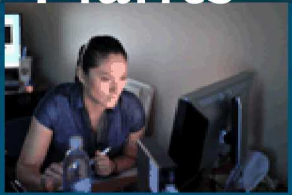
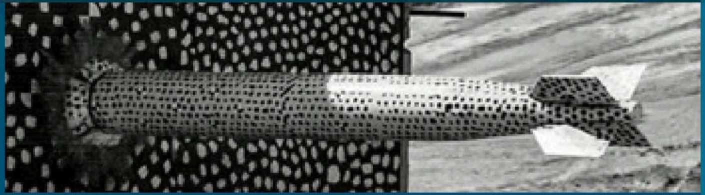
SAND2020-10949C

Fusing Diverse Datasets with Machine Learning to Inform System State Conditions: An Application for Photovoltaic Plants





Thushara Gunda

INFORMS 2020

TB44. Statistical Machine Learning for Energy Analytics

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Acknowledgements

Funding support from DOE SETO Agreement 34172

Colleagues –

- Nicole Jackson, Hector Mendoza, Michael Hopwood, Rachel Homan, Cliff Hansen
- Andy Walker, Gerald Robinson, Roger Hill, Jal Desai
- Christopher Downs, Laura Kraus, Ryan Jones, …

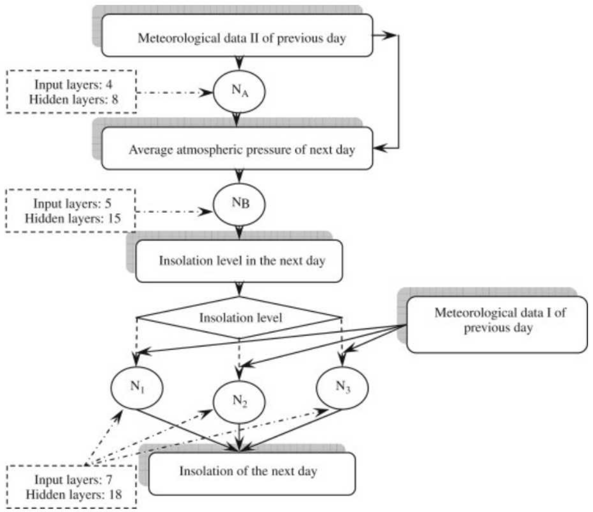
Views expressed in this presentation are solely my own and do not reflect those of the DOE or any other US Government entity.

Outline

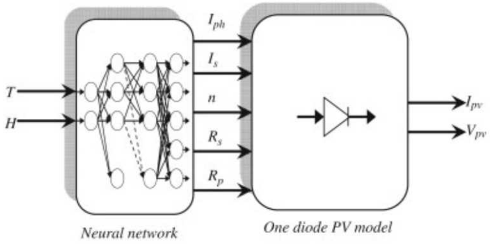
- Machine learning & Photovoltaics
- PV & Failures
  - Failure characterization using IV Traces
  - Identification of common failure modes within text records
  - Fusion of datasets to evaluate weather impacts
- PV plants as a complex system
- Future work

Machine Learning & Photovoltaics

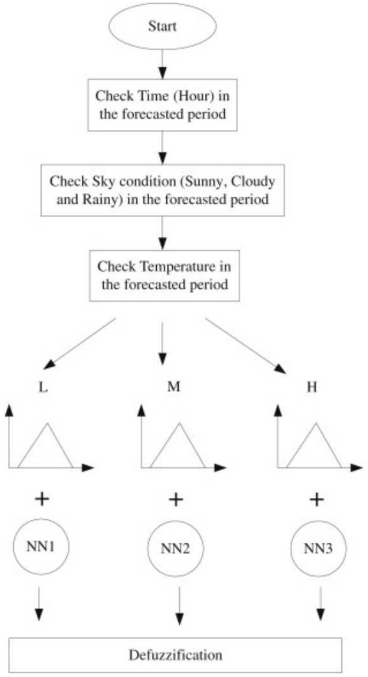
- ML has been leveraged in PV for decades
- Multiple applications, including
  - Radiation and weather forecasting
  - Sizing of PV systems
  - Simulation of PV systems and controls
- Across stand-alone, grid-connected, and hybrid systems



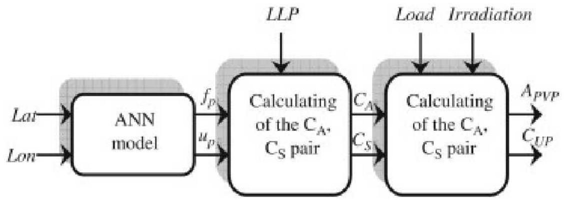
(Kemmoku et al., 1999)



(Karatepe et al., 2006)

