Undergraduate Research on Big Data Analysis: Towards Large-Scale Electric Vehicle Integration Studies

Lyric Haylow, Diana Wallison, Jonathan Snodgrass, Thomas Overbye
Department of Electrical and Computer Engineering
Texas A&M University
College Station, TX
{lyrich237339, diwalli, snodgrass, overbye}@tamu.edu

***Abstract*— With the increase in electric vehicles (EV) and heavy-duty electric vehicles (HDEV) on the roads today, an appropriate increase in charging infrastructure needs to be made to keep up with the growing load demand they're making. To appropriately plan for developing new charging infrastructure, preliminary studies with synthetic grids that can accurately simulate how the electric grid may react are necessary. However, previous synthetic models have been insufficient as they're either too restricting or assume ideal conditions for distribution or transmission grid aspects. With new large-scale combined transmission and distribution (T&D) grid models, the impact of both EVs and HDEVs can be accurately measured, and corresponding changes to the grid can be ensured. In a recent study to test how the I45 highway from Dallas to Houston within the U.S. state of Texas can withstand 96 test case scenarios, full T&D synthetic model simulations resulted in roughly 67.9 GB of resultant data. This paper aims to present methods for aggregating, summarizing, and analyzing the optimal power flow (OPF), generator cost, least marginal price (LMP), and overload results into easily interpretable sizes.**

***Index Terms*-- Electric Vehicle, Synthetic Networks, Charging Infrastructure, Electric grid infrastructure.**

I. INTRODUCTION

Traditional synthetic grid models that simulate how the electric grid reacts at any given time to different factors are insufficient for analyzing how electric vehicles (EV) and heavy-duty electric vehicles (HDEV) may affect the electric grid. With modern incentives toward a clean energy future, EVs are a natural step in the right direction. They're the fastest-growing car in the U.S. [1, 2], and their unique appeal as a vehicle that doesn't rely on environmentally detrimental fuel has resulted in increased governmental interest in their success. The replacement of a fuel source dependent on gas with an electric power option does lead to an increase in load, however, and that's even more noticeable with HDEVs. As the number of EVs increases in size and number, the resulting load demand will increase, which will encourage future challenges towards maintaining their popularity. One of the significant challenges of electrifying heavy-duty vehicles is the limited charging infrastructure [3]. While there are definitively significant incentives for transitioning away from conventional gas-powered vehicles, the cost of building additional supporting infrastructure necessary for both light-duty and heavy-duty EVs is daunting. Some examples would be appropriate increases in charging stations, electric grid line capacity, and generated power.

Preliminary studies into how EVs affect the electric grid are necessary to quantify and locate the developments required for EVs to remain successful. The cost of building new equipment necessary for an EV to function isn't insignificant, and investors don't want to waste their money. Accordingly, several research groups have begun developing highly synthetic combined models that can run large-scale combined T&D simulations [4]. These models are needed to evaluate how EVs will affect the electric grid, as they'll be able to accommodate the harsh challenges of simulating what amounts to the load of a house on wheels. A natural consequence of the thoroughness a top-down analysis achieves, especially on the effect EVs and HDEVs will have, is the amount of resultant data generated. There are many critical variables to consider that have an absolute impact on the electric grid, and it's essential that a large amount of data is generated to ensure results are promising. The author is focused on large data analysis and condensing the large amounts of information generated into summarized sections that can aid in troubleshooting the model's functionality and ascertaining how the results can be interpreted. The main focus of this paper is to discuss the practices used in simplifying a project's generated data into interpretable portions, as well as give examples of actual results.

The rest of the paper is structured as follows: Section 2 covers the motivation for the pursuit of developing a combined T&D model and the aspects regarding how it's uniquely suited to analyzing how EVs and HDEVs affect the electric grid. The methodology proposed for large data analysis, as well as examples of how they're applied and their results, is covered in Section 3. Lastly, a summary is provided in Section 4.

