Validation of Wind and PV Power Generation Using Historical and Forecast Weather Data

Seung Jun Lee, Farnaz Safdarian, Thomas J. Overbye
Department of Electrical and Computer Engineering
Texas A&M University
College Station, TX
{sl47745, fsafdarian, overbye}@tamu.edu

***Abstract*—This paper presents a comparative analysis of renewable energy power output using forecast weather with different margins and historical weather data as benchmarks for selected days. The analysis evaluates the accuracy and performance trends of solar and wind forecasts against historical data, focusing on uncertainties at various forecast horizons. The benchmark hourly power generation data is used to compute relative errors, providing insights into temporal variations. Visualizations highlight the alignment between forecast trends and historical patterns, qualitatively assessing weather impacts on generation. The results for wind generation indicate that a seven-day forecast can achieve an accuracy of 80 percent, while a five-day forecast can reach an accuracy of approximately 90 percent. However, forecasts beyond ten days are only about 50 percent or less accurate compared to actual data. In contrast, the results for solar generation show comparable levels of accuracy throughout the entire forecast date but with higher average error values. This information is useful for grid operation, planning, electricity market, reliability and resilience studies and energy management.**

***Index Terms*—Renewable energy, historical and forecast Weather data, prediction, synthetic power grids**

I. Introduction

The rising global demand for energy-intensive technologies, such as data centers, cryptocurrency mining, and EV charging stations, is driving a continuous increase in energy needs. To meet this demand, renewable energy sources have been increasingly utilized due to their benefits, including economic and environmental advantages. However, renewables are inherently dependent on weather conditions, which pose several challenges for their use as a reliable and consistent energy supply. Variability in weather patterns, seasonal changes, and geographical differences impact the effectiveness of renewable energy, creating obstacles to adapting these sources for high-energy-demand systems. Therefore, understanding the direct impact of weather and forecasts on power systems, as well as long-term renewable energy forecasting, has become crucial for efficient integration.

Accurately simulating weather conditions is crucial for improving forecasts of demand and power generation. This topic has garnered significant attention, particularly for solar energy production [1], [2] and wind energy [3]–[5]. Initial studies were made in 1991 [6], emphasizing the need to include weather variables in power system assessments, particularly for evaluating contingencies and outages during extreme weather. Recent studies, such as [7], underscore the role of weather in assessing the resilience of power system infrastructure.

The growing interest in integrating weather data into power system planning and operations led to tools like "Renewables.ninja" [8], [9] that provide estimates of global renewable energy output based on weather data for 2019 as a sample year. Similarly, the PLUSWIND database [10] offers hourly wind speeds and estimated generation data for nearly all U.S. wind plants from 2018 to 2021, enabling detailed analysis of wind generation's geographic and temporal variations. later, a more general study [11], suggested a strategy for directly including weather data to optimal power flow problems based on renewable and thermal weather-dependent power outputs that can use any available historical or forecast weather resources.

The long-term forecast of renewable power generation with the idea of possible uncertainties can significantly improve planning studies, grid stability analysis, and meeting the demand for renewable energy. Having weather forecast data and their impact on the grid allows operators to proactively prepare for potentially critical weather conditions, helping to prevent outages.

Many studies have focused on forecasting renewable energy generation output. Reference [12] utilizes Numerical Weather Prediction (NWP) data to forecast wind power generation at a wind farm located in northwest China. They employ a training model based on actual data to predict wind power output for the following seven days. The results of the study highlight the influence of wind turbine type and wind direction on the accuracy of the forecasts.

The authors in [13] developed multiple machine learning models for long-term forecasting, which were trained using historical load and weather data. This study analyzes the contributions of specific weather and temporal features to enhance the model's accuracy.

The study referenced in [14] utilized wind turbine characteristic curves and power system simulation tools, such as

TABLE I: Historical and forecast solar power generation simulation results

Date	Time	California (Historical)	California(Forecast)	Florida (Historical)	Florida (Forecast)	Texas (Historical)	Texas (Forecast)
7/15/2024	10:00:00 AM	2688.1	6364.4	1323.0	791.5	2876.6	3445.6
7/15/2024	11:00:00 AM	4247.9	6179.1	2075.2	1260.0	3968.3	3906.0
7/15/2024	12:00:00 PM	6612.7	7396.3	4003.4	2734.4	5489.8	4922.7
7/15/2024	1:00:00 PM	9348.5	9030.1	4616.5	3294.3	7493.8	6521.2
7/15/2024	2:00:00 PM	12771.0	11518.3	3086.1	1846.3	7030.8	5245.8
7/15/2024	3:00:00 PM	11057.6	10138.0	2046.9	1138.1	5173.3	3291.2
7/15/2024	4:00:00 PM	9787.8	8539.7	1485.0	882.2	3579.7	2120.0
7/15/2024	5:00:00 PM	6347.7	5761.9	1045.7	720.6	2464.8	1595.8
7/15/2024	6:00:00 PM	4189.4	3903.5	750.5	538.3	2291.8	1756.4

PSS_E and DigiSilent PowerFactory. It applied short-term Numerical Weather Prediction (NWP) forecasts to generate wind generation forecasts for short intervals of five minutes, achieving an error rate as low as 15%.

The study from [15] utilizes minute-by-minute solar generation data obtained from a solar station, along with weather data that correlates with solar generation, to develop a Recurrent Neural Network model. This model is then used to process short-term forecasts of solar generation.

However, a common limitation in the literature is the lack of metrics to validated weather forecast with different margins from the study day. Additionally, some studies do not have high resolution measurements and geographical coordinates, which limits their ability to capture location-specific differences in weather conditions. Furthermore, some models lack detailed information about each generator, such as the power curves for wind or photovoltaic (PV) systems. Power curves define the relationship between environmental conditions, such as wind speed or solar irradiance, and electrical energy generation. Without properly incorporating these curves, forecast weather models may fail to produce accurate predictions of energy generation.

This paper aims to highlight the accuracy of power output forecasts with different distance ahead from the study dates for both solar and wind energy by comparing them with historical benchmark data and find the expected uncertainties for each forecast. The forecast datasets is from National Oceanic and Atmospheric Administration (NOAA) that forecasts the weather up to the next 16 days and we propose a metric to find how the uncertainties of forecasts change as they get closer to the study date. Generator data, their geographical coordinates, types, classes and power curves of renewables are extracted from [16]

The paper introduces the data and models used for the simulation (Section II), describes the methodology for comparing forecast and historical data (Section III), presents simulation results to describe the accuracy of each forecast data (Section IV), and concludes with an introduction to future research directions using this data (Section V).

II. Data Preparation

A. Weather Data

NOAA publishes weather forecast datasets known as the Operational Model Archive and Distribution System (NOMADS) in various geographic resolutions at [17]. The dataset used in this paper has a resolution of 0.25 degrees, which is approximately available at each 30 km. The data is updated every 6 hours and includes parameters such as dew point, temperature, and wind speed, radiation, and cloud coverage at various levels, including 10 and 100 meters above ground. The data includes weather forecasts for up to 16 days, with hourly intervals for the first five days and three-hour intervals thereafter.

The historical data used in this analysis comes from European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Version 5 (ERA5) [18]. ERA5 includes weather measurements found in weather forecast data and provides data coverage for the entire world. The time range of the ERA5 dataset spans hourly data from 1940 to the present. According to reference [19], ERA5 is particularly well-suited for analyzing historical wind and solar generation performance. The ERA5 dataset includes wind data at 10 meters and 100 meters, as well as irradiance data necessary for calculating solar and wind power.

The NOAA data is in GRIB (Gribbed Binary) format, while ERA5 is available in NetCDF (Network Common Data Form) format. In order to use the weather data for the simulation, files need to be converted to a format that is more efficient and suitable for power system study. To achieve this the PoWer Weather (PWW) file format is introduced at [20], which reduces the file size to about one third of the original files by storing the data more efficiently. Additionally, [21] provides an in-depth explanation of processing the weather data into the PWW file format.

B. Grid data

Please note that, because real power grid information is classified in the US as Critical Energy Infrastructure Information (CEII), this study utilizes the copper plate model derived from EIA-860 data to generate renewable energy outputs, which are then mapped to synthetic generator locations [22]. Power flow weather (PFW) models are employed to establish the relationship between input weather measurements and output power data [23]. These PFW models are developed using the types and models of renewable resources detailed in EIA-860 data. Furthermore, the data has been validated in [19] to confirm that the PFW models accurately represent the weather-induced effects on the power grid.

III. Methodology

Reference [11] outlines a methodology for directly incorporating weather measurements into optimal power flow models. It utilizes power curves, renewable resource types, and the geographical coordinates of generators, mapped to the nearest weather data points.

For the study, a representative day from each season —spring, summer, and winter—was selected along with forecast data saved from up to 16 days prior to these study dates. This approach captures seasonal weather variability across the study region. To compare forecast data with what happened in reality, historical weather data from the same days are used as a benchmark. As mentioned, for forecast data NOAA data and as benchmarks ERA5 data are selected. After the weather data are loaded in time step simulation and the output power of renewables are calculated from both forecast data and the benchmarks, hourly generation data is used for each study day to find the relative error between the forecast and actual generation, providing insight into the model's performance throughout the day. Calculating the maximum and minimum relative errors highlights deviations from the benchmark, enabling an assessment of forecast reliability and uncertainty. Generation data for each forecast date and benchmark is plotted for visualization. These visualizations provide insights into weather behavior and its influence on power generation trends.

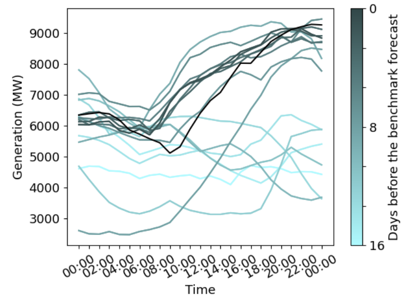
The error calculation was performed using the mean absolute error equation, as referenced in [24], for each state's historical and forecast generation data. In this context, f(x) represents the historical generation, g(x) signifies the forecast generation, and n indicates the number of data points. To compute the average error across all states for the individual forecast data, the sum of absolute error is divided by the number of data points (equation 1). Additionally, the maximum and minimum errors were determined through this process to assess the improvements in accuracy as the forecast date approaches the benchmark.

$$\frac{1}{n} * \sum_{i=1}^{n} \frac{|f(x_i) - g(x_i)|}{|f(x_i)|} * 100 \quad (1)$$

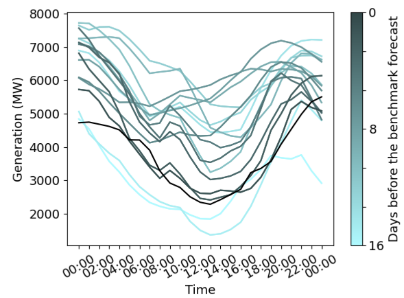
IV. Case Study and Simulation Results

For this study, EIA-860 data at the end of year 2023 is used for simulations with grid model available at [21]. Selected study dates are April 25th, 2024, July 15th, 2024 and October 15th, 2024. The power generation outputs for both solar and wind are calculated using time step simulations in PowerWorld Simulator Version 23 is used to load weather measurements of forecast and historical benchmark data in PWW format, calculate power outputs and find the differences from the selected historical benchmark data by comparing the results.

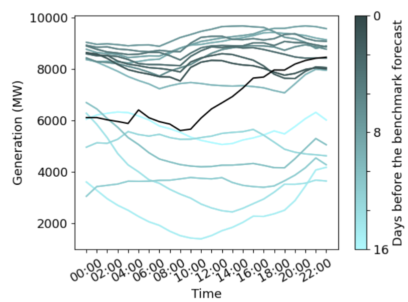
In Figures 1 and 2, the data points for the forecast from the start date and end date are illustrated. The color of the plot indicates the distance of the data from the benchmark, with lighter shades representing a greater deviation and darker shades indicating closer proximity. The y-axis represents wind power generation, while the x-axis denotes the time point in hours.



(a) Wind forecast of April 25th, 2024



(b) Wind forecast of July 15th, 2024



(c) Wind forecast of October 15th, 2024

Fig. 1: Wind forecast result

A. Analyzing Wind Power Results

As it can be observed from comparing Figures 1 and 2, forecasting wind power output is more challenging compared to predicting power output from solar energy. This can be related to the more unpredictable patterns for wind speed. This increases the uncertainty in the disparity between benchmark wind power and the forecast.

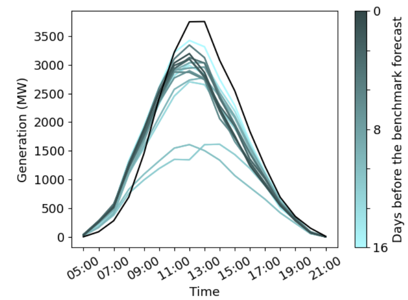
The data presented in Table 2 illustrates the relative error between forecast and benchmark renewable power generation, revealing significant fluctuations in error values using different forecast margins in successive data days, particularly in the 10 days leading up to the benchmark date. These sharp variations highlight significant inconsistencies in the forecast model's accuracy during this period.

However, a noticeable pattern emerges after the 10-day mark before the benchmark date, where the errors exhibit a linear decline, approaching the benchmark data. This consistent

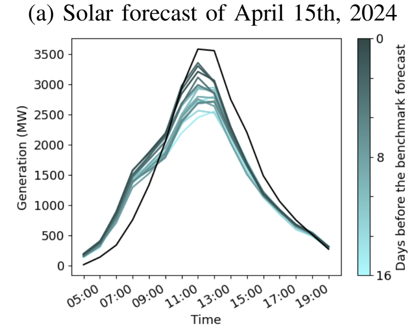
TABLE II: Error values for wind forecast of 2024

Date	Max Error (%)	Min Error (%)	Relative Error (%)	Date	Max Error (%)	Min Error (%)	Relative Error (%)	Date	Max Error (%)	Min Error (%)	Relative Error (%)
04 / 01	53	16	36	07 / 01	47	1	21	10 / 03	32	2	17
04 / 02	51	1	25	07 / 02	43	1	22	10 / 04	78	41	65
04 / 03	37	2	23	07 / 03	100	24	58	10 / 05	65	3	46
04 / 04	61	2	22	07 / 04	106	23	62	10 / 06	46	6	23
04 / 05	61	27	46	07 / 05	140	21	80	10 / 07	57	34	46
04 / 06	49	4	25	07 / 06	75	7	50	10 / 08	48	1	30
04 / 07	62	2	28	07 / 07	132	21	68	10 / 09	38	2	19
04 / 08	52	1	18	07 / 08	134	4	68	10 / 10	58	8	33
04 / 09	62	9	36	07 / 09	139	7	67	10 / 11	58	14	36
04 / 10	17	1	7	07 / 10	157	19	74	10 / 12	65	5	39
04 / 11	45	1	15	07 / 11	91	4	47	10 / 13	45	4	27
04 / 12	35	2	11	07 / 12	83	2	49	10 / 14	60	8	35
04 / 13	29	1	9	07 / 13	68	1	43	10 / 15	53	5	32
04 / 14	29	1	7	07 / 14	53	1	30	10 / 16	50	1	28
04 / 15	35	1	9	07 / 15	45	1	15	10 / 17	47	3	26
04 / 16	35	1	9	07 / 16	22	1	10	10 / 18	43	2	23

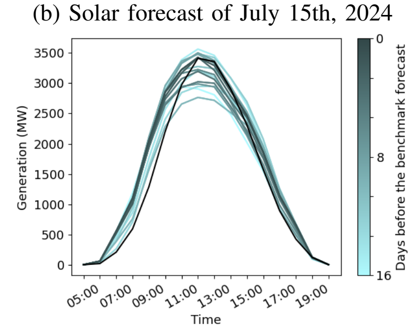
downward trend observed across all three months suggests an improvement in the alignment of forecasts with the benchmark data over time. The transition from a more unpredictable behavior to a more consistent decline indicates that the model has become more predictable using forecasts from 10 days prior to the benchmark date. In Figure 1, as the forecast day gets closer to the studied benchmark day, the plots become progressively darker, in a way that color intensity correlates negatively with error margins and distance from the study dates revealing a trend where the forecast power generation increasingly aligns better with the benchmark plot (black). More details on relative errors on each days of forecast before the study day is presented in Table 2. Despite the initial larger relative errors noted in Table 1, the overall behavior of power generation can be estimated from the plots using relative errors as uncertainties, as the forecast time gets closer to the benchmark day. The results suggest that forecasts made before the 10-day mark can provide more valuable insights into generation performance trends. Although the earlier forecasts are less precise as expected, but are useful for capturing the general behaviors and trends of power generation, allowing for preliminary planning and decision-making based on expected patterns.



(a) Solar forecast of April 15th, 2024



(b) Solar forecast of July 15th, 2024



(c) Solar forecast of October 15th, 2024

Fig. 2: Solar forecast result

B. Analyzing Solar Power Results

Solar forecasting demonstrates stable patterns, unlike the unpredictable fluctuations in wind forecasting. The monthly plots reveal trends closely aligned with the benchmark data. Peak solar generation aligns more closely with historical data as the forecast date approaches. However, most of the forecast plots share a similar shape throughout all the results. This behavior is illustrated in Figure 2, and Table 3 presents this phenomenon numerically as well.

Table 3 shows minimal fluctuations in error values, reflecting consistent forecast accuracy. The values from both the maximum and relative error columns fall within a similar range, which demonstrates the consistent behavior of all forecast data. In contrast, the minimum error shows improvement as the days progress, as the peak generation is becoming closer to the actual peak generation data. Solar forecasts are less affected by unpredictability compared to wind forecasts, as indicated by the smoother error trends in Table 3.

TABLE III: Error values for solar forecast of 2024

Date	Max Error (%)	Min Error (%)	Relative Error (%)	Date	Max Error (%)	Min Error (%)	Relative Error (%)	Date	Max Error (%)	Min Error (%)	Relative Error (%)
04 / 01	97	1	39	07 / 01	83	9	72	10 / 03	59	2	52
04 / 02	76	5	47	07 / 02	82	6	84	10 / 04	75	3	49
04 / 03	81	4	37	07 / 03	93	3	74	10 / 05	59	1	44
04 / 04	62	7	39	07 / 04	86	4	94	10 / 06	59	5	27
04 / 05	98	10	47	07 / 05	90	6	75	10 / 07	90	1	42
04 / 06	90	20	54	07 / 06	91	6	77	10 / 08	89	1	52
04 / 07	86	10	34	07 / 07	92	1	86	10 / 09	86	1	39
04 / 08	68	3	38	07 / 08	81	1	82	10 / 10	87	1	54
04 / 09	98	1	43	07 / 09	71	2	75	10 / 11	80	1	51
04 / 10	95	4	48	07 / 10	82	2	83	10 / 12	77	1	48
04 / 11	88	2	49	07 / 11	85	4	75	10 / 13	86	3	49
04 / 12	98	4	47	07 / 12	99	4	96	10 / 14	74	2	50
04 / 13	74	4	46	07 / 13	95	1	88	10 / 15	76	2	49
04 / 14	88	1	45	07 / 14	98	2	85	10 / 16	80	1	52
04 / 15	82	5	48	07 / 15	40	2	90	10 / 17	70	1	49
04 / 16	89	5	48	07 / 16	41	3	94	10 / 18	86	1	53

V. Conclusions and Future Work

This paper presented a comparison between forecast power output and historical data for both photovoltaic (PV) solar energy and wind energy based on forecast weather measurements with different daily distances from selected study days and historical weather measurements as a benchmark of what happened in reality. Weather measurements were used to calculate renewable power outputs, which were then compared.

Each dataset was compared on an hourly basis to compute the differences between forecast and benchmark weather measurements and related power outputs. The results of the simulation were visualized by plotting each data point of generation, illustrating the trends between historical and forecast data. Consequently, minimum, maximum, and relative errors were calculated. The results show larger fluctuations in forecasts farther from the study dates, with error rates declining after the 10-day mark, indicating an improved accuracy as the forecast time approached closer to the benchmark date. However, the forecasts before this period still reflected changes in renewable power generation, showcasing both upward and downward trends compared to the historical data.

The comparisons help find uncertainties in the weather forecast with different distances from study days and reality and can be used for power system operation and planning studies. Future research will analyze a broader range of study dates to improve the understanding of forecasting uncertainties and enhance predictive accuracy.

VI. Acknowledgements

This work was partially supported through funding provided by Texas A&M Engineering Experiment Station (TEES) Smart Grid Center and partially as part of the "Electric Grid Resilience" Project (60NANB24D210) funded by the National Institute of Standards and Technology of US Department of Commerce.

References

[1] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.

[2] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Brussels, Belgium, Oct 2011, pp. 528–533.

[3] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium 2010*, 2010, pp. 1–8.

[4] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using gaussian processes and numerical weather prediction," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 656–665, 2014.

[5] S. Al-Yahyai, Y. Charabi, and A. Gastli, "Review of the use of numerical weather prediction (nwp) models for wind energy assessment," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 3192–3198, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032110001814

[6] R. Billinton and L. Wenyuan, "A novel method for incorporating weather effects in composite system adequacy evaluation," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1154–1160, 1991.

[7] M. Panteli and P. Mancarella, "Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1733–1742, 2017.

[8] S. Pfenninger, I. Staffell, and M. Jansen, "Renewables. ninja-a model for the global output of weather-dependent renewable energy sources," in *EMS Annual Meeting Abstracts*, vol. 15, 2018.

[9] "Renewables.ninja," https://www.renewables.ninja/, accessed: 2024-01-04.

[10] D. Millstein, S. Jeong, A. Ancell, and R. Wiser, "A database of hourly wind speed and modeled generation for us wind plants based on three meteorological models," *Scientific Data*, vol. 10, no. 1, p. 883, 2023.

[11] T. Overbye, F. Safdarian, W. Trinh, Z. Mao, J. Snodgrass, and J. Yeo, "An approach for the direct inclusion of weather information in the power flow," 2023.

[12] C. Xiang, W. Fu-Jun, L. Tian-Qi, C. Zhen-Huan, L. Xiao-Hu, G. Tie-Ying, F. Suilin, and F. Zheng, "Wind power prediction considering the layout of the wind turbines and wind direction," in *2012 Asia-Pacific Power and Energy Engineering Conference*. Shanghai, China: IEEE, Mar 2012.

[13] J. Yang, M. Tuo, J. Lu, and X. Li, "Analysis of weather and time features in machine learning-aided ERCOT load forecasting," in *2024 IEEE Texas Power and Energy Conference (TPEC)*. College Station, TX: IEEE, Feb 2024.

[14] Z. Dong, K. P. Wong, K. Meng, F. Luo, F. Yao, and J. Zhao, "Wind power impact on system operations and planning," in *IEEE PES general meeting*. IEEE, 2010, pp. 1–5.

[15] S. M. Awan, Z. A. Khan, and M. Aslam, "Solar generation forecasting by recurrent neural networks optimized by levenberg-marquardt algorithm," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*. Washington, DC: IEEE, Oct 2018.

[16] "Form eia-860 detailed data with previous form data (eia-860a/860b)." [Online]. Available: https://www.eia.gov/electricity/data/eia860/

[17] "Nomads information." [Online]. Available: https://nomads.ncep.noaa.gov/info.php?page=overview

[18] Ecmwf, "Era5 houly weather data," Dec 2024. [Online]. Avail-

able: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download

[19] F. Safdarian, J. Cook, S. J. Lee, and T. J. Overbye, "Calculation and validation of weather-informed renewable generation in the us based on era5 hourly weather measurements," in *Power and Energy Conference at Illinois (PECI)*. Champaign, IL: IEEE, Apr 2024.

[20] F. Safdarian, M. Stevens, J. Snodgrass, and T. J. Overbye, "Detailed hourly weather measurements for power system applications," in *2024 IEEE Texas Power and Energy Conference (TPEC)*. College Station, TX: IEEE, Feb 2024.

[21] "Tamu grid test cases." [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/

[22] J. Cook, F. Safdarian, J. Snodgrass, and T. J. Overbye, "Using power flow application capabilities to visualize and analyze us energy information administration generation data," in *Submitted to the 2024 IEEE Texas Power and Energy Conference (TPEC)*, College Station, TX, Feb 2024. [Online]. Available: https://overbye.engr.tamu.edu/publications

[23] F. Safdarian, J. Cook, K. Zhgun, T. J. Overbye, and J. Snodgrass, "Power flow modeling of the impacts of weather and other resiliency hazards with a focus on transmission planning," in *58th Hawaii International Conference on System Sciences*, Waikoloa, HI, January 2025.

[24] S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (mae) into systematic and unsystematic components," *PloS one*, vol. 18, no. 2, 2023.