Evaluation of Performance Metrics for Electric Grid Operational Scenarios

Yijing Liu, Matthew Gaskamp, Zeyu Mao, Diana Wallison
Komal S Shetye, Katherine Davis and Thomas Overbye
Texas A&M University
College Station, Texas

Donald Morrow
Grid Focus LLC
Cary, North Carolina
don@gridfocusllc.com

2020 52nd North American Power Symposium (NAPS) | 978-1-7281-8192-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/NAPS50074.2021.9449731

***Abstract*—Today's power grid infrastructure is evolving faster than ever, as is its complexity. To effectively address the challenges of grid modernization, upcoming power engineers need more first-hand, practical experience operating these power systems. Several interactive operational scenarios have been developed in order to allow the users to get hands-on experience with a simulated power grid. This paper develops several metrics to evaluate the trainees' performance during an operational scenario. Operating reserve should be quickly available to maintain system reliability. Voltage schedules also play an important role in keeping equipment within system operating limits. Transmission line loading constraints prevent the system from exceeding thermal and stability limits after credible contingency events. In this work, these three metrics are monitored throughout the interactive simulation and used to guage user performance. To give general insight on how well a trainee performs, an overall Performance Score (PS) combing these three metrics is calculated at the end of assessment.**

***Index Terms*—Grid modernization, synthetic power grids, operational scenario, reliability, voltage stability, evaluation metrics**

I. INTRODUCTION

Students and engineers who (will) work in the power industry must be ready to analyze and operate large-scale power grids. To that end, prior training or experience in such systems can be quite valuable. Though several small cases are available to the public and can be used for teaching and training purposes, access to actual power grid data sets can be hard to obtain. There is an acute need to expose students to large-scale scenarios where they can get hands-on experience operating synthetic grids in a variety of conditions.

Power system models that allow interactive simulations are critical in developing operational scenarios. Previous works [1–4] describe the development of large-scale synthetic electric grids and their applications in university power system courses. They bridge the gap between publicly available test cases and complexity of actual power grids. Various scenarios utilizing a 2000-bus synthetic case are assigned to the lab portion of a undergraduate class throughout the semester. Starting in Fall 2018 undergrad students in the Power System Control and Operation course work collaboratively in their final lab to operate this 2000-bus synthetic Texas grid [4].

Human factors play an important role in electric grid operations [5–8]. This paper builds upon the operational scenarios from [1–4] to provide students with the opportunity to operate large scale synthetic grids, while also gaining insight into the students' decision making process. The interactive power system simulation package of [9], which models dynamics using a ½ cycle time step, was used for the simulation, in which numerous communication and interactive control action capabilities are included. Participants are provided with the data and information needed to support precise and timely decision-making throughout the real-time simulation scenarios. These custom designed scenarios can be used to give people hands on operational experience, as well as test new control interface designs, algorithms and scenarios to measure their impact on human factors.

Performance evaluation is not only valuable to the researcher, but also acts as quantitative performance feedback to the user and is hence provided to the student/engineer in the process of teaching/training. The motivation of this paper is to present the evaluation metrics that can assess the performance of a participant with a rating score. Analyses and visualization tools have been developed to enhance participants' situational awareness and decision-making in normal and abnormal power system conditions [10]. Work [11] developed and described grid resilience metrics and analysis methods, which allows for historical data to serve as the basis for creating grid resilience metrics. Reference [12] presented a project to enhance operator training programs by evaluating operator performance relative to a reference operator model obtained using optimal control theory. In [13], the authors defined a training evaluation methodology aiming to reduce the gap between required knowledge and knowledge of operators. Under certain situations, power imbalance may cause a rapid deterioration in power system security and stability, including large voltage deviations [14]. Line overloads and voltage violations [10] have impacts on power system security and limit effective and timely system operation. As a result, line flows and voltage magnitude are used as evaluation metrics. Real power reserves are calculated throughout the testing period, and used as another metric. These metrics are then combined and an overall Performance Score (PS) is given to the participant.

