Towards predicting site annualized probabilistic performance ratios using the kWh Analytics database

Dr. E. Victor Garcia, Alexander Barriga, Dr. Ben Browne, Adam Shinn, Helena Li, Zhe Yuan
kWh Analytics, Inc.

Abstract

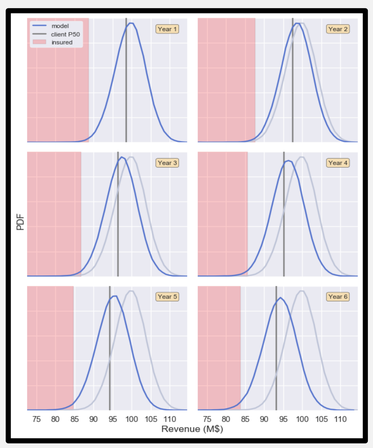
We use the kWh Analytics production history and system metadata databases containing hundreds of thousands of PV systems, along with satellite GHI and DNI measurements, to build quantile regression models that predict annualized weather-adjusted performance ratios. Thus we quantify performance ratio probability density functions directly, which are a function of soiling, shading, snow, wiring losses, light-induced degradation, depending upon region, equipment selection, and system age. We will observe how these systematic differences correlate with PV system parameters, to identify clearly the largest sources of difference. These models can also be used as an independent way to predict energy production for any given PV system given accurate system metadata.

Motivation

Reducing the cost of financing solar PV projects is critical for scaling solar PV in the USA and abroad in light of our warming planet. Thus, predicting the annualized distribution of performance ratio for any given site given PV system design and climate is an integral part of characterizing and reducing PV production risk.

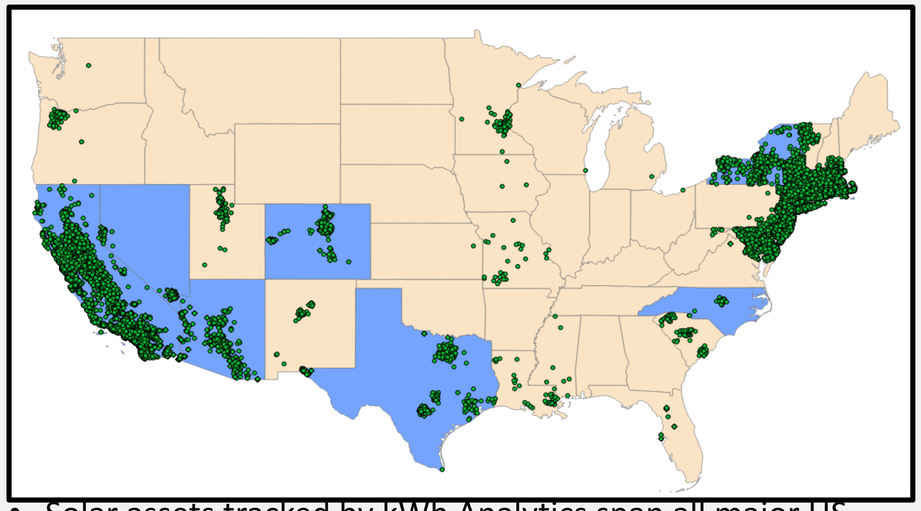
For the kWh Analytics Solar Revenue Put, we insure PV system production risk. Quantile regression allows us to model the probabilistic distribution of performance ratio for any given site directly using historical data from similar sites.

Thus, we can assess the production risks for any given PV site, given location and equipment using historical data from previously built similar PV sites.



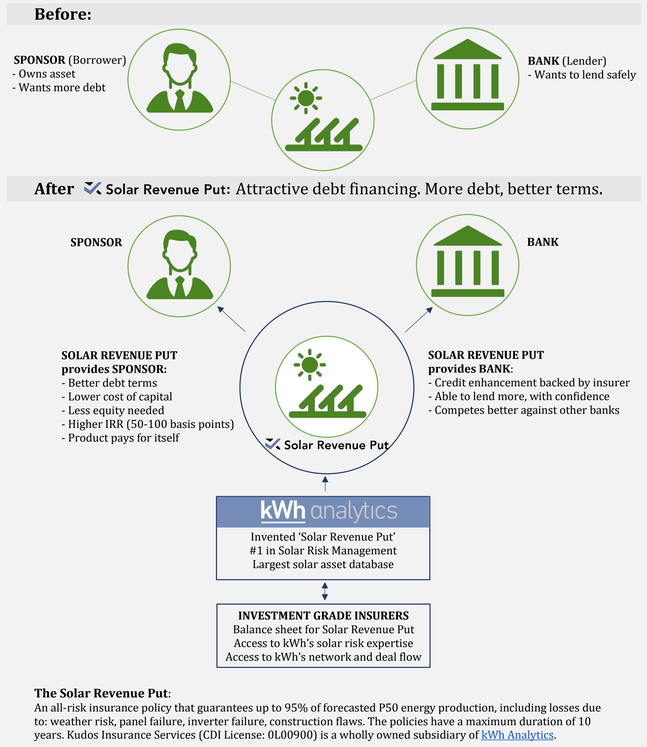
Left: Inter-annual distribution of revenue production for a given solar site. Pink is the insured range, blue probability density function is our insurance model, and grey probability density function is the 1st year prediction our our insurance model. The grey bar is the client P50 revenue.

Production Data



- Solar assets tracked by kWh Analytics span all major US solar markets
- > 200,000 systems
- Residential, Commercial/Industrial, Utility PV systems, spanning a diversity of manufacturers and climates.

kWh Analytics Solar Revenue Put



The Solar Revenue Put:
An all-risk insurance policy that guarantees up to 95% of forecasted P50 energy production, including losses due to: weather risk, panel failure, inverter failure, construction flaws. The policies have a maximum duration of 10 years. Kudos Insurance Services (CDI License: 0L00900) is a wholly owned subsidiary of kWh Analytics.

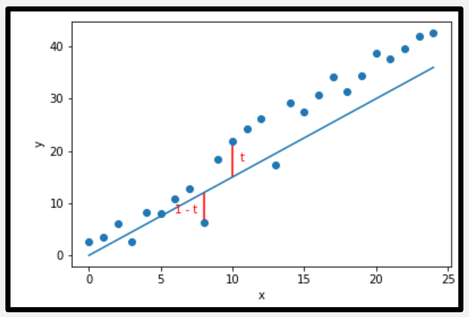
Quantile Regression

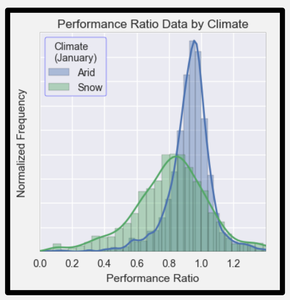
Quantile Loss Score

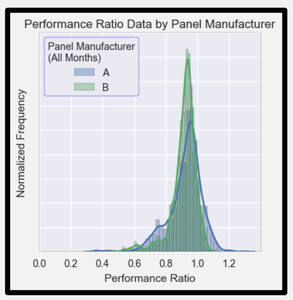
$$L_t(y^p, y) = \sum_{i=y_i<y_i^p}^{N} (1-t)\left|y_i - y_i^p\right| + \sum_{i=y_i\geq y_i^p}^{N} (t)\left|y_i - y_i^p\right|$$

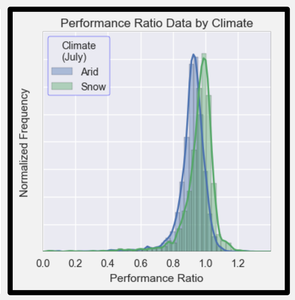
- t: quantile to estimate.
- i: Data point index
- y_i: Observed performance ratio
- y^p: Predicted performance ratio
- N: Number of observations

Right: Quantile loss score is an extension of mean absolute error. The model is biased to fit performance ratios at the quantile *t*.

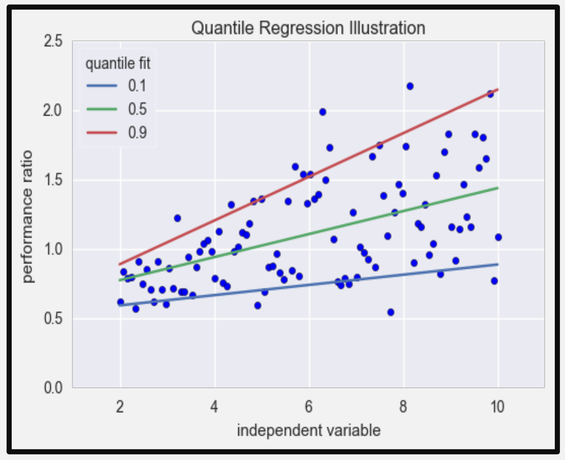








Above: The probability distribution of performance ratio can vary by panel manufacturer or climate type at different times of year. Thus, a generalized quantile regression model can reproduce the annualized performance ratio probability density function for any given site.



Left: Similar to linear regression, using quantile regression, we predict the PV system annualized performance ratio given an independent variable at quantile 0.5 (50th percentile, green line). Quantile regression also allows for the prediction at quantile 0.1 (blue line, PV systems with the 10th percentile worst PRs) or 0.9 (red line, PV systems with the 90th percentile best PRs).

Model	Quantile Loss Score	Quantile Loss Score (with weather-related features)	
		Train	Test
Linear Quantile Regression	0.0148	0.0137	0.0141
Random Forest	0.0145	0.0088	0.0132
Gradient Boosting	0.0141	0.0110	0.0126
Baseline	0.01534		

Left: We can also use non-linear models such as decision trees to model the performance ratio probability distribution for sites given weather-related and equipment-related features.