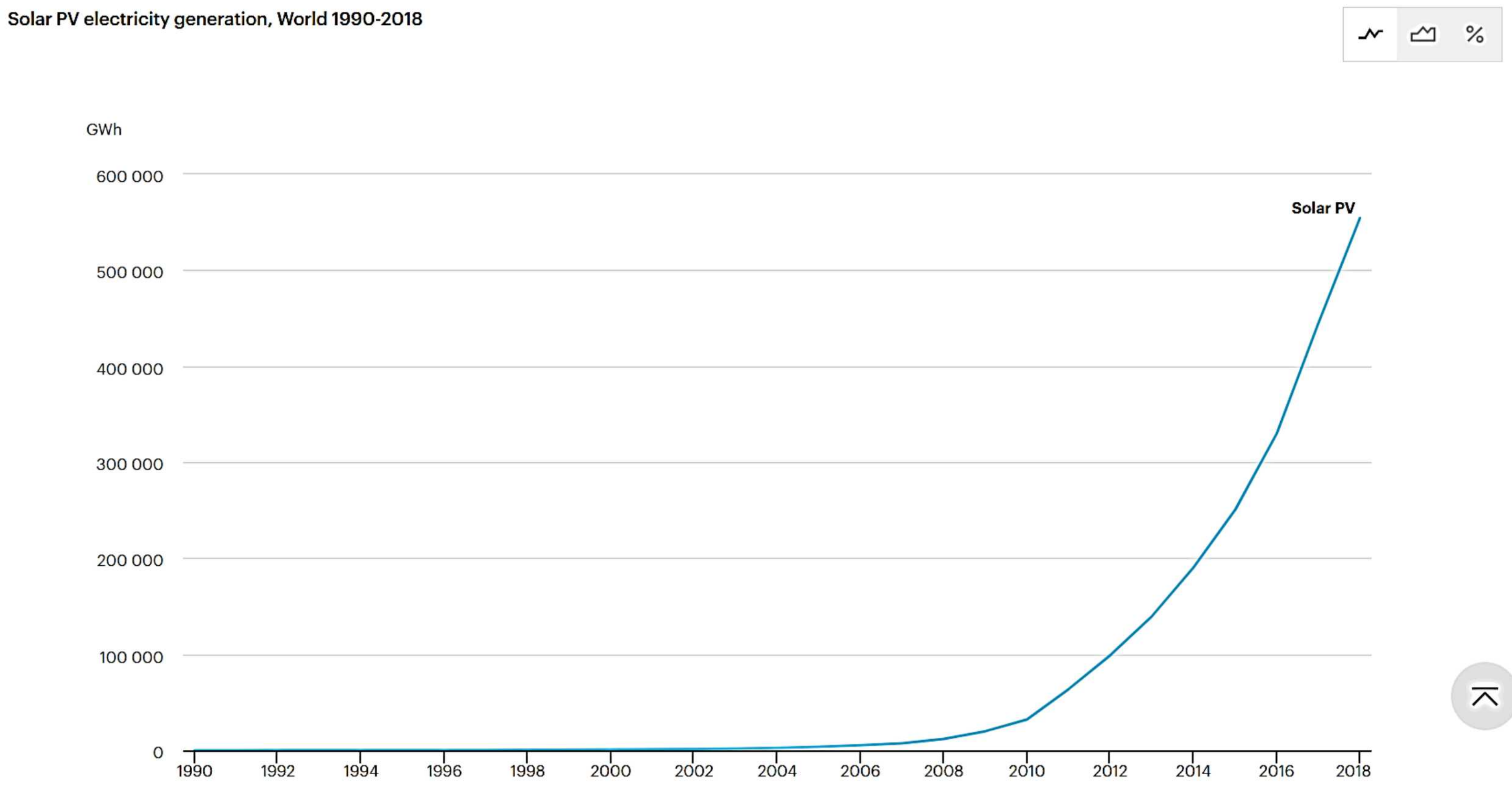


(Chen et al., 2013)



(Mellit et al., 2003)

Photovoltaics

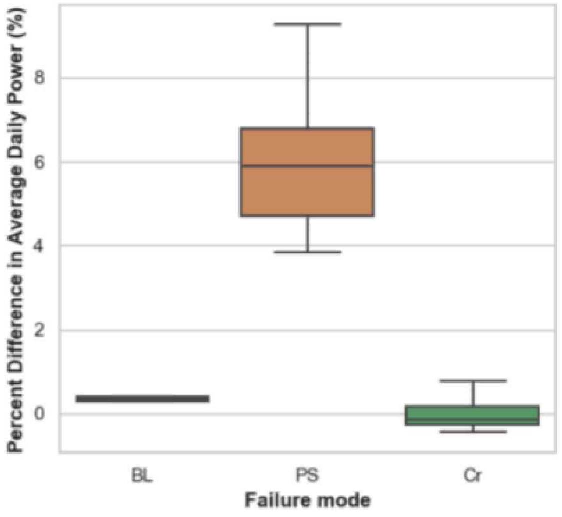


(IEA, 2020)

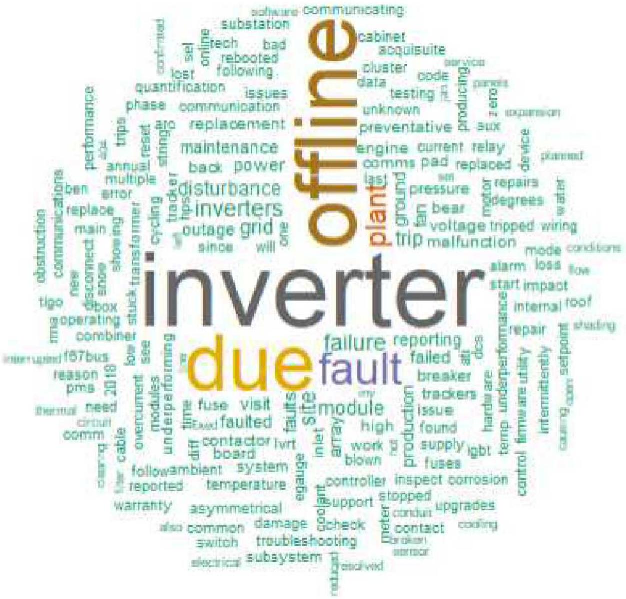
- PV production has demonstrated exponential growth over the last decade
- Understanding and dealing with the PV system aging process is creating a new and unknown set of challenges

