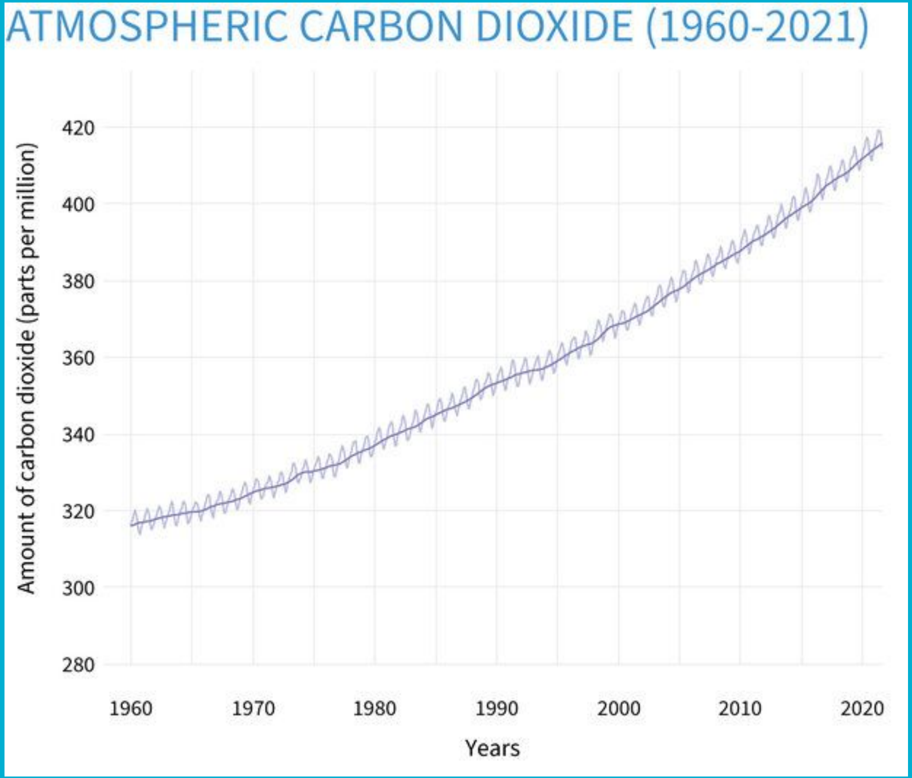
Making solar power affordable: Using data to maximize cost efficiency

By Zachary Brown

Carbon emissions over time



Source: climate.gov

Project goal

To provide recommendations to homeowners in Texas that will help maximize the cost efficiency of their solar panel installation

How to make solar panels affordable

- Purchase the largest configuration of solar panels that makes sense for the house
- Design the solar panel installation with a relatively high inverter loading ratio
- Identify and secure any rebate or grant available
- Consider buying a solar panel model with lower conversion efficiency
- Consider scheduling the installation for July or December

How to make solar panels affordable

- Purchase the largest configuration of solar panels that makes sense for the house
- Design the solar panel installation with a relatively high inverter loading ratio
- Identify and secure any rebate or grant available
- Consider buying a solar panel model with lower conversion efficiency
- Consider scheduling the installation for July or December

How to make solar panels affordable

- Purchase the largest configuration of solar panels that makes sense for the house
- Design the solar panel installation with a relatively high inverter loading ratio
- Identify and secure any rebate or grant available
- Consider buying a solar panel model with lower conversion efficiency
- Consider scheduling the installation for July or December

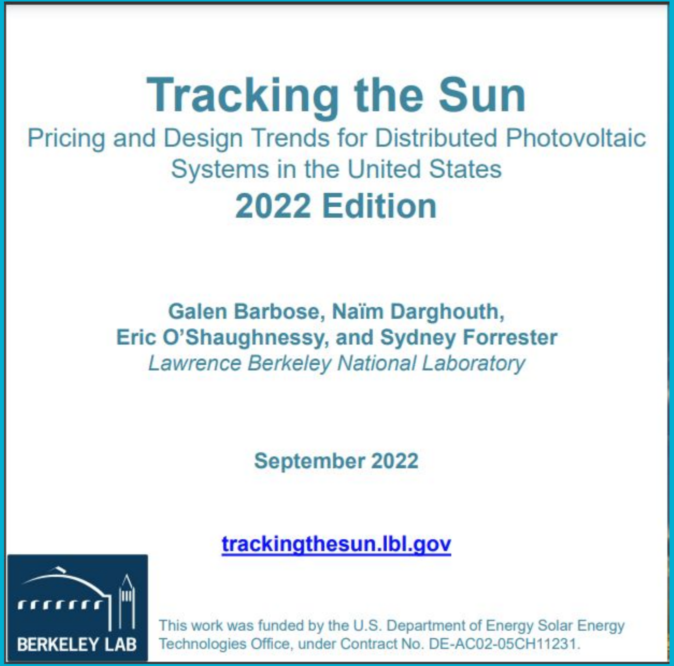
How to make solar panels affordable

- Purchase the largest configuration of solar panels that makes sense for the house
- Design the solar panel installation with a relatively high inverter loading ratio
- Identify and secure any rebate or grant available
- Consider buying a solar panel model with lower conversion efficiency
- Consider scheduling the installation for July or December

How to make solar panels affordable

- Purchase the largest configuration of solar panels that makes sense for the house
- Design the solar panel installation with a relatively high inverter loading ratio
- Identify and secure any rebate or grant available
- Consider buying a solar panel model with lower conversion efficiency
- Consider scheduling the installation for July or December

The data



Tracking the Sun

Pricing and Design Trends for Distributed Photovoltaic Systems in the United States

2022 Edition

Galen Barbose, Naïm Darghouth,
Eric O'Shaughnessy, and Sydney Forrester
Lawrence Berkeley National Laboratory

September 2022

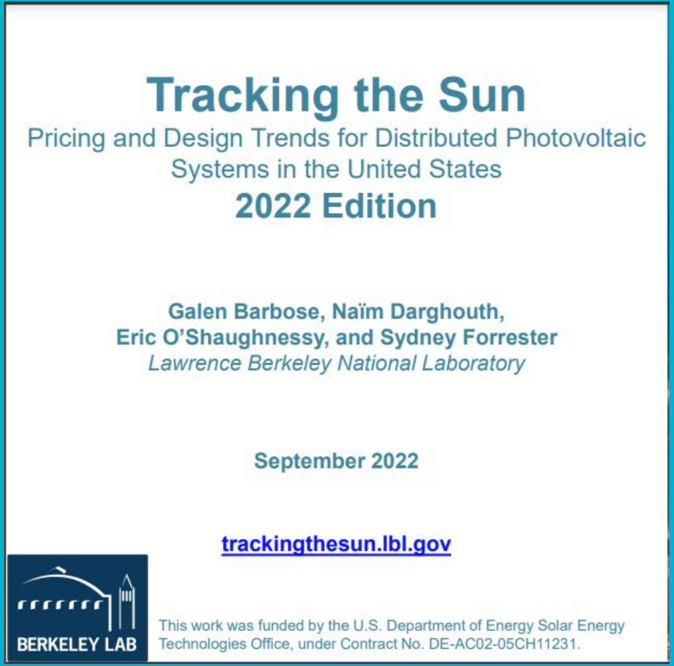
trackingthesun.lbl.gov

BERKELEY LAB

This work was funded by the U.S. Department of Energy Solar Energy Technologies Office, under Contract No. DE-AC02-05CH11231.

- Residential installations only
- 2020/2021 installations
- 26 states included in the data, all were used in this analysis

The data



Tracking the Sun

Pricing and Design Trends for Distributed Photovoltaic Systems in the United States

2022 Edition

Galen Barbose, Naïm Darghouth, Eric O'Shaughnessy, and Sydney Forrester

Lawrence Berkeley National Laboratory

September 2022

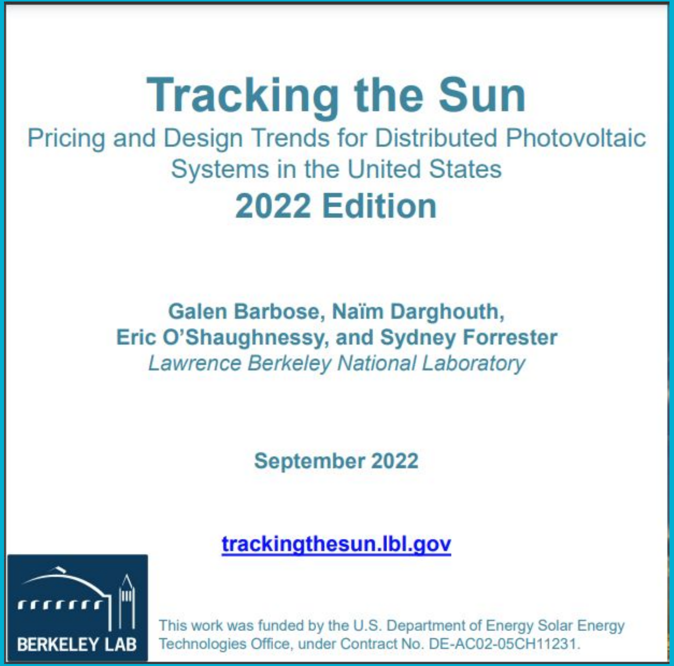
trackingthesun.lbl.gov

BERKELEY LAB

This work was funded by the U.S. Department of Energy Solar Energy Technologies Office, under Contract No. DE-AC02-05CH11231.

