Second Stage in Feasibility Assessment of Synchronous Operations of the North American Eastern and Western Interconnections

Final Project Report

S-102G

Power Systems Engineering Research Center
Empowering Minds to Engineer
the Future Electric Energy System

Second Stage in Feasibility Assessment of Synchronous Operations of the North American Eastern and Western Interconnections

Final Project Report

Project Team
Thomas J Overbye, Project Leader
Adam Birchfield
Farnaz Safdarian
Jonathan Snodgrass
Texas A&M University

Graduate Students
Jordan Cook
Seri Kang
Texas A&M University

PSERC Publication 24-01

April 2024

For information about this project, contact:

Thomas J Overbye
Professor and Holder of the O'Donnell Foundation Chair III
Department of Electrical and Computer Engineering
Texas A&M University
308C, Wisenbaker Engineering Building
College Station, Texas 77843-3128
Phone: 979-458-5001
Email: overbye@tamu.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
527 Engineering Research Center
Tempe, Arizona 85287-5706
Phone: 480-965-1643
Fax: 480-727-2052

Acknowledgments

We would like to thank our entire Industrial Advisory group, with special thanks to SPP.

Executive Summary and Abstract

The main goal of this project is an economic analysis of interconnecting Eastern Interconnection (EI) and Western Electricity Coordinating Council (WECC) ac interconnect under various load and weather conditions. The potential benefit would be the possibility of more renewable energy transfer across the interconnection or overall less renewable curtailment and accordingly less operation cost after the connection of these grids. This project builds on PSERC S-92G work on the stability analysis of these two grids. Due to the use of Critical Energy/Electricity Infrastructure Information (CEII), this report primarily outlines the general methodologies developed, reserving CEII-specific details for a separate report that includes a variety of power flow and stability models for the EI and WECC, using 2022 series EI cases and 2023 series WECC cases with the combined case having more than 123,000 buses. To facilitate combining EI and WECC cases, some WECC bus numbers were changed to avoid number overlap with the EI. The cases were then combined by adding ac interconnections at the nine interconnection points previously identified in S-92G. During this process, almost all the buses were assigned to more than 47,000 electrical substations and all the substations were geo-mapped.

A strategy is proposed to directly use weather data initially from sources like International Civil Aviation Organization (ICAO), the World Meteorological Organization (WMO), and later from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis fifth generation data (ERA5) data to calculate the output power of renewable resources. For this goal, the US Energy Information Administration Form 860 (EIA-860) dataset is used to find the types of wind turbines and solar cells. Additionally, EIA-860 data are used to create power systems using a copper plate model, which allows for generation to flow with zero impedance. Various power flow structures, such as buses, substations, and weather-dependent models, are used, and advanced visualization techniques like geographic data views and contour mapping are explored, enhancing the understanding of complex generation data. The coupling of weather information to power flow studies is highlighted as a useful tool for calculating expected wind and solar generation outputs based on weather conditions. To create load time series in each area, Federal Energy Regulatory Commission (FERC) regions are assigned to the areas and FERC form 714 is used to create load time series in each region. A clustering strategy is proposed to find interesting load and weather scenarios for economic analysis.

Also, the correlation of renewable generation across the EI and the WECC is calculated with consideration of load for the potential economic benefits of this interconnection. It is 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 can be shared between the EI and WECC grids. Further analysis found that there are days when there is a larger difference in the wind and solar available capacities on each side. Additionally, an interesting behavior of negative correlation of wind generation between some areas in EI and WECC was found as well as zero or negative correlation for the highest producers of renewable energy, which allows even more power transfer between EI and WECC grids. Also, a negative correlation between load demand and wind generation across the EI and WECC grids refers to the need for power transfer between the two grids. This shows that a joined grid would better utilize renewable energy with its varying nature and differences in production across EI and WECC grids, promoting long-term sustainability.

Also, a strategy for scenario selection with the goal of studying a comprehensive variety of weather and load scenarios for power flow simulations is presented. With large power systems and large amounts of available data, it is computationally expensive to run simulations on all scenarios and it is required to select a wide variety of scenarios with different impacts on the operation, considering load and weather for renewable generation output. Using the K-Means method for clustering, representative points are strategically chosen to simulate various solar, wind, and load conditions. The two selected representative points include an average and an outlier. Choosing these two points allows for baseline data analysis as well as anomalies, which can cause stress in the grid. The method is then demonstrated to show its functionality and how it captures the diversity of a dataset. The resulting clusters help finding interesting scenarios by addressing the variability that is inherent in power systems. This leads to improving grid reliability by preparing for a range of scenarios.

Project Publications:

[1] T.J. Overbye, S. Kunkolienkar, J. Snodgrass, A.B. Birchfield, "On the Existence of Dominant Inter-Area Oscillation Modes in the North American Eastern Interconnect Stability Simulations," submitted to 2024 Hawaii International Conference on System Sciences (HICSS), January 2024, Honolulu, HI.

[2] K. Zhgun, F. Safdarian, J. Griffin, T.J. Overbye, "Identifying Problematic AC Power Flow Alternative Solutions in Large Power Systems", 2023 North American Power Symposium, October 2023.

[3] J. L. Wert, T. Chen, F. Safdarian, J. Snodgrass, T. J. Overbye, "Calculation and Validation of Weather-Informed Renewable Generator Capacities in the Identification of Renewable Resource Droughts," IEEE PowerTech 2023, Belgrade, Serbia, June 2023.

[4] T.J. Overbye, K.R. Davis, A.B. Birchfield, "The Electric Grid and Severe Resiliency Events." National Academy of Engineering Bridge Journal, vol. 53, no. 2, 2023, pp. 73-79, June 2023.

[5] A.B. Birchfield, T.J. Overbye, "A Review on Providing Realistic Electric Grid Simulations for Academia and Industry," Curr Sustainable Renewable Energy Rep, June 2023.

[6] T.J. Overbye, S. Kunkolienkar, "On the Existence of Distinct Inter-Area Electro-Mechanical Modes in North American Electric Grids," 2023 Kansas Power and Energy Conference (KPEC), April 2023.

[7] S. Kunkolienkar, F. Safdarian, J. Snodgrass, T.J. Overbye, "Quantification of Area Sparsity in Large-Scale Electric Grids," 2023 Kansas Power and Energy Conference, April 2023.

[8] 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.

[9] F. Safdarian, M. Stevens, J. Snodgrass, and T. J. Overbye, "Detailed Hourly Weather Measurements for Power System Applications," IEEE Texas Power and Energy Conference (TPEC), 2024.

Student Theses:

[1] J. L Wert, “EXTRACTING VALUE FROM ELECTRIC GRID INFORMATION TOWARDS A RESILIENT GRID AMIDST DECARBONIZATION EFFORTS,” (PhD Thesis, August 2023)

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

The overall purpose of this project has been to study the economic analysis and dynamic feasibility of an interconnection of the North American Eastern and Western electric grids. Currently most of the electricity used in North America (NA) is supplied by four major interconnects, shown in Figure 1-1, with each operating at 60 Hz but asynchronous with each other. These interconnects are the Eastern Interconnection (EI), the Western Interconnection (WI) or Western Electricity Coordinating Council (WECC), Texas (ERCOT), and the Quebec Interconnection. All of these ac networks are internally synchronized and are linked to each other only through dc ties. However, for several years between 1967 and the early 1970's the EI and WI were operated as a single synchronous system [1], which included 94% of the US generating capacity [2]. This interconnection was motivated by the November 1965 Northeast Blackout, which left 30 million people without power across eleven US states, and Canada. The interties functioned well at first but soon became unstable due to oscillations on the western side and large inadvertent exchanges. This led to overloading of transmission facilities, major system breakups, and reduced transmission capacity. Interconnecting large grids especially with ac ties is a big challenge that needs rigorous assessment and planning in terms of feasibility and potential benefits.

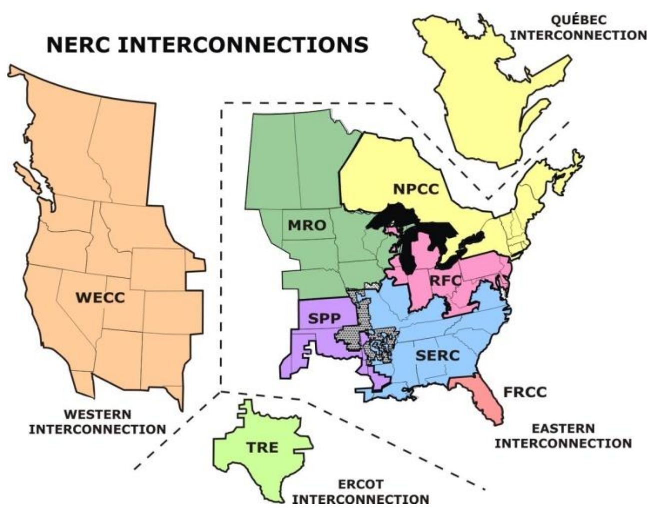

Figure 1-1 North American Electric Interconnects (Source NERC)

There have been several studies and implementations around the world of joining large grids with dc ties, and some examples with ac ties. In 1991, the continental Europe grid was broken into two synchronous grids separating western and central Europe due to political issues, and reconnected in 2000 with the emergence of favorable conditions [3]. This was done after extensive steady-state and dynamics studies [4]. For further expansion, [5] studied the feasibility of connecting this synchronous grid with the Baltic States. This involved creating a merged static and dynamics model of the two grids. Some of the issues found in this process were the emergence of very low frequency (~ 0.07 Hz) oscillations, as well as transfer capability limitations due to local congestion. Reference [6] considered possible scenarios for interconnecting Korea and Japan using 765 kV HVAC within the Korean peninsula and 180 kV HVDC interconnection between the islands. Results of power flow studies for load increase scenarios for the ac ties and different power injections for dc were shown. The need for, political issues with, and advantages of both schemes were discussed. In [7], two candidates were evaluated for the future Chinese "super grid", to enable bulk capacity long distance power transmission, i.e. 1) the ultra-high-voltage ac (UHVAC) synchronous power grid, 2) the extra high-voltage ac (EHVAC) asynchronous super power grid. This work provided qualitative assessments of both schemes considering security, economic, and environmental factors based on which the EHVAC asynchronous method was found to be superior, with a caveat that better studies are needed to verify the results. The benefits of the ac connections were lower short circuit currents compared to the asynchronous system, while the main disadvantage was the susceptibility to cascading failures.

References [8], [9] studied the economic aspects of interconnecting EI and WI grids with focus on resource planning aspects or the use of HVDC for transmission expansion and design [10]. These works are part of a larger effort comprising of research and industry members that proposed four different high-capacity wide-area transmission infrastructure designs to expand the US grid [11]. This study was focused on leveraging dc systems i.e. upgrading the existing back-to-back (B2B) dc ties and/or building long HVDC lines or overlays. While this included rigorous analyses considering future capacity, carbon policies, etc., a key area of improvement mentioned in [11] is performing contingency and stability analyses.

The previous project S-92G mainly focused on the dynamic aspects of an ac interconnection of North American EI and WI using both real grids and synthetic grid data with the project report available at [12], and a paper at [13]. The two synthetic grids [14] used there are available at [15], geographically located over the EI and WI footprints. The 70,000-bus eastern synthetic grid and the 10,000-bus western case bear no relation to the actual grids except that generation and load profiles are similar, based on public data. The transmission lines are entirely fictitious. These test systems are meant to reflect heavy load, i.e. peak summer conditions.

1.2 Overview of the Project and Report

This project builds on the previous work of S-92G for the stability analysis of an EI and WECC ac interconnect. The main goal of this project has been to study the economic impact of interconnecting EI and WECC grids in a variety of operating points including variation in load and weather scenarios. As such the project required the use of Critical Energy/Electricity Infrastructure Information (CEII) that cannot be disclosed in public reports such as this one. So the focus of this report is mostly to provide a description of the underlying methodologies that have been developed during the project since they should be applicable in many situations. A separate report then provides the CEII specific details.

By way of an overview of the methodology, the project utilized a variety of different power flow and stability models for the EI and WECC, with the final results done using a 2022 series EI cases with almost 96,000 buses and 2023 series WECC cases with more than 27,400 buses. To facilitate combining the cases the WECC cases were renumbered to avoid number overlap with the EI. The case were then combined by adding ac interconnections at the nine interconnection points previously identified in S-92G and given in Table 1-1. This result in a grid with more than 123,000 buses. During this process almost all of the buses were assigned to more than 47,000 electrical substations and all the substations were geo-mapped. A oneline for this grid is shown in Figure

1-2 with the nominal voltage of the ac transmission lines indicated by the different colors (green for above 700 kV, orange for 500 kV, red for 345 kV, blue for 230 kV, and black for lower voltages; the HVDC lines are not shown). This study did not involve ERCOT.

Table 1-1 Initial Interconnection between EI and WECC grids

No.	Location	East Bus (kV)	West Bus (kV)	Capacity
1	Fort Peck, MT	230	161	200 MVA
2	Miles City, MT	230	230	239 MVA
3	Rapid City, SD	230	230	287 MVA
4	Laramie, WY	345	345	800 MVA
5	Stegall, NE	230	230	150 MVA
6	Sidney, NE	230	230	240 MVA
7	Lamar, CO	345	230	239 MVA
8	Blackwater, NM	230	345	400 MVA
9	Artesia, NM	230	345	277 MVA

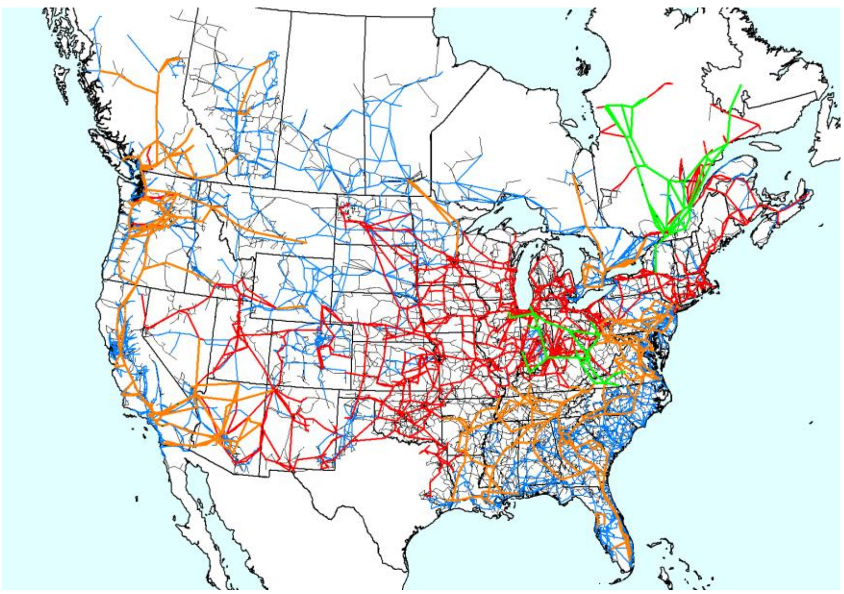

Figure 1-2 North American grid

For the economic analysis, estimated cost characteristics were developed for all of the generators in the US portion of the footprint. This was facilitated by mapping the EI and WECC generators to their corresponding generators in the US Energy Information Administration Form EIA-860 (EIA-860), which provided more information about each device including its fuel type and whether it had been retired. This information is visualized in Figure 1-3 with the size of each oval proportional to the total substation generation, and color used to indicate the primary fuel type (orange for natural gas, black for coal, red for nuclear, green for wind, blue for hydro and yellow for solar).

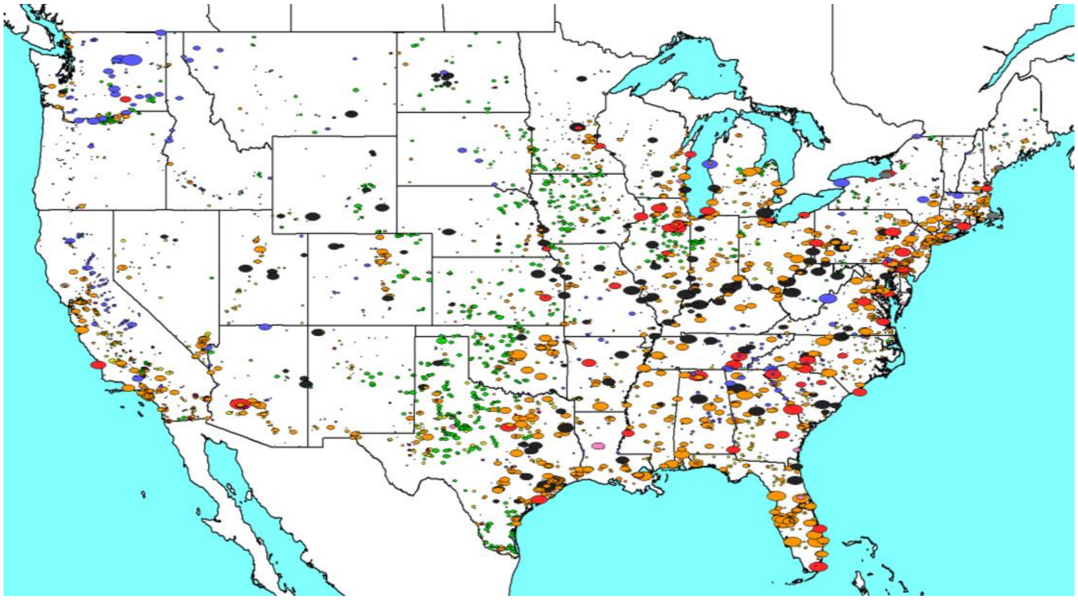

Figure 1-3 US Substation Generation by Primary Fuel Type

Since the intent of the study was to look at the operation of a combined grid under many different conditions, and the grid itself has a large amount of wind and solar, and important part of the study was utilizing realistic weather information. This was done primarily by considering historical data. At the beginning of the study this was done using measurements from weather stations across United Stares are gathered from the two of the main resources International Civil Aviation Organization (ICAO) and the World Meteorological Organization (WMO), [16]. Later in the study the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis fifth generation data or ERA5 data was utilized [17, 18]. Weather measurements are then used to calculate the output power of renewable resources based on the approach presented explained in [22]. To create load time series in each area, Federal Energy Regulatory Commission (FERC) regions are assigned to the areas and FERC form 714 [19] is used to create load time series in each region.

The CEII specific portion of the study then involved performing a variety of economic and stability analyses considering a variety of different operating points. The initial studies were performed using the existing and anticipated new transmission in each of the interconnects, and the potential nine previously mentioned ac interconnections. The overall results showed economic benefits in doing the interconnection, and no stability concerns. Then, working in consultation with the industrial advisors, the additional of new transmission lines added for the specific purpose of increasing the EI-WECC ac transfer capacity were considered. As expected, this did provide additional economic benefit and enhanced stability. However, these potential benefits need to be weighed against associated with developing these new lines.

1.3 Report Organization

As noted, the purpose of this report is to provide some of the general research results developed during this project to help facilitate the CEII portion of the project reported elsewhere. Therefore this report is organized as follows. Chapter 2 describes the economic analysis of the studied electric grids under a variety of load and weather scenarios and provides the optimal power flow and cost results with and without the interconnections as well as the correlation study of renewables and load in EI and WECC to study the potential benefits of the interconnection. Chapter 3 explains how to use US Energy Information Administration generation data for power flow analysis and visualization. Chapter 4 provides a scenario selection strategy for improved power flow studies. Chapter 5 provides a dynamic analysis using a simulation-based approach to study the extent to which specific oscillation patterns exist in large-scale electric grids and the last chapter provides a conclusion.

2. Economic Analysis of Interconnecting East and West Interconnections

2.1 Introduction

The most recent studies prompt the question,” Is there a potential benefit in utilizing renewable generation with this interconnect?” If there is a possibility of renewable power transfer across the EI and WECC grids, the renewable curtailment would be decreased. Given that the output of solar and wind generation is contingent on regional weather patterns, it is crucial to investigate such weather variations. This chapter summaries the potential benefits of synchronizing the EI and WECC grids into a single grid. By utilizing historical weather data, projecting power output from existing wind and solar generators, and applying Pearson correlation coefficients, this chapter inquiries into the prospective advantages of a synchronized grid, particularly in light of the unique relationships between solar generation, wind generation, and load demand.

2.2 Potential Renewable Generation Benefits of Interconnection

Several studies have run statistical analyses on weather, such as [20], which studies wind and solar output in Europe with spatial correlation, or [21] 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.

Renewable generation is non-dispatchable, meaning that it cannot be increased nor decreased based on changing electrical needs, unlike other sources, such as coal or natural gas. Without advanced electrical energy storage devices, wind and solar power must be used at the time it is generated. If a grid has high amounts of renewable generation but not a high demand for power at the time, that energy may be wasted, but in the event of a joined grid, it will be more efficiently utilized. This joined grid has the potential to increase the amount of used renewable generation and decrease the reliability of less sustainable energy sources.

Load demand and renewable generation have been studied in the literature (such as [22] and [23]), but these studies use renewable generation output instead of direct weather data. This chapter uses historical weather data to study the output of renewable generation with the latest generators’ data. By utilizing decades of historical weather data, the research explores various scenarios, enabling the examination of modern generation capabilities in conjunction with historical weather patterns. This allows the execution of simulations using weather data from recent decades on a designated grid with specified wind and solar capacities, facilitating the examination of the impact of significant weather changes in the current grid.

2.3 Weather Data to Generation Output

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 unprocessed meteorological data is incorporated in this work. Based on the strategy as shown in [24], weather data in operation and planning problems, such as

Optimal Power Flow (OPF), could refine the accuracy of renewable generation forecasting and capture rapid climatic shifts, without significantly increasing computational complexity.

The applied methodology in this work incorporates a plethora of meteorological variables into hourly evaluations of wind and solar generation output. This utilizes data from 1972 to 2022, sourced from meteorological stations across the U.S. [25] Historical weather data is collected at an hourly granularity from various sources worldwide [26], [27].

Using the geographic coordinates for each generator provided by the U.S. Energy Information Administration's form 860 (EIA-860) [28], each generator is partnered with the nearest weather station. For missing data, interpolation of nearest station data is utilized. Moreover, the EIA-860 dataset provides granular data on power generators which is instrumental in classifying similar wind and solar units into collective power plants by location.

This study uses six energy input-output models to quantify the influence of weather on renewable generators. The first model, which incorporates local wind speeds and turbine power curves as per [29], [30], and [31], calculates wind power plant output based on the wind measurements. Subsequent models two, three and four adapt this framework to accommodate different wind turbine classifications and power curves based on [32]. The fifth model estimates solar PV generation using local solar radiation, cloud coverage and PV characteristics such as tilt angle, azimuth angle and power point tracking as specified in [25], and the final model, drawing from [33], updates the thermal generator capacity based on temperature. The results of these methodologies have been validated in [34].

It is important to mention that the purpose of using historical data from the last years 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. Also, the generation mix in this study is based on the end of year 2021 and all interesting weather scenarios are studied on the same grid.

2.4 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 [35] and [36], 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 in one region is being compared to weather data in another region, Pearson's correlation coefficient is chosen for its linear characteristic. Pearson's correlation coefficient is calculated in (7_1):

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \qquad (2\_1)$$

where r is the Pearson correlation coefficient, x_i is each sample of x, $\bar{x}$ is the mean of x variable y_i is each sample of y, $\bar{y}$ is the mean of y variable.

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 also increase proportionally in the y-dataset. If two datasets have a correlation of -1, it means that if the variables in the x-dataset increase, the variables decrease proportionally in the y-dataset. If two datasets have a correlation of 0, it means that there is no increase or decrease in y as x increases, or basically that there is no relationship between the two datasets. Reference [37] provides more details about the applied correlation method in this study.

Data is analyzed using the Pandas and Numpy packages in Python [38], [39]. All heat maps shown in this work were created using the Seaborn package in Python [40]. All other graph curves generated using the Matplotlib package in Python [41].

2.5 Renewable Power Correlation between Eastern and Western Interconnect Grids

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).

Therefore, the eastern corridor will be studied mainly in this section. The studied states near the seam include AZ, CO, ID, MT, NM, UT and WY that are considered in WECC grid and AR, IA, KA, LA, MN, MO, NE, ND, SD and OK in EI grid. 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.

Using the Pearson correlation method discussed in the previous section and the Pandas Python package [38], 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 2-1, 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 2-2, the correlation in solar generation output has centered around 0.8, while the correlation in wind generation output has centered around 0.5. Therefore, according to Figure 2-2, solar correlation continues to remain higher than wind correlation.

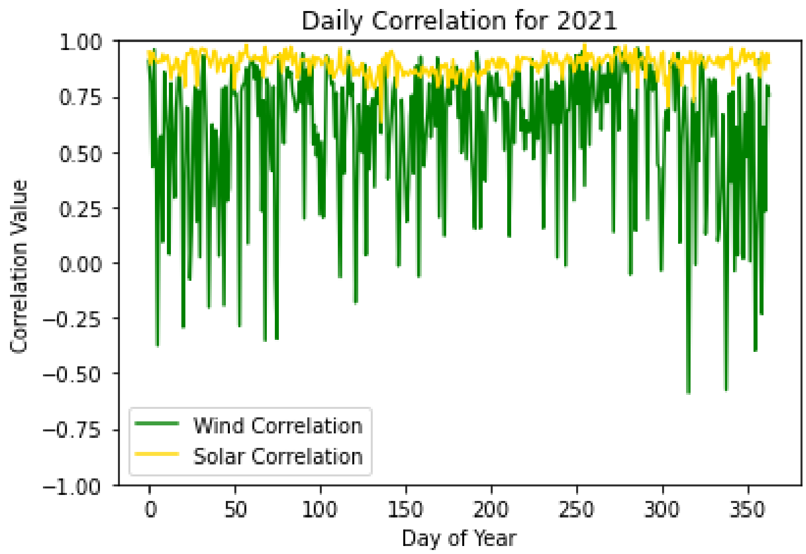

Figure 2-1 Daily correlation between the electric real power generated from wind and solar in studied states of EI and WECC in 2021

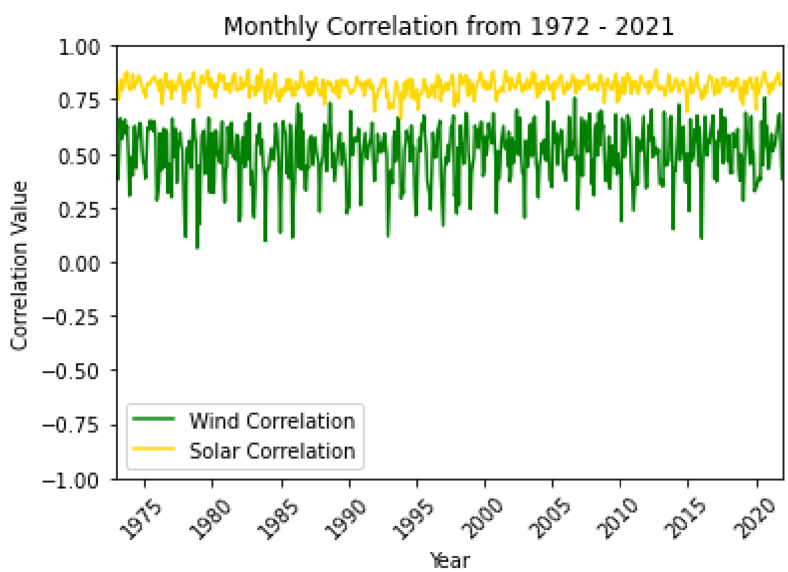

Figure 2-2 Monthly correlation between the electric real power generated from wind and solar in studied states of EI and WECC from 1972-2021

Figure 2-2 shows the long term need for the joined EI and WECC grids, as the correlation values in wind generation vary greatly and do not center around 1. This is especially highlighted in daily correlation as seen in Figure 2-1, as the wind generation changes substantially, often times even dipping into negative correlation, which possibly means that one side of the grid is producing far more wind energy and is therefore possible to share power across the seam.

2.6 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 2-2, there are certainly days that merit further study in a synchronous grid consideration. In Figures 2-3 to 2-5, 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 [42] in both the EI and WECC were compared and their output power from renewable correlations are calculated. Figure 2-3 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. It is important to note that Figures 2-3 is an example of an outlying date that particularly highlights the possible benefits of a joined grid.

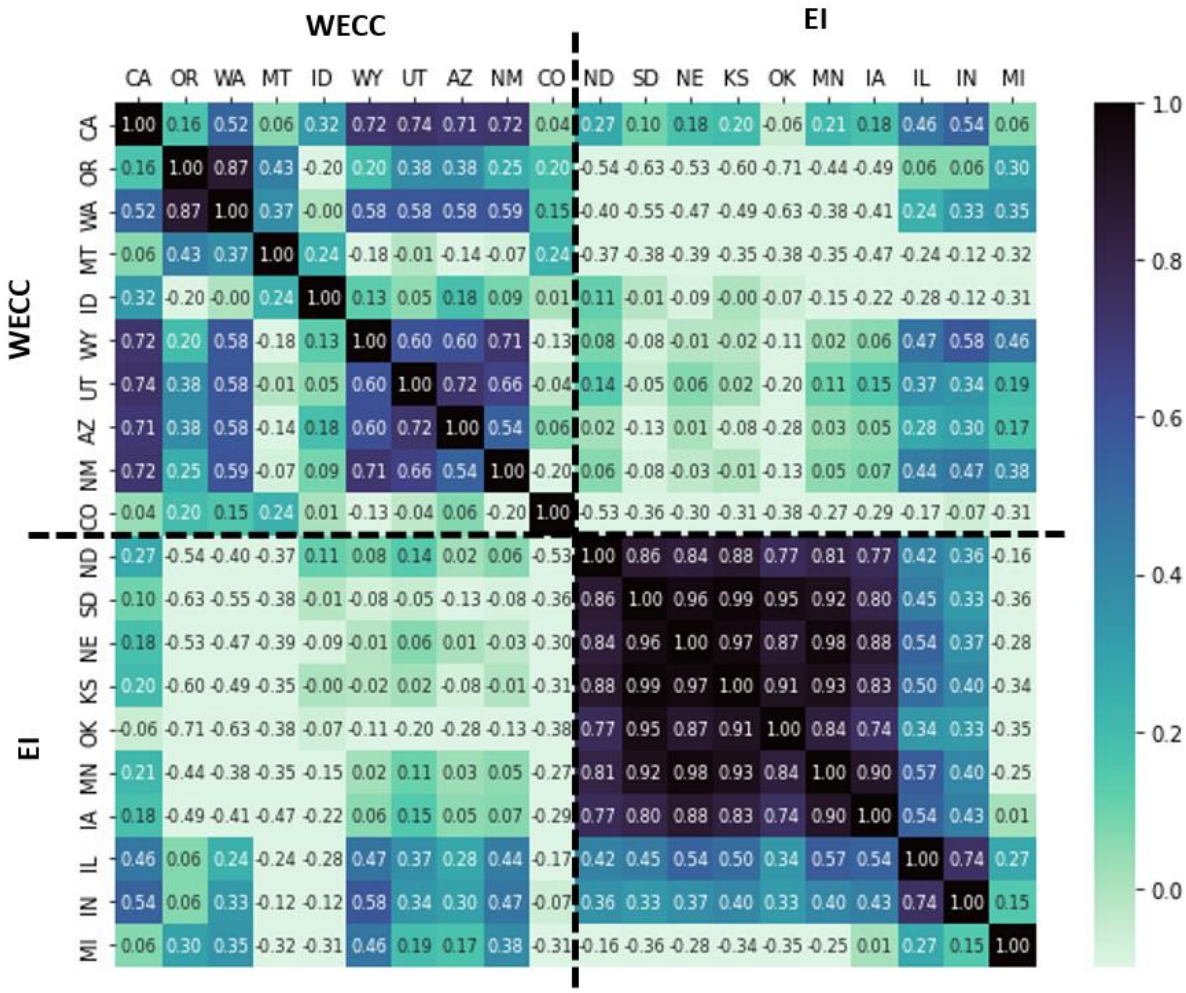

Figure 2-3 Correlation of Wind Output from March 9, 1998

Next, the top 10 states for installed solar generation [43] in each grid are compared to one another. Solar generation 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 [44]. 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 2-4, which shows solar generation correlation of WECC and EI states in a sample day, 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.

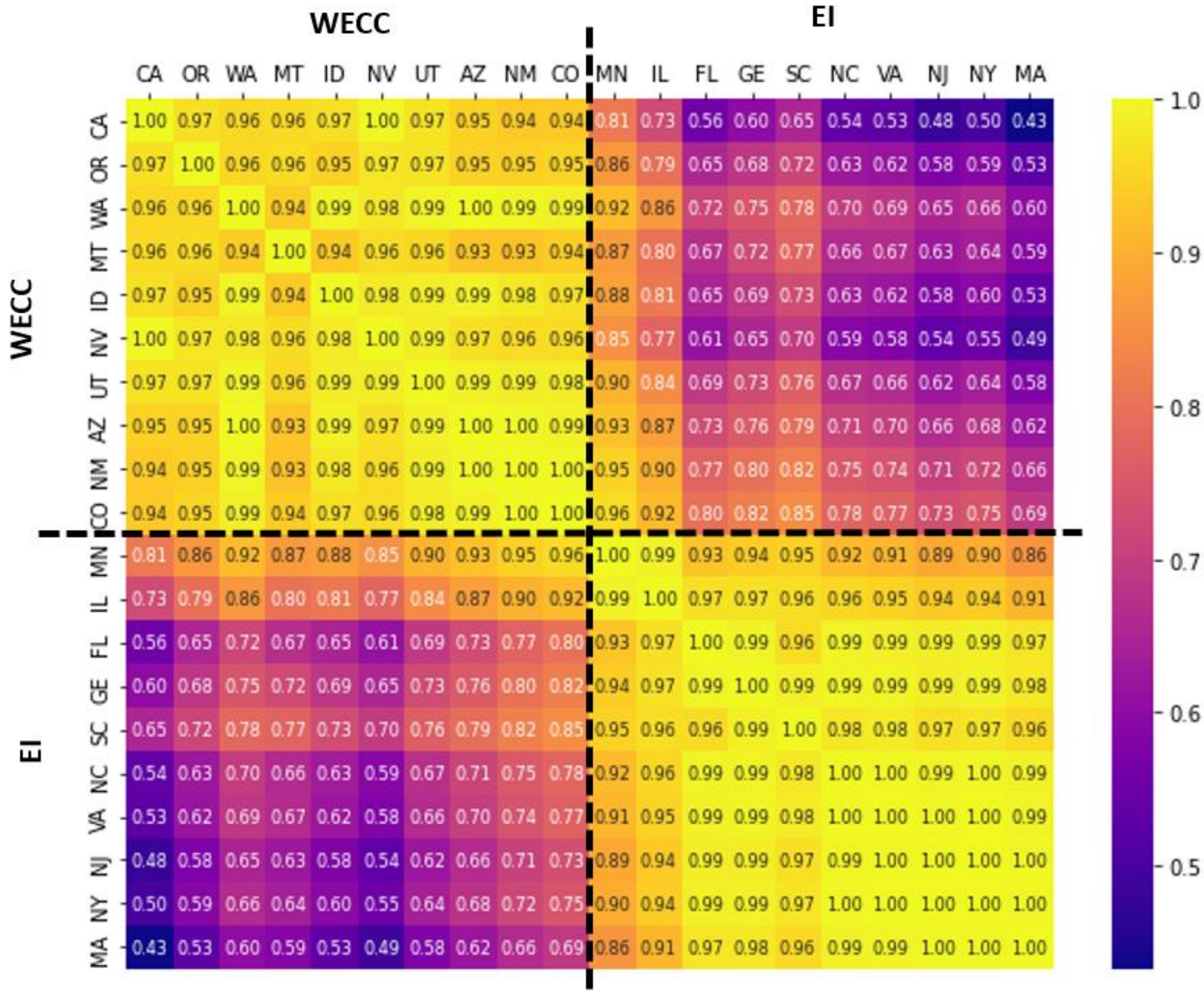

Figure 2-4 Solar Generation Correlation from April 8, 2015

Finally, the three states with the highest renewable energy capacity (solar and wind combined) are correlated considering both wind and solar sources. Florida, despite having a substantial solar capacity, is excluded from consideration due to having zero wind capacity. Figure 2-5 shows the summed solar and wind renewable generation correlation for the same US states since 1972. Although there is less of a defined differential in this heat map, there is still more correlation in the states inside each individual grid than the comparison between states on different sides of the seam. Note that the EI and WECC grids are separated with a dotted line and are organized spatially. Even with the combined solar and wind capacities, states inside each part of the grid experienced a higher correlation than states in different grids. Performing a cross-comparison, there is either a low correlation or more often a negative correlation.

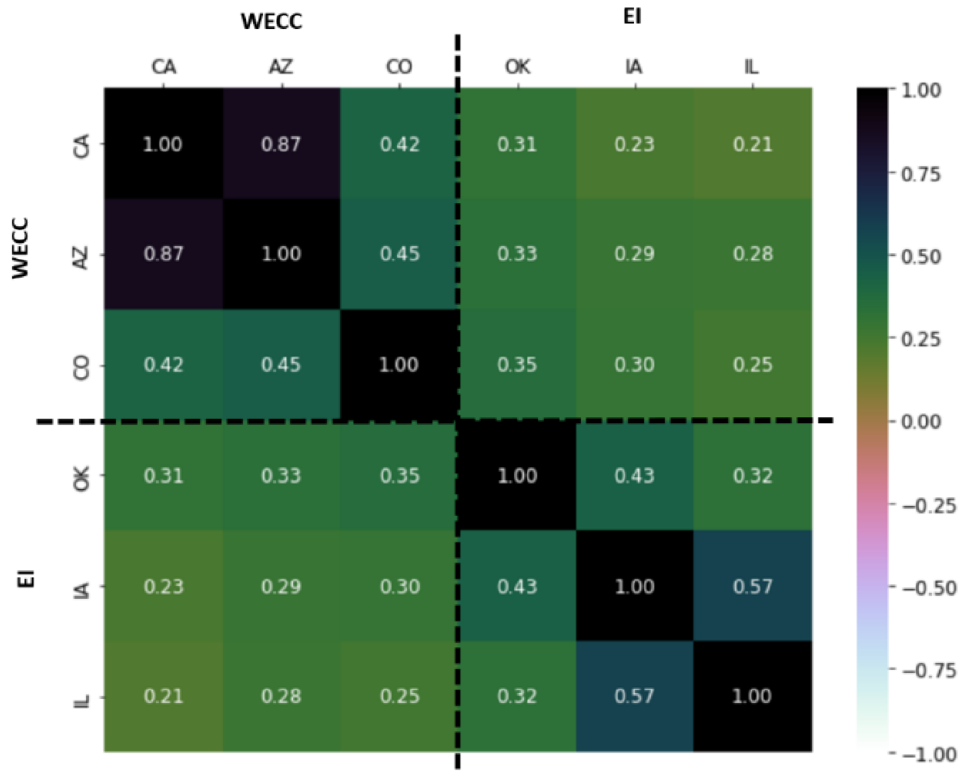

Figure 2-5 Overall Renewable Generation Correlation in Top 3 Producing States ranging from 1972 – 2021

2.7 Load and Wind Generation Correlation

Another comparison that should be considered in the possibility of joining the EI and WECC grids is the relationship between load demand and renewable generation, specifically wind generation. Factors such as peak load demand and wind generation patterns, along with the geographical distribution of renewable resources further contribute to the complexity of this comparison.

The Federal Energy Regulatory Commission (FERC) provides FERC Form No. 714, containing load data for every electric utility with a planning area having an annual peak demand for power 200 MW or more. This dataset is utilized when studying load-use and its changes [45]. As seen in Figure 2-12, the load typically peaks in the afternoon and decreases at night. While load exhibits patterns that are commonly predictable, renewable generation, particularly wind generation, is far more erratic. Solar does see a common pattern with its renewable generation due to the nature of the sun and capturing its energy. As for a general pattern, wind tends to increase at nights when solar generation is not available.

Because of this variable behavior, it is important to study the differences between the EI and WECC in terms of load and renewable generation. Figure 2-6 shows the normalized load for the year 2022. The data has been normalized to make direct comparisons between the two grids. This normalization is crucial due to the significant difference in load demand between the EI and

WECC. By scaling the data as a percentage of each grid, similarities or differences become more discernible, providing a clearer understanding of their respective characteristics.
As seen in this image, both the EI and WECC exhibit similar patterns of load use.

While Figure 2-6 might not show any notable difference in load demand between the two grids, Figure 2-7 does show a notable difference in wind generation trends. It is important to mention that the EI grid has much more installed wind generation than the WECC grid, but the WECC wind generation stayed fairly constant during the year 2021. The EI had a decrease in its wind generation in the summer months and then increased again in the fall into spring. Figure 2-8 correlates the load demand and wind generation on a 24 basis for the entire year of 2021. From fall to winter, both grids are consistent with an erratic correlation. On the other hand, the spring to summer season notices a difference: the EI grid stabilizes above 0 and the WECC grid stabilizes below zero. This pattern is an important finding, as these differences further communicate the benefit of a joined grid.

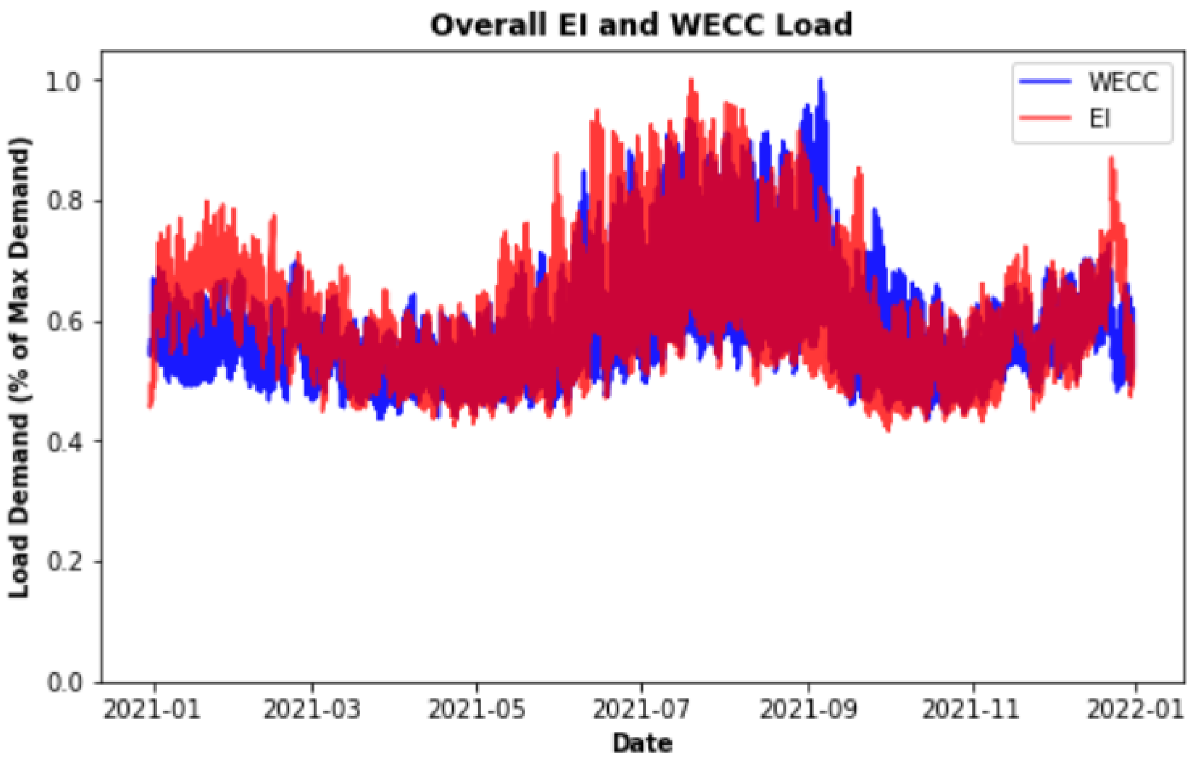

Figure 2-6 Normalized Load in the EI and WECC Grids

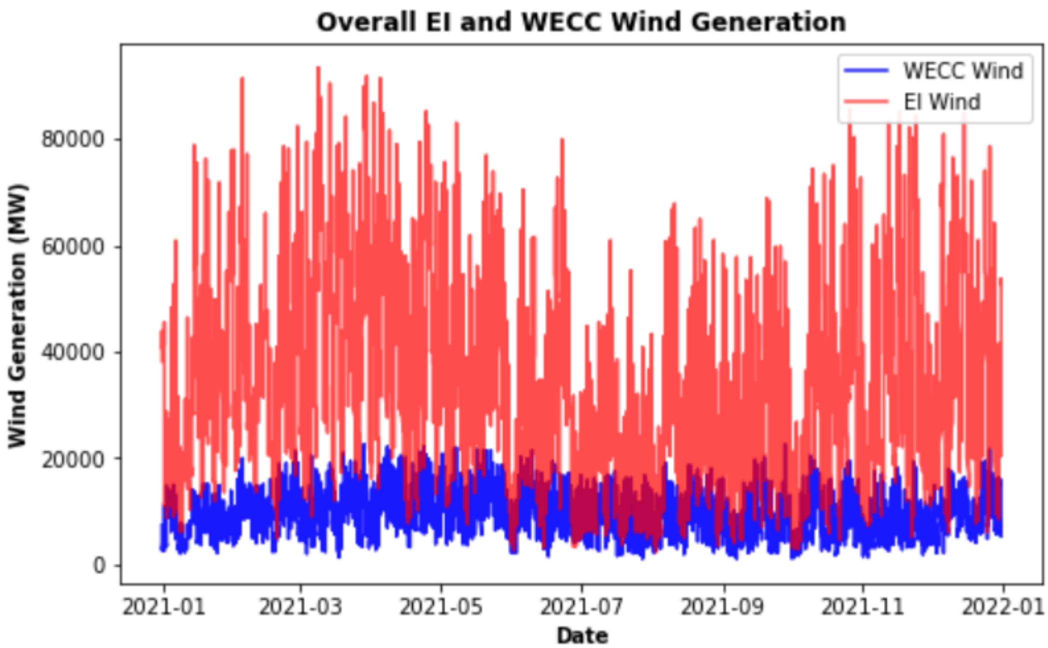

Figure 2-7 Wind Generation in the EI and WECC Grids

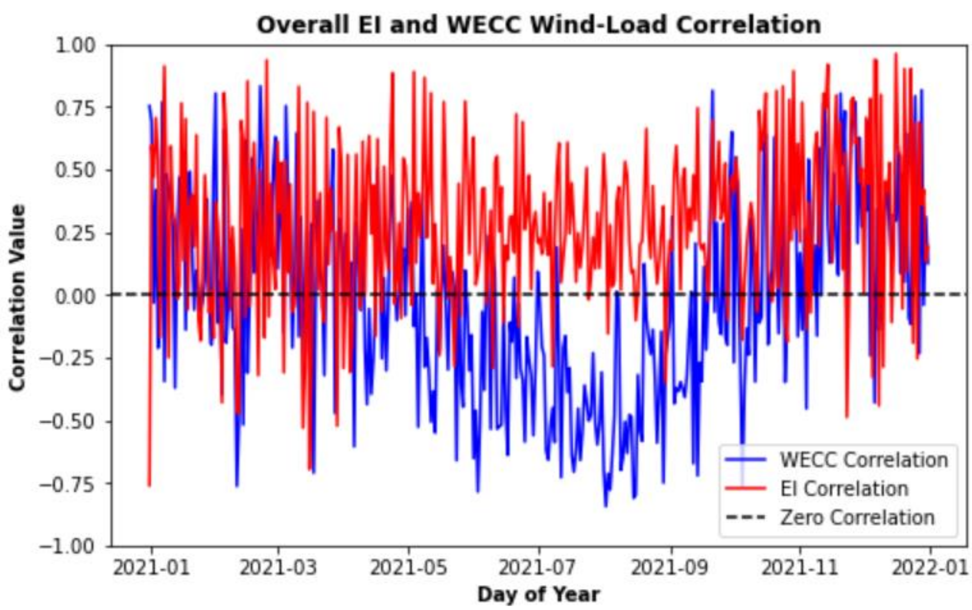

Figure 2-8 Load-Wind Correlation in the EI and WECC Grid

The difference in correlation values illustrates an advantage of a joined grid. The high correlation between load and wind generation signifies that as load increases, so does the amount of wind generation—a highly desirable characteristic in sustainability considerations. When correlation is negative, it means that either there is high load and little wind generation, or high wind generation and little load—neither of which are beneficial for sustainability. Since wind energy is non-dispatchable and cannot be stored in large amounts, it must be used as it is produced. Therefore, if

wind energy is generated in large quantities during times of high power demand, it can be utilized, reducing the need for less clean or non-renewable energy sources.
As seen in Figure 2-8, there are times when the EI or when the WECC have low correlation between load and wind generation. In a connected grid, these periods would more greatly allocate renewable resources or mitigate gaps in renewable energy production.

2.8 Studied Weather Scenarios

Different weather scenarios can change the flow between the EI and the WECC, especially if renewables are more available on one side. Using historical weather data, we can create realistic weather scenarios to simulate. Some of these scenarios of interest include a day where there's a high amount of wind and solar generation, a day with a low amount, and most interestingly – days where there is a large difference across the seam of the two grids. The purpose of simulating past scenarios is not intended to recreate the grid at the moment in the past, but rather to study future events. If a weather event has already occurred, it is likely to happen again at some point in the future. The following table shows the weather scenarios that were studied:

Table 2-1 Studied Weather Scenarios

Scenario Number	Description	Date
W1	The date with the highest East to West renewable difference in an hour	12/5/2013 (max occurs at 3am with 41.5 GWh difference)
W2	The date with the highest West to East renewable difference in an hour	12/19/2008 (max occurs at 3pm with a 10.5 GWh difference)
W3	The highest East to West renewable difference in a 24-hour period	3/9/1998 (827 GWh in total for the day)
W4	The highest West to East renewable difference in a 24-hour period	12/1/2013 (119 GWh in total for the day)
W5	The date with the highest overall renewable generation in an hour	4/15/2008 (66 GWh at 1pm)
W6	The date with the lowest overall renewable generation in an hour	10/12/2005 (1.4 GWh at 8pm)
W7	A date with average renewable generation	11/7/1976 (634 GWh in total for the day)

These interesting weather scenarios can be visualized with the following graphs to show the renewable output for the date studied.

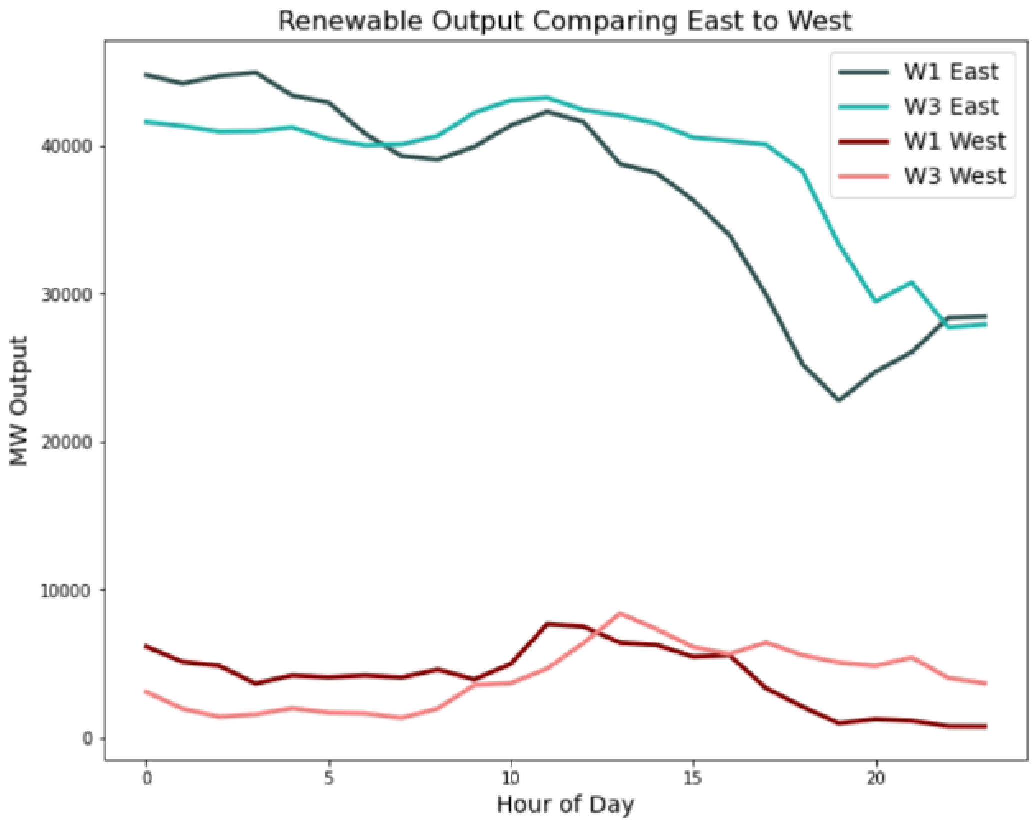

Figure 2-9 Weather Scenarios

2.9 Studied Load Scenarios

The case provides the maximum load value at each bus. Different regions are also provided in the case. Using FERC 714 load data, we can study the load-use patterns in many regions and normalize the data. FERC Form No. 714 is a collection of data from the United States electric utility balancing authorities and planning areas. The goal is to use the latest trends and scale the maximum load based on their region. The load can be studied on an hourly basis for up to a year.

Case Regions
NPCC
PJM
MISO
MRO
CAISO
SERC
SPP
FRCC
Northwest
Southwest

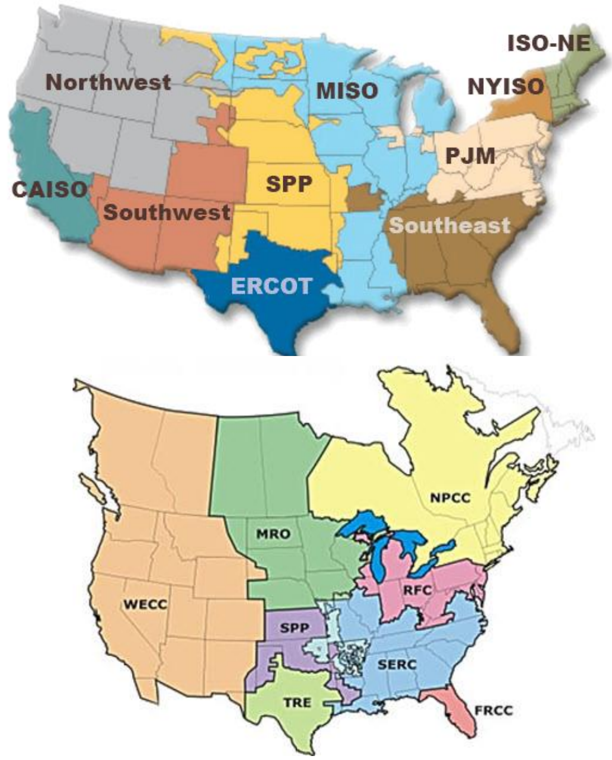

Figure 2-10 Reference Maps

Sometimes an entire area is posted (like SPP, MISO, CAISO for example) on the public submission portal. Other times they are not (such as FRCC, WECC, SERC). In this instance, the data of utilities or co-ops in the region can be summed and then normalized. For example, FRCC is comprised of Florida Power and light, Duke Energy Florida, Orlando Utilities Commission, Tampa Electric Company, and so on. If the load is not tailored perfectly for the region in question, it does not affect the analysis, as the load data is normalized and simply utilized to study the load demand changes in a general area on a certain date. The studied FERC 714 load data simply provides trends.

When normalizing the data, all values in the dataset are placed between 0 and 1 (with 1 being the peak that region experienced in that year and each other value being a ratio based on the peak). This allows a cross-comparison between areas that experience different levels of demand. This is also useful for the study because the max load value is already provided in the case. Having a data array that is a ratio of the maximum demand allows for time-series studies with realistic load-use patterns. The following images show examples of how the load is normalized in different regions on different days.

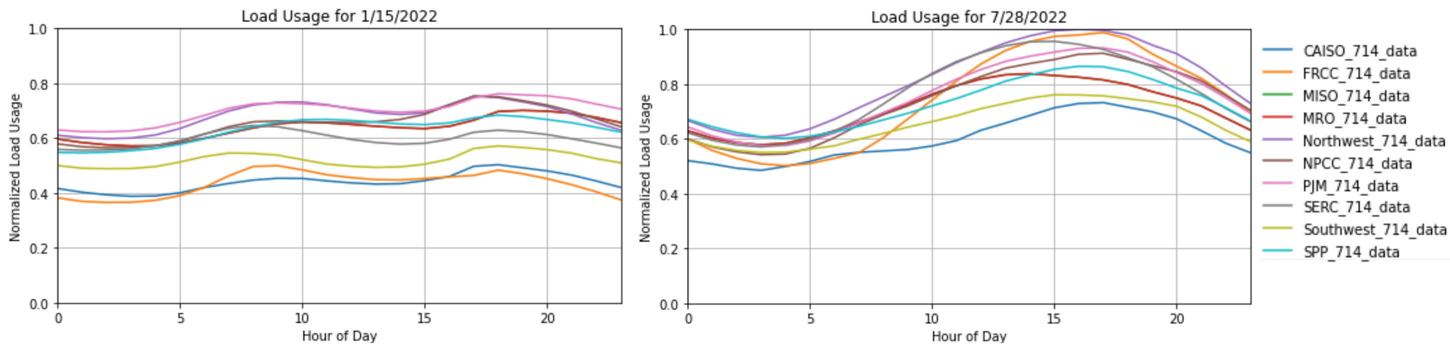

Figure 2-11 Load scenarios

Several days of interest were included in this study when considering load scenarios.

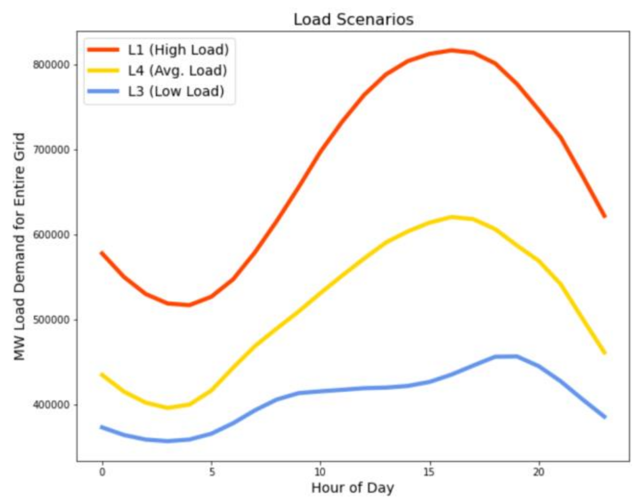

Figure 2-12 Overall high, average and low load scenarios

2.10 Cost Curves

This study requires that the cost of generators is known. Using EIA data, each generator is randomly placed within a realistic cost value for its fuel type. Given the considerable quantity of generators, the Law of Averages holds true. Wind and solar generation are assumed to have no fuel cost. The following table shows the EIA data that provides the average cost in US dollars $ for generators in the last decade by year and fuel type.

Year	Total			
	Nuclear	Fossil Steam	Hydro-electric	Gas Turbine and Small Scale
2011	24.70	35.09	8.88	44.54
2012	27.42	37.20	11.34	35.67
2013	27.29	37.92	10.88	37.92
2014	26.79	39.04	11.90	42.60
2015	25.71	37.26	13.42	33.24
2016	25.36	36.08	10.98	30.19
2017	24.38	35.41	10.29	31.76
2018	23.86	35.86	10.65	32.43
2019	23.73	36.66	10.80	28.33
2020	21.92	34.86	12.71	24.55
2021	22.74	35.66	12.30	30.18

Figure 2-13 EIA cost curve data

3. Using Power Flow Application Capabilities to Visualize and Analyze US Energy Information Administration Generation Data

3.1 Introduction

The US Energy Information Administration Form 860 (EIA-860) [46] dataset serves as a valuable and publicly accessible resource for professionals and researchers in the power and energy sector, offering extensive information on electrical generators in the United States with a capacity of one megawatt or more. This dataset, released annually and available on the EIA website, includes details on traditional fossil fuel facilities as well as wind and solar power capacities. The section presents a method for translating EIA-860 data into a generic power flow simulator format for visualization and analysis. Utilizing this data and weather measurements, the section demonstrates scenarios like the impact of a partial solar eclipse in October 2023. The power system model follows a copper plate model [47], allowing generation to flow with zero impedance. Various power flow structures, such as buses, substations, and weather-dependent models, are used, and advanced visualization techniques like geographic data views and contour mapping are explored, enhancing the understanding of complex generation data. The coupling of weather information to power flow studies is highlighted as a useful tool for calculating expected wind and solar generation outputs based on weather conditions.

3.2 Methodology

This section details the creation of a power flow case using only the EIA-860 dataset for enhanced power flow visualization and analysis. Employing a copper plate approach, all generators connect directly to a slack bus using low-impedance lines, simulating the absence of a real transmission grid. The process involves defining basic data structures like areas, buses, generators, substations, and zones. Schedule 1 represents utility data, introducing a fictitious slack area and a super area covering all regions for the slack area. Schedule 2, representing plant data, employs substations and buses, with a single bus assigned to each substation. Hypothetical slack bus and slack substation are introduced. Generator models from Schedule 3 are allocated for each dataset generator, including a fictional slack bus generator. The slack bus, vital for maintaining reactive and real power equilibrium, is crucial in this case. Optional generators, based on proposed and retired generation data, can be included in scenarios, allowing customization for different cases, as further explained in subsection E.

3.2.1 Using EIA-860 Utility Data to Define Areas

The EIA-860 Schedule 1 data is utilized to establish areas in the grid being constructed. The "Utility ID" column serves as the area number, with the "Utility Name" as the area name. Additional fields like "Street Address," "City," or "Zip" can be stored but are non-essential for this case. This information is then transferred to the "Area" component in the power flow package. To enhance organization and hierarchy, all areas are controlled by a super area. After adding the super area, it is crucial to create an area designated for the slack bus, with a distinct area number like 999999 to prevent overlap with other bus numbers.

3.2.2 Using EIA-860 Data to Create Substations

The EIA-860 Schedule 2 data is used to create substations and buses in the case model (buses will be created in the following section). The “Plant Code” becomes the substation number, and the “Plant Name” becomes the substation name. The geographic coordinates need to be saved. As mentioned previously, other fields can be stored, but are not necessary. This can now be sent to **Substation** in the power flow package.

Additionally, a slack bus substation will need to be created, named, and numbered (with a large value). This can have its latitude and longitude left blank since it represents a virtual location.

3.2.3 Creating Zones

Zones provide another way to group the information in the created case study. The Federal Information Processing System (FIPS) code is a two-digit number that uniquely identifies states and territories in the United States [48]. The FIPS code can be the zone number, while the state’s postal abbreviation can be the zone name.

3.2.4 Using EIA-860 Plant Data to Create Buses

Using EIA-860 Schedule 2 data, buses are generated, distinct from substations. Utility codes set the bus's area, state fields determine the zone, and the Plant Code serves as both bus and substation number. "Utility ID" designates the area number, "Plant Code" as the bus number, and "Plant Name" as the bus name. Any additional data from Schedule 2 can be stored or deleted. After adding this data to the Bus section in the power flow package, a slack bus must be created, with a large number, a proper name, and matching slack area, zone, and substation numbers. Proper specification of the slack bus in the case is crucial for the power flow package to operate correctly.

3.2.5 Using EIA-860 Generator Data to Create the Generators

In EIA-860 Schedule 3_1 data, there are three sheets: "Operable" lists generators in service, "Proposed" includes upcoming generators, and "Retired and Canceled" covers out-of-service generators. For the "Operable" sheet, generators are sorted based on the "Synchronized to the Transmission Grid" column, including only those with "Y" or "X" (synchronized). "Proposed" and/or "Retired and Canceled" sheets can be copied with aligned column names by sorting generators based on the "Plant Code," which serves as the bus number. To accommodate power flow package requirements, a new primary label is created by combining the bus number and "Generator ID" for uniquely identifying generators.

$$= \text{CONCATENATE}(BusNumber, "\_", GenID)$$

Another basic ID will need to be created by taking the mod of the row number to have a unique value for each generator (once again changing the column name as necessary).

$$= \text{MOD}(\text{ROW}(BusNumber), 100)$$

Other changes need to be made to the EIA-860 Schedule 3 data to make the generators ready to be added to the case model that is being created. The following table shows that changes that need to be made:

Table 3-1 Changes made to Generator Spreadsheet

New Col?	Original Name	New Name	Value
N	Prime Mover	Unit Type Code	original values
N	Energy Source 1	Fuel Type Code	original values
N	Minimum Load (MW)	Min MW	original values
Y		Max MW	*See below equation
N	Status	Status	Yes
Y		Set Volt	1.0
Y		Gen MW set point	0
Y		Max Mvar	0
Y		Min Mvar	0
Y		Gen Mvar set point	0
Y		AGC	No
Y		AVR	No

New Col? indicates whether a new column needs to be created (Y) or if an existing column is used (N). **Original Name** shows what the column is currently called in the spreadsheet. **New Name** is what the column is renamed to. **Value** is the data held in each column, which can be kept the same or changed as indicated.

The "Prime Mover" is the machine that converts the energy into electricity, which can be described as a unit type in a power flow package. The "Energy Source" is where the energy is being obtained (like coal, gas, wind, solar, etc.) and can also be known as the fuel type. A generator has a minimum amount of power it can output as indicated by the "Minimum Load". Each also have a maximum power output. The EIA-860 data has two fields that could be used: either the “Summer Capacity (MW)” or the “Winter Capacity (MW)”. Whether to use one or the other depends on the application. One approach, particularly when studying wind and solar outputs modified by the actual weather is to just pick the largest value. For this approach, the entries in that column need to be set to the maximum of those two fields (change the arguments in the equation as necessary):

$$* = \mathrm{MAX}(SummerCapacity, WinterCapacity)$$

"Status" refers to whether or not the generator is in use, so therefore they are all set to yes. The "Set Volt" is the value at which the voltage is set, and since per unit values are used, all generators are set to 1.0. The "Gen MW Set Point" is initially set to 0 since all generators are started out as not outputting any power, but this can be changed based on the application. The "Min Mvar" and "Max Mvar" specify the allowable reactive power output from the generator and since nothing is flowing, it is set to zero (as well as the "Gen Mvar set point". "AGC", which stands for automatic generation control, is set to "No" so that the generator can be set on manual control. "AVR", which stands for automatic voltage regulation, is also set to zero since the voltage regulation plays a role in reactive power regulation and that's already been set to zero.
At this point, the generator data is ready to be loaded in the power flow package for the **Generators** section.

3.2.6 Creating Lines to Connect the Buses

The final step involves establishing low-impedance lines to connect all buses to the slack bus, forming a copper plate model for a power flow solution. Despite the absence of a real transmission grid, this allows for operational simulations. Creating branches involves assigning the sending bus number in the proposed model to all implemented bus numbers, directing them to the previously created slack bus. Impedances and line limits are set to zero or very low values. Note that zero capacity lines are suitable for models allowing line limit capacities in soft constraints or assuming zero as infinity; otherwise, line capacities should be set to a large value. With these elements in place, operation simulations can be run, starting with all generator outputs at zero but adjustable to any value, especially their maximum.

3.2.7 Setting up Power Flow Weather Models

It is helpful to utilize the EIA-860 information given in the Schedule 3_2 wind and 3_3 solar files (although these only have data for the operating generators, not the proposed ones) when creating new power flow weather (PFW) models. These files should have the same unique IDs as before (which is a combination of the plant code and generator ID).

For wind turbines, the important fields are the “Wind Quality Class” and the “Turbine Hub Height”. These should be stored in columns and then when input into the power flow package, organized based on the wind quality class. New PFW models should be inserted for these generators based on the wind quality class.

For the solar models, the important columns are “Single-Axis Tracking” (usually a change in tilt), “Dual-Axis Tracking” (which rotate horizontally and change tilt), “Fixed Tilt” (which is a mounted panel with fixed angle and orientation), “Azimuth Angle” (which is the angle from due north), and “Tilt Angle” (which is the angle a panel is tilted up from being parallel to Earth). A new field will need to be created that numerically describes the tracking with an integer code. The data can be input into the power flow package now, creating new photovoltaic (PV) models for these generators. The tracking and angle data should be added to the PFW models.

Table 3-2 Solar Panel Parameters

Tracking Type	Integer Code
No Tracking	0
Single-Axis with Fixed Tilt Angle	1
Single-Axis with Fixed Azimuth	2
Dual-Axis	3

3.3 Simulations and Visualizations

Utilizing the power flow case created based on the EIA-860 data a wide variety of simulations can be performed including the direct inclusion of weather information as described in [75]. Using the actual generator data is useful for several studies such as emission calculations [49].

To enhance the visual analysis of power grid operations, there are many useful tools to employ, such as the Geographical Data View (GDV) [65], contour mapping [67], and broad visualizations

of the transmission grid [71]. These tools are used to increase situational awareness of the grids across diverse scenarios and temporal spans [50], [51].
Figure 3-1 shows the retired generators in the 2023 EIA-860 form as ovals, where the size is proportional to the generator capacity and the color refers to the fuel type. In this figure, red refers to nuclear units, black refers to coal, brown refers to natural gas, green to wind, blue to hydro, yellow to solar, and magenta to energy storage. Figures 3-2 to 3-4 use the same color code and show renewable generators at the end of 2022.

Figure 3-1 Retired Generators in 2023

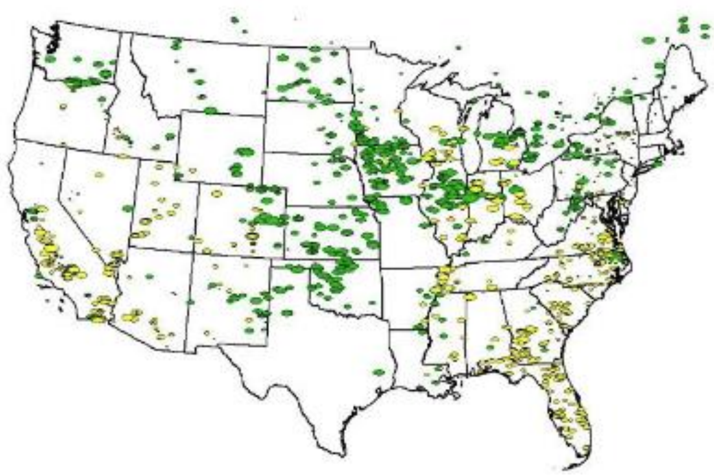

Figure 3-2 Overall renewable generator maximum capacity in the US

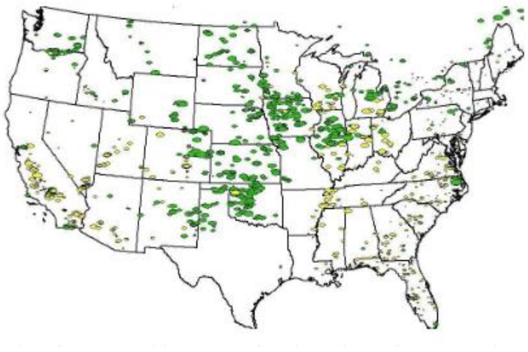

Figure 3-3 High renewable generation at 1 pm of April 15, 2008 (UTC)

Figure 3-4 Low renewable generation at 8 pm of October 12, 2005 (UTC)

Having all weather measurements can help in visualizing interesting scenarios, such as the solar eclipse than happened on October 14, 2023. Figure 3-5 visualizes the annular eclipse by contours where the color shows the percentage of solar radiation, with dark blue referring to 0% and dark red referring to 100%. Figures 3-5 and 3-6 shows the impact of the eclipse on the solar generation.

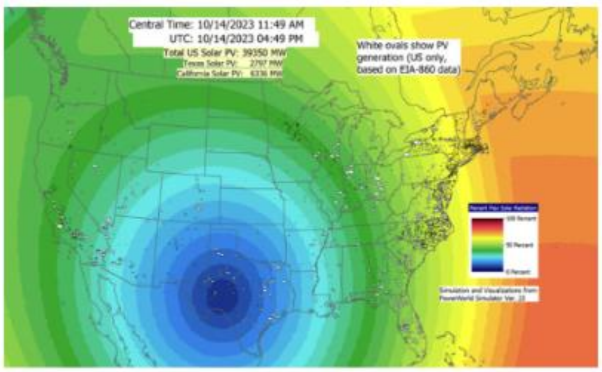

Figure 3-5 Solar Eclipse in the United States in 2023

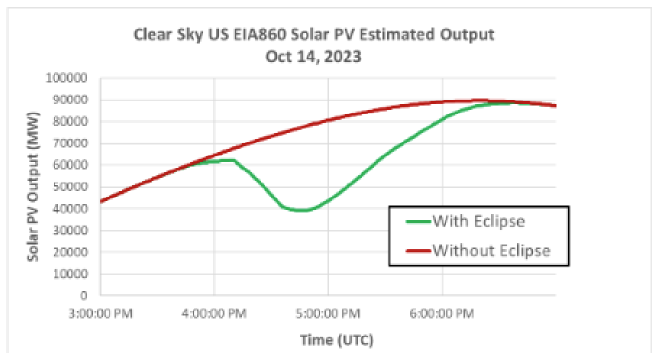

Figure 3-6 Potential "clear sky" impact of the eclipse on US generation

3.4 Conclusion

This section has utilized the EIA-860 data to develop a case study model for power flow analysis, and it has presented a series of informative visualizations. The EIA-860 dataset is a valuable resource and by following the procedure outlined in this section, others can replicate the model and further contribute to the knowledge of this field. The significance of data visualization enables effective communication to a broad audience and increases the situational awareness of a large grid.

4. Enhancing Power Flow Studies Through Representative Scenario Selection

4.1 Introduction

When conducting planning studies on the electrical grid, both load and renewable generation vary throughout the year. Typically, the grid witnesses its peak demand for load during the summer, with winter ranking second in terms of high demand. On the other hand, load is typically the lowest in the spring. The traditional approach to scenario usage involves utilizing the peak load scenarios in summer and winter, along with the low load scenarios in spring. However, this section departs from the traditional approach by introducing a novel method for identifying scenarios to be examined in planning studies, moving away from the conventional focus on specific operating points.

It is essential to align the level of realism with the specific objectives of the planning study when running power flow simulations. In the past, weather has only been modeled implicitly. This includes modeled load values, real and reactive power output, transmission line limits, transformer limits, and more.

Weather has long been affecting the grid and although the implicit methods were adequate in the past, weather data now needs to be modeled explicitly. The two primary reasons for this shift is a large increase in amount of renewable generators (specifically wind and solar) with the continued expansion of their capacity and the occurrence of extreme weather events with the need to plan for worst-case scenarios [52].

Incorporating weather data into power flow simulations is just one characteristic of the evolving landscape in grid modeling. Equally important is the inclusion of realistic load values and trends, reflecting the dynamic nature of power demand in various regions. The demand for electrical power exhibits fluctuations influenced by a multitude of factors. One of these is the significant impact that weather has on the load in an electrical grid [53]. For example, during the summer in southern areas of the United States, there is a surge in electricity due to air conditioning usage. On the other hand during the same season in more northern areas, there is less of a surge because it stays relatively cooler. Another factor is the economic or commercial operations in an area. Industrial hubs exhibit distinct load profiles compared to residential areas. Understanding these nuances is vital to understanding the infrastructure [54].

Another crucial consideration is the utilization of actual data in these simulations. For the most authentic and reliable outcomes, it is imperative to employ genuine, real-world data. The research in [55]. also uses publicly available data to create synthetic time series of load at buses in a synthetic system. Additionally, [56] also develops scenarios by determining generation and load characteristics. Similar to these sections, this section also uses publicly available data to create scenarios; the difference here is that a clustering method is used to reduce the number of scenarios to be tested.

The research in [57] highlights the importance of clustering in reducing the number of scenarios to aid with computational complexity. The research in [58] and [59] also mention the benefits in the context energy and power, with the both considering unit commitment and the second highlighting the wind generation scenarios. [60] compares several methods to reduce the amount of tested scenarios and does a case study on a 24-bus case.

The work in this section presents a new method of three-dimensional clustering to study wind generation, solar generation, and load demand and choose the scenarios for testing. Load and weather-related time series are created as explained in chapter 2. This is different from other clustering methods, as it will choose the central scenario as well as the outlier. There is a great need to study outliers because those scenarios are when the grid is under stress (the importance of outlying data in [61]). The scenario selection method will be demonstrated using Texas data for its renewable generation capacity and its unique grid situation. This method is aimed at enhancing the resiliency of the power grid by offering scenarios for testing in order to plan and prepare for all possible events. Most of this chapter is based on what is published in [62].

4.2 The Problem

A common challenge faced with such vast datasets is that there is an abundance of information, but difficulty in finding meaningful representation of what is presented. For instance, studying a single year on an hourly basis yields 8,760 data points. When encompassing multiple factors for study, the number of potential scenarios grows exponentially. To illustrate, the combination of a year's worth studying hourly of two dimensional data leads to 80 million permutations for analysis. This complexity is compounded by the inclusion of decades-long datasets, dating back to 1972 for weather data and the early 2000s for load data.

This sheer magnitude of data points presents a challenge, due to computational limitations and resource constraints (time, amount of people working, available computers, etc.). The need for a discerning approach in data point selection arises from the practical need to optimize the testing process and work efficiently, ensuring a focused and meaningful analysis within the constraints of available resources.

In the past, the load values at the summer peak are often used when running these simulations, as well as scenarios in the winter or the spring for lows. Although it is important to plan for demand that would cause the grid stress, it is necessary to consider weather, as the dependency on renewable generation grows. This approach also couples wind and solar weather data with the load to get accurate, realistic scenarios to run as operating points.

4.3 Proposed Solution

K-Means Clustering, as explained in [63], is a popular unsupervised machine learning algorithm that aims to group data points together based on their characteristics. The solution proposed here uses this K-Means clustering method in three dimensions to aid in data point selection. This method uses three characteristics:

1. Calculated Output of Solar Generation
2. Calculated Output of Wind Generation
3. Total Load

For every *n*-clusters, there will be *2n*-scenarios for testing. This is because two points will be chosen from each cluster - the point closest to the centroid and the point furthest away from the centroid ('closeness' refers to the Euclidian distance in three dimensions). This allows for both average data to be studied as well as the outlying data.

The inclusion of both average data points and outlying data points is crucial for obtaining a comprehensive understanding of the underlying patterns and dynamics within a dataset. Analyzing average points allows for grasping central tendencies and typical behaviors exhibited by the features studied in the power grid, therefore providing a baseline understanding of the general trends. On the other hand, the examination of outlying points is equally essential, as these instances often carry valuable information about anomalous conditions. The study of outliers aids in identifying potential scenarios when the grid faces periods of stress. Typically if a grid can withstand the troublesome times, it can withstand baseline operations. Thus, the dual exploration of averages and outliers enhances the robustness of this analysis, offering a more complete perspective.

When choosing the amount of scenarios that should be studied, it is important to also consider the data properties. As explained in [64], the elbow method is a way to determine the minimum number of clusters a dataset should have. When calculating the variance that is preserved for an increasing number of clusters, there is a point in the curve, known as an 'elbow' or 'knee' that is a leveling off-point. The integer that corresponds to this change is often considered the minimum number of clusters that is ideal for a dataset. It is important to note that the 'elbow' point does not exclude the need to have more clusters, as the amount that is chosen for each study must reflect the necessary level of depth needed to understand a situation.

Since this method gives a data point that is only a singular time snapshot, it would also be beneficial to conduct studies based on the entire day for the selected time point. This is particularly helpful when trying to study how the grid morphs or changes based on past and future conditions. This is only something to keep in mind based on the studies that are being performed.

4.4 Case Study and Results

To effectively demonstrate the method proposed in this section, a case study will be showcased utilizing Texas data. The choice of Texas is particularly apt, given its distinctive characteristics as both a self-contained grid with large renewable penetration and a singular state. This allows us to have consistent data from one entity, which significantly reduces the amount of discrepancies in the time-series data used.

For this case study, data from 2021 was used since that is the latest publicly available load data provided by FERC. Although for each study, this can be tailored as necessary, even concatenating data sets that span multiple years or decades. A single year is easier to demonstrate and visualize for the purposes of this section.

This raw load and generation data can be seen in Figure 4-1. Something to note is that solar generation is relatively low compared the wind generation and total load, so pre-processing has to be done on the raw data so that the algorithm takes the solar generation input into account (such as normalizing the data so that all variables are given equal consideration in the clustering process). The different scales of solar, wind, and load can be seen in Figure 4-4, as each box and whisker plot shows the MW values from the dataset.

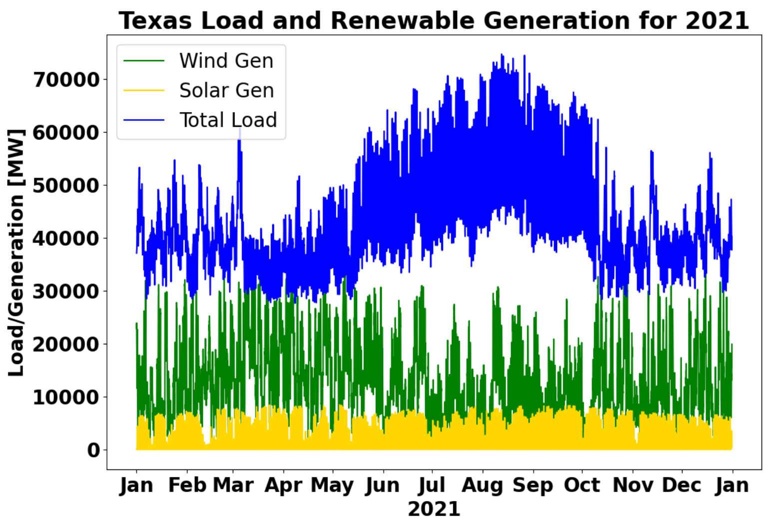

Figure 4-1 Texas Load and Generation 2021

From here, the elbow method is used. The following result can be seen in Figure 4-2, which leads to a selection of 4 clusters to capture the data, as that is the integer that has the sharpest angle change.

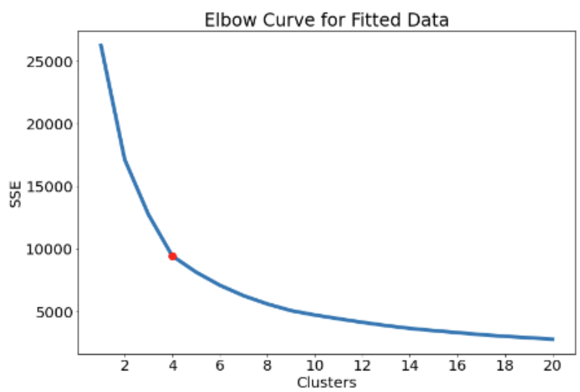

Figure 4-2 Elbow Plot for Texas 2021 Data

As seen in Figure 4-3, the data is split into four base clusters. The characteristics of these clusters can be generalized in Table 4-1.

Table 4-1 Cluster Characteristics

Color	Solar	Wind	Load
Green	Low	Low	Low
Gray	Low	High	Low
Pink	Low	High	High
Teal	High	High	Varies

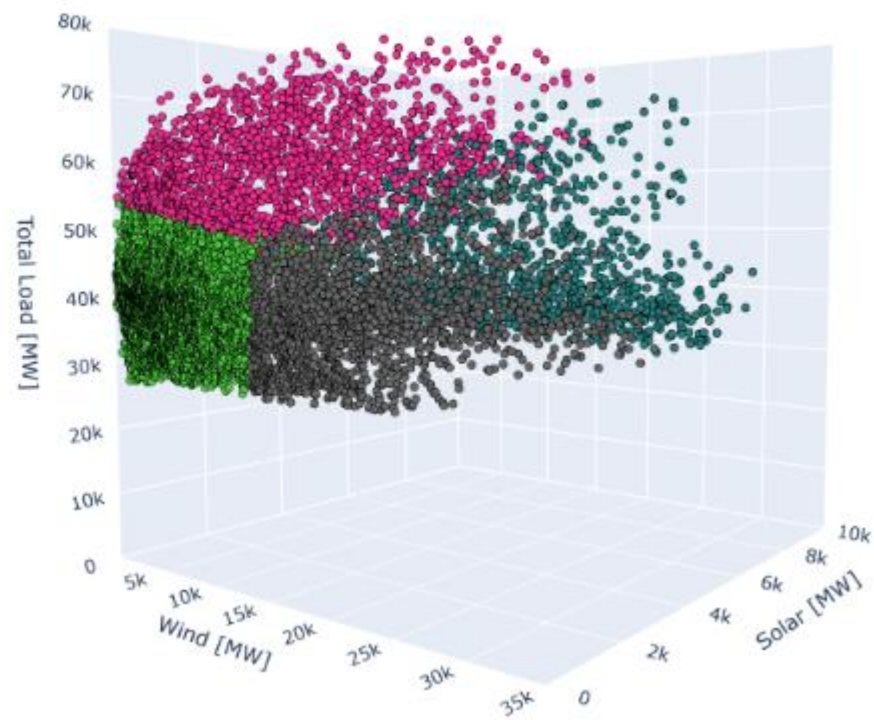

Figure 4-3 3D Clustering of Data

It is important to mention that there is only one cluster for the case of high solar generation independent of total load and wind generation due to the fact that there is a relatively low density of data points with high solar generation. This is because solar has many points around zero because of the nature of solar generation at night. Wind varies around much, going from 0 MW to upwards of around 30,000 MW. Load never dips below 25,000 MW for any time points, so this also is represented in the clusters.

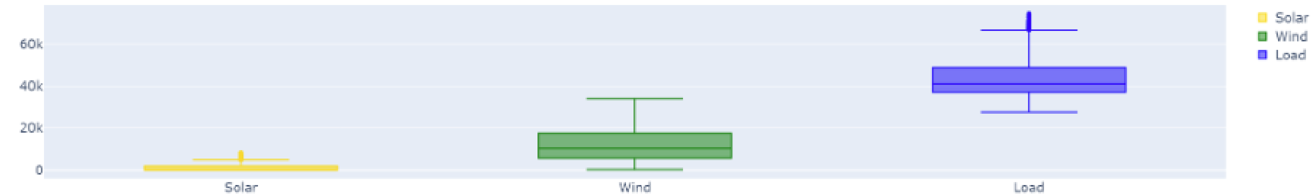

Figure 4-4 Box & Whisker Plot for 3 variables

Finally looking at Figure 4-5 one can see X's that mark the data that is closest to the centroids of the clusters. On the other hand, the diamonds that mark the points that are furthest from their respective centroid. These furthest points are the extremes of typical operating conditions that could potentially affect the stability of the grid. These selected points reduce the amount of data points that we have to simulate, while still having a diverse range of scenarios.

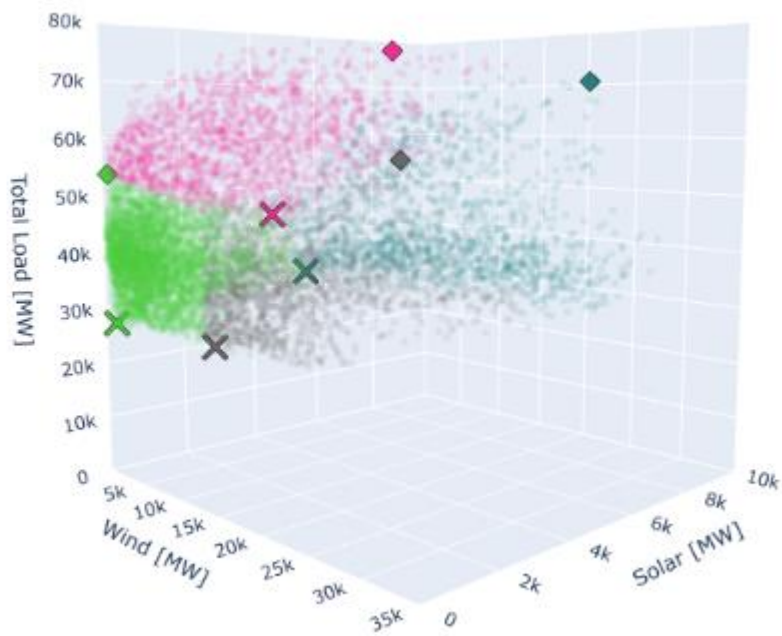

Figure 4-5 Closest and Furthest Data Points

Although the centroid in the green cluster is close to the pink cluster, it is still a scenario that merits testing, especially since no other points in its proximity are being tested. Both the pink and teal outlier point will merit interesting scenarios that would not normally be tested when studying typical grid operations.

Figure 4-8 shows a bar plot of the different scenarios, with the centroid on top being directly compared to the outlier in each cluster. This allows one to visualize and easily understand what is being changed in each of the proposed scenarios.

Figure 4-6 Centroid vs. Outlier Bar Plot

4.5 Conclusion

This section showed an initial approach for scenario selection using K-Means clustering. It used both average points of operation that is often seen as well as the outliers that could potentially cause some stress in the grid. This method was demonstrated and visualized using data from 2021 in the state of Texas.

The results help finding interesting scenarios based on using clusters discovered in the raw data. Future work includes adding more dimensions into this study, such as time itself, region, or more weather factors. Adding time as a dimension for clustering could potentially help in finding scenarios for different seasons, adding regions could help in identifying areas of weakness in a grid, and adding more weather measurements would make the model more accurate. Validation could also be done by studying these selected points in PowerWorld and seeing the effects these scenarios have on the grid.

References

[1] 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, 2019.

[2] N. Cohn, S. B. Biddle, R. G. Lex, E. H. Preston, C. W. Ross, and D. R. Whitten, “On-line computer applications in the electric power industry,” Proceedings of the IEEE, vol. 58, no. 1, pp. 78–87, 1970.

[3] I. Toljan, E. Banovac, and S. Teˇsnjak, “Development of UCTE and Reconnection of 1st and 2nd Pan-European Synchronous Zones,” in Proceedings of the 4th IASME/WSEAS international conference on Energy & environment. World Scientific and Engineering Academy and Society (WSEAS), 2009, pp. 419–424.

[4] H. Breulmann, E. Grebe, M. L¨osing et al., “Analysis and damping of inter-area oscillations in the UCTE/CENTREL power system,” CIGRE Session, no. 38-113, 2000.

[5] M. Luther, I. Biernacka, D. Preotescu et al., “Feasibility Aspects of a Synchronous Coupling of the IPS/UPS with the UCTE,” CIGRE Session, no. C1 204, 2010.

[6] S. S. Lee, J. K. Park, and S. I. Moon, “Power system interconnection scenario and analysis between Korean peninsula and Japan,” in Fig. 5. Total MW Flow on the AC Ties from West to East 2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No.03CH37491), vol. 3, 2003, pp. 1455–1460 Vol. 3.

[7] Z. Xu, H. Dong, and H. Huang, “Debates on ultra-high-voltage synchronous power grid: the future super grid in China?” IET Generation, Transmission Distribution, vol. 9, no. 8, pp. 740–747, 2015.

[8] A. L. Figueroa-Acevedo, “Opportunities and benefits for increasing transmission capacity between the US eastern and western interconnections,” PhD Dissertation, Iowa State University, 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, pp. 513–521, 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, D. Osborn, J. Feltes, H. Kirkham, D. Duebner, and Z. Huang, “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, pp. 390–398, 2017.

[11] A. Bloom (National Renewable Energy Laboratory), “Interconnection Seams Study,” in TransGrid-X 2030 Symposium, Ames, Iowa, 2018.

[12] T.J. Overbye, K.S. Shetye, H. Li, W. Trinh, J. Wert, *Feasibility Assessment of Synchronous Operations of the North American Eastern and Western Interconnections*, PSERC Report S-92G, January 2021.

[13] T. 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,” 2022.

[14] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, “Grid structural characteristics as validation criteria for synthetic networks,” IEEE Transactions on Power Systems, vol. 32, no. 4, pp. 3258–3265, July 2017.

[15] https://electricgrids.engr.tamu.edu/
[16] “Current Weather and Wind Station Data”. [Online]. Available: https://aviationweather.gov/adds/dataserver current/current/metars.cache.csv
[17] “ERA5 hourly data on single levels from 1940 to present”. [Online]. Available https://cds.climate.copernicus.eu/cdsapp!/dataset/reanalysis-era5-single-levels?tab=form
[18] F. Safdarian, M. Stevens, J. Snodgrass, T. J. Overbye, “Detailed Hourly Weather Measurements for Power System Applications”, 2024 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, Feb. 2024
[19] https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-no-714-annual-electric/data
[20] I. Graabak and M. Korp˚as, “Variability characteristics of European wind and solar power resources—a review, Energies, vol. 9, no. 6, p. 449 2016
[21] 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
[22] M. A. Abdullah, A. Agalgaonkar, and K. M. Muttaqi, “Probabilistic load flow incorporating correlation between time-varying electricity demand and renewable power generation,” Renewable energy, vol. 55, pp. 532–543, 2013.
[23] X. Xu, Z. Yan, M. Shahidehpour, H. Wang, and S. Chen, “Power system voltage stability evaluation considering renewable energy with correlated variabilities,” IEEE Transactions on Power Systems, vol. 33, no. 3, pp. 3236–3245, 2017.
[24] 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.
[25] “Current Weather and Wind Station Data”. [Online]. Available: https://aviationweather.gov/adds/dataserve current/current/metars.cache.csv
[26] “ERA5 hourly data on single levels from 1940 to present”. [Online]. Available https://cds.climate.copernicus.eu/cdsapp!/dataset/reanalysis-era5-single-levels?tab=form
[27] “Weather Station Identifiers”. [Online]. Available: http://www.weathergraphics.com/identifiers/
[28] (2019) “U.S. Energy Information Administration (EIA)”. [Online]. Available: https://www.eia.gov/electricity/data/eia860/
[25] G. M. Masters, Renewable and efficient electric power systems. John Wiley & Sons, 2013.
[30] 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.
[31] 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.
[32] 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.
[33] A. De Sa and S. Al Zubaidy, “Gas turbine performance at varying ambient temperature,” Applied Thermal Engineering, vol. 31, no. 14-15, pp. 2735–2739, 2011.

[34] J. L. Wert, T. Chen, F. Safdarian, J. Snodgrass, and T. J. Overbye, “Calculation and Validation of Weather-Informed Renewable Generator Capacities in the Identification of Renewable Resource Droughts,” in IEEE PowerTech 2023, 2023.
[35] 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.
[36] 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
[37] “Correlation coefficient: Simple definition, formula, easy steps,” https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/, accessed: 2023-11-08.
[38] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134
[39] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers,P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R´ıo, M. Wiebe, P. Peterson, P. Gerard-Marchant, ´ K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “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
[40] 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
[41] J. D. Hunter, “Matplotlib: A 2d graphics environment,” Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.
[42] 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
[43] S. E. I. Association, “Solar state by state,” Online, 2023, accessed: 2023-11-11. [Online]. Available: https://www.seia.org/states-map
[44] L. Mart´ın, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, “Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning,” Solar Energy, vol. 84, no. 10, pp. 1772–1781, 2010.
[45] F. E. R. Commission, “Form no. 714 - annual electric balancing authority area and planning area report,” December 2022.
[46] (2019) “U.S. Energy Information Administration (EIA)”. [Online]. Available: https://www.eia.gov/electricity/data/eia860/
[47] 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.
[48] “Appendix d - usps state abbreviations and fips codes,” https://www.bls.gov/respondents/mwr/electronic-data- interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm, accessed: 2023-11-13.

[49] J. L. Wert, F. Safdarian, D. Wallison, J. K. Jung, Y. Liu, T. J. Overbye, and Y. Xu, "Spatiotemporal operational emissions associated with light-, medium-, and heavy-duty transportation electrification," IEEE Transactions on Transportation Electrification, 2023.
[50] T. J. Overbye, K. S. Shetye, J. Wert, W. Trinh, A. Birchfield, T. Rolstad, and J. D. Weber, "Techniques for maintaining situational awareness during large-scale electric grid simulations," in 2021 IEEE Power and Energy Conference at Illinois (PECI). IEEE, 2021, pp. 1–8.
[51] J. L. Wert, F. Safdarian, T. J. Overbye, and D. J. Morrow, "Case study on design considerations for wide-area transmission grid operation visual storytelling," in 2022 IEEE Kansas Power and Energy Conference (KPEC). IEEE, 2022, pp. 1–6
[52] I. Staffell and S. Pfenninger, "The increasing impact of weather on electricity supply and demand," Energy, vol. 145, pp. 65–78, 2018.
[53] T. Hong, W.-K. Chang, and H.-W. Lin, "A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data," Applied energy, vol. 111, pp. 333–350, 2013.
[54] E. Veldman, M. Gibescu, H. J. Slootweg, and W. L. Kling, "Scenario-based modelling of future residential electricity demands and assessing their impact on distribution grids," Energy policy, vol. 56, pp. 233–247, 2013.
[55] H. Li, J. H. Yeo, A. L. Bornsheuer, and T. J. Overbye, "The creation and validation of load time series for synthetic electric power systems," IEEE Transactions on Power Systems, vol. 36, no. 2, pp. 961–969, 2021.
[56] H. Li, J. H. Yeo, J. L. Wert, and T. J. Overbye, "Steady-state scenario development for synthetic transmission systems," in 2020 IEEE Texas Power and Energy Conference (TPEC). IEEE, 2020, pp. 1–6.
[57] J. Hu and H. Li, "A new clustering approach for scenario reduction in multi-stochastic variable programming," IEEE Transactions on Power Systems, vol. 34, no. 5, pp. 3813–3825, 2019.
[58] H. Keko and V. Miranda, "Impact of clustering-based scenario reduction on the perception of risk in unit commitment problem," in 2015 18th International Conference on Intelligent System Application to Power Systems (ISAP). IEEE, 2015, pp. 1–6.
[59] E. Du, N. Zhang, C. Kang, J. Bai, L. Cheng, and Y. Ding, "Impact of wind power scenario reduction techniques on stochastic unit commitment," in 2016 Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO). IEEE, 2016, pp. 202–210.
[60] Y. Dvorkin, Y. Wang, H. Pandzic, and D. Kirschen, "Comparison of scenario reduction techniques for the stochastic unit commitment," in 2014 IEEE PES General Meeting—Conference & Exposition. IEEE, 2014, pp. 1–5.
[61] J. W. Osborne and A. Overbay, "The power of outliers (and why researchers should always check for them)," Practical Assessment, Research, and Evaluation, vol. 9, no. 1, p. 6, 2004.
[62] J.S. Cook, M. Briones, F. Safdarian, and T.J. Overbye, "Enhancing Power Flow Studies Through Representative Scenario Selection", 2024 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, Feb. 2024.
[63] J. Ortega, N. Almanza-Ortega, A. Vega-Villalobos, R. Pazos-Rangel, J. C. Zavala-Diaz, and A. Mart ́ınez-Rebollar, The K-Means Algorithm Evolution, 04 2019.

[64] H. Humaira and R. Rasyidah, "Determining the appropriate cluster number using elbow method for k-means algorithm," in Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia, 2020.