Additional Insights in Creating Large-Scale, High Quality Synthetic Grids: A Case Study

Farnaz Safdarian, Adam B. Birchfield, Komal S. Shetye, Thomas J. Overbye
Dept. of Electrical and Computer Engineering
Texas A&M University
College Station, TX, USA
fsafdarian@tamu.edu, abirchfield@tamu.edu, shetye@tamu.edu, overbye@tamu.edu

***Abstract*—In this paper, a synthetic network has been developed to generate a fictitious but realistic power system model and improved based on the actual generation data on the Midwest U.S. footprint with capability to represent characteristic features of actual power grids, without revealing any confidential information. This synthetic network model is available online and can be shared freely for teaching, training, and research purposes. Geographic Data views (GDVs) and validation metrics derived from the North American Eastern Interconnect show the effectiveness and authenticity of the developed grid.**

Keywords—large scale synthetic grids, power system characteristics, load data, generation data, transmission system

I. INTRODUCTION

A gap between academia and industry in the field of power engineering comes from the fact that data of actual power systems are confidential and access to those are restricted even for research purposes. There is often a lack of access for many researchers to some information concerning the actual electric grid and its associated data. For example, the US power flow cases and several structural information of the actual grid are considered to be critical energy/electricity infrastructure information (CEII) [1] with restricted availability. However, in order to improve modern power system models, operation and planning optimization problems such as power flow, economic dispatching, unit commitment and generation expansion planning, complex electricity market models with emerging distributed energy resources, dynamics and transient stability studies, geomagnetic disturbance studies, and advanced algorithms, there is a strong need for access to diverse, large and complicated power systems that are available for research and publications.

Several IEEE test cases are established and widely used to represent a portion of the American Electric Power System (in the Midwestern US) [2]. A test system proposed in [3] is based on structural attributes and data from the ISO - New England. Reference [4] develops an approximate model of the European interconnected system using actual transmission networks to study the effects of cross-border trades. However, until recently, there was limited work focusing on the creation of complicated and realistic synthetic large-scale power system models using publicly available data that can mimic the full complexity of modern electricity grids for more accurate power system studies. Our previous work [5-7] introduced creation of synthetic power systems based on census data [8] and U.S. Energy Information Association (EIA) generation data [9], which span on the actual geographic footprints and provided realistic test cases for power system studies without revealing any sensitive information. Additional complexities can still be added into synthetic models to extend their applications.

In this paper, a synthetic network has been developed to generate a fictitious but realistic power system model based on the actual generation data that are statistically similar to the actual U.S. power system on the Midwest U.S. footprint with capability to represent characteristic features of actual power grids, without revealing any confidential information. This synthetic network model is available at [10] and can be shared freely for teaching, training, and research purposes.

II. BACKGROUND

The synthetic grid models are created using metrics derived from the North American Eastern Interconnect (EI) and publicly available data, provided by the U.S. Census and Energy Information Administration. Reference [6] outlines fundamental steps for the creation of synthetic power system models including geographic load, generator substations, and assignment of transmission lines.

The overall approach for building these networks is summarized below, and described more in [6]:

1. Substation Planning

This step includes locating and sizing load and generator substations using public data and statistics derived from the U.S. electric grids. Considering humans as the primary consumers of electricity, population data in geographic latitude and longitude coordinates is the main base of synthetic loads. Additionally, publicly available generator data from the U.S. EIA, which includes data for all generators in the U.S., is used to site and size generators in the synthetic case.

Then, a clustering technique is employed, which ensures the synthetic substations meet realistic proportions of load and generation, among other constraints. Within each substation, loads are usually connected to the lowest voltage level, and generators are often connected to the highest voltage level through a generator step up (GSU) transformer. In addition, transformers are added in each substation to connect multiple nominal voltage levels.

2. Transmission Planning

This step is the most challenging and computationally expensive step. First, transmission line electrical parameters are calculated based on the assigned voltage levels and the percentages of substations with each voltage level as well as line length, which are determined based on the distance between substations. Conventions employed by grid planners are considered as metrics to make the parameter selections more realistic. Reference [7] presents a methodology to generate synthetic line topologies with realistic parameters. This step includes several structural statistics to be used in characterizing real power system networks, including connectivity, Delaunay triangulation overlap, DC power flow analysis, and line intersection rate and considers N-1 contingencies to improve the system reliability.

