Creating Hourly-Resolution Time-Varying Synthetic Load Data for U.S. Electric Grid Models

Jonathan M. Snodgrass*, Zachary Harrison*, Christopher L. DeMarco, Bernie Lesieutre
Department of Electrical and Computer Engineering
University of Wisconsin-Madison, USA
{jsnodgrass2, zpharrison, cdemarco, lesieutre}@wisc.edu

***Abstract*—Real electrical grid data can be difficult to obtain and is rarely in the desired or useful format needed to conduct research or test applications. This paper presents a comprehensive methodology for creating synthetic hourly load data for the continental United States, addressing this critical need in power systems research. Beginning with the creation of synthetic substation locations using U.S. Census Bureau tract data and Geographic Information Systems (GIS), we develop a framework that combines publicly available load data aggregated to the state level with typical load profiles by class—commercial, industrial, and residential. We fit hourly actual load data to these profiles and disaggregate them to the bus level within each state by leveraging land use databases and census information. Finally, we add reactive power components to each bus based on load class characteristics. The resulting dataset provides realistic, hourly-resolution active and reactive power demands for every synthetic bus location across the United States, enabling researchers to conduct time sequence power flows, optimal power flows with transition constraints, and other advanced power systems analyses without the barriers of data availability or proprietary restrictions.**

***Index Terms*—Load, Synthetic Systems, Load Class, Active/Reactive Power, Network, Substation Location, GIS, Time Series**

I. Introduction

FINDING quality, publicly available power systems data is a common challenge researchers encounter. This paper addresses this by creating synthetic hourly load data for the continental United States. Our goal is to specify active and reactive power demands for each bus within every state for every hour of the year, enabling time sequence power flows, optimal power flows with transition constraints, and contingency analysis. This work is part of the larger EPIGRIDS project to create a synthetic U.S. grid system.

Hourly load datasets have become critical for power systems planning, particularly with accelerating renewable energy adoption. Variable renewable generation, combined with temporally varying loads such as AI datacenters and electric vehicles, demands time-series load data representing both typical operations and extreme scenarios. Resource adequacy studies depend on time-varying datasets, especially when analyzing renewable energy droughts—extended periods of low renewable generation spanning days to weeks [1]–[3]. Weather-informed renewable capacity calculations are essential for identifying these drought conditions [1], and their impacts on resiliency require careful planning consideration [2], with sophisticated methods needed to characterize events across U.S. regions [3]. Grid resilience scenarios spanning hours to days require properly modeled time series load to assess system performance under weather-related hazards affecting transmission planning [4].

Despite these needs, most power system test cases provide only static load snapshots. Traditional IEEE test systems [5] and Reliability Test Systems [6]–[8] represent single operating instances. Synthetic grids like the ACTIVSg series [9] provide primarily static snapshots, focusing on transient stability rather than time-varying patterns [10]. Recent work by [11] developed time-series load profiles using building stock databases and weather data, while the TX-123BT system [12] provides five years of hourly data for Texas. However, comprehensive time-series load data for large-scale, multi-state synthetic grids remains limited.

Our methodology creates state-level load profiles separated by commercial, industrial, and residential classes, then disaggregates to the bus level. This enables selecting any hour to obtain load levels for every bus, or identifying when specific load levels occur—critical capabilities for contingency analysis and system studies. Finding reliable hourly data for entire years is difficult, especially for small regions like states or counties, as most public data covers larger ISO or utility territories. Our dataset provides year-long synthetic load data disaggregated by state and to the bus level across the continental United States, simplifying research that currently requires extensive data gathering efforts.

The remainder of this paper is organized as follows: Section II describes the collection of all data sources required for the methodology, including state-level load data, typical load profiles, census data, and land use information. Section III presents the complete methodology, including the geographic framework for synthetic substation placement, state-level load class decomposition, bus-level disaggregation, reactive power assignment, and post-processing techniques. Section IV provides comprehensive results and validation, including generated load profiles, error analysis across multiple regions,

*Jonathan Snodgrass and Zachary Harrison performed this work at the University of Wisconsin-Madison as students. Jonathan Snodgrass is currently a Senior Research Engineer at Texas A&M University (Snodgrass@tamu.edu) and Zachary Harrison is an Applied Scientist at Zillow.

and discussion of performance factors. Finally, Section V summarizes our contributions and discusses future research directions.

