Evaluation of Solar Forecasting 2 Forecasts Using the Solar Forecast Arbiter

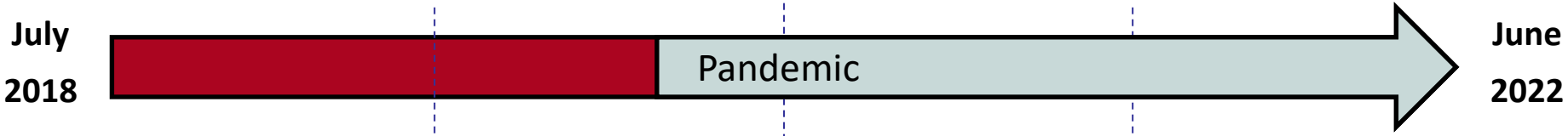
William F. Holmgren, Clifford W. Hansen, Justin Sharp, Aidan Tuohy, Adam Wigington, David P. Larson, Leland J. Boeman, Antonio T. Lorenzo, Anastasios Golnas

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

Solar Forecasting 2 Overview: $12M total, 8 awards, 3+ years

Lead Org	Topic	Title
UArizona	1	Open Source Evaluation Framework for Solar Forecasting
PNNL	2	Development of WRF-Solar v2
NREL	2	Probabilistic Cloud Optimized Day-Ahead Forecasting System based on WRF Solar
BNL	2	Advancing WRF-Solar Model to Improve Solar Irradiance Forecast in Cloudy Environments
UC San Diego	2	HAIMOS Ensemble Forecasts for Intra-day and Day-Ahead GHI, DNI and Ramps
NREL	3	Solar Uncertainty Management and Mitigation for Exceptional Reliability in Grid Operations
Johns Hopkins	3	Coordinated Ramping Product and Regulation Reserve Procurements in CAISO and MISO using Multi-Scale Probabilistic Solar Power Forecasts
EPRI	3	Probabilistic Forecasts and Operational Tools to Improve Solar Integration

Project Timeline for TA2/3 Evaluations



Year 1

- Teams discuss evaluation needs and constraints
- TA 2/3 research starts
- Solar Forecast Arbiter alpha

Year 2

- TA 2/3 research continues
- SFA development continues
- SFA reengages TA 2/3 teams on evaluations
- SFA betas

Year 3

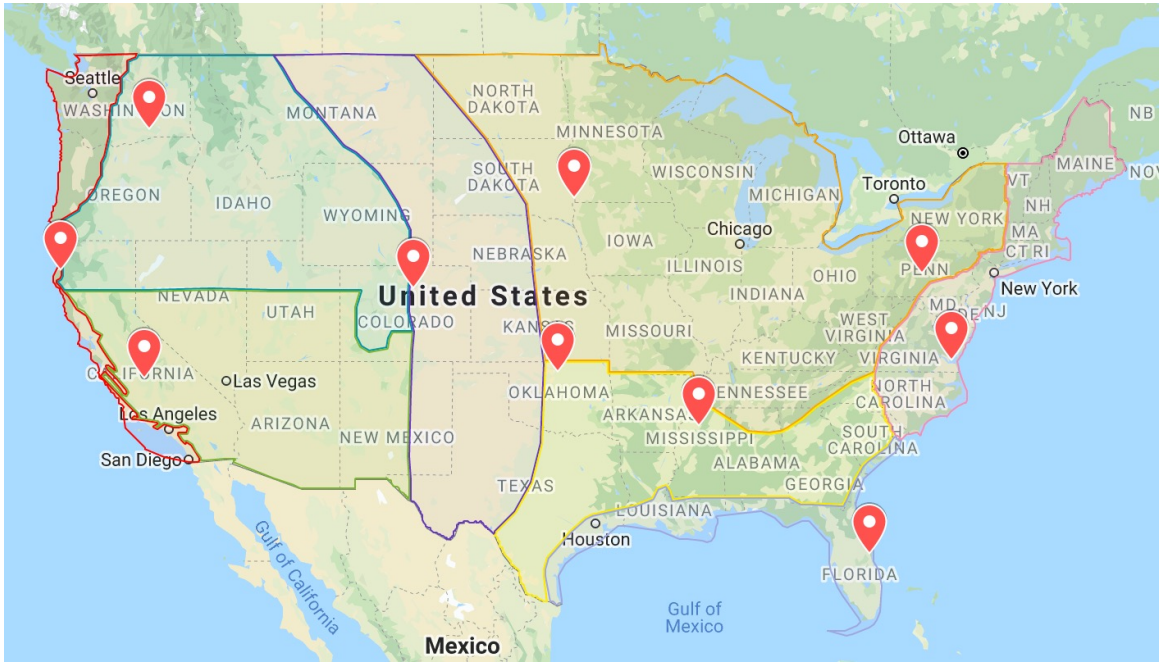
- SFA 1.0.0
- Iterate TA2 plan
- Iterate TA3 plan
- SFA bug fixes
- TA 2/3 research
- TA2 teams deliver retrospective forecasts

Year 4

- TA3 teams deliver operational forecasts
- SFA improvements, fixes
- Present results (today)
- Write papers

Site selection

- 10 sites
- Determined w/ TA2 teams
- At least 1 site per SFA climate zone
- Table Mountain, CO. SURFRAD
- Hanford, CA. SOLRAD
- Humboldt State, CA. MIDC
- Richland, WA. PNNL
- Sioux Falls, SD. SURFRAD
- Lamont, OK. ARM
- Goodwin Creek, MS. SURFRAD
- Cocoa Beach, FL. DOE RTC
- Langley, VA. NASA
- Penn State, PA. SURFRAD

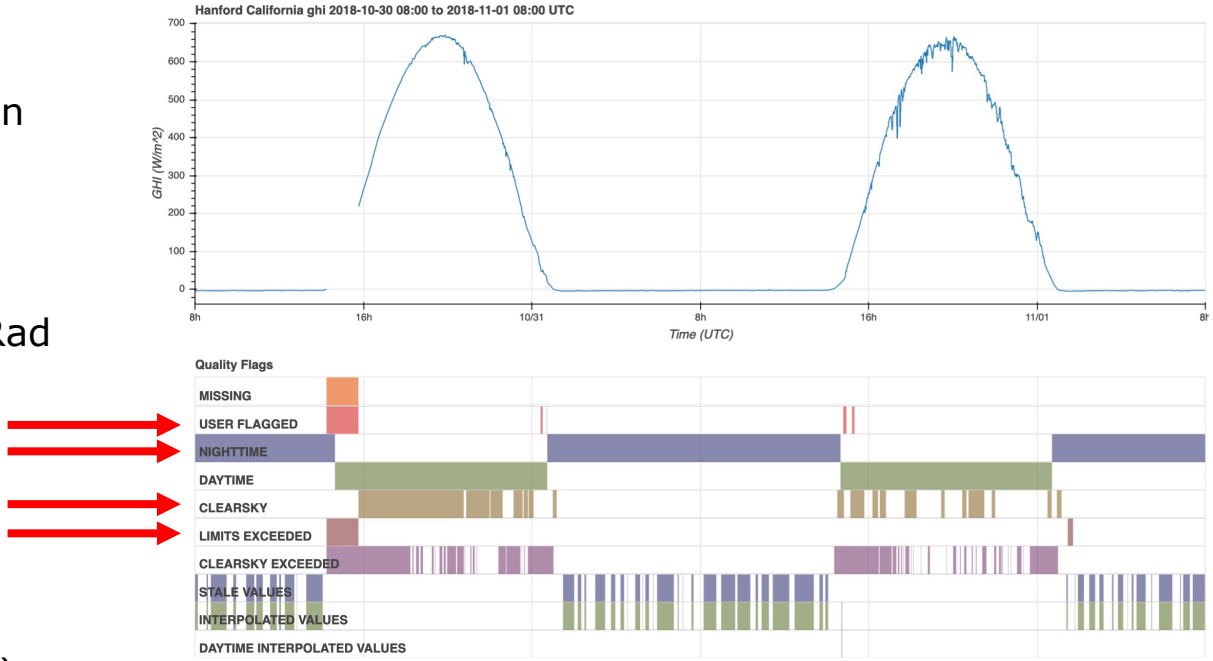


TA2/3 sites

Observation Data

Data filtering procedures

- SFA includes data validation toolkit
- Toolkit applied to all data when uploaded
- Separately we applied QCRad 3-component consistency test.
- Analysis excludes
  - User Flagged
  - Limits Exceeded
  - Nighttime (if > 5 min. in hour)



Observation data, flags from SFA Dashboard

Topic Area 2 Forecasts

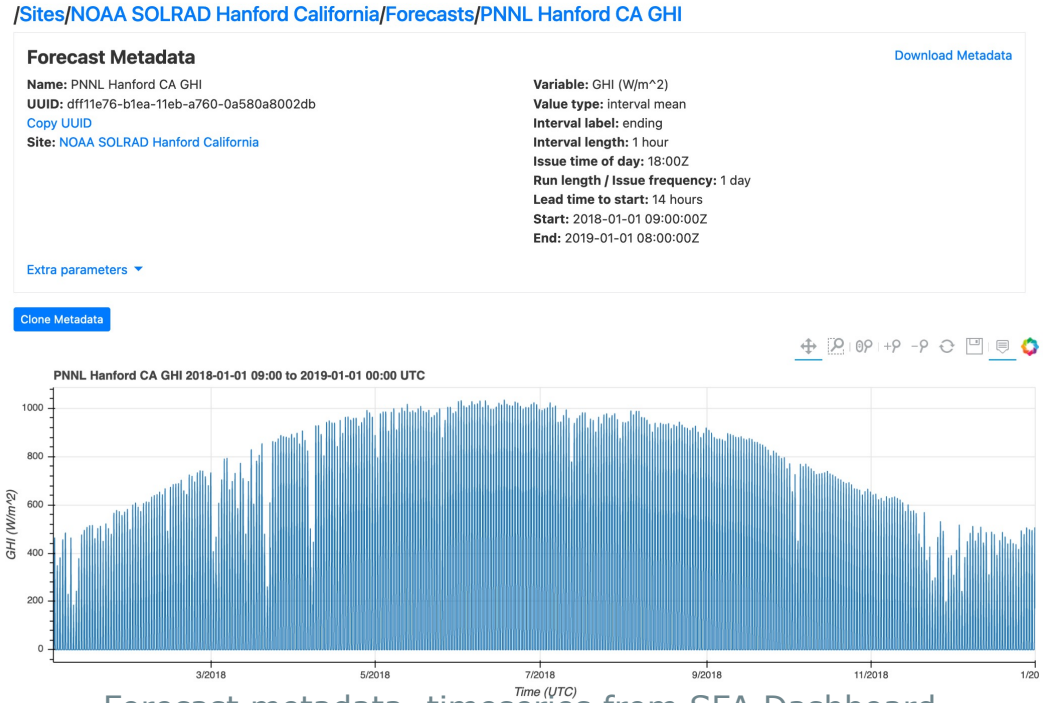
	New forecast	Reference forecast
PNNL	WRF Solar v2 + Vaisala post processing	WRF Solar v1 + fixes + Vaisala post processing
NREL	WRF Solar v2 ensemble mean	WRF Solar v1
BNL	WRF Solar v2	WRF Solar v1
UCSD	HAIMOS	NAM + Larson model

* Each WRF Solar model was configured differently!

Topic Area 2 Forecasts

Retrospective analysis of day ahead forecasts

- Jan – Dec 2018
- Variables: GHI, DNI
- Forecasts issued at 10 am local for each site
- Midnight to midnight (run length = 24h)
- Lead time = 14 hours
- Interval mean
- Interval length = 1h
- Interval label = ending



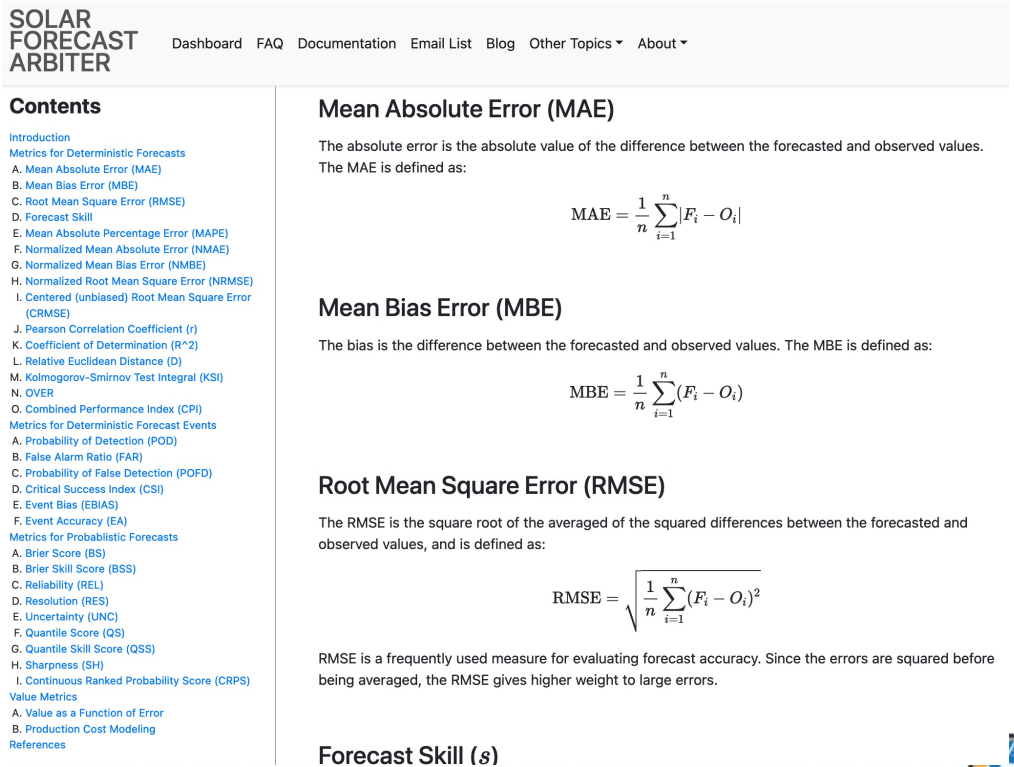
Forecast metadata, timeseries from SFA Dashboard

Topic Area 2 Metrics

"Give us your best forecast"

- MBE
- MAE
- RMSE
- CRMSE
- Pearson correlation
- Relative Euclidean distance
- Skill
- KSI
- OVER
- CPI

Metrics documentation



SOLAR FORECAST ARBITER

Dashboard FAQ Documentation Email List Blog Other Topics About

Contents

Mean Absolute Error (MAE)

The absolute error is the absolute value of the difference between the forecasted and observed values. The MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|F_i - O_i|$$

Mean Bias Error (MBE)

The bias is the difference between the forecasted and observed values. The MBE is defined as:

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)$$

Root Mean Square Error (RMSE)

The RMSE is the square root of the averaged of the squared differences between the forecasted and observed values, and is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)^2}$$

RMSE is a frequently used measure for evaluating forecast accuracy. Since the errors are squared before being averaged, the RMSE gives higher weight to large errors.

Forecast Skill (s)

Example TA2 Report

TA2 Evaluation Table Mountain GHI

Clone report parameters

This report of forecast accuracy was automatically generated using the Solar Forecast Arbiter.

This report can be downloaded as a standalone HTML file, standalone HTML file without timeseries or PDF file. The download is a ZIP archive that includes checksums for the report file and a PGP signature that can be used to verify the authenticity of the report. The Solar Forecast Arbiter PGP key ID is 0x22bd497c0930f8b0.

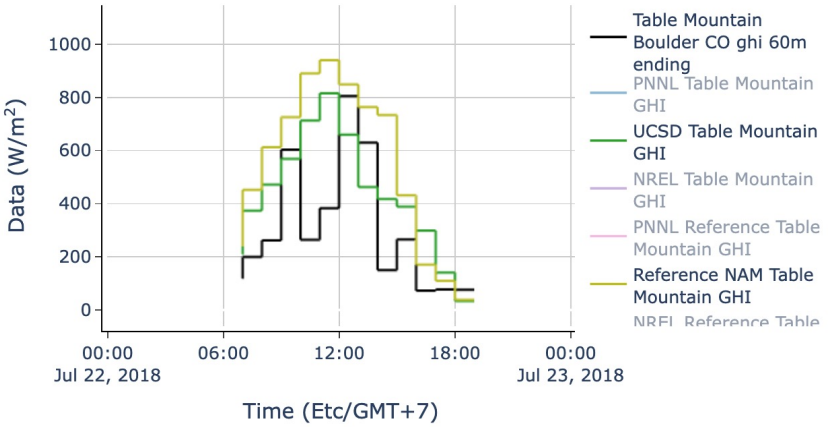
Report Metadata

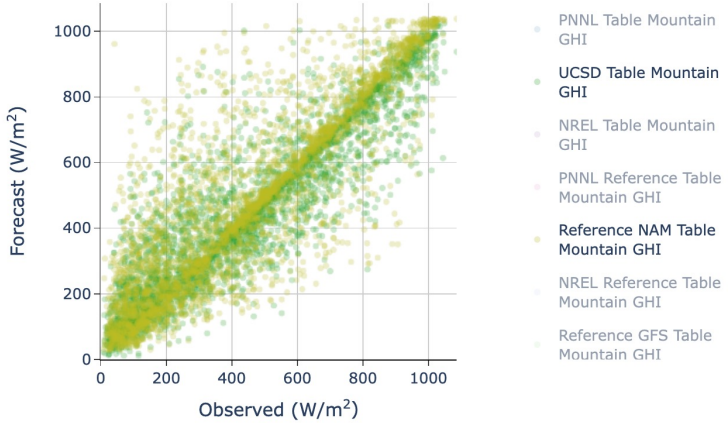
- Name: TA2 Evaluation Table Mountain GHI
- Start: 2018-01-01 07:00:00+00:00
- End: 2019-01-01 06:59:00+00:00
- Generated at: 2021-12-15 21:15:05+00:00

Data

This report includes forecast and observation data available from 2018-01-01 07:00:00+00:00 to 2019-01-01 06:59:00+00:00.

Time series plots





Example TA2 Report

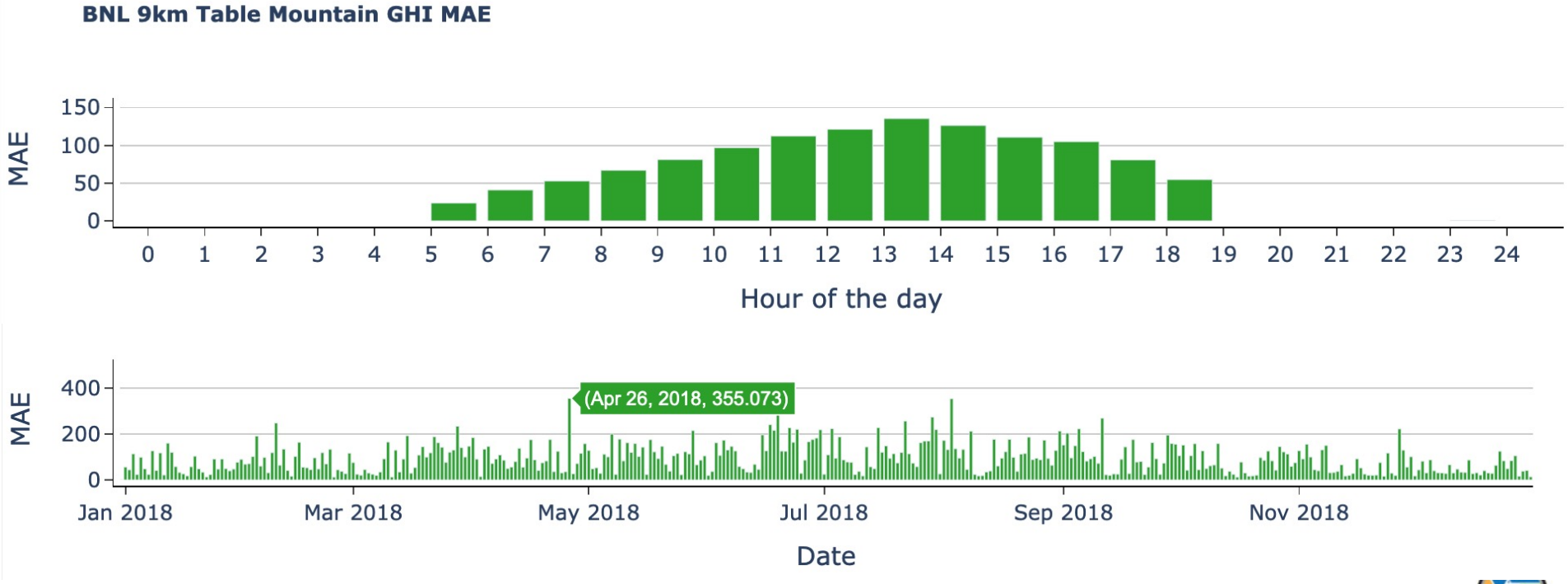
▾ Table of total metrics

Forecast	MBE	MAE	RMSE	CRMSE	r	Rel. Euc. Dist.	Skill	KSI	OVER	CPI
PNNL Table Mountain GHI	25.9	100	150	148	0.848	0.188	9.19e-03	32.7	16.3	87.3
UCSD Table Mountain GHI	1.43	80.6	119	119	0.904	0.141	0.215	28.8	6.45	68.1
NREL Table Mountain GHI	5.77	109	159	159	0.822	0.215	0.07	34.7	12.2	91.1
PNNL Reference Table Mountain GHI	28.3	101	151	149	0.846	0.189	nan	32.4	16.3	87.9
Reference NAM Table Mountain GHI	32.4	96.5	151	148	0.856	0.163	nan	32.4	9.33	86
NREL Reference Table Mountain GHI	51.7	105	171	163	0.822	0.217	nan	52	30	106
Reference GFS Table Mountain GHI	-15	116	177	176	0.787	0.224	nan	26.5	5.82	96.4
BNL 9km Table Mountain GHI	44.2	98.9	152	145	0.857	0.182	0.158	44.2	22	92.5
BNL 9km Reference Table Mountain GHI	82.5	117	180	160	0.824	0.266	nan	82.5	56.9	125

Metrics downloadable in csv/json

Example Report

Plots of metrics by hour, date, etc.