II. MOTIVATION

To assess the importance of current studies regarding HDEVs and explain the novelty and justification for the extent of their work, there are three questions that will be addressed. The first question to address would be why there is an intense focus on EVs. Firstly, the transition to EVs is happening, and their allure will likely stay strong for the foreseeable future. Their elimination of CO2 is a strong point in their favor, as they help with future governmental projections for greenhouse gas emissions to be net-zero by 2050 [5], therefore giving them strong governmental support. They also help reduce smog in cities, as they're primarily centralized around urban centers for their charging stations [6]. They're also attractive in performance, as they're proven to be a fun and smooth car to drive and relatively cheaper than the average gas-powered car. The first Tesla vehicle was built primarily as a sports car, with the Tesla Roadster able to contend with other popular sports vehicles at the time, and they've maintained the car's smooth performance, making it even more likely that EVs will have a continued future. Perhaps the most pertinent consideration for EVs would be their ability to save power generated by renewable energy sources and to supply electricity to the grid in an emergency [7]. With the implementation of two-way chargers and the ability to change the charging rate of EVs, they can effectively work as remote batteries capable of supporting the electric grid in times of emergency. They also help flatten the load curve for the electric grid, as most EVs are primarily charged at night rather than the day, and they can function as storage for excess energy generated from renewable energy [8, 9].

The second question the scientific community may ask is why there is an insistence on running combined T&D simulations to test how EVs and HDEVs affect the electric grid. The answer to that relates directly to the fact that simulations on EVs are being run. Traditionally, to simulate the effect that light-duty EVs would have on the electric grid, there would be a distribution level analysis, with the assumption that the transmission grid would work under ideal conditions. That would work for that scenario, considering that most average EVs would be charged at locations such as houses or central parking locations. These locations are typically fed from distribution nodes, so assuming ideal conditions for the transmission grid would be feasible. Conversely, the size and charging demand of HDEVs mean they can be connected to both high-power transmission nodes and smaller-scale distribution nodes [10]. If an accurate simulation of how both scales of vehicles will impact the electric grid is to be effective, then a combined T&D model is vital. Assuming the condition that either part of a T&D model as ideal no longer stands, since an analysis of both EVs and HDEVs simultaneously would affect any grid aspect that would be conventionally assumed as ideal. A failure to do so would introduce significant inaccuracies in projected grid stability, leading to faulty investment decisions and an underbuilt charging infrastructure. Considering these realities, a combined T&D grid model, such as the example shown in Fig. 1, is absolutely necessary for preliminary studies on how light EVs and HDEVs can affect the electric grid.

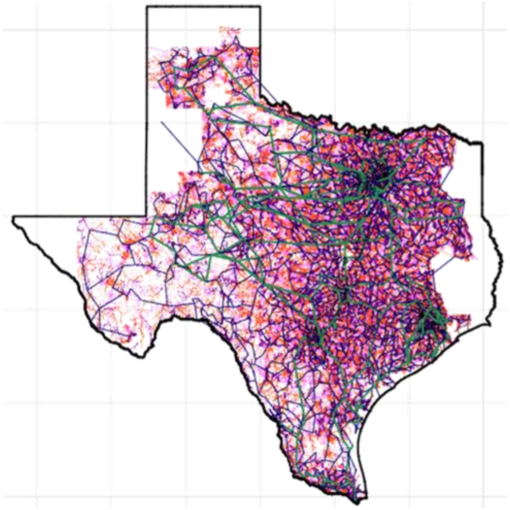

Figure 1: The syn-texas-TDgird-v02 synthetic T&D system of Texas [4].