PV & Failures

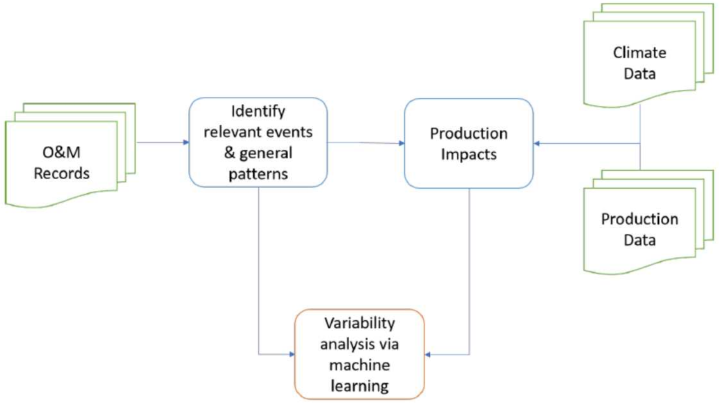
- More recently, machine learning is being increasingly used for understanding failures within the PV industry



Failure Characterization



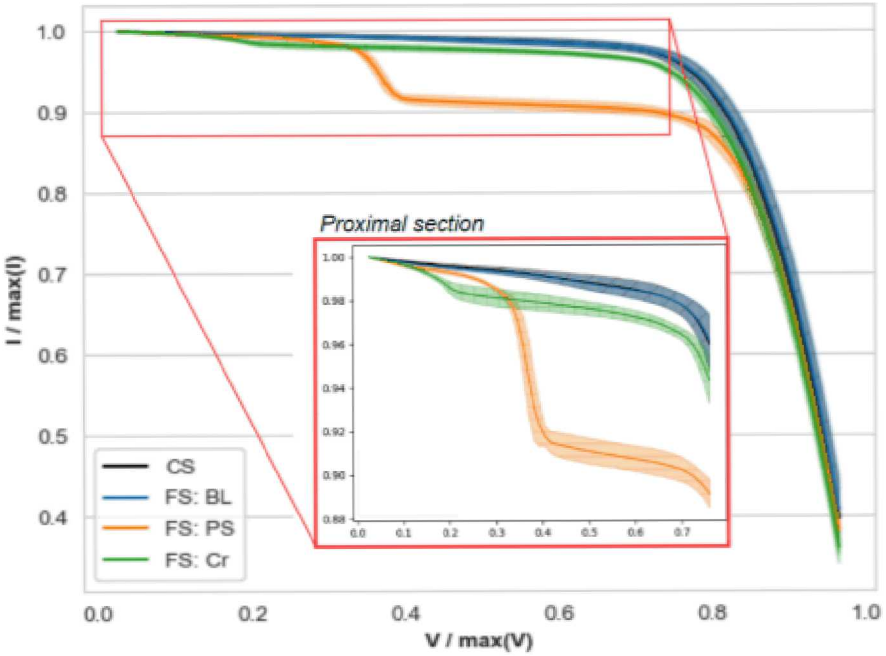
Common Failure Modes



Weather Impacts

Failure Characterization

- Extended current approaches of feature extraction to consider the entire IV trace using 3 NN architectures
- Data processing including quality checks, normalization, and interpolation
- Multi-headed LSTMs and 1D CNNs had comparable high accuracies. Single-headed LSTM outperformed when only considering proximal regions.



NN Architecture	Sampling Region	Num. Predictors	Average accuracy, % (SD) across 20 tests			
			BL	PS	Cr	Total
Multi-headed LSTM	Entire Curve	4	100.0 (0)	100.0 (0)	98.0 (2.8)	99.3 (1.0)
	Proximal	4	99.1 (2.8)	100.0 (0)	97.6 (5.1)	98.9 (1.7)
Single-headed LSTM	Entire Curve	4	74.5 (37.5)	68.2 (42.0)	23.7 (36.3)	51.8 (15.8)
	Proximal	4	70.3 (44.6)	95.0 (21.8)	76.2 (37.4)	79.0 (20.8)
1D CNN	Entire Curve	4	99.4 (2.4)	100.0 (0)	100.0 (0)	99.8 (0.8)
	Proximal	4	100.0 (0)	100.0 (0)	94.0 (10.3)	98.0 (3.2)
1D CNN	Entire Curve	2[a]	71.2 (40.8)	80.0 (40.0)	62.7 (42.7)	68.9 (27.4)

[a] Utilizing only I_{CS}, I_{FS} out of the normal set: I_{CS}, I_{FS}, ϵ_{FS}, and δ_I

(Hopwood et al., 2020)

O&M Data

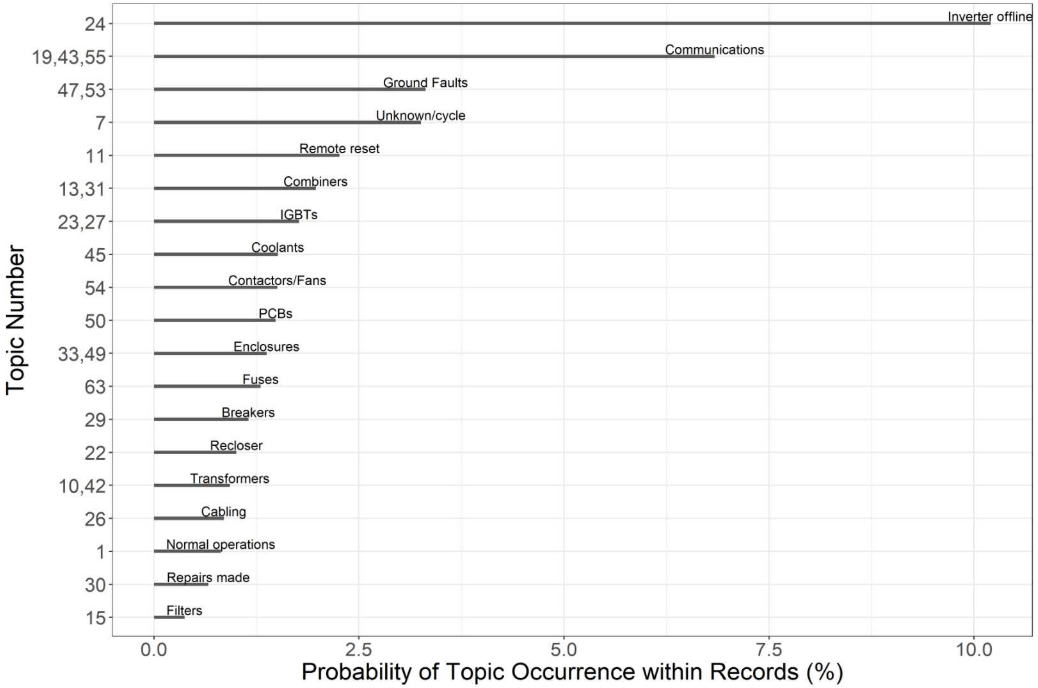
- Maintenance logs contain valuable contextual information
- However, significant diversity in the structure and detail of the logs makes it challenging to ascertain needed insights
- Most contain a general comment field to capture issue description

Context	Where	What	When	Why	How
WO type	Site ID	General desc.	Event Start	Priority	Completion desc.
WO Status	Asset	Failure cat.	Event End	Production impact	Effort (labor)
Warranty info				Impact level	

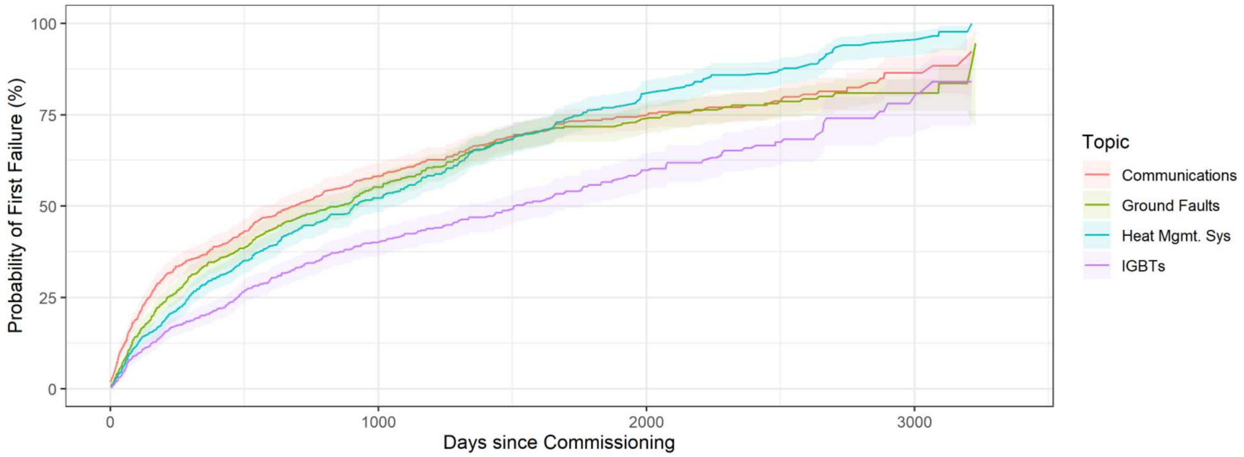
randid	WOType	WONumber	FailureCategories	Cause	ImpactLevel	FullDesc	CompletionActivity
C2S16	Corrective	17_014989	Hardware Malfunction	NA	Partial	Circuit 2 trackers stuck in flood	Hardware Adjustmer
C2S1	Corrective	17_017409	Software Problem	NA	Partial	Trackers going into flood stow	Software/Firmware /
C2S16	Corrective	17_015278	Hardware Malfunction	NA	Partial	Circuit 2 Trackers went to Flooc	Power Cycle
C2S16	Corrective	17_015727	Software Problem	NA	Partial	Circuit 2 in flood stow due to m	Power Cycle
C2S16	Corrective	17_018812	Hardware Malfunction	NA	Partial	Circuit 2 in flood stow. Power C	Power Cycle
C2S1	Corrective	17_019236	Hardware Failure	NA	Mixed	All trackers in intermittent floo	Hardware Replacem
C2S1	Corrective	17_025187	Software Problem	NA	Partial	Tracker NCUs indicate in flood	Software/Firmware /

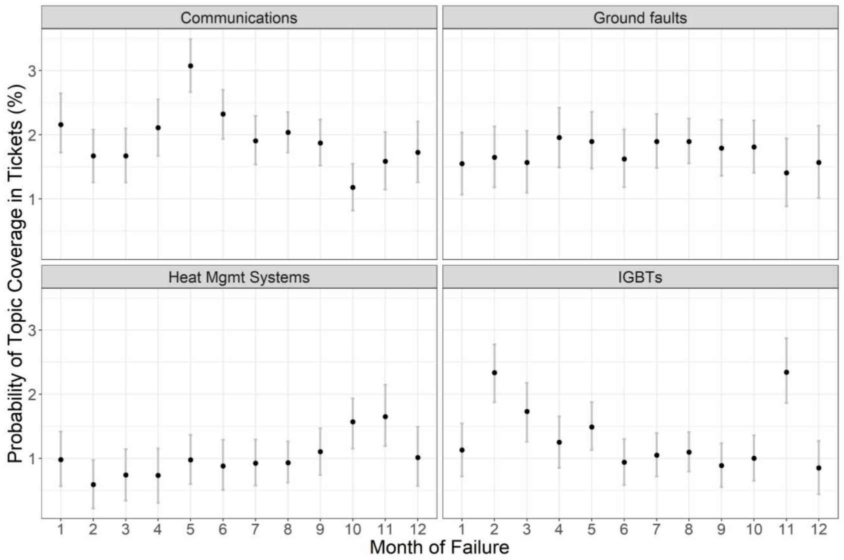
(Gunda et al., In review)

Identification of Common Failure Modes



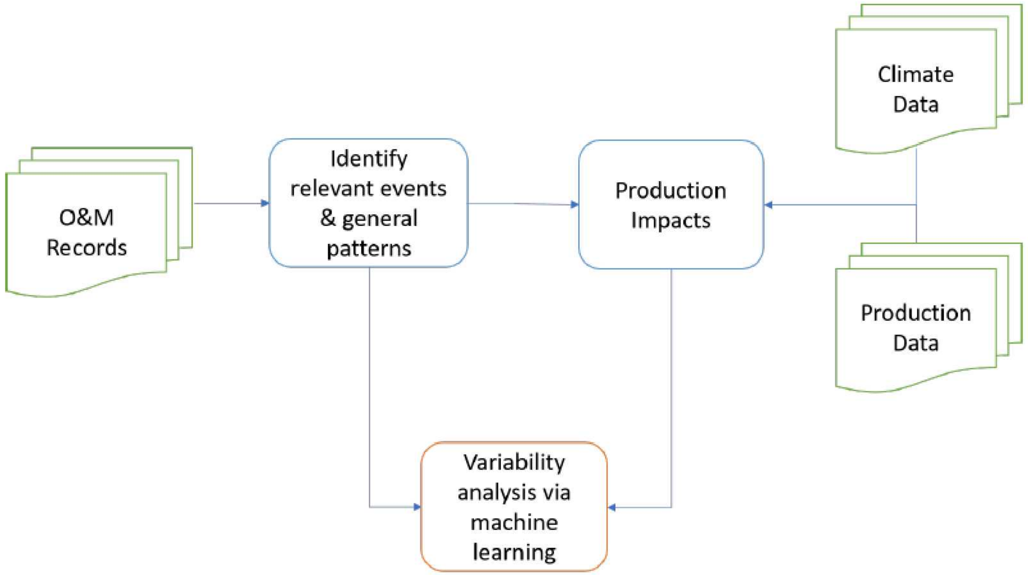
- Focused on inverters – one of the leading causes of PV system failures
- Used ML in 2 ways:
  - Single Vector Decomposition to identify inverter-related records
  - Latent Dirichlet Allocation to group "like" records
- Temporal patterns in clustered records were evaluated using survival analysis and estimateEffects





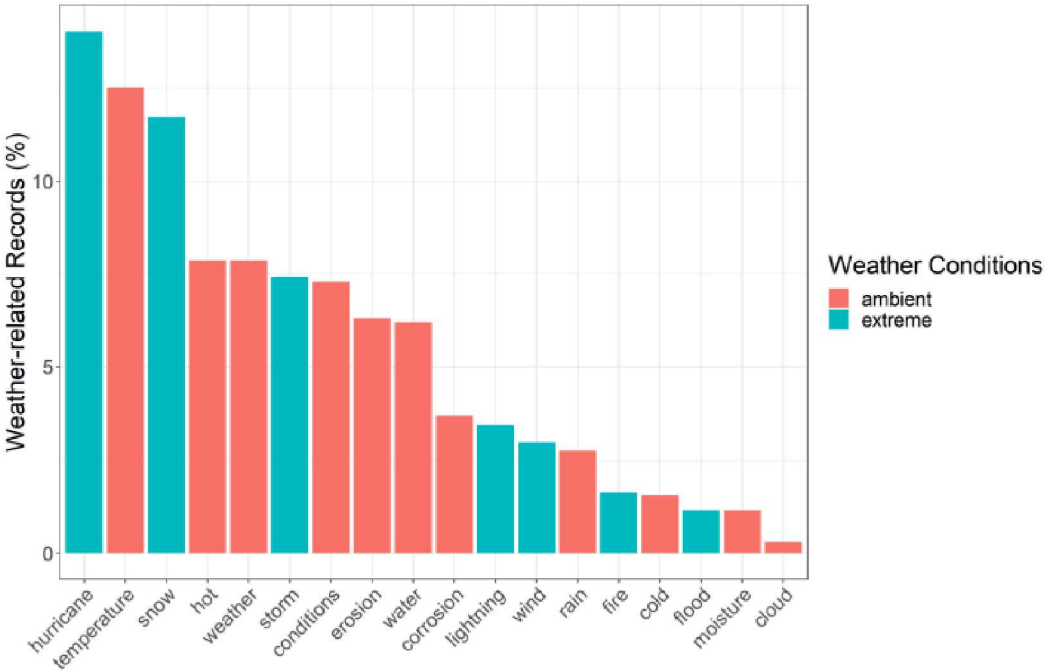
Weather Impacts

(Jackson & Gunda, In review)



- Fused 3 datasets:
  - Production data
  - Weather data
  - O&M records
- Lot of processing involved!

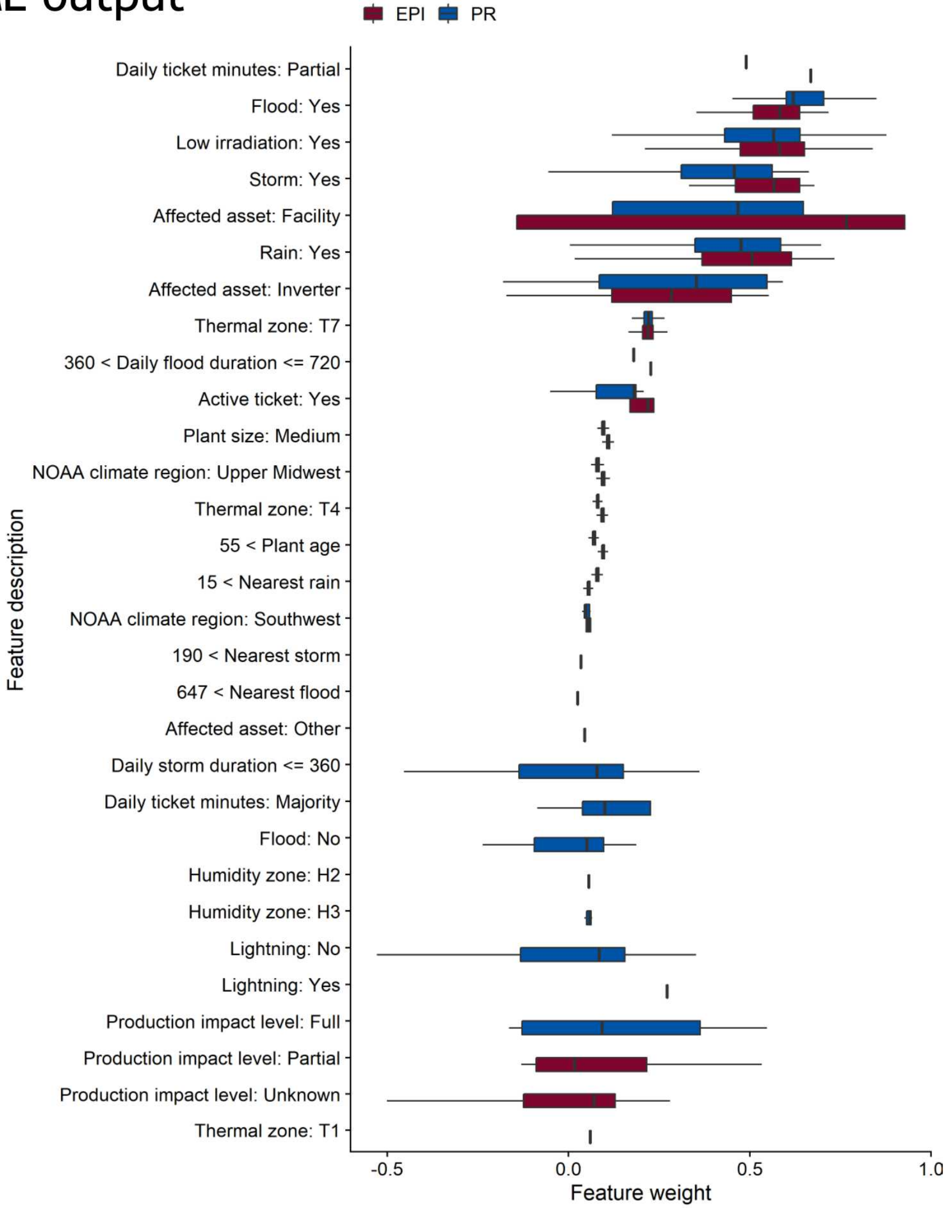
- O&M data was used to constrain analysis
- Unlike inverter-analysis, significant diversity emerged when considering weather events
  - Key Term Identification was used to identify relevant records
  - Quality control and categorization was conducted manually



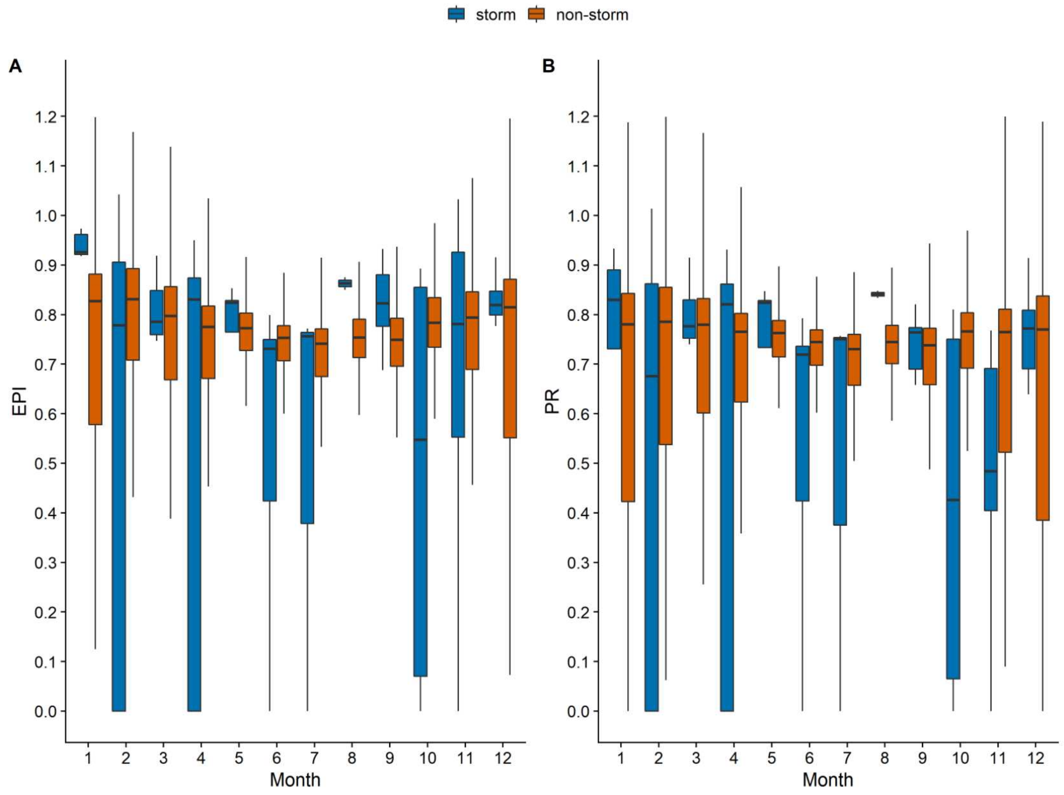
Weather Impacts: Storms

(Jackson & Gunda, In review)

LIME output



- Variations in performance evaluated using statistics and random forest (RF) implementation
- RF analysis incorporated weather, O&M, and metadata information
- LIME analysis was used to characterize important features



PV Plants as Complex Systems

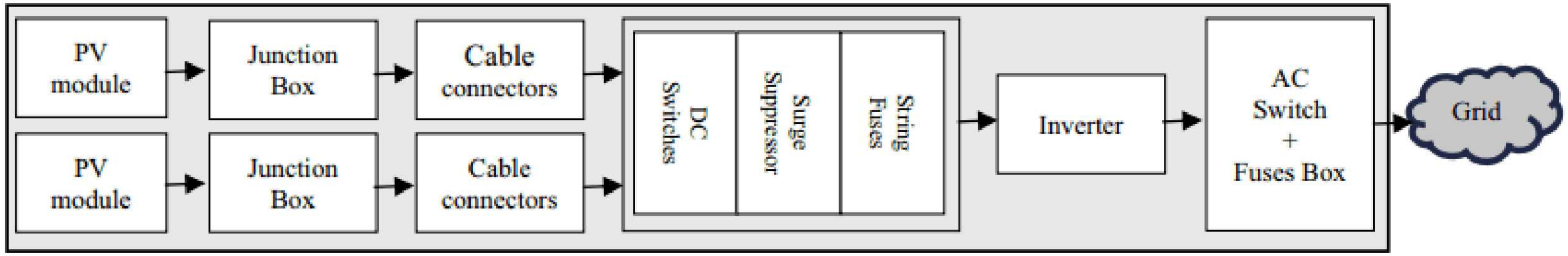
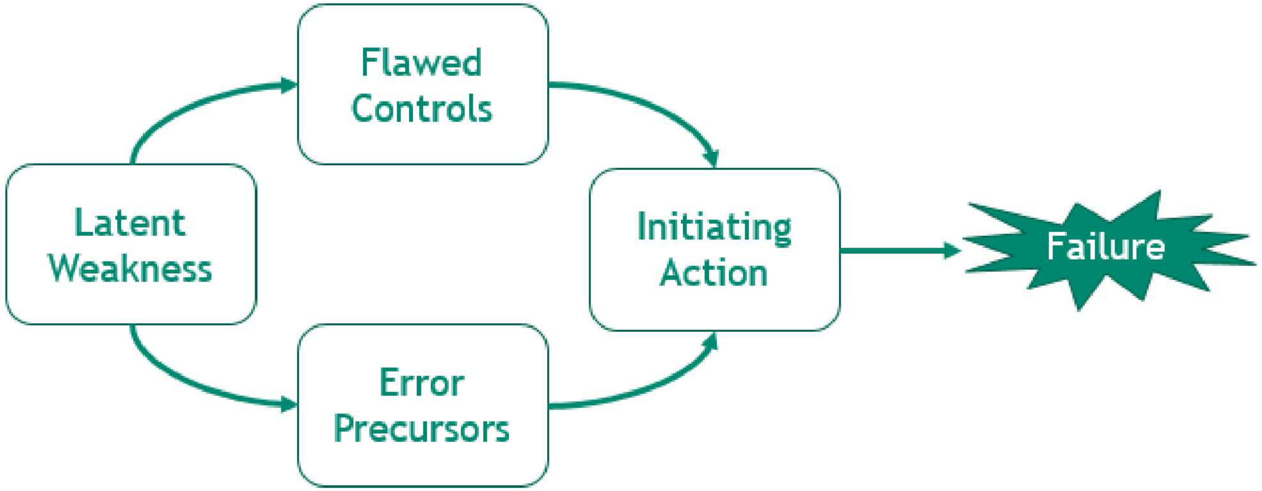


Fig. 1. Simplified schematic diagram of photovoltaic plant.

Cristaldi et al., (2015)

- Lots of parts and interconnections
- Both human and machine elements in laten weaknesses and controls
- Precursors can be maintenance-oriented
- Actions can be chronic or acute
- Non-linear pathways with lots of uncertainty and dynamic components

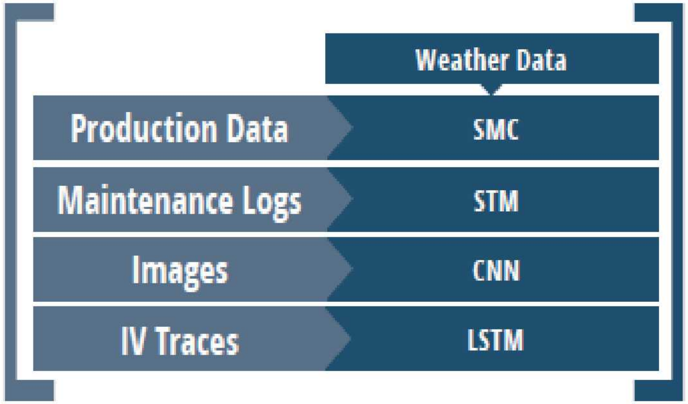
→ COMPLEX SYSTEM

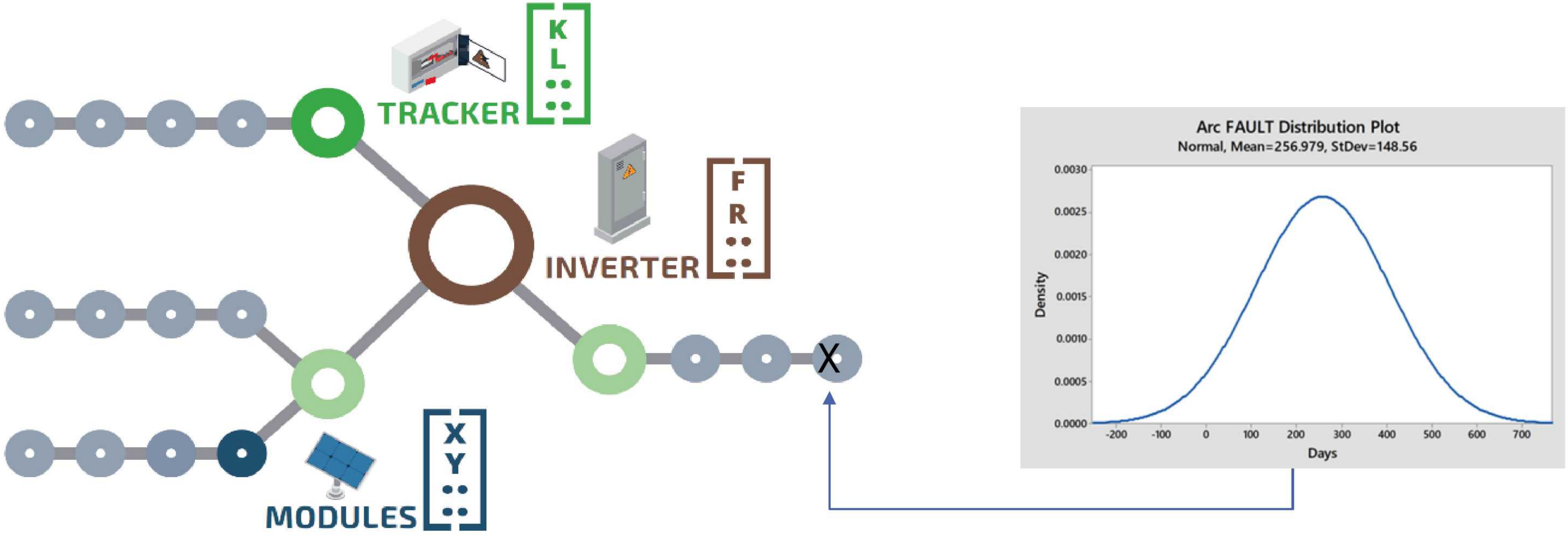


Adapted from the DOE Human Performance Improvement Handbook

Future Work

- ➢From failure characterization to failure prediction
- ➢Stop evaluating data and analyses in silos
- ➢Better integrate system network architecture into algorithm approach





References

Chen, S. X., Gooi, H. B., & Wang, M. Q. (2013). Solar radiation forecast based on fuzzy logic and neural networks. *Renewable Energy*, *60*, 195-201.

Gunda et al., "A Machine Learning Evaluation of Maintenance Records for Common Failure Modes in PV Inverters". Currently in review.

Hopwood, M. W., Gunda, T., Seigneur, H., & Walters, J. (2020). Neural Network-Based Classification of String-Level IV Curves From Physically-Induced Failures of Photovoltaic Modules. *IEEE Access*, *8*, 161480-161487.

IEA, 2020: https://www.iea.org/data-and-statistics?country=WORLD&fuel=Renewables%20and%20waste&indicator=SolarGen

Jackson, N. and T. Gunda. "Evaluation of Extreme Weather Impacts on Utility-scale Photovoltaic Plant Performance." Currently in review.

Karatepe, E., Boztepe, M., & Colak, M. (2006). Neural network based solar cell model. *Energy conversion and management*, *47*(9-10), 1159-1178.

Kemmoku, Y., Orita, S., Nakagawa, S., & Sakakibara, T. (1999). Daily insolation forecasting using a multi-stage neural network. *Solar Energy*, *66*(3), 193-199.

Mellit, A., Benghanem, M., Arab, A. H., & Guessoum, A. (2003, June). Modelling of sizing the photovoltaic system parameters using artificial neural network. In *Proc. of IEEE, CCA* (Vol. 1, pp. 353-357).

Thank you for your time!

Thushara Gunda: tgunda@sandia.gov