Towards Operational Validation: Mapping Power System Inputs to Operating Conditions

Eran Schweitzer*, Ti Xu†, Adam B. Birchfield†,
Anna Scaglione*, Thomas J. Overbye†, Robert J. Thomas‡, Zhifang Wang§

***Abstract*—Validation is a critical component in the development of synthetic models, which aims to convince a user that a model achieves its claims of realism. While users of power system test cases are primarily interested in operational results, which could be considered outputs, it is more convenient and feasible to control distributions of data inputs, both structural and otherwise. Validation metrics, are therefore generally focused on input rather than operation features. This paper investigates the link between input and output statistics in power system. Better understanding of this relationship allows validation to continue to focus largely on input statistics, while at the same time offering some assurance about operational behavior.**

***Index Terms*—Model Validation, Network Analysis, Power System Topology, Synthetic Power Systems**

I. Introduction

As part of the effort to expand the current collection of publicly available synthetic power systems funded by ARPA-E [1], a critical component is careful examination and selection of the methodology and metrics used to validate cases. Newly created systems will be judged on how well they depict the real power grid they are supposed to mimic. Demonstrating their validity is therefore essential. In our previous work [2], several tests are introduced, which validate cases generated based on the methodology in [3]. The work in [4] aims to expand these tests, and introduces a distinction between structural and operational metrics.

A large part of the validation challenge lies in the gap between the quantities that can be easily controlled and those that are of interest. While users of power system test cases are primarily interested in operational results, it is more convenient and feasible to control distributions of data inputs, both structural and otherwise. It is therefore, easier to formulate validation of input statistics, as is true of most of the metrics in [2] for example. What users desire however, are certain assurances about the operational statistics.

The North American Energy Reliability Corporation (NERC), and specifically its Model Validation Working Group, have spent a lot of time and effort considering validation of both steady-state and dynamic cases [5], [6], which resulted in issuing a recent standard [7], requiring data validation in planning studies. The essence of these documents is that given the same input data taken from measurements of a real event, simulations should produce outputs (which we refer to in this paper as operational behavior) that match said real event. Synthetic systems offer an additional challenge, since no real events can be measured on them. Our general tool for addressing this challenge is to consider statistical distribution of operating behaviors and try to achieve similar ones.

The tuning nobs available when creating a synthetic case, are system input data, which consists of topology, equipment parameters and load, rather than the simulation outputs. Our work in [4] raises the question: are structural and operational statistics related, and if so, how? In this paper, we expand from strictly structural to consider statistics on the input dataset more broadly. We present a methodology for tackling the question of how input statistics influence the observed output statistics, namely power system operational behavior. As these links become clearer, both conceptually and numerically, validation metrics on the input statistics will offer progressively more information about the operations that will result.

The goal of this paper is to demonstrate the application of this methodology by example. We focus on two metrics, graph cycle distribution and generator rating, and test their impact using the ACTIVSg2000 case [8], which was created on the ERCOT geographic footprint. In addition to comparing various modifications of synthetic cases among themselves, real ERCOT data is used for further reference. The specific examples were chosen because they both impact operational behavior and as such serve to illustrate our point. First, the cycle distribution is analyzed with respect to its impact on steady state operations. Second, the impact of the relationship between generators' rating and their capacity is considered with respect to transient stability behavior. Future work will expand the range of metrics tested in this manner.

The remainder of the paper is structured as follows. In Section II we describe the testing methodology used to determine the link between input and output statistics. Sections III and IV form the bulk of the paper and discuss the application and findings from two experiments using the testing methodology. Finally, section V offers concluding remarks.

II. Methodology

Our objective is to evaluate the impact of a specific feature on the operational behavior of the power system. The main

* School of Electrical, Computer and Energy Engineering Arizona State University, Tempe, AZ, USA. {eran.schweitzer, anna.scaglione}@asu.edu
† Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. {abirchfield, overbye}@tamu.edu
‡ School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. rjt1@cornell.edu
§ Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA, USA. {atharih, zfwang}@vcu.edu

challenge is to isolate a single feature, while the obvious benefit is drawing conclusions on a feature's latent impacts. To this end, we aim to create two (or more) cases that are identical except for the feature under examination. By comparing the operations of the two systems, we can therefore infer the impact of the altered feature. It should be emphasized that since simulations are used to demonstrate impact, we cannot claim proof. Rather, we collect supporting evidence for a hypothesis, with the rationale that the likelihood it is correct increases with afirmative results.

Holding everything but a single variable fixed is the classic experimental design. The features we are referring to however, are not a deterministic number but rather distributions of a parameter or structural feature. Our methodology can therefore be restated as maintaining all input *statistics* but one constant, and observing result in the operational output.