The third and last question that may be asked pertains to the need for large-scale grid models. To answer this question, the full implications of attempting to accurately simulate the load that EVs and HDEVs introduce must be considered. With EVs, their location at any given time is not static, nor is it predictable where they may charge and introduce load. Unlike conventional load sources that connect to one location on the grid and have a relatively consistent demand for power, like houses, an EV can be plugged into the grid anywhere an individual finds themselves. When HDEVs are also incorporated into the same analysis, their use as long-range vehicles that routinely cross large distances makes it even more important for the scale of an attempted study to include a similar size. It's also necessary to account for the tremendous load that HDEVs introduce. Assuming a truck driver goes the average speed limit of 55 to 60 mph and adheres to federal law that they are limited to driving no more than 11 hours within a 24-hour time frame, that makes it around 605 to 660 miles they can traverse in a single day [11, 12]. Coupled with the assumption that HDEVs consume an average of 2kWh per mile [13], that calculates to 1210-1320 kWh of power they consume per working day. That's roughly the energy necessary to power a 1500 sqft house for an entire month. Considering that individual HDEVs have the mobile load equivalent of small neighborhoods, accurately ascertaining their exact impact on any part of the electric grid, despite significant distance changes on a regular basis, becomes imperative. In short, large-scale grid models need to be large to cover the possible area that both light EVs and HDEVs are likely to traverse and accurately reflect which parts of the grid will be affected by the massive load changes that EVs introduce.

III. METHODS

The goal is to provide meaningful conclusions about simulated results from T&D models or, at the very least, to provide the means by which those results can be summarized so that those conclusions can be drawn. Previous cases of combined T&D systems typically consisted of small models that could have the values they derived directly inspected to check for correctness [14-16]. With the new large-scale T&D models becoming available, the number of lines alone is typically in the millions, making the visual inspection method of results unfeasible. This also introduces the need for an automated method of debugging results.

These observations come from a recent project to develop a unifying co-simulation infrastructure integrating transportation demand, grid assets, land use, demographics, and emissions to optimally accelerate EV development and measure the impact of EV integration. Ninety-six electrification scenarios of urban and long-haul truck charging demand in the state of Texas in the U.S. were developed and integrated into the combined T&D simulation. The aim was to use these scenarios and their various parameters, shown in Table 1, to develop charging strategies to minimize operational, infrastructure, and environmental costs. A future paper will discuss the combined T&D simulation methodology shown in Fig. 2 in more detail, but in short, PowerWorld Simulator's OPF analysis is used for the transmission system, and OpenDSS is used for the unbalanced distribution system power flow. However, running a singular scenario through this T&D model generated roughly 317,000 files, consisting of folders with 24 hours of feeder details, branch details, bus and substation data, and more. That amounted to an average of 700 MB of files, and as a base case was necessary, that resulted in roughly 30.749 million files, with 67.9 GB of data between them. That's assuming the simulation for all 96 scenarios is run only once, and seeing as how this project was to see what changes could be made to alleviate the worst cases, the entire simulation was run many more times. All told, a few hundred million files were created, and at least a terabyte worth of data was generated. Simply put, there was no other option than to create an automated method of processing and debugging results.

Table 1: Scenario Parameters for integrating EVs and HDEVs into the synthetic Texas T&D grid.

Parameters	**96 Scenarios with five parameters**			
I-45 Charging Location	The midpoint between Houston and Dallas		Near Dallas	
Charging Logic	Upon Arrival		Starting at Midnight	
Season	Peak (Summer)		Shoulder (spring/fall)	
Charging Rate	100 kW	200 kW	300 kW	
EV market adoption rate	25%	50%	75%	100%

