Sandia National Laboratories

Battery Archive – A Standardized Data Repository to Accelerate Evaluation of Energy Storage Systems

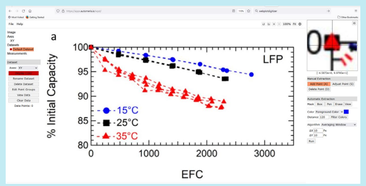
Lara Kassabian, Joseph Lubars, Yuliya Preger, Irving Derin, Steve Verzi, Venkat Subramanian, Valerio De Angelis

Contact: ypreger@sandia.gov, vdeange@sandia.gov

Background and Objectives

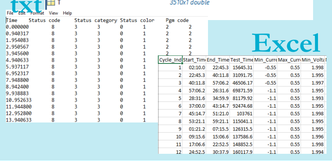
- **Problem**: Energy storage stakeholders require battery data to improve system performance and safety, but it is difficult to access, even from publicly funded studies. A significant amount of standardized data are needed to support AI/ML for performance/safety prediction.
- **State of the field**: In 2021, we launched Battery Archive as the first public repository for standardized battery data, focusing on commercial Li-ion cells. The site has been used by 50K+ from high schoolers to utility and Fortune 500 employees (top reason for access: get data to test their own battery state of health prediction with ML).
- **This work**: 1) Enable battery module data processing
  2) Evaluate best practices for battery machine learning for lifetime prediction across a variety of datasets

Before Battery Archive



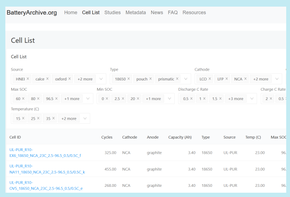
Matlab

txt



Excel

With Battery Archive

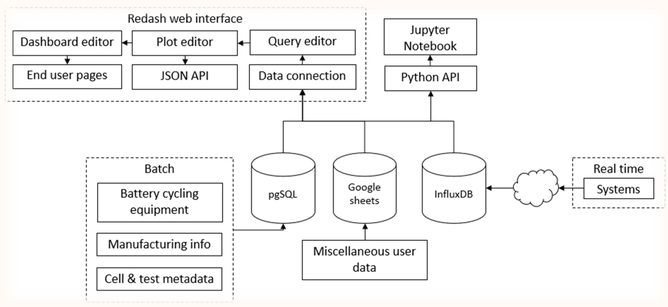


OE Mission Alignment: Enable more reliable critical infrastructure by aggregating data to support development of better models of energy storage system performance and safety.

Milestone 1: Extend Battery Archive Beyond Cell Data to Module Data

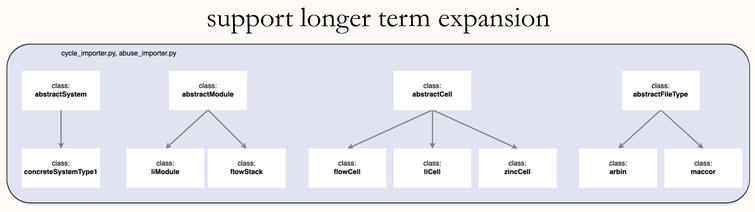
Battery Archive software package (issued Sandia copyright)

Representative model (system architecture is not tied to a single database schema or visualization tool)

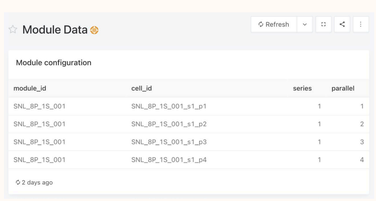


Methodology: Key considerations for module vs. cell data

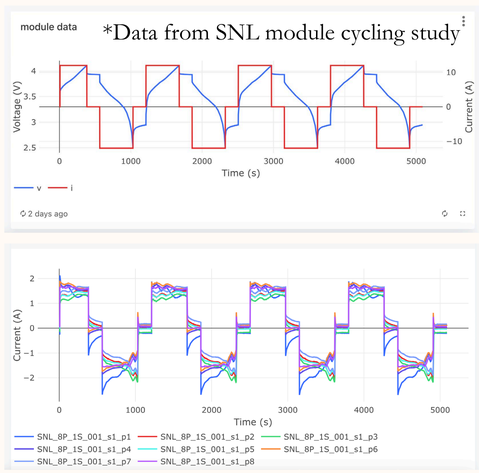
Restructuring importer code to deal with different data types and support longer term expansion