3. Reactive Power Planning

At this stage, AC power flow solvable synthetic cases are created, with varying network sizes and complexities. Then, the test cases are augmented with additional complexities like voltage control devices such as switched shunts.

4. Key Considerations and Challenges

Key challenges include geography constraints such as lakes, mountains, and urban areas, as well as network topology parameters, increasing power flow feasibility in base and N-1 contingency conditions, intractability of n^2 possible combinations of branches (where n is the number of buses), many competing metrics to meet, and consideration of contingency conditions that increases computation even more. References [11, 12] present some metrics that are extracted from the North American Eastern Interconnect for validating synthetic grids for achieving realistic data sets.

III. LARGE-SYSTEM EXAMPLE

A synthetic grid is created in the US Midwest footprint mainly on Midcontinent Independent System Operator (MISO) and Southwest Power Pool (SPP) coverage area based on the method explained in the previous section. All simulations are carried out using PowerWorld simulator [13] on an Intel(R) Core(TM) i7 2.59 GHz laptop with 32GB of RAM. The case is built as a 27,163-bus case for power flow studies and general analysis and research purposes. The number of buses are estimated based on the size of the grid. The geography is complex and diverse in terms of vegetation and civilization, and includes 19 areas, divided by U.S. states. The transmission network is built with seven voltage levels: 500 kV, 345 kV, 230 kV, 161 kV, 138 kV, 115 kV and 69 kV. These voltage levels and percentages of substations including each voltage level is approximated from [14].

Figure 1 shows a one-line diagram of the case with 500 kV and 345 kV highlighted with thicker lines. Orange shows 500 kV lines and red shows 345 kV lines, blue shows 230 kV, blach refers to 161 kV, 138 kV and 115 kV lines and green shows 69 kV lines. Figure 2 shows nominal voltage of transmission lines over the number of buses, and Table 1 gives a summary of the case.

TABLE I. A SUMMARY OF CASE STATISTICS OF 27-BUS SYSTEM

Number of substations	13,074
Number of buses	27,163
Number of areas	19
Number of transmission lines	28,550
Number of transformers	10,651
Number of loads	14,054
Number of generators	4,224
Number of shunts	1,961
Total design load (GW)	154 GW

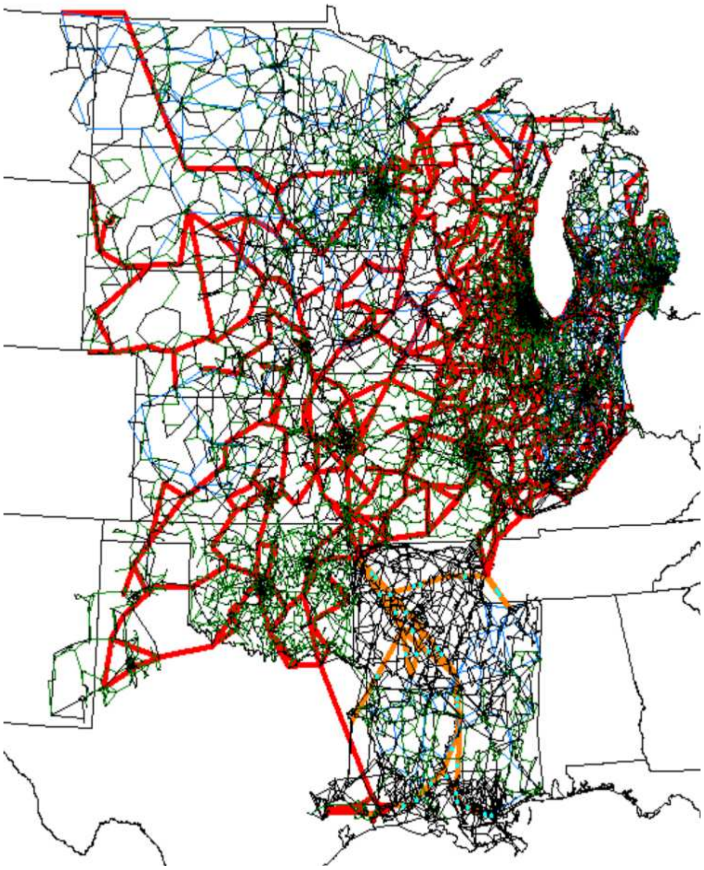

Fig. 1. One-line Diagram of the 27K-Bus Case a figure caption.

