CaseBuilder and the EIA-860 Dataset: Towards a Rapid Power-Case Creation and Modification System

Eric Keller, William Yun, Jonathan M. Snodgrass, Thomas J. Overbye
Department of Electrical and Computer Engineering
Texas A&M University, College Station, TX, USA
{erickeller, williamyun, snodgrass, overbye}@tamu.edu

***Abstract*—This paper introduces CaseBuilder, a modular, object-oriented workflow used to convert the EIA-860 database into a copper plate model of the generation fleet of the United States. Casebuilder improves on previous, more manual methods of creating cases from datasets. CaseBuilder includes a suite of data cleaning functions and multiple methods of interacting with the chosen power flow modeling software. It serves as the foundation for future, automated, case generation and modification workflows.**

In this example, using a publicly available generation inventory and power-flow modeling software, we recreated a case that researchers and engineers use to study the output of the US generation mix across both weather and time. We can now automate the production of updated versions of these cases, and to generate past versions of this case from previous versions of the dataset to examine what, if anything, has changed across the available timeframe. We can also rapidly iterate on this case, to explore potential changes to the generation fleet.

***Index Terms*—Power system modeling, case generation, EIA-860, data automation, SimAuto, AUX files, renewable generation, scenario analysis.**

I. Introduction

The U.S. power grid is undergoing a rapid transformation driven by large-scale integration of variable renewable energy. In a single day, solar generation can drop from supplying 40% of grid demand during midday to near zero within hours as the sun sets, requiring conventional generation to ramp up 13–17 GW in just 3 hours, a phenomenon known as the "duck curve" [1]. This intraday variability historically occurred only over seasonal timescales. Operators manage these fluctuations using short-term forecasts (1 to 3 days), but planners designing the grid five to ten years ahead lack equivalent tools. They cannot predict specific weather patterns a decade in advance, yet those patterns will fundamentally determine grid reliability as renewables displace dispatchable generation.

Current planning practices treat this uncertainty crudely. System operators and regional utilities employ worst-case heuristics for near-term planning: assume solar and wind operate at 5% of nameplate capacity and ensure fossil units can compensate entirely. For long-term planning with higher renewable penetration and retiring fossil capacity, planners often treat renewables as fully dispatchable, modeling them as if their output were controllable like conventional thermal units. This assumption is demonstrably false, but without data-driven scenarios that reflect realistic weather variability, no practical alternative exists.

The consequences extend beyond technical reliability. Regulators and policymakers need quantitative answers: How many hours per year will renewables produce below critical thresholds? What dispatchable capacity (storage, gas, nuclear) is required to maintain reliability? Without these numbers, utilities cannot justify infrastructure investments to state legislatures or public utility commissions.

Historically, utilities and regional reliability coordinators have built a handful of static stress cases, typically six per year representing summer and winter high- and low-load conditions, as well as spring and fall shoulder seasons. These cases sufficed when most generation was dispatchable. Today, however, grid uncertainty arises from weather-driven variability, requiring analysis of *hundreds or thousands* of scenarios to represent different weather patterns, renewable droughts (extended low-output periods), and combinations of retirements and new resources. Yet the software and data pipelines used to construct these power-flow cases have barely changed. Creating a single case using publicly available generation inventories and commercial power-flow analysis tools remains a manual, error-prone process that can take days of spreadsheet cleaning, scripting, and repeated verification.

To address this bottleneck, CaseBuilder was developed as a software package to rapidly create, modify, and rectify cases. CaseBuilder automates the previously manual process of assigning generator and bus data, applying geographic coordinates, linking plants to regions, and producing clean, solvable cases in standard formats, including PowerWorld, PSSE, and PSLF. By translating raw generation inventory data into structured Python objects and using batched file operations, CaseBuilder reduces case creation time from hours to minutes and provides the foundation for scalable, data-driven planning-case generation.

Results presented here were done using PowerWorld Simulator Version 24.

Copyright ©2026 IEEE. Personal use of this material is permitted. This paper was presented at the Texas Power and Energy Conference (TPEC), College Station, TX, February 2026.

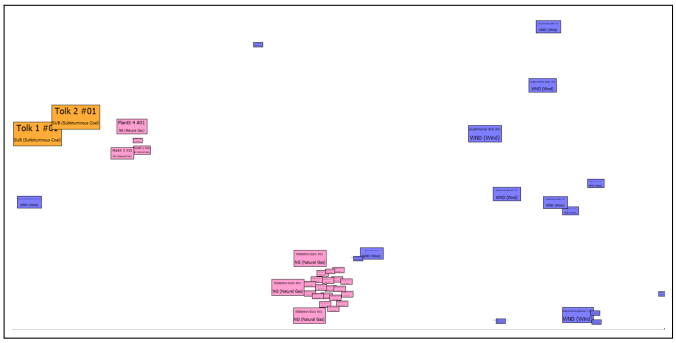

Fig. 1. Geographic data view (GDV) of generation in the Texas Panhandle from the EIA-860 2024 Release. Pink represents gas generators, blue represents wind, grey is distillate fuel oil. Scaled by maximum MW output.

II. Previous Work

This work builds on a sequence of projects that connect weather data, renewable resource modeling, and public generator inventories to power system planning studies. In this section we briefly summarize that prior work and explain how CaseBuilder extends it.

A. Weather-Integrated Power Flow and Renewable Droughts

Traditional power flow studies treat renewable generation using fixed injections or simple derates, which cannot capture spatially and temporally correlated weather patterns. Previous work developed a framework in which hourly weather measurements are mapped to generator locations and used to adjust generation, line ratings, and other parameters inside the power-flow model itself [2]. That paper showed that explicitly coupling weather and power flow can reveal stress conditions that are not apparent under static assumptions, motivating the need for large numbers of weather-driven cases rather than a small set of stylized snapshots.

Subsequent work used this weather-integrated framework to construct long time series of renewable output and identify periods of extended underperformance, or renewable resource droughts [3], [4]. In these studies, historical weather reanalysis data was combined with a fixed generator portfolio to *backcast* renewable production across many decades. For example, by applying 1985 weather to the 2021 generator fleet, it was possible to identify how that fleet would have performed during historical weather patterns. The resulting data sets allowed detection and characterization of multi-day wind and solar droughts, quantification of their severity and frequency in different regions, and assessment of their implications for grid resilience and resource adequacy. A key requirement of this approach is the ability to generate consistent power-flow cases that reflect a specific generator inventory while being reused across many years of historical weather.

B. Manual Historical Cases and Planned Generation Data

To support these weather-driven studies with publicly available data, a detailed manual workflow for converting the EIA-860 generation inventory into power system cases was developed [5]. The EIA-860 dataset [6], published annually by the U.S. Energy Information Administration, provides a comprehensive inventory of all U.S. electric generators with nameplate capacity of 1 MW or greater, including plant-level data (location, operator), generator-level data (capacity, fuel type, operating dates), and renewable-specific details (turbine class, tracking configuration). This work demonstrated how to clean and merge these files, interpret technology-specific fields, and construct copper-plate cases suitable for visualization and analysis in commercial tools. The resulting cases, along with other synthetic networks, are distributed through the Texas A&M synthetic grid repository [7]. However, the EIA-860 case construction process itself remained largely manual, limiting the number of historical years and variants that could be created in practice.

Generators, or more accurately, plants, with generation smaller than 1 MW may, in the aggregate, change the generation profile of the United States. The dataset contains generators smaller than 1 MW, but there may be KW-level privately held generators whose aggregate output could change the profile of the state level generation. This effect is not modeled by the case created here, but may be in future work.

More recently, interconnection queue data from seven U.S. ISOs was consolidated into a unified database, demonstrating how this information can be used for power system planning models [8]. This work shows how large-scale proposed generation portfolios can be cleaned, standardized, and prepared for integration with PowerWorld-style planning cases, enabling analysis of future high-renewable scenarios based on active queue entries and technology mixes. Together with the EIA-860 based cases, this provides both historical and prospective views of the generation fleet.

C. Contribution of CaseBuilder

CaseBuilder automates and generalizes the earlier manual EIA-860 workflow and links it directly to these weather-driven and planning studies. Instead of hand-building a single case, CaseBuilder provides a modular pipeline that ingests both EIA-860 files and interconnection queue data, constructs the necessary power system objects, and exports cases in bulk. This automation makes it feasible to generate dozens of historical cases across many EIA-860 reporting years, as well as multiple planning cases based on interconnection queue data, providing the large case ensembles required by prior weather-in-the-power-flow and renewable drought analyses.

III. EIA-860 Dataset

CaseBuilder uses the EIA-860 dataset as its primary data source for several reasons. Although regional reliability coordinators such as WECC or ISOs/RTOs maintain detailed planning cases that include transmission topology and loads, these datasets are not publicly available. Access typically requires nondisclosure agreements because they contain proprietary utility data and critical infrastructure information. In addition, they are aggregated from numerous utility submissions and often contain inconsistent naming, duplicate objects, and incomplete generator attributes. In contrast, the EIA-860 dataset is fully

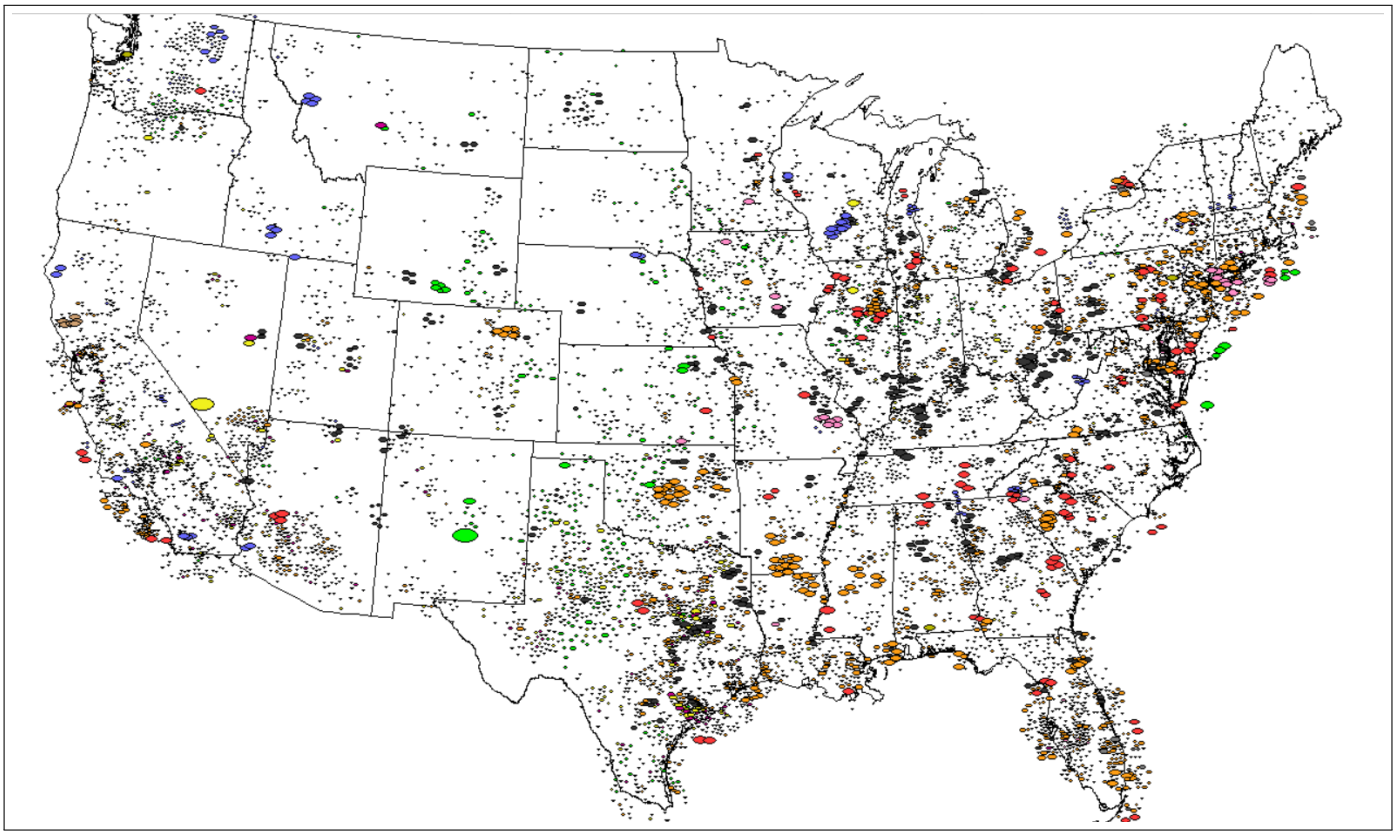

Fig. 2. GDV of the continental US from the 2023 EIA-860 dataset. Wind generation is in green, solar in yellow, gas in orange, and nuclear in red. Generator indicators are scaled to output and spread out to lessen overlap between indicators.

public, uniform across regions, and standardized nationwide, making it ideal for automated case creation. By focusing on EIA-860, CaseBuilder provides a consistent baseline that can later be aligned with or used to validate WECC or ISO/RTO cases within GridWorkshop.

The dataset is released annually, with early-release versions typically made available mid-year. CaseBuilder uses four primary files (Plant, Generator, Wind, and Solar) to construct copper-plate cases where each generator connects to a single slack bus through individual branches.

The dataset presents multiple obstacles to creating these cases. The first and most trivial is that the column names on the dataset need to be translated into names usable by the power-flow software. The second, more serious problem, is that not all of the content is usable as presented. Empty cells have to be filled; some cells are blank (containing " ") and are therefore strings in what are supposed to be numeric columns. Any commas within cells containing strings have to be stripped out, as commas are used as a delimiter by various programs and in python. Third, not all of the relevant data is contained within a single file, and files have multiple sheets which must be included for a complete historical inventory. Simply concatenating the inventories would result in duplicate generators with incomplete information. Finally, not all of the required fields for object creation in the power system modeling software are present in the case: reactive power values are largely missing.

IV. CaseBuilder Design and Workflows

CaseBuilder has a two-tier class hierarchy. The primary class is the "case" class, of which all other classes are subclasses. CaseBuilder includes classes for the major node types in the power modeling software (shunts, loads, generators) and the primary branch type (AC lines). Each object has a set of key fields that must be present to successfully create the objects, called key fields. Other fields, called optional fields, are handled with optional arguments. The names fed into the class are evaluate against valid field names; matches are kept, and mismatches discarded. Created objects are appended to a list in the case class, which is checked to prevent duplicate objects. The package is flatter than some other methods of creating cases. The idea was that the nodes and branches are what the case is fundamentally about and the organization hierarchies are downstream of the properties of these objects, so the amount of organizational overhead necessary to create a case object with this package was minimized. The case class contains all functions common to the classes used in the software. It's chief use is in creating the aux files for batch insertions, checking for duplicates, and inserting single objects. It can also insert objects and make changes directly.

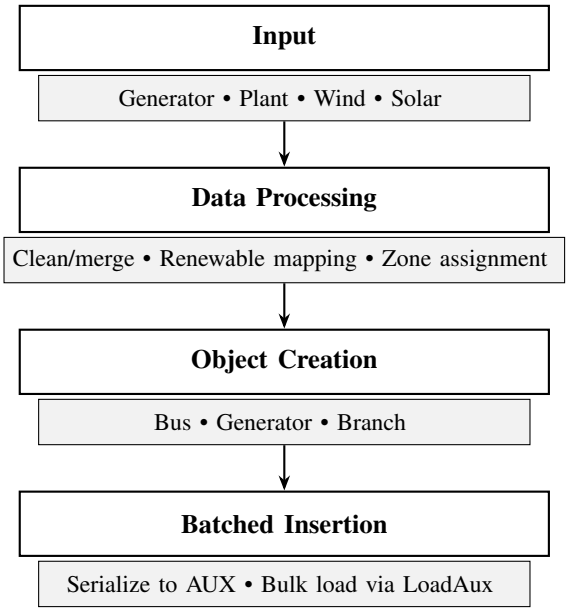

Fig. 3. CaseBuilder layered architecture emphasizing batched AUX insertion for 100× speedup over row-wise SimAuto calls.

A. Loading and Cleaning EIA Data

1) File Loading and Merging: CaseBuilder loads the EIA-860 Plant, Generator, Wind, and Solar files. The Plant file provides geographic coordinates that are used to map weather effects to renewable generators. The Generator file contains all generation types, while the Wind and Solar files provide technology-specific details for renewable units including turbine class and tracking configuration. These files span multiple sheets (Operable, Proposed, Retired) that must be merged without creating duplicate entries.

2) Data Quality and Cleaning: Several data quality issues require correction. Column names must be translated to PowerWorld-compatible field names. Empty cells and blank strings in numeric columns are filled with appropriate defaults. Commas within string fields are removed to prevent delimiter conflicts. Plant names are sanitized to remove special characters that would cause parsing errors.

3) Renewable Technology Mapping: Wind generators are classified by turbine class and mapped to PowerWorld generator types (IEC Class 1 maps to W1, Class 2 to W2, etc.). Solar generators are classified as either fixed-tilt (P1) or single-axis tracking (P2) based on the tracking configuration indicated in the EIA-860 Solar file. These classifications are merged back into the main generator dataframe by matching plant names and generator IDs.

4) Bus Naming and Numbering: Bus names are created by combining the sanitized plant name with the EIA-860 generator ID. Bus numbers are assigned sequentially from the DataFrame index. All buses default to $138\,\mathrm{kV}$ nominal voltage, representing a medium-voltage transmission level suitable for copper-plate models.

B. CaseBuilder: Object batch insertion

Generator reactive power limits require careful handling because EIA-860 does not consistently report power factors, so defaults are applied: if the reported power factor is missing, invalid (≤ 0 or > 1.0), or unrealistically low (< 0.70), it defaults to 0.95. This default aligns with FERC Order 827, which requires new large generating facilities to maintain a power factor of at least 0.95 [9]. This ensures all generators have reasonable reactive capability without manual data cleaning. Reactive limits are then calculated from nameplate capacity S and power factor PF:

$$Q_{\max} = \sqrt{\left(\frac{S}{\mathrm{PF}}\right)^2 - S^2}, \tag{1}$$

$$Q_{\min} = -\sqrt{\left(\frac{P_{\min}}{\mathrm{PF} + 0.05}\right)^2 - P_{\min}^2}. \tag{2}$$

The offset in $Q_{\min}$ prevents unrealistically wide reactive ranges when minimum load is significant. For bidirectional resources (pumped hydro, batteries), symmetric limits $Q_{\min} = -Q_{\max}$ are used. Each generator's minimum output is set to the EIA-860 "minimum load" value.

For example, consider a 500 MW wind farm in the EIA-860 dataset with no reported power factor. CaseBuilder applies the default power factor of 0.95, yielding $Q_{\max} = 164$ MVAr and $Q_{\min} = -82$ MVAr (assuming $P_{\min} = 0$). The generator connects to a dedicated 138 kV bus, which is then connected to the system through an individual branch.

Buses, generators and branch creation calls are written into an AUX file, which is then loaded and executed. A slack bus and slack generator with sufficient capacity are inserted last, to allow the case to solve.

By default, CaseBuilder builds a copper-plate model in which all generators connect to a single slack bus through individual branches. This configuration assumes infinite transmission capacity, allowing researchers to isolate and study generation capacity, mix, and availability without introducing transmission-level constraints. The slack bus serves as the system's voltage reference and automatically balances any mismatch between total generation and network losses, ensuring that power flow equations remain solvable even in the absence of explicit loads.

Other cases can be created with Casebuilder's toolset. There are classes for loads and shunts, which were tested but unused in this copper plate model. The original case being mimicked didn't include them, so they weren't included here. Branch limits, too, are included as modifiable parts of the branch class, but the limits were set unrealistically high to focus on how generation changes with respect to the weather.

Using the a method of inserting objects defined by fields in a dataframe directly into cases by calling Powerworld's simauto functions, the 36,000+ bus case too approximately 6 hours to assemble. The AUX batch insertion takes less tham 30 minutes.

Figure 4 shows an example case generated by CaseBuilder from EIA-860 data, containing over 30,000 generators dis-

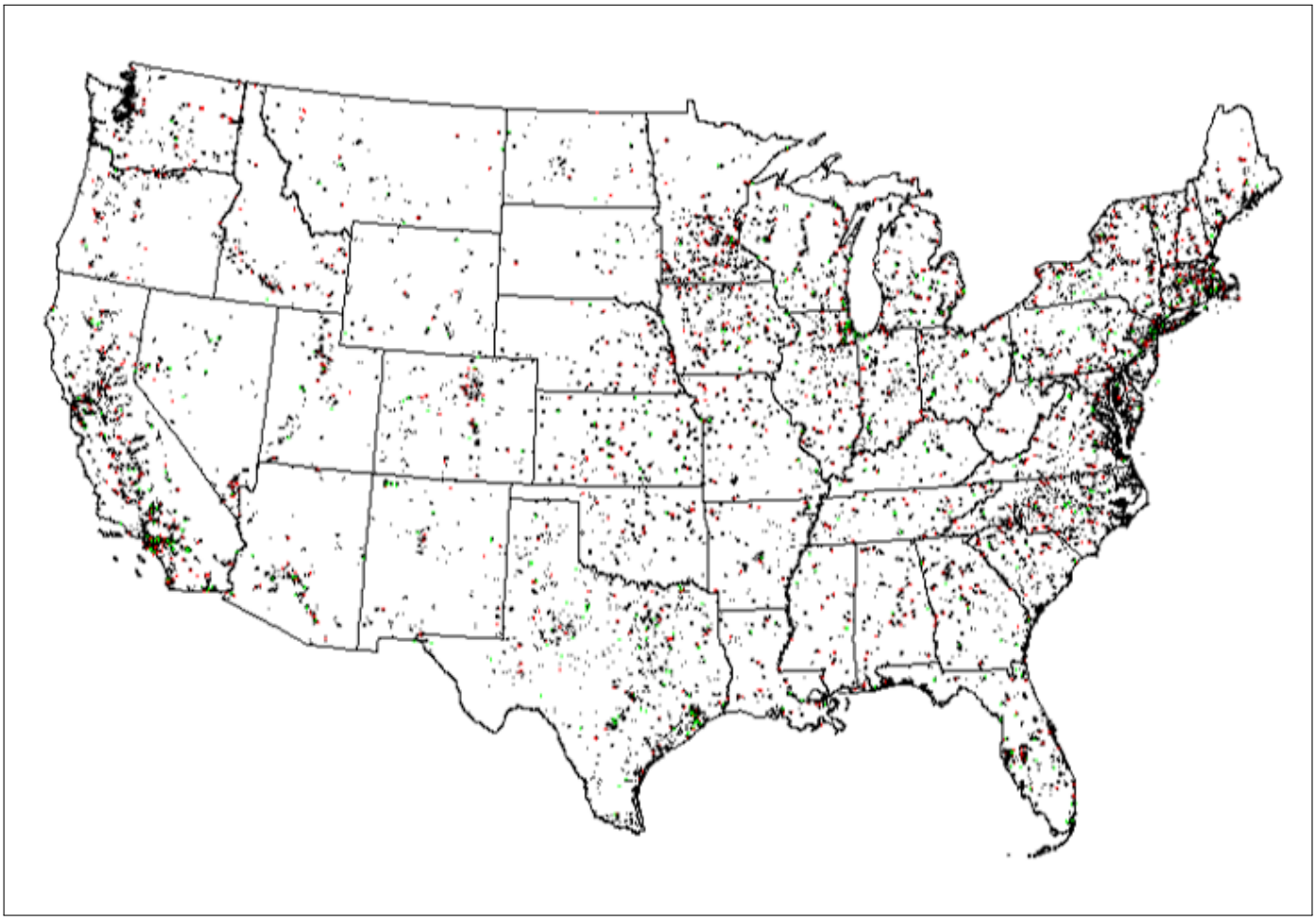

Fig. 4. A geographical display of generators within the continental United States taken from the EIA860 copper plate case as generated by Casebuilder. Alaska and Hawaii are included, but not shown.

tributed across the United States. The batched AUX file insertion approach reduces case creation time from hours to minutes for national-scale cases. The class-based structure ensures that identical EIA inputs always produce identical cases, and new object types can be added by extending the base class definition.

V. Conclusion and Future Work

CaseBuilder provides a significant improvement over previous manual methods for creating EIA-860 power system cases. By automating data cleaning, object creation, and batch insertion through AUX files, CaseBuilder replaces a manual, error-prone process with a more repeateable and scalable one. The method using aux commands to create the case directly is time intensive (Figure 5), but batch insertion reduces this time frame to minutes.

THe bottleneck on case size for the excel or csv to case commands is RAM, as larger files either need to be broken into chunks with each chunk fed invididually or must be run on a host that can load the full dataset in memory. Speed is determined by CPU speed and core count.

CaseBuilder's object-based design allows for modularity and serves as the foundation of future case-creation, correction, and analysis tools. To demonstrate CaseBuilder's capability at scale, 25 historical EIA-860 cases spanning 2001 to 2025 are currently being generated to analyze how renewable resource droughts have evolved as the generation fleet has changed over time. Several potential future planning cases are also being created by ingesting the consolidated interconnection queue data [8], enabling forward-looking resilience studies under different renewable deployment scenarios. In the immediate term, historical EIA-860 cases from previous releases of the dataset (2001-2025) are being created. The ease and speed of creating cases enables quick determination of what differences, if any, these datasets would cause in the analysis. Current development projects include correcting anomalously long or short line lengths, aligning branch limits with resistance and vice-versa, and automatically inserting generation models for renewables. Other features will include load modeling, case comparison and correction, and error detection. Combined, these tools will enable rapid creation of scenarios for existing and future cases, both synthetic and real, where needed.

Fig. 5. Progress bars of the aux command method fo creating the EIA860 case in Powerworld

Casebuilder, as its name suggests, is a toolset for creating cases. Classes, fields, and functions will be added as needed, and all objects found in a real case are intended to be supported by Casebuilder. Casebuilder is intended to be a single module in a broader system. It's toolset will be expanded to include modification of existing records, the insertion of power flow models, and integration into a scenario creation and evaluation workflow. Whether modifying existing cases or simply creating variations of the same cases is faster has yet to be determined. Our intend is that we will be able to expand Casebuilder to work with a broad range of commercial software, but it currently only works with Powerworld.

ACKNOWLEDGMENT

Funding was provided in part by the Department of Energy (DOE) Office of Cybersecurity, Energy Security, and Emergency Response (CESER) through project DE-CR0000067, and in part by The Department of Energy (DOE) Grid Deployment Office (GDO) under project number DE-EE0010491. This work is also part of the "Electric Grid Resilience" Project (60NANB24D210) funded by the National Institute of Standards and Technology of US Department of Commerce. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] California Independent System Operator (CAISO), "What the duck curve tells us about managing a green grid," California Independent System Operator, Folsom, CA, USA, CAISO Fast Facts, 2016. [Online]. Available: https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf

[2] T. J. Overbye, F. Safdarian, W. Trinh, Z. Mao, and J. Yeo, "An approach for the direct inclusion of weather information in the power flow," in *Proc. 56th Hawaii Int. Conf. System Sciences (HICSS)*, Jan. 2023.

[3] J. L. Wert, F. Safdarian, A. Gonce, T. Chen, D. Cyr, and T. J. Overbye, "Wind resource drought identification methodology for improving electric grid resiliency," in *Proc. IEEE Texas Power and Energy Conf. (TPEC)*, College Station, TX, Feb. 2023.

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

[5] J. S. Cook, F. Safdarian, J. Snodgrass, and T. J. Overbye, "Using power flow application capabilities to visualize and analyze US energy information administration generation data," in *Proc. IEEE Texas Power and Energy Conf. (TPEC)*, College Station, TX, Feb. 2024.

[6] U.S. Energy Information Administration, "Form EIA-860 annual electric generator report," U.S. Energy Information Administration, Washington, D.C., USA, Tech. Rep., 2024. [Online]. Available: https://www.eia.gov/electricity/data/eia860/

[7] Texas A&M University. Texas A&M electric grid test case repository. [Online]. Available: https://electricgrids.engr.tamu.edu

[8] S. Dzakpasu, T. J. Overbye, and J. Snodgrass, "Visualization and analysis of united states independent system operator generator interconnection requests," in *Proc. North American Power Symposium (NAPS)*, Hartford, CT, Oct. 2025.

[9] Federal Energy Regulatory Commission, "Reactive power requirements for non-synchronous generation," Federal Energy Regulatory Commission, Order No. 827, 155 FERC ¶ 61,277, Jun. 2016. [Online]. Available: https://www.ferc.gov/media/order-no-827