



Exceptional service in the national interest

HARMONIZING CALCULATIONS OF EXPECTED YIELD

Joshua S. Stein

Sandia National Laboratories

PVPMC Workshop, May 13-15, 2025



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

SAND2025-05780C



EXPECTED YIELD – WHY & HOW?

- **Goal:** Maximize energy production from a PV plant.
- **Problem:** How to quantify lost energy (recoverable vs. nonrecoverable)?
- **Financial Yield** (banks, insurers, asset owners/managers)
 - **Basic:** $P50 \times GHI_{\text{actual}} / GHI_{\text{TMY}}$
 - **Empirical:** Fit a regression model (e.g., PVUSA) on proforma model then calculate expected yield with actual weather
 - **Software:** re-run original proforma model with actual weather
- **Operational Yield** (performance engineers, analytics software)
 - Simple for anomaly detection: PVWatts, no clipping, default losses
 - Advanced for realistic system performance: root cause, underperformance





EXPECTED YIELD – WHY & HOW?

- **Goal:** Maximize energy production from a PV plant.
- **Problem:** How to quantify lost energy (recoverable vs. nonrecoverable)?
- **Financial Yield** (banks, insurers, asset owners/managers)
 - **Basic:** $P50 \times GHI_{\text{actual}} / GHI_{\text{TMY}}$
 - **Empirical:** Fit a regression model (e.g., PVUSA) on financial model then calculate expected yield with actual weather
 - **Software:** re-run original financial model with actual weather
- **Operational Yield** (performance engineers, analytics software)
 - Simple for anomaly detection: PVWatts, no clipping, default losses
 - Advanced for realistic system performance: root cause, underperformance

We surveyed ~20 stakeholders how they calculate expected yield.

We implemented several methods and will compare the results.



MODELS AND METHODS



Compare results of the two approaches on a well monitored system at Sandia

Models considered include...

- **PVWatts**

- $P_{dc} = P_{stc} * (G/1000) * (1 + \gamma * (T_{mod} - 25))$

- **PVUSA**

- $P_{ac} = G * (A + B * G + C * T_{amb} + D * WS)$

- **ML Models (Random Forest, Neural Network)**

- Feature sets include:

- GHI, T_{amb} , WS
 - POA, T_{amb} , WS

Method

Fit

- Fit model to proforma yield

Evaluate

- Run fitted model with **measured weather** (1-min) to estimate **expected yield**.

Compare

- Compare **expected yield** to **measured yield** to evaluate errors.

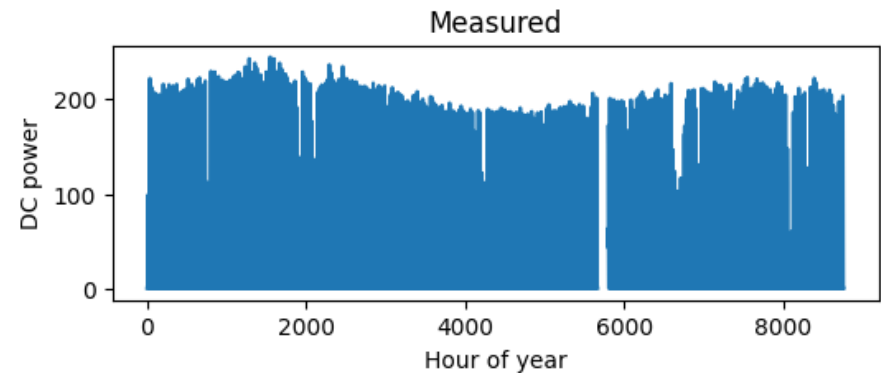
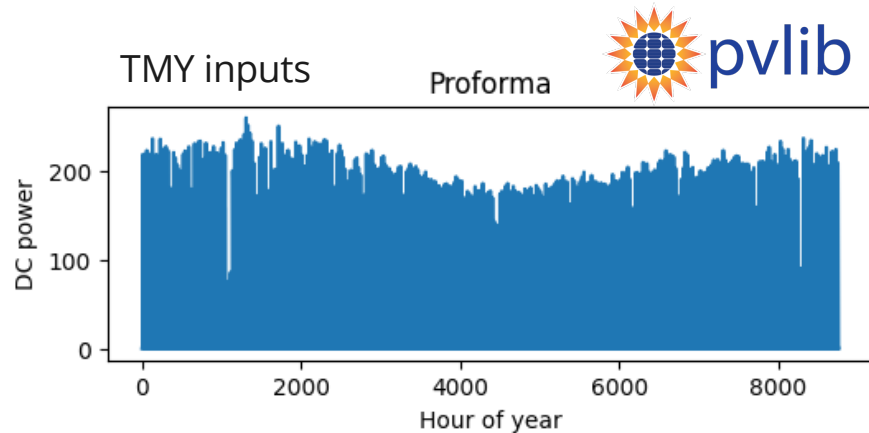
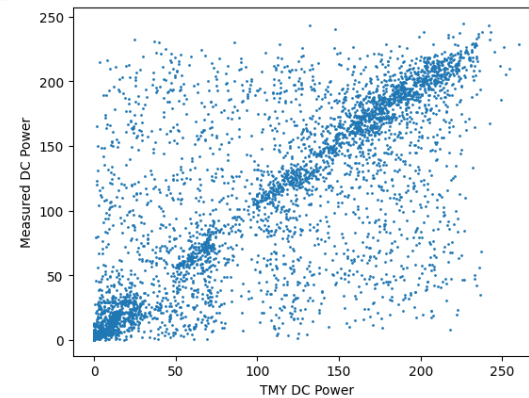
*This comparison **does not** assume that selected feature sets or ML-model architecture are optimized*



PROFORMA YIELD VS. MEASURED YIELD

We use a generic pvlb model scaled to a measured system in Albuquerque

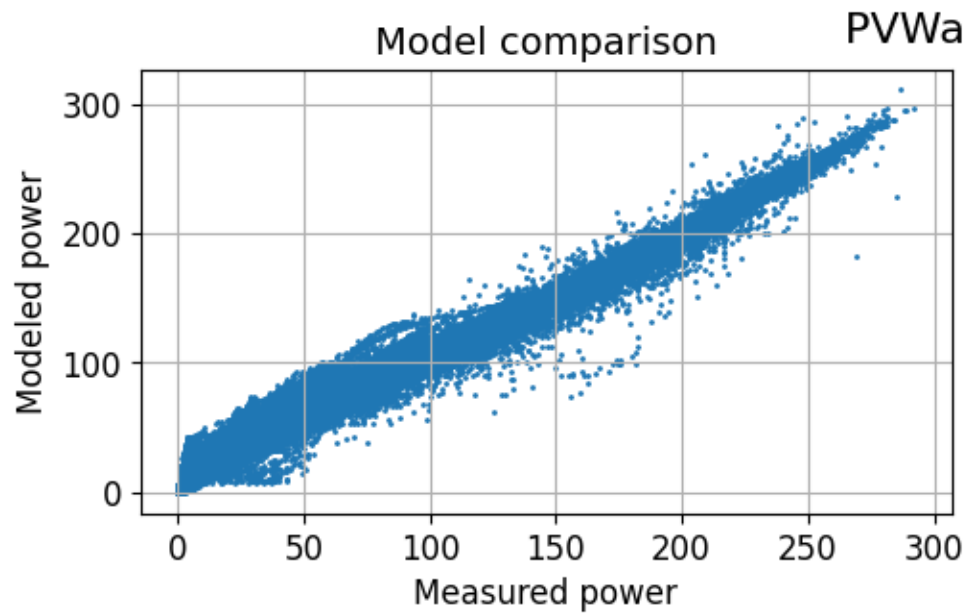
- Scale the results to approximately match the annual yield on a real system at Sandia.
- Filter out nighttime and obvious shading of POA sensor.
- PV system has no clipping and very little row-row shading (less complex than typical utility-scale plants).



