Open Source Machine Learning Applied to PV Monitoring Applications

SAND2016-4645C

PV Systems Symposium
Santa Clara, CA May 10, 2016

Joshua Stein, PhD.
Birk Jones, PhD.
Sandia National Laboratories, New Mexico (USA)

PVPerformance MODELING COLLABORATIVE

Exceptional service in the national interest

Sandia National Laboratories

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

SAND2016-XXXX C

Introduction

Motivation

1. Advanced PV monitoring
2. Embed most accurate Machine Learning Techniques into monitoring systems
3. Low cost embedded devices can be deployed easily
4. Open-source code for multiple applications

Machine Learning Advantages

1. Rapid adaptable learning, adaptable
2. Design information not needed
3. Accurate results

Presentation

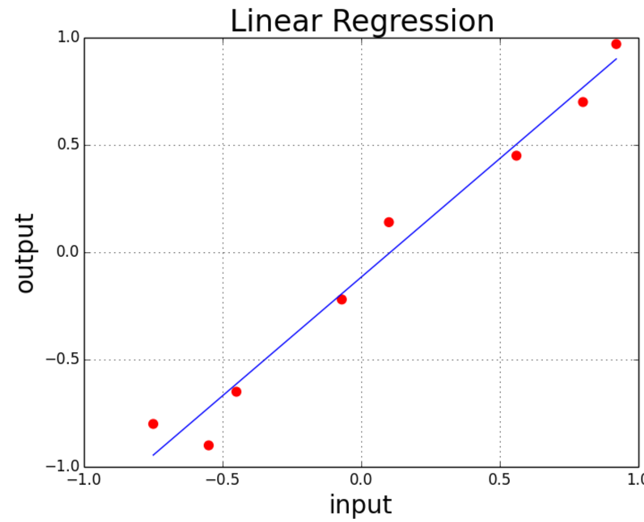
1. Experiment Setup
2. Machine Learning Techniques
3. Experiment Results
4. Demonstration of embedded system description
5. Conclusions

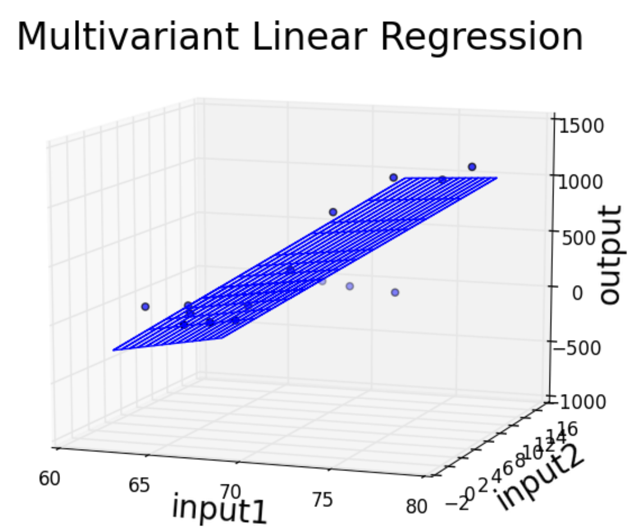
Experiment Set-up

- Purpose
  - Analyze PV data (voltage, current, and power)
  - Apply algorithms to estimate outputs
- Methodology
  - Compare Two Algorithms
    - Least Squares Regression
    - Gaussian Process Regression
  - Training Data
    - Variable test days from 1 – 15 days
    - POA irradiance (pyranometer) and module temperature
    - Data Set Types
      - Complete set (1 minute data)
      - 33% of data randomly selected
      - Histogram bins set based on irradiance and cell temperature statistics (IEC 61853-1)
  - Testing Data is constant
    - 10 days at one minute intervals

Least-Squares Linear Regression

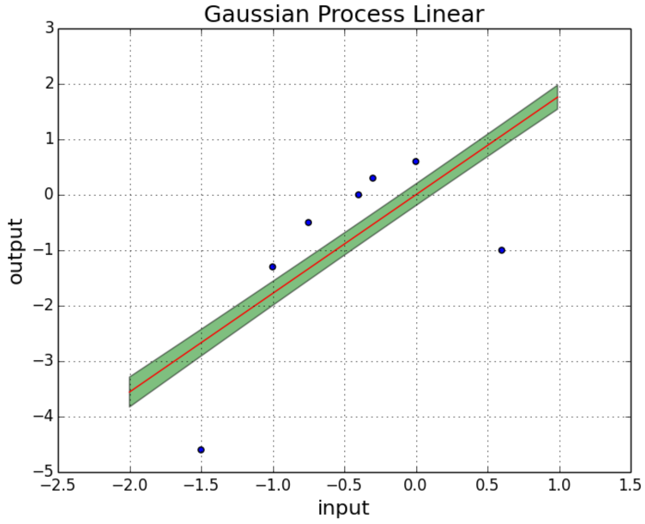
- Common form of Machine Learning
  - Relationship between:
    - One dependent variable and
    - One or more independent variables
  - Solution minimizes:
    - $\sum(\hat{y} - y)^2$
  - Gradient Descent that discovers the best:
    - B_0, β_1
  - Goal fit a function to the data set
    - $f(x,\beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$
- Applied to PVUSA Rating Methodology
  - Present study only uses POA irradiance and module temp (no wind speed)

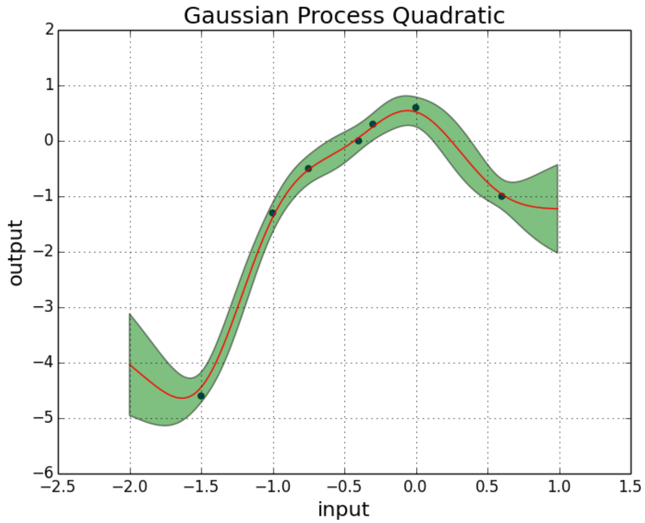




Gaussian Process Regression

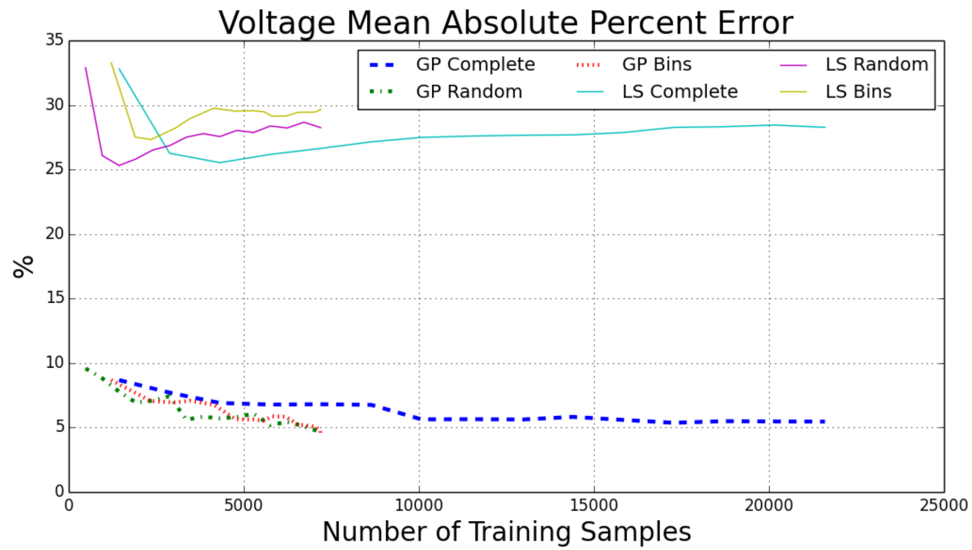
- *GP is the generalization of the Gaussian probability distribution*
- *Parameterized by:*
  - *Mean function* – $\mu(x)$
  - *Covariance function (kernel)* - $K(x, x')$
  - $f(x) \sim \mathcal{GP}(\mu(x), K(x, x'))$
- *Gaussian Process Regression*
  - *Predictive distribution*
    - $p(y_*|x_* X, y) = \mathcal{N}(\mu_*, {\sigma_*}^2)$
  - *where*
    - $\mu_* = K_{*N}(K_N + \sigma^2 I)^{-1} y$
    - ${\sigma_*}^2 = K_{**} - K_{*N}(K_N + \sigma^2 I)^{-1} K_{*N} + \sigma^2$
    - *and*
      - K_N *(covariance)*
      - K_{*N} *(training set covariance)*
      - K_{**} *(training-test set covariance)*