The following three sections are as follows: In section II, a set of metrics are adopted for performance evaluation. Section III presents evaluation of two users simulating the same system and scenario, independently, and comparing their performance. Section IV presents concluding remarks and directions for future work.

II. EVALUATION METRICS

This section provides some background on the definition and importance of operating reserve, voltage schedules and line congestion management. Evaluation metrics are developed correspondingly. At the end, a percentage score, Performance Score, is defined to give an overall score of the user's simulated grid operation.

A. Operating Reserve

Many of the properties of the power systems, including generation output, load fluctuations, and transmission component availability are variable. To maintain power system reliability, additional capacity above what is needed to meet current demands should be available either on-line or on-standby so that it can be called on to respond if load increases or generation decreases. This capacity is referred to as the operating reserve and is used to maintain the real power balance of the system. In most research and practice, the amount of operating reserve is greater than the capacity of the largest generator and a proportion of the peak load [15].

The operating reserve is equal to total operating capacity minus the electric load. Electricity capacity is the maximum electric output a generator can produce under certain conditions, and the total operating capacity is the net electricity capacity from all generators in the system [16]. Both dispatchable power sources (1) and renewable power sources (2) provide operating capacity when they are online.

$$P_{capacity} = P_{max} \tag{1}$$

$$P_{capacity} = P_{current} \tag{2}$$

Due to renewable energy's variable and uncontrollable nature, its maximum capacity is irrelevant and not used in (2).

When the difference between operating capacity and system demand (MW$_{OC}$ - MW$_{SD}$) is less than the minimum reserve requirement ($\min R$), an energy emergency situation is defined in light of N-1 reliability. Otherwise, the system is operating in the nominal energy adequate condition. A similar approach is used in real-world agencies like ERCOT.

B. Voltage Schedules

The system voltage schedule is part of the system operating plan to keep equipment within system operating limits. Different resources are operated to provide voltage control to maintain safe system voltage levels for the reliable transfer of real power to serve load [17]. Generators connected with the bulk electric system are issued a voltage magnitude (V$_{mag}$) schedule to be maintained. Though both generator internal voltages and network properties affects bus voltages, the voltage level of each bus is more dependent on the generator which is electrically nearest to that bus [18]. Thus generator bus voltages are monitored and used for evaluation without sacrificing accuracy. V$_{mag}$ of other buses (i.e. buses not connected with a generator) are also important, so they are monitored as well. Considering the dynamic nature of the power system, the voltage schedule target value and a tolerance band for V$_{mag}$ is monitored throughout the simulation.

- **Target Set Point**: V$_{mag}$ under normal conditions. It is defined as slightly above nominal voltage (i.e., 1.02 p.u.)
- **Tolerance Band**: a range of V$_{mag}$ which provides the max/min value under normal conditions. Excursions outside of a normal schedule are not precluded during abnormal conditions but they are limited in extent. Outside a bandwidth of say ± 0.05 around the target value (i.e., V$_{mag}$ outside range(0.97, 1.07)) is defined as "Warning" and outside a bandwidth of say ± 0.08 (i.e., V$_{mag}$ outside range (0.94, 1.10)) is defined as "Violation".

C. Transmission Line Loading

Transmission system constraints are set as a specific level or finite amount of power that can be transferred between two elements on the electric power grid to ensure that grid is operated in a secure manner [19, 20]. Congestion would arise when there is a need to increase flow across a transmission line, but such higher transmission throughput is thwarted by constraints. Transmission congestion may result in potential reliability problems when these constraints limit access to operating reserves required for secure operations. Different strategies can be used to relieve transmission congestion (i.e., changing the operation of the transmission system).

Transmission line loading is constantly monitored throughout the dynamic simulation. A transmission element loading of 75% is often considered very high in practice and a "Warning" is given under this situation. Full utilization (100% of transmission capacity) is rarely achievable due to reliability considerations, so a "Violation" is defined when a transmission line approaches its practical maximum utilization (i.e., exceeds 90% of its capacity).