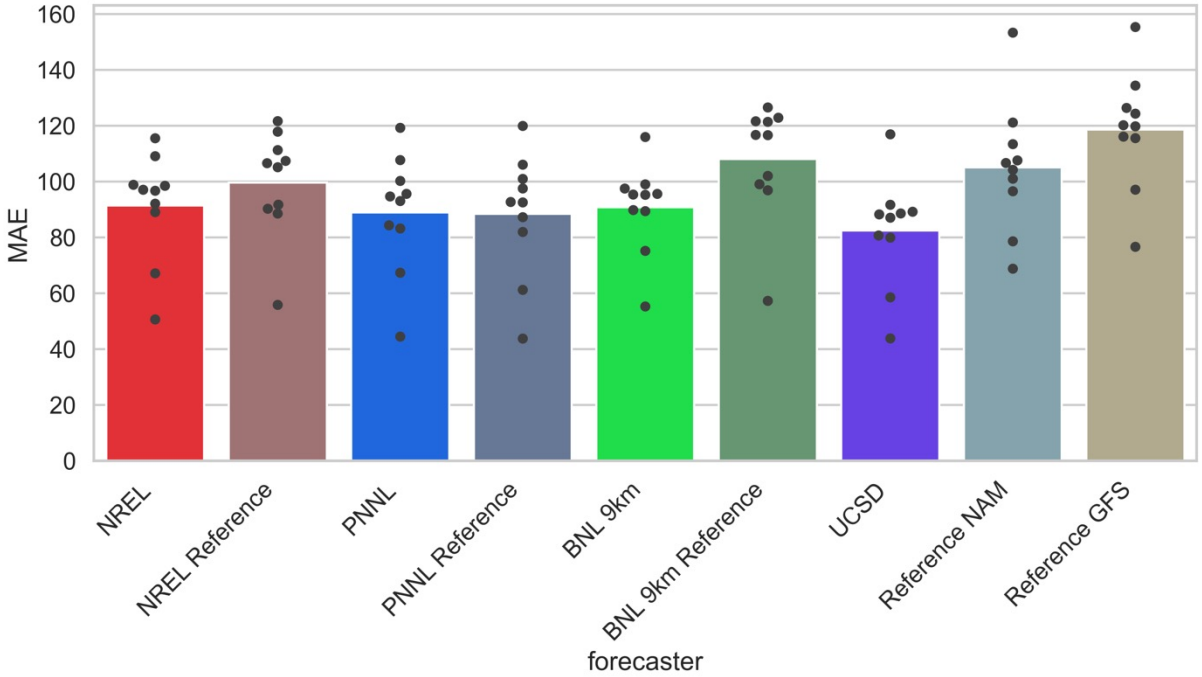


Topic Area 2 results

Total TA2 **GHI MAE** scores for 2018

Bars = average MAE across all sites

Points = MAE at each site

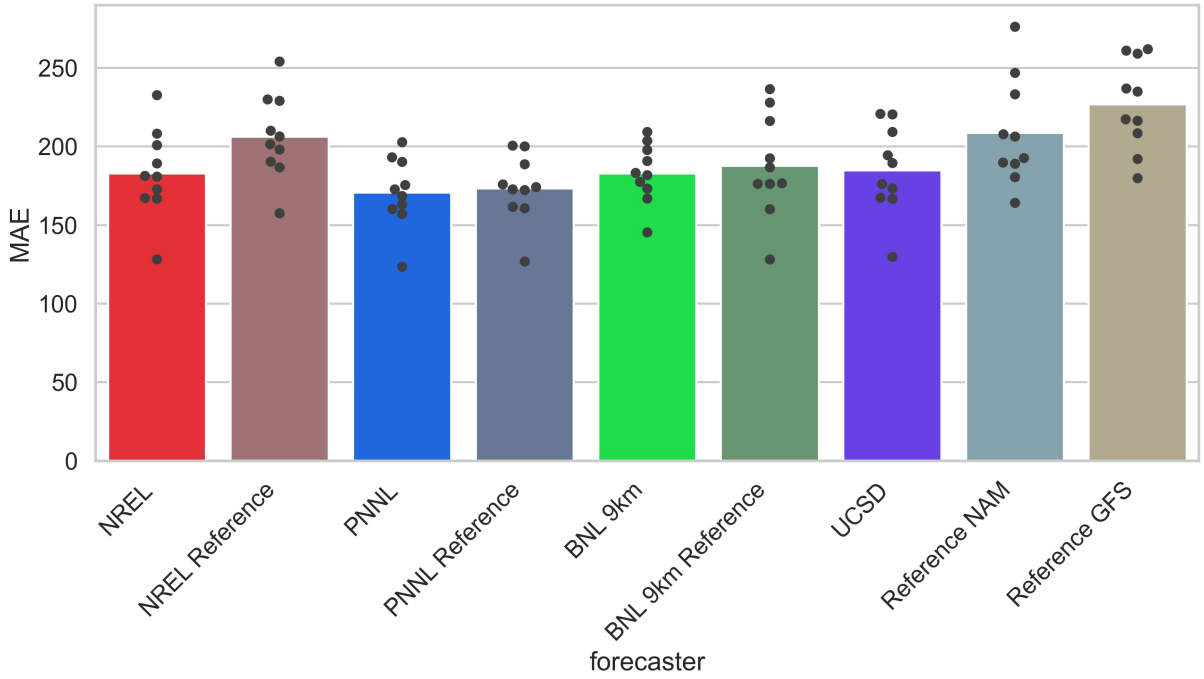


Topic Area 2 results

Total TA2 **DNI MAE** scores for 2018

Bars = average MAE across all sites

Points = MAE at each site

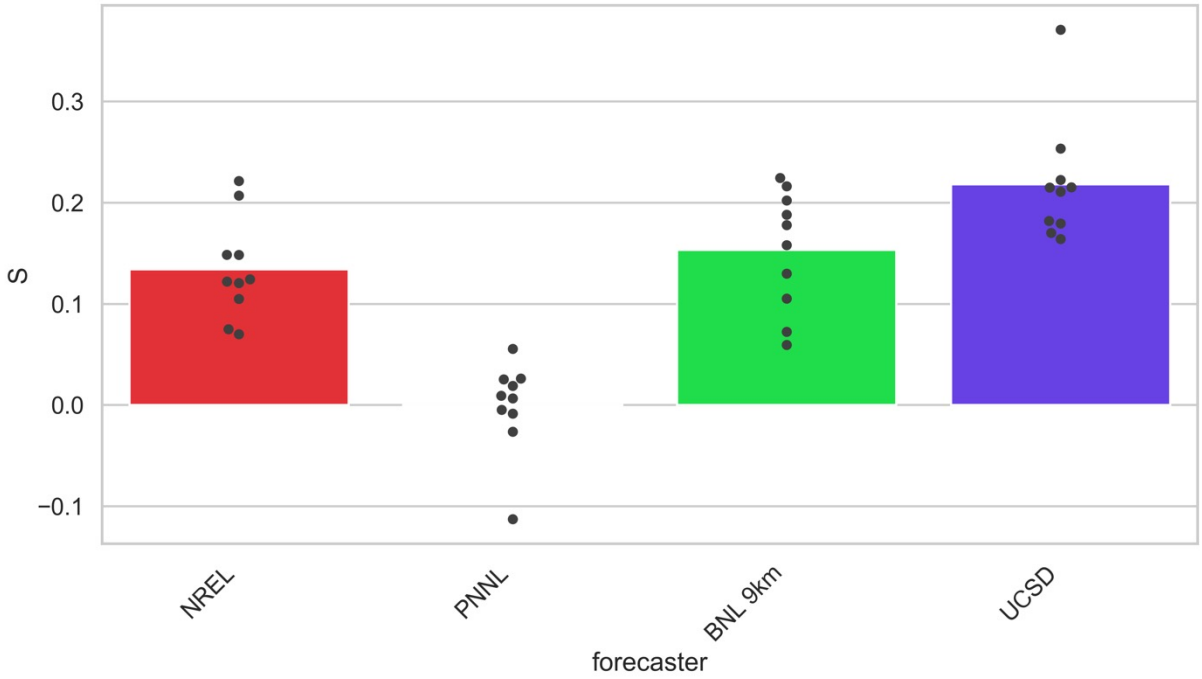


Topic Area 2 results

Total TA2 **GHI skill** scores for 2018

Bars = average skill across all sites

Points = skill at each site

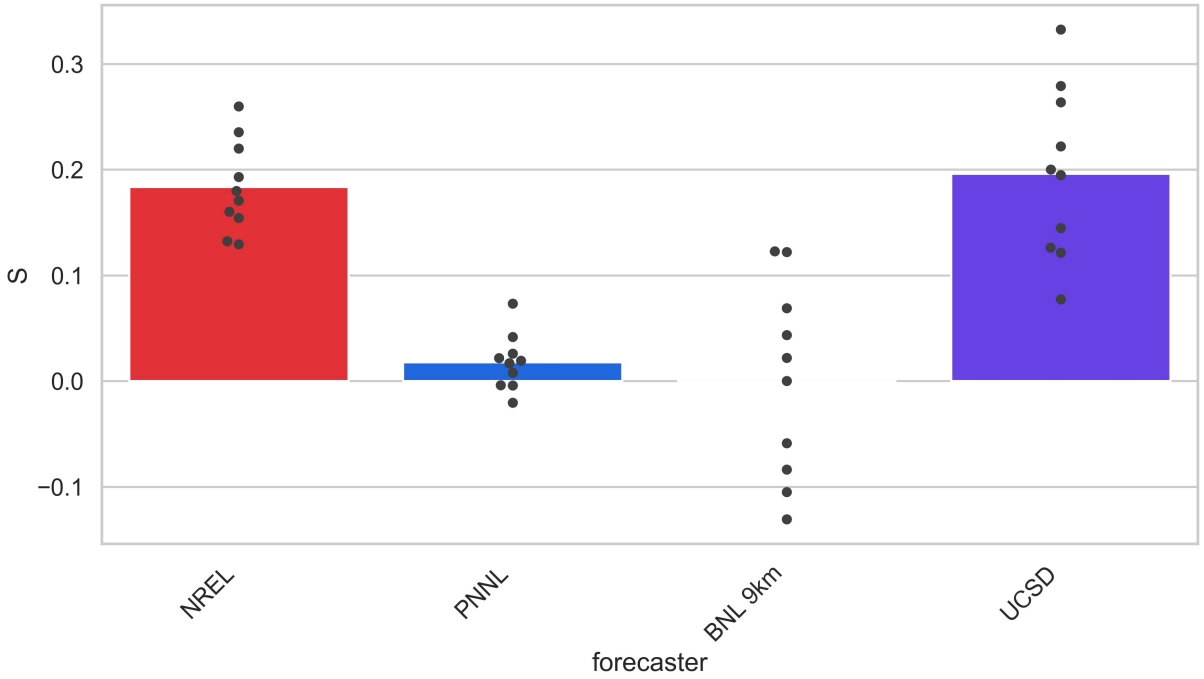


Topic Area 2 results

Total TA2 **DNI skill** scores for 2018

Bars = average skill across all sites

Points = skill at each site

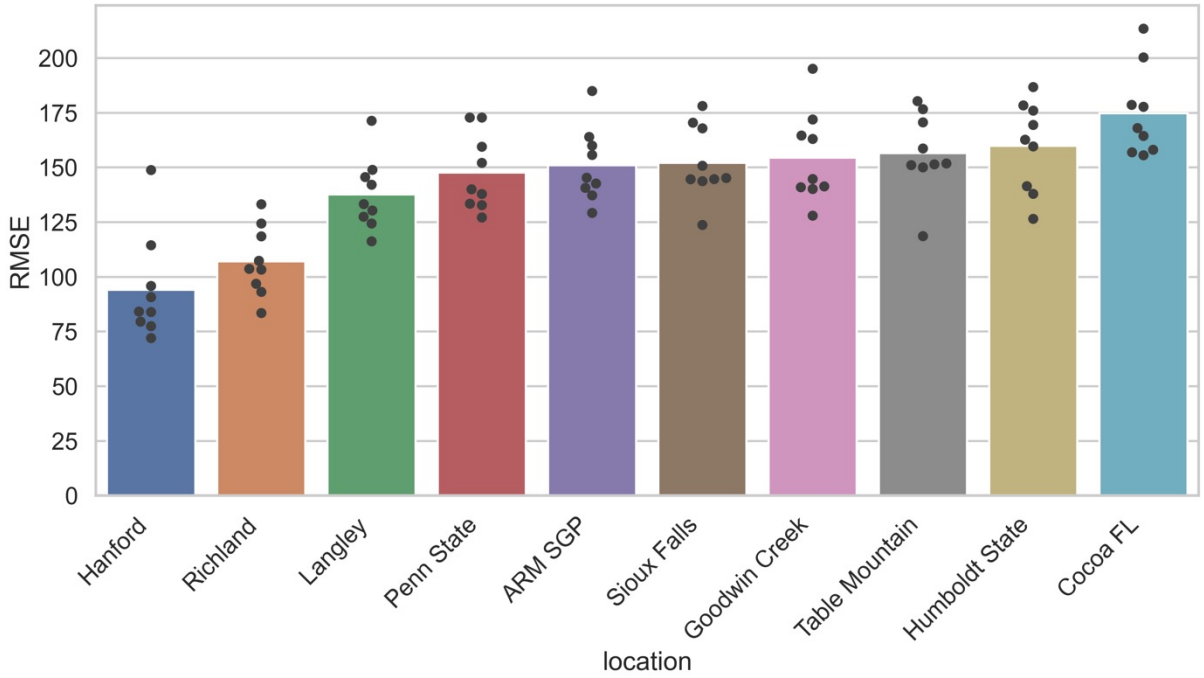


Topic Area 2 results

Total TA2 **GHI RMSE** scores for 2018

Bars = average RMSE across all forecasts

Points = RMSE of each forecast

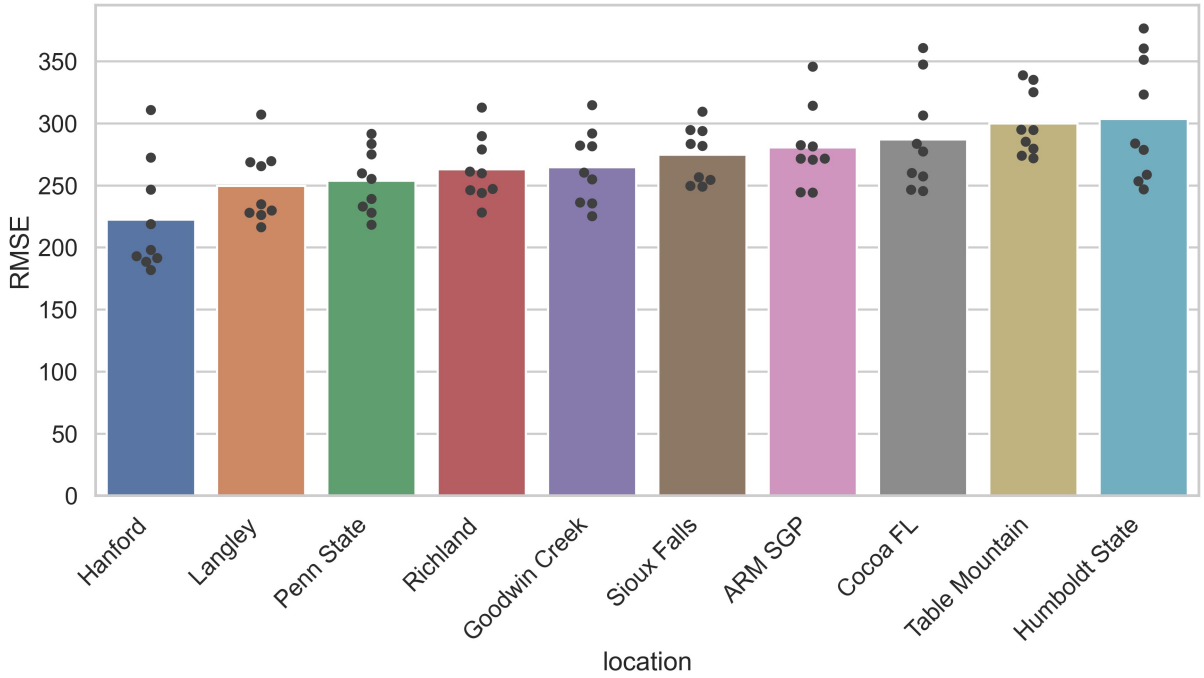


Topic Area 2 results

Total TA2 **DNI RMSE** scores for 2018

Bars = average RMSE across all forecasts

Points = RMSE of each forecast



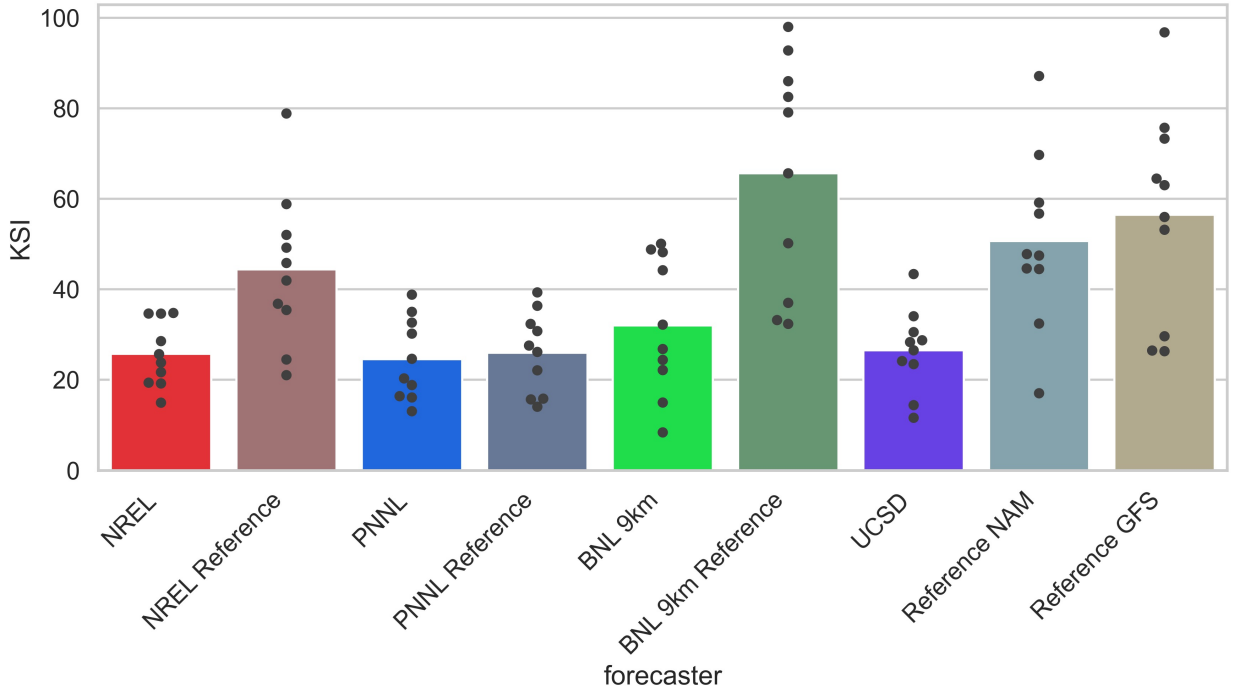
Topic Area 2 results

Total TA2 **GHI KSI** scores for 2018

Bars = average KSI across all sites

Points = KSI at each site

Greater improvement over reference than seen in MAE or RMSE



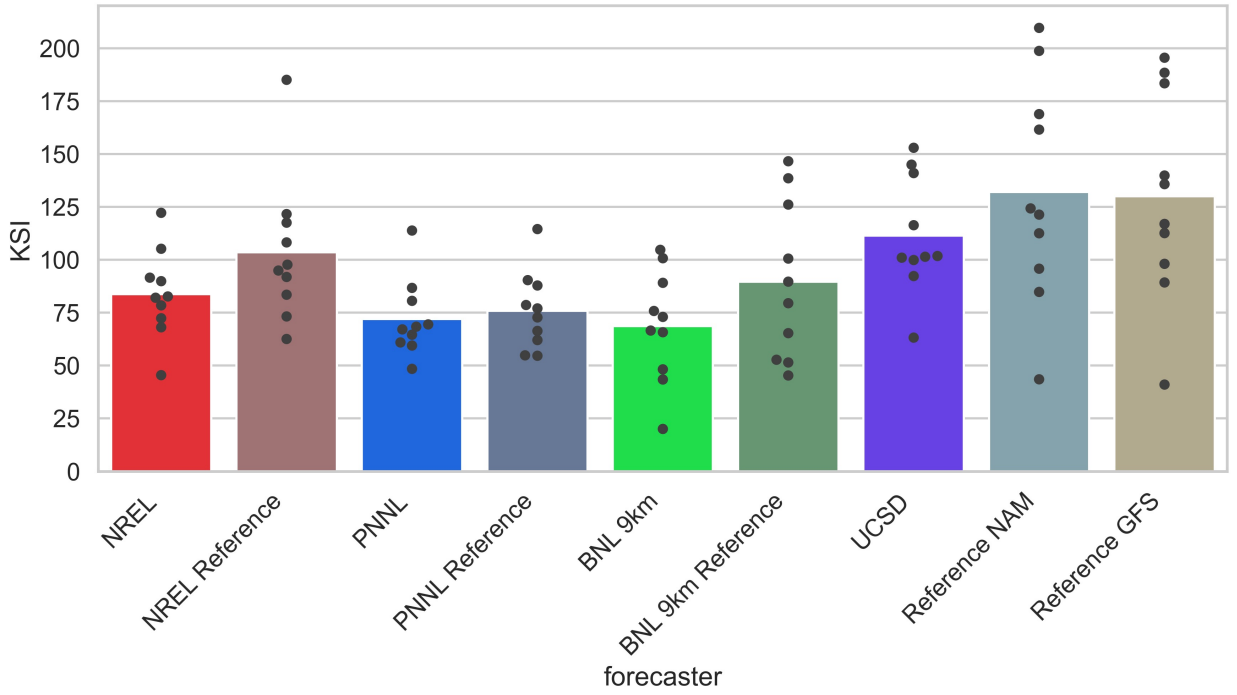
Topic Area 2 results

Total TA2 **DNI KSI** scores for 2018

Bars = average KSI across all sites

Points = KSI at each site

Greater improvement over reference than seen in MAE or RMSE



Topic Area 3 Forecasts

Operational analysis of day ahead forecasts

- Sep – Nov 2021
- Variables: GHI, DNI
- Forecasts issued at 10 am local for each site
- Midnight to midnight (run length = 24h)
- Lead time = 14 hours
- Interval mean
- Interval length = 1h
- Interval label = ending
- 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 98, 99
- Exclude periods when API unavailable

/Sites/NREL MIDC Humboldt State University/Probabilistic Forecasts/NREL Humboldt State GHI

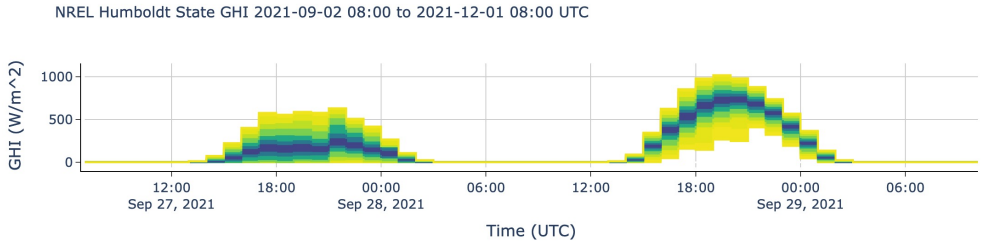
Probabilistic Forecast Metadata Download Metadata

Name: NREL Humboldt State GHI
UUID: 73a34dfa-efea-11eb-8c10-0a580a800235
Copy UUID
Site: NREL MIDC Humboldt State University

Variable: GHI (W/m^2)
Value type: interval mean
Interval label: ending
Interval length: 1 hour
Issue time of day: 18:00Z
Run length / Issue frequency: 1 day
Lead time to start: 14 hours
Axis: y
Start: 2021-09-01 09:00:00Z
End: 2021-12-11 08:00:00Z

Extra parameters

Clone Metadata



Forecast metadata, timeseries from SFA Dashboard

Topic Area 3 Metrics

- Quantile score (QS)
- Quantile skill score (QSS)
- Continuous Ranked Probability Score (CRPS)
- Continuous Ranked Probability Skill Score (CRPSS)

No instruction to optimize metrics

Metrics documentation

SOLAR FORECAST ARBITER

Dashboard FAQ Documentation Email List Blog Other Topics About

Contents

Metrics for Probablistic Forecasts

Probablistic forecasts represent uncertainty in the forecast quantity by providing a probability distribution or a prediction interval, rather than a single value.

In the probabilistic metrics below, we adopt the following nomenclature:

- $F_i(x)$: cumulative distribution function of a probability forecast for a continuous value x at each time i
- f_i : probability forecast for an event (e.g. $x >= x_0$) at time i
- o_i : indicator for whether an event occurred at time i: $o_i = 1$ if an event occurs at time i and $o_i = 0$ otherwise
- f_k : discrete values that appear in the set of probability forecasts f_i
- N_k : conditional sample size. The number of times each forecast value f_k appears in the set of probability forecasts f_i
- $n = \sum_{k=1}^{K} N_k$: number of forecast events
- $p(f_k) = \frac{N_k}{n}$: marginal distribution of the forecasts. The frequency of each forecast value f_k in the set of probability forecasts f_i
- $\bar{o}_k = p(o = 1 \| f_k) = \frac{1}{N_k} \sum_{i \in N_k} o_i$: the conditional average observation. Average of o_i at the N_k times when $f_k = f_i$
- $\bar{o} = \frac{1}{n} \sum_{i=1}^{n} o_i = \frac{1}{n} \sum_{k=1}^{K} N_k \bar{o}_k$: sample climatology of an event

Brier Score (BS)

The BS measures the accuracy of forecast probability for one or more events (Brier50). For events with binary outcomes, BS is defined as:

$$\mathrm{BS} = \frac{1}{n} \sum_{i=1}^{n} (f_i - o_i)^2$$

Topic Area 3 Reference Forecasts

Persistence ensemble (PeEn)

Time of day PeEn computes statistics from the past N days of observation data at the corresponding time of day.

1. Pull previous N days of data.
2. Resample data to desired interval length.
3. Bin data by desired times of day. Assuming no data gaps, there are N values in each bin.
4. For each bin, compute the desired percentiles.
5. Associate each bin with the forecast date time (e.g. first bin is midnight tomorrow, second bin is 1 am tomorrow, etc). This is the forecast.

SFA uses N = 30.

```
504
505 def persistence_probabilistic_timeofday(observation, data_start, data_end,
506                                          forecast_start, forecast_end,
507                                          interval_length, interval_label,
508                                          load_data, axis, constant_values):
509     r"""
510     Make a probabilistic persistence forecast using the *observation* from
511     *data_start* to *data_end*, matched by time of day (e.g. to forecast 9am,
512     only use observations from 9am on days between *data_start* and
513     *data_end*). This is a common variant of the Persistence Ensemble (PeEn)
514     method. [1]_ [2]_ [3]_
515
```

Reference forecast code on GitHub

solarforecastarbiter/reference_forecasts/persistence.py

Total CRPSS scores at 9/10 sites

Topic Area 3 Results

GHI

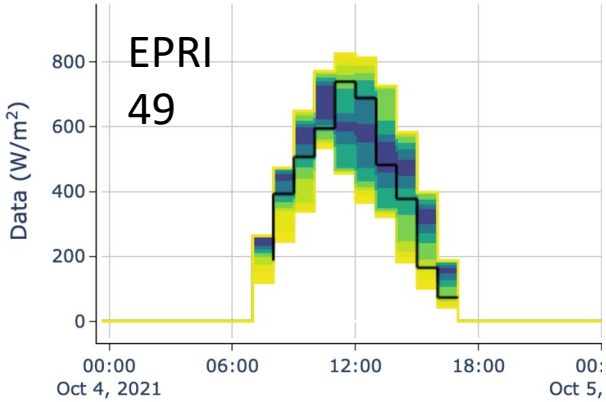
forecaster	EPRI	JHU	NREL
Cocoa Beach	4.2	-1.9	24.1
Goodwin Creek	15.1	-10.1	55.3
Hanford CA	33.5	-60.3	41.0
Humboldt State	30.5	-45.8	40.1
Lamont OK	31.1	-57.5	44.9
Penn State	37.3	22.3	49.0
Richland WA	43.8	-8.6	53.1
Sioux Falls	41.8	3.4	55.9
Table Mountain	30.4	-43.9	40.7
mean	29.7	-22.5	44.9

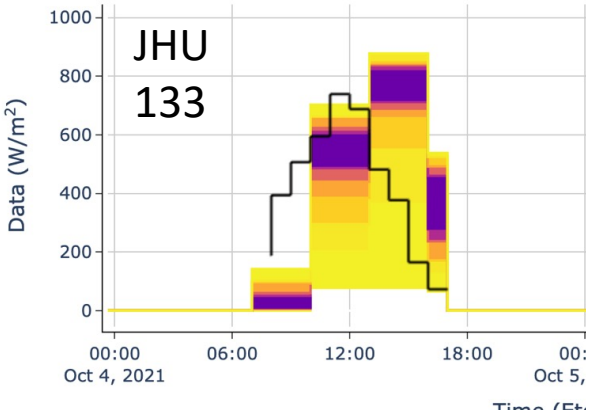
DNI

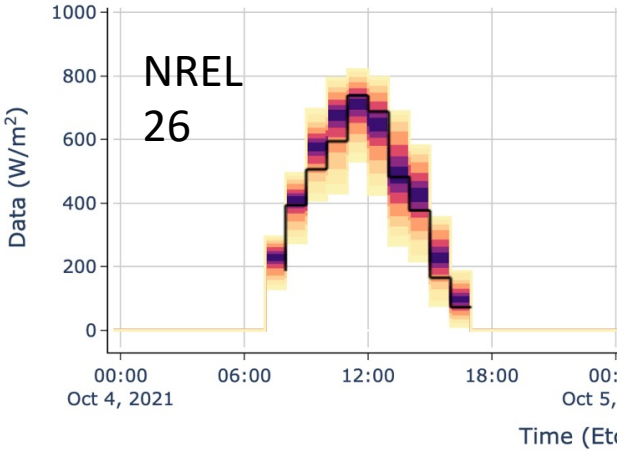
forecaster	EPRI	JHU	NREL
Cocoa Beach	23.1	26.0	25.6
Goodwin Creek	28.8	38.9	58.3
Hanford CA	28.1	-6.7	20.7
Humboldt State	42.7	36.4	44.7
Lamont OK	42.7	26.3	44.7
Penn State	50.2	53.4	55.6
Richland WA	50.3	25.0	52.0
Sioux Falls	44.5	40.6	56.4
Table Mountain	38.5	18.8	41.2
mean	38.8	28.7	44.4

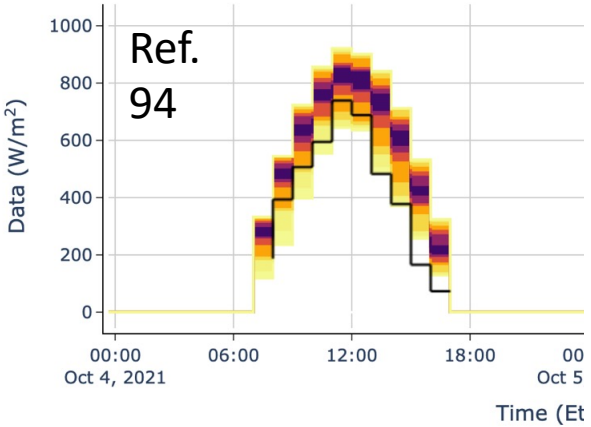
Topic Area 3 Results

- Team
- CRPS for this day



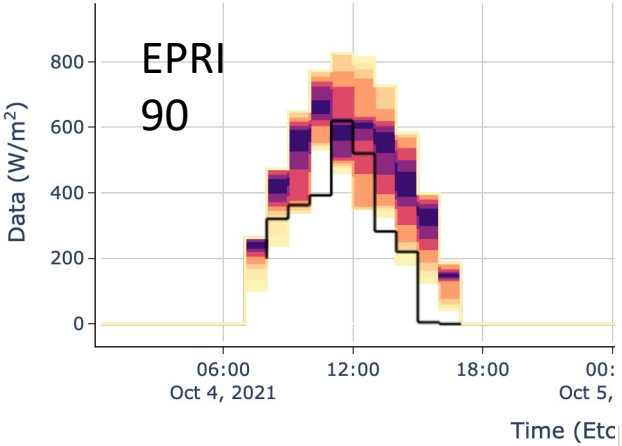


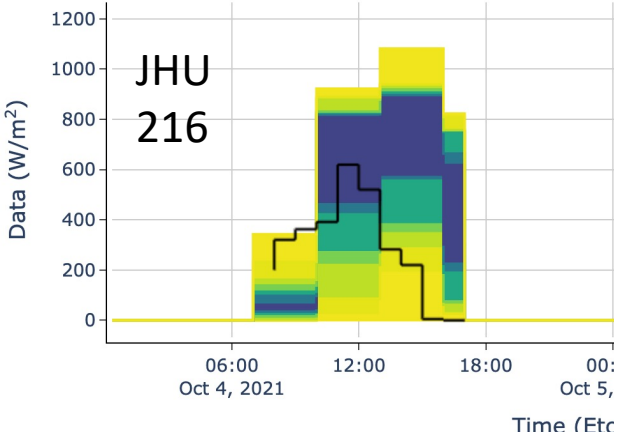


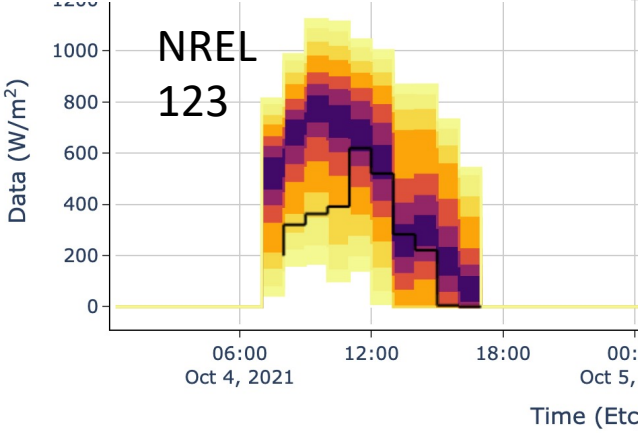


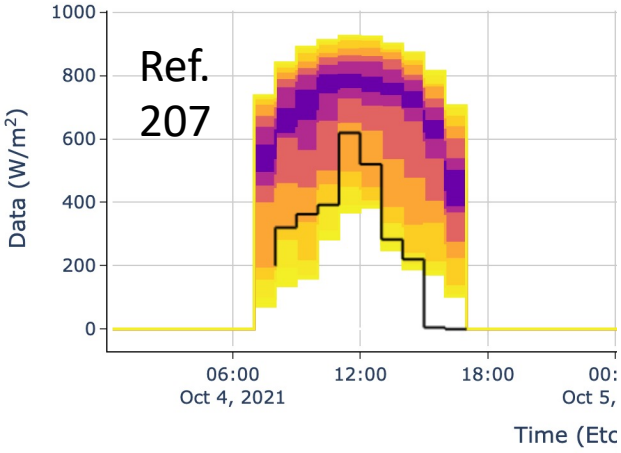
Topic Area 3 Results

- Team
- CRPS for this day

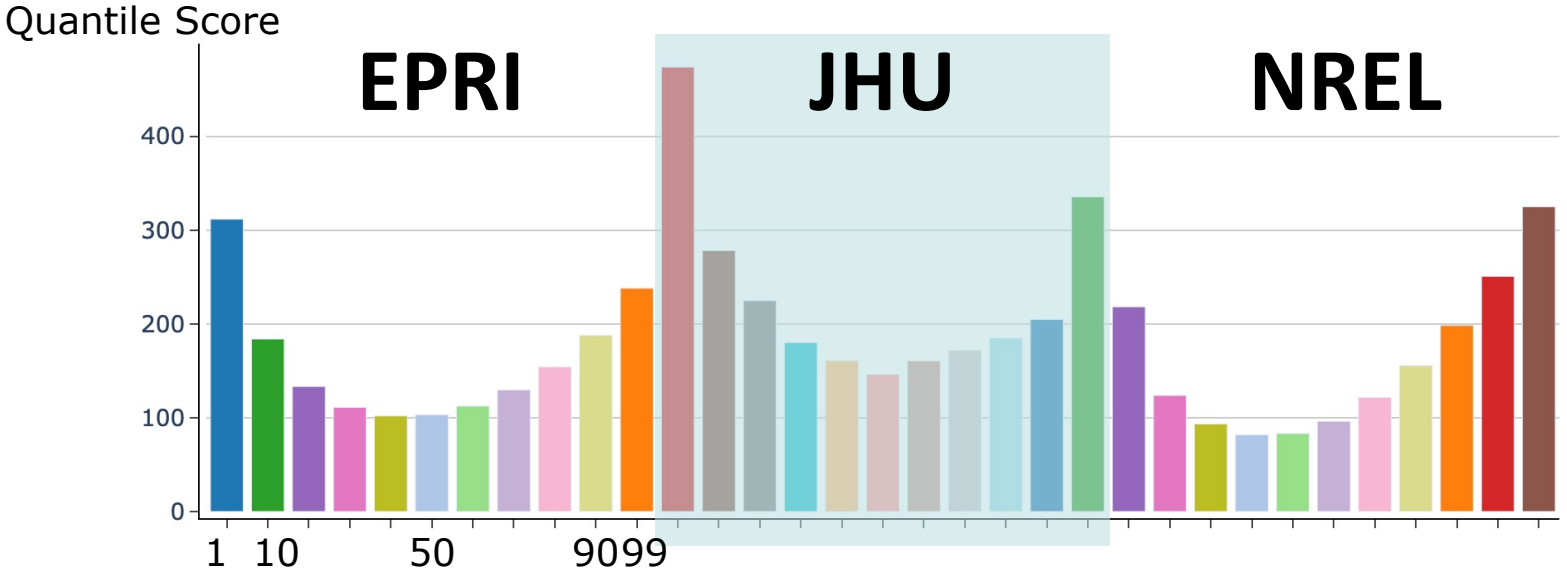








Topic Area 3 Results



Summary

- TA2: Deterministic forecasts show little MAE or RMSE improvement relative to **sophisticated** reference forecasts
- …but distribution metrics suggest forecasts are improved
- TA3: Significant probabilistic forecast skill for 2/3 teams (GHI) 3/3 teams (DNI)
- 3rd party validation provides useful insight
- SFA reports, input data, paper draft to be made public soon