II. Data Sources and Collection

Our work depended on the collection of reliable and accurate data from multiple sources. This section describes the four primary categories of data required: state-level hourly load measurements, typical load profiles by customer class, geographic and census information, and land use classifications.

A. State-Level Hourly Load Data

To obtain hourly load data for each state, we used data from ISOs and EIA [13], which was organized by ISO service territories rather than states. In some cases, we aggregated or subtracted sub-regions. When regions encompassed multiple states, we obtained total annual demand for each state from EIA, calculated each state's percentage of the regional total, and applied this to hourly regional data.

For the Midwest, we used MISO [14] and PJM [15] data. For Wisconsin, we subtracted MISO regional breakdowns to obtain Wisconsin and Michigan combined, then used Michigan's percentage to isolate Wisconsin data. Minnesota and Iowa were extracted from grouped regional data using the percentage method. Illinois combined its MISO region with PJM's Chicago area data. Fig. 1 shows the resulting hourly profiles for each Midwest state.

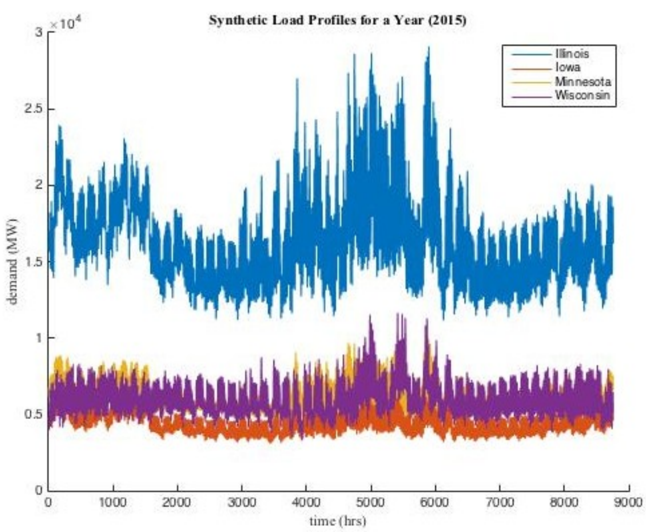

Fig. 1: Calculated load profiles for an entire year for the Midwest region, showing state-level hourly data aggregated from ISO sources.

B. Typical Load Profiles by Class

Beyond state-level hourly data, we also needed typical load profiles disaggregated by load class. These profiles represent the characteristic shape of daily load curves for commercial, industrial, and residential customers. We obtained these profiles from several publicly available sources including NYSEG [16], PGE [17], SCE [18], and AEP Ohio [19]. These profiles distinguish between weekdays and weekends, as well as seasonal variations, providing the necessary granularity to capture realistic load behavior.

Each profile type provides a normalized 24-hour load shape with values that sum or average to unity, which can then be scaled to match actual load levels. The profiles account for the characteristic differences between load classes: residential loads typically peak in the evening hours when occupants return home, commercial loads peak during business hours, and industrial loads tend to be relatively flat with sustained operation throughout the day and night.

We also explored load profiles from the National Renewable Energy Laboratory (NREL) [20], which provides commercial and residential hourly load profiles for all TMY3 locations in the United States. However, as discussed in Section IV, these profiles yielded higher errors in our fitting process.

C. Geographic and Census Data

Due to CEII concerns, we used U.S. Census Bureau tract data rather than real substation locations to subdivide states for synthetic substation placement. The number of census tracts closely agrees with substation counts nationally and by state, providing a practical foundation. Census tracts represent small, permanent geographic subdivisions corresponding to neighborhoods or communities. Census tract boundary data from the U.S. Census Bureau includes population and geographic boundary information essential for load allocation.

D. Land Use and Land Cover Data

To disaggregate state-level loads to individual synthetic bus locations, we utilized the National Land Cover Database (NLCD) [9]. This database provides detailed land use classifications across the United States, allowing us to estimate the proportion of commercial, industrial, and residential activity in each census tract. Figures 2 and 3 show examples of the land cover database and its classification key.

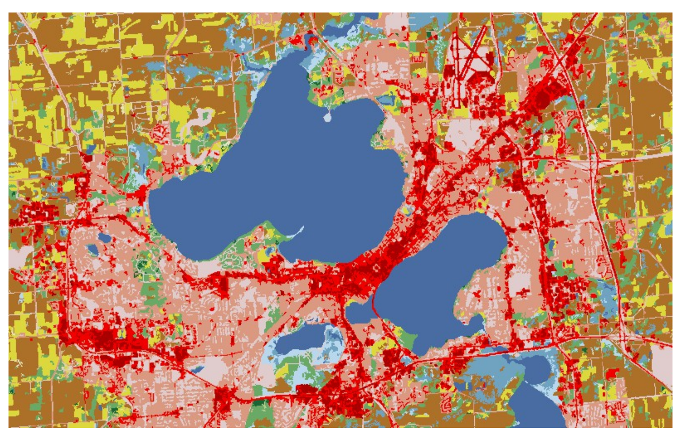

Fig. 2: National Land Cover Database 2011 (NLCD) showing land use classifications across a region.

The NLCD classifies land cover into categories such as developed (open space, low/medium/high intensity), forest,

Fig. 3: Key for the NLCD, showing different land use categories including developed areas, forests, agriculture, and water bodies.

shrub, grassland, pasture, cultivated crops, and wetlands. For our purposes, we focused on the developed categories to identify commercial, industrial, and residential areas. The database provides this information at a spatial resolution of 30 meters, offering sufficient detail to characterize each census tract's land use composition.

III. Methodology

This section presents our methodology for creating synthetic hourly load data from geographic framework through final power profiles.

A. Geographic Framework: Synthetic Substation Locations

We establish geographic locations for load assignment by creating synthetic substation locations across the continental United States.

1) Census Tract-Based Placement: Using census tract data from Section II-C, we created synthetic substation locations by calculating each tract's geographic centroid as GPS coordinates for the synthetic bus. This leverages the correlation between census tract and substation density, ensuring realistic spatial distribution. Census tract boundary data from the U.S. Census Bureau yields a comprehensive set of synthetic bus locations covering the continental United States with density patterns naturally corresponding to population centers.

2) Geographic Feature Handling: For tracts including water bodies or sensitive environmental areas, we modified centroid calculations to avoid unsuitable placements. Using NLCD data, we adjust centroids to weight toward areas suitable for electrical infrastructure, excluding water, wetlands, or other unsuitable categories, effectively shifting locations to realistic positions within tracts.

Fig. 4 shows synthetic substation locations across the continental U.S., demonstrating good coverage and density patterns corresponding to population centers and geographic features.

B. State-Level Load Class Decomposition

We decompose aggregate state load into commercial, industrial, and residential components, essential for realistic bus-level assignment.

Fig. 4: Synthetic substation locations across the continental U.S. Each dot represents a synthetic substation.

1) Load Profile Fitting Methodology: We fit actual hourly state load to typical load profiles through daily optimization. The fitting process determines scaling factors for each load class such that scaled profiles sum to match actual state load. For hour h of day d, we seek scaling factors α_c, α_i, and α_r for commercial, industrial, and residential classes such that:

$$L_{state}(h,d) = \alpha_c \cdot P_c(h,d) + \alpha_i \cdot P_i(h,d) + \alpha_r \cdot P_r(h,d) \quad (1)$$

where $L_{state}(h,d)$ is actual state load, and $P_c(h,d)$, $P_i(h,d)$, $P_r(h,d)$ are normalized typical profiles. This is solved as a non-negative least-squares problem daily, allowing load class composition to vary seasonally.

2) Annual Load Class Curves: After fitting the typical profiles to actual data for each day of the year, we concatenated the results to create continuous annual load curves for each class at the state level. This yields three time series—one for commercial, one for industrial, and one for residential load—that sum to match the actual state load for every hour of the year.

These annual curves capture both the daily patterns characteristic of each load class (e.g., the evening peak for residential customers) and the seasonal variations driven by weather and economic activity. The curves maintain realistic relationships between load classes while matching known state-level totals. Each state thus has three distinct load class time series, each with 8,760 hourly values (for a non-leap year), that can be used for bus-level disaggregation.

C. Bus-Level Load Disaggregation

With annual load class curves established at the state level, the next step is to disaggregate these curves to individual bus locations. This disaggregation creates unique load profiles for each synthetic substation based on the land use characteristics of its census tract.

1) Spatial Allocation by Land Use: For each synthetic bus location (corresponding to a census tract centroid), we determined the mix of commercial, industrial, and residential load based on the land cover database described in Section II-D. Each census tract was assigned percentages p_c, p_i, and p_r representing the proportion of each load class in that tract, where $p_c + p_i + p_r = 1$.

As a first approximation, load was apportioned to each synthetic substation within a state based on the relevant census tract's percentage of the state's total population. For more refined studies, the percentage of load in each class (residential, commercial, industrial) was estimated for each census tract using the land use data associated with that census tract. The percentage of total state load apportioned to each census tract was determined from the land use intensity database.

This approach distributes load realistically by location character rather than uniformly by population. For example, commercial districts receive higher commercial load proportions even with small populations.

Table I shows bus-level allocation for Dane County, Wisconsin, with state peak load distributed across census tracts. Bus 55025000408 has 71% commercial, 21% industrial, and 7% residential loads (business district character), while buses 55025000401 and 55025000503 are predominantly residential (100% and 92%). Power factors reflect load class mixes, with higher commercial/industrial percentages showing factors closer to unity.

TABLE I: Bus-level load allocation for Dane County, Wisconsin.

Bus Name (Tract ID)	MVA Load	MW Load	Mvar Load	PF	% State Peak	Com (%)	Ind (%)	Res (%)
55025000205	4.98	4.63	1.84	0.93	0.000318	0.11	0.18	0.70
55025000300	6.59	6.13	2.42	0.93	0.000421	0.30	—	0.70
55025000401	2.76	2.52	1.12	0.91	0.000173	—	—	1.00
55025000402	9.24	8.73	3.04	0.95	0.000600	0.33	0.45	0.23
55025000405	11.93	11.09	4.44	0.93	0.000715	0.58	—	0.42
55025000406	3.66	3.33	1.53	0.91	0.000229	0.23	0.05	0.71
55025000407	5.53	5.06	2.23	0.92	0.000349	0.20	0.10	0.69
55025000408	19.41	18.22	6.71	0.94	0.001251	0.71	0.21	0.07
55025000501	8.73	8.25	2.86	0.95	0.000567	0.28	0.26	0.45
55025000503	6.81	6.16	2.90	0.91	0.000423	0.22	0.08	0.92

2) Active Power Time Series Generation: Bus load is determined by the tract's share of state load in each class. Let n be total buses in a state and w_j be the weight for bus j based on population or land use intensity. For load class k ($k \in \{c, i, r\}$), load at bus j for hour h is:

$$L_j^k(h) = L_{state}^k(h) \cdot p_j^k \cdot \frac{w_j}{\sum_{j=1}^{n} w_j \cdot p_j^k} \tag{2}$$

where $L_{state}^k(h)$ is state-level load for class k at hour h, and p_j^k is the proportion of load class k at bus j. Total active power at bus j is:

$$P_j(h) = L_j^c(h) + L_j^i(h) + L_j^r(h) \tag{3}$$

Fig. 5 shows the overall process, illustrating data flow from state-level information through load class fitting to final bus assignments.