D. Performance Score (PS)

Finally, we compute a novel metric to quantify the participant's operation of the scenario using the variables calculated in previous steps II-A - II-C:

$$RI = \lfloor \frac{100}{\alpha+\beta+\gamma} \times (\alpha \mathrm{e}^{-0.05a} + \beta \mathrm{e}^{-0.05b} + \gamma\sqrt{1-c^2}) \rceil^{1} \tag{3}$$

where a denotes as the count of buses that have voltage magnitude violations, b is the count of transmission lines that have line loading violations and c equals to the ratio of energy emergency duration to total simulation duration; α is the weight coefficient of the V_{mag} term, β is the weight coefficient of the line loading term and γ is the weight coefficient of the operating reserve term.

In (3), the weight coefficient needs careful consideration to describe and express the physical meaning of assigned score weights meaningfully and accurately. Initially, all three terms are assigned with the same weights ($\alpha = \beta = \gamma = 1$). But the authors will discuss the impact of different weights in III and might put different scalings on them in future work.

A natural exponential decay function ($\mathrm{e}^{-0.05x}$, $x > 0$) is used for the score calculation of V$_{mag}$ and the transmission lines to give a bigger penalty for the first several limiting elements. The mathematical model $\mathrm{e}^{-0.05x}$, $x > 0$ first decreases

---

[1] The symbol $\lfloor x \rceil$ represents rounding x to the nearest integer.

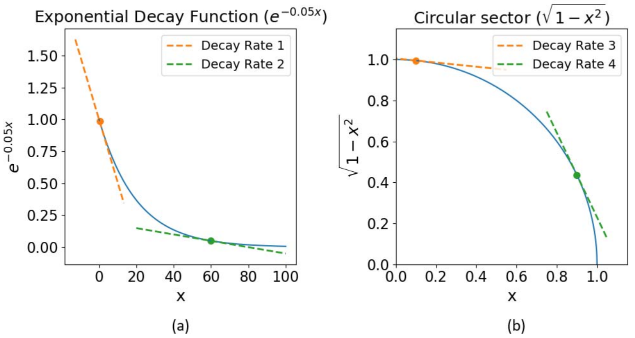

Fig. 1: Decay rate of exponential function and circular sector

rapidly with a small *0.05x* ($0 < x < 30$, as shown in Fig.1 (a) Decay Rate 1) and then decays much slower with a larger *0.05x* value ($x \geq 30$, Fig.1 (a) Decay Rate 2). The coefficient 0.05 is used in this function since each evaluation term with 10 violations is considered a passing score with a percentage score of $\mathrm{e}^{-0.05\times10} = 60\%$.

A circular sector function ($\sqrt{1-x^2}, 0 < x < 1$) is used in calculating the score of operating reserve. This function has a rather slow decay rate with *x* near the origin ($0 < x < 0.5$, as shown in Fig.1 (b) Decay Rate 3) and the function declines fast after $x > 0.5$ (Fig.1 (b) Decay Rate 4). We don't give a big penalty to a small *x* (which is the ratio of energy emergency period to total simulation duration) since energy emergency condition is more like an alert than an actual violation.

III. EVALUATION RESULTS

In this section, the performance metrics are demonstrated by comparing the dynamic simulation results of two participants each having run a single-user voltage control scenario utilizing the 2000-bus system - ACTIVSg2000 [21] (Fig. 2). During the scenario, which lasted 1800 seconds (30 minutes), the entire system load was first decreased for 900 seconds from 100% to 93% and in the last 900 seconds increased back to 100%. During the evaluation process, a software package that was developed in the Python programming language is used to collect case information such as Bus Nominal Voltage, Generator's maximum MW limit, etc. This software package, named EasySimAuto (ESA), greatly simplifies interfacing with Simulator's application programming interface (API) [22].

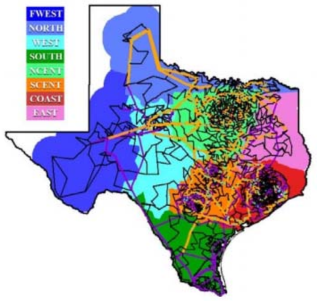

Fig. 2: One-line diagram of the 2000-bus case [23]

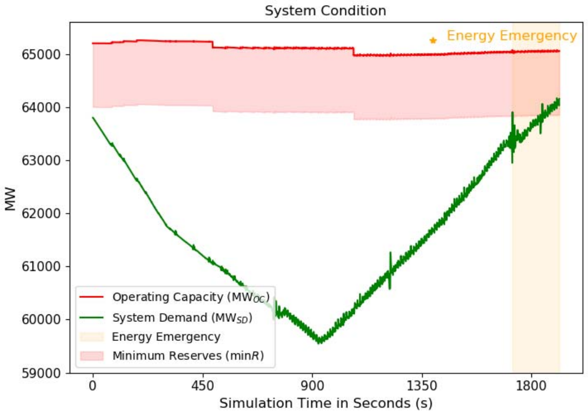

Fig. 3: Participant A: Operating Reserve

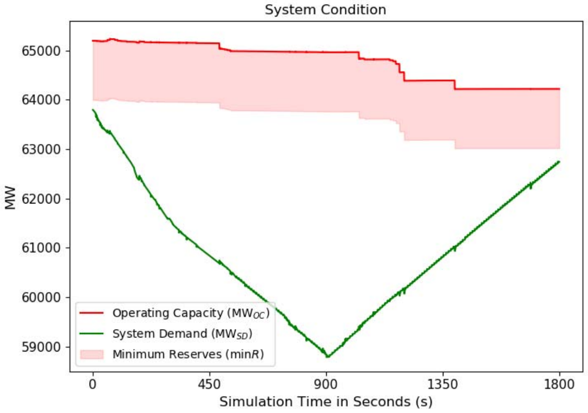

Fig. 4: Participant B: Operating Reserve

A. *Operating Reserve*

Although generation should meet demand, operating capacity always includes real power reserve to help maintain reliability in case of contingency events. The display in Fig. 3 shows MW_{OC}, MW_{SD}, and the $\min R$ for the simulation results from Participant A. Fig. 3 provides a snapshot of operating reserve evaluation results collected from Participant A after one of the interactive scenario tests. In this example, we can see that (MW_{OC} -MW_{SD}) $> \min R$ when the simulation starts, which is a sign of N-1 reliability. (MW_{OC} -MW_{SD}) varies continuously when Participant A adjusts generator voltage set-point. System demand also changes greatly through the simulation as defined in the scenario, and forms a big "V" shape in Fig. 3. (MW_{OC} -MW_{SD}) drops below $\min R$ at around 1620s, which triggers the energy emergency condition. The total amount of time when the system is under energy emergency condition is counted and used for the calculation of Performance Score. c = 0.1 with a corresponding score of $\frac{100}{3} \times \sqrt{1-0.1^2} = 33.16$. Fig. 4 shows the operating reserve from Participant B operating the same interactive scenario. With Participant B changing generator voltage set-point and the load variation defined in the scenario, the difference between (MW_{OC} -MW_{SD}) and $\min R$ varies through