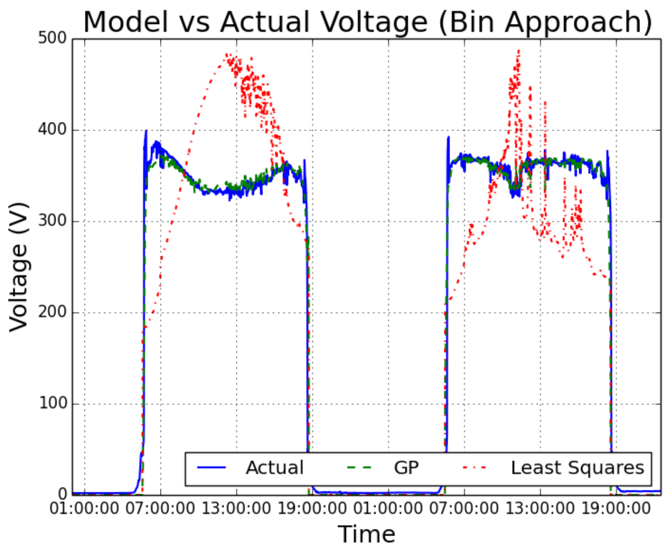


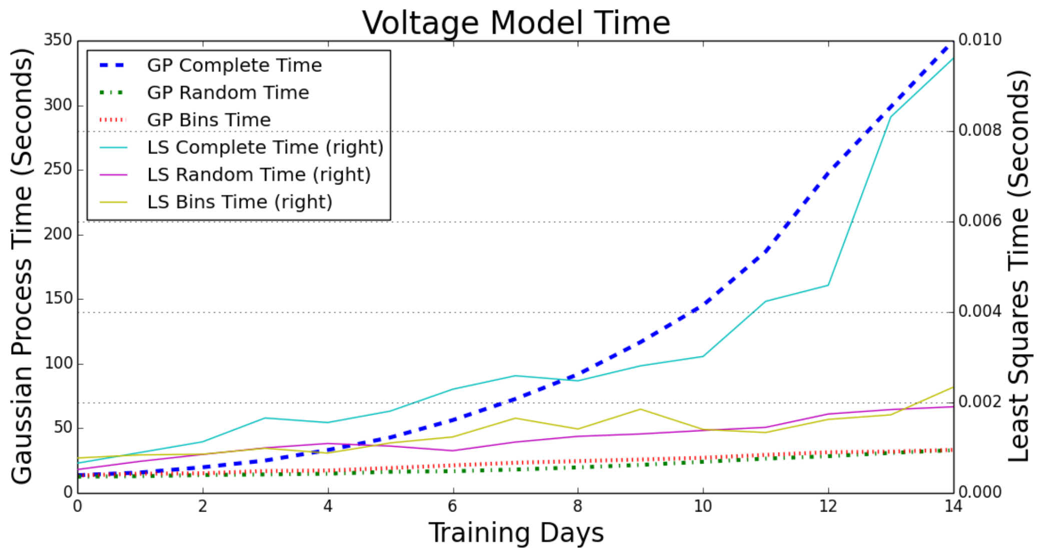


Voltage Model Results

- GP converges to 5% Mean Absolute Percent Error (MAPE)
- LS MAPE over 25%
- Random and statistical bin sampling have similar results

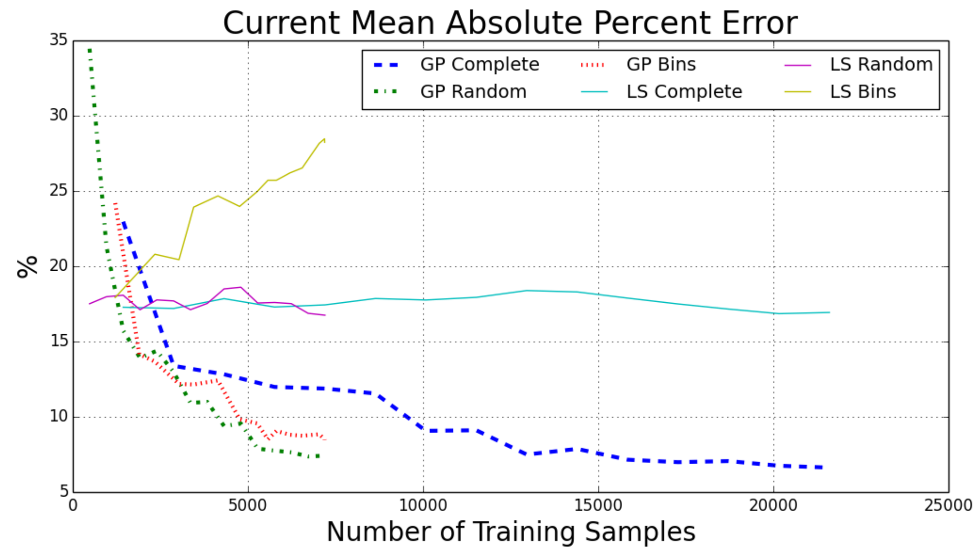


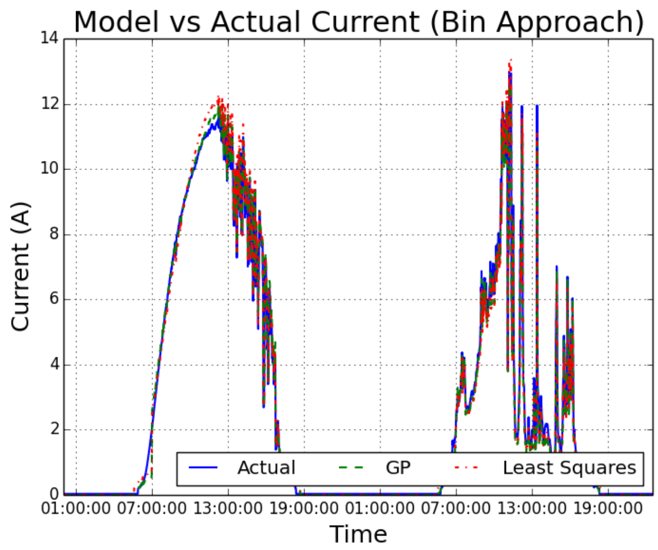


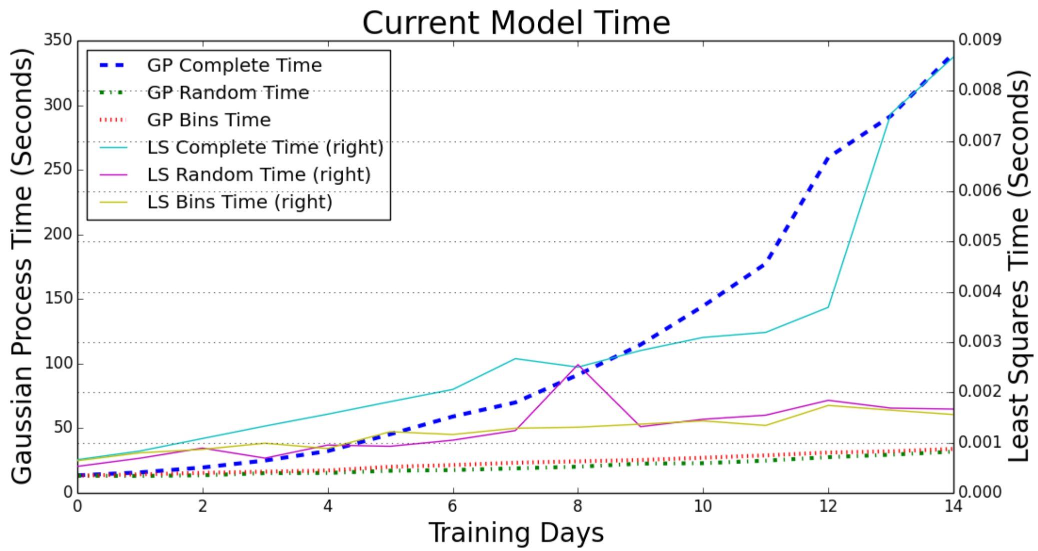


Current Model Results

- GP converges to 6-7% MAPE
- LS MAPE over 15%
- Random and statistical bin sampling have similar results

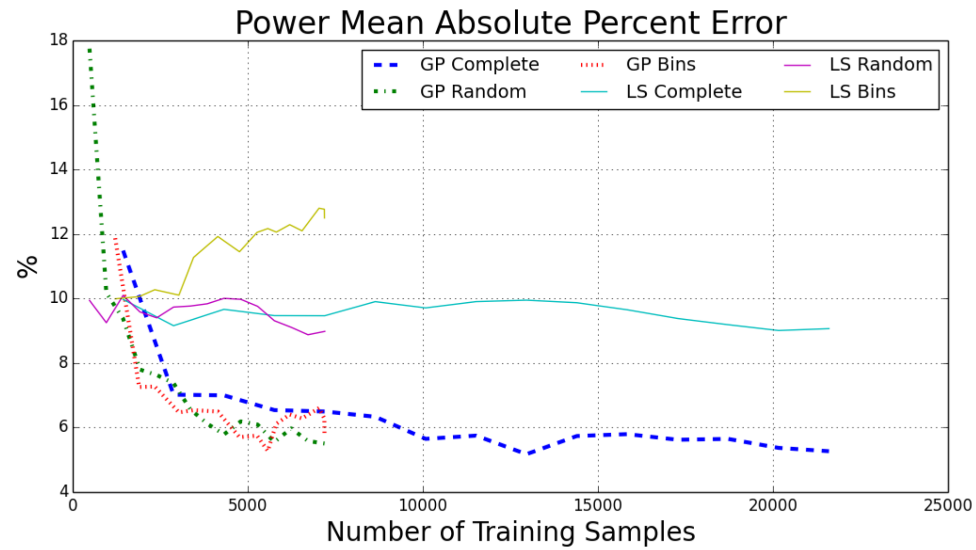


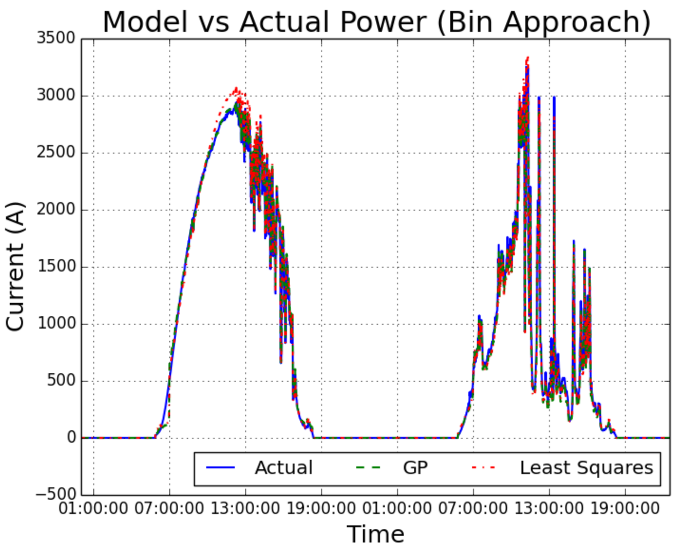


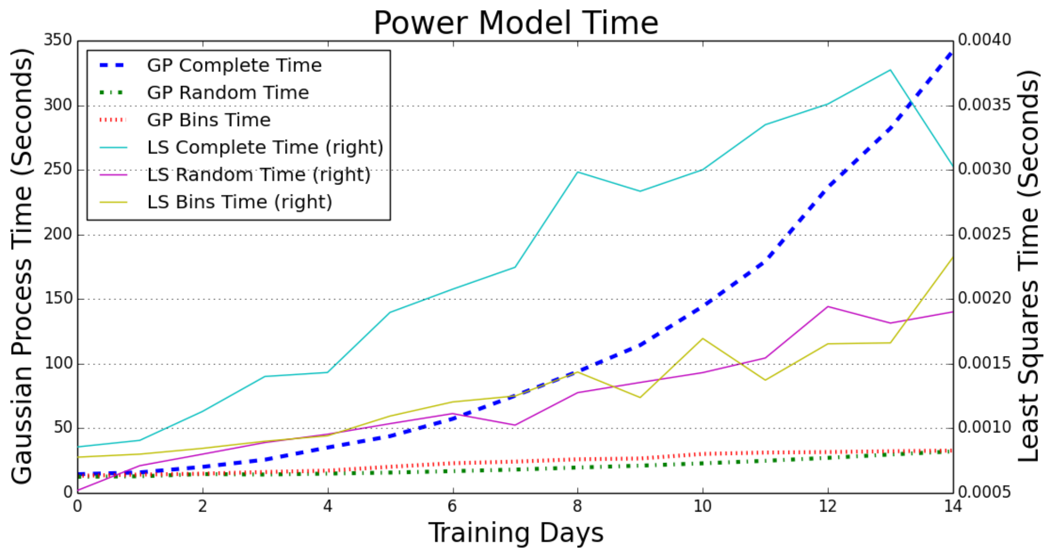


Power Model Results

- GP converges to about 5% MAPE
- LS MAPE over 8%
- Random and statistical bin sampling have similar results







Demonstration

Sandia National Laboratories

Embedded Device

- Raspberry Pi
- Modbus connection (Python code)
- MySQL database
- Gaussian Process Estimate (Python code)
- Web-based interface
- Material Cost: < $300

Query real-time results through web-based interface

Conclusions & Future Work

Conclusion

- Gaussian Process provides an accurate and reliable model of PV parameters (Voltage, Current, & Power)
- The algorithm can be embedded into monitoring systems
- Inputs are flexible

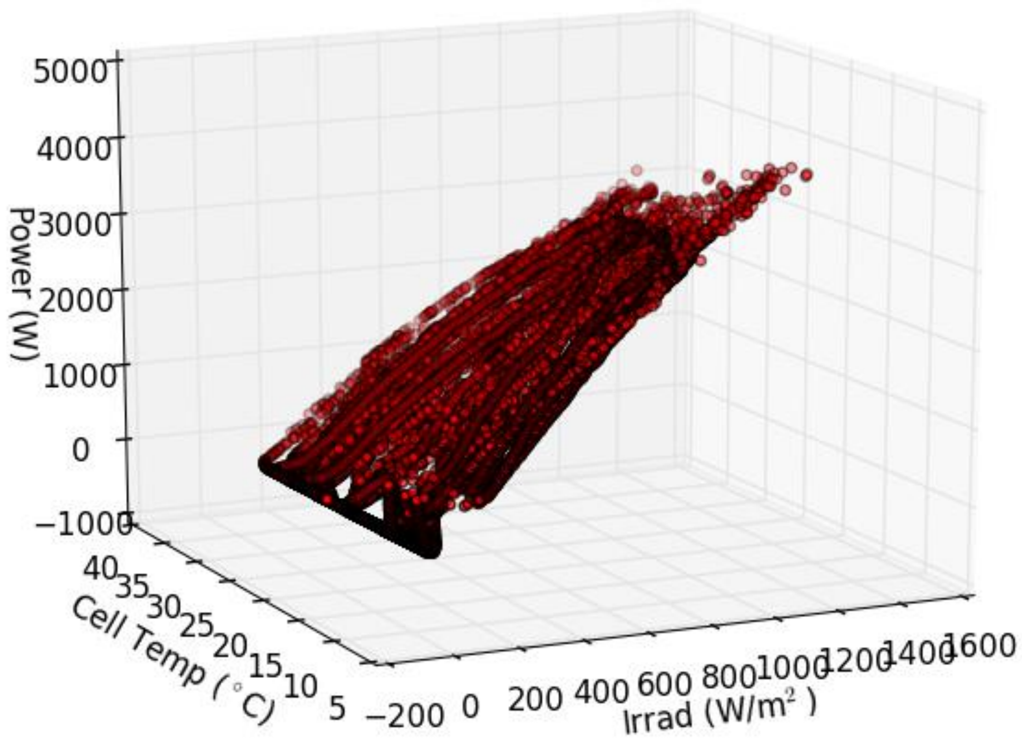
Future Work

- PVSC presentation & paper
- Open source distribution
- Interested partnerships??

Joshua S. Stein
jsstein@sandia.gov

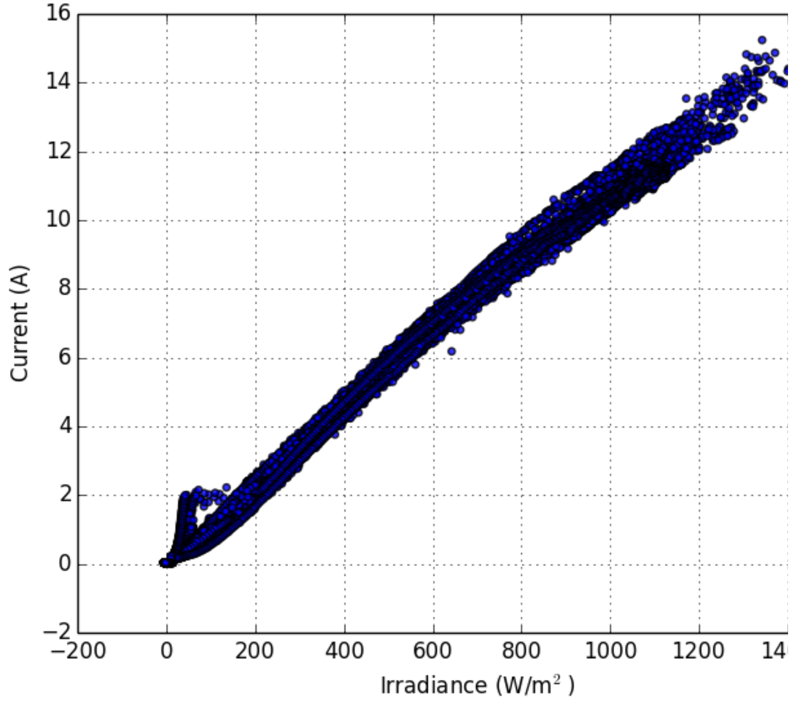
C. Birk Jones
cbjones@sandia.gov

Extra

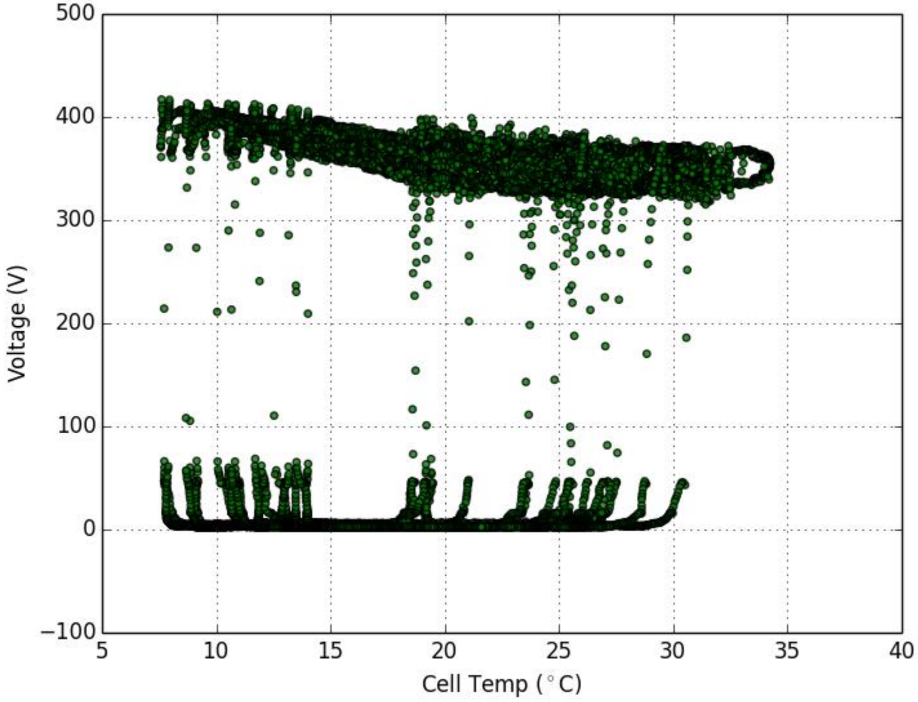


Power (W)
5000
4000
3000
2000
1000
0
−1000
Cell Temp (°C)
40
35
30
25
20
15
10
5
−200
0
200
400
600
800
1000
1200
1400
1600
Irrad (W/m²)

Extra



16
14
12
10
8
6
4
2
0
−2
−200
0
200
400
600
800
1000
1200
14
Current (A)
Irradiance (W/m²)



500
400
300
200
100
0
−100
5
10
15
20
25
30
35
40
Voltage (V)
Cell Temp (°C)