Analysis of Economic Criteria in the Creation of Realistic Synthetic Power Systems

Desiree Phillips, Ti Xu, Thomas Overbye
Dept. of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL, 61801 USA
{dphllps2, txu20, overbye}@illinois.edu

***Abstract*—In creating entirely fictitious electric grids, there need to be metrics associated with their design that qualify their use in power systems research. In specifying these measures, the questions "What metrics matter for a given study and why?" should be answered. This paper extends the previously developed synthetic network base models for energy economic studies and aims to outline which economic parameters should be prioritized. Using a synthetic, 200-substation representation of Illinois, the optimal power flow (OPF) is analyzed in two stages to suggest what specifications need to be met to validate the model.**

***Index Terms*—power flow, optimal power flow (OPF), synthetic networks**

I. INTRODUCTION

One problem that arises in power systems research is the lack of publicly available realistic transmission system models. The electric grid is one of the United States' critical infrastructures, so it is not surprising that for national security purposes, transmission network information is proprietary. The actual system data that can be obtained is typically restricted by a non-disclosure agreement, meaning any findings that come from analysis of these networks are not publishable. Of the test cases that are publicly available, information is often lost due to equivalencing, the device models do not meet a required level of complexity, or they do not contain data that is needed for a specific type of study (e.g.: economic).

Several authors, along with an IEEE working group, have looked at developing power system test cases for economic analysis [1], [2]. Beyond computational analysis of the systems created, there has been no look at what economic parameters are sufficient measures of feasibility. In addition, there is no standardized methodology for the creation of these test cases [3].

To increase the number and accuracy of available models, research is being done in using statistics obtained from publicly available power systems data to develop entirely fictitious synthetic networks. Previous work has addressed geographically realistic power grids and applying them to geomagnetic-disturbance studies and energy economics, but once extended, these models can examine wholesale markets, multi-system operations and optimization, and grid dynamics [4], [5].

Past analysis looked into developing a synthetic Texas network whose behavior matched the economic trends of the actual ERCOT system [6]. The paper illustrated use of the synthetic test case for planning and reserve deployment by comparing the real and synthetic results, but did not look into developing economic measures of realism.

As an extension of this economic work, this paper will analyze a synthetic, entirely fictitious, 200-substation representation of Illinois (IL) from an optimal power flow (OPF) perspective. The goal of this contribution is to examine additional economic details of the power system, in particular the locational marginal prices (LMPs), to validate the economic parameters being created through the statistical analysis of real networks. Note that this work is attempting to capture the behavior of a real network, not its exact values.

The paper reads as follows. Creation of the Illinois synthetic network will be covered in Section II and the metrics of concern for power system economic studies are introduced in Section III. Analysis of the resulting operating costs of the synthetic network are covered in Section IV, with conclusions and future directions being discussed in Section V.

II. CREATION OF SYNTHETIC NETWORK BASE MODELS

Creating synthetic network base models begins with the application of statistics summarized from actual system models and publicly available data. Once a synthetic network base model is constructed, additional complexities can be added to improve the realism of the model by including data necessary for various types of studies.

Based on the statistics summarized from a recent Eastern Interconnect (EI) planning case [7], the substations are categorized into three different groups:

- Type G (about 4.5 %): purely contain generators;
- Type L (about 91 %): purely contain loads;
- Type H (about 4.5 %): contain both generators and loads.

Analysis of the groups has shown that random assignment of G/L/H can be insufficient for emulating regional system behavior; there needs to be some basis for substation assignment [8].

This work was supported by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000714; the U. S. Department of Energy Consortium for Electric Reliability Technology Solutions (CERTS); and the Institute for Sustainability, Energy, and Environment (iSEE).

This paper was presented at the IEEE PowerTech, Manchester, UK, June 2017.

Only high-level summaries about actual load information is available to public. Given the highly linear relation between the electricity demand and population [7], load substations are cited using postal code population data publicly available through the U.S. census database [9]. This data contains the geographic coordinates and population of each postal code, with the size of the load (in MW and Mvar) dependent upon the population size.

On the supply side, the U.S. Energy Information Administration (EIA) maintains a yearly survey of the nation's generators and has already made the 2014 survey data available to the public [10]; information collected includes the fuel type, generation capacity, and geographic coordinates. Some generators are selected and assigned to the Type H substations. The remaining generators are clustered into several non-overlapping subsets, each of which forms a Type G substation.

Whether generation or load, clustering of adjacent substations may be done to adjust the size (number of buses) of the network created. The resulting load (generation) is simply a sum of the load (generation) which created it. This clustering, however, is not done arbitrarily, as it may create large load/generation centers and unrealistic substation groupings. To prevent this, there is a maximum number of substations that can be clustered and generation with differing governor types cannot be combined.

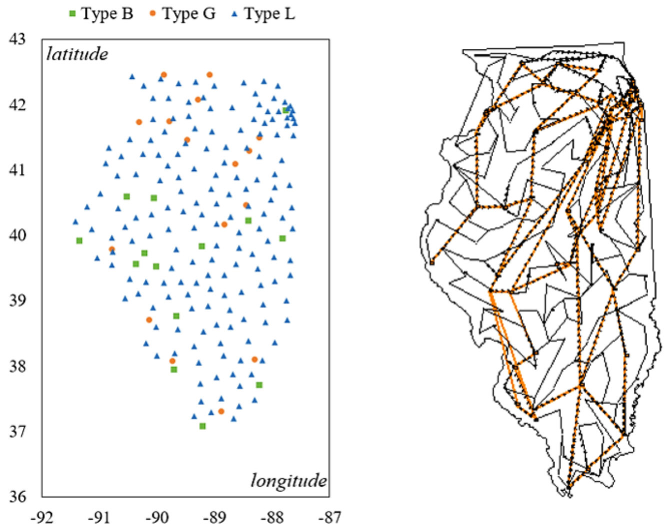

Fig. 1. Synthetic substation and line assignment results.

Upon the completion of the substation assignment process, synthetic transmission lines are connected to these substations. About 10-20% of substations are assigned to a higher system voltage with probabilities proportional to load and large Type H/G substations are more likely to be at a higher voltage level. Within each substation, loads are usually connected to the lowest voltage level, and generators are often connected to the highest voltage level through a step-up transformer. An iterative penalty-based algorithm is proposed in [5] to connect buses in each voltage level, while satisfying some structural statistics summarized from the actual models.

Line electrical parameters required for power flow analysis include series impedance, shunt admittance, and MVA limits are obtained from available datasheets and reference manuals. References [5], [11] provides a basis for the specification of the transmission line and tower configurations based on voltage level. From there, specific conductor properties can be obtained from the corresponding conduction sheet. Figure 1 shows the synthetic substations and line assignment results for the IL-200 case, a synthetic Illinois grid with 200 substations.

Once the system base case is developed, additional statistical data regarding system economics and production costs can be added to the model for optimal power flow (OPF) studies. The AC-OPF problem referred to here aims to minimize the energy production costs given system-dependent equality and inequality constraints. Since the constraints are all a function of the line/generation parameters previously specified, the object function, i.e. the generation cost curves $C(P)$, are all that requires defining. Statistics summarized in [2] provided the basis for determining cost model coefficients by fuel type and capacity. For this work, a quadratic cost model was applied to each generator:

$$C(P) = \begin{cases} a_0 + a_1 P + a_2 P^2 & \text{Approach A} \\ a_0 + c_f (b_1 P + b_2 P^2) & \text{Approach B.} \end{cases}$$

where a_0 denotes the no-load cost for each generator, c_f is the fuel cost, and b_1 and b_2 are fuel-dependent cost coefficients. There are two approaches in order to allow for flexibility in the input economic statistics. In particular, approach B allows the user to directly manipulate the generation fuel costs while approach A is a more straightforward curve-fitting implementation. For further information on the origins of the economic data that provided the coefficients for this network, we refer the readers to [4]–[6].

III. DEFINING ECONOMIC CRITERIA

Selecting the criteria to be used in algorithmically creating economic data will be divided into two stages. The first stage will involve economic metrics that do not require an OPF solution before they can be examined. These metrics will be studied in a "pre-OPF" stage using statistical comparisons with publicly available data. The second stage comes after the OPF has been solved and is validation of the pre-OPF metrics. This "post-OPF" analysis will look at the economic results and decide if the criteria chosen in the pre-OPF step are both necessary and sufficient measures of realism.

A. *Pre-OPF Analysis*

Figure 2 shows the dispatch ("supply") curve for different regional Independent System Operators (ISOs) [12]. For day-ahead auctions, ISOs look at an aggregate of the individual supply curves and use OPF techniques to determine the suppliers they will use to meet their customers' demand. The shape of the curve visualizes the idea that as electric demand increases, plants with lower operating costs are brought on first. In this case, Electric Reliability Coucil of Texas' (ERCOT) marginal costs are more uniformly low due to a combination of (1) low-cost fuel resources and (2) retiring older units and replacing them with more energy-efficient generation.

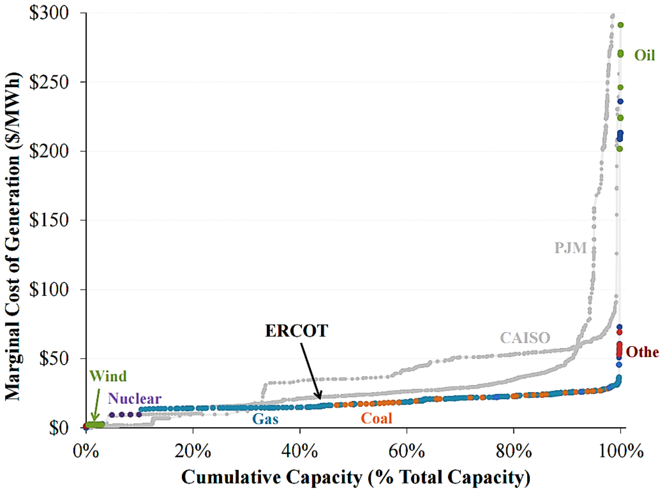

Fig. 2. Supply Curves of various regional ISOs [12].

In developing realistic synthetic networks, there is one aspect of the supply curve to be preserved: the shape of the curve. As shown in Fig. 2, the curve's trend is typical of many power systems, regardless of the region being examined. In particular, the supply curve for each system is monotonically increasing, showing that the least expensive generating units are turned on first to meet an increasing demand. The pre-OPF portion of this work consists of using the synthetic network's supply curve to validate further economic analysis. Depending on the topology of the electric network and the generation mix, the curve may not be as flat/steep as seen here. However, the trends related to operator and market characteristics should be consistent across all developed grids.

B. Post-OPF Validation

Once all cost parameters have been defined and the OPF run, it is now time to look at the results and perform post-OPF validation. For this portion of the analysis, the statistical features of the locational marginal prices (LMPs) are of primary interest. LMPs are examined because they are directly affected by system demand, nearby congestion, and energy and reserve offers [13].

Since reasonable computational times are essential when calculating the LMPs, the OPF results using the AC and DC power flows are both examined. For many power system studies, use of the DC power flow gives a good approximation with the added benefits of speed and simpler implementation [14]. As with the pre-OPF features, the exact values may differ depending on the grid being developed, but the overall trend (distribution) of the values should be similar to those seen in real cases.

IV. IL-200 CASE STUDY

Figure 3 shows the generation mix of the IL-200 case, while Table I provides some of it's generation and load characteristics. With regards to the cost curves, Approach B was used to allow for input of fuel cost data using parameter c_f. A merit-order method is used to develop a commitment scheme for the IL-200 system. While there are more robust/complex unit commitment methods, the focus of this work is validation of the economic parameters, not the optimality of unit scheduling. Based on the production costs of a given unit, generators are de-committed until a specified reserve margin has been reached. During each iteration of generation decommission, a power flow is solved to ensure system feasibility.

TABLE I
IL-200 CHARACTERISTICS

Buses	452
Lines	641
Generators	202
Gen. Capacity	50.6 GW
Total Load	25.7 GW

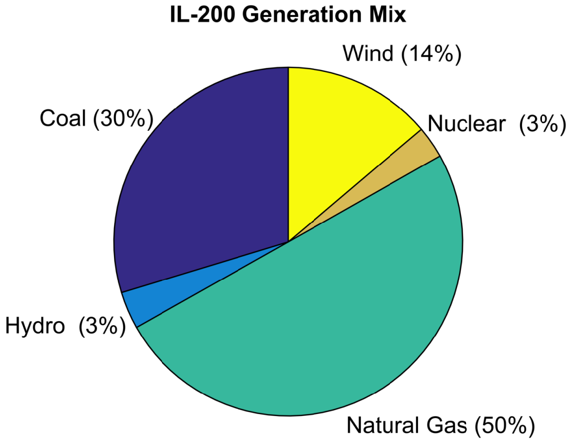

Fig. 3. Generation Mix of the IL-200 Case.