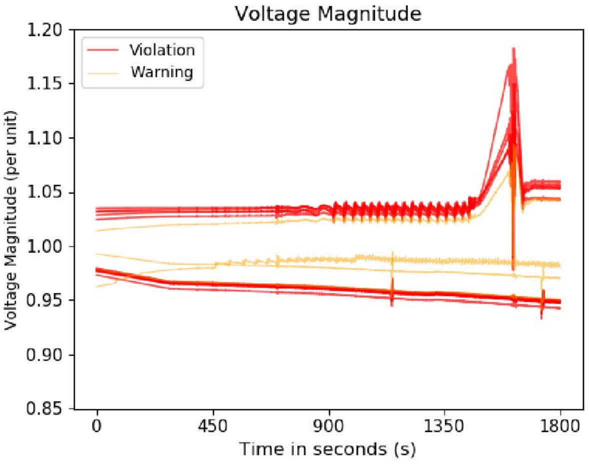

Fig. 5: Participant A: Voltage Magnitude
Number of Violation(s): 14

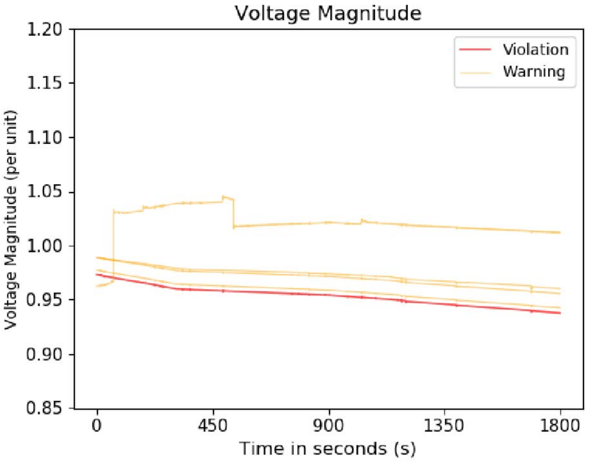

Fig. 6: Participant B: Voltage Magnitude
Number of Violation(s): 1

the simulation period, and (MW_{OC} -MW_{SD}) $> \min R$ all the time. As a result, B has a good score in this part ($\frac{100}{3} \times \sqrt{1-0^2} = 33.33$).

Though Participant A and B are exposed to the same scenario, they take different control actions, which leads to different results (A ends with an energy emergency situation, whereas B is still under energy adequate condition). Hence, this is effectively a human factors testing experiment, where the effectiveness of various user interfaces, scenarios, etc. can be evaluated based on user response and simulation impacts.

B. Voltage Schedule

Voltage magnitude for all buses are constantly monitored throughout the simulation. Only buses for which the Nominal Bus Voltage is above 115kV are selected in Performance Score calculation since in actual system operations, it is more critical to maintain the voltages of the higher voltage network, which is why they also have narrower limits than the lower voltage network. The dynamic simulation results from Participant A