- Residential installations only
- 2020/2021 installations
- 26 states included in the data, all were used in this analysis

The data



Tracking the Sun

Pricing and Design Trends for Distributed Photovoltaic Systems in the United States

2022 Edition

Galen Barbose, Naïm Darghouth, Eric O'Shaughnessy, and Sydney Forrester

Lawrence Berkeley National Laboratory

September 2022

trackingthesun.lbl.gov

BERKELEY LAB

This work was funded by the U.S. Department of Energy Solar Energy Technologies Office, under Contract No. DE-AC02-05CH11231.

- Residential installations only
- 2020/2021 installations
- 26 states included in the data, all were used in this analysis

Cost Efficiency

Relevant fields:

- Total installed price
- System size (DC)
- Rebate or grant

Cost efficiency metric:

- Price per KW

$$\text{Price per KW} = \frac{\text{total installed price - rebate or grant}}{\text{system size (DC)}}$$

Feature Engineering

1. **Dummy Variables**
   Categorical features with more than 30 instances dummied
2. **Train-Test Split**
   Dataset split into 75% training set and 25% test set
3. **Missing Value Imputation**
   Simple imputer using most common value fit on train set, applied to train and test sets
4. **Feature Scaling**
   Standard scaler fit to train data, applied to train and test sets
5. **Feature Selection**
   Select 400 best features with f-regression trained on train, applied to train and test

Feature Engineering

1. **Dummy Variables**
   Categorical features with more than 30 instances dummied

2. **Train-Test Split**
   Dataset split into 75% training set and 25% test set

3. **Missing Value Imputation**
   Simple imputer using most common value fit on train set, applied to train and test sets

4. **Feature Scaling**
   Standard scaler fit to train data, applied to train and test sets

5. **Feature Selection**
   Select 400 best features with f-regression trained on train, applied to train and test

Feature Engineering

1. **Dummy Variables**
   Categorical features with more than 30 instances dummied

2. **Train-Test Split**
   Dataset split into 75% training set and 25% test set

3. **Missing Value Imputation**
   Simple imputer using most common value fit on train set, applied to train and test sets

4. **Feature Scaling**
   Standard scaler fit to train data, applied to train and test sets

5. **Feature Selection**
   Select 400 best features with f-regression trained on train, applied to train and test

Feature Engineering

1. **Dummy Variables**
   Categorical features with more than 30 instances dummied
2. **Train-Test Split**
   Dataset split into 75% training set and 25% test set
3. **Missing Value Imputation**
   Simple imputer using most common value fit on train set, applied to train and test sets
4. **Feature Scaling**
   Standard scaler fit to train data, applied to train and test sets
5. **Feature Selection**
   Select 400 best features with f-regression trained on train, applied to train and test

Feature Engineering

1. **Dummy Variables**
   Categorical features with more than 30 instances dummied

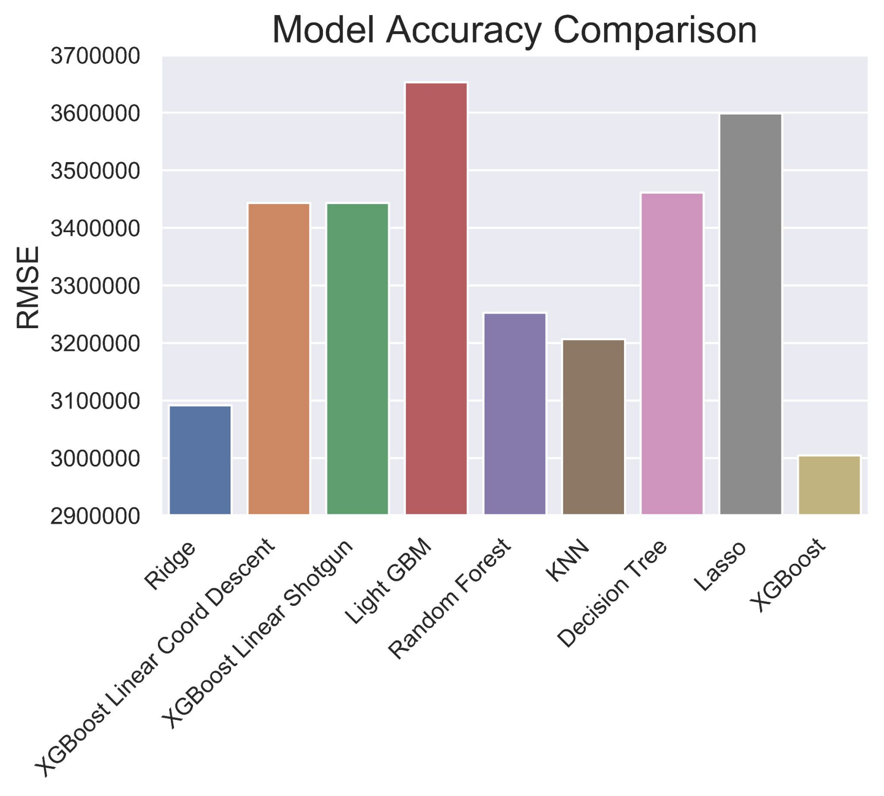
2. **Train-Test Split**
   Dataset split into 75% training set and 25% test set

3. **Missing Value Imputation**
   Simple imputer using most common value fit on train set, applied to train and test sets

4. **Feature Scaling**
   Standard scaler fit to train data, applied to train and test sets

5. **Feature Selection**
   Select 400 best features with f-regression trained on train, applied to train and test

Initial Model Screening

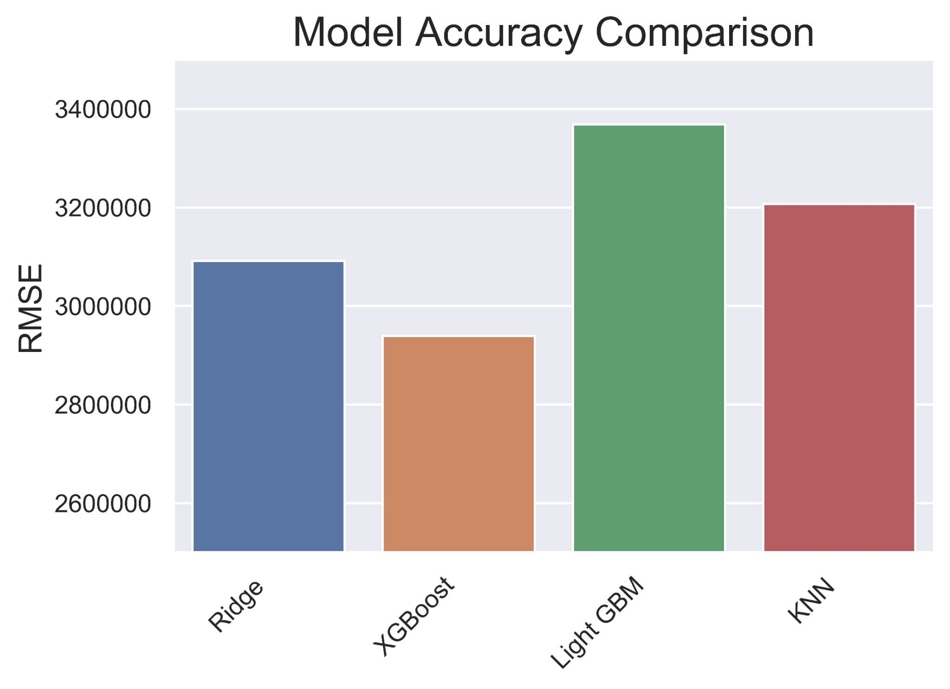


- Hyperparameter tuned models on 10% train set
- Retrained on 80% of data with set hyperparameters
- Tested on remaining 20% of dataset