A. IL-200 Supply Curve

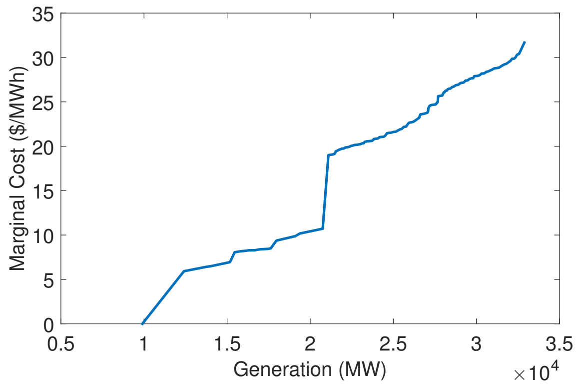

Fig. 4. Supply Curve of the IL-200 Case.

The supply curve of the IL-200 case is shown in Fig. 4. The curve is monotonically increasing, a necessary feature, and there is a steep rise in cost of 8.29 $/MWh at 20.7 GW. Due to the generation mix and parameter assignment of Section II, many of the individual incremental cost curves overlap. This overlap, when summed to create the aggregate cost curve, can create discontinuities and sharp rises in cost [15]. Note that the curves in Fig. 2, in particular the CAISO curve at 35% capacity, have similar jumps in cost.

B. AC versus DC LMPs

The LMPs resulting from using an AC versus DC power flow are shown in Fig. 5. Not only is the shape of their waveforms similar to one another, they are comparable to what is typically seen in the OPF literature as well [16]. The standard deviation of the DC-OPF is lower due to the asymptotic behavior seen in the ends of the AC LMP curve.

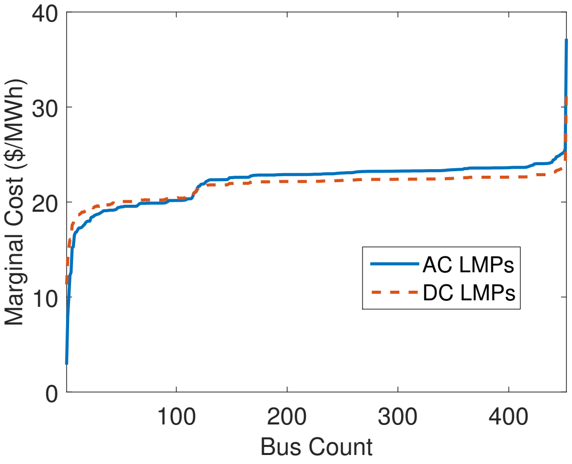

Fig. 5. Comparison of the IL-200's LMPs using the AC and DC power flow.

C. Comparison to MISO and Synthetic ERCOT Results

Table II shows the LMP statistics for the IL-200 case using the AC and DC power flow (PF), the ERCOT synthetic network (using approach B), and the Midcontinent Independent System Operator (MISO) data for Illinois [17]. As shown in Figure 6, the IL-200's LMP discontinuity occurs in the MISO IL data as well, however the MISO curve isn't as smooth due to an order of magnitude difference between the number of data points available.

TABLE II
OPF LOCATIONAL MARGINAL PRICE STATISTICS

Measure	AC PF	DC PF	ERCOT B	MISO
Mean ($/MWh)	22.11	21.69	23.87	26.67
S.Deviation ($/MWh)	2.42	1.46	2.69	2.29

The standard deviation of the IL-200 LMPs are close to those calculated in the synthetic ERCOT model using approach B. Various regions of the power grid may operate at different price points, so it is not imperative that the average LMPs match between IL and ERCOT.

