Maximizing MWh:

A Statistical Analysis of the Performance of Utility-Scale Photovoltaic Projects in the United States

Mark Bolinger[1], Joachim Seel[1], Manfei Wu[2]

[1]Lawrence Berkeley National Laboratory

[2]Goldman School of Public Policy, University of California at Berkeley

March 2016

This research was supported by funding from the U.S. Department of Energy's SunShot Initiative.

Context for this report

LBNL's latest *Utility-Scale Solar* report (utilityscalesolar.lbl.gov) found that capacity factors for utility-scale PV projects vary by more than a factor of two, ranging from ~15% to ~35% (in AC terms)

- Sample consists 128 utility-scale PV projects totaling 3,201 MW_{AC} that achieved commercial operation in 18 different states over the seven-year period from 2007-2013, and that were operating throughout the entirety of 2014
- "Utility-scale" defined as any ground-mounted project larger than 5 MW_{AC}

This report investigates what drives this high degree of variation

- First known use of multivariate regression techniques to analyze empirical variation in project-level performance

Understanding performance is important because the profitability of the utility-scale sector depends directly on how well these projects perform over time, which in turn affects the extent to which the sector can raise much-needed investment capital

- The relative importance of performance will only increase in the future as the federal investment tax credit ("ITC") is eventually phased down to 10% (from 30% currently)

To normalize performance across projects, we focus on net AC capacity factor in 2014 ("NCF")

Since some projects in the sample have been operating since 2007 (see table), we considered using a longer-term (e.g., cumulative) measure of capacity factor, but ultimately decided to focus on a single year – 2014 – for three reasons:

1) The bulk of our project sample is very young (see table)

- 37% of projects and 52% of capacity achieved commercial operations in 2013
- For these 2013 projects, 2014 was the *only* full year for which we could calculate a capacity factor
- Given inter-year variation in insolation, it seemed inappropriate to compare capacity factors in 2014 (for 2013-vintage projects) to capacity factors from 2008-2014 (for earlier projects)

2) Data availability favors a focus on 2014

- Vaisala provided us with site-specific insolation estimates for 2014, but not for earlier years

3) Focusing on 2014 allows us to isolate the potential influence of time

- Comparing how older and newer projects performed in 2014 could shed light on empirical module degradation rates

COD Year	Projects in Sample	Capacity in Sample (MW_{AC})
2007	2	19
2008	1	10
2009	3	54
2010	7	144
2011	32	464
2012	36	834
2013	47	1,676
Total	**128**	**3,201**

Possible drivers of AC capacity factors in 2014

1) **Solar resource strength:** Vaisala provided average annual global horizontal irradiance ("GHI") estimates in 2014 for each site in our sample
2) **Tracking:** Single-axis tracking increases the "plane of array" irradiance (dual-axis even more so, though there are no dual-axis projects in our sample)
3) **Inverter Loading Ratio ("ILR"):** At a higher DC:AC ratio, the inverter operates closer to (or at) full capacity for more of the day, which boosts AC capacity factor
4) **Project vintage ("COD Year"):** All else equal, module degradation could result in older projects having lower 2014 capacity factors than newer projects
5) **Orientation (tilt and azimuth):** Matters most for fixed-tilt projects. We lack good data, but assume that tilt and azimuth will be optimized across projects to maximize generation, and therefore will not add any explanatory power to our models.
6) **Temperature:** All PV modules perform better at lower operating temps, and modules with lower power temperature coefficients are not as negatively affected by higher operating temps. Unfortunately, we lack good data on coefficients and temps.

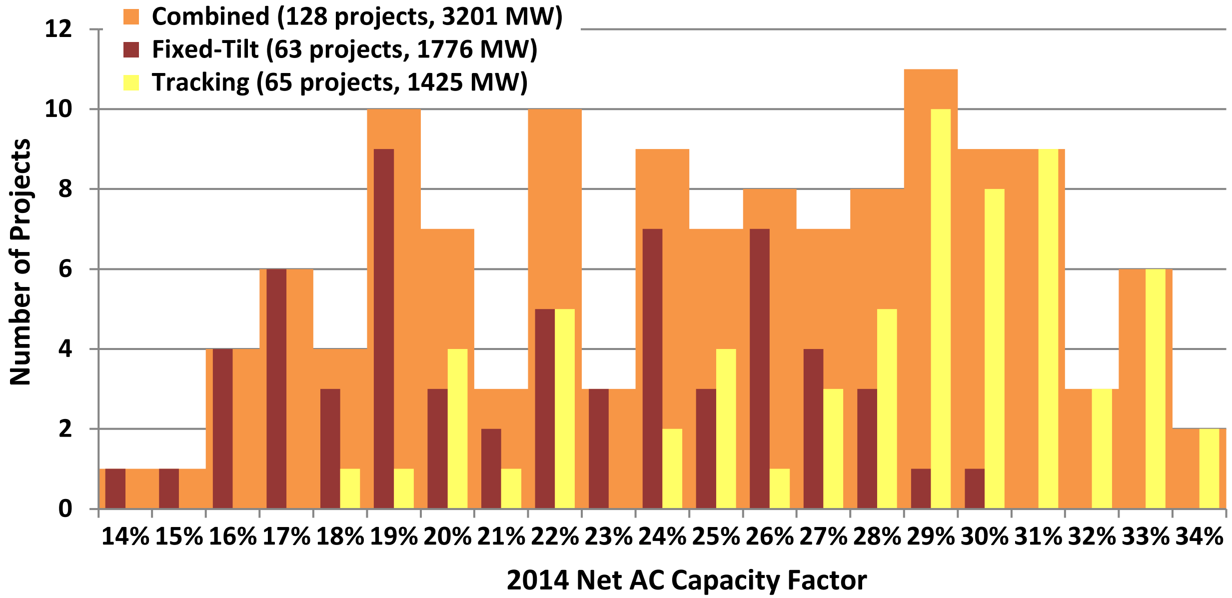
Our regression models include the first four drivers in this list as independent variables, but omit the last two due to data limitations

Sample descriptive statistics and histogram

	Project Capacity (MW_{AC})	2014 Net Capacity Factor	2014 GHI ($kWh/m^2/day$)	ILR (MW_{DC}/MW_{AC})
Minimum	5.2	14.8%	3.73	1.05
10^{th} percentile	7.6	18.4%	3.84	1.11
25^{th} percentile	10.0	20.7%	4.65	1.17
Median	15.0	25.7%	5.39	1.24
75^{th} percentile	20.0	29.9%	5.61	1.28
90^{th} percentile	38.5	31.8%	5.81	1.35
Maximum	250.0	34.9%	6.02	1.50

- 2014 net capacity factor (“NCF”) varies by more than a factor of two
- Sample is split almost evenly between fixed-tilt and tracking projects

- Somewhat bi-modal NCF distribution due to influence of fixed-tilt vs. tracking
- Though not shown, fixed-tilt projects have a slightly higher mean ILR (1.25) than tracking (1.21), which makes sense given slight redundancy between tracking and high ILR

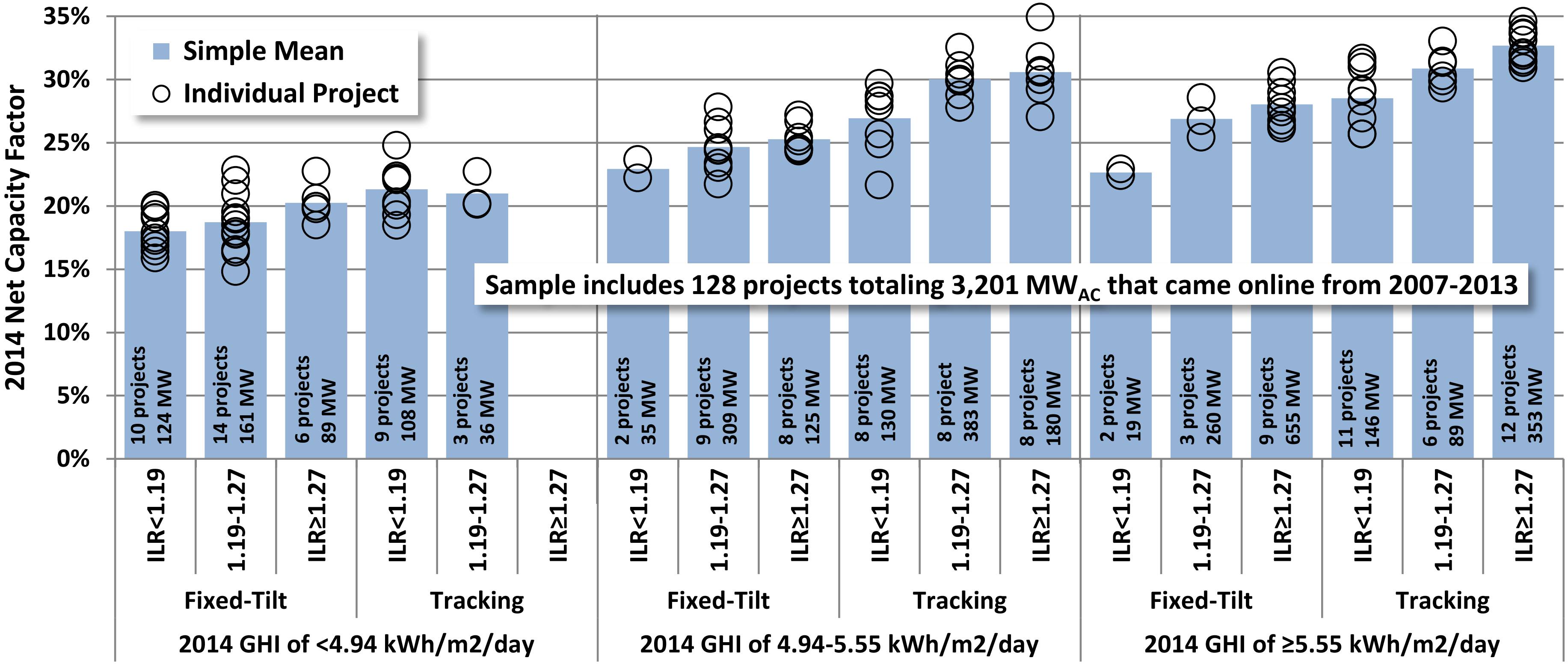


Sample description by state
(sorted in descending order by mean capacity factor)

State	# of Projects	Capacity (MW_{AC})	Mean 2014 NCF	Mean 2014 GHI ($kWh/m^2/day$)	Mean ILR (MW_{DC}/MW_{AC})	% Tracking
AZ	19	594	30.4%	5.62	1.27	89.5%
CA	41	1,561	28.1%	5.52	1.26	46.3%
NV	8	284	27.9%	5.41	1.23	50.0%
CO	4	62	27.3%	5.30	1.14	100.0%
NM	13	154	26.9%	5.63	1.17	61.5%
TX	6	84	23.0%	4.85	1.18	83.3%
GA	3	54	22.5%	4.65	1.27	0.0%
TN	1	8	21.0%	4.36	1.26	0.0%
NC	3	54	20.9%	4.26	1.21	66.7%
FL	4	53	20.0%	4.88	1.12	50.0%
DE	2	22	19.9%	3.94	1.19	50.0%
IN	4	39	19.4%	3.88	1.35	0.0%
MD	3	39	19.2%	3.90	1.27	0.0%
IL	2	28	19.0%	3.77	1.18	50.0%
NY	1	32	18.5%	3.93	1.20	0.0%
OH	2	18	18.4%	3.75	1.20	50.0%
NJ	11	107	17.5%	3.85	1.20	9.1%
PA	1	10	15.9%	3.76	1.15	0.0%
Total US	**128**	**3,201**	**25.5%**	**5.08**	**1.23**	**50.8%**

- 83% of MW and 66% of projects are in the “top 5” states for NCF & GHI
- CA alone accounts for 49% of capacity and 32% of projects
- Single-axis tracking is generally more prevalent in high-GHI states

Sample broken out by 3 of 4 independent variables
(“COD Year” not shown)



- Thresholds for GHI and ILR bins are based on 33rd and 66th percentiles for each, in order to evenly distribute the sample among bins
- Generally positive relationship between 2014 NCF and ILR, tracking, and GHI, particularly among the mean values (there is more variability among the individual projects)

We also included three “interactive” variables

In addition to the four primary independent variables described so far – i.e., 2014 GHI, Tracking, ILR, and COD Year – we also included three “interactive” independent variables to try to capture the nuances in how GHI, tracking, and ILR influence capacity factor in relation to one another:

1) **Tracking x 2014 GHI:** We would expect single-axis tracking to provide more of an NCF boost at sites with a stronger solar resource

2) **Tracking x ILR:** We would expect the benefit of tracking to diminish at higher ILRs, given that tracking and a high ILR are, at least to some extent, redundant

3) **2014 GHI x ILR:** As with tracking, we would expect higher ILRs to provide more of an NCF boost at sites with a stronger resource

Build-up of robust multivariate regression model

In all five models, the dependent variable is net AC capacity factor in 2014.

	Model 1	Model 2	Model 3	Model 4	Model 5
2014 GHI	0.0604***	0.0521***	0.0478***	0.0479***	0.0423***
	(0.00280)	(0.00265)	(0.00199)	(0.00202)	(0.00245)
Tracking		0.0337***	0.0429***	0.0421***	0.0405***
		(0.00443)	(0.00296)	(0.00303)	(0.00285)
ln(ILR)			0.2391***	0.2146***	0.2158***
			(0.01911)	(0.02206)	(0.03393)
COD Year				0.0033**	0.0023*
				(0.00138)	(0.00132)
Tracking x 2014 GHI					0.0186***
					(0.00501)
Tracking x ln(ILR)					-0.0015
					(0.04686)
2014 GHI x ln(ILR)					0.0575*
					(0.02952)
Constant	0.2546***	0.2375***	0.2328***	0.2172***	0.2202***
	(0.00243)	(0.00268)	(0.00198)	(0.00640)	(0.00598)
Observations	128	128	128	128	128
R^2	0.716	0.811	0.920	0.926	0.938
Root MSE	0.02746	0.02248	0.01467	0.01423	0.01316