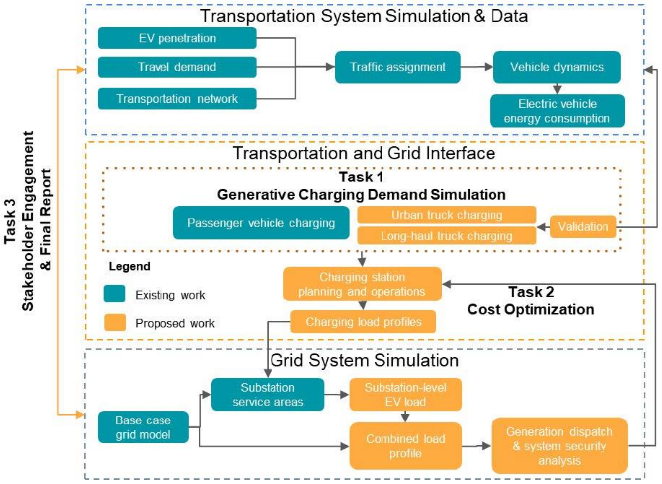

Figure 2: Flowchart of the methodology for the combined T&D EV simulation used.

For processing, there are three overarching steps that need to be taken, and some advice that can help along the way. Since the 96 scenarios are a niche case and unlikely to apply in the same ways to other results from T&D simulations, the steps are general enough to be applicable in alternate situations but also produce informative products. The first step lies in identifying the most immediately relevant information. While all generated results are likely meaningful in some way or another, that does not mean they're pertinent to whatever end result is being looked for. From the ten folders created for each scenario, only 4 of them, namely the optimal power flow (OPF), least marginal price (LMP), generator cost, and overloads folders, were used. Considering that the goal was to measure the impact of HDEVs, the relevant aspects were how power flow was affected around the state, the increased cost of generators and electricity prices, and the immediate overloads created by the introduction of the HDEVs. Other created files, such as branch and bus details, logs involved in processing, and more, were deemed irrelevant to the study and put aside. From this step alone, 159,320 files were set aside, meaning almost a 50% reduction in total data size. A variation of expected files reduced across different test cases is expected, yet any reduction in processing is beneficial, and the benefit scales with the amount of data needing processing.

The following two steps pertain to compressing the dataset to two different levels of analysis. They address the need to troubleshoot potential errors in processing and identify general trends that can prove meaningful when drawing conclusions. The second step, and first level of analysis, involves aggregating all resultant and relevant data from a single category into singular files per scenario. This compresses the relevant data further and makes it available for visual inspection. For example, within the HDEV project, the T&D model simulates 24 hours of the expected effect for each scenario. That means the folders of relevant data chosen are divided into 24 sections, each with their expected impact within that hour. For solely the overloads folder, 24 subfolders for each hour were created, and each hourly folder held more than 6000 files. Of the data inside each, only the value of how many amps a line or transformer was overloaded was considered relevant.

Table 2: Snippet of the second layer of analysis of simulated overloads

Line	MVAOver_1	MVAOver_2	MVAOver_3	MVAOver_4	MVAOver_5	MVAOver_6...
line.l(p1lv-p3ulv)	0.0018	0	0	0.005	0.0007	0.0038
line.l(p4lv-p5ulv)	0.007	0	0	0	0	0.0017
line.l(p2lv – p3ulv)	0.0046	0.0017	0.0021	0.0023	0.0032	0.0070
line.l(p1lv – p4ulv)	0.0046	0.0017	0.0021	0.0024	0.0032	0.0070
line.l(p6lv – p7ulv)	0.0183	0.0128	0.0135	0.0140	0.0156	0.0225
line.l(p2lv – p7ulv)	0.0039	0.0011	0.0015	0.0018	0.0026	0.0062

Any element that wasn't over 100% load for that specific part was no longer needed, as only the number of overloaded lines and transformers was significant. As shown in Table 2, merging the 24 hours' worth of overloads with consideration of individual line connections allows for evaluating any particularly overloaded lines and any potential errors from running the simulation. The second column describes the beginning and end buses, effectively the name of singular lines. The last 24 columns, of which six are shown as an example, show the number of megavolt-amperes over the limit that line is at for that hour. For the other three chosen categories considered relevant for the HDEV project, the load in M.W. at a substation was taken from the OPF folder, the generator production cost for an individual substation from the Generator_Cost folder, and the bus marginal cost per megawatt from the LMP folder. This first level of analysis with these given values allowed for an interview of how each scenario affected the electric grid. Any likely computational error can be caught by this stage as well, so long as values are questioned whether they are justifiable.