Database clearly conveying relationships and hierarchy of cells in modules



Results: Dashboards with module and cell data



Milestone 2: Evaluate Best Practices for AI/ML Lifetime Prediction Across a Variety of Datasets

Methodology

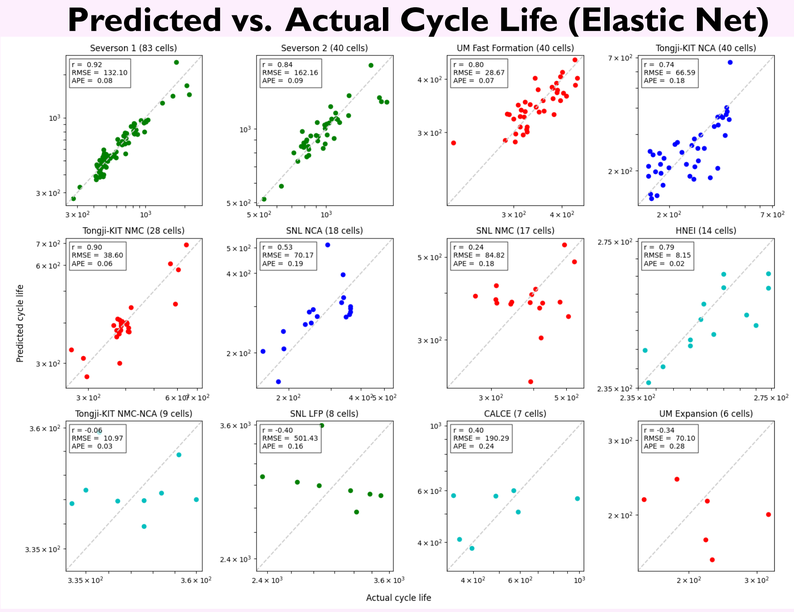
- **State of the field**: Most battery ML studies focus on demonstrating the lowest RMSE with a single dataset and highly tuned model
- **Our objective**: We evaluated the impact of datasets (different # of cells, variety of degradation mechanisms), models, features, and outliers on successful lifetime prediction.
- We pre-processed each dataset in a consistent manner by removing anomalous cycles
- Then, we calculated the cycle life as the number of cycles required to reach 85% of the initial discharge capacity
- The underlying code can be readily used to clean and generate features for new datasets as they are added to Battery Archive

Datasets

Dataset Name	Chemistry	Cells	Cycle Life Range
Severson 1	LFP	83	278-2156
Severson 2	LFP	40	524-1874
UM Fast Formation	NMC	40	209-439
Tongji-KIT NCA	NCA	40	155-419
Tongji-KIT NMC	NMC	28	263-642
SNL NCA	NCA	18	156-362
SNL NMC	NMC	17	270-527
HNEI	NMC-LCO	14	237-274
Tongji-KIT NMC-NCA	NMC-NCA	9	332-360
SNL LFP	LFP	8	2360-3500
CALCE	LCO	7	337-984
UM Expansion	NMC	6	153-321

Key Results

Predicted vs. Actual Cycle Life (Elastic Net)



Color Key: Green = LFP, Red = NMC, Blue = NCA, Cyan = Other

Innovation and Impact

- Battery Archive can now accommodate performance data (Li-ion single cell and module cycling) and safety data (mechanical indentation)
- Systematic ML across a variety of datasets from Battery Archive showed that successful battery lifetime prediction is highly dependent on the dataset, models, features, and outliers (publication under review)
- Having a standardized repository of public battery data has conservatively generated at least $100 million worth of value to users (based on cost of users generating similar data with a typical testing lab)

Future Work

- Extend Battery Archive to calorimetry data to enable AI/ML for prediction of safety events
  - Agree on standard data formats with leading battery safety institutions and secure initial datasets
  - Update software to accommodate new data type
- Continue uploading more examples of existing data types to Battery Archive

U.S. DEPARTMENT of ENERGY
NNSA National Nuclear Security Administration

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.
SAND No. SAND2026-16085PE

Acknowledgements: This material is based upon work supported by the U.S. Department of Energy, Office of Electricity (OE), Energy Storage Division.

Sandia National Laboratories