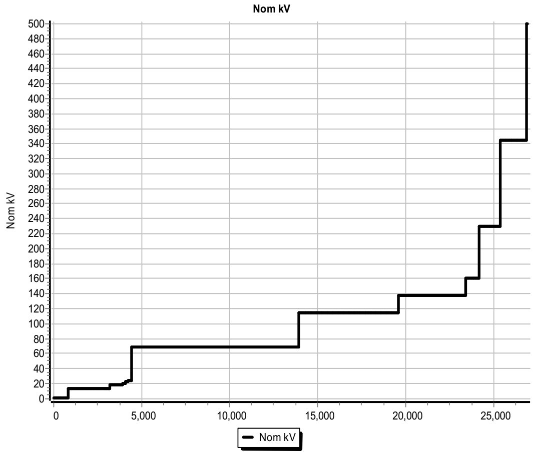

Fig. 2. Nominal voltage of transmission lines over the number of buses of the 27K-bus case.

The types and number of generators in the case is extracted from EIA-860. All generator with a capacity larger than 8 MW are included in the case. Table II shows a summary of the number and overall capacity of generators grouped by fuel types.

TABLE II. THE TYPE AND NUMBER OF GENERATORS

Fuel Type	Number of Units	MW Capacity
Natural Gas	1,268	128,513
Coal	187	65,296
Nuclear	27	27,106
Wind	427	46,398
Solar	30	735
Hydro	163	7684
Petroleum	101	3,917
Other	204	34,130
Total	2,407	313,779

Figure 3 shows the geographic data view (GDV) [15] of generators. The size of ovals is proportional to the MW capacity of the units and the colors show the fuel types of generation substations where black refers to coal, brown to natural gas, green to wind, yellow to solar, dark blue to hydro, red to nuclear, dark magenta to petroleum and brown to other types such as biomass.

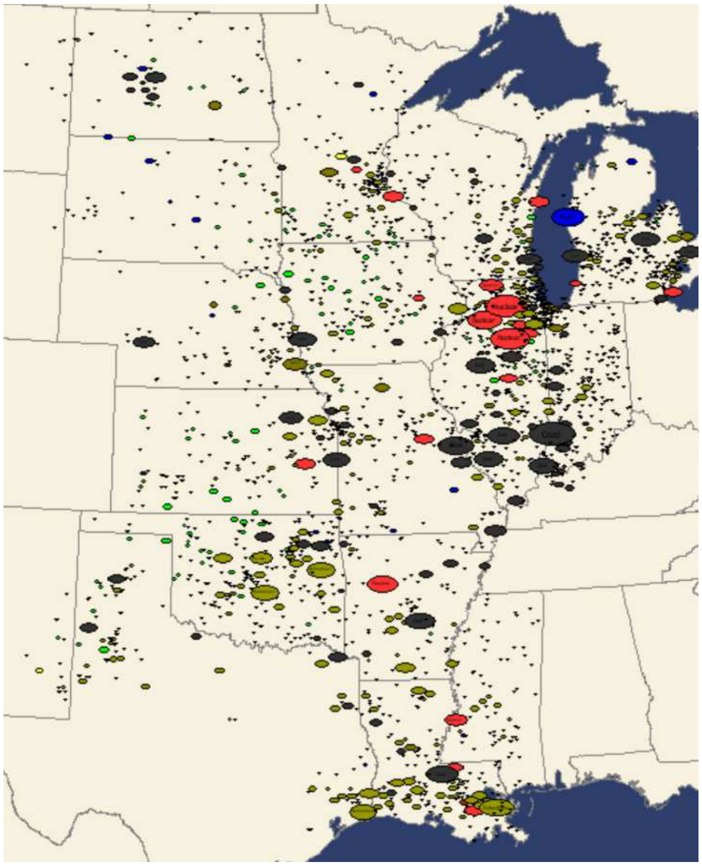

Fig. 3. MW capacity and fuel type of generation substations of 27k- bus system.

Figure 4 shows the set points of generating units where the size of rectangles is proportional to the output active power in MW and color is proportional to output reactive power in MVAR. Dark red shows reactive power set points near maximum and dark blue shows reactive power of generators near minimum.

As it can be observed from Figure 4, most generators are not at their maximum or minimum MVAR limits. This gives the generators more freedom to regulate the voltages of nearby buses by injecting or consuming reactive power.

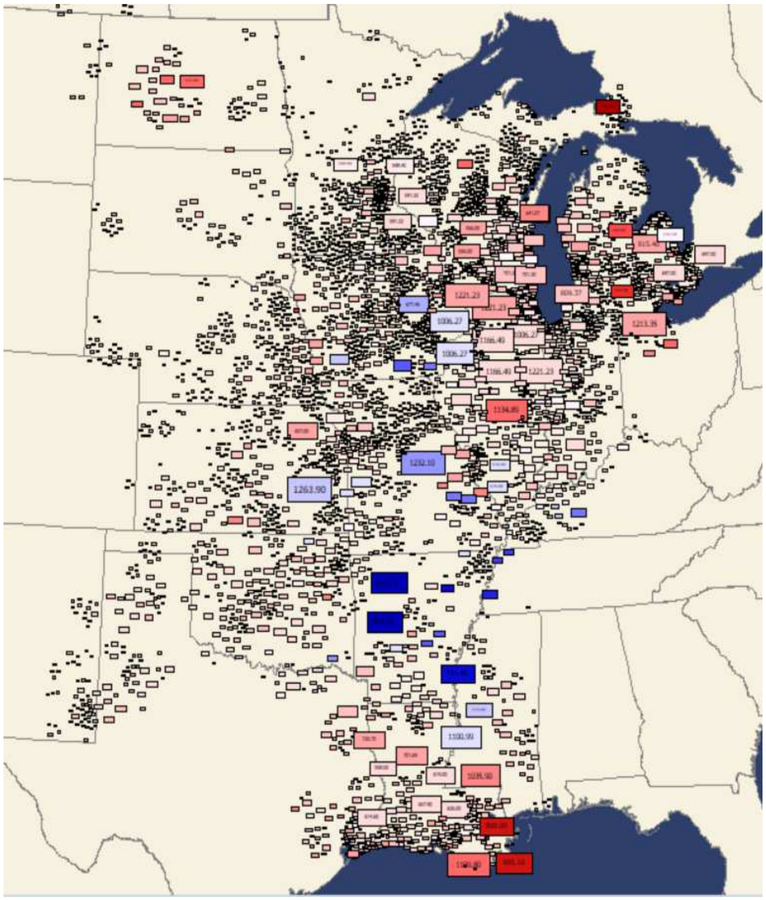

Fig. 4. Generator’s real and reactive power output of 27k-bus system.

As explained in the previous section, load is placed at every location, with a MW and MVAR amount proportional to the population. Statistics derived from the Eastern Interconnect are used to calculate values for the amount of load consumed per person. Also, industrial load is added to the locations with a high share of industrial load. Finally, a load substation is placed for every load and sized using the same MW and MVAR values for the load it is connected to. To complete the substation, a voltage level is assigned based on typical values for the region the substation resides in and its MW load. Additionally, a step-down transformer is assigned with electrical parameters that are based on median values for Eastern Interconnect transformers that share the same voltage level. Figure 5 shows load substations where their size is proportional to the load in MW. As it can be observed, the load size varies in different areas and the goal of this GDV is to match the load size and its variation to the actual cases.

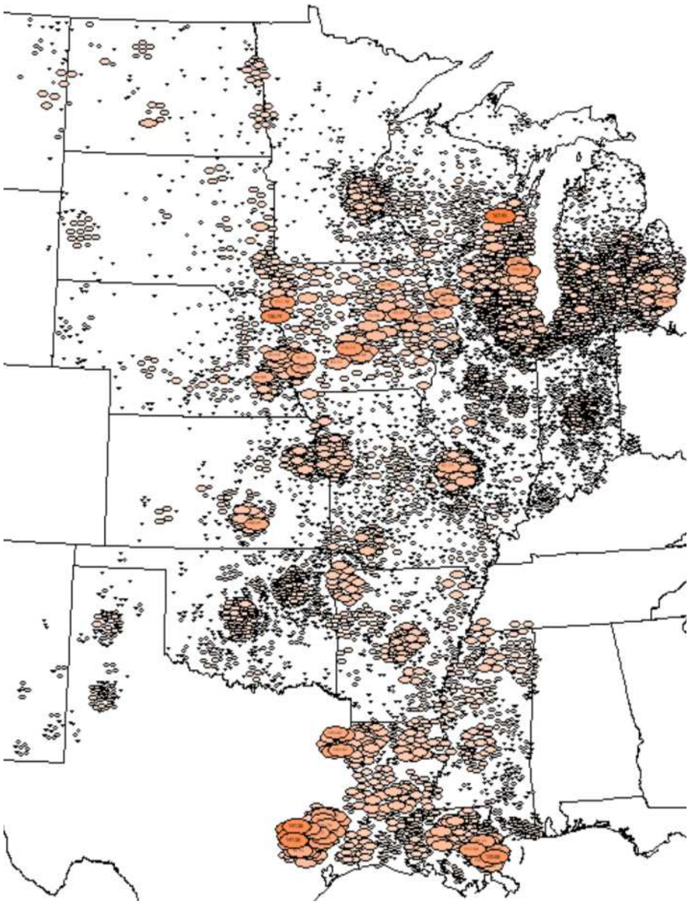

Fig. 5. Load substations of 27k-bus system.

Another important challenge with improving this case was to match the overall power flow patterns across the geographical footprint of U.S. Midwest.

In order to have an organized view of the flows, GDV summary objects are used [15]. These GDVs group the electric grid objects geographically and show the summary GDVs based on a vertical and horizontal grid covering the entire system footprint. Such summaries could be used with actual values or with the previously mentioned differences between solutions.

Figure 6 shows the net flows between areas. The size and thickness of arrows is proportional to the size of power flow, and size of rectangles are proportional to the net injection to the areas in MW, where magenta refers to net import and yellow refers to net export.

Figure 7 shows the overall flow pattern of electricity using a 6 by 10 grid of GDV summary objects on the 27K bus system. The size of each GDV is proportional to the net injection where magenta refers to net import and yellow refers to net export. The arrows show the direction of flow and the thicker and darker they are, the more active power flow between GDV summary objects.

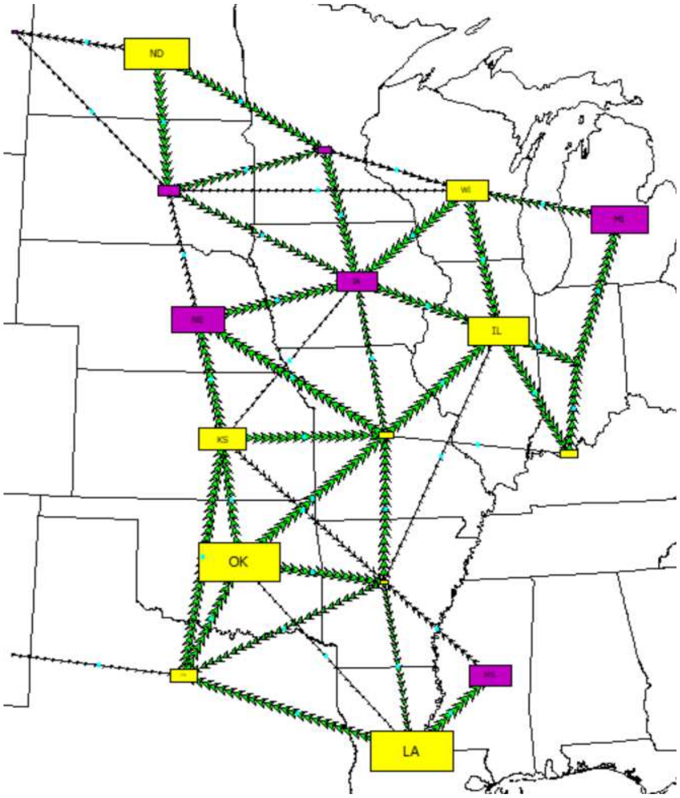

Fig. 6. Area GDV with interfaces of 27k-bus system.