Given the complex nature of power systems, creating identical cases but for one feature is not trivial. Power system operations is highly optimize to the present scenario. For this reason, we define *identical* as sharing all statistical properties *except* the operational ones. Specifically, algorithms like unit commitment (UC) and optimal power flow (OPF) may be run to achieve the most "realistic" operating point. Therefore, it is entirely possible that two cases in a test will have different generation dispatch. Despite this "additional" difference, we consider the cases identical, since the same algorithm is used to determine the commitment and dispatch. In essence, UC and OPF are already part of operations and are thus part of the observed output of the experiment.

III. Cycles

In [4] we introduced the minimum cycle distribution of the power system graph as a structural metric. Cycles bases have long been part of electrical analysis, in fact, Kirchhoff's work [9] uses fundamental cycle bases for the application of his famous voltage law. Cycle bases are however, non-unique. We therefore, adopt minimum cycle bases using the algorithm from [10].

A minimum cycle bases is one where the sum of the weights of all cycles is minimum. When unit weight is assigned to each edge, this translates to meaning that it is the collection of the smallest cycles that form a basis for the graph. While this collection is also non-unique, its distribution is. A cycle distribution is a count of how many cycles of each size form the basis of the graph. For the distribution to change, the count between different cycle sizes must change, which means that some cycle is broken into smaller constituent parts. This, however, contradicts the definition of a minimum cycle distribution, and explains why the distribution is unique even if the cycles making it up can be selected in different ways.

We observe that the original ACTIVSg2000 case has a slightly larger mode in the cycle distribution than other interconnects. Considering that cycles imply parallel paths, an intuitive reasoning based on KCL suggests that there might be a relationship between the cycle basis of the power system graph and the resulting loading distribution.

TABLE I
Negative Binomial Parameters to Cycle Distribution

Case	p	r
ACTIVSg2000	0.27	12.66
ERCOT	0.74	1.76

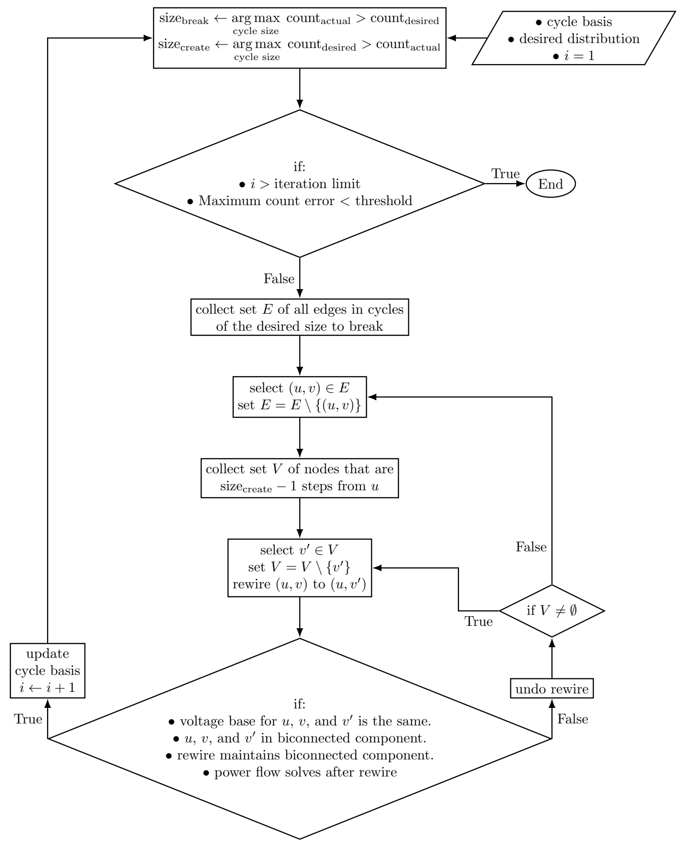


Figure 1. Flowchart describing how cases are rewired to target a specific cycle distribution

A. Modifying Cycles

To test this hypothesis, we start with the ACTIVSg2000 case and rewire some edges to target a cycle distribution more similar to the ERCOT system. A simple greedy approach to achieve this is outlined in Figure 1. In [4] we showed that the minimum cycle lengths can be well fit by a Negative Binomial distribution,

$$f(k; p, r) = \frac{\Gamma(k + r)}{k!\Gamma(r)} p^k (1 - p)^r \quad k = 0, 1, \ldots, \qquad (1)$$

where $\Gamma(\cdot)$ is the gamma function, $p \in (0, 1)$, and $r > 0$. Note that since cycle must be size 3 or greater, we map cycle size as $k = \text{cycle size} - 3$. The parameters fit to the ACTIVSg2000 and ERCOT cases are shown in Table I.

Using the fit parameters, Equation (1), and the total number of cycles[1], we can calculate how many cycles we wish to see

[1]The total number of cycles does not change since by Euler's equation it is $M - N + 1$ where M is the number of edges, N the number of nodes, and 1 is the number of connected components.

TABLE II
KL-DIVERGENCE OF ACTIVSG2000 CYCLE DISTRIBUTION

	original case	modified case
$D_{KL}(p_{\text{ACTIVSg2000}} \| q_{\text{ERCOT}})$	0.2050	0.0157
$D_{KL}(p_{\text{ACTIVSg2000}} \| q_{\text{ACTIVSg2000}})$	0.0240	0.2316

at any size. The procedure in Figure 1 selects the cycle size that is most *over represented* in the current cycle basis as the type of cycle to break, while the most *under represented* cycle is the one to target. Edges participating in cycles of the appropriate size to break are selected, and one by one, new neighbors are sought, until one is found which fulfills the requirements. These are, namely:

- the voltage basis of the nodes will be the same (as it does not make sense to rewire transformers).
- System biconnectivity is unaltered.
- A power flow solution exists.

As an implementation note, the calculation of a minimum cycle basis is rather expensive. For this reason, in the main loop of Figure 1, an update algorithm is used, which we developed as a modification of [11]. At the end of the procedure the full cycle distribution is calculated again to catch any possible errors caused during successive updates[2].

Figure 2 shows the minimum cycle distribution of the original and modified ACTIVSg2000 cases, as well as the original ERCOT case for reference. It is visually clear, that the desired change in the distribution has been achieved.

We evaluate the change numerically by considering the Kullback-Leibler (KL) divergence [12] between the empirical distribution and the desired Negative Binomial distribution parametrized by the ERCOT values in Table I. The empirical distribution simply refers to the histogram of the data, which could have bins with no associated weight. This differs from the chosen Negative Binomial *model*, which will map any value in its domain to an associated weight, irrespective of the underlying data used to initially fit the model. For example, there might not be cycles of size 52 in the dataset, but (1) can certainly be evaluated at $k = 52 - 3$. We adopt the following notation:

- p_i refers to the *empirical* cycle distribution for case i. Whether the original or modified case is intended is indicated either with a superscript or elsewhere.
- q_i refers to the Negative Binomial distribution with parameters for case i from Table I.

Using this convention, the change in KL-divergence values is tabulated in Table II. The numbers strongly support the visual from Figure 2 that the cycle distribution for the modified ACTIVSg2000 case closely matches that of the original ERCOT case, and furthermore no longer matches the original ACTIVSg2000 case.

[2]We experimentally found that the update did not perform perfectly. However, the error, on the order of 0.1%, is minimal.

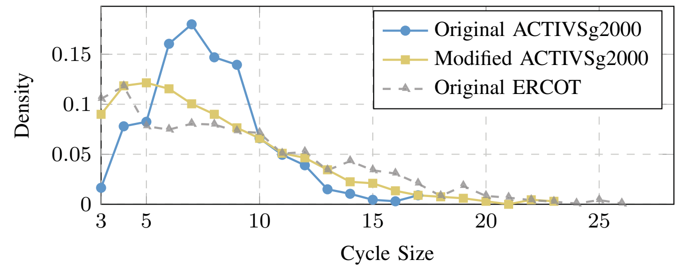


Figure 2. Cycle distribution for the ACTIVSg2000 case before and after modification. The targeted ERCOT case is also shown for reference.

TABLE III
ACTIVSG2000 OPERATIONAL VALUES

Quantity	Original	Modified Case
Cost [$]	1,220,002.12	1,222,871.83
Average LMP [$]	19.62	19.70
Losses [MW]	1389.19	1539.84
Average L_s	1.5498	1.6050
Average $\Delta\theta$ [degree]	1.50	1.63
$D_{KL}(L_s \| \text{Exp}(\mu_{L_s}))$	0.1625	0.0979

B. Effects On Operations

Having established the desired structural change, we turn to look at how operations are (or are not) impacted. As mentioned in Section II, a simple UC determines an economically optimal dispatch for both the original and modified cases. Subsequently, the ACOPF is solved using Matpower [13], which also accounts for losses, voltage deviations, etc., that are not captured by the DC model used for UC. One necessary technical note is that in all cases, line limits are removed since they are not considered during topology manipulation, and could therefore introduce problematic constraints that are merely artifacts of the modification algorithm. As all cases are handled the same, we are still comparing two identical cases except for a topological manipulation. Once the operating point for both cases is established, statistics between the two can be compared.

In [4] we proposed the distribution of a line's load with respect to its Surge Impedance Loading (SIL) rating as a good way to describe the loading of the system. SIL represents an impedance matched termination for a lossless line, in the sense that the voltage profile is flat at the nominal voltage, when the line delivers its SIL [14]. We use L_s to denote the ratio between actual loading and SIL,

$$L_s = \frac{\text{Actual MVA flow}}{\text{SIL MVA rating}}. \tag{2}$$

We further showed that for all interconnect cases the exponential distribution is a good model for L_s. Figure 3 shows the distribution of L_s for the original and modified ACTIVSg2000 cases. One can see that the peak of the distribution shifts in a desirable direction, where desirable means more similar to the ERCOT case, although the change is admittedly not very significant.

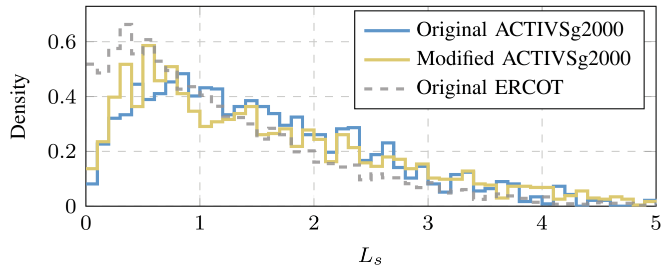


Figure 3. Distribution of L_s for ACTIVSg2000 case before and after modification. The ERCOT case is also shown for reference.

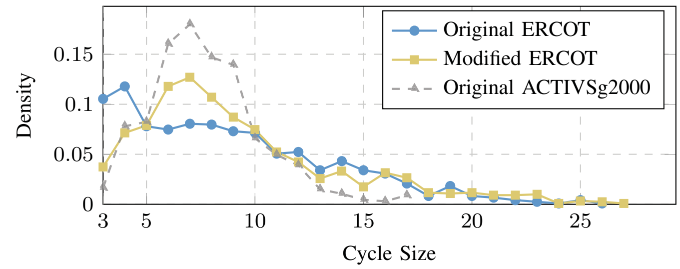


Figure 4. Cycle distribution for the ERCOT case before and after modification. The targeted original ACTIVSg2000 case is also shown for reference.

The values in Table III, however, provide further evidence to the changes. First, losses increase following the modification, which agrees with the simultaneously observed increase in average L_s. That loading in the system increase is further supported by a slight increase in the average $\Delta\theta$ between adjacent buses, suggesting higher real power flows. Furthermore, it is noteworthy that L_s behaves more like an Exponential distribution, which is seen by the decrease in the KL-divergence between L_s and the Exponential distribution parametrized by the Maximum Likelihood Estimator, which in the case of Exponential is simply the mean value, μ_{L_s}.

The change also manifests itself in the cost to operate the system, which increases as well. Further insight is possible in this particular case, since the generator commitment happens to not change between original and modified cases. Letting, C_0 and l_0 be the original total cost and losses respectively, C_m and l_m the modified total cost and losses respectively, and $\bar{p}$ be the average of all the LMP prices in the modified case,

$$(C_m - C_0) - \bar{p}(l_m - l_0) = -\$97.46. \tag{3}$$

In words: the difference in cost between the two scenarios can be largely attributed to losses.

C. Interpretation

Each of the impacts from the previous sections is on its own rather small. However, the culmination of all of these effects strongly suggest that altering the cycle distribution did in fact change the operational behavior of these power system. Furthermore, this change largely agrees with the initial hypothesis, which related cycles to current paths. The basic intuition that cycles imply parallel current paths, suggests that a higher density of larger cycles means more parallel paths, and therefore, reduces loading on any individual branch. Decreases in losses, L_s, $\Delta\theta$, and cost, all support this hypothesis. Finally, as the cycle distribution changed, the loading distribution, measured by L_s, also shifted. The data represented here is one example of a rewiring. However, out of 25 additional runs only 1 showed a different trend in terms of losses, and non showed a different trend in terms of average L_s.

While the direction of change is clear, its magnitude is fairly small. This test shows that the cycle distribution impacts the loading distribution of a power system, however, it is clearly one of many marginal effects. In other words, the cycle distribution affects but does not *determine* the loading distribution. The implication of our finding is that a correct cycle distribution is not *sufficient* for correct operation statistics. The closer resemblance to an Exponential distribution following the modification provides some evidence for the *necessity* of a correct cycle distribution to achieve a correct loading distribution. Since we are only looking at a marginal effect, rather than the full joint distribution of all variables, we cannot make a definitive claim of necessity. The impact, however, is clear and therefore, if this feature is not matched others may have to be adversely manipulated to achieve desirable operational behaviors.

TABLE IV
KL-DIVERGENCE OF ERCOT CYCLE DISTRIBUTION

	original case	modified case
$D_{KL}(p_{\text{ERCOT}} \vert q_{\text{ACTIVSg2000}})$	0.5243	0.4224
$D_{KL}(p_{\text{ERCOT}} \vert q_{\text{ERCOT}})$	0.0584	0.0891

D. A Reverse Experiment

In an effort to further strengthen the argument, another test is performed where the ERCOT case is modified to try and get its cycle distribution closer to that of the original ACTIVSg2000 case. The results of the cycle modification are shown in Figure 4. While it appears visually that movement in the "right" direction is made, closer numerical evaluation show this is far less successful a modification than in the ACTIVSg2000 case. Table IV shows the KL-divergence for the original and modified cases similar to Table II. While the modified ERCOT case is *more* similar to the original ACTIVSg2000 case it is still quite different. At the same time, the modified case remains closer, at least in the KL-divergence sense, to the original ERCOT case than to the original ACTIVSg2000 case.

With the caveat that the structural change is less successful, a similar and opposite trend is seen in the operational statistics. Table V shows that as the cycle distribution is pushed towards something more like the original ACTIVSg2000 case losses *decrease*, as does average L_s, and $\Delta\theta$. These changes similarly translate to a *decrease* in the total cost, however, in this case the unit commitment does change and therefore it is no longer possible to attribute the change in cost directly to the losses. Since the structural change is much smaller as previously

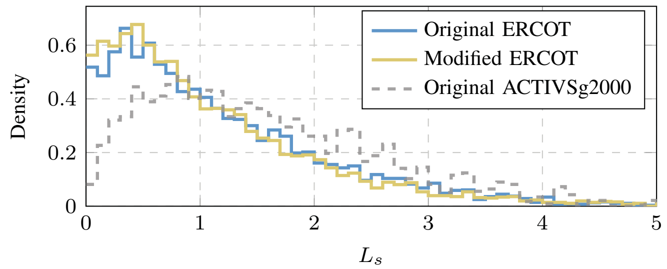


Figure 5. Distribution of L_s for the ERCOT case before and after modification. The original ACTIVSg2000 case is also shown for reference.

TABLE V
ERCOT OPERATIONAL VALUES

Quantity	Original	Modified Case
Cost [$]	1,740,845.88	1,739,953.28
Average LMP [$]	20.15	20.12
Losses [MW]	1570.72	1412.80
Average L_s	1.2065	1.1422
Average $\Delta\theta$ [degree]	2.42	2.41
$D_{KL}(L_s \vert \text{Exp}(\mu_{L_s}))$	0.0358	0.0278

discussed, the observed changes are also smaller compared to those in Section III-B. Still, the fact that the changes are consistent with those in Section III-B further supports the conclusion that structural cycles impact the loading distribution in the system.

E. Results from Modified Algorithm

Following the observations discussed above, a tweak was made to the ACTIVS algorithm [3] to address the cycles issue. A second 2000 bus case on the ERCOT footprint was created, which we will refer to as `ACTIVSg2000 v2`. Figure 6a shows that the algorithm tweak in fact alters the cycle distribution to be more similar to that observed in the ERCOT case. The fit parameters to the Negative Binomial distribution are,

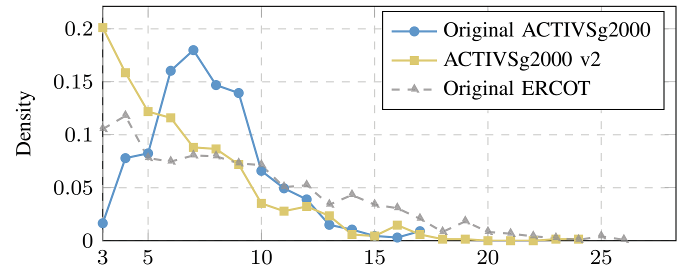
$$p = 0.69, \qquad r = 1.53,$$

which are significantly closer to the ERCOT values in Table I.

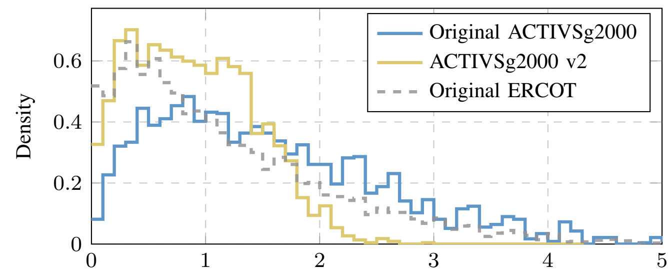
Since there are many other elements involved in the generation of cases, conclusions from a direct comparison are limited. For example, the two version have different load, which obviously impacts the loading distribution. Nonetheless, figure 6b shows the L_s distribution has also changed and in a manner consistent with the change seen in Figure 3, i.e., the peak shifting somewhat to the left. While the change cannot be strictly attributed to the change in cycle distribution, the agreement with the previous findings serve as further support for the hypothesis.

IV. DYNAMICS EXAMPLE

As the ACTIVS cases begin to integrate data for dynamic simulations [15], there is a need to consider validation methods and metrics in this area as well.

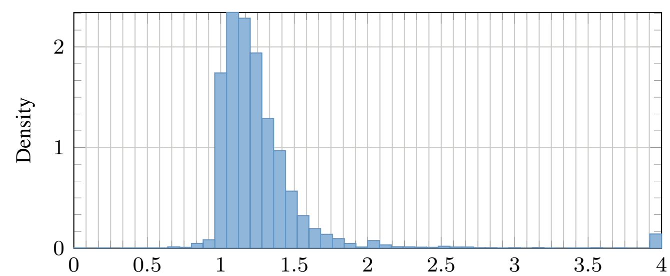


(a) Cycle Size

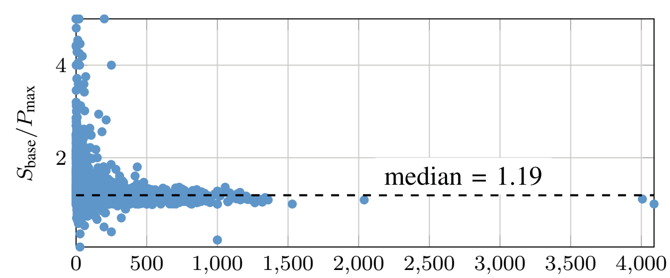


(b) L_s

Figure 6. Cycle and L_s distributions following a tweak to the ACTIVS algorithm to address the cycle distribution.



(a) $S_{\text{base}}/P_{\text{max}}$



(b) P_{max} [MW]

Figure 7. Relationship between generator rating and capacity. (a) shows the distribution of the ratio between the two, while (b) shows that this ratio depends on the capacity of the generator.

A. Experiment Description

The capacity of a generator, P_{max}, and its rating, S_{base}, are generally not exactly the same, as seen in Figure 7, based on data from an Eastern Interconnect (EI) case with over 8,000 generators. The ratio in the ACTIVSg2000 case between S_{base} and P_{max} is currently set to roughly match the median at 1.2. Since many generators operate at or close to capacity we consider the impact the $S_{\text{base}}/P_{\text{max}}$ has.

To test this relationship, two modified ACTIVSg2000 cases

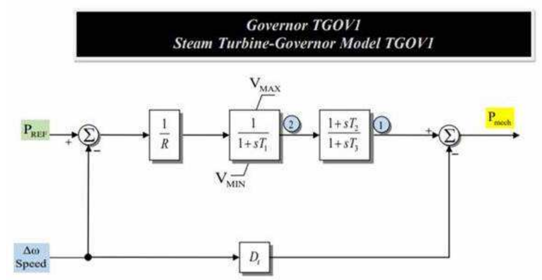


Figure 8. Governor block diagram from PowerWorld Simulator, showing valve limits (V_{MAX}, V_{MIN}). These are per unit values in $T_{\mathrm{rate}} \approx S_{\mathrm{base}}$.

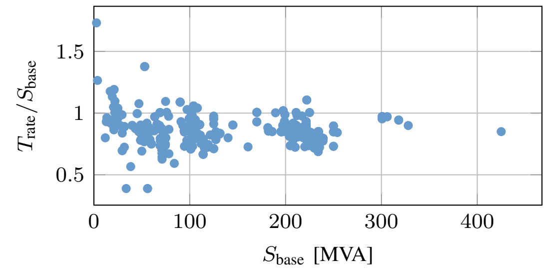


Figure 9. Ratio of turbine rating to generator rating versus generator rating. This plot helps justify the simplifying assumption made in simulation that $T_{\mathrm{rate}} = S_{\mathrm{base}}$.

are created, with $S_{\mathrm{base}}/P_{\mathrm{max}}$ set to 1.1 and 1.0, respectively. In each of the resulting three cases, $N-1$ contingencies are run and statistics about the resulting transient behavior collected. Three types of contingencies are used to model both large and small disturbances. These are:

1) generator outage,
2) line fault followed by trip, and
3) three-phase load bus faults.

Results are evaluated with respect to the Frequency Response, also known as β, which we calculate as:

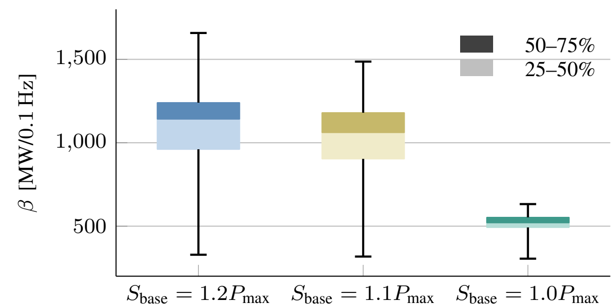
$$\beta = \frac{\text{MW lost}}{10(60\,\text{Hz} - M_f)}, \tag{4}$$

where M_f is the minimum frequency during the contingency event. The factor of 10 is due to the traditional reporting units of MW/0.1 HZ.

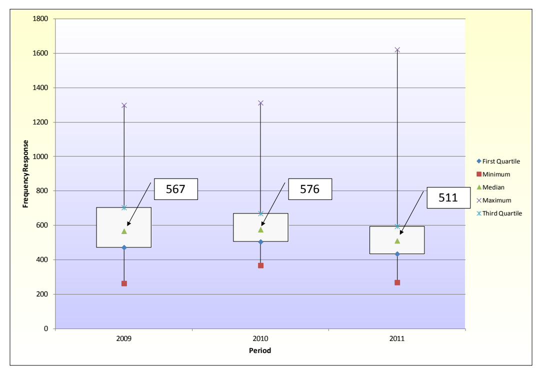
B. Rational

Many governor models, such as the one shown in Figure 8, limit the output power via valve limits. These per-unit values are based on the turbine rating, T_{rate}. This rating is, in turn, very closely related to the generator MVA base, S_{base}. In fact, out of 5336 models in the EI model, only around 12.9% make use of a T_{rate} different from S_{base}. About 1.3% of the governor models have $T_{\mathrm{rate}} > S_{\mathrm{base}}$, while the remaining 11.6% have $T_{\mathrm{rate}} < S_{\mathrm{base}}$. All of these cases are shown in Figure 9, where it is evident that the ratio strays substantially from 1 only for fairly small machines. Considering the small number of affected machines and the relative small impact of the change, we make the simplifying assumption that $T_{\mathrm{rate}} = S_{\mathrm{base}}$. As such, it is a logical hypothesis that changing S_{base} will impact transient behavior.

This simplification reduces the performance of generators that would otherwise have $T_{\mathrm{rate}} > S_{\mathrm{base}}$, however, as Figure 9 illustrates, the impact is minimal. At the same time it improves the performance of generators with $T_{\mathrm{rate}} < S_{\mathrm{base}}$. First of all, these ratios are also mostly close to one. Secondly, performance is improved, we do not have to worry about generating unreasonably constrained cases. This is similar in spirit to removing the line limits in Section III.



(a) Frequency Response β for ACTIVSg2000 cases.



(b) ERCOT Frequency Response taken from [16]

Figure 10. Frequency Response following contingencies in three different ACTIVSg2000 cases and on the ERCOT system. The results in (a) suggest that a mixture of the three cases will help achieve results similar to those in (b).

C. Results

The statistics collected from the contingency runs are reported in Figure 10a. As expected, more extreme responses occur as the $S_{\mathrm{base}}/P_{\mathrm{max}}$ ratio is reduced, which is evident by the smaller maximum β values. When comparing Figures 10a and 10b, we see that neither of the three cases quite matches data from the ERCOT interconnect reported in [16]. On the one hand, the average β from the case where $S_{\mathrm{base}} = P_{\mathrm{max}}$ matches the ERCOT data far better. On the other hand, the extreme values are far more accurate in the higher ratio cases. Beyond confirming the intuition that the ratio of $S_{\mathrm{base}}/P_{\mathrm{max}}$ affects transient behavior, these results also suggest that it is important to capture more of the distribution as shown in Figure 7. Including more generators with ratios close to, or even below, 1 will help bring the median β closer to realistic levels, while keeping some generators at higher ratios will help help keep some of the extreme values.

V. Conclusion

This paper takes a deeper look at how synthetic power grid models are validated, and the implications of chosen metrics. Our goal is to evolve a process that is useful for validating synthetic power system models however derived, and this work is a step in that process. The majority of power system test case users are interested in how the systems operate, which we view as system outputs. From the test case creation perspective, it is desirable, to place constraints on system inputs: parameters including topology, generator ratings, etc. This paper formulates an approach, illustrated by example, for evaluating the link between the statistical distribution of an input and the statistics of operational behavior. While direct functional relationships are not derived, demonstrating connection between input values and operation helps justify the validation constraints placed on the inputs. Testing validation metrics in this manner can identify impactful ones, thus reducing the burden of meeting unnecessary criteria. At the same time, each added critera increases confidence that the resulting system will operate in the desired manner, on the basis of observations for how it pushes the system towards particular operation.

The novelty in the work is its emphasis on *distributions*. Rather than asking how a given parameter or feature impacts system behavior, we show that a *variety* in parameter values is needed to achieve the desired operations, and furthermore, the distribution of values itself makes a difference. Ultimately, a "joint distribution," which combines any possible operational criteria, with system parameters, is the goal. Unfortunately, merely listing all of the criteria is impractical if not impossible, leaving a full description of the distribution out of reach.

This paper focused on the topological cycle distribution and the distribution of generator $S_{\text{base}}/P_{\text{max}}$, which are both marginal distribution of the joint distribution in all variables and criteria. Section III showed how the the cycle distribution, a system parameter, impacts operational distributions of line loading (L_s), as well as, losses, and operation cost. Section IV showed how $S_{\text{base}}/P_{\text{max}}$, another system parameter, impacts the frequency response (β) distribution of under various contingency scenarios. The results presented suggest that the input distributions (cycles, and $S_{\text{base}}/P_{\text{max}}$) have an impact on the output distributions (L_s, β, etc.). It is also clear from the results, that the input distributions considered are not solely responsible for the observed output statistics. We therefore conclude that the input distributions are *marginal* to the observed outputs. Finally, since the outputs considered here are marginals of the complete joint distribution, by matching them the solution gets closer to the desired result.

The motivation we advocate, is that matching more and more operational marginal distributions, strengthens the argument that the power system case in general converges towards the desired joint distribution. Matching the operational distributions can, in turn, be achieved by matching specific system parameter distributions, shown via the methodology presented, to impact said operations.

VI. Acknowledgements

This work was funded in part by the Advanced Research Projects Agency- Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000714. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] ARPA-E Grid Data. [Online]. Available: https://arpa-e.energy.gov/?q=arpa-e-programs/grid-data

[2] A. B. Birchfield, E. Schweitzer, M. H. Athari, T. Xu, T. J. Overbye, A. Scaglione, and Z. Wang, "A metric-based validation process to assess the realism of synthetic power grids," *Energies*, vol. 10, no. 8, 2017.

[3] 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, 2017.

[4] E. Schweitzer, A. Scaglione, and R. Thomas, "The validation of synthetic power system cases," in *IREP'2017 Symposium*, 2017.

[5] "Procedures for validation of powerflow and dynamics cases," NERC Model Validation Working Group, Tech. Rep., 2012.

[6] "Power System Model Validation," NERC Model Validation Task Force of the Transmission Issues Subcommittee, Tech. Rep., December 2010.

[7] "Steady-State and Dynamic System Model Validation," NERC MOD-033-1, Tech. Rep., 2014.

[8] TAMU Electric Grid Test Cases. [Online]. Available: https://electricgrids.engr.tamu.edu/

[9] G. Kirchhoff, "Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird," *Annalen der Physik*, vol. 148, no. 12, pp. 497–508, 1847.

[10] F. Berger, P. Gritzmann, and S. de Vries, "Minimum cycle bases for network graphs," *Algorithmica*, vol. 40, no. 1, pp. 51–62, 2004.

[11] J. D. Horton, "A polynomial-time algorithm to find the shortest cycle basis of a graph," *SIAM Journal on Computing*, vol. 16, no. 2, pp. 358–366, 1987.

[12] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951.

[13] R. Zimmerman, C. Murillo-Sánchez, and R. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb 2011.

[14] J. D. Glover, M. Sarma, and T. Overbye, *Power System Analysis & Design*, 5th, Ed. Cengage Learning, 2012.

[15] T. Xu, A. Birchfield, K. Shetye, and T. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *IREP'2017 Symposium*, 2017.

[16] "Frequency Response Initiative Report: The Reliability Role of Frequency Response," North American Electric Reliability Corporation (NERC), Tech. Rep., October 2012.