Large-Scale Weather Correlations for a Possible Interconnection of North American Power Grids

Jordan Cook, Farnaz Safdarian, Jonathan Snodgrass, Thomas J. Overbye
Texas A&M University
Department of Electrical and Computer Engineering
College Station, TX
{jordancook, fsafdarian, snodgrass, overbye}@tamu.edu

***Abstract*—This paper introduces a strategy for analyzing the correlations among large-scale solar power, wind power, as well as identifying patterns and detailed renewable power generation from weather data. The aim is to investigate the potential advantages of connecting neighboring, yet currently disconnected grids by assessing the feasibility of transferring renewable power between them. The case study is on two of the largest grids in North America and the potential of joining them. The insight gained from this research is aimed at reducing the overall operation cost by taking advantage of renewable energy trading between the two large grids as well as enhancing grid resiliency and reliability in the pursuit of more sustainable energy in the future.**

***Index Terms*—Weather integration, renewable generation, power flow, correlation, sustainable energy**

I. INTRODUCTION

The landscape of power generation has experienced a remarkable surge of renewable generation in the most recent years. In 2022, the total amount of renewable generation surpassed both coal and nuclear generation [1]. The four main sources of renewable generation are wind, hydroelectric, solar, and biomass. Wind and solar energy depends on the weather in the area and will be the primary focus of this paper. Because of this dependence on weather, it has become necessary to directly include weather measurements in power flow studies. [2]

Due of the variable nature of weather, renewable energy is nondispatchable, meaning that generation cannot be changed by the demand of power in the area. Therefore, it is essential to study patterns of weather in the context of renewable generation to more efficiently utilize this resource.

One method of studying the weather involves correlating time-series data related to renewable generation. The research in [3] correlates the amount of wind generation in different areas of China, finding that the correlation is higher in areas that are closer spatially, while [4] found similar results in different geographical settings. Correlation can also be used to quantify variability in solar PV farms [5]. Reference [6] found negative correlation between wind generation and solar generation, indicating the two variables are inversely related.

In North America, there are four major electric grids: the Eastern Interconnect (EI), the Western Electricity Coordinating Council (WECC), the Electric Reliability Council of Texas (ERCOT), and the Quebec Interconnect. These grids are AC networks that operate separately and are only linked to each other through DC ties. From 1967-1975 the EI and WECC were joined but were ultimately disconnected due to overloads and breakups [7]. There is still consideration on a connection of these two grids. References [8] and [9] study the potential of building more transmission for the EI and WECC interconnection. Additionally, reference [10] explore using direct current (DC) in a joined grid. Reference [11] focuses on upgrading the existing DC tie lines and even building some long distance DC ties. On the other hand, there have been some recent studies in an AC interconnection. The work in [12] studied the characteristics of possible tie-points, power flow on an interconnect, stability, frequency response, and more. Reference [13] reviewed the various challenges that a synchronous grid may face, provided dynamic assessments and improved situational awareness of large systems. Furthermore, [14] presented some stability considerations.

The most recent studies prompt the question, "Is there a potential benefit in utilizing renewable generation with an interconnect between the EI and WECC?" Given that the output of solar and wind generation is contingent on regional weather patterns, it is crucial to investigate such weather variations. Several studies have run statistical analyses on weather, such as [15], which studies wind and solar output in Europe with spatial correlation, or [16] which runs statistical procedures on regional weather data in Italy. However, it is worth mentioning that there is a significant lack of research on weather patterns in the context of wind and solar generation in North America, specifically across the major grids.

This paper presents a strategy for analyzing correlations among large-scale solar and wind power to derive detailed insights into renewable power generation from weather data. The aim is to study the potential benefits of connecting neighboring, yet disconnected, grids by exploring the possibility of transferring renewable power between them. This approach is useful regardless of region or AC versus DC connection.

The paper will begin by detailing the process in which hourly weather data can be translated into renewable generation output in section II. Next, section III will explain the correlation method, offering insights into its application within the context of weather, particularly relating these variables over time and space. The methods presented in this paper will be demonstrated with a case study, involving the EI and WECC grids in North America in section IV and results in section V considering the possibility of joining the two largest grids in that continent. Finally, the paper will conclude with remarks in section VI.

