

Making Regional Probabilistic Solar Forecasts from Deterministic Forecasts with Open-Source Tools and Data

William B. Hobbs

Southern Company, Birmingham, AL, 35203, USA

Abstract—Grid operators in areas with lots of solar generation need solar power forecasts to support decision making. Most forecasts that are available are deterministic, and most of the ones that are probabilistic are for single plants, meaning don’t have combined probabilities for aggregations of all the plants in a region. Additionally, few methods for producing probabilistic forecasts are easy to replicate or deploy for grid operators. This work presents a method for producing regional probabilistic forecasts from an existing deterministic forecast using open source tools and freely available data. The method is demonstrated using a deterministic forecast that is produced with similar tools and data. Reference code to replicate this work, which is freely available, is also introduced.

I. INTRODUCTION

Grid operators in regions with meaningful amounts of solar generation typically make use of solar power forecasts. These forecasts can cover minutes, hours, or days in the future, and are used to support operational decisions like scheduling other generation assets. The most common type of forecasts is deterministic, with one estimated value for each time interval in the future. Probabilistic forecasts, which include information about uncertainty or the range of possible outcomes, may be able to help improve decision making related to economics and reliability. However, there are less tools and products available for probabilistic forecasts than for deterministic forecasts, and this is even more the case when considering aggregations of solar plants across a grid operator’s region.

The work here presents a method for producing regional probabilistic forecasts from an existing deterministic forecast using open source Python tools, such as pvlib [1], XGBoost [2], scikit-learn [3], and and freely available data from NOAA. In practice, the deterministic forecast used in this method would likely be the most accurate forecast that a grid operator has available to them, but for the purposes of demonstration, a deterministic forecast was produced for this work using similar tools and data as the probabilistic forecast, with methods similar to the reference forecasts in [4]. This work focuses on day-ahead forecasts, i.e., a forecast provided by approximately 6 AM local time covering the following day, approximately 18–42 hours ahead, but it is expected that the methods could be applied to intraday forecasts, e.g., 4 hours-ahead, and multi-day-ahead forecasts.

Reference Python code to replicate this work is available in a repository on GitHub under an open-

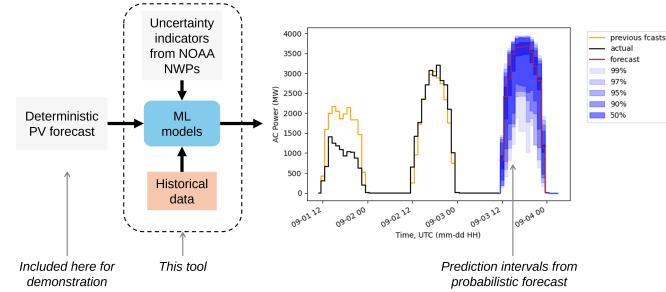


Fig. 1. Schematic overview of this work. Additional feature inputs the ML models, such as solar position, are covered in Subsection III-B.

source license, accessible at <https://github.com/williamhobbs/solar-fleet-forecast-probability-tool>.

II. OVERVIEW

An existing solar fleet in the Southern Company Balancing area was modeled, with several years of deterministic forecasts produced for the day-ahead horizon using archived outputs from the NOAA National Centers for Environmental Prediction (NCEP) HRRR [5]. A machine learning (ML) probabilistic forecast was produced using a mix of Random Forest and XGBoost quantile regressions, after normalizing all power values to clear sky. Features for training those regression models include common variables like actual generation, deterministic forecasts, solar position, month, and two more novel features: variation in weighted average cloud cover from the NCEP GEFS ensemble [6], and aggregated spatial variation from a downsampled version of the HRRR.

For demonstration purposes, a realistic fleet of solar plants was modeled using metadata from EIA Form 860 (e.g., approximately 50 plants in the Southeast US). A deterministic day-ahead forecast was produced for this regional fleet using the HRRR model, with GRI2B files being downloaded and processed using Brian Blaylock’s Herbie python package [7], available on GitHub at <https://github.com/blaylockbk/Herbie>. pvlib was used to convert this to a power forecast for each site, and then the forecast for each site was summed. Actual generation was modeled using HRRR analysis (F00).

A schematic overview of this work is illustrated in Fig. 1.

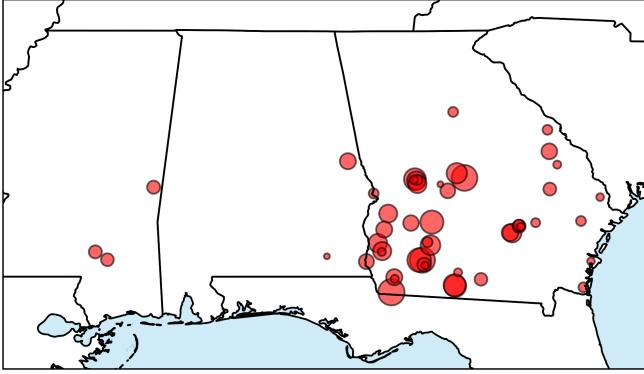


Fig. 2. A map of the solar plants being forecasted for, covering the region of the Southern Company Balancing Area. Marker size corresponds to plant AC capacity, which totals to about 3,700 MW for the whole fleet.