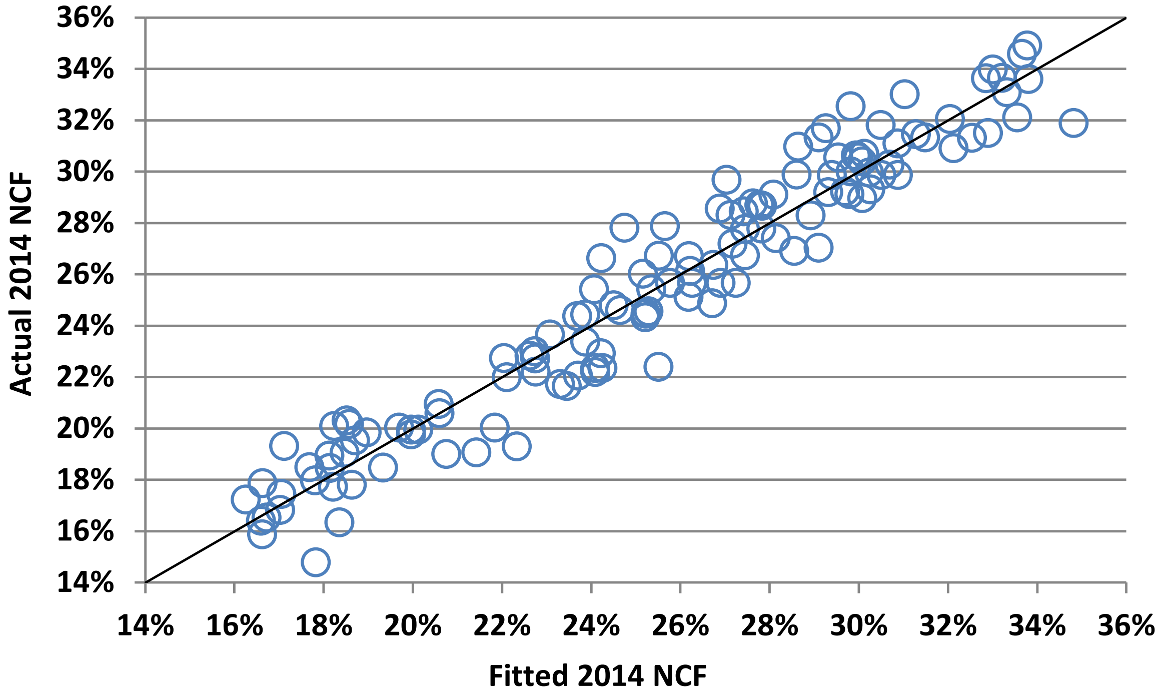
Robust standard errors are in parentheses, below the coefficients

***p<0.01, **p<0.05, *p<0.1

All continuous variables are centered around their means. The COD Year variable is based in the earliest possible year, 2007.

- We transformed ILR by taking its natural logarithm (reflects diminishing marginal benefit due to more "power clipping" at higher ILRs)
- We centered the independent variables
- We used "robust" regression techniques
- For Models 1-4, we added variables in order of largest to smallest influence
- Model 5 is preferred specification

Model 5 fitted vs. actual: highly linear and tight fit

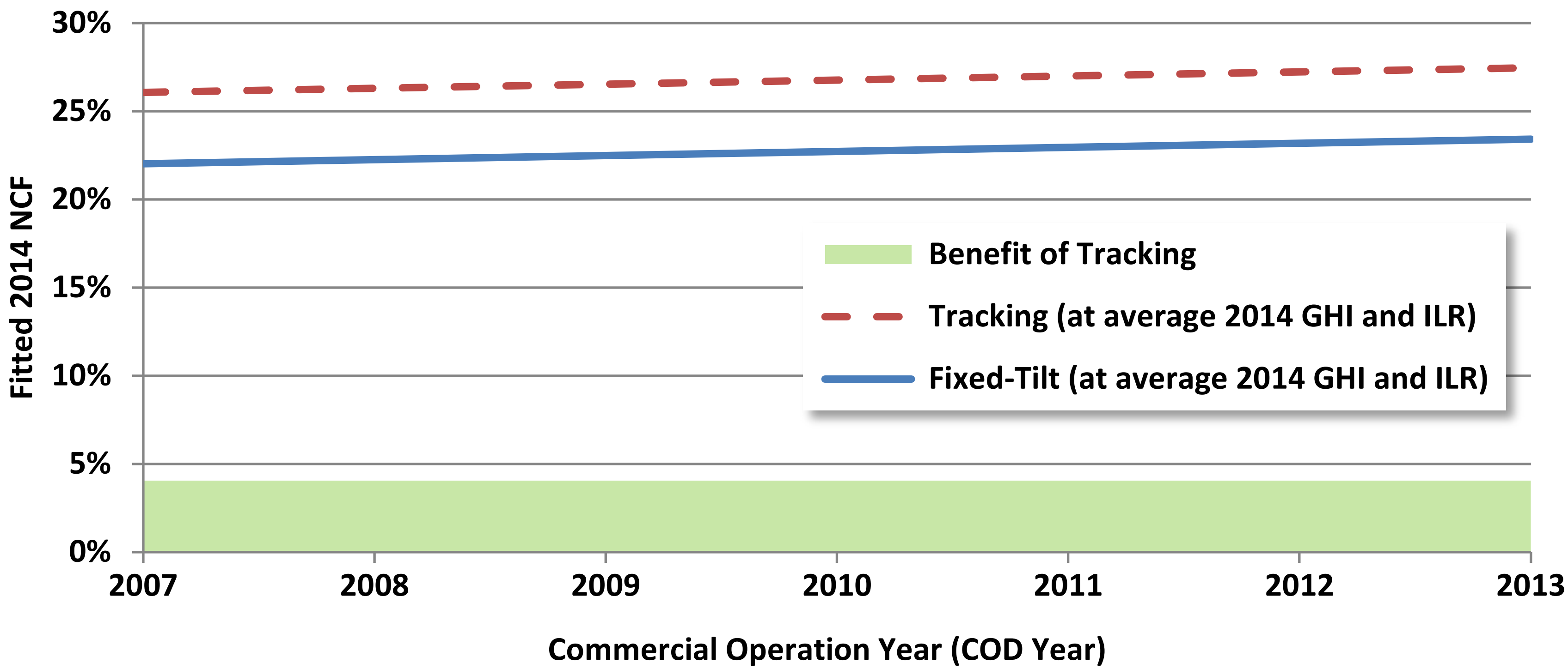


- $R^2 = 0.938$
- Root MSE = 0.013
- Residuals do not exhibit heteroskedasticity
- All but one of the independent variables [Tracking x ln(ILR)] are statistically significant (and most are highly significant)

In addition to using robust regression techniques and visually inspecting the plotted residuals, we also conducted several statistical tests to ensure that our data and model specification conform to the underlying assumptions of ordinary least squares (OLS) regression:

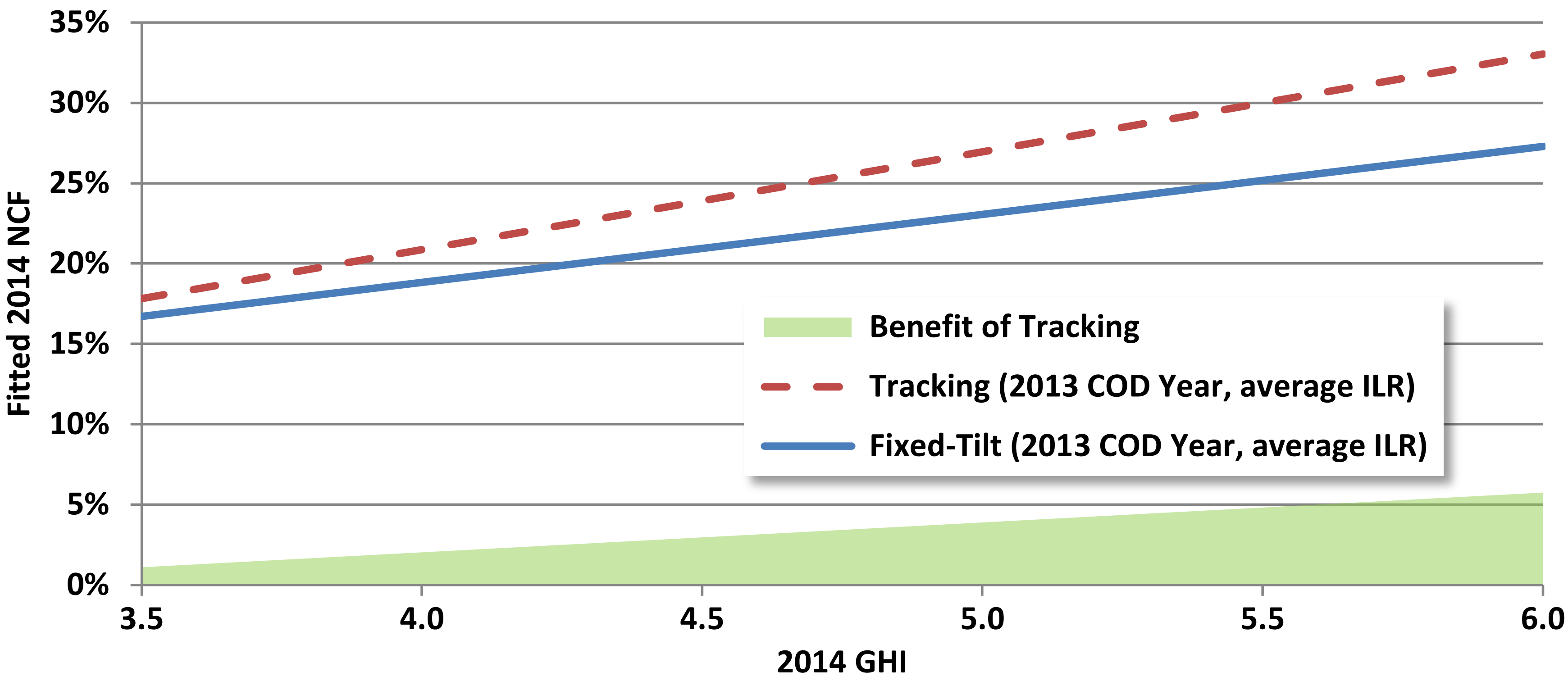
- The Breusch-Pagan and White tests both indicate no heteroskedasticity problems
- The Shapiro-Wilk test indicates normality in the residuals
- The Variance Inflation Factor test suggests no multicollinearity problems
- The linktest finds good model specification

Model 5 interpretation: influence of COD Year
(at average GHI and ILR)



