A Methodology for the Creation of Geographically Realistic Synthetic Power Flow Models

Kathleen Gegner, *Student Member, IEEE*, Adam B. Birchfield, *Student Member, IEEE,* Ti Xu, *Student Member, IEEE*, Komal Shetye*, Member, IEEE,* Thomas J. Overbye*, Fellow, IEEE*
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Abstract— **Confidentiality concerns hinder research on the U.S. electric grid. To enable greater innovation, our research seeks to create entirely fictitious synthetic power system networks that capture the functionality, topology, and defining characteristics of the actual U.S. transmission system, and thus provide realistic test cases for research, without revealing any sensitive information. Creation of these models relies only on publicly available data and statistics derived from the actual grid. This paper outlines two fundamental steps for the creation of synthetic power system models: geographic load and substation placement and assignment of transmission line electrical parameters. We use public census data to build a set of fictitious load substations, allocated according to population, across real geographic regions of the United States.**

I. Introduction

One major problem in power systems research is the lack of public access to realistic electric transmission system models. These safeguards protect the United States' most sensitive infrastructure, but hinder innovation of the electric grid. Therefore, to bridge the gap between protecting U.S. national security and fostering innovation, the intent of this research is to build freely available, entirely synthetic transmission system models that are statistically similar to the actual U.S. transmission system and which span its geographic footprint, but are ultimately entirely fictitious; hence present no security concern. To do this, the synthetic models are created using statistics derived from an actual planning case for the North American Eastern Interconnect and publicly available data, provided by the U.S. Census and Energy Information Administration.

Realistic synthetic networks have wide reaching applications, including research in power system visualization, grid operation optimization, dynamics and stability studies, numerical algorithm testing, exploratory initiatives for planners and policy makers, geomagnetic disturbance studies, and educational purposes.

Previous work by Wang et al. [1] [2] developed a method for generating power system topologies with a modified small world algorithm. We expand on their technique to create topologically and geographically realistic networks using public data and geographic information, to site loads, generators, substations, and transmission lines.

II. Geographic Siting and Sizing of Loads, Generators, and Substations

The electric infrastructure in the United States has one overarching purpose, and that is to provide electricity to people. With humans as the primary consumer of electricity, it seemed logical to base synthetic loads on population data. To do this, we used publicly available 2010 U.S. Census data [3] that contains the population for each U.S. ZIP code and the latitude and longitude coordinate of each ZIP code center. Using this data provided a means to estimate load while capturing the geographic characteristics needed for making the synthetic case realistic. Additionally, publicly available generator data from the U.S. Energy Information Administration (EIA) [4], which includes data for all generators in the U.S., was used to site and size generators in the synthetic case.

A. Loads

Loads were considered first. A load was placed at every ZIP code center, with a MW and Mvar amount proportional to the population at each ZIP code. A 2012 NERC confidential planning model of the Eastern Interconnect and publicly available population data were used to calculate values for the amount of load used per person, found to be 2.01 kW and 0.574 kvar, shown in Fig. 1. Then, the population data for each ZIP code was scaled by these values, to determine the amount of load at each ZIP code, in MW and Mvar.

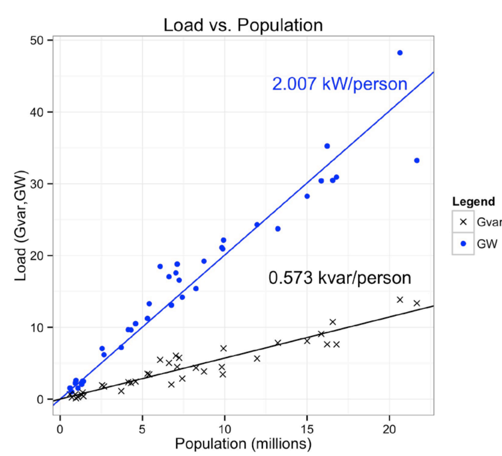

Fig. 1: Load, in Gvar and GW, compared to the total population in millions of people, for the 37 states in the Eastern Interconnect. [5]

Finally, a load substation was sited at every load calculated above and sized using the same MW and Mvar values for the load it is connected to. Step-down transformers were added at each substation, with electrical parameters determined by selecting median parameter values from a list of transformers at the same voltage level, in the Eastern Interconnect case.

This simple approach offers a good starting place for developing synthetic loads and their substations in the U.S. electric grid, as seen by the almost linear relationship in Fig. 2. Certainly, there are some issues at boundary population densities, where too many synthetic loads may be created in very low density, rural areas, and too few synthetic loads created in very high density, urban areas, but considering how varied the number of substations are in the actual grid, our approach still seems to make sense. Future work may explore patterns, if any, in how many urban and rural load substations are appropriate, based on population density.

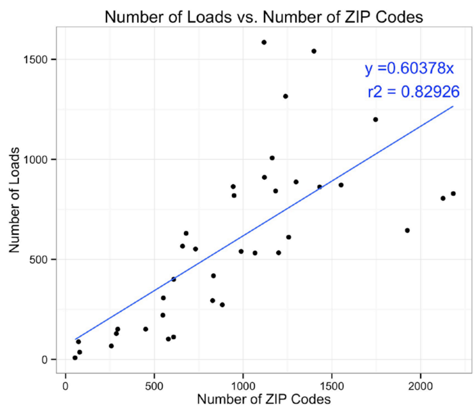

Figure 2: Total number of loads compared to the number of ZIP codes in each of the 37 states of the Eastern Interconnect. [5]

B. Generators

The U.S. Energy Information Administration (EIA) provides publicly available data for all electric generators operating in the U.S. electric grid [4]. We used this data directly to site and size generator substations in our synthetic network. Public data, used for each generating unit, includes its nameplate capacity, technology type, latitude and longitude coordinate, name, and utility owner.

To create realistic generators in the synthetic case, more information about each generator is needed, beyond what is provided in the public EIA data. Three of these additional parameters include a generator's actual dispatched MW value, maximum Mvar capacity, and minimum Mvar capacity. Additionally, in the real Eastern Interconnect model, generator technology types are categorized by 9 different governor models – cross-compound, diesel, gas, gas turbine, general purpose, hydroelectric, nuclear, steam, wind – rather than the 27 technology types used in the EIA generator data.

To better relate the public EIA data and real Eastern Interconnect, we first sorted the publicly known generators by their technology types into six of the nine governor models, shown in Table I. More research on governor modeling is necessary to ensure proper sorting of all technology types in the public data, especially for the additional three governor model types, diesel, cross-compound, and general purpose that were not considered.

Next, real Eastern Interconnect data was used to relate a generator's governor type and maximum MW capacity to its (1) dispatched MW value, (2) maximum Mvar capacity, and (3) Mvar range. After testing different regression models for the data, a linear regression model, with no y-intercept, was chosen to calculate each of the parameters above, (1-3). Fig. 3 shows how well the different regression models fit the data using the coefficient of determination (r^2) as a metric for assessing best fit. The results, using a linear regression model with no y-intercept, are summarized in Table II. It should be noted that in reality the dispatched MW would depend on fuel costs, generator outages, and demand, rather than the simple dependence on governor type and nameplate MW capacity that we assume. However, for the sake of simplifying our approach and due to lack of data on the aforementioned factors, this approach is sufficient for now, as we continue to experiment with our methodology.

Table II: Scaling factors applied to a generator's total maximum MW capacity to determine dispatched MW and Mvar limits, for various governor types.

Governor Type	Dispatched MW as fraction of MW capacity	Max Mvar as fraction of MW capacity	Mvar range as fraction of MW capacity
Steam	0.867	0.466	0.588
Gas	0.689	0.509	0.620
Gas Turbine	0.569	0.560	0.624
Hydro	0.923	0.384	0.433
Nuclear	0.983	0.368	0.450
Wind	0.226	0.213	0.357

Table I: Summary of technology types used from the EIA generator data and the governor model assigned to them.

Governor Type	Gas	Gas Turbine	Hydroelectric	Nuclear	Steam	Wind
Public Generator Data Technology Type	• Natural gas internal combustion engine • Natural gas with compressed air storage • Other natural gas	• Natural gas fired combustion turbine • Natural gas fired combined cycle	• Conventional hydroelectric • Hydroelectric pumped storage	Nuclear	• Conventional steam coal • Coal integrated gasification combined cycle • Natural gas steam turbine	Onshore Wind

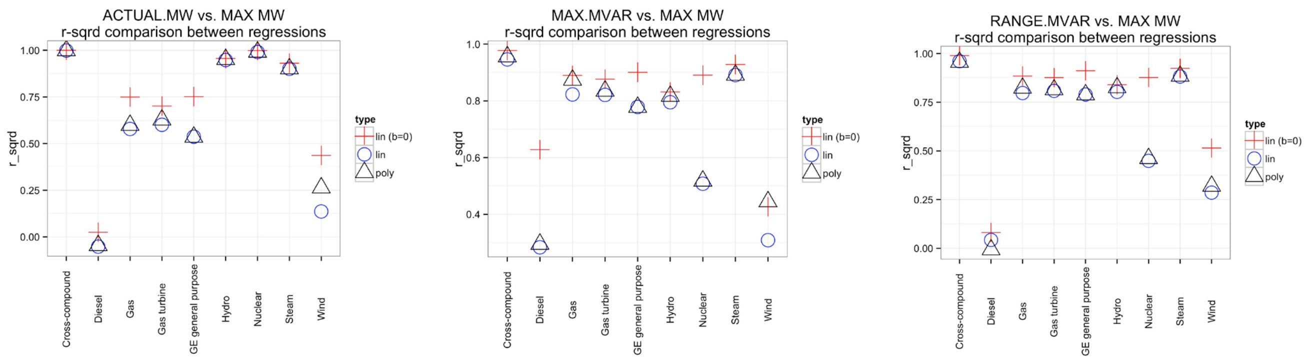

Figure 3: Comparison of regression models using the coefficient of determination, r^2, as a metric for assessing best fit. An r^2 value of 1 corresponds to a model that exactly fits the data. [5]

The calculated values for dispatched MW and Mvar are then appended to the EIA generator data so that every generator has a latitude and longitude coordinate, governor type, and value for maximum MW capacity, dispatched MW, maximum Mvar capacity, and range of Mvar capacity.

Finally, a generator substation is sited at every generator listed in the modified EIA data and is sized to fit the same MW and Mvar constraints given by the generating units they are connected to. A governor step up transformer is assigned at each generator substation, with parameters determined by selecting median parameter values from a list of transformers with the same governor type and voltage level, in the Eastern Interconnect case.

As an example, Table III provides a data sample for three generators listed in the EIA data, with dispatched MW values and Mvar limits obtained using the approach outlined above. Latitude and longitude coordinates, plant name, and utility owners, are not included in this table, but are available in the actual EIA data.

Table III: Sample generator data, developed using the approach for siting and sizing generators, described above.

Technology Type	Governor Assigned	Total MW Capacity	MW Dispatch	Max Mvar	Mvar range
Conventional Steam Coal	Steam	1350	1170	629.7	752.8
Conventional Hydroelectric	Hydroelectric	666.7	589.3	256.3	289.0
Nuclear	Nuclear	3494	3442	1287	1571

C. *Intermediary Substations*

All created synthetic substations are either load or generator substations, since in the Eastern Interconnect, load and generator substations make up about 90% of all substations. The other 10% are intermediary substations that are not directly connected to loads or generators, but are used for switching, connecting transmission lines, and etc. Our synthetic network ignores these intermediary substations, for now.

III. Clustering to Control Synthetic Network Size

In power systems, models of the grid are identified by the number of buses in their system. To allow our synthetic cases to be similarly identified, while providing flexibility in the size of networks that can be created, we employ a simple clustering algorithm based on substation adjacency.

Neighboring, like typed substations are combined to form a cluster. The algorithm limits the number of substations that can be combined to avoid unrealistic massive load or generator centers. Additionally, generator substations can only be clustered if they share the same governor type.

As an example, a XXXX bus case for the state of Tennessee is reduced to a 150 bus case, by combining 628 load substations into 90 equivalent ones, and 76 generator substations into 15 equivalent ones. The remaining buses are left as a cushion for transformers. A limit of 200 Mvar is set for load clusters, representing about 2 million people.

IV. Transmission Line Creation

A. *Transsmission Line Electrical Parameters*

One defining characteristic of transmission lines is their impedance. In previous topological research, Wang et al. [2] randomly assigned transmission line impedances using a statistical model. However, for our network we assign line impedances according to their voltage level and length, since geographic distances are known.

First, the voltage level is used to select the type of transmission tower, conductor type, and other transmission line conventions given in [8]. Then, properties for each conductor are determined from [9] and the following parameters are obtained.

1. Distance between phases (D), in meters
2. Conductor type
3. Number of conductors per phase, referred to as bundle number
4. Distance between conductors in a bundle (d), in meters
5. Single conductor (solid or stranded) outer radius (r_c), in meters
6. Single conductor (solid or stranded) geometric mean radius (GMR), in meters

Next, calculations are performed for per mile inductance and capacitance values using (1) and (2), and the equivalent distance between phases, D_{eq}, with (3), assuming horizontal spacing shown in Fig. 5. Additionally, the equivalent bundle spacings, D_{SL} and D_{SC}, are calculated using equations shown in Table IV. The per mile resistance, r, is found in [9], assuming a temperature of 50° C. The permittivity of free space, ε is 8.854 x 10^{-12} F/m.

$$l = 2 \times 10^{-7} \left(\ln \left(\frac{GMD}{D_{SL}} \right) \right) \times \left(\frac{1609\ m}{1\ mi} \right) \text{ H/mile} \qquad (1)$$

$$c = \frac{2\pi\varepsilon}{\ln\left(\frac{GMD}{D_{SC}}\right)} \times \left(\frac{1609\ m}{1\ mi} \right) \text{ F/mile} \qquad (2)$$

$$D_{eq} = \sqrt[3]{D \times D \times 2D} \text{ meters} \qquad (3)$$

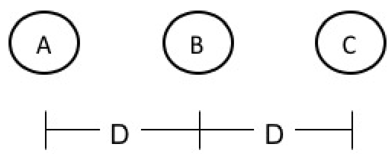

Figure 5: Transmission tower horizontal spacing configuration for a three phase system.

Table IV: Summary of equations for calculating D_{sl} and D_{sc} for various bundling numbers (conductors per bundle).

Conductors per bundle	1	2	3
D_{SL}	GMR	$\sqrt{GMR \times d}$	$\sqrt[3]{GMR \times d \times d}$
D_{SC}	r_c	$\sqrt{r_c \times d}$	$\sqrt[3]{r_c \times d \times d}$

Resistance, reactance, and admittance values are calculated next using (4-6), where L is the length of the line in miles and f is the frequency of the grid, 60 Hz. These values are then input to a software package, like Power World, which assumes a long-line transmission line model shown in Fig. 6. For more information about transmission model types see [10].

$$R = r \times L \ ohms \qquad (4)$$

$$X = 2\pi f \times l \times L \text{ ohms} \qquad (5)$$

$$\mathrm{B} = 2\pi f \times \frac{1}{c} \times L \text{ Siemens} \qquad (6)$$

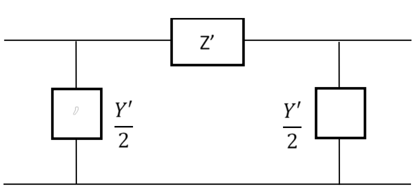

Figure 6: Long-line transmission line model [10].

Lastly, the MVA limit for each transmission line is calculated (7) using the voltage level, V_{line}, number of conductors, n, and allowed amperage through each conductor, I_{max}, found from [9].

$$MVA_{max} = \sqrt{3} \times I_{max} \times V_{line} \times n \qquad (7)$$

As an example of the process described above, a 50 mile, 345 kV line is considered. A Cardinal conductor is used with 2 conductor bundling and per bundle conductor spacing, d, of 1.5 ft. The equivalent spacing between phases, D_{eq}, is determined by the tower selected in Fig. 7. Spacing values are shown in Table V. The calculated values for R, X, B, and MVA limit are shown in Table VI, assuming a system base of 100 MVA and Z_{base} of 1190 Ω.

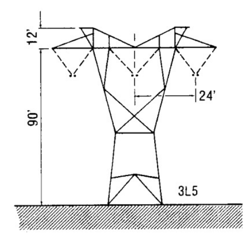

Figure 7: 345 kV transmission tower, with phase spacing, D, of 24 ft. [8]

Table V: Calculated spacing values found using equation (3) and equations in Table IV.

Parameter	Value
D_{SL}	0.246 m
D_{SC}	0.273 m
D_{eq}	30.24 m

Table VI: Calculated actual and per unit values for resistance, line impedance, shunt admittance, and maximum MVA allowed on the line.

Parameter	Value (actual)	Value (per unit)
R	2.82 Ω	2.37×10^{-3}
X	29.2 Ω	2.45×10^{-2}
B	3.59×10^{-4} S	0.427
MVA_{max}	1207 MVA	

B. *Topology Generation Concepts*

The next step in building synthetic networks, which will be addressed more thoroughly in future research, is construction of network topologies. Wang et al. [1][2] presented a topology generation algorithm for arbitrary nodes based on a modified Watts-Strogatz small world model. Their research pointed us in two important directions. The first being the idea of leveraging computational geometry techniques for network creation, especially the Delaunay triangulation technique, and second they emphasized the importance of mean nodal degree as a metric for comparing networks.

In real power networks, the average number of transmission lines attached to a node (mean nodal degree) does not increase with system size as noted in the RT-*nested-SmallWorld* method [1][6][7]. This same finding is observed with Delaunay triangulation which validates our idea of using it to build network topologies that closely mimic characteristics of the real grid.

Delauanay triangulation is a computational geometry technique that links nodes based on nearest-neighbor concepts. The Delauany graph is distinguished from other triangulations in that no triangle's circumcircle contains another point, as demonstrated in Fig. 7. Because of this, the triangles are nicely shaped and connect nearby neighbors together, which is needed in power systems.

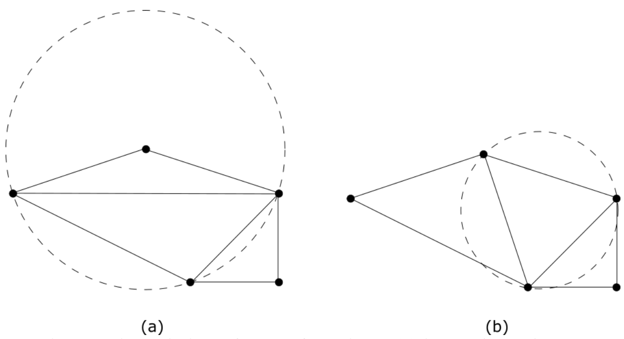

Fig. 7. Triangulation of a set of 5 points. (a) is not the Delaunay triangulation, because at least one triangle's circumcircle contains another point. (b) is the Delaunay triangulation of these points, because no triangle's circumcircle contains another point.

Additionally, we propose expanding the Deluanay triangulation approach to consider not just geometric relations, but also line power-flows, since transmission line electrical parameters and geographic distances are known. Ultimately, the expanded Delaunay triangulation approach will ensure that the synthetic topologies more closely resemble real power network topologies. Future work will address this.

V. Conclusions and Future Work

This paper presented a methodology to site loads, generators, substations, and transmission lines, as a first step towards creating statistically and geographically realistic synthetic transmission networks.

Future work will expand on the methodology outline in this paper to include:

1. Developing a transmission topology generation technique that ensures lines are not overloaded and that an AC power flow solution exists.
2. Creating various network sizes, like a single state's system, the Eastern Interconnect, or the entire North American grid.
3. Adding more components to capture the complexity of the grid, including, generator step up transformers, transient models, dynamic models, and other operational constraints.

As we have worked on this project, it has become even clearer that building these realistic, publicly available, and entirely fictitious synthetic cases, is not only possible, but also necessary to enable greater innovation of the electric grid.

Acknowledgment

We would like to thank the Fox Family and National Science Foundation for their support with this research.

References

[1] Z. Wang and R. J. Thomas, "On Bus Type Assignments in Random Topology Power Grid Models," in *2015 48th Hawaii International Conference on System Sciences (HICSS)*, 2015, pp. 2671–2679.

[2] Z. Wang, A. Scaglione, and R. J. Thomas, "Generating statistically correct random topologies for testing smart grid communication and control networks," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 28–39, 2010.

[3] U.S. Census Bureau. Census 2010, ZIP Code Tabulation Areas Gazetteer File. https://www.census.gov/geo/maps-data/data/gazetteer2010.html

[4] U.S. Energy Information Association. Form EIA-860, 2014. http://www.eia.gov/electricity/data/eia860/index.html

[5] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

[6] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," *Phys. Rev. E*, vol. 69, no. 2, p. 025103, Feb. 2004.

[7] Z. Wang, A. Scaglione, and R. J. Thomas, "The node degree distribution in power grid and its topology robustness under random and selective node removals," in *Communications Workshops (ICC), 2010 IEEE International Conference on*, 2010, pp. 1–5.

[8] General Electric Company, Project UHV, and Electric Power Research Institute, *Transmission Line Reference Book - 345 kV and Above*, 2nd ed. Electric Power Research Institute, 1982, pp. 44, 55-57.

[9] J. Glover, M. Sarma and T. Overbye, *Power System Analysis and Design*. Stamford, CT: Cengage Learning, 2012, pp. 191-192, 204-205

[10] J. Glover, M. Sarma and T. Overbye, *Power System Analysis and Design*. Stamford, CT: Cengage Learning, 2012, pp. 816-817

[11] D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf