Data-driven Design of PV Module Materials: Informed by a Non-relational Data Warehouse & Analytics Environment with > 3.4 GW of PV Plant Datasets

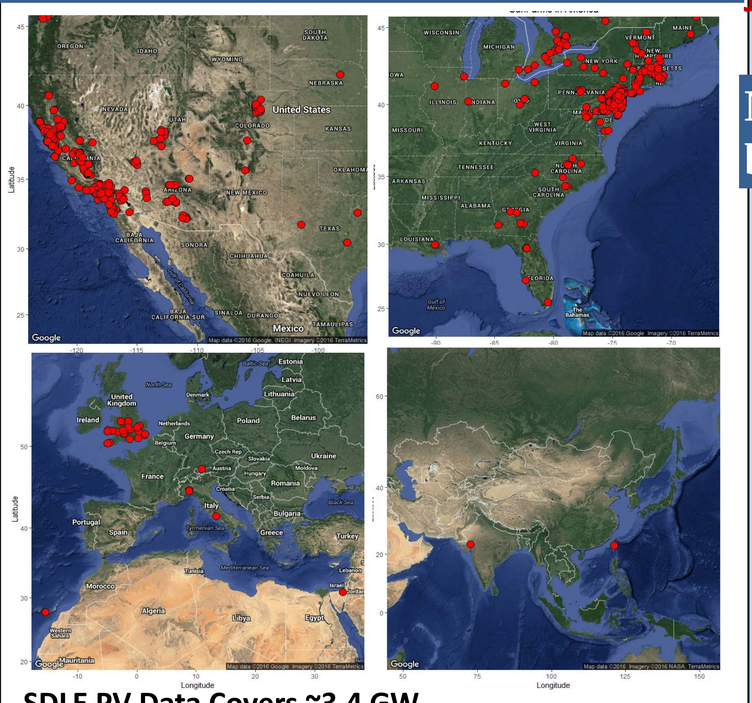
Roger H. French[1], Laura S. Bruckman, and Timothy J. Peshek
[1] Solar Durability & Lifetime Extension Center, Case Western Reserve University, Cleveland OH, USA

Objective

A. Bringing operating PV plants into research
- As a critical epidemiological population
- Non-relation data warehouse – Energy-CRADLE
- Currently hosting data from over 3.4GW
- From 780 field developed power plants

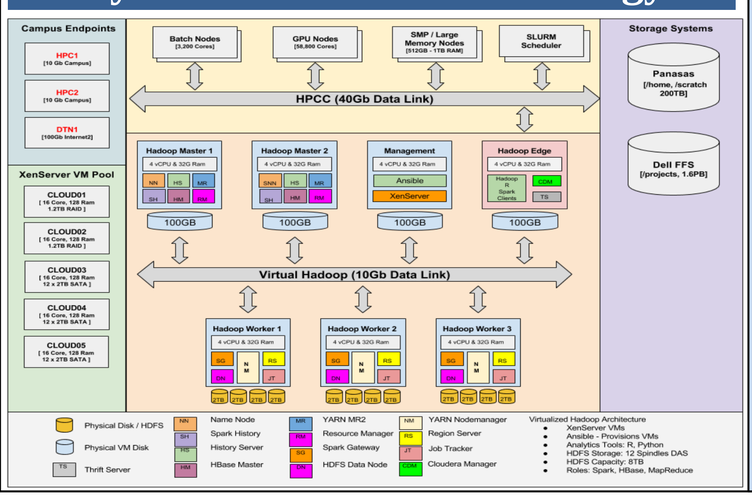
B. Using data science methods
- Such as machine learning, predictive network modeling
- Aggregate and integrate DuraMat research results
- Into common system-level models of PV modules
- Exposed to real-world conditions and lab-based accelerated exposures

Global SunFarm Network

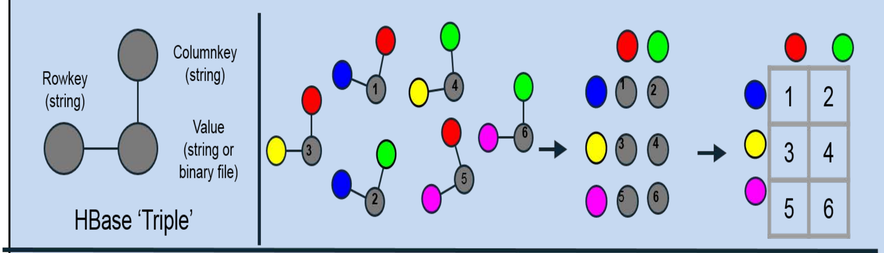


SDLE PV Data Covers ~3.4 GW
- 787 PV Plant Sites
  encompasses 1.92% of Global PV Plant Generation
- 5638 PV Plants (Inv. & Modules)
- Distributed in 13 different climate zones
- 60 PV Module Brands/Models
- 38 PV Inverter Brands/Models
- Single Modules to 265 MW power plants
- Going Back Up To 15 years