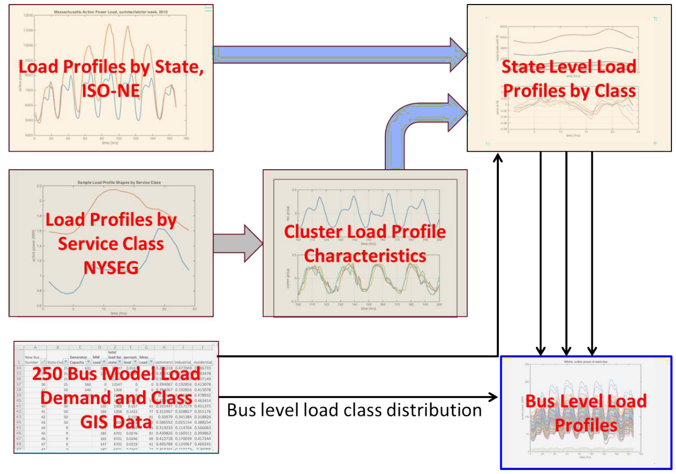

Fig. 5: Flowchart for creating bus-level load profiles.

D. Reactive Power Assignment

We add reactive power components based on load class composition and power factor characteristics. Different load classes exhibit different power factors: residential loads (with inductive devices) have lower power factors, commercial loads (with power electronics and correction equipment) have higher factors, and industrial loads have the highest due to regulatory and economic incentives.

Power factor ranges (Table II) were assigned based on literature and utility data. For each bus and hour, power factors were randomly assigned within these ranges for each load class.

TABLE II: Power Factor Ranges by Load Class

Load Class	Power Factor Range
Residential	0.90 – 0.95 lagging
Commercial	0.95 – 0.98 lagging
Industrial	0.98 – 1.00 lagging

1) Reactive Power Calculation: Once power factors were assigned, we calculated reactive power for each load class component at each bus. For a load component with active power P and power factor pf, the reactive power Q is given by:

$$Q = P \times \left(\frac{\sqrt{1 - pf^2}}{pf} \right) \tag{4}$$

The total reactive power at bus j for hour h is then the sum of reactive power contributions from all load classes:

$$Q_j(h) = Q_j^c(h) + Q_j^i(h) + Q_j^r(h) \tag{5}$$

where each term is calculated using (4) with the appropriate active power component and power factor for that load class.

This approach ensures that reactive power profiles maintain realistic relationships with active power profiles, with appropriate variation based on load class composition at each bus location.

E. Post-Processing and Quality Control

Initial load profiles showed intermittent jumps between consecutive days in commercial and industrial classes due to day-by-day fitting and concatenation. We developed a smoothing technique for natural day transitions.

We examine the last five hours of one day (hours 20–24) and first five hours of the next (hours 1–5), adding corrective terms z and y with quadratic corrections using constants α and β:

$$z(20:24) = \alpha \times \frac{[0\ 1\ 4\ 9\ 16]}{16}, \quad y(1:5) = \beta \times \frac{[16\ 9\ 4\ 1\ 0]}{16} \tag{6}$$

The smoothing problem minimizes total load change while enforcing positive second derivatives around midnight, forcing both days to have the same concavity. Second derivatives were estimated using four points (last two hours of first day, first two hours of second day) and two points (last and first hours). Constraints applied to all three load classes yield correction constant pairs (α_1, β_1), (α_2, β_2), (α_3, β_3). The objective function is:

$$\min \left[(\alpha_1 + \alpha_2 + \alpha_3)^2 + (\beta_1 + \beta_2 + \beta_3)^2\right] \tag{7}$$

This technique produces smooth day transitions suitable for time-domain simulations.

IV. Results and Validation

We present generated load profiles and validation through error analysis across U.S. regions.

A. Generated Load Profiles

Our methodology produces comprehensive profiles at state and bus levels. Figs. 6 and 7 show hourly active and reactive loads for 6,673 buses in the Midwest (Illinois, Iowa, Minnesota, Wisconsin), demonstrating scale and complexity with unique profiles varying throughout the year showing expected daily cycles and seasonal variations.

B. Error Analysis and Regional Performance

We validated by comparing summed bus-level loads to original state data. Since our fitting process matches state data, primary error sources are profile-to-pattern fit quality.

New England achieved average errors under 2% using NYSEG profiles, suggesting good fit with moderate climate and mixed economic activity. Midwest states showed higher errors, typically under 3% but occasionally exceeding 10%. Fig. 8 shows Wisconsin's error distribution. Higher Midwest errors likely stem from extreme weather (cold winters, hot humid summers) differing from New York profiles, particularly affecting residential heating and cooling.

Similar errors occurred in Ohio, Michigan, and Indiana using Ohio Power Company data. Western states (California, Washington, Oregon) produced comparable results with PGE and SCE profiles. Region-specific profiles (PGE for Northern California, SCE for Southern California) provided reasonable

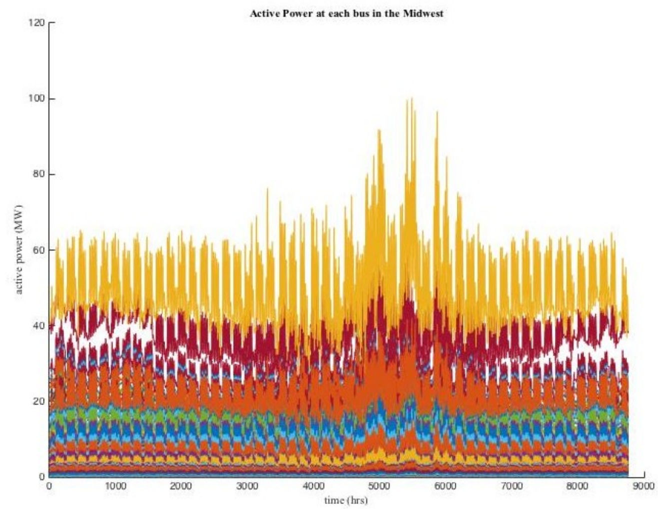

Fig. 6: Active load for 6,673 Midwest buses showing hourly variation throughout the year.

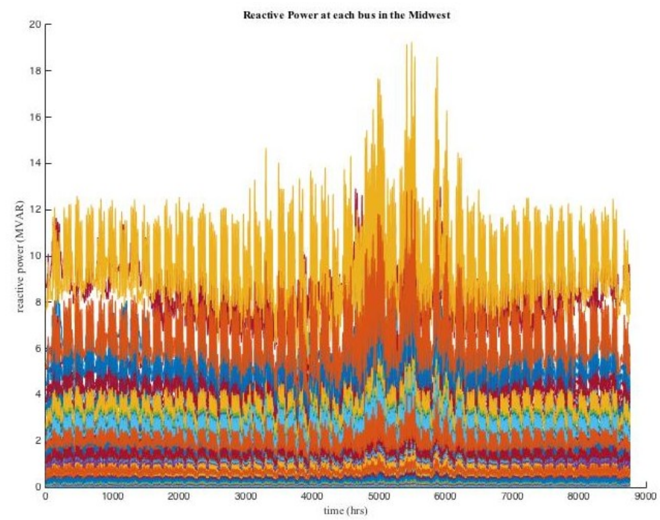

Fig. 7: Reactive load for Midwest buses.

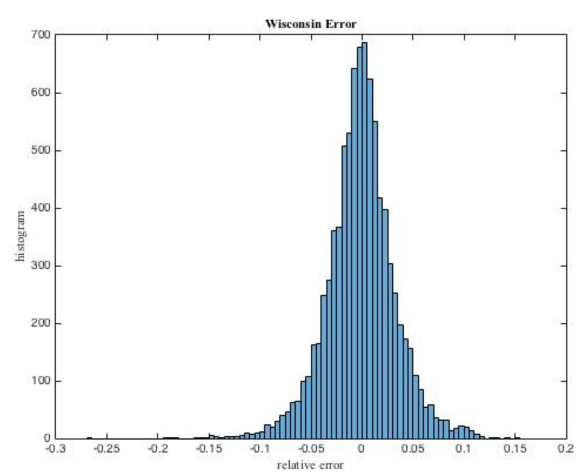

Fig. 8: Wisconsin error distribution histogram.

accuracy, confirming benefits of matching profiles to regional climate and economics.

NREL profiles showed higher errors (exceeding 20%) consistently across the U.S., possibly due to differences between building models generating NREL profiles and actual building stock and behavior patterns.

Despite varying error levels, the data captures essential characteristics—daily cycles, seasonal variations, spatial distribution, and realistic active-reactive power relationships—critical for many applications.

V. Conclusions and Future Work

We present a comprehensive methodology for creating synthetic hourly load data across the continental United States. Our pipeline creates synthetic substation locations using census tract data, decomposes state load into commercial, industrial, and residential components through optimization-based fitting, spatially disaggregates using land cover data, assigns reactive power based on load class power factors, and applies smoothing for continuous trajectories. Validation across multiple regions shows errors typically under 3%. The dataset, publicly available at https://electricgrids.engr.tamu.edu/electric-grid-test-cases [21], enables time sequence power flows, optimal power flows with transition constraints, and advanced analyses without data availability barriers.

Future work includes developing region-specific typical load profiles to better capture local weather and economic characteristics, testing profiles in regions with similar patterns, and exploring load-dependent power factor models for residential loads. We will continue expanding the public repository to enable renewable resource adequacy studies, resilience analysis, and integration of emerging loads such as electric vehicles and data centers.

Acknowledgements

This work was supported by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000717.

References

[1] 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*, Belgrade, Serbia, June 2023, pp. 1–6.

[2] F. Safdarian, J. L. Wert, D. Cyr, and T. J. Overbye, "Power system resiliency and reliability issues from renewable resource droughts," in *Kansas Power and Energy Conference 2024*, April 2024.

[3] K. Zhgun, F. Safdarian, and T. J. Overbye, "Classification and identification of renewable energy droughts in the united states," in *2025 IEEE Texas Power and Energy Conference (TPEC)*, College Station, TX, February 2025.

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

[5] R. Christie, "Power systems test case archive," University of Washington, Seattle, 1993, available at: https://labs.ece.uw.edu/pstca/.

[6] IEEE Reliability Test System Task Force, "Ieee reliability test system," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, no. 6, pp. 2047–2054, Nov./Dec. 1979.

[7] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, and C. Singh, "The ieee reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.

[8] A. Barrows *et al.*, "The ieee reliability test system: A proposed 2019 update," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 119–127, Jan. 2020.

[9] 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, pp. 3258–3265, 2017. [Online]. Available: https://dx.doi.org/10.1109/TPWRS.2016.2616385

[10] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *Proc. 10th Bulk Power Systems Dynamics and Control Symposium (IREP 2017)*, Espinho, Portugal, Sept. 2017.

[11] H. Li, J. Yeo, A. 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. 1, pp. 961–969, Jan. 2021.

[12] S. Jiang, Z. Li, and M. B. Cain, "A synthetic texas power system with time-series weather-dependent spatiotemporal profiles," *Scientific Data*, vol. 10, no. 1, p. 436, 2023.

[13] U.S. Energy Information Administration, "U.S. Electric System Operating Data," 2018, accessed: 2018. [Online]. Available: https://www.eia.gov/realtime_grid/

[14] Midwest ISO Energy, "Market Reports," 2016, accessed: 2016. [Online]. Available: https://www.misoenergy.org/Library/MarketReports/Pages/MarketReports.aspx

[15] PJM Interconnection, "DataMiner2," 2018, accessed: 2018. [Online]. Available: https://dataminer2.pjm.com/feed/hrl_load_metered/definition

[16] New York State Electric & Gas Corporation, "Load Profiles," accessed: 2018. [Online]. Available: https://www.nyseg.com/suppliersandpartners/servicesandresources/loadprofiles

[17] Pacific Gas and Electric, "2000 Static Load Profiles," 2000, accessed: 2000. [Online]. Available: https://www.pge.com/nots/rates/2000_static.shtml

[18] Southern California Edison, "SCE Dynamic Load Profiles," 2018, accessed: 2018. [Online]. Available: https://www.sce.com/regulatory/load-profiles/dynamic-load-profiles

[19] AEP Ohio, "Load Profiles," 2018, accessed: 2018. [Online]. Available: https://www.aepohio.com/account/service/choice/cres/LoadProfiles.aspx

[20] Office of Energy Efficiency & Renewable Energy, "Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States," 2013, accessed: 2013. [Online]. Available: https://openei.org/doe-opendata/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-s

[21] Texas A&M University Electric Grid Test Case Repository, "Electric grid test cases," 2025, accessed: 2025. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases