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

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

Abstract—**To enable greater innovation in power systems, 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 generator substation placement and assignment of transmission line electrical parameters.**

I. Introduction

One issue 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 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.

The overall approach we propose for building these networks is summarized below, and described more fully in this paper:

1. Site and size load and generator substations using public data and statistics derived from the U.S. electric grid
2. If desired, reduce network size with a clustering algorithm that combines like-typed, nearby substations
3. Calculate transmission line electrical parameters based on voltage level and line length, using conventions employed by grid planners
4. Create a transmission line network topology (future)
5. Create AC power flow solvable synthetic cases, with varying network sizes and complexities (future)

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. Statistics derived from the Eastern Interconnect were used to calculate values for the amount of load consumed 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 total load at each ZIP code, in MW and Mvar.

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. To complete the substation, a voltage level was assigned based on typical values for the region the substation resides in and its MW load. Additionally, a step down transformer was assigned with electrical parameters that are based on median values for Eastern Interconnect transformers that share the same voltage level.

978-1-5090-0261-0/16/$31.00 ©2016 IEEE

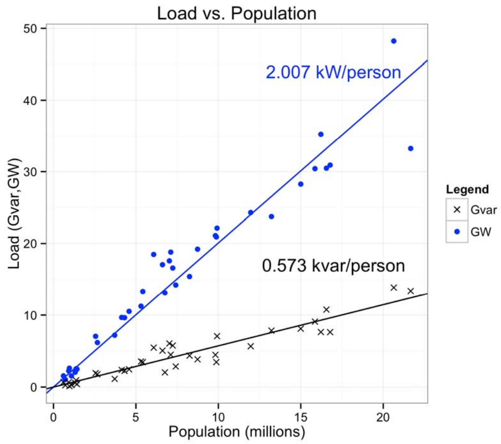

Fig. 1: Load, in Gvar and GW, compared to the total population in millions of people, for the 37 states in the Eastern Interconnect. [5]

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. 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 seems adequate. Future work may explore if any well-defined patterns exist in how many loads are appropriate in areas with very large or very small population densities.

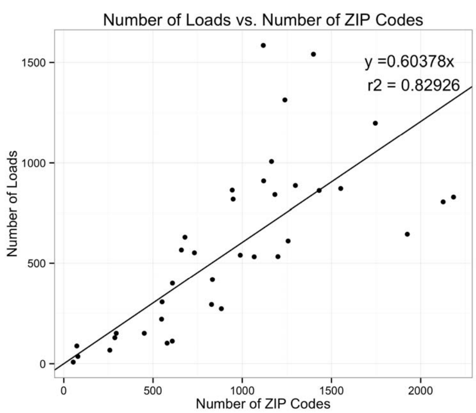

Fig. 2: Total number of loads compared to the number of ZIP codes in each of the 37 states of the Eastern Interconnect. [5]

B. Generators

The 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 was needed, beyond what was 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 models of the real Eastern Interconnect, 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, statistics for the Eastern Interconnect were 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 based on Eastern Interconnect data, a linear regression model, with a y-intercept of 0, was chosen to calculate each of the parameters above, (1-3), with results shown in Table II. To justify our model choice, 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. 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. Also, the poor model fit for diesel and wind is largely ignored because of how few diesel generators exist in the U.S. electric grid and because wind's intermittency is already known to deviate from typical dispatching models.

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

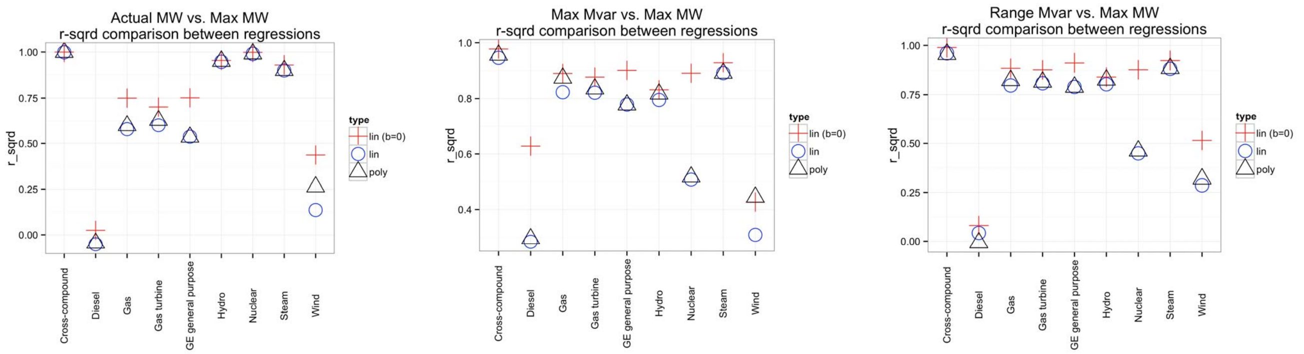

Fig. 3: Comparison of regression models using the coefficient of determination, r^2, as a metric for assessing best fit. Regression models include linear, linear with a y-intercept of 0, and 2nd order polynomial. An r^2 value of 1 corresponds to a model that exactly fits the data. [5]

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

The calculated values for dispatched MW and Mvar were 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 was sited at every generator listed in the modified EIA data and was sized to fit the same MW and Mvar constraints given by the generating units they are connected to. To complete the substation, a voltage level was assigned based on typical values for the region the substation resides in and the MW generating capacity of its generating units. Additionally, a governor step-up transformer was assigned with electrical parameters that are based on median values for Eastern Interconnect transformers that share the same governor type and voltage level.

As an example of our approach, 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.

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.

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

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, with a total load or generation amount equal to the sum of the individual loads and generators that make it up. 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 fictitious 816 bus case for the state of Tennessee is reduced to a 150 bus case, by combining 628 load substations (derived from population data) into 90 equivalent ones, and 76 generator substations (derived from EIA generator data) into 15 equivalent ones. The remaining buses are left as a cushion for transformers and other connections. A limit of 100,000 people is set for each load cluster.

IV. Transmission line creation

After loads and generators have been sited and sized, the next steps are to calculate transmission line electrical parameters and to construct their network topology. This section presents information for the former, and highlights a few concepts for future work in network topology generation.

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, known by what substation the line is connected to, and length, since geographic distances are known.

First, the voltage level was used to select the type of transmission tower, conductor type, and other transmission line conventions given in [6]. Then, properties for each conductor were determined from [7] and the following parameters were obtained.

1. Distance between phases (D), in meters
2. Conductor type
3. Number of conductors per phase, bundle
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 were performed for per mile inductance and capacitance values using (1) and (2), and the equivalent distance between phases, D_{eq}, with (3), assuming tower configurations with horizontal phase spacing shown in Fig. 5. Additionally, the equivalent bundle spacings, D_{SL} and D_{SC}, were calculated using equations shown in Table IV. The per mile resistance, r, is found in [7], assuming a temperature of 50° C. The permittivity of free space, ε is 8.854 x 10^{-12} F/m.

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

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

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

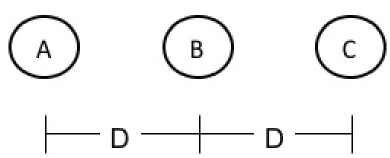

Fig. 5: Transmission tower horizontal spacing configuration for a three phase system, with A, B, and C phases.

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 were 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 were then input to a software package, like PowerWorld, which assumes a long-line transmission line model shown in Fig. 6. For more information about transmission line model types see [8].

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

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

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

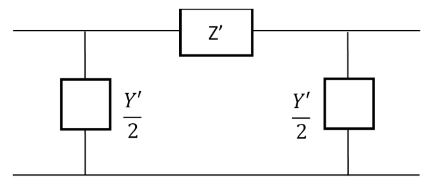

Fig. 6: Long-line transmission line model [10].

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

$$MVA_{max} = \sqrt{3} \times I_{max} \times V_{line} \times n$$

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.

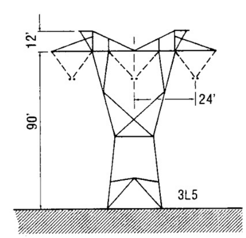

Fig. 7: 345 kV transmission tower, with phase spacing, D, of 24 ft. [6]

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 to expand the topology generation algorithm proposed in [1][2] by using computational geometry techniques that take advantage of the geometric relations known from geographically sited substations. One such geometric technique is called Delaunay triangulation, which links nodes in a network based on nearest-neighbor concepts. The Delaunay 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, which is needed in power systems. Another benefit for using Delaunay triangulation to generate power system topologies is the property Delaunay graphs share with real power networks, where 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][10][11].

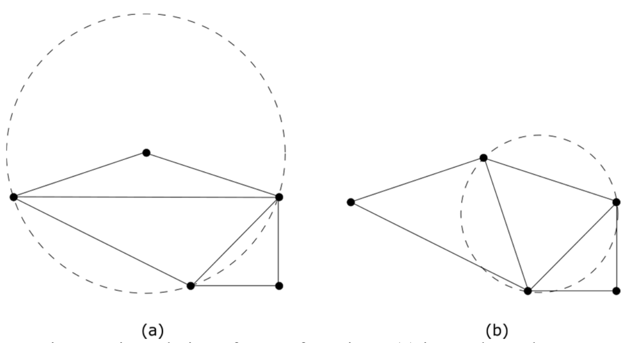

Fig. 7. Triangulation of a set of 5 points. (a) is not the Delaunay triangulation, because at least one triangle's circumcircle contains another point. (b) is the Delaunay triangulation of these points, because no triangle's circumcircle contains another point.

Overall, the observed benefits for using Delaunay triangulation make it an obvious choice for building synthetic topologies, though we propose going even one step further: expanding the Delaunay triangulation approach to consider not just geometric relations, but also line power-flows, since transmission line electrical parameters are known. Ultimately, the enhanced Delaunay triangulation approach will ensure that the created synthetic topologies more closely resemble real power network topologies. Future work will address this.

V. Conclusions and Future Work

This paper presented a foundational step for building geographically and statistically realistic, though entirely fictitious, synthetic networks for the U.S. electric grid. A methodology was proposed to geographically site loads, generators, and substations and assign transmission line electrical parameters. Ultimately, the validity of our approach will be decided when a full synthetic case has been made, but we believe the decisions we made – (1) basing loads on population data, which share a nearly direct relationship, (2) using actual generators operating in the U.S. electric grid and their locations, obtained from publicly available EIA data, and (3) assigning transmission line electrical parameters using conventions employed by transmission planners – will ensure as close to a realistic representation of the U.S. electric grid as possible, though still entirely fictitious.

Future work will expand on the methodology outlined 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 synthetic 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, transient models, more precise electrical parameters for step-down and generator step-up transformers, and other operational constraints.
4. Developing verification standards to assess if our synthetic cases are sufficiently realistic.

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

Acknowledgment

We would like to thank the National Science Foundation for their support with this research, under grant IIP-0968983.

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] 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.

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

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

[9] 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

[10] 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.

[11] 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.