Common Research & Analytics Data Lifecycle Environment for Energy



Hadoop/Hbase & NoSQL DB Abstraction [2]



Combines Lab data (Spectra, Images etc.) With Time-series Data (PV Power Plant Data)

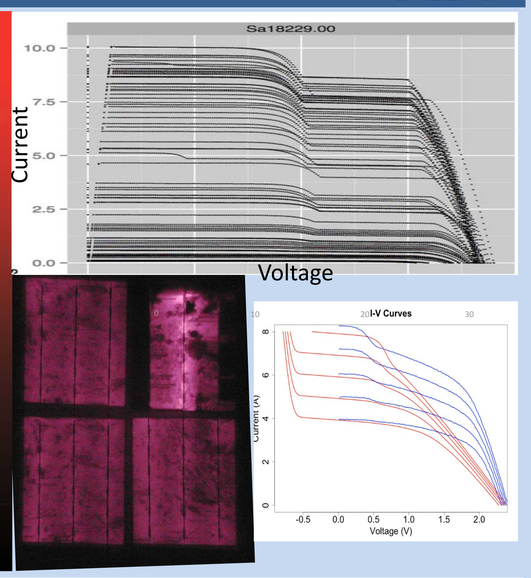
High Performance PV Data Analytics: Petabyte Data Warehouse In A Petaflop HPC Environment
- In-place Analytics: Distributed R-analytics in Hadoop/HDFS
- In-memory Data Extraction: To Separate HPC Compute Nodes

JPV In Press

A non-relational data warehouse for the analysis of field and laboratory data from multiple heterogeneous photovoltaic test sites

Yang Hu, *Member, IEEE*, Venkat Yashwanth Gunapati, Pei Zhao, Devin Gordon, Nicholas R. Wheeler, Mohammad A. Hossain, *Member, IEEE*, Timothy J. Peshek, *Member, IEEE*, Laura S. Bruckman, Guo-Qiang Zhang, *Member, IEEE*, and Roger H. French, *Member, IEEE*

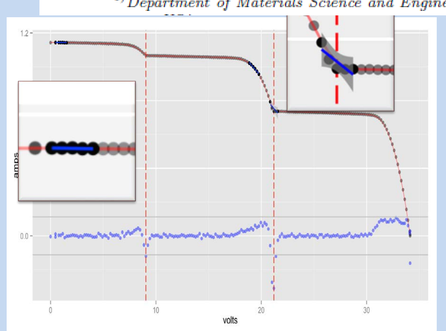
Metastability of PV Module Materials Identified by Machine Learning 2.2 million I-V curves[2][3]

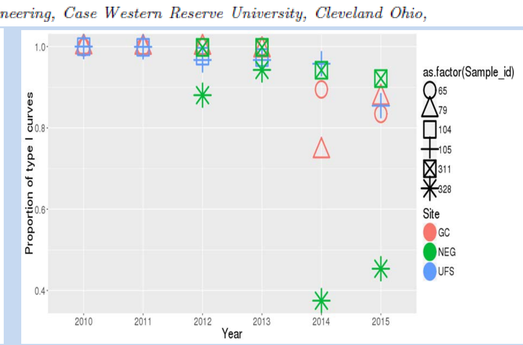


Insights into metastability of photovoltaic materials at the mesoscale through massive $I-V$ analytics

JVSTB

Timothy J. Peshek,[1, a)] Justin S. Fada,[1] Yang Hu,[1] Yifan Xu,[1] Mohamed A. Elsaeiti,[1] Erdmut Schnabel,[2] Michael Köhl,[2] and Roger H. French[1]
[1)] *Department of Materials Science and Engineering, Case Western Reserve University, Cleveland Ohio,*





Underlying Machine Learning Procedures:
Local linear regression fitting + Residual Thresholding
Classify I-V curve into five categories
- **Type I** : V_{oc} only
- **Type II** : V_{oc} + one bypass diode turns on
- **Type III** : V_{oc} + two bypass diodes turn on

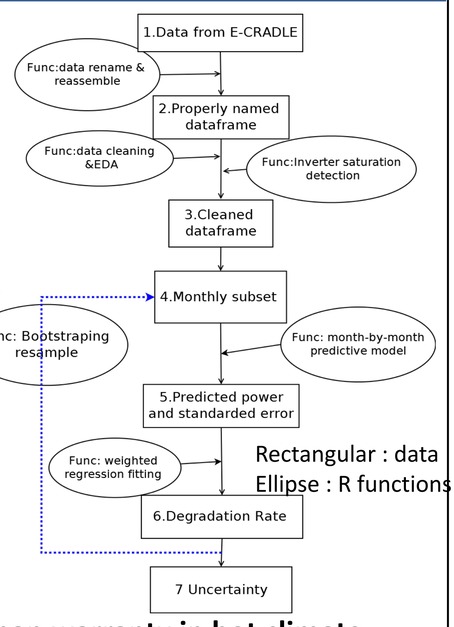
Detecting heterogeneity in PV modules from massive real-world "step" I-V curves: A machine learning approach

Yang Hu[1)], Erdmut Schnabel[2)], Michael Köhl[2)], Roger H. French[1)],and Timothy J. Peshek[1)]

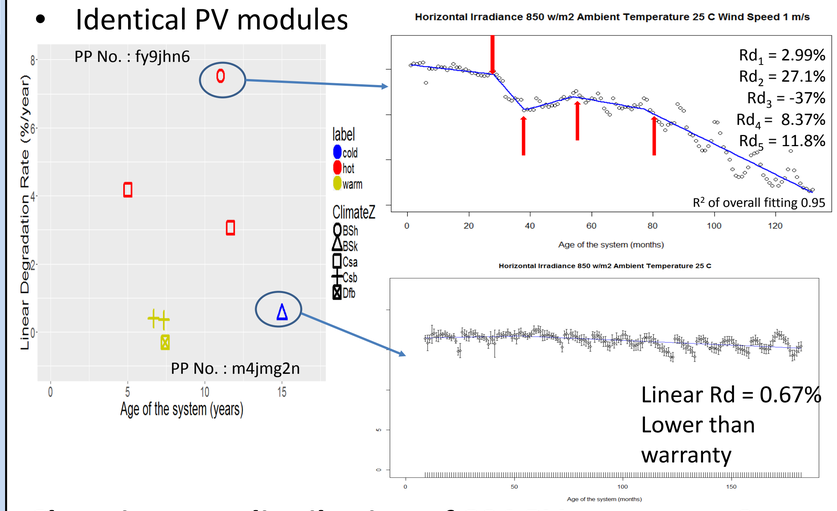
PVSC 2016 proceedings

Pipelined Time-Series Analysis Studies

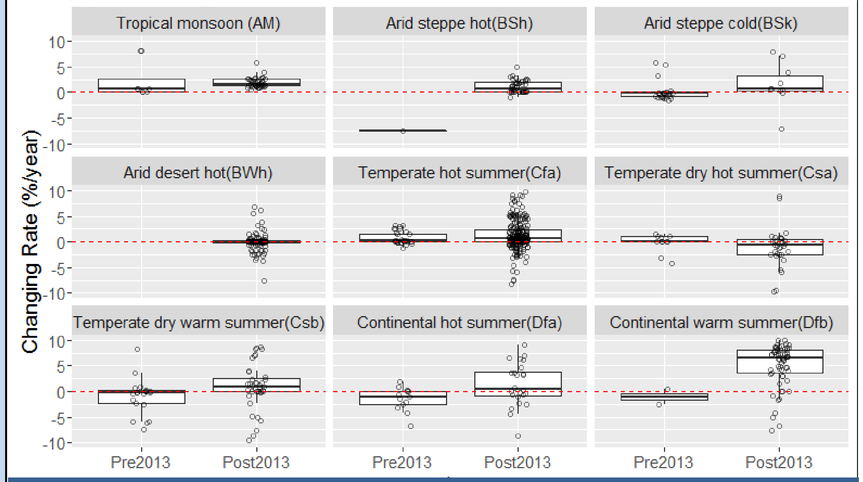
Month-by-Month Rd Model
- Use all data noon time, morning and afternoon
- Categorize data by age
- Every 30 days from the first operation date considered a month
- Develop predictive model for each month
- Linear regression models
- These models serve as a snapshot of the system status
- Use monthly regression models to normalize system performance under the same climate condition
- **Do not assume linear degradation rate**
- Look at the profile of the monthly predicted value
- Use bootstrap approach resample the data to estimate the uncertainty



PV systems degrade faster than warranty in hot climate
- Seven PV systems
- Identical PV modules



Changing rate distribution of 804 PV systems at 9 Koppen-Geiger climate zones



Acknowledgements

- This work is supported by the Bay Area Photovoltaic Consortium Prime Award No. DE-EE0004946, Subaward Agreement No. 60220829-51077-T.
- The SDLE center was established through funding through the Ohio Third Frontier, Wright Project Program Award Tech 12-004.

References

[1] Hu, Yang et. al " A non-relational data warehouse for the analysis of field and laboratory data from multiple heterogeneous photovoltaic test sites" Journal of Photovoltaics, IEEE in press
[2]Peshek, Timothy J., et al. "Insights into metastability of photovoltaic materials at the mesoscale through massive I–V analytics." *Journal of Vacuum Science & Technology B* 34.5 (2016): 050801.
[3] Hu, Yang et. Al "Detecting heterogeneity in PV modules from massive real-world "step" I-V curves: A machine learning approach." PVSC 2016 proceedings