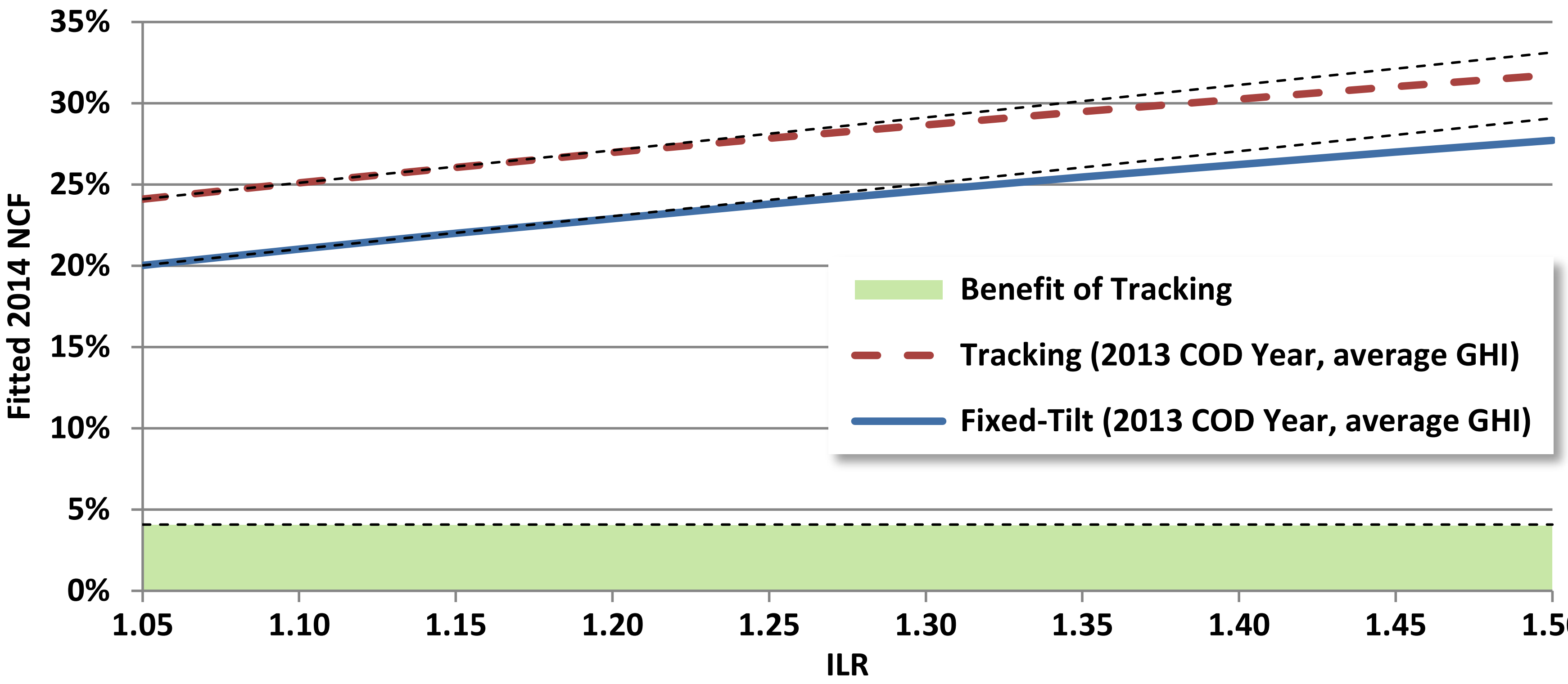
- Constant term tells us that a fixed-tilt project built in 2007 with average 2014 GHI and ILR is predicted to have a 2014 NCF of 22.02%
- Single-axis tracking adds 4.05% to this same project's 2014 NCF, for 26.07% in total
- Each successive COD Year adds 0.23% to both fixed-tilt and tracking projects, such that a 22.02% and 26.07% NCF for a fixed-tilt and tracking project, respectively, built in 2007 becomes 23.41% and 27.46% for projects built in 2013

Model 5 interpretation: influence of GHI
(for 2013 project with average ILR)



- Rest of interpretation focuses on projects built in 2013
- For each 1 kWh/m^2/day change in 2014 GHI, the model predicts a 4.23% and 6.08% absolute change in 2014 NCF for fixed-tilt and tracking projects, respectively
- This means that the benefit of tracking increases at sites with a stronger solar resource, which makes intuitive sense and is consistent with the greater prevalence of tracking at high-GHI sites seen earlier

Model 5 interpretation: influence of ILR
(for 2013 project with average GHI)



- Due to its logarithmic specification, ILR has a slightly diminishing marginal effect on 2014 NCF as it increases (as shown by the slight deviation from the linear thin dashed lines, included only for visual comparison)
- For both fixed-tilt and tracking projects, 2014 NCF increases by 1.00% when moving from an ILR of 1.05 to 1.10, but by only 0.73% when moving from an ILR of 1.45 to 1.50
- This diminishing effect potentially reflects greater amounts of power clipping at higher ILRs

Model 5 interpretation: marginal influence of GHI/ILR

Fitted 2014 NCF of a Fixed-Tilt Project with a 2013 COD

ILR\GHI	3.5	4.0	4.5	5.0	5.5	6.0	*(B) 5.5-4.0*
1.05	14.8%	16.4%	18.1%	19.7%	21.4%	23.1%	*5.0%*
1.10	15.3%	17.1%	18.9%	20.7%	22.5%	24.3%	*5.4%*
1.15	15.9%	17.8%	19.7%	21.7%	23.6%	25.5%	*5.8%*
1.20	16.4%	18.5%	20.5%	22.6%	24.6%	26.7%	*6.1%*
1.25	16.9%	19.1%	21.3%	23.4%	25.6%	27.8%	*6.5%*
1.30	17.4%	19.7%	22.0%	24.3%	26.5%	28.8%	*6.8%*
1.35	17.9%	20.3%	22.7%	25.0%	27.4%	29.8%	*7.2%*
1.40	18.4%	20.8%	23.3%	25.8%	28.3%	30.8%	*7.5%*
1.45	18.8%	21.4%	24.0%	26.6%	29.1%	31.7%	*7.8%*
1.50	19.2%	21.9%	24.6%	27.3%	30.0%	32.6%	*8.1%*
(A) 1.45-1.10	*3.44%*	*4.24%*	*5.03%*	*5.83%*	*6.62%*	*7.42%*	

Fitted 2014 NCF of a Tracking Project with a 2013 COD

ILR\GHI	3.5	4.0	4.5	5.0	5.5	6.0	*(B) 5.5-4.0*
1.05	15.9%	18.5%	21.1%	23.7%	26.3%	28.8%	*7.8%*
1.10	16.5%	19.2%	21.9%	24.6%	27.4%	30.1%	*8.2%*
1.15	17.0%	19.9%	22.7%	25.6%	28.4%	31.3%	*8.6%*
1.20	17.5%	20.5%	23.5%	26.5%	29.4%	32.4%	*8.9%*
1.25	18.0%	21.1%	24.2%	27.3%	30.4%	33.5%	*9.3%*
1.30	18.5%	21.7%	24.9%	28.1%	31.3%	34.5%	*9.6%*
1.35	19.0%	22.3%	25.6%	28.9%	32.2%	35.6%	*9.9%*
1.40	19.4%	22.9%	26.3%	29.7%	33.1%	36.5%	*10.3%*
1.45	19.9%	23.4%	26.9%	30.4%	33.9%	37.5%	*10.6%*
1.50	20.3%	23.9%	27.5%	31.1%	34.8%	38.4%	*10.8%*
(A) 1.45-1.10	*3.40%*	*4.20%*	*4.99%*	*5.79%*	*6.58%*	*7.37%*	

Benefit of Tracking for a Project with a 2013 COD

ILR\GHI	3.5	4.0	4.5	5.0	5.5	6.0
1.05	1.13%	2.06%	2.99%	3.91%	4.84%	5.77%
1.10	1.12%	2.05%	2.98%	3.91%	4.84%	5.77%
1.15	1.11%	2.04%	2.97%	3.90%	4.83%	5.76%
1.20	1.11%	2.04%	2.97%	3.90%	4.82%	5.75%
1.25	1.10%	2.03%	2.96%	3.89%	4.82%	5.75%
1.30	1.10%	2.02%	2.95%	3.88%	4.81%	5.74%
1.35	1.09%	2.02%	2.95%	3.88%	4.81%	5.74%
1.40	1.08%	2.01%	2.94%	3.87%	4.80%	5.73%
1.45	1.08%	2.01%	2.94%	3.87%	4.80%	5.73%
1.50	1.07%	2.00%	2.93%	3.86%	4.79%	5.72%

Orange-shaded rows (A):

- The benefit of a higher ILR increases at higher GHI levels
- Moreover, for any given GHI, the benefit of a higher ILR is slightly greater for fixed-tilt than for tracking projects (e.g., 4.24% vs. 4.20% for a GHI of 4.0)

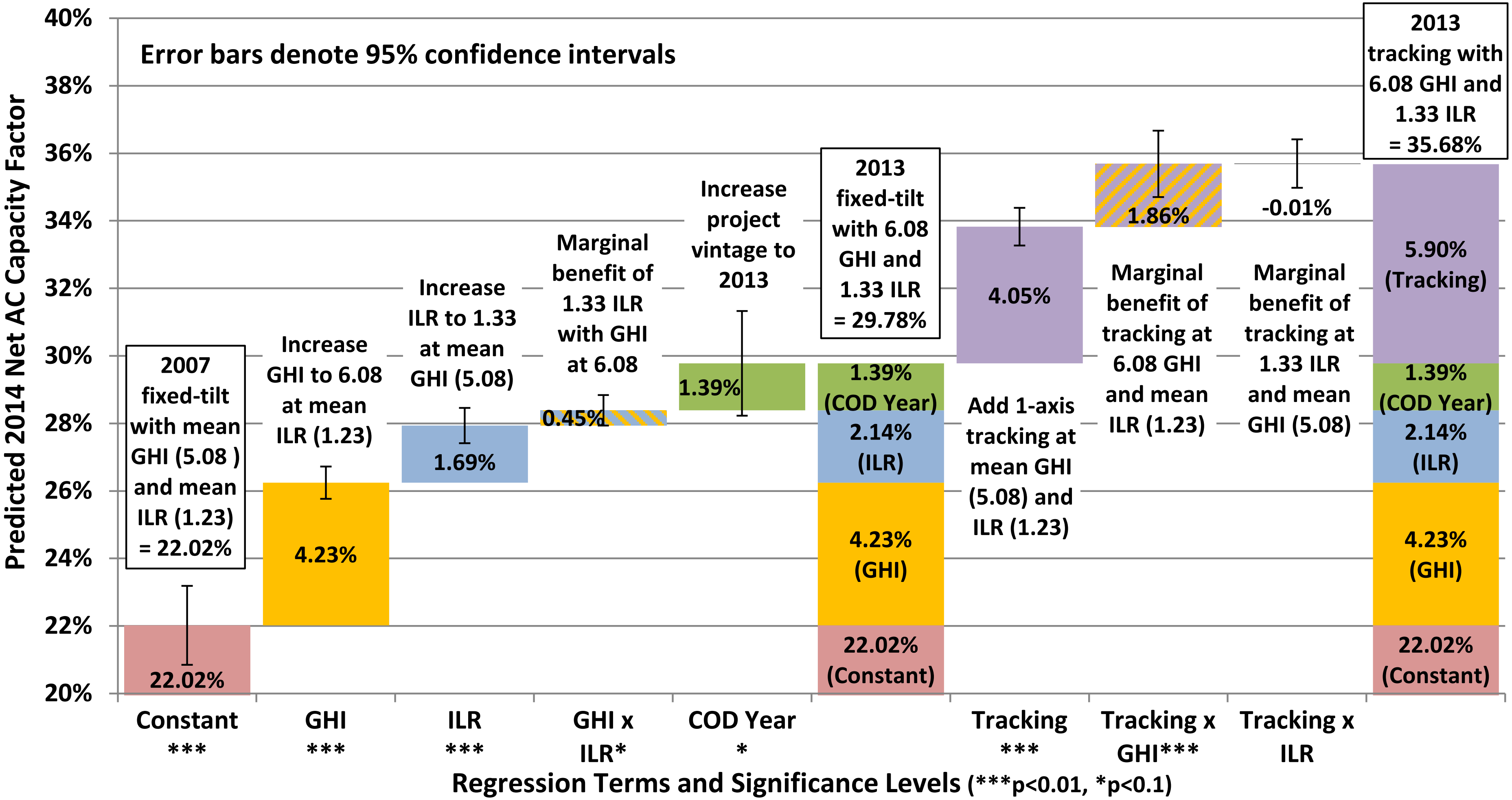
Blue-shaded columns (B):

- The benefit of a higher GHI increases with higher ILRs
- Moreover, for any given ILR, the benefit of a higher GHI is greater for tracking than for fixed-tilt projects (e.g., 8.9% vs. 6.1% for an ILR of 1.20)

Bottom third of table affirms earlier story:

- For a given ILR, the benefit of tracking increases with GHI
- For a given GHI, the benefit of tracking *decreases* slightly as ILR increases

Model 5 interpretation: summary graphic



Graph progresses from a 2007 fixed-tilt project with average GHI and ILR (on the far left) to a 2013 fixed-tilt project with a higher GHI and ILR (in the middle) to a 2013 tracking project with the same higher GHI and ILR (on the far right)

Conclusions

- The rapid deployment of utility-scale PV in the United States in recent years has resulted in a diverse fleet of operating projects that exhibit significant variation in empirical AC capacity factors
- The regression models developed for this analysis find that just 3 highly significant independent variables – GHI, Tracking, and ILR – can explain 92% of this variation (with GHI alone able to explain 71.6%)
- Adding a 4th independent variable (COD Year) and 3 interactive terms (Tracking x GHI, Tracking x ILR, GHI x ILR) improves the model further and reveals interesting relationships between these independent variables (e.g., the performance benefit of tracking increases with a higher GHI but diminishes with a higher ILR)
- The model would presumably be improved with good data on power temperature coefficients and module operating temperatures
- The empirical data and statistical modeling results presented in this paper can provide a useful indication of the level of performance that solar project developers and investors can expect from various project configurations in different regions of the country
- Moreover, the tight relationship between actual and fitted capacity factors from this relatively simple model should instill confidence among investors that the projects in this sample, at least, have largely performed as expected to date

Questions?

- Read the full report: https://emp.lbl.gov/publications/maximizing-mwh-statistical-analysis
- Watch a youtube video summary: https://www.youtube.com/user/EETDEMP/videos
- Contact the authors:
  - ➢ Mark Bolinger (MABolinger@lbl.gov)
  - ➢ Joachim Seel (JSeel@lbl.gov)

Interested in perusing our other renewable energy reports?

- ➢ Check out http://emp.lbl.gov/reports/re

This research was supported by funding from the U.S. Department of Energy's SunShot Initiative.