The third step in summarizing large data analysis consists of combining all scenarios from the first layer of analysis per category. For instance, all 96 scenario files for tracking how OPF changes are combined into one large table that can track hourly change. This second layer of analysis allows for a comparison of how each scenario relates to the other. Within the HDEV project, this method enabled a direct comparison of different scenarios to ascertain which parameters affect the electric grid the most. Note that the process of combining the differing scenarios depends on the category being handled. For the generator production costs, all the costs across all generators were summed together under each hour.

IV. SUMMARY

The incorporation of heavy-duty electric vehicles (HDEV) into the transportation network will require extensive growth of charging infrastructure, and to ensure that future investments into such are smartly made, it's crucial to accurately predict the necessary locations for new development. To that end, a unifying co-simulation infrastructure integrating transportation demand, grid assets, land use, demographics, and emissions to optimally accelerate electric vehicle (EV) development as well as measure the impact of EV and HDEV integration was developed for a local project. This would be a large-scale combined transmission & distribution (T&D) simulation, and the results would be to develop charging strategies that minimize operation, infrastructure, and environmental costs for the U.S. state of Texas. Given five parameters regarding charging rate, season, EV market adoption rate, charging logic, and initial charging location, 96 urban and long-haul truck charging demand electrification scenarios were developed and integrated into the combined T&D simulation. Generated data for such a large-scale model resulted in roughly 30.749 million data files being compiled, so a methodology was developed to automate processing and debugging results.

Three general steps were taken for the method developed for aggregating, summarizing, and analyzing the resultant data. The first would be to decide on relevant data for the study being made. Within the HDEV project, this step allowed for an almost 50% reduction in necessary files. The second step involves aggregating all relevant data from singular categories into individual scenarios in tables. This allows for troubleshooting for potential processing errors and makes visual inspection marginally easier. Since the HDEV project was concerned with the impact of EVs on the grid, one of the categories selected was the optimal power flow at bus nodes. As an example, by organizing a table of all bus nodes on one axis, and a 24-hour scale on the other axis, the change in M.W. of load over time and location could be seen. The third step involves further combining each scenario summary file into a final table per category. This allows for comparison of how each scenario, and hence the different parameters between them, could be related to another. General trends can also be observed at this point in time. From this last step, a resultant five files were made that gave accurate representations of how the electric grid simulation withstood HDEV integration.

An important note would be that it's highly encouraged for parallel processing of separate scenarios to be implemented. As each scenario's results were individual, this allowed for a division of the automated processing, cutting computational time from an average of 10 hours to an average of 50 minutes. Future work will focus on the methodology for the combined T&D simulation discussed in this paper. Additional work will also focus on long-term grid planning to accommodate future EV scenarios and incorporating newly built EV charging stations into synthetic grids used for research. Overall within this paper though, an efficient and automated method for processing was developed for aggregating, summarizing, and analyzing large amounts of data.

ACKNOWLEDGMENT

This work is supported through funding provided by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) Grant No. DE-EE0009665, administered through a sub-award from ElectroTempo. The authors gratefully acknowledge ElectroTempo for the EV charging datasets and their guidance on the project.

REFERENCES

[1] "Americans Buy Nearly 1.2 Million Electric Vehicles to Hit Record in 2023, According to Latest Kelley Blue Book Data." Blue Book. https://mediaroom.kbb.com/2024-01-16-Americans-Buy-Nearly-1-2-Million-Electric-Vehicles-to-Hit-Record-in-2023,-According-to-Latest-Kelley-Blue-Book-Data#:~:text=Electric%20vehicles%20(EVs)%20represent%20the,up%20from%205.9%25%20in%202022. (accessed July 15, 2024).

[2] "Trends in electric cars." IEA. https://www.iea.org/reports/global-ev-outlook-2024/trends-in-electric-cars (accessed July 15, 2024).

[3] I. Mahmud, B. M. Medha, and M. Hasanuzzaman, "Global challenges of electric vehicle charging systems and its future prospects," Research in Transportation Business & Management, vol. 49, p. 101011, 2023, doi: 10.1016/j.rtbm.2023.101011.

[4] C. Mateo et al., "Building and validating a Large-Scale combined transmission & distribution synthetic electricity system of Texas," International Journal of Electrical Power & Energy Systems, vol. 159, 2024, doi: 10.1016/j.ijepes.2024.110037.

[5] (2021). The Long-Term Strategy of the United States: Pathways to Net-Zero Greenhouse Gas Emissions by 2050. [Online] Available: https://www.state.gov/tackling-the-climate-crisis-together/longtermstrategy-3/

[6] R. Challa, D. Kamath, and A. Anctil, "Well-to-wheel greenhouse gas emissions of electric versus combustion vehicles from 2018 to 2030 in the US," *J Environ Manage*, vol. 308, p. 114592, Apr 15 2022, doi: 10.1016/j.jenvman.2022.114592.

[7] J. L. Wert, F. Safdarian, D. Wallison, J. K. Jung, Y. Liu, and T. J. Overbye, "Spatiotemporal Operational Emissions Associated With Light-, Medium-, and Heavy-Duty Transportation Electrification," IEEE Transactions on Transportation Electrification, vol. 10, no. 1, pp. 859-874, 2023, doi: 10.1109/TTE.2023.3275050.

[8] K. Mladen and H. Abu-Rub, "Flexible Integration of EVs and P.V.s into the Electricity Grid," presented at the Qatar Foundation Annual Research Conference Proceedings, 2016.

[9] R. Jovanovic, S. Bayhan, and I. S. Bayram, "A multiobjective analysis of the potential of scheduling electrical vehicle charging for flattening the duck curve," Journal of Computational Science, vol. 48, 2021, doi: 10.1016/j.jocs.2020.101262.

[10] B. Borlaug *et al.*, "Heavy-duty truck electrification and the impacts of depot charging on electricity distribution systems," *Nature Energy,* vol. 6, no. 6, pp. 673-682, 2021, doi: 10.1038/s41560-021-00855-0.

[11] "Summary of Hours of Service Regulations." https://www.fmcsa.dot.gov/regulations/hours-service/summary-hours-service-regulations (accessed July 15, 2024).

[12] R. El Helou, S. Sivaranjani, D. Kalathil, A. Schaper, and L. Xie, "The impact of heavy-duty vehicle electrification on large power grids: A synthetic Texas case study," Advances in Applied Energy, vol. 6, p. 100093, 2022.

[13] S. Kothari. "Tesla Semi: Everything we know in March 2024." TopElectricSUV. https://topelectricsuv.com/news/tesla/tesla-semi-all-we-know-feb-2022/ (accessed July 15, 2024).

[14] K. Balasubramaniam and S. Abhyankar, "A combined transmission and distribution system co-simulation framework for assessing the impact of Volt/VARcontrol on transmission system," Chicago, IL, USA, 2017.

[15] T. A. Papadopoulos, G. A. Barzegkar-Ntovom, E. O. Kontis, and E. A. Koukoulianta, "Combined Transmission and Distribution Test System for Small-Signal Stability Analysis: Initial Results," presented at the 2022 57th International Universities Power Engineering Conference (UPEC), 2022.

[16] C. González-Morán, P. Arboleya, R. R. Mojumdar, and B. Mohamed, "4-Node Test Feeder with Step Voltage Regulators," International Journal of Electrical Power & Energy Systems, vol. 94, pp. 245-255, 2018, doi: 10.1016/j.ijepes.2017.06.02