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

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

***Abstract*—The electrical generator data across the United States with a minimum capacity of one megawatt is publicly accessible via the US Energy Information Administration form 860 (EIA-860). It is a valuable resource, holding significance for power and energy researchers as well as industry professionals. This paper describes the development and application of a power flow model utilizing the EIA-860 dataset. Beyond the model development process, this research emphasizes the critical role of effective data visualization techniques in communicating essential information about electricity generation. By showcasing the dataset's versatility and analytical power, this paper allows users to make informed decisions to advance power flow and generation adequacy studies.**

***Index Terms*—EIA-860, data visualization, US electrical generators, renewable energy, public data, power systems analysis, visualization, situational awareness**

I. Introduction

In today's era of data-driven decision-making, the US Energy Information Administration form 860 (EIA-860) [1] emerges as a valuable and publicly accessible resource for professionals in the power and energy sector, as well as researchers seeking insights into the dynamic landscape of electrical generation in the United States. This dataset encompasses a wealth of information on electrical generators having a capacity of one megawatt or more. This extensive amount of information makes the dataset a useful tool for those seeking to analyze and understand the nation's electrical power infrastructure.

The EIA-860 dataset provides details of the nation's generators including traditional fossil fuel facilities and the ever-expanding roster of wind and solar power capacities. This data is provided annually, with an early release in June for the prior year and the final release in September. It can be accessed and downloaded off the EIA website [1]. Additionally, form EIA-860M provides preliminary monthly updates.

The data is provided in several Schedules, accessed easily via Excel spreadsheets or pandas data frames [2]. Schedule 1 provides utility information, Schedule 2 provides plant information with geographic coordinates, Schedule 3 contains generator information with more details for wind turbines and solar cell models, Schedule 4 provides ownership information, and Schedule 6 contains information on boilers and associated equipment. Schedule 3, the one most heavily utilized in this paper, is separated into overall generator information, wind generation information, and solar generation information. Since the data is divided into utilities, plants, units, and by US states, there is a natural mapping of the data into power flow structures.

This paper presents a method of putting the EIA-860 data into a generic power flow simulator format to leverage power flow visualization and analysis capabilities. Next, using the EIA-860 generator data and weather measurements, interesting scenarios can be visualized, such as the partial solar eclipse that occurred on October 14, 2023. The power system model built in this paper, as defined in [3], is a copper plate model [4], meaning that generation can flow to any point in the grid with zero impedance. The procedure presented in the paper can be applied to many different power flow packages, with the results given here illustrated using PowerWorld Simulator [5]. The paper's EIA-860 based power flow cases can be accessed at [6].

In this paper, several standard power flow structures are used that are common to essentially all power flow studies [7]. This includes buses, substations, generators, areas, zones, and weather-dependent models (which are wind and solar generators). It is important to include zones for system hierarchy, as explained in [8].

Effective visualization of power flow applications plays a pivotal role in understanding the potentially complex aspects of generation data. The findings in [9] explain many of the techniques that aid in power flow visualizations. There are basic ways to do this, such as the single-line diagram, time series bar charts [10], or scatter-plot matrices. One of the most advanced techniques is the geographic data view (GDV), as explained in [11] and [12]. The other advanced technique is contour mapping, especially when studying voltages or frequencies across different locations [13] [14]. The methods introduced in [15] and [16] explain how the third dimension can be used to represent more data. Delaunay triangulation technique is also useful when understanding the grid's state overall [17]. Dynamic line flows can also be shown in a gird [18]. GreenGrid combines GDVs and transfers them into topological parameters. [19]. Similar to this technique is GreenCurve, which is a space-filling curve that displays multivariate graphs [20].

There is also the useful ability of coupling weather information directly to the power flow studies. The research in [21] explains how this is accomplished. By using weather data from the nearest weather station, the expected output of

 Accepted to TPEC 2024, Texas, February 2024.

wind and solar generation can be calculated (which directly depends on the weather). Additionally, contours of factors such as temperature, wind speed, and solar irradiance can be created for visualization purposes.

II. Methodology

This section explains how a power flow case can be created just utilizing the EIA-860 dataset with the goal of leveraging existing power flow visualization and analysis capabilities. As noted, the approach utilizes a copper plate approach in which there is not real transmission grid; rather all the generators are connected to the slack bus utilizing low-impedance lines. The first step in developing the power flow case is to define the basic data structures of areas, buses, generators, substations and zones. The EIA-860 Schedule 1 data, which includes utility data, is represented by "areas". A fictitious slack area is introduced, along with a super area that encompasses all regions for the slack area. The EIA-860 Schedule 2 data, which is plant data, is represented with substations and buses with a single bus assigned to each substation. Furthermore, a hypothetical "slack bus" and "slack substation" are introduced. Generator models are from the EIA-860 Schedule 3 data, allocating one generator model for each of the generators in the dataset. Additionally, a fictitious slack bus generator is introduced. The slack bus plays a crucial role in a power system as it maintains the equilibrium of reactive and real power through the emission and absorption processes, therefore it must be included in this case.

It is also worth noting that optional generators may be included in the scenarios, as there is data about the proposed and retired generation. Each procedure can be customized, such as one case with only the generation that is currently installed or another that includes the proposed generation. This is explained more in subsection E.

A. Using EIA-860 Utility Data to Define Areas

The EIA-860 Schedule 1 data is used to set up the areas in the grid that is being built. The column "Utility ID" in the EIA-860 Excel spreadsheet is used as the area number, while "Utility Name" is used as the area name. Other fields, such as the "Street" Address", "City", or "Zip" can be stored, but it is not vital for this case. This now can be sent to **Area** in the power flow package.

Now, all areas need to be controlled by a super area. This adds organization and hierarchy to the model. Once the super area has been added, it is vital to create an area that will hold the slack bus. This bus should have a recognizable area number, something high such as 999999 so there is no overlap with other bus numbers. A separate area for this slack bus must be created.

B. Using EIA-860 Plant 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.

C. 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 [22]. The FIPS code can be the zone number, while the state's postal abbreviation can be the zone name. This can now be sent to **Zone** section in the power flow package. The slack bus zone also needs to be created and given a large, recognizable zone number. This ensures that the slack bus zone is distinct from the other zones and allows for clear organization within the case.

D. Using EIA-860 Plant Data to Create Buses

Using the EIA-860 Schedule 2 data, the buses are created. The difference here from substations (discussed in subsection B) is that the utility code is used to set the bus's area, the state fields are used to set the zone, and the Plant Code is both used as the bus and substation number. "Utility ID" should be the area number, "Plant Code" should be the bus number, and "Plant Name" should be the bus name. The substation number should have the same value as the bus number. As previously stated, there will only be a single bus in each substation for creating the proposed grid. Any other data from EIA-860 Schedule 2 can be stored or deleted as desired. Once this data has been added to the **Bus** section in the power flow package, an additional slack bus needs to be created, with a large number, a proper name, and the same slack area, zone, and substation numbers that were previously given to the slack in the above sections. It is crucial to ensure the case specifies the proper slack bus for the power flow package to function as intended.

E. Using EIA-860 Generator Data to Create the Generators

In the EIA-860 Schedule 3_1 data, there are several sheets. The main sheet is labeled "Operable", which contains only the generators that were in service at the end of the year. There are also two additional sheets: "Proposed", which includes all the generators expected to go into commercial operation within the decade, and "Retired and Canceled", which includes the generators that have gone out of service.

When initially viewing the "Operable" sheet, the generators need to be sorted based on the column titled "Synchronized to the Transmission Grid". If there is an "N" in this field, it indicates that the generator cannot be synchronized with the AC electric power system's frequency, phase, or voltage. If there is "Y" in this column, it can be synchronized, and "X" refers to either. This case is only including generation that can be synchronized, not stand alone backup generators.

At this point, the "Proposed" and/or "Retired and Canceled" sheets can be copied; however, the column names must align with each other, therefore some columns will need to be moved or deleted. The "Plant Code" will serve as the bus number, so the generators should be sorted by this number.

Since the EIA-860 generators are uniquely identified by their plant code (which corresponds to the bus number in the proposed model) but power flow packages use two-character ID's, a new, primary label needs to be created. This label can be created by combining the bus number and the "Generator ID".

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

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:

—c—¿X—¿X—¿X—

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):

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

"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.

F. Creating Lines to Connect the Buses

The final step to connect all buses to the slack bus is to implement low impedance lines, hence creating a copper plate model. This allows for a power flow solution, albeit one with no real transmission grid. When creating the branches, the sending bus number in the proposed model will include all of the bus numbers already implemented. Each of these branches are sent to the slack bus number, which was already created. All impedances are set to zero or very low values. All the line limits on each of the three phases will also be set to zero. Please note that having zero capacity lines is only useful for models that allow line limit capacities in soft constraints or models that assume zero as infinity. Otherwise, we should set line capacities to a large value.

At this point, all the necessary puzzle pieces are there to run operation simulations on the created test case. Initially, all generator outputs are zero, but they can be set to any value, particularly their maximum.

G. 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 I: 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

III. 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

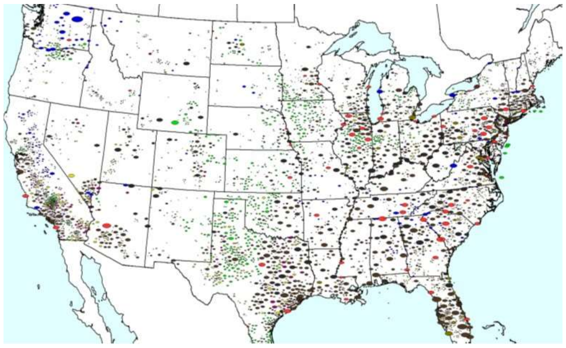

Fig. 1: Generators in the contiguous US in 2023 where size of ovals are proportional to the generator capacity and color refers to their fuel type. In this Figure, red refers to nuclear units, black refers to coal, brown refers to natural gas, green to wind, blue to hydro, yellow to solar, and magenta to energy storage

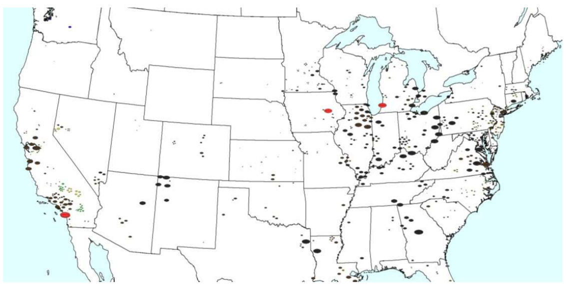

Fig. 2: Retired Generators in 2023 where the size of ovals are increased compared to the previous figure and lay-out is used to avoid overlaps with a similar color code as Figure 1

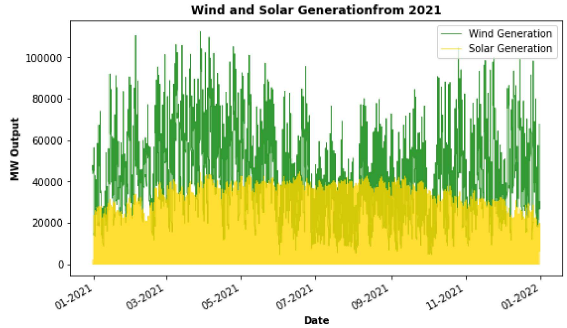

Fig. 3: Renewable Generation output for US states using 2021 Data

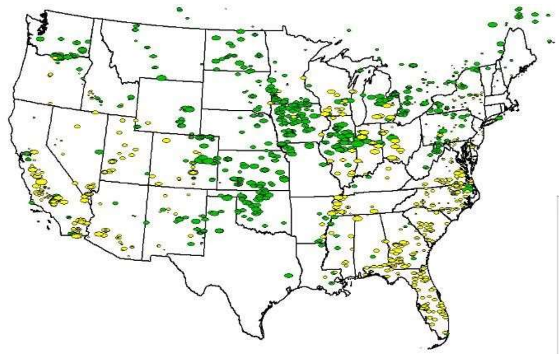

Fig. 4: Overall renewable generator maximum capacity in the US by generator model at the end of 2022 where size of ovals are proportional to the generation capacity, green color refers to wind turbines, and yellow refers to solar cells

[21]. Using the actual generator data is useful for several studies such as emission calculations [23].

To enhance the visual analysis of power grid operations, there are many useful tools to employ, such as the Geographical Data View (GDV) [11], contour mapping [13], and broad visualizations of the transmission grid [17]. These tools are used to increase situational awareness of the grids across diverse scenarios and temporal spans [24], [25].

Figure 1 shows a GDV of the 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. However, Figure 1 includes retired units. Figure 2 shows the retired generators in a separate GDV with the same color code.

Another example of the usefulness of the EIA-860 can be seen in renewable generation. Since wind and solar generation are time-varying, the output at different points in time can be visualized. Figure 3 shows hourly renewable generation values in 2021. —Figures 4, 5, and 6 visualize the renewable generation at different timepoints where the size of the ovals are proportional to the generation capacity.

As explained in section II, subsection C, the US states are setup to be zones. This allows the total generation in each state to be studied. This is also useful when studying renewable generation because weather data can be separated out into states as well. While Figure 3 shows the combined wind and solar generation in the United States, Figures 7 and 8 display the wind and solar production for different states in 2021. This also highlights the importance of the monthly supplements (EIA-860M), given the frequent addition of new renewable generators.

Having all weather measurements can help in visualizing interesting scenarios, such as the solar eclipse than happened on October 14, 2023. Figure 9 shows the impacted areas on the map. Figure 10 visualizes the annular eclipse by contours where the color shows the percentage of solar radiation,

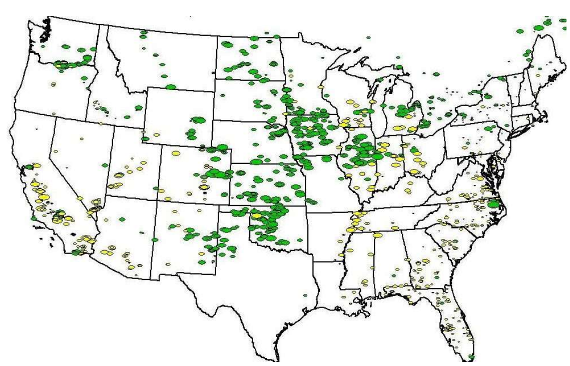

Fig. 5: Renewable generation based on the generation model at the end of 2022 and with the weather measurements in a day with high overall availability of the renewable resources at 1 pm of April 15, 2008 with the same color key as Figure 4

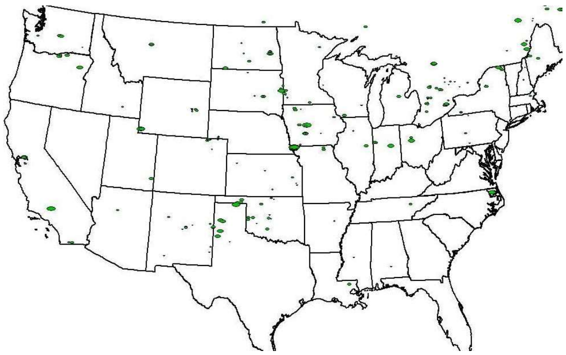

Fig. 6: Renewable generation based on the generation model at the end of 2022 and with the weather measurements in a day with low overall availability of the renewable resources at 8 pm of October 12, 2005 (UTC) with the same color key as Figure 4

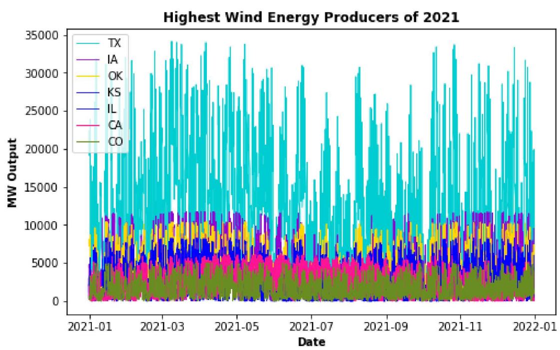

Fig. 7: Wind Energy production in the United States for 2021, separated by states with the highest production

with dark blue referring to 0% and dark red referring to 100%. Figure 11 shows the impact of the eclipse on the solar generation.

IV. Conclusion

This paper 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 paper, 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.

Fig. 9: Affected regions in Solar Eclipse in the United states in 2023

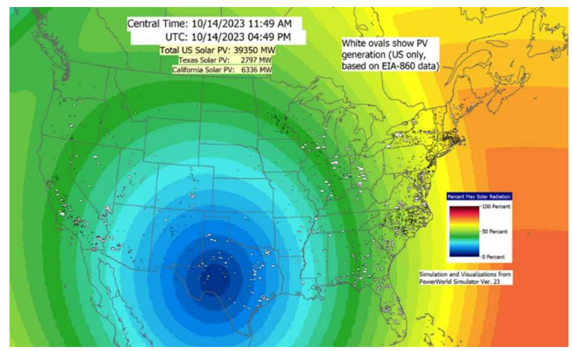

Fig. 10: Solar Eclipse in the United states in 2023

Acknowledgment

This work was partially supported through funding provided by the Power Systems Engineering Research Center (PSERC) through project S-99 and partially by Advanced Research Projects Agency–Energy (ARPA-E) for grid optimization projects.

References

[1] (2019) "U.S. Energy Information Administration (EIA)". [Online]. Available: https://www.eia.gov/electricity/data/eia860/

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

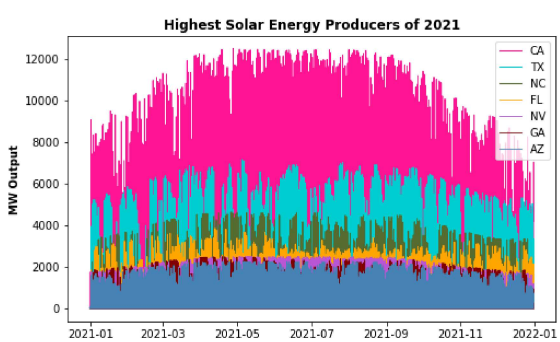

Fig. 8: Solar Energy production in the United States for 2021, separated by states with the highest production

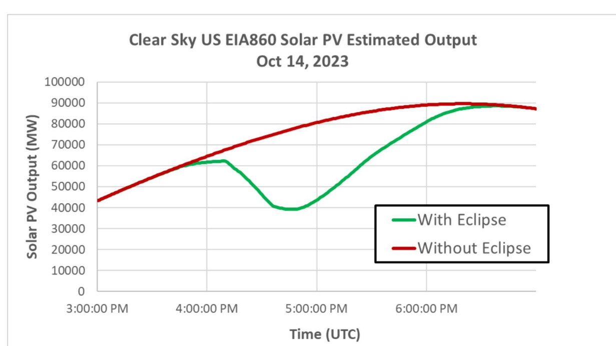

Fig. 11: Potential "clear sky" impact of the eclipse on US generation

[3] A. B. Birchfield and T. J. Overbye, "A review on providing realistic electric grid simulations for academia and industry," *Current Sustainable/Renewable Energy Reports*, pp. 1–8, 2023.

[4] 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.

[5] T. J. Overbye, P. W. Sauer, C. M. Marzinzik, and G. Gross, "A user-friendly simulation program for teaching power system operations," *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 1725–1733, 1995.

[6] [Online]. Available: https://electricgrids.engr.tamu.edu/

[7] W. Group, "Common format for exchange of solved load flow data," *IEEE Transactions on Power Apparatus and Systems*, no. 6, pp. 1916–1925, 1973.

[8] B. Stiller, T. Bocek, F. Hecht, G. Machado, P. Racz, and M. Waldburger, "NERC Libraries of Standardized Powerflow Parameters and Standardized Dynamics Models," North American Electric Reliability Corporation, Tech. Rep., 10 2015.

[9] M. T. Fischer and D. A. Keim, "Towards a survey of visualization methods for power grids," *arXiv preprint arXiv:2106.04661*, 2021.

[10] D. V. Nga, O. H. See, C. Y. Xuen, L. L. Chee *et al.*, "Visualization techniques in smart grid," *Smart Grid and Renewable Energy*, vol. 3, no. 03, p. 175, 2012.

[11] 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.

[12] T. J. Overbye, E. M. Rantanen, and S. Judd, "Electric power control center visualization using geographic data views," in *2007 iREP Symposium-Bulk Power System Dynamics and Control-VII. Revitalizing Operational Reliability*. IEEE, 2007, pp. 1–8.

[13] J. D. Weber and T. J. Overbye, "Voltage contours for power system visualization," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 404–409, 2000.

[14] T. J. Overbye, D. A. Wiegmann, A. M. Rich, and Y. Sun, "Human factors aspects of power system voltage contour visualizations," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 76–82, 2003.

[15] Y. Sun and T. J. Overbye, "Visualizations for power system contingency analysis data," *IEEE transactions on power systems*, vol. 19, no. 4, pp. 1859–1866, 2004.

[16] F. Milano, "Three-dimensional visualization and animation for power systems analysis," *Electric Power Systems Research*, vol. 79, no. 12, pp. 1638–1647, 2009.

[17] T. J. Overbye, J. Wert, K. S. Shetye, F. Safdarian, and A. B. Birchfield, "Delaunay triangulation based wide-area visualization of electric transmission grids," in *2021 IEEE Kansas Power and Energy Conference (KPEC)*. IEEE, 2021, pp. 1–6.

[18] T. J. Overbye and J. D. Weber, "Visualizing the electric grid," *IEEE Spectrum*, vol. 38, no. 2, pp. 52–58, 2001.

[19] P. C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin Jr, R. Guttromson, and J. Thomas, "A novel visualization technique for electric power grid analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 3, pp. 410–423, 2009.

[20] P. C. Wong, H. Foote, P. Mackey, G. Chin, Z. Huang, and J. Thomas, "A space-filling visualization technique for multivariate small-world graphs," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 5, pp. 797–809, 2011.

[21] 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.

[22] "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.

[23] 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.

[24] 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.

[25] 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.