Further Tuning

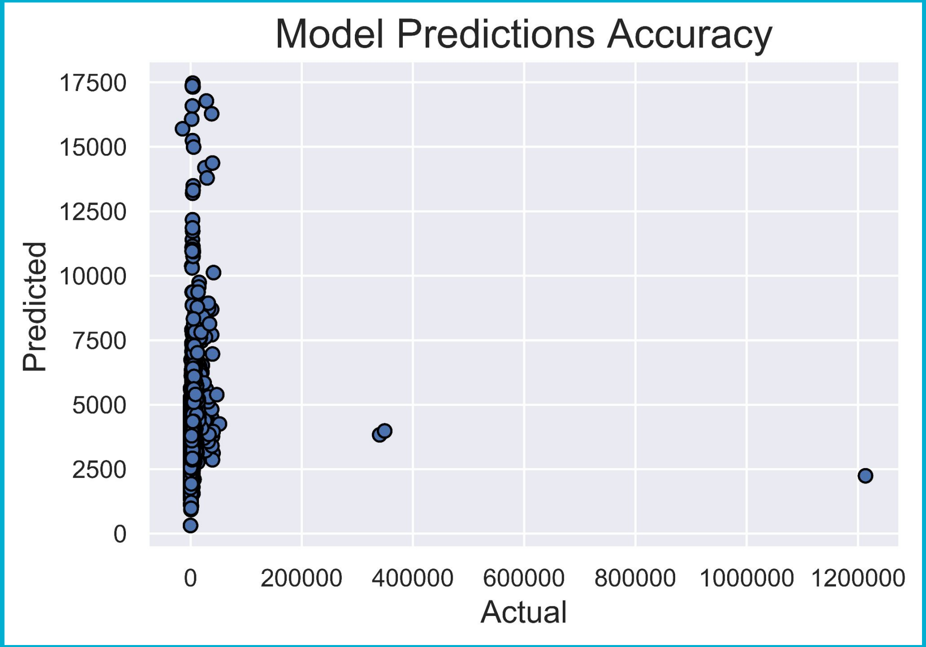
- Four models hyperparameter tuned with 80% of the data for training
    - Two of four hyperparameters locked for XGBoost and Light GBM

- XGBoost had the best performance and was used for the remainder of the project



Model Performance

XGBoost Regressor	RMSE (million)
Training	2.96
Test	35.05

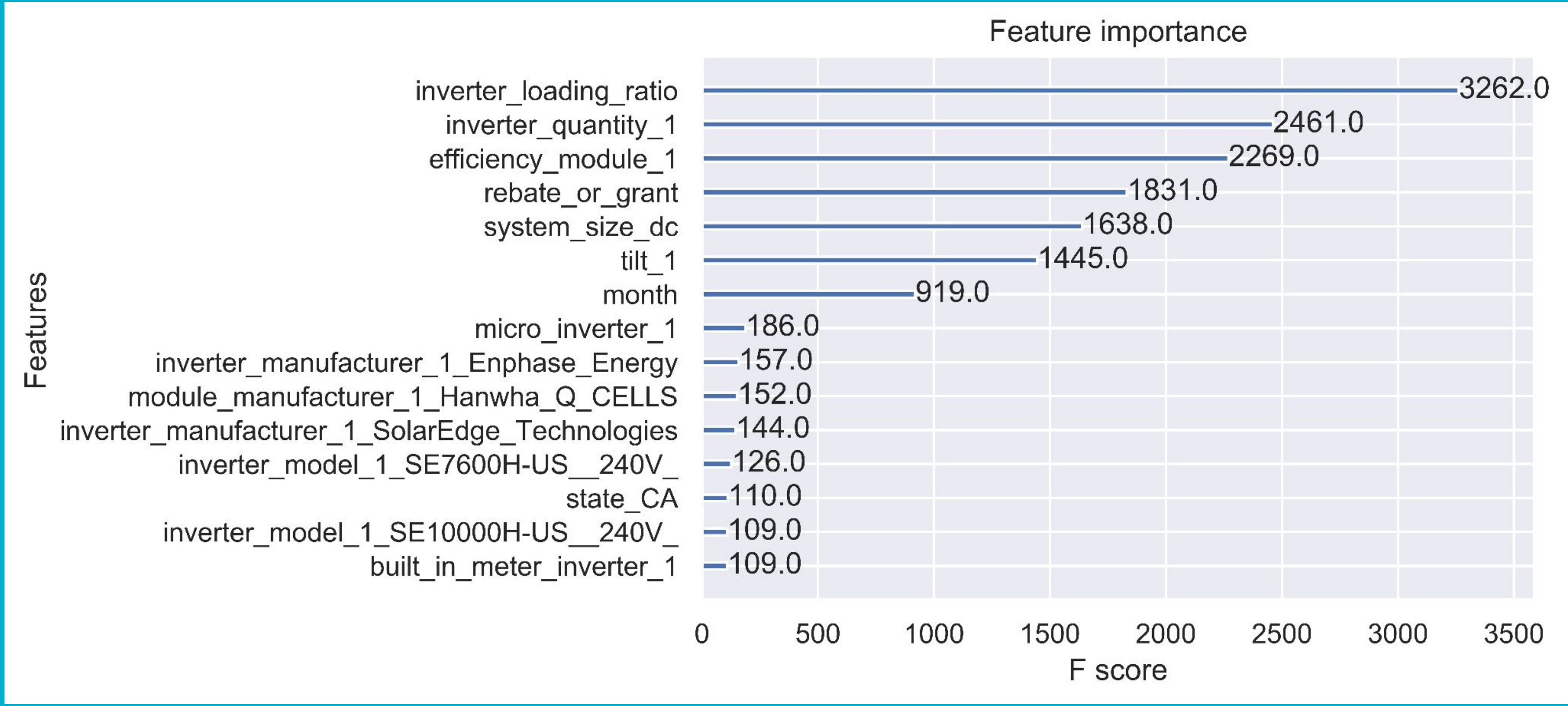


Outliers

	134435	2022447	1840658
data_provider_1	Salt River Project	Utah Office of Energy Development	New York State Energy Research and Development...
system_id_1	50806	SolarPV--0000002563	253904
installation_date	2020-03-30	2020-07-01	2020-08-07
system_size_dc	10.08	7.54	6.8
total_installed_price	3427200.0	2631400.0	8255000.0
rebate_or_grant	0.0	0.0	1476.0
customer_segment	RES	RES	RES
expansion_system	0	0	0
multiple_phase_system	0	0	0
tracking	-1	-1	-1

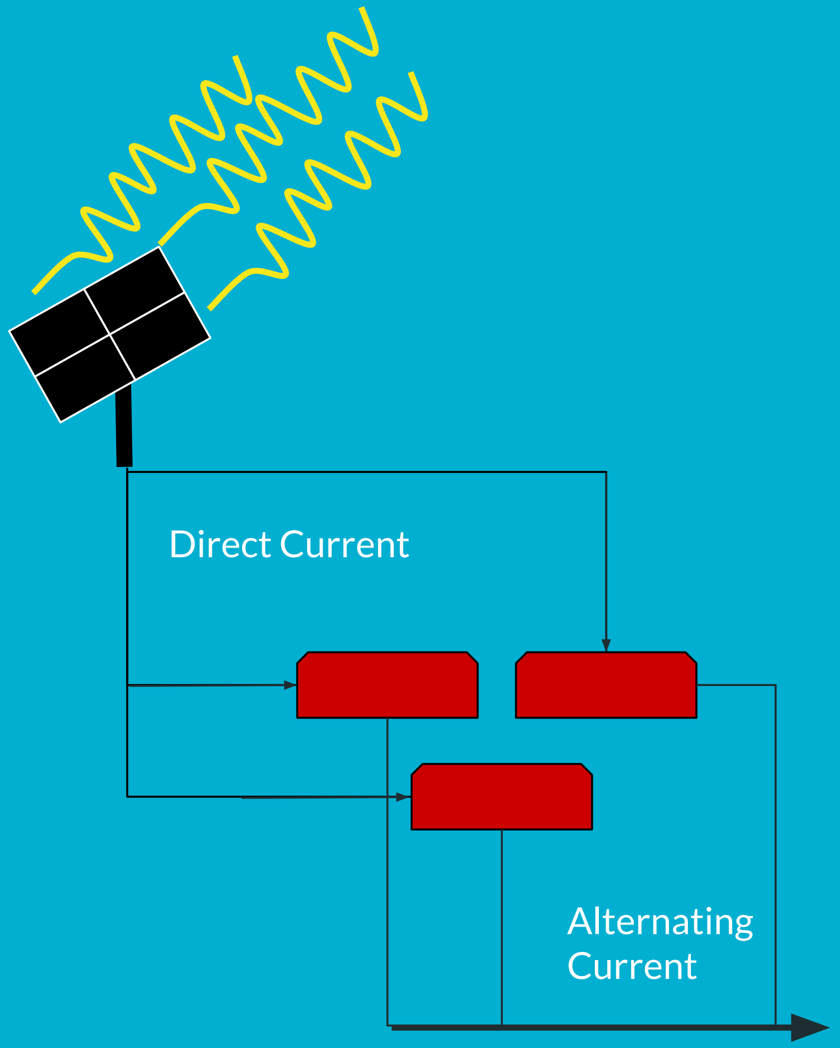
	108019	108020	108142	108175	108233
data_provider_1	Arizona Public Service	Arizona Public Service	Arizona Public Service	Arizona Public Service	Arizona Public Service
system_id_1	107903	107904	108026	108059	108117
installation_date	2020-06-17	2020-06-17	2020-06-19	2020-06-22	2020-06-23
system_size_dc	5.76	8.75	4.725	3.55	5.85
total_installed_price	17488.26	22631.0	18972.0	9900.0	20475.0
rebate_or_grant	0.0	0.0	0.0	0.0	0.0
customer_segment	RES	RES	RES	RES	RES
expansion_system	0	0	0	0	0
multiple_phase_system	0	0	0	0	0
tracking	-1	-1	-1	-1	-1