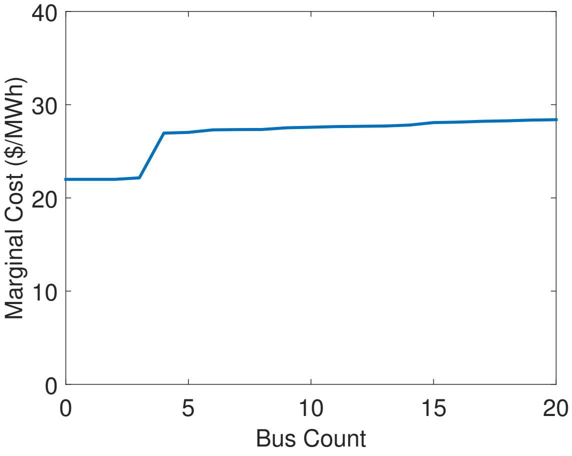

Fig. 6. Distribution of the MISO LMPs in Illinois [17].

V. CONCLUSION

This paper looks at the features that should be considered when developing synthetic networks which contain generation parameters that come from the statistical analysis of the actual grid. These preliminary results, while currently unable to capture the complete picture, are indicative of feasible system behavior. The supply curve exhibits trends that also appear in various ISOs, and the distribution of the AC and DC LMPs are consistent with economic results obtained from actual electric grids.

Future work will look at how the location of the synthetic network and the generation mix affects the range of values allowed by the economic parameters. Previous work has looked into mimicing the behavior of the grid the synthetic network was based off of, but it may be worth examing how regional grid differences manifest themselves in the statistical parameter constraints.

Additionally, going from a network of 200 substations to 2000 substations may introduce system complexities that this iteration of synthetic networks cannot capture. Immediate future work has been focused on looking into what parameters may become significant in larger, more-detailed synthetic networks.

REFERENCES

[1] C. Schaffner and X. P. Zhang, "Test systems for Economic Analysis An introduction," in *IEEE PES General Meeting*, July 2010, pp. 1–2.

[2] 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, Jan 2016.

[3] J. Marmolejo Saucedo and R. Rodríguez Aguilar, "A proposed method for design of test cases for economic analysis in power systems," *Journal of Applied Research and Technology*, vol. 13, no. 3, pp. 428–434, June 2015.

[4] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical Considerations in the Creation of Realistic Synthetic Power Grids for Geomagnetic Disturbance Studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502–1510, March 2017.

[5] 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. PP, no. 99, pp. 1–1, 2016.

[6] 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 *Hawaii International Conference on System Sciences (HICSS-50)*, 2017.

[7] K. M. Gegner, A. B. Birchfield, Ti 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)*, Feb 2016, pp. 1–6.

[8] 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*, Jan 2015, pp. 2671–2679.

[9] (2010) Census Gazetteer Files: ZIP Code Tabulation Area. U.S. Census Bureau. [Online]. Available: https://www.census.gov/geo/maps-data/data/gazetteer2010.html

[10] (2014) Form EIA-860 Detailed Data. U.S. Energy Information Association. [Online]. Available: http://www.eia.gov/electricity/data/eia860/index.html

[11] J. Glover, M. Sarma, and T. Overbye, *Power System Analysis and Design*, 5th ed. Cengage Learning, 2011, ch. Appendix, pp. 814–817.

[12] S. Newell, K. Spees, J. Pfeifenberger, R. Mudge, M. DeLucia, and R. Carlton, "ERCOT Investment Incentives and Resource Adequacy," The Brattle Group, Inc, Tech. Rep., 2012.

[13] L. L. Lai, Ed., *Power System Restructuring and Deregulation*. John Wiley & Sons Ltd, 2001, ch. 3, pp. 88–89.

[14] A. J. Wood, B. F. Wollenberg, and G. Sheblé, *Power Generation, Operation, and Control*. John Wiley & Sons, 2014, ch. 8, pp. 352–354.

[15] Steven Stoft, *Power System Economics:Designing Markets for Electricity*. Wiley-IEEE Press, 2002, ch. 6, pp. 60–73.

[16] T. J. Overbye, Xu Cheng, and Yan Sun, "A comparison of the AC and DC power flow models for LMP calculations," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, Jan 2004.

[17] (2016, October) LMP Consolidated Table. Midcontinent Independent System Operator. [Online]. Available: https://www.misoenergy.org/MarketsOperations/RealTimeMarketData/Pages/LMPConsolidatedTable.aspx