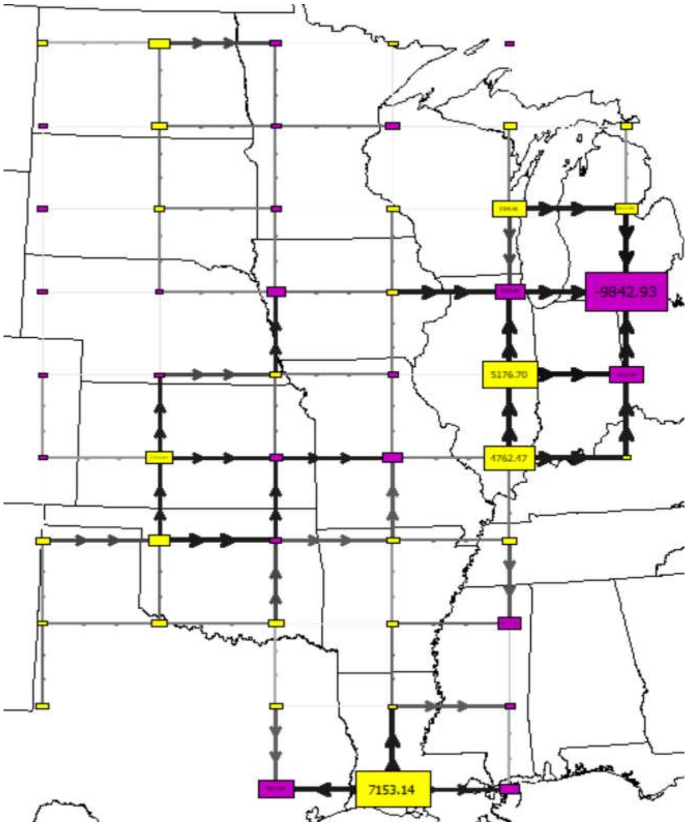

Fig. 7. Grid-based visualization of 27k-bus system.

These GDVs are very important to visualize and validate with actual flows. Before the industrial load adjustments, the power flow was as shown in Figure 8, i.e. a large export from the south to the north, which is not representative of what happens in the actial grids. Comparing these GDVs with real cases, it was realized that flow patterns of Figure 8 were not realistic and industrial load should be also added tp some areas including southern areas. This is an interesting example of how GDVs helps with the situational awareness and validating the cases.

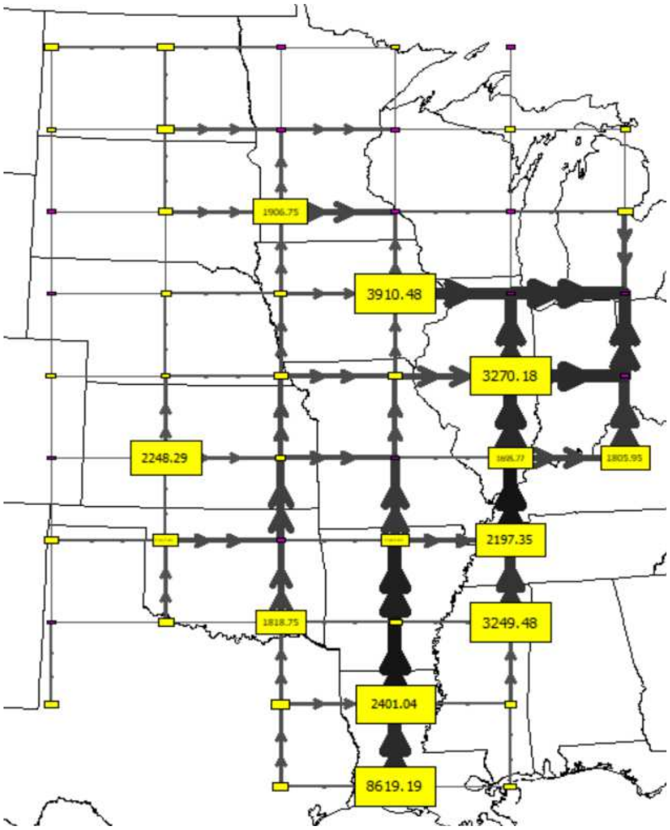

Fig. 8. Grid-based visualization of 27k-bus system.

An important challenge with adjusting load and flow patterns was to have a solvable ACPF problem without any mismatches in the nodal power balance and avoid getting alternative solutions. In general, the power flow equations can have normal solution, and a potentially alternative solution. Alternative solutions (also called low-voltage solutions) may occur when initial state of voltage magnitudes of some buses are close to the other side of curve and this is not a desired solution [16, 17]. Figure 9 shows very low voltages in red.