Most Important Features



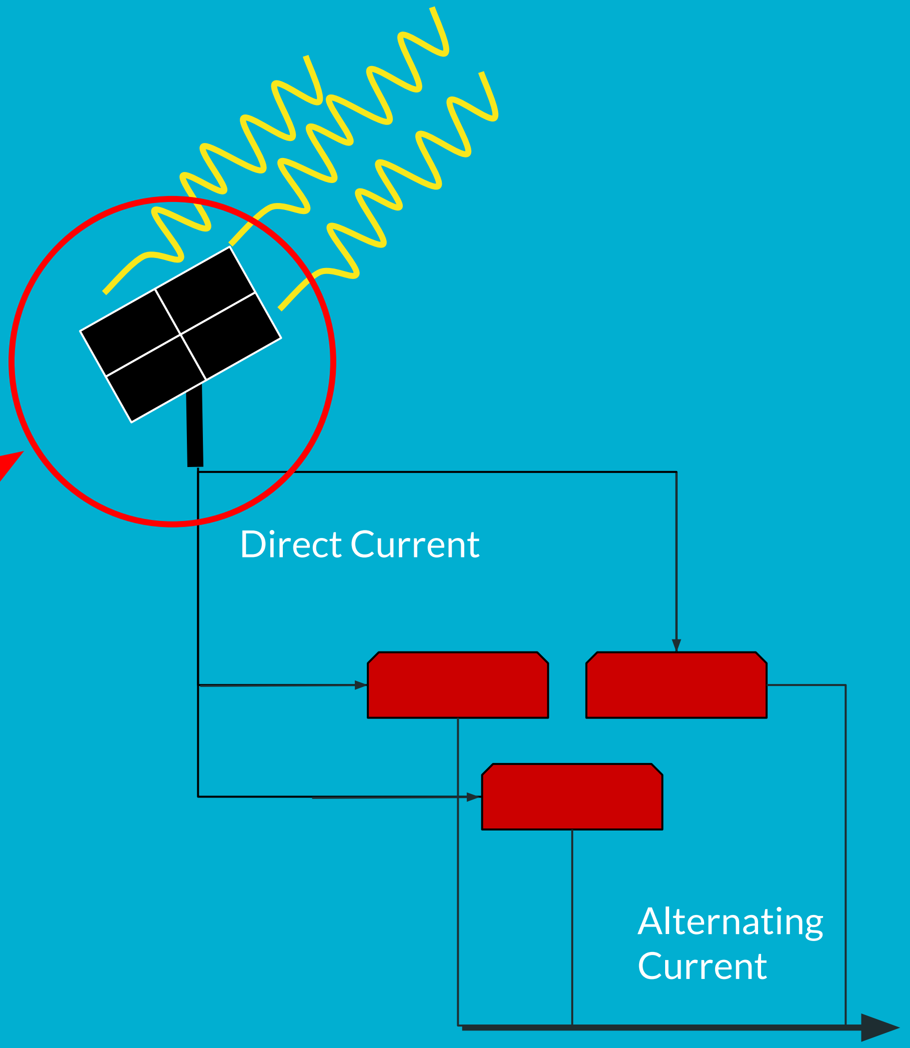
Solar Panel Assembly

- Inverter loading ratio
- Inverter quantity
- Photovoltaic module efficiency
- Rebate or grant
- System size
- Tilt



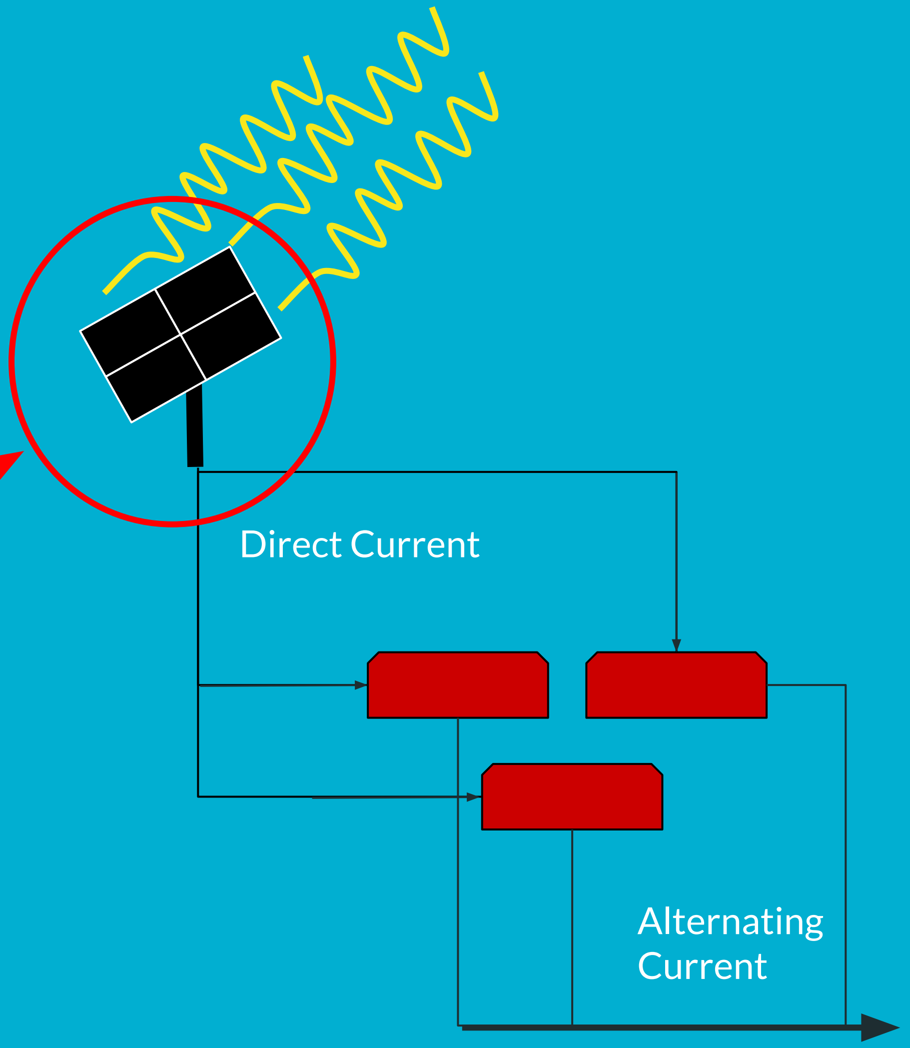
Solar Panel Assembly

- Inverter loading ratio
- Inverter quantity
- Photovoltaic module efficiency
- Rebate or grant
- System size
- Tilt



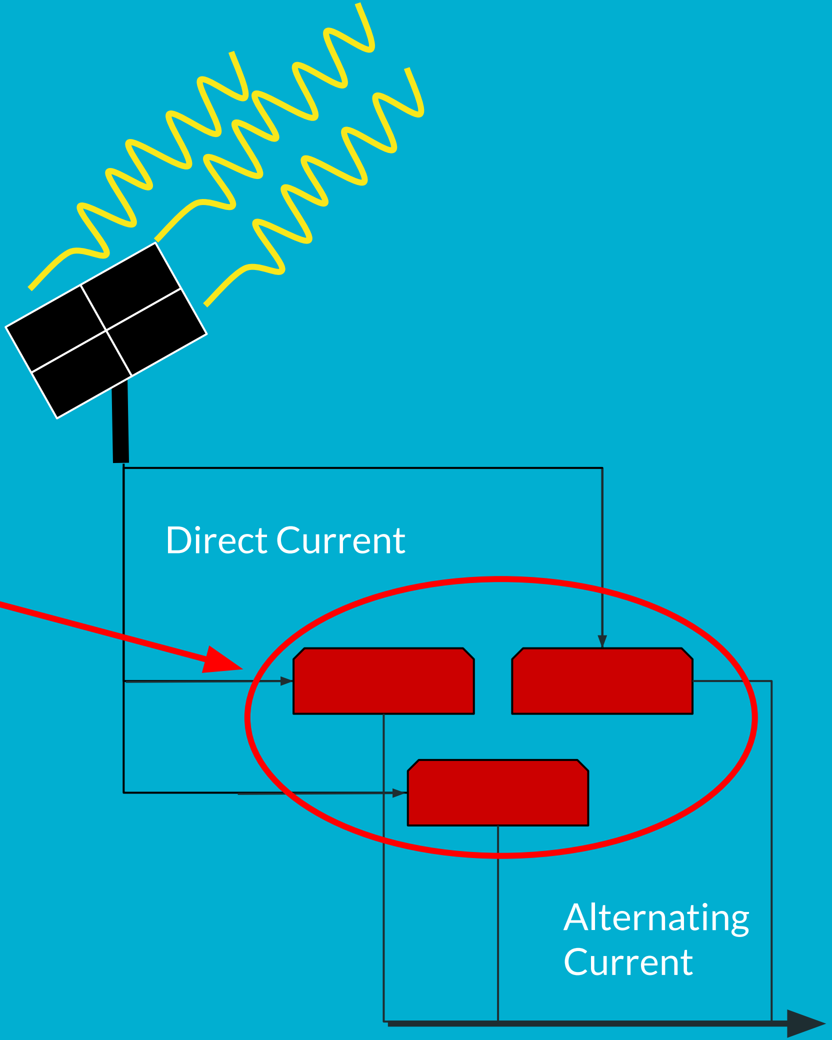
Solar Panel Assembly

- Inverter loading ratio
- Inverter quantity
- Photovoltaic module efficiency
- Rebate or grant
- System size
- Tilt



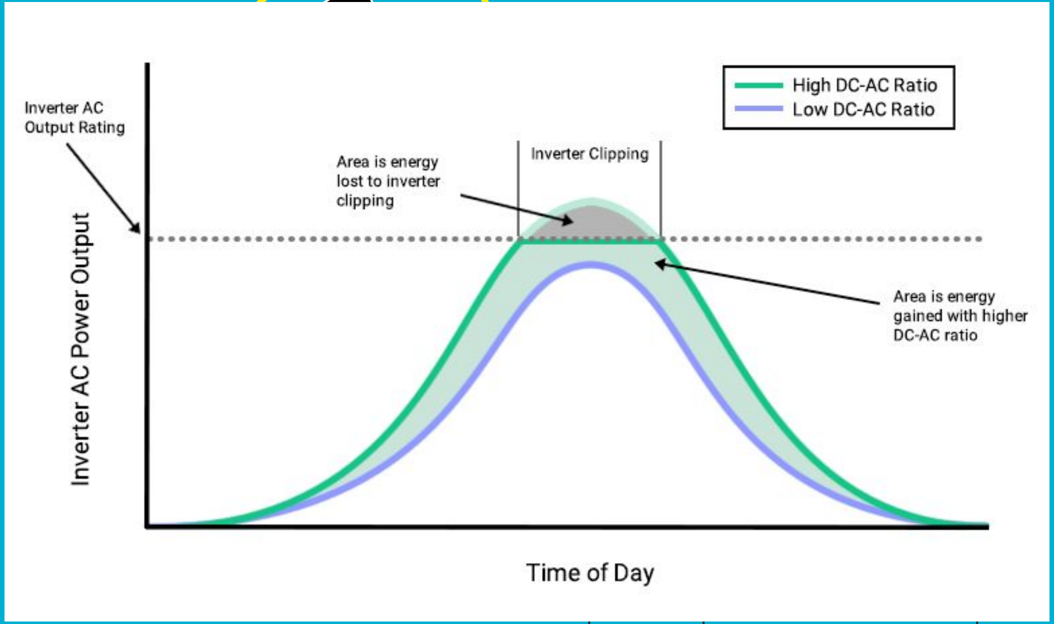
Solar Panel Assembly

- Inverter loading ratio
- Inverter quantity
- Photovoltaic module efficiency
- Rebate or grant
- System size
- Tilt



Solar Panel Assembly

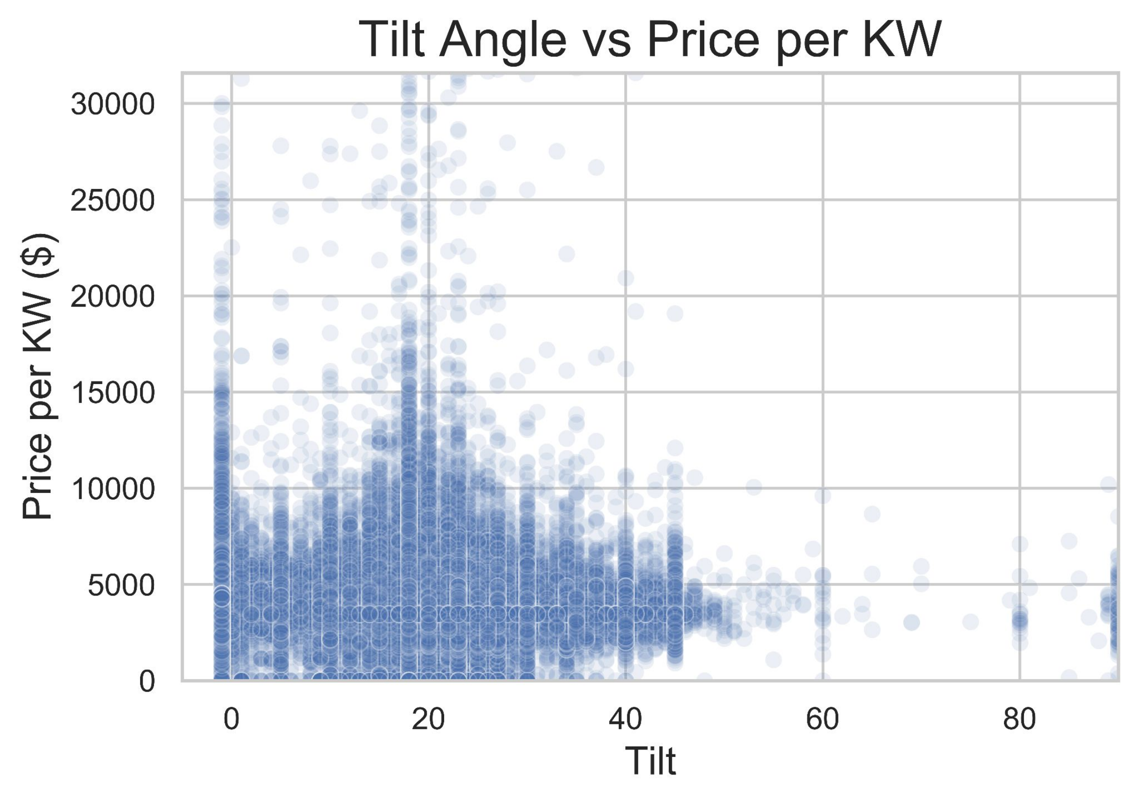
- Inverter loading ratio
- Inverter quantity
- Photovoltaic module efficiency
- Rebate or grant
- System size
- Tilt



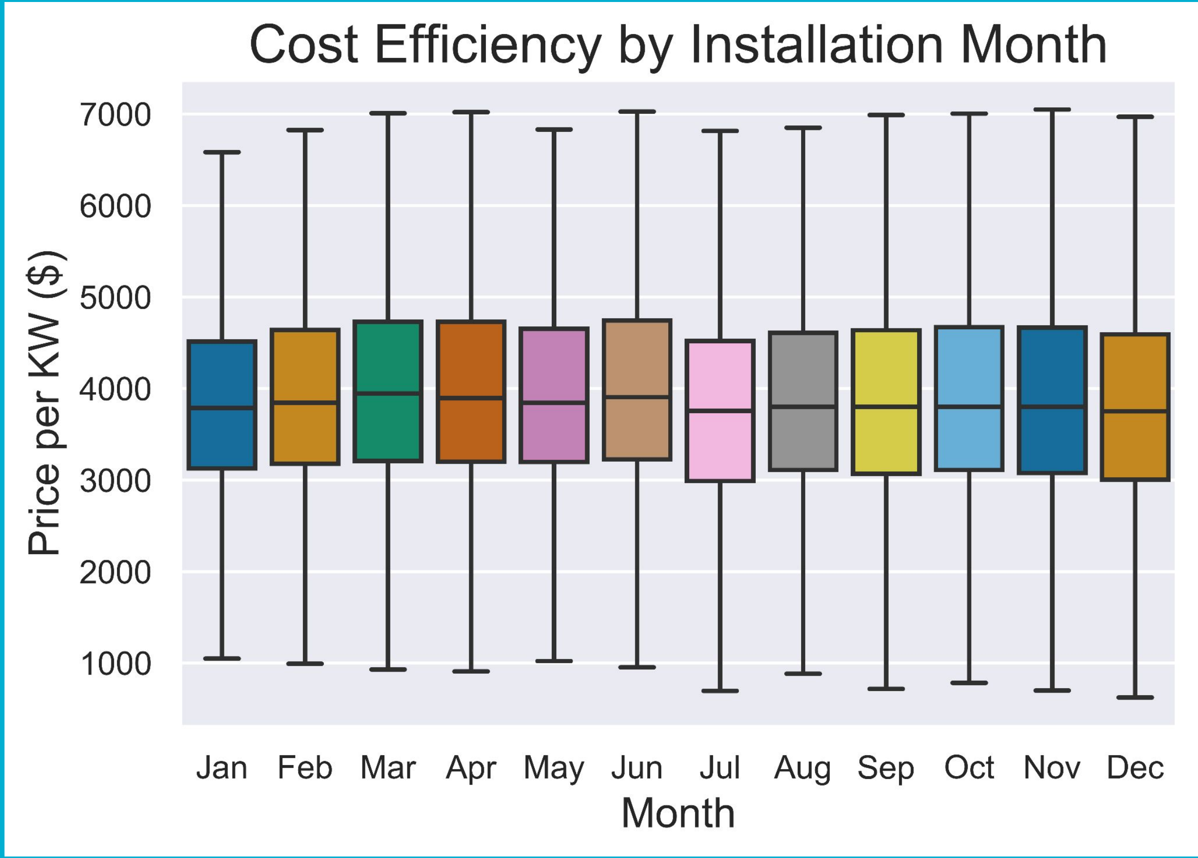
Alternating Current

Image source: https://aurorasolar.com/blog/choosing-the-right-size-inverter-for-your-solar-design-a-primer-on-inverter-clipping/

Tilt


Tilt Angle vs Price per KW
Price per KW ($)
30000
25000
20000
15000
10000
5000
0
0
20
40
60
80
Tilt

Installation Month



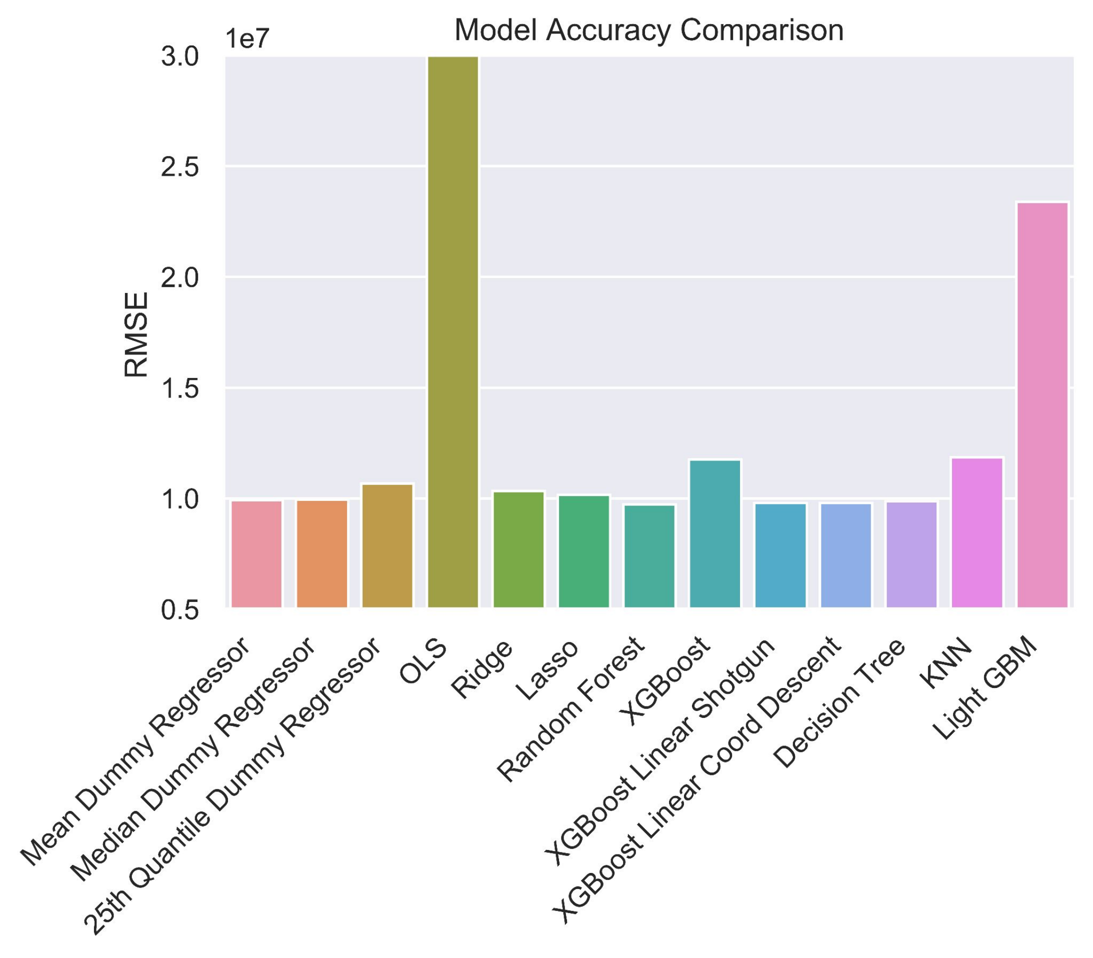
Summary

- XGBoost regressor - 3M train RMSE, 35M test RMSE
  - Outliers likely due to typos in final installed price
- Cost efficiency drivers:
  - Inverter loading ratio ↑ price per KW ↓
  - Inverter quantity ↑ price per KW ↓
  - PV module efficiency ↓ price per KW ↑
  - Rebate ↑ price per KW ↓
  - System Size ↑ price per KW ↓
  - Tilt
  - Month - July and December most efficient

Future Work

- Recommendation tool
- Inputs:
  - Budget
  - Monthly electricity usage
  - Location/electrical supplier
- Outputs:
  - System size
  - Inverter quantity
  - Available rebates/grants
  - Time to recover investment

Appendix

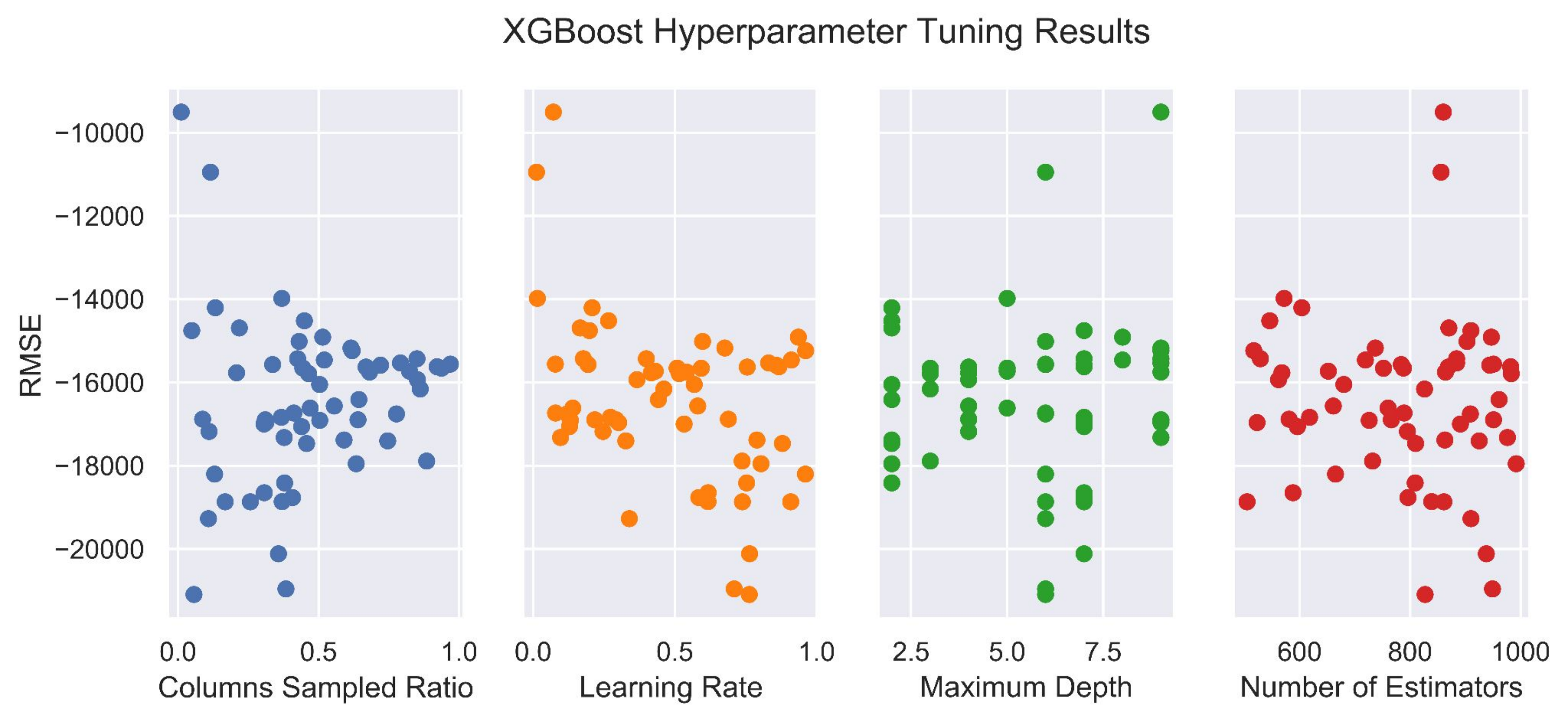


Model Accuracy Comparison
1e7
3.0
2.5
2.0
1.5
1.0
0.5
RMSE
Mean Dummy Regressor
Median Dummy Regressor
25th Quantile Dummy Regressor
OLS
Ridge
Lasso
Random Forest
XGBoost
XGBoost Linear Shotgun
XGBoost Linear Coord Descent
Decision Tree
KNN
Light GBM

Appendix

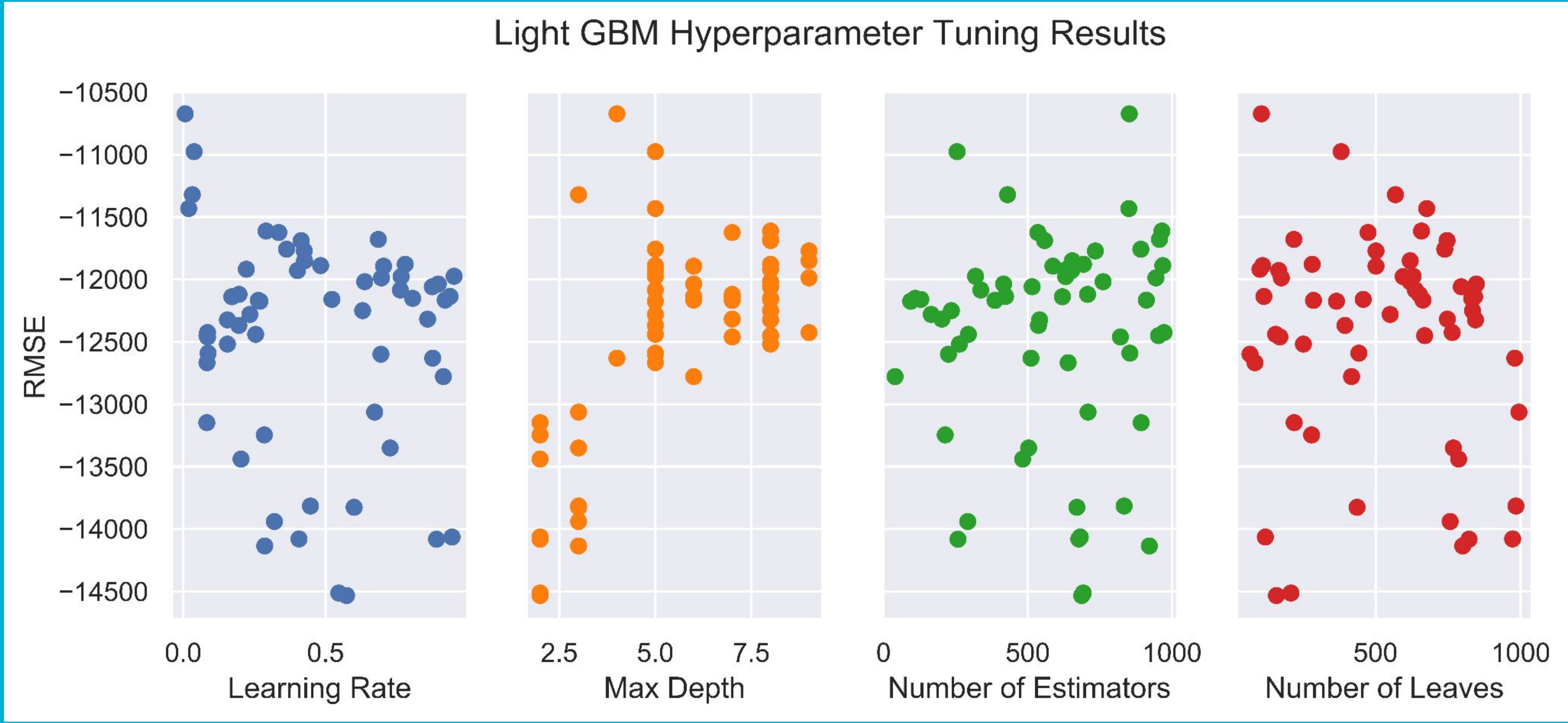
Model	Hyperparameters	RMSE
Mean Dummy Regressor	N/A	9,923,409
Median Dummy Regressor	N/A	9,941,871
25th Quantile Dummy Regressor	N/A	10,670,605
Ordinary Least Squares	N/A	1.2650560e+19
Ridge Regression	alpha	10,327,496
Lasso Regression	alpha	10,155,465
Random Forest Regression	max_features, max_depth, min_samples_leaf, n_estimators	9,720,312
XGBoost Regressor	n_estimators, max_depth, eta, colsample_bytree	11,759,799
Linear XGBoost Regressor - shotgun updater	reg_lambda, reg_alpha, feature_selector	9,788,061
Linear XGBoost Regressor - coordinate descent updater	reg_lambda, reg_alpha, feature_selector	9,789,561
Decision Tree Regressor	max_depth, min_samples_leaf	9,865,568
K Nearest Neighbors Regressor	n_neighbors	11,856,353
Light GBM Regressor	num_leaves, n_estimators, max_depth, learning_rate	23,385,497

Appendix



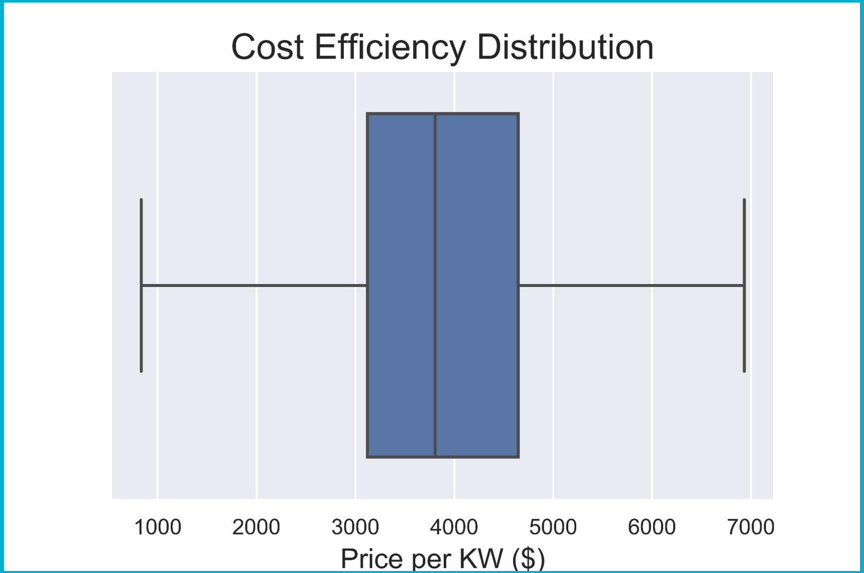
XGBoost Hyperparameter Tuning Results
RMSE
−10000
−12000
−14000
−16000
−18000
−20000
0.0
0.5
1.0
Columns Sampled Ratio
0.0
0.5
1.0
Learning Rate
2.5
5.0
7.5
Maximum Depth
600
800
1000
Number of Estimators

Appendix

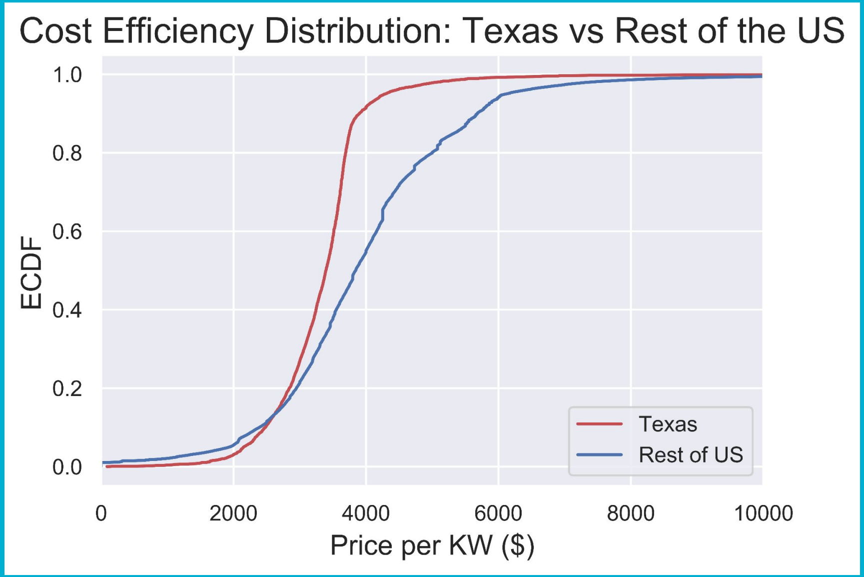


Light GBM Hyperparameter Tuning Results
RMSE
−10500
−11000
−11500
−12000
−12500
−13000
−13500
−14000
−14500
0.0
0.5
Learning Rate
2.5
5.0
7.5
Max Depth
0
500
1000
Number of Estimators
500
1000
Number of Leaves

Appendix

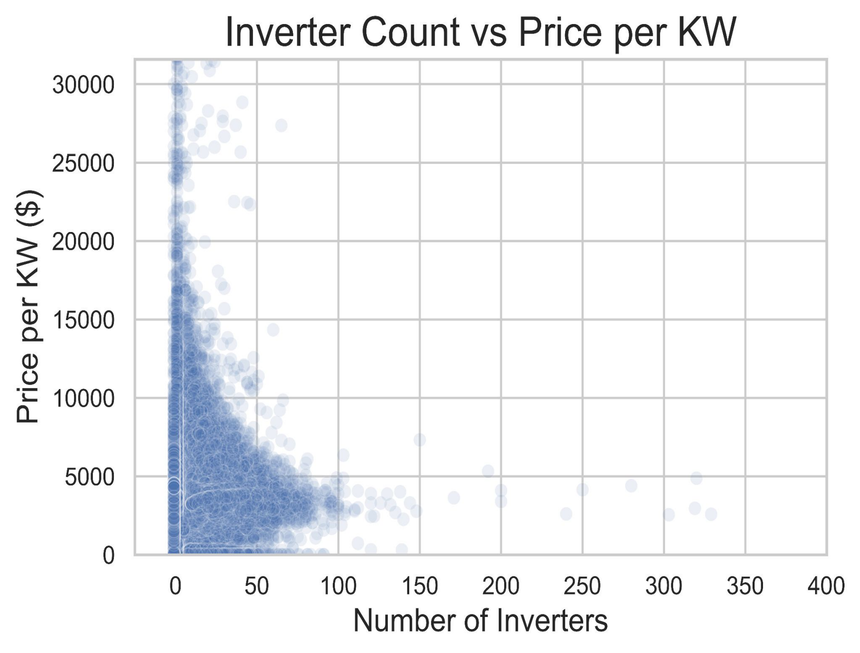


Cost Efficiency Distribution
1000
2000
3000
4000
5000
6000
7000
Price per KW ($)

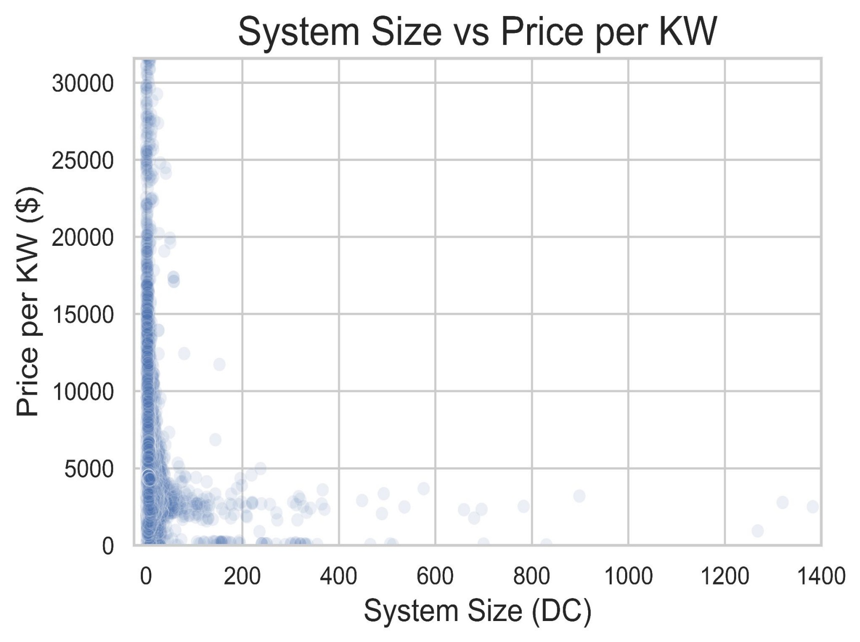


Cost Efficiency Distribution: Texas vs Rest of the US
ECDF
1.0
0.8
0.6
0.4
0.2
0.0
0
2000
4000
6000
8000
10000
Price per KW ($)
Texas
Rest of US

Appendix

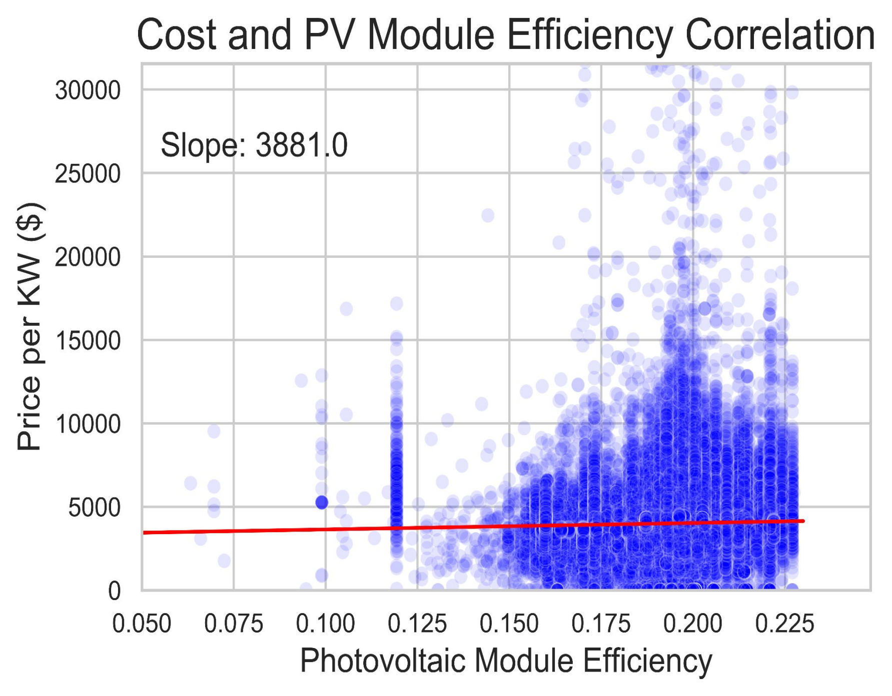


Inverter Count vs Price per KW
Price per KW ($)
30000
25000
20000
15000
10000
5000
0
0
50
100
150
200
250
300
350
400
Number of Inverters

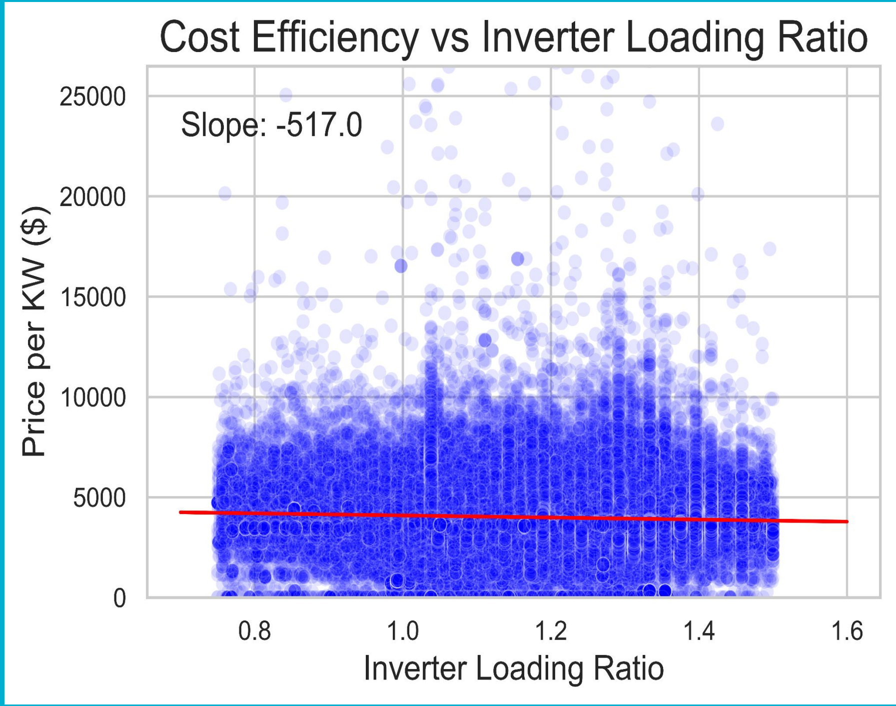


System Size vs Price per KW
Price per KW ($)
30000
25000
20000
15000
10000
5000
0
0
200
400
600
800
1000
1200
1400
System Size (DC)

Appendix



Cost and PV Module Efficiency Correlation
Slope: 3881.0
Price per KW ($)
30000
25000
20000
15000
10000
5000
0
0.050
0.075
0.100
0.125
0.150
0.175
0.200
0.225
Photovoltaic Module Efficiency



Cost Efficiency vs Inverter Loading Ratio
Slope: -517.0
Price per KW ($)
25000
20000
15000
10000
5000
0
0.8
1.0
1.2
1.4
1.6
Inverter Loading Ratio