II. WEATHER DATA TO GENERATION OUTPUT

To calculate the amount of power being generated, there are many components that need to be known, such as gen-

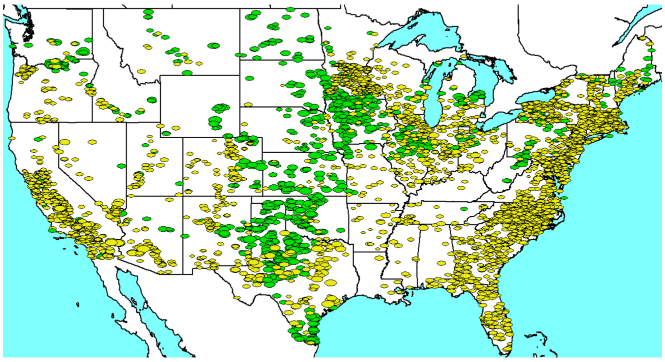

Fig. 1: Renewable Generators in the US

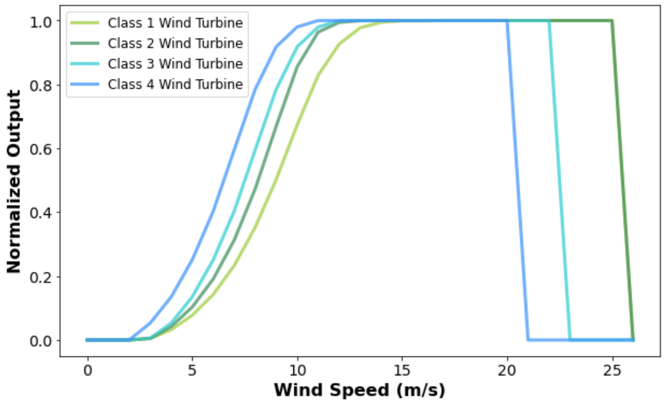

Fig. 2: Curves Used to Determine WTG Output [30]

eration capacity, bus-level load, geographic location of grid components, and the weather measurements at each location and each time. The generation data in this study is based on the end of 2022; all weather scenarios are studied on the same grid. Historical weather data, ranging from 1972 to 2022, is used to study scenarios of interest. It is important to mention that the purpose of using historical data is not to simulate past scenarios or recreate the grid at a moment that has previously occurred. Rather, it is meant to study future events and possibilities. If a weather event has already occurred, it is likely to happen again at some point in the future. This approach allows for a deeper understanding of potential challenges and aids in the goal of being prepared for all possible weather outcomes.

A. Generator Information

The first step in determining how much energy is being produced from renewable generators is knowing the capacity of all the generators in an area. In the United States, the Energy Information Administration provides form 860 (EIA-860) which provides extensive information on the country's electrical generators [17]. It is publicly accessible and contains no critical energy/electric infrastructure information (CEII). This data is divided into utilities, plants, units, and US states. Because of this, there is a natural mapping of this data into power flow structures.

By using the methods presented in [18], the EIA-860 data is used to create a Copper Plate model [19] to develop a case that aids in both visualization and analysis. The EIA-860 data also contains parameters of wind and solar generators, such as wind class for wind turbines or the type of tracking for solar PV plants.

Using a Geographic Data View (GDV) [20], Figure 1 shows the overall renewable generator capacity in the contiguous United States at the end of 2021. The green ovals represent a wind turbine generator and the gold ovals represent a solar plant. The size of each oval is proportional to the capacity of the generator being represented.

B. Weather Data

Instead of the conventional method of relying on aggregated outputs of renewable generators disclosed from utilities, which is implicit or indirect inclusion of weather in a model or simulation, the detailed direct integration of meteorological data is incorporated in this work. Based on the strategy shown in [2], weather data in operation and planning scenarios, such as Optimal Power Flow (OPF), could refine the accuracy of renewable generation forecasting as well as capture rapid climatic shifts, without significantly increasing computational complexity.

The applied methodology in this paper incorporates meteorological variables into hourly outputs of wind turbines and solar cells. This utilizes data from 1972 to 2022, sourced from weather stations across the U.S. [21] Historical weather data is collected at an hourly granularity from various sources worldwide [22]. There are other weather data sets that are useful, such as ERA5 [23], MERRA5 [24], or HRRR [25]. These could potentially be useful for future work and comparisons.

The International Civil Aviation Organization (ICAO) and the World Meteorological Organization (WMO) serve as primary providers of meteorological data, which encompasses thousands of weather stations across the globe. Real-time weather data from weather stations, identified via ICAO codes, is used from [21].

C. Renewable Generation Based on Weather Measurements

Using the geographic coordinates for each generator provided by the EIA-860 data, each generator is partnered with the nearest weather station. For missing data, interpolation of nearest station data is utilized. This study uses several energy input-output models to quantify the influence of weather on renewable generators.

1) Wind Generation: The first class of wind turbines, which incorporates local wind speeds and turbine power curves as per [26], [27], and [28], calculates wind turbine generator (WTG) output based on the wind measurements. Subsequent wind turbine classes two, three and four adapt this framework to accommodate different power curves based on [29]. Figure 2 shows these wind classes. Please note that the curves are linearized in a piece-wise manner.

From the curves shown in Fig. 2, a scalar value between 0 and 1 is found by using the wind speed at a time point and the class of the WTG. This scalar is than multiplied by the maximum capacity of the generator, which results in the generated power at that time point for that wind turbine.

2) Solar Generation: The solar model estimates PV generation using local solar radiation, cloud coverage and PV characteristics such as tilt angle, azimuth angle and power point tracking as specified in [26].

III. Correlation Method

For comparing two vectors, correlation can be used to quantify the relationship or connection between the two measurements. These relationships are represented with a scalar. Two main types of correlation coefficients can be calculated: the Pearson correlation coefficient or the Spearman's rank correlation coefficient. As explained in [31] and [32], Spearman's is best used in monotonic datasets (a function that does not stop increasing or a function that does not stop decreasing), while Pearson's is ideal for linear functions in a dataset.

In this study, since weather data from one region is being compared to that of another, Pearson's correlation coefficient is chosen for its linear characteristics. Pearson's correlation coefficient ranges from -1 to 1. If two datasets have a correlation of 1, that means that if the variables in the x-dataset increase, the variables in the y-dataset will also increase proportionally. Conversely, if two datasets have a correlation of -1, it means that as the variables in the x-dataset increase, the variables in the y-dataset will decrease proportionally. If two datasets have a correlation of 0, it means that there is no increase or decrease in y as x increases, or simply that there is no relationship between the two datasets. Reference [33] provides more details about the applied correlation method in this study.

Because weather data is obtained each hour, this results in vectors of wind and solar generation data organized by state and generation type. The inherent organization of this data, both by state and type of generation, makes it particularly suitable for correlation comparisons. The vectors are allow for versatile organization by region or over different time scales. This extensive and well-organized dataset facilitates a meaningful correlation analysis, providing insights into the relationships between weather conditions and renewable energy generation at various scales. To further understand patterns, correlation values can be taken over long periods of time at intervals to get a vector, which aids in seeing if there are patterns.

Data is analyzed using the Pandas and Numpy packages in Python [34], [35]. All heat maps shown in this paper were created using the Seaborn package in Python [36]. All other graph curves generated using the Matplotlib package in Python [37].

IV. Case Study

To demonstrate the concepts discussed thus far, this section will showcase their the practical application by analyzing the two largest grids in North America and the possibility of their joining. The focus is on evaluating the potential benefits of sustainability and efficient utilization of renewable generation in light of weather differences. Figure 3 shows the geographical distribution of EI and WECC areas.

When studying the possibility of the EI and WECC connection, it is important to specifically study the states that are near or on the border between the two grids and possible power transfers among them.

Although the population is less dense in the central states compared to states that are on the East and West Coast, the majority of power transfer that occurs will likely be a result of generation and demand in this particular area. The actual power flow within a grid depends on many factors, such as load demand, generation capacity, transmission capacity, and market dynamics. So in the event of a synchronous grid, it is unlikely that a state such as California (on the west coast) will supply a state such as Maine (far away in the Northeast).

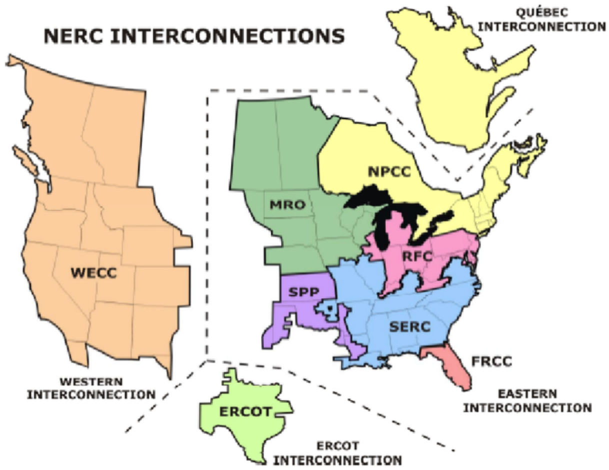

Fig. 3: North American four major electric grids [38]

TABLE I: States Studied

WECC States	EI States
Arizona (AZ)	Arkansas (AR)
Colorado (CO)	Iowa (IA)
Idaho (ID)	Kansas (KA)
Montana (MT)	South Dakota (SD)
New Mexico (NM)	Minnesota (MN)
Utah (UT)	Missouri (MO)
Wyoming (WY)	Nebraska (NE)
	North Dakota (ND)
	Oklahoma (OK)

Therefore, the eastern corridor will be studied mainly in this section. Table I shows the studied states near the seam. It is important to note that although some states are split between the EI and WECC, they are placed in either in the EI grid or WECC grid based on the location of the majority of their land mass for simplicity and availability of data. This assumption is valid, as there are no major sites (such as large cities or large generators) that are in these areas. Additionally, Canada is not included in this study.

V. Analysis and Results

A. *Renewable Power Correlation between Grids*

Using the Pearson correlation method discussed in section III and the Pandas Python package [34], the renewable generation in the year 2021 is analyzed on a daily basis for the entire year. A correlation coefficient is calculated for each day, and an array is produced to illustrate the entire year (so each correlation array has 24 time points, resulting in 365 coefficients). As seen in Figure 4, solar generation has a high correlation value overall, as this is to be expected due to the proximity of studied states near the seam and the nature of sunlight traveling across a region, often with different time zones. On the other hand, wind generation experienced a far more erratic pattern, some days with a correlation near one and other days with a negative correlation.

To analyze the weather data for a longer time period and possibly cut out some of the noise, the solar and wind generation correlations are calculated on a monthly basis dating back to 1972 (so each correlation array has around 720 time points due to the number of hours in a month, resulting in 588 coefficients because of the number of months studied). As seen in Figure 5, the correlation in solar generation output has centered around 0.8, while the correlation in

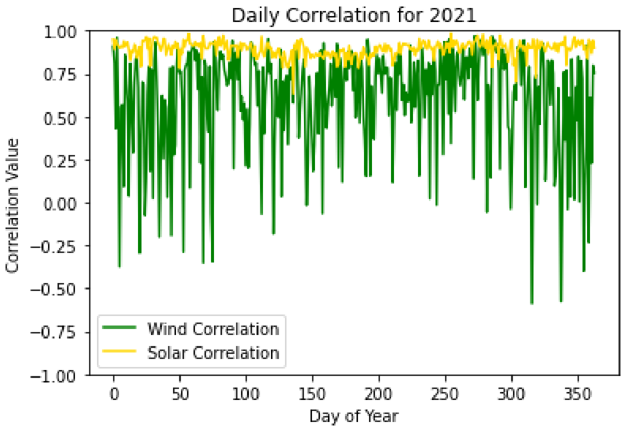

Fig. 4: Daily correlation between the electric real power generated from wind and solar in studied states of EI and WECC in 2021

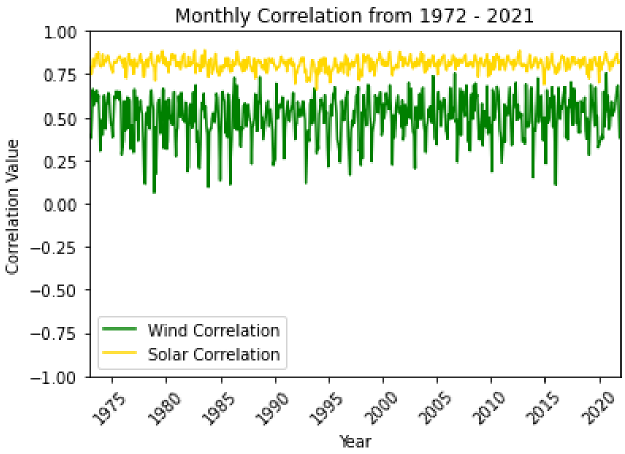

Fig. 5: Monthly correlation between the electric real power generated from wind and solar in studied states of EI and WECC from 1972-2021

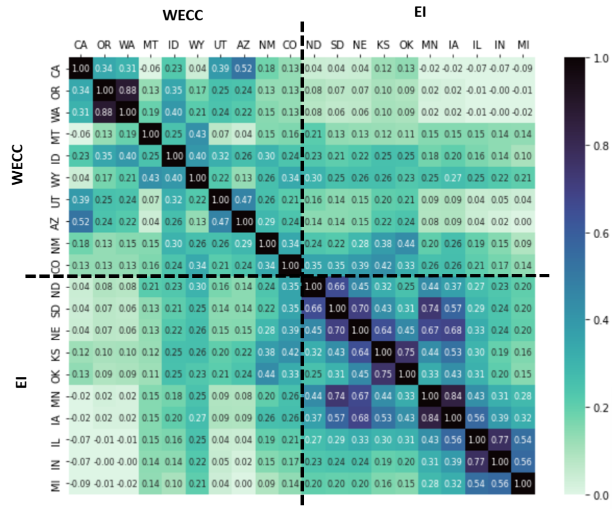

Fig. 6: Wind Generation Correlation from 1972 - 2021

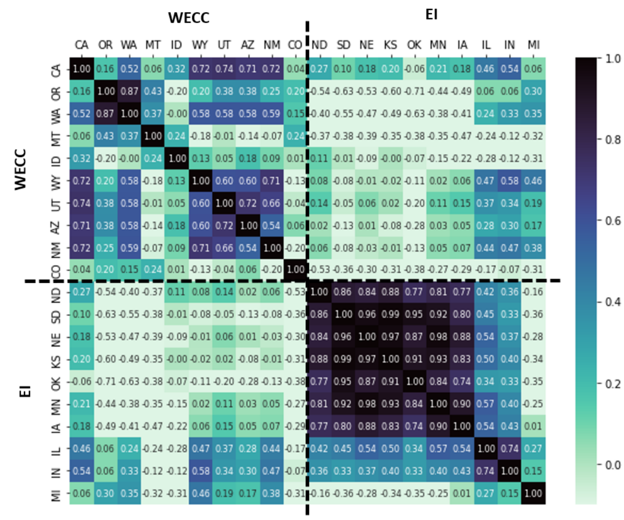

Fig. 7: Correlation of Wind Output from March 9, 1998

wind generation output has centered around 0.5. Therefore, according to Figure 5, solar correlation continues to remain higher than wind correlation.

Figure 5 shows that the correlation values in wind generation vary greatly and do not center around 1, therefore indicating that there could be potential benefits from high transfer capacity between the EI and WECC grids. This is especially highlighted in daily correlation as seen in Figure 4, as the wind generation changes substantially, often times even dipping into negative correlation, which possibly means that one side of the grid is producing significantly more wind energy and is therefore possible to share power across the seam.

B. Renewable Power Correlation Between US States

To further understand the relationships between renewable generation output, there is a need for a more detailed study on the individual renewable correlation between each state's renewable output. As seen in Figure 4, there are certainly days that merit further study in a synchronous grid consideration. In Figures 6, 7, 8, 10, and 11, the studied states are arranged in spatial order based on their geographic location (from West to East).

The top 10 states for installed wind generation [39] in both the EI and WECC can be compared and their output power from renewable correlations are calculated. Figure 6 shows the correlation of wind generation dating back to 1972, with EI and WECC states divided with a dotted line. From this heat map, it can be seen that the highest correlation values occur around the diagonal. This is expected, as the wind generation in states that are near each other is logically going to be similar, as wind patterns usually do not change drastically in close regions. Some notable points are Oregon and Washington, which have a very high correlation to one another compared to the rest of the map. Additionally, the states on the EI side exhibit more similarity than states on the WECC side.

To do a more in-depth analysis, Figure 7 shows only 24 hours of correlation study on a day of particular interest. In this heat map, there is a clear division in wind generation output between the EI and WECC grids. This highlights the possible benefits of a connected grid, as a day such as this would yield a high amount of power transfer between the grids to better utilize the wind generation.

Next, the top 10 states for installed solar generation [40] in each grid are compared to one another. Solar generation

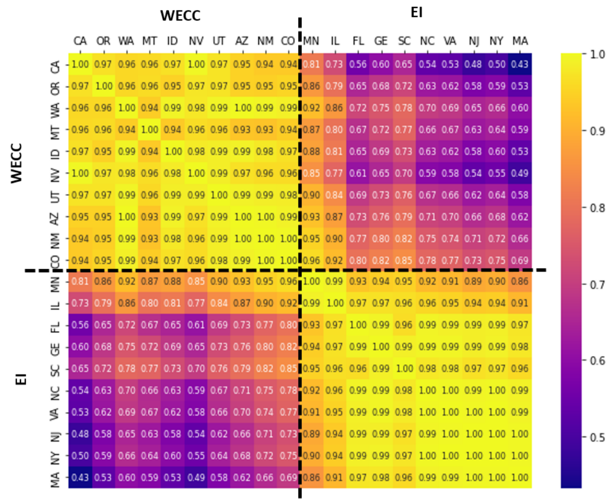

Fig. 8: Solar Generation Correlation from April 8, 2015

correlation has a less specific pattern due to the nature of sunlight. Given clear sky conditions, the time at which a solar generator outputs its highest amount of power is solar noon, the time of day when the sun is at its highest point in the sky. In a grid that has a vast range of latitudes and longitudes, solar noon occurs at very different times, increasingly so across the states on the opposite coasts of the United States. As shown in Figure 8, states that are in close proximity to one another have a high correlation of solar generation, but there is still a big difference between the EI and WECC grids. Simply due to the behavior of the Earth's rotation of the sun and time differences, there is a notable differential in solar power and therefore a need for power transfer between the EI and WECC grids.

As seen from Figures 6 and 7, there is a strong relationship between Oregon's and Washington's wind generation. Using a geographic data view (GDV) [20] and knowing the geographic location of all US generators over one megawatt [17], Figure 9 shows the capacity of wind generation in these two states. The majority of generation here occurs in the Columbia River Gorge due to that area having consistently high winds (as also explored in [2]). Moreover, there tends to be a high correlation between flat, prairie states, as seen in Figures 6 and 7. These states also have high wind energy production because of the consistently high winds and flat terrain. These two unique areas pose an intriguing comparison, considering they both have high wind generation but are in different climates. Figure 9 shows the correlations between the Columbia River Gorge (CRG) and the prairie states, separated by a dotted line with a scenario of particular interest. There is a very high correlation within the CRG states and within the prairie states, but the interesting finding is that the two regions are negatively correlated. This means that there is indeed a relationship between the two, but that the wind generation has an inverse relationship.

It is important to note that Figure 10 is an outlying day and those two areas do not always experience extreme differences. Figure 11 shows the correlation between the same states in the CRG and prairie spanning the entire time there has been weather data in those areas. In this lengthy analysis, there is still a high correlation within the individual areas, but around zero correlation between the two.

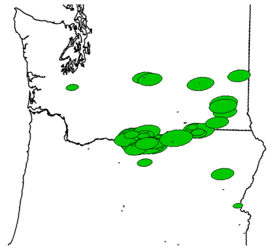

Fig. 9: GDV for Oregon and Washington

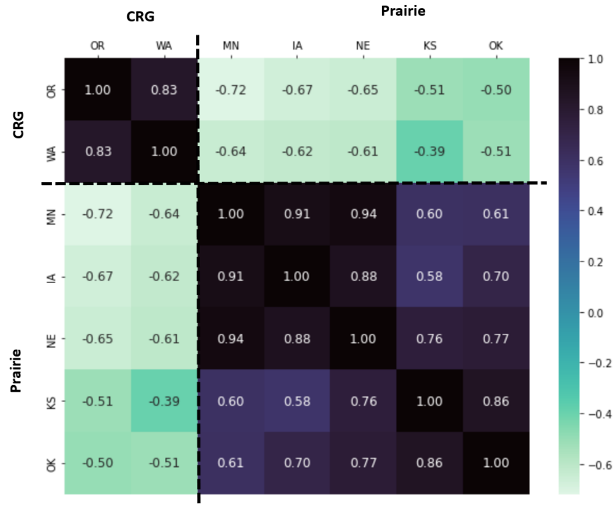

Fig. 10: Wind Generation Correlation in the CRG and Prairie on December 19, 2008

VI. Conclusion

This paper discussed a novel approach of analyzing weather data in the context of renewable generation. First, the explanation covered the process of obtaining hourly generation data for wind and solar energy, utilizing raw weather data and generator data sourced from publicly accessible datasets. Next, the correlation method was covered, considering the weather data vectors. Finally, these techniques were demonstrated on the EI and WECC grids of North America. It was found that there is typically a solar correlation of around 0.75 and a wind correlation of 0.5 which shows that the renewable resource generation could be shared between the EI and WECC grids. Further analysis found that there are days when there is a large difference in the wind and solar available capacities on each side. Additionally, an interesting behavior of negative correlation of wind generation between states in the Columbia River Gorge and prairie states was found as well as zero or negative correlation for the highest producers of renewable energy. This shows that a joined grid would better utilize renewable energy due to its varying nature and differences in production across EI and WECC grids, therefore promoting long-term sustainability. The methods in this paper were demonstrated on the EI and WECC grids, but they can be used in any grid when studying transmission upgrade potential.

Acknowledgment

This work was partially supported through funding provided by the Power Systems Engineering Research Center (PSERC) through project S-99, S-102G and partially by

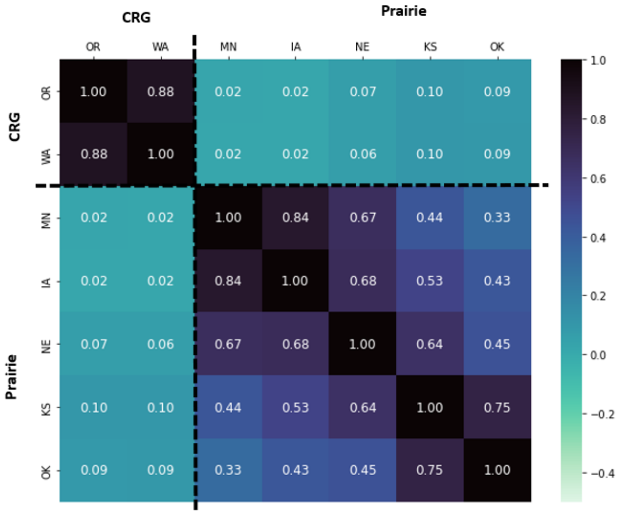

Fig. 11: Wind Generation Correlation in the CRG and Prairie ranging from 1972 to 2021

Advanced Research Projects Agency–Energy (ARPA-E) for grid optimization projects.

References

[1] K. Antonio, "Renewable generation surpassed coal and nuclear in the u.s. electric power sector in 2022," March 2023, [Online; posted 27-March-2023]. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=55960

[2] T. J. Overbye, F. Safdarian, W. Trinh, Z. Mao, J. Snodgrass, and J. H. Yeo, "An Approach for the Direct Inclusion of Weather Information in the Power Flow," *Proc. 56th Hawaii International Conference on System Sciences (HICSS)*, 2023.

[3] G. Ren, J. Wan, J. Liu, and D. Yu, "Spatial and temporal correlation analysis of wind power between different provinces in China," *Energy*, vol. 191, p. 116514, 2020.

[4] S. Liu, G. Li, H. Xie, and X. Wang, "Correlation characteristic analysis for wind speed in different geographical hierarchies," *Energies*, vol. 10, no. 2, p. 237, 2017.

[5] T. E. Hoff and R. Perez, "Quantifying pv power output variability," *Solar Energy*, vol. 84, no. 10, p. 1782–1793, 2010.

[6] J. Widen, "Correlations between large-scale solar and wind power in a future scenario for sweden," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 2, p. 177–184, 2011.

[7] J. Cohn, "When the grid was the grid: the history of North America's brief coast-to-coast interconnected machine [scanning our past]," *Proceedings of the IEEE*, vol. 107, no. 1, pp. 232–243, 2018.

[8] A. L. Figueroa-Acevedo, "Opportunities and benefits for increasing transmission capacity between the us eastern and western interconnections," Ph.D. dissertation, 2017.

[9] Y. Li and J. D. McCalley, "Design of a high capacity inter-regional transmission overlay for the u.s." *IEEE Transactions on Power Systems*, vol. 30, no. 1, p. 513–521, Jan 2015.

[10] M. A. Elizondo, N. Mohan, J. O. Brien, Q. Huang, D. Orser, W. Hess, H. Brown, W. Zhu, D. Chandrashekhara, Y. V. Makarov, and et al., "Hvdc macrogrid modeling for power-flow and transient stability studies in north american continental-level interconnections," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 4, p. 390–398, Dec 2017.

[11] A. Bloom, J. Novacheck, G. Brinkman, J. McCalley, A. Figueroa-Acevedo, A. Jahanbani-Ardakani, H. Nosair, A. Venkatraman, J. Caspary, D. Osborn, and et al., *The value of increased HVDC capacity between eastern and western U.S. grids: The Interconnections seam study*, Oct 2020.

[12] K. S. Shetye, T. J. Overbye, H. Li, J. Thekkemathiote, and H. Scribner, "Considerations for interconnection of large power grid networks," in *2021 IEEE Power and Energy Conference at Illinois (PECI)*, 2021, pp. 1–8.

[13] T. J. Overbye, K. Shetye, H. Li, W. Trinh, and J. Wert, "Feasibility assessment of synchronous operations of the North American eastern and western interconnections," *PSERC Publication*, pp. 21–02, 2021.

[14] T. J. Overbye, K. Shetye, J. Wert, H. Li, C. Cathey, and H. Scribner, "Stability considerations for a synchronous interconnection of the north american eastern and western electric grids," *Proceedings 55th Hawaii International Conference on System Sciences (HICSS)*, Jan 2022.

[15] I. Graabak and M. Korpås, "Variability characteristics of european wind and solar power resources—a review," *Energies*, vol. 9, no. 6, p. 449, 2016.

[16] G. A. Giorgio, M. Ragosta, and V. Telesca, "Application of a multivariate statistical index on series of weather measurements at local scale," *Measurement*, vol. 112, pp. 61–66, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224117305031

[17] (2019) "U.S. Energy Information Administration (EIA)". [Online]. Available: https://www.eia.gov/electricity/data/eia860/

[18] 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 *2024 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2024, pp. 1–6.

[19] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Network flow and copper plate relaxations for ac transmission systems," in *2016 Power Systems Computation Conference (PSCC)*. IEEE, 2016, pp. 1–8.

[20] T. J. Overbye, J. L. Wert, K. S. Shetye, F. Safdarian, and A. B. Birchfield, "The use of geographic data views to help with wide-area electric grid situational awareness," in *2021 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2021, pp. 1–6.

[21] "Current Weather and Wind Station Data". [Online]. Available: https://aviationweather.gov/adds/dataserver_current/current/metars.cache.csv

[22] "Weather Station Identifiers". [Online]. Available: http://www.weathergraphics.com/identifiers/

[23] Energy Systems Integration Group, "Weather DataSet Needs for Planning and Analyzing Modern Power Systems," Oct. 2023.

[24] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle *et al.*, "The modern-era retrospective analysis for research and applications, version 2 (merra-2)," *Journal of climate*, vol. 30, no. 14, pp. 5419–5454, 2017.

[25] S. G. Benjamin, S. S. Weygandt, J. M. Brown, M. Hu, C. R. Alexander, T. G. Smirnova, J. B. Olson, E. P. James, D. C. Dowell, G. A. Grell *et al.*, "A north american hourly assimilation and model forecast cycle: The rapid refresh," *Monthly Weather Review*, vol. 144, no. 4, pp. 1669–1694, 2016.

[26] G. M. Masters, *Renewable and efficient electric power systems*. John Wiley & Sons, 2013.

[27] V. Sohoni, S. Gupta, R. Nema *et al.*, "A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems," *Journal of Energy*, vol. 2016, 2016.

[28] P. Giorsetto and K. F. Utsurogi, "Development of a new procedure for reliability modeling of wind turbine generators," *IEEE Transactions on Power Apparatus and Systems*, no. 1, pp. 134–143, 1983.

[29] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015.

[30] PowerWorld Developers, "Power Flow Weather Models: WindClass1, WindClass2, WindClass3, WindClass4, WindBasic," https://www.powerworld.com/WebHelp/s.

[31] J. Hauke and T. Kossowski, "Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data," *Quaestiones geographicae*, vol. 30, no. 2, pp. 87–93, 2011.

[32] N. Faizi and Y. Alvi, "Chapter 6 - correlationfor datasets, please refer to companion site: https://www.elsevier.com/books-and-journals/book-companion/9780443185502," in *Biostatistics Manual for Health Research*, N. Faizi and Y. Alvi, Eds. Academic Press, 2023, pp. 109–126. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780443185502000025

[33] A. Edwards, "The correlation coefficient," *An introduction to linear regression and correlation*, vol. 4, pp. 33–46, 1976.

[34] T. pandas development team, "pandas-dev/pandas: Pandas," Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134

[35] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[36] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. [Online]. Available: https://doi.org/10.21105/joss.03021

[37] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[38] E. Ela, M. Milligan, and B. Kirby, "Operating reserves and variable generation," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2011.

[39] N. D. of Environment and Energy, "Wind facilities' installed capacity by state," Online, 2023, accessed: 2023-11-10. [Online]. Available: https://neo.ne.gov/programs/stats/inf/205.htm

[40] S. E. I. Association, "Solar state by state," Online, 2023, accessed: 2023-11-11. [Online]. Available: https://www.seia.org/states-map