The initial state of the system is very important for achieving a feasible solution. If initial voltage levels are violated lower than 0.9 per unit (pu) or higher than 1.1 pu, alternative solutions may occur. In order to avoid low voltage solution, the initial state can be changed to a flat start where all voltage magnitudes are changed to 1 PU and if any change need to be applied in the grid, such as increase or decrease in the load, it should be applied in small steps and voltage magnitude of buses should be regulated after each step with the use of reactive power control devices such as switched shunts.

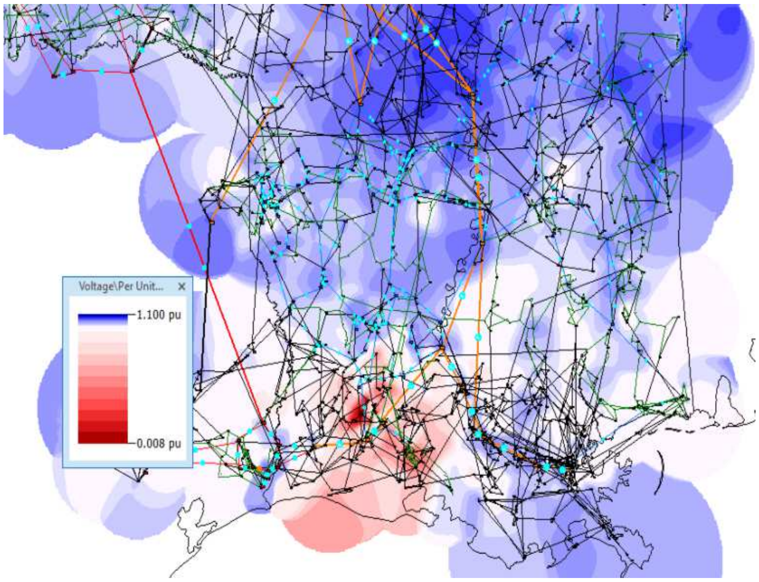

Fig. 9. Showing a low-voltage solution of 27k-bus system.

Another development to the 27K grid is considering a price responsive demand model [18]. The load benefit model is considered a piecewise linear model, and each load entity has four offer steps. Up to 3-10% of the load is assumed to have high prices up to 3000 $/MWh which may be shed. From 3-10% to around 25-40% of each load has a price up to 180 $/MWh. From 25-40% to around 50-75% of each load has a price up to 140 $/MWh and the last step has a price up to 10 $/MWh. Generators' cost is determined based on actual costs depending on their fuel types and locations. Following curves in Figure 10 shows the load marginal benefit curve and incremental cost curve of the whole area.

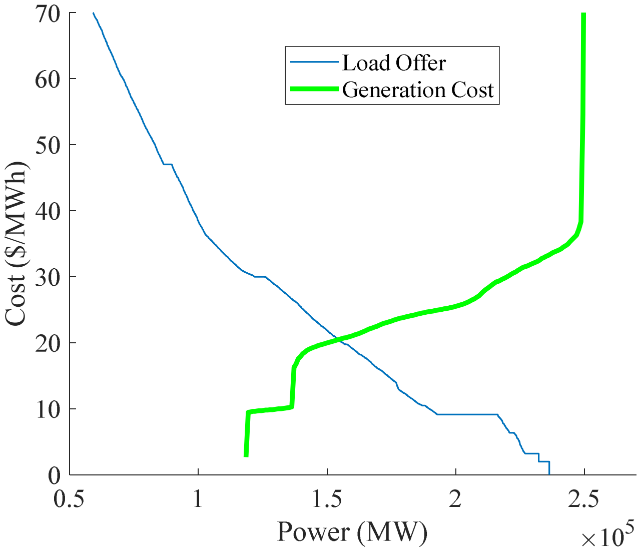

Fig. 10. Super area incremental cost and marginal benefit curves of 27k-bus system.

Table III shows some validation metrics of grid proportions, generators, load, and substations of the 27K bus case. Validation metrics are derived from the North American Eastern Interconnect.

TABLE III. Validation Metrics of the 27K bus System

Validation Metric	**Criteria**	27K Bus Case
Buses per Substation	Mean 1.7-3.5	2.08
Substations in kV Range	<200 kV, 85-100%	99.8%
	>201 kV, 7-25%	13%
Substations with Load	75-90%	89.5%
Load per Bus	Mean 6-18 MW	6 MW
Load Power Factor	Mean 0.93-0.96	0.96
Generator Substations	5-25%	12.9%
Generator MW Maximum Capacities	25-200 MW, 40+%	36%
	200+ MW, 2-20%	8.5%
Committed Generators	60-80%	79%
Shunt Capacitors and Reactors	10-25% of subs shunts 30-50% above 200 kV	12% 37%

IV. Conclusion

A large case study with around 27K buses on the geographical footprint in Midwest U.S. is created based on actual publicly available generation and census data, and is validated based on validation metrics achieved from the North American Eastern Interconnect and visualizations from creating GDVs. Visualizations help with the situational awareness of the grid. The comparisons with actual data show the effectiveness and authenticity of the grid for accurate research to improve modern power system models, operation and planning optimization problems, complex electricity market models with emerging distributed energy resources, dynamics and transient stability studies, geomagnetic disturbance studies, and advanced algorithms.

Acknowledgment

This work is partially supported through funding provided by the U.S. National Science Foundation (NSF) in Award 1916142, the U.S. Department of Energy (DOE) under award DE-OE0000895, the US ARPA-E Grant No. DE-AR0001366, and the Power Systems Engineering Research Center (PSERC).

References

[1] https://www.ferc.gov/enforcement-legal/legal/major-orders-regulations/critical-energyelectric-infrastructure-information

[2] R. Christie, "Power systems test case archive: 118 bus power flow test case," University of Washington, Department of Electrical Engineering,[Online], Available: www. ee. washington. edu/research/pstca/pf118/pg tca118bus. htm, 1993.

[3] D. Krishnamurthy, W. Li, and L. Tesfatsion, "An 8-zone test system based on ISO New England data: Development and application," IEEE Transactions on Power Systems, vol. 31, no. 1, pp. 234-246, 2015.

[4] Q. Zhou and J. W. Bialek, "Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades," IEEE Transactions on power systems, vol. 20, no. 2, pp. 782-788, 2005.

[5] T. Xu, A. B. Birchfield, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Application of large-scale synthetic power system models for energy economic studies," in Proceedings of the 50th Hawaii International Conference on System Sciences, 2017.

[6] K. M. Gegner, A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," in 2016 IEEE Power and Energy Conference at Illinois (PECI), 2016: IEEE, pp. 1-6.

[7] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," IEEE Transactions on power systems, vol. 32, no. 4, pp. 3258-3265, 2016.

[8] U.S. Census Bureau. Census 2019, ZIP Code Tabulation Areas Gazetteer File. https://www.census.gov/geo/maps-data/data/gazetteer2019.html

[9] U.S. Energy Information Association. Form EIA-860, 2019. http://www.eia.gov/electricity/data/eia860/index.html

[10] https://electricgrids.engr.tamu.edu/electric-grid-test-cases/datasets-for-arpa-e-perform-program/

[11] V. Krishnan et al., "Validation of synthetic US electric power distribution system data sets," IEEE Transactions on Smart Grid, vol. 11, no. 5, pp. 4477-4489, 2020.

[12] A. B. Birchfield et al., "A metric-based validation process to assess the realism of synthetic power grids," Energies, vol. 10, no. 8, p. 1233, 2017.

[13] PowerWorld » The visual approach to electric power systems (https://www.powerworld.com/)

[14] https://upload.wikimedia.org/wikipedia/commons/d/d4/UnitedStatesPowerGrid.jpg

[15] Overbye, T. J., Wert, J. L., Shetye, K. S., Safdarian, F., & Birchfield, A. B. (2021, February). The Use of Geographic Data Views to Help With Wide-Area Electric Grid Situational Awareness. In *2021 IEEE Texas Power and Energy Conference (TPEC)* (pp. 1-6). IEEE.

[16] K. S. Shetye, T. J. Overbye, A. B. Birchfield, J. D. Weber, and T. L. Rolstad, "Computationally Efficient Identification of Power Flow Alternative Solutions with Application to Geomagnetic Disturbance Analysis," in 2020 IEEE Texas Power and Energy Conference (TPEC), 2020: IEEE, pp. 1-6.

[17] T. J. Overbye and R. P. Klump, "Effective calculation of power system low-voltage solutions," IEEE Transactions on Power Systems, vol. 11, no. 1, pp. 75-82, 1996.

[18] Su, C. L., & Kirschen, D. (2009). Quantifying the effect of demand response on electricity markets. *IEEE Transactions on Power Systems*, *24*(3), 1199-1207.