III. METHODS

A. Setup

A list of solar plants is pulled from EIA Form 860 data [8], making reasonable assumptions about temperature coefficients, ground coverage ratios, and tracker systems based on thin film or crystalline designations. The list was filtered on sites in the Southern Company Balancing Area. Resulting sites are shown in Fig. 2, totaling about 3,700 MW of AC capacity.

Actual weather was based on HRRR analysis (F00). The NREL NSRDB [9] was initially used for actual weather, but issues were found during high solar zenith angles on overcast days, as documented in [10], [11], and [12], which resulted in unrealistically large forecast errors.

The PVWatts [13] DC power and inverter models in pvlib were used for modeling DC and AC power with both actual weather clear sky irradiance conditions for each plant. Clear sky power was used to normalize actual power to a clear sky index, and later was used to similarly normalize forecasts.

B. Features

Features used for training the quantile regression models are described in the following subsection.

1) *Deterministic forecast:* In a real-world implementation of the work described here, the input deterministic forecast would be based on the best-available forecast, which could be a commercial forecast or product of an ensemble. A key aspect of that forecast is that it is well-tuned to match the fleet of plants being forecasted. For the purposes of demonstration in this work, the NOAA HRRR was used with pvlib for day-ahead forecasts. Pairing this with modeled actual generation using HRRR F00 and the same pvlib configuration should reasonably replicate the results of a well-tuned operational forecast with real generation data.

Herbie was used to retrieve GRIB2 files with initialization times of 06:00Z (midnight Central Standard Time) and lead times of 27 to 44 hours (covering daytime in the day ahead). Variables included global horizontal irradiance (GHI), ambient temperature, and wind speed. The HRRR has a native resolution of approximately 3 km × 3 km, which was found to

add too much noise in day-ahead forecasts. To smooth out the HRRR, the model was coarsened by calculating the mean across a 10 × 10 grid (approximately 30 × 30 km).

2) *Ensemble spread in cloud cover from GEFS:* The first novel feature used was the variation in cloud cover across 30 ensemble members of the NOAA GEFS model. Herbie was used to retrieve total cloud cover (TCC) for each GEFS member, then for each member, calculate average cloud cover over all solar plants, weighted by each plant's AC capacity. Finally, calculate standard deviation of those averages across all 30 ensemble members. This is calculated as:

$$X_{TCC,j} = \frac{\sum_i w_i \times TCC_i}{\sum_i w_i} \quad (1)$$

$$\sigma_{X_{TCC,j}} =$$

$$\sqrt{\sum (X_{TCC,j} - \mu)^2 / N}$$

where the weighted average total cloud cover in ensemble member j is $X_{TCC,j}$, w_i is the AC capacity of each of i plants, and TCC_i is the total cloud cover for each plant. Then the standard deviation of weighted average TCC, $\sigma_{X_{TCC}}$, is calculated from each value $X_{TCC,j}$, the average of all values of X_{TCC} , μ , and the number of ensemble members, N (30 in the case of GEFS).

An example illustration, using the first three of the 30 ensemble members, is shown in Fig. 3 for a relatively high uncertainty day, and in Fig. 4 for a low uncertainty day.

The intention here was to represent the uncertainty in the future state of the atmosphere, and to weight it geographically according to the solar fleet. This is done in a way that is independent from the deterministic forecast, leaving room for separate selection of the best available deterministic forecast.

3) *Spatial variation from HRRR:* The second novel feature was a form of spatial variation from a downsampled version of the HRRR. Similar to the coarsening performed for the deterministic forecast, the HRRR model was downsampled with a 10 × 10 window size, but in the case the maximum and minimum GHI for each window was calculated, representing the range of irradiance forecasted in the neighborhood of each plant. These values were used to model a form of worst-case and best-case for the whole fleet for each forecast interval, which were then normalized to clear sky, and the difference between the two was calculated. Spatial variation, denoted V , is calculated as follows:

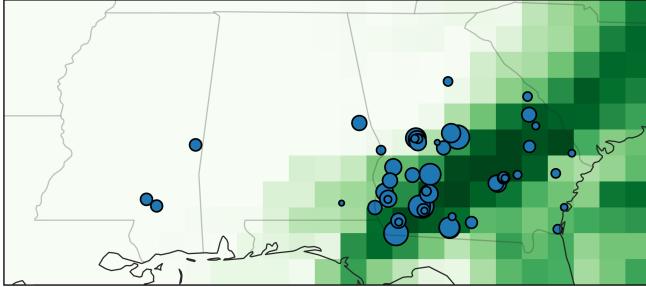
$$P_{min,i} = f(\min_k GHI_{i,k}, \dots) \quad (3)$$

$$P_{max,i} = f(\max_k GHI_{i,k}, \dots) \quad (4)$$

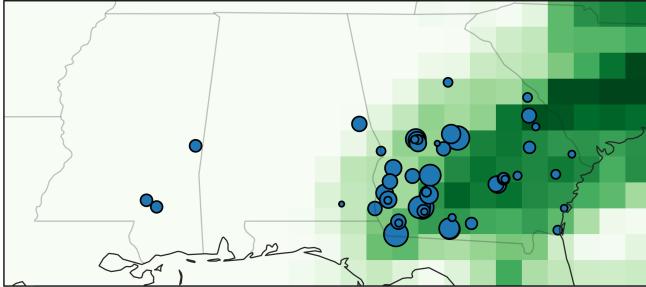
$$V = \frac{\sum_i P_{max,i} - \sum_i P_{min,i}}{\sum_i P_{cs,i}} \quad (5)$$

$P_{min,i}$ is the minimum power for plant i , calculated as a function of the minimum global horizontal irradiance (GHI) value from the k grid cells in the 10 × 10 window around the

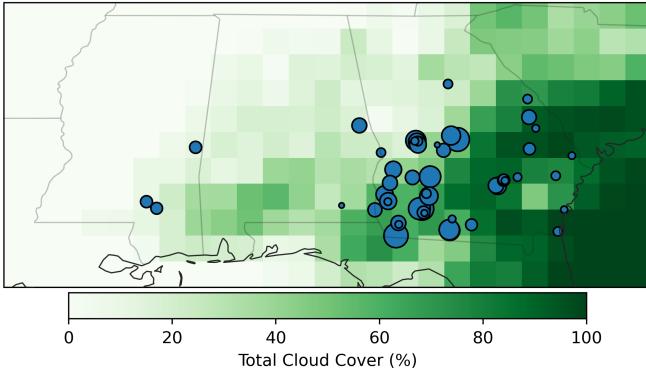
GEFS - Ensemble Member p01
Initialized: 06:00 UTC 10 Apr 2021
Valid: 18:00 UTC 11 Apr 2021



GEFS - Ensemble Member p02



GEFS - Ensemble Member p03



0 20 40 60 80 100
Total Cloud Cover (%)

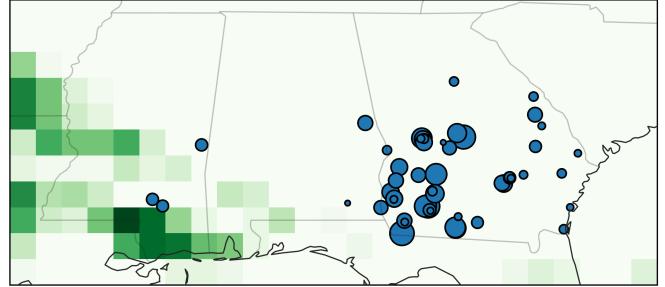
Fig. 3. Maps of GEFS Total Cloud Cover (TCC) for three of 30 ensemble members for a sample forecast initialization and valid time on a day with relatively high standard deviation of weighted average TCC: April 11, 2021. Solar plants are overlaid as blue markers, with marker size indicating AC capacity. The ensemble spread in cloud cover feature is calculated by first calculating the weighted average cloud cover in each member (weighted by plant AC capacity) and then calculating the standard deviation of across all 30 ensemble members.

plant. $P_{max,i}$ is similar, but for the maximum GHI value. And $P_{cs,i}$ is the clear sky modeled power for each plant.

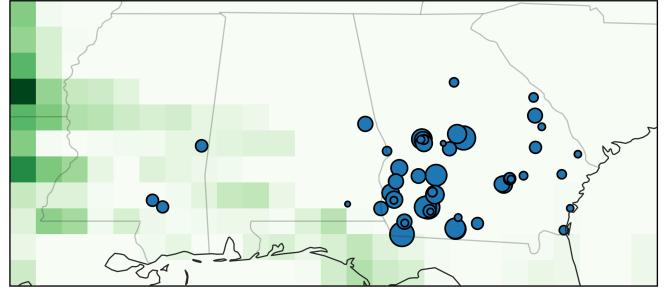
Example data for $\sum_i P_{max,i}$, $\sum_i P_{min,i}$, $\sum_i P_{cs,i}$, along with forecasted and actual power, and are shown in 5, covering the two days from Figs. 3 and 4. A sample map of GHI before downsampling is shown in 6. A hypothetical plant is labeled, along with the 10×10 downsampling region it would fall in. The average irradiance, used in the deterministic forecast, is approximately 500 W/m^2 , while the minimum is below 200 W/m^2 and the maximum is above 800 W/m^2 .

4) Additional features: Additional features that were found to improve results included month of the year and solar azimuth and zenith calculated for central point in the region.

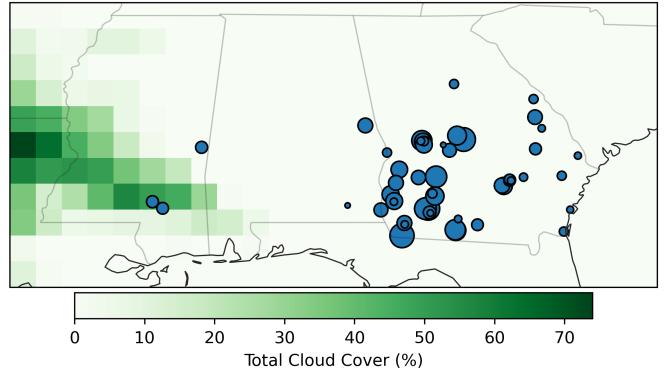
GEFS - Ensemble Member p01
Initialized: 06:00 UTC 11 Apr 2021
Valid: 18:00 UTC 12 Apr 2021



GEFS - Ensemble Member p02



GEFS - Ensemble Member p03



0 10 20 30 40 50 60 70
Total Cloud Cover (%)

Fig. 4. Maps of GEFS Total Cloud Cover (TCC), similar to 3, but on a day with low standard deviation of weighted average TCC: April 12, 2021.

Other features that were tested and not found to be helpful included hour of the day and individual maximum and minimum values from the spatial variation step (rather than combining them by calculating the difference).

C. Need for Machine Learning Models

Initial explorations of the relationship between forecast error and one or both of standard deviation of weighted average TCC and HRRR spatial variation showed promise. For example, in the three days illustrated so far (April 10-12, 2021), errors tend to be high when one or both of those values are high, as can be seen in Fig. 7.

This is also visible in a scatter plot of data from the same period, shown in Fig. 8.

However, when expanding the data set to cover many months, these correlations weakened. This can be seen in Fig. 9, where two years of data (2021-2022) are included. There

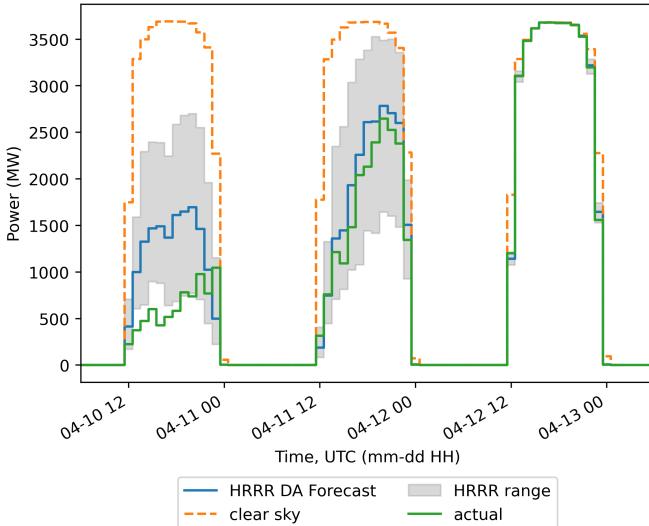


Fig. 5. Plot showing the spatial variation in output from the HRRR (gray), along with day-ahead forecasted power (blue), actual power (green), and modeled clear sky power (orange), covering April 10-12, 2021.

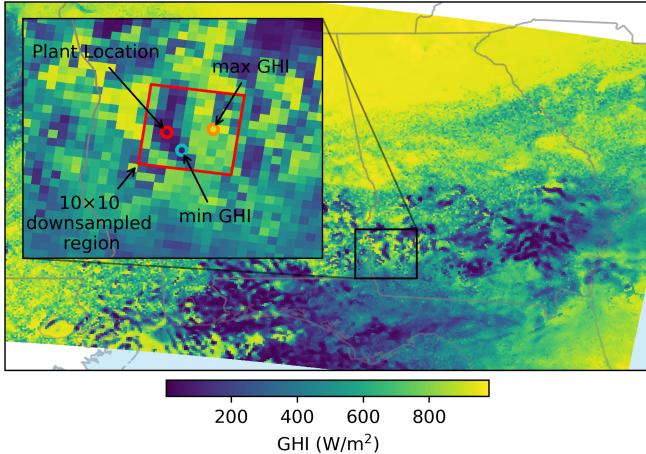


Fig. 6. Map of GHI from HRRR for the day-ahead forecast valid April 9, 2021, 18:00 UTC covering the Southeast US. The inset includes a hypothetical plant location (red circle), the window of 10×10 HRRR grid cells that are downsampled, and approximate locations of the minimum (cyan) and maximum (orange) GHI that is returned when downsampling.

are examples where one or both indicator values are low but error is high.

This complex behavior is what lead to the use of machine learning models, specifically quantile regressions.

D. Quantile Regression models

These features, along with the target of clear sky normalized actual generation, were used with a random forest quantile regression from quantile-forest [14], which is based on scikit-learn [3], and an extreme gradient boosting quantile regression [2], available at <https://github.com/dmlc/xgboost>. To avoid over-fitting, the XGBoost model was initially trained with 2021 data and tested with 2022, then finally evaluated with 2023 data. The random forest model was also trained with 2021, tested with 2022, and re-evaluated with 2023.

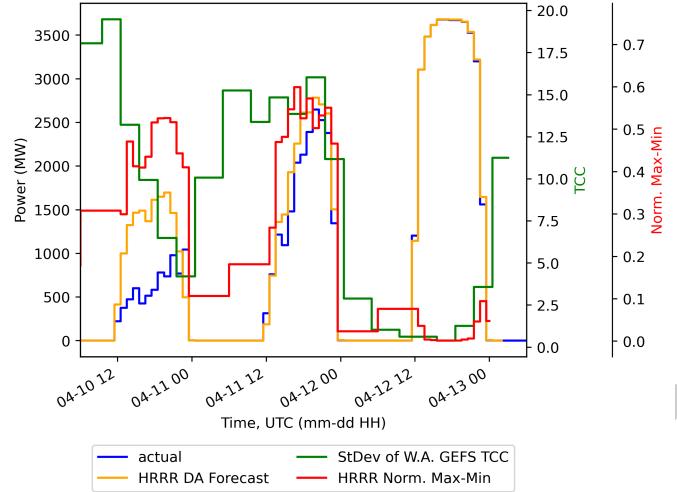


Fig. 7. Uncertainty indicators, standard deviation of weighted average TCC and normalized range of spatial variation from HRRR, along with forecasts and actuals for April 10-12, 2021.

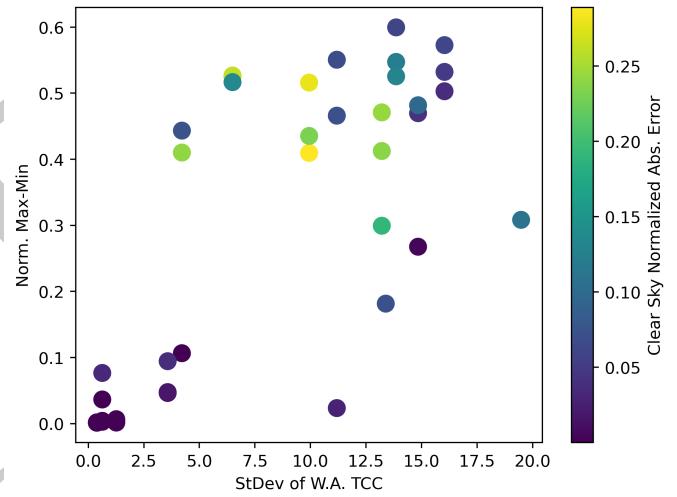


Fig. 8. Scatter plot of the uncertainty indicators from 7 along with absolute error normalized to clear sky, covering April 10-12, 2021.

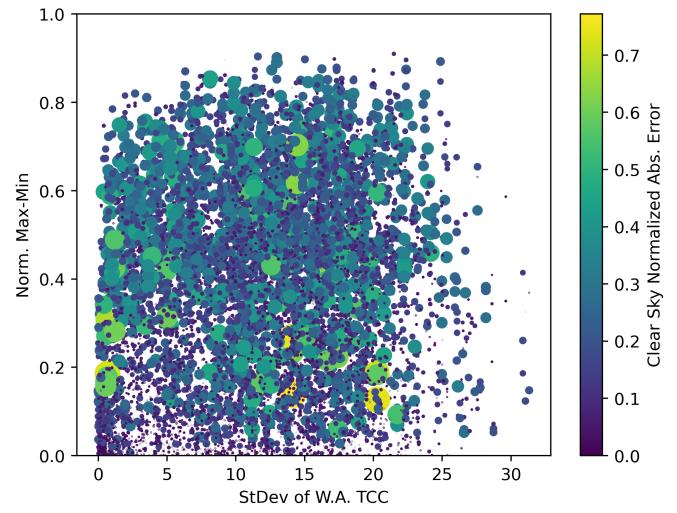


Fig. 9. Similar to 8, but covering all of 2021 and 2022. To assist in visualizing data, marker area has been scaled with the square of error.

TABLE I
COVERAGE (PICP) AND WIDTH (PINAW) FOR ALL TIME INTERVALS AND FOR
INTERVALS WITH SOLAR ZENITH LESS THAN 75 DEGREES.

Target PI	All Intervals		zenith < 75	
	PICP	PINAW	PICP	PINAW
0.995	0.991	0.59	0.991	0.68
0.99	0.989	0.54	0.988	0.63
0.98	0.981	0.49	0.979	0.57
0.95	0.954	0.41	0.949	0.47
0.5	0.510	0.13	0.485	0.15

Target prediction intervals were set, semi-arbitrarily, at 99% (i.e., 0.005 to 0.995 quantiles), 97%, 95%, 90%, and 50%, along with a central forecast (0.5 quantile). The models tended to be under-dispersed, so prediction intervals used in training were manually increased to get coverage that was close to the target prediction intervals.

The random forest model was found to perform better than XGBoost, with the exception of some overcast days, so I hybrid output was created by selecting the narrowest prediction interval from the two models for each forecast time step.

E. Final processing

Resulting outputs were denormalized with clear sky values to return them to units of power.

IV. REFERENCE PERSISTENCE ENSEMBLE FORECAST

As a reference for comparison, a simple probabilistic persistence forecast was produced. It was based on the probabilistic persistence forecast in [4], with a method often referred to as a Persistence Ensemble, or PeEn [15]–[17]. The observed clear sky index of power was used as an input, and then outputs were de-normalized back to power.

Because we used a relatively high clear sky model, there are enough hours with clear sky index of zero to skew the bottom of the wider prediction intervals to zero. To avoid this, hours where observed clear sky index was zero were excluded. This resulted in less than 3% of daytime hours being filtered out. An alternative approach could have been to apply a Persistence Ensemble with a time of day component. This would likely require a month of year component as well, or similar adjustment to account for variations in the relationship between solar position and time of day throughout the year.

The persistence ensemble was calculated based on observations from 2021 and 2022, so that it could be applied with 2023 data, similar to our machine learning model.

V. RESULTS

Prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW), calculated similarly to [18], are shown in Table I, for all time intervals and for time intervals where solar zenith is less than 75 degrees (i.e., close to sunrise and sunset). A three-day sample time series plot from April 10-12, 2021, is shown in Fig. 10, illustrating high and low uncertainty days.

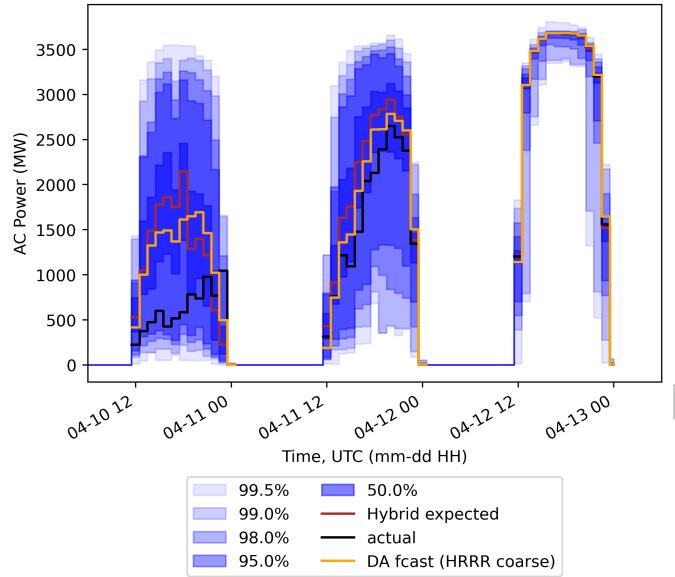


Fig. 10. Day-ahead probabilistic forecast outputs for a sample time range in 2021. Prediction intervals covering approximately 50% to 99% are shown in the shaded areas, along with the original deterministic day-ahead (DA) forecast, the new central machine learning (ML) forecast, and actual modeled generation.

TABLE II
COVERAGE (PICP) AND WIDTH (PINAW) FOR THE ML FORECAST AND THE
PERSISTENCE ENSEMBLE (PeEn), BOTH EXCLUDING ZERO CLEAR SKY INDEX INTERVALS.

Target PI	ML Forecast		PeEn	
	PICP	PINAW	PICP	PINAW
0.995	0.991	0.59	0.997	0.81
0.99	0.969	0.54	0.991	0.80
0.98	0.951	0.46	0.979	0.79
0.95	0.904	0.37	0.943	0.77
0.5	0.441	0.12	0.494	0.45

A. Comparison with Reference Persistence Ensemble

Because the persistence ensemble excluded daytime intervals with observed clear sky index of zero, PICP and PINAW were recalculated for the ML forecast after first excluding these same hours.

The coverage fractions (PICP) went down slightly after filtering out hours with zero power. This indicates that the ML forecast model's performance is not perfectly uniform, consistent with results from intervals with solar zenith less than 75 degrees, shown in Table I. To compensate for PICPs not matching target PIs, we can manually compare resulting PICPs from the persistence forecast and ML forecast. For example:

- The ML target PI of 0.98 has an actual PICP of 0.95 and a PINAW of 0.46. Compare that with the persistence target PI of 0.95 that has an actual PICP of 0.94 and a PINAW of 0.77. The ML forecast has a roughly 40% narrower PINAW.
- Looking at the target PI 0.995 in the ML forecast, it has PICP of 0.991 and PINAW of 0.59. That PICP matches

that of the target 0.99 PI for the persistence forecast, which has a PINAW value of 0.80, so the ML forecast has a 25% narrower PINAW.

- And the lowest PI: the target PI 0.5 in the ML forecast has a slightly lower PICP (0.44 vs 0.49 in the persistence forecast), but the PINAW is 0.12 vs 0.45 in the persistence forecast, an over 70% narrower PINAW.

VI. DISCUSSION

This work has presented a method for producing a probabilistic forecast for a regional fleet of PV sites from an existing deterministic forecast. For demonstration purposes, a relatively simple deterministic forecast was developed and actual output of the fleet was modeled. These and other factors present several potential areas for, as well as some considerations for operational deployment of this type of forecast system

A. Areas for Future Work

A list of potential areas for future work and improvement includes:

1) *Hyperparameter tuning*: The hyperparameters, or settings, used for the Random Forest and XGBoost machine learning models have only been manually tuned. More robust optimization, or tuning, could result in better performance.

2) *Extending to intraday and multi-day-ahead forecasts*: This work has focused on day-ahead forecasts (provided by approximately 6 AM local time covering the following day), but it is expected that the methods could be applied to intraday and multi-day-ahead forecasts. The ML models trained on day-ahead could be re-used for other forecast horizons, but it is possible that they would overestimate uncertainty for shorter horizons (intraday) and underestimate uncertainty for longer horizons (two days-ahead and longer), so retraining different models for different horizons may be needed.

3) *Exploring extrapolation*: Random Forest and XGBoost models cannot extrapolate beyond conditions they have been trained on. The work presented here used two years of training data, which may be enough to cover a broad enough range of conditions for this not to be an issue, but future work could test that. Additional models that can extrapolate could be considered or developed.

Additionally, a process could be implemented to check to see if input conditions fall outside the range of conditions previously trained on, allowing users to be alerted.

4) *Possible feature improvements*: Weights for the TCC feature are currently based on AC nameplate for each plant. It is possible that weighting clear sky modeled power in each interval could improve the weighting. For example, tracking plants have higher output relative to nameplate early in the morning than fixed-tilt plants, and this could be included in weighting.

The regions over which the HRRR is downsampled could be optimized in terms of window size (10×10 was chosen arbitrarily) and centering (the regions are based on a simple grid downscaling and are not centered on each plant).

Additional temporal uncertainty features could be considered, for example, by looking at the standard deviation of TCC or variation in the HRRR over time. The features currently used put more emphasis on where clouds appear rather than when they will appear, and accounting for both uncertainties could improve performance.

Other features, such as surface pressure and relative humidity, could be added. These two examples were used in an Analog Ensemble in [19] to produce a probabilistic forecast from a deterministic one. And for the existing NWP-based features, additional statistics could be considered. For example, the maximum and minimum range of weighted average TCC instead of standard deviation.

5) *Including additional weather models*: For example, recent updates to ECMWF [20] include free access to a version of deterministic and ensemble irradiance forecasts. The updated forecast data are only available starting in the first quarter of 2024, so forecasters using this data may have to wait until 2025 or later to have enough data to use for training and evaluation, although datasets such as [21] and [19] may help with data needs.

6) *Probabilistic net load forecasts*: Solar forecasts are only one component of what grid operators need to support decision making – they also need load (demand) forecasts, and ideal probabilistic forecasts would include the net of load and solar (and wind, for systems with wind). As covered in [22] and [23], load and renewable generation in a region are influenced by the same weather, and therefore are not independent. That means that their uncertainties need to be combined using information about their correlation, e.g., using copulas. This approach, as used in [22] and [23]. Additionally, ensemble forecasts, such as GEFS, could be used, but as ensembles are typically under-dispersed, i.e., lacking spread compared to observed outcomes [21], additional processing would likely be required.

B. Operational Deployment Considerations

To deploy this forecasting system in an operational environment, the primary considerations would involve model training with realistic deterministic forecasts and observed power. Additionally, accurate plant availability could be critical, assuming that forecast error due to plant outages is not intended to be included in the probabilistic forecast. As the number of solar plants in most areas is increasing significantly over time, the changes in the makeup of a fleet over time may also need to be accounted for. Careful data processing and quality control could be needed for all of these areas.

One option for getting representative deterministic forecasts and observations for model training could be to work with forecast providers to obtain re-forecasts, or historical forecasts that would have been produced, for current fleets. Actual power data could come from a combination of real observations and satellite-based irradiance, which is often available from commercial forecast vendors.

VII. AVAILABLE CODE

All code used to produce the results in this work is available in a repository on GitHub at <https://github.com/williamhobbs/solar-fleet-forecast-probability-tool> with an open-source BSD 3-Clause License. Everything is coded in Python using Jupyter notebooks, split across 12 notebooks that walk through topics such as creating the deterministic forecasts and modeled actuals, introducing the novel features used to help predict uncertainty (spatial variation in irradiance from HRRR and ensemble spread in total cloud cover from GEFS), training the ML models, and running and evaluating the resulting forecast. The repository also includes an example notebook with details on how someone might implement a real-time version of the models presented here.

The repository does not include full data files from HRRR and GEFS that would be needed to make forecasts for different plants. That set of files is too large to distribute, about 27 GB in total, but the repository does include code to re-download and process those files. Because those files have been subsetted to only cover a portion of the Southeast US, users would likely need to download their own files anyway if they wanted to model different plants.

For users that want to explore working with the group of plants used here, timeseries data files are included in the repository so that the process of downloading HRRR and GEFS files can be skipped.

ACKNOWLEDGMENT

Thanks to Will Holmgren at DNV for troubleshooting and other assistance with setting up a version of the Solar Forecast Arbiter reference forecasts, which inspired portions of this work. Thanks to David Larson and Dan Kirk-Davidoff at EPRI for inspiration and feedback related to tools and datasets to use in this work. A portion of this work used code generously provided by Brian Blaylock's Herbie python package (<https://doi.org/10.5281/zenodo.4567540>) – thanks to Brian for publishing and maintaining that package, and for responding to requests that helped make this work possible.

REFERENCES

- [1] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, p. 884, 2018. [Online]. Available: <https://doi.org/10.21105/joss.00884>
- [2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 785–794. [Online]. Available: <http://doi.acm.org/10.1145/2939672.2939785>
- [3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: <https://www.jmlr.org/papers/v12/pedregosa11a.html>
- [4] C. W. Hansen, W. F. Holmgren, A. Tuohy, J. Sharp, A. T. Lorenzo, L. J. Boeman, and A. Golnas, "The solar forecast arbiter: An open source evaluation framework for solar forecasting," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 2452–2457. [Online]. Available: <https://doi.org/10.1109/PVSC40753.2019.8980713>
- [5] D. C. Dowell, C. R. Alexander, E. P. James, S. S. Weygandt, S. G. Benjamin, G. S. Manikin, B. T. Blake, J. M. Brown, J. B. Olson, M. Hu *et al.*, "The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. part i: Motivation and system description," *Weather and Forecasting*, vol. 37, no. 8, pp. 1371–1395, 2022. [Online]. Available: <https://doi.org/10.1175/WAF-D-21-0151.1>
- [6] X. Zhou, Y. Zhu, D. Hou, B. Fu, W. Li, H. Guan, E. Sinsky, W. Kolczynski, X. Xue, Y. Luo *et al.*, "The development of the ncep global ensemble forecast system version 12," *Weather and Forecasting*, vol. 37, no. 6, pp. 1069–1084, 2022. [Online]. Available: <https://doi.org/10.1175/WAF-D-21-0112.1>
- [7] B. K. Blaylock, "Herbie: Retrieve numerical weather prediction model data," 2023. [Online]. Available: <https://doi.org/10.5281/zenodo.4567540>
- [8] "Form eia-860 detailed data with previous form data (eia-860a/860b)," 2023. [Online]. Available: <https://www.eia.gov/electricity/data/eia860/>
- [9] "National solar radiation database." [Online]. Available: <https://nsrdb.nrel.gov/>
- [10] G. Buster, M. Bannister, A. Habte, D. Hettinger, G. Maclaurin, M. Rossol, M. Sengupta, and Y. Xie, "Physics-guided machine learning for improved accuracy of the national solar radiation database," *Solar Energy*, vol. 232, pp. 483–492, 2022. [Online]. Available: <https://doi.org/10.1016/j.solener.2022.01.004>
- [11] D. Yang, "Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB)," *Journal of Renewable and Sustainable Energy*, vol. 13, no. 1, p. 016101, 01 2021. [Online]. Available: <https://doi.org/10.1063/5.0030992>
- [12] A. R. Jensen, W. B. Hobbs, K. S. Anderson, and W. F. Holmgren, "Missed clouds in the nsrdb at low sun elevation angles," to appear in 2024 IEEE 52nd Photovoltaic Specialists Conference (PVSC).
- [13] A. P. Dobos, "PVWatts version 5 manual," National Renewable Energy Laboratory, Golden, CO, Tech. Rep. NREL/TP-6A20-62641, 2014. [Online]. Available: <https://doi.org/10.2172/1158421>
- [14] R. A. Johnson, "quantile-forest: A python package for quantile regression forests," *Journal of Open Source Software*, vol. 9, no. 93, p. 5976, 2024. [Online]. Available: <https://doi.org/10.21105/joss.05976>
- [15] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied energy*, vol. 157, pp. 95–110, 2015. [Online]. Available: <https://doi.org/10.1016/j.apenergy.2015.08.011>
- [16] D. Yang, "A universal benchmarking method for probabilistic solar irradiance forecasting," *Solar Energy*, vol. 184, pp. 410–416, 2019. [Online]. Available: <https://doi.org/10.1016/j.solener.2019.04.018>
- [17] K. Doubleday, V. V. S. Hernandez, and B.-M. Hodge, "Benchmark probabilistic solar forecasts: Characteristics and recommendations," *Solar Energy*, vol. 206, pp. 52–67, 2020. [Online]. Available: <https://doi.org/10.1016/j.solener.2020.05.051>
- [18] J. Boland and A. Grantham, "Nonparametric conditional heteroscedastic hourly probabilistic forecasting of solar radiation," *J*, vol. 1, no. 1, pp. 174–191, 2018. [Online]. Available: <https://doi.org/10.3390/j1010016>
- [19] D. Yang, W. Wang, and T. Hong, "A historical weather forecast dataset from the european centre for medium-range weather forecasts (ecmwf) for energy forecasting," *Solar Energy*, vol. 232, pp. 263–274, 2022. [Online]. Available: <https://doi.org/10.1016/j.solener.2021.12.011>
- [20] E. C. for Medium-Range Weather Forecasts, "Ecmwf releases a much larger open dataset," 2024, accessed on: 2024-06-03. [Online]. Available: <https://www.ecmwf.int/en/about/media-centre/news/2024/ecmwf-releases-much-larger-open-dataset>
- [21] W. Wang, D. Yang, T. Hong, and J. Kleissl, "An archived dataset from the ecmwf ensemble prediction system for probabilistic solar power forecasting," *Solar Energy*, vol. 248, pp. 64–75, 2022. [Online]. Available: <https://doi.org/10.1016/j.solener.2022.10.062>
- [22] B. Li, J. Zhang, and B. F. Hobbs, "A copula enhanced convolution for uncertainty aggregation," in *2020 IEEE power & energy society innovative smart grid technologies conference (ISGT)*. IEEE, 2020, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/ISGT45199.2020.9087644>
- [23] M. Beichter, K. Phipps, M. M. Frysztacki, R. Mikut, V. Hagenmeyer, and N. Ludwig, "Net load forecasting using different aggregation levels," *Energy Informatics*, vol. 5, no. Suppl 1, p. 19, 2022. [Online]. Available: <https://doi.org/10.1186/s42162-022-00213-8>