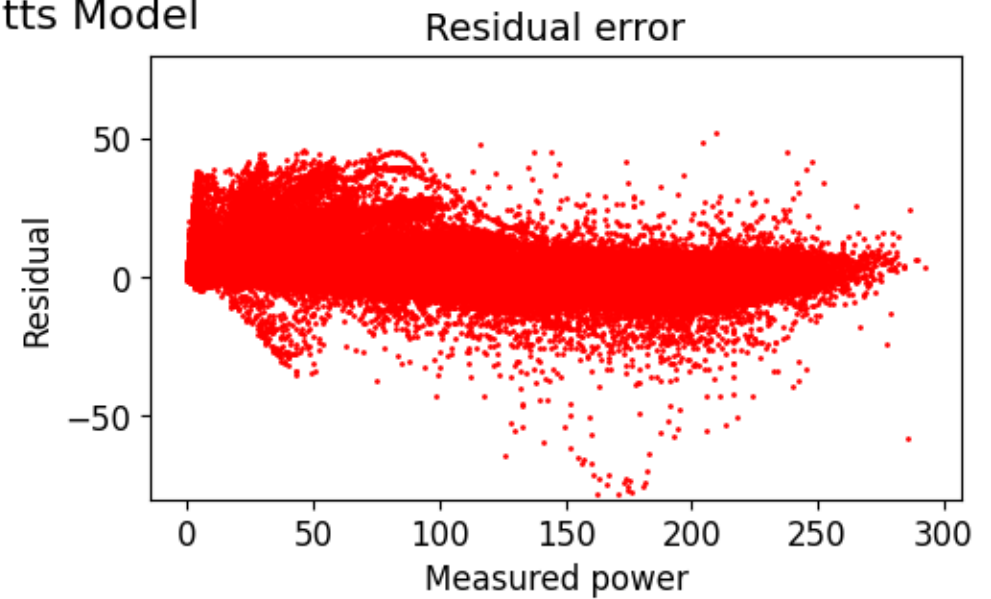
PV WATTS COMPARISON



PVWatts shows only slight systematic bias



PVWatts Model

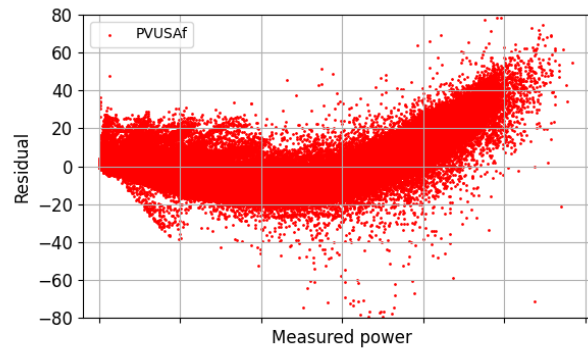
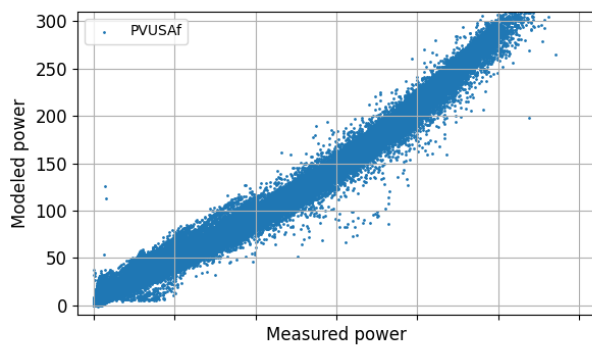


PVUSA COMPARISON

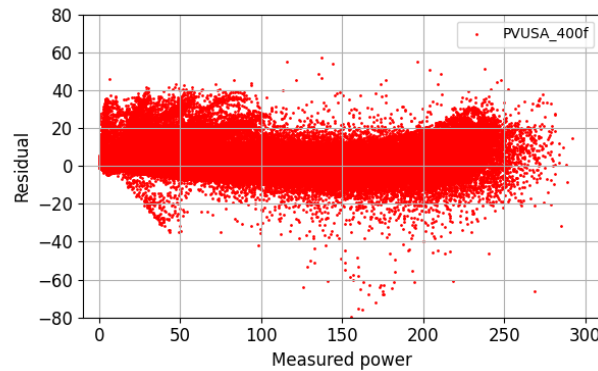
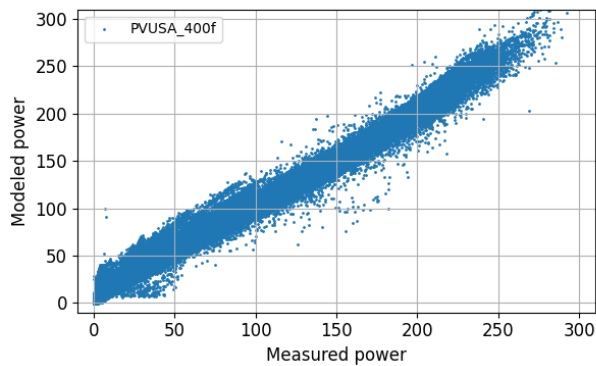


PVUSA shows systematic bias that is reduced by fitting to only high irradiance periods (POA>400 W/m²)

PVUSA Models



unfiltered



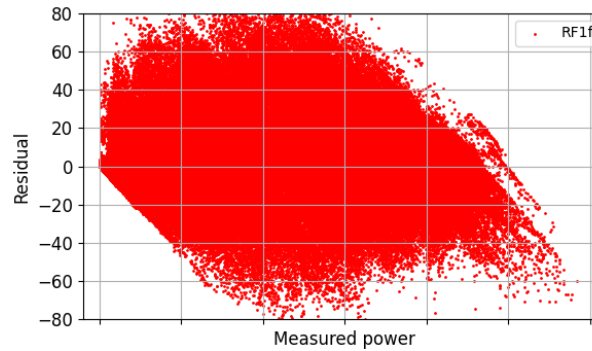
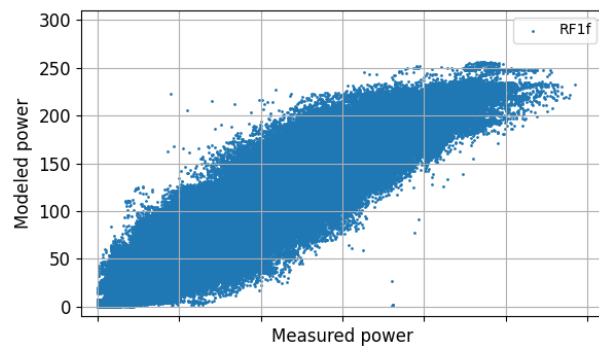
POA>400

ML - RANDOM FOREST

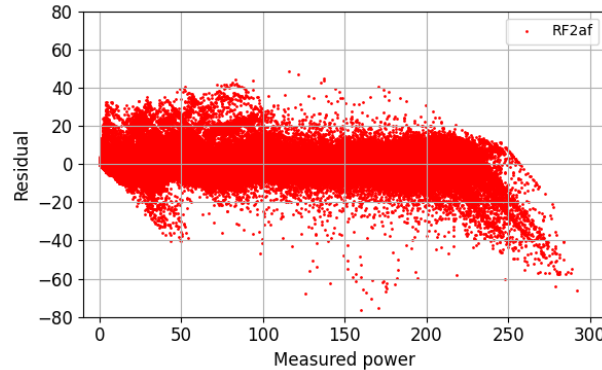
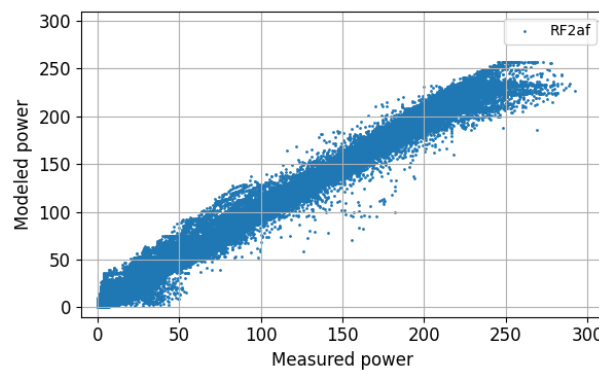


POA irradiance greatly increases model accuracy compared with GHI

Random Forest Models



GHI, T_{amb} , WS



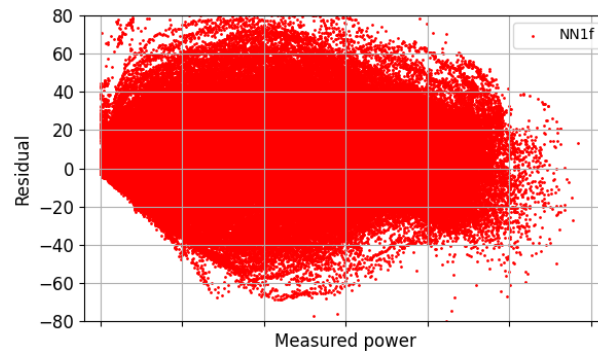
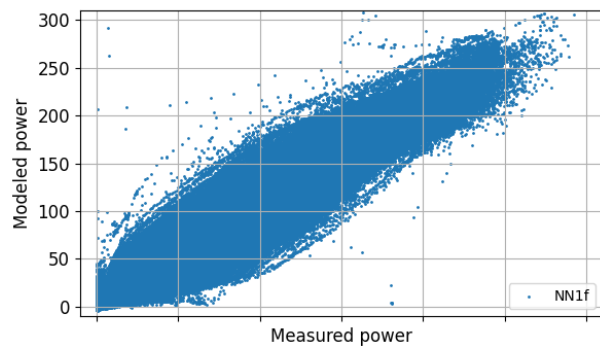
POA, T_{amb} , WS

ML – NEURAL NETWORK

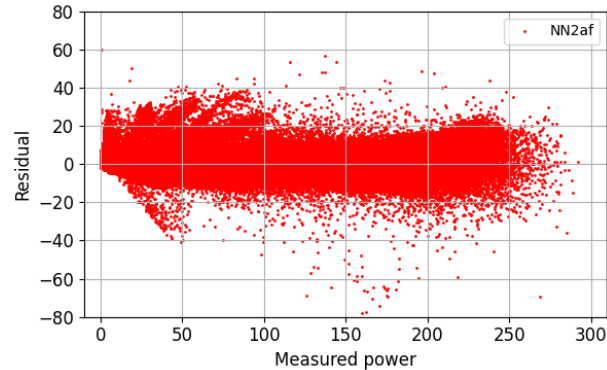
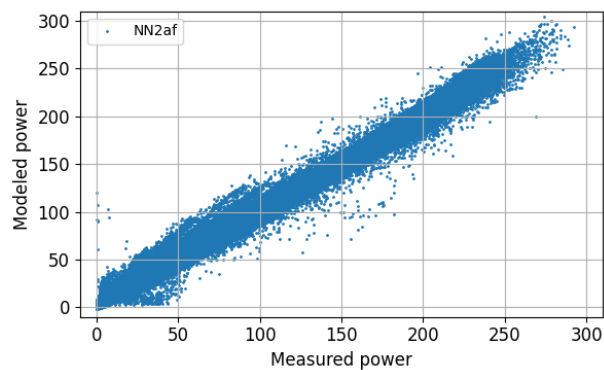


ML models show very little systematic bias

NN Models



GHI, T_{amb} , WS



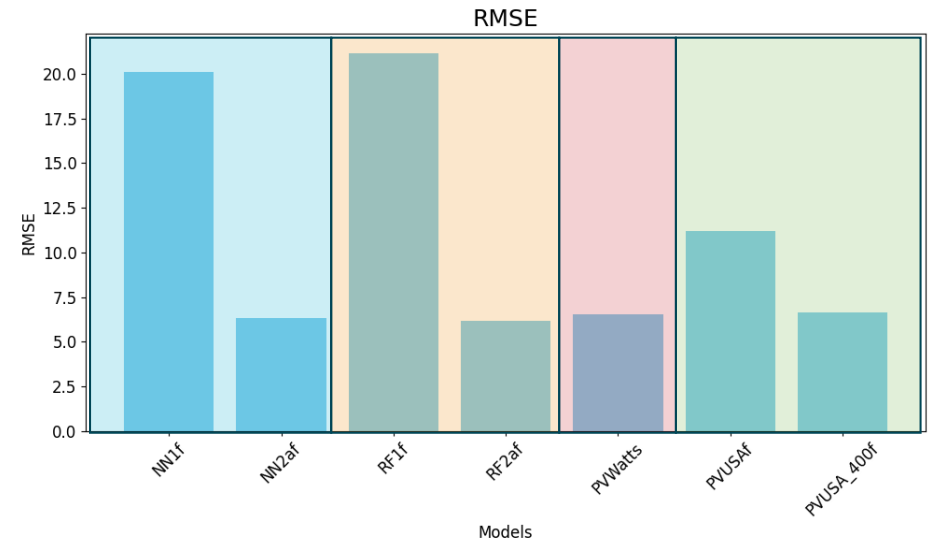
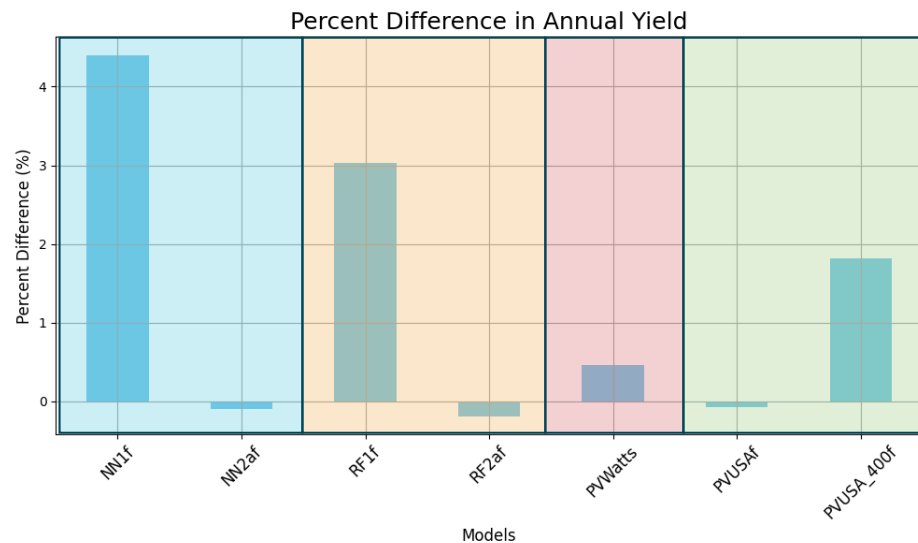
POA, T_{amb} , WS

ERROR COMPARISON



Bias and random errors are not necessarily correlated

- ML models trained to POA, T_{amb} , WS have lowest bias errors.
- Filtered PVUSA has higher bias errors and lower RMSE
- PVWatts performs well in both bias and RMSE



CONCLUSIONS



Simple model comparison demonstrates significant differences in expected yield.

- Choice of model significantly affects the expected yield.
 - Fixed equation models (e.g., PVWatts & PVUSA) introduce more systematic errors than data-driven ML-models (e.g., random forest or neural network).
 - Filtering training data affects bias and random errors differently
 - Obvious→ Using POA irradiance as a feature is much better than GHI
 - Model differences in expected yield contribute to variations between PV analytics tools.
 - Industry would benefit from more standardization in this area.