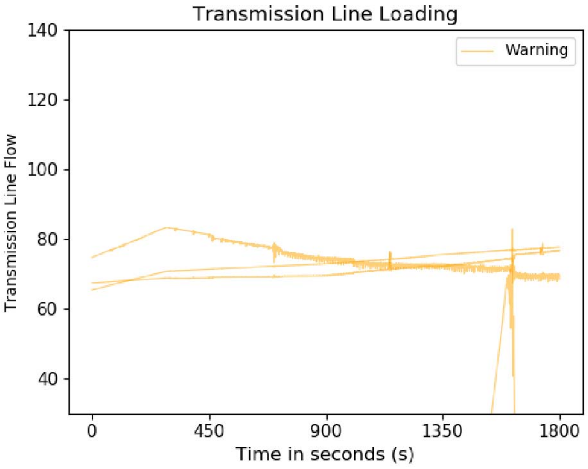

Fig. 7: Participant A: Transmission Line loading
Number of Violation(s): 0

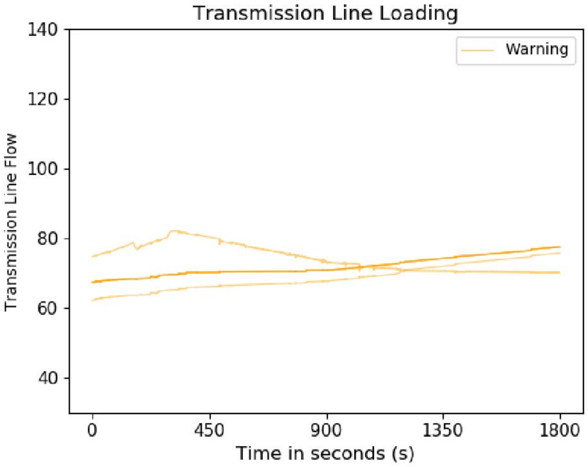

Fig. 8: Participant B: Transmission Line loading
Number of Violation(s): 0

are plotted in Fig. 5 with two different colored lines. All the buses with a V_{mag} “Violation” are plotted and marked red in the figure. Several representative buses (i.e., two buses that have the minimum/maximum V_{mag} value, and and two buses chosen at random) with a “Warning” are plotted and marked orange. The buses well within V_{mag} ± 0.05 tolerance band around the target value are not plotted in this figure for the sake of simplification. Around 1150s and 1620s, Participant A closed a large generator and opened it immediately when A noticed that this generator consumes a large amount of reactive power and these actions caused the voltage spikes and dips seen in the figure. For Participant A, a = 14. As mentioned in II-D, this work initialized the performance score calculation with $\alpha = \beta = \gamma = 1$ so Participant A has a corresponding score of $\frac{100}{3} \times \mathrm{e}^{-0.05\times 14} = 16.55$. As depicted in Fig. 6, Participant B has fewer V_{mag} violations and the voltage magnitude curves have lower envelope fluctuation than Participant A. For Participant B, a = 1 and the score equals to $\frac{100}{3} \times \mathrm{e}^{-0.0.5\times 1} = 31.70$.

C. Transmission Line Loading

Fig. 7 shows line flows on selected abnormal transmission lines from Participant A's dynamic results. The transmission system has many lines that have line loading warnings throughout this operation test, but there's no line violating limits defined in II-C. Selected four lines with "Warning" (i.e., two buses that have the minimum/maximum transmission line flow value, and other two random lines) are plotted and marked orange in Fig. 7. As discussed in III-B, the close and open generator actions caused the rapid increase in transmission line flow around 1150s and 1620s. For Participant A, no transmission line has a Violation, which means $b = 0$ and the score is $\frac{100}{3} \times \mathrm{e}^{-0.05\times 0} = 33.33$. As shown in Fig. 8, Participant B did as good as A in controlling line flows and got $b = 0$ with a score of $\frac{100}{3} \times \mathrm{e}^{-0.05\times 0} = 33.33$ for this part of evaluation.

D. Performance Score

The final step of the evaluation process is to calculate the Performance Score using (3). Table I provides details on Performance Score calculations for Participant A and B given that $\alpha = \beta = \gamma = 1$. Participant B did a better job than A in maintaining operating reserve, V_{mag}, and transmission line flows and B got a total score of $\lfloor 98.36 \rceil = 98$. Participant A had more violations than B in two sections and got a total performance score of $\lfloor 83.04 \rceil = 83$. Note that Participant A got some peak values in V_{mag} and line flow dynamics results, which may bring stability issues to the equipment if this happens in a real world control room. Participant A and B have taken quite different control actions during the scenario tests and these actions affect the simulated grid differently. These metrics help reveal the impacts of participants' decision-making process.

TABLE I: Performance Score Calculation Details
$(\alpha = \beta = \gamma = 1)$

Participant	Operating Reserve	V_{mag}	Line Loading	PS
A	33.16	16.55	33.33	83
B	33.33	31.70	33.33	98

TABLE II: Performance Score Calculation Details
$(\alpha = 0.4, \beta = 1.3, \gamma = 1.3)$

Participant	Operating Reserve	V_{mag}	Line Loading	PS
A	13.27	21.52	43.33	78
B	13.33	41.21	43.33	98

As a comparison, Table II shows how weight coefficients impact Performance Score calculation. With new coefficients $\alpha = 0.4, \beta = 1.3, \gamma = 1.3$, more penalties are given to the actual violations (i.e., V_{mag} and line flow violations) and less penalties are given to the system condition alerts (i.e., Operating Reserve). For Participant A, the total Performance Score has dropped since more weight has been put on the V_{mag} term, on which A got many violations. For Participant B, the total Performance Score doesn't change since they got almost perfect scores on all three terms and the sum of three weighted terms doesn't vary as the weights change. As such, the impact of weight coefficients on total Performance Score has been revealed.

IV. CONCLUSIONS

In this paper, several evaluation metrics were applied to measure a participant's performance during an interactive simulation. A single-user voltage control case was utilized to show that the adapted metrics are able to indicate level of voltage stability, level of reliability in light of real power reserves, and transmission system loading conditions. A Performance Score was calculated at the final step of dynamics results evaluation. The participant could benefit from the real-time grid operation training and this performance evaluation report.

An ongoing work is to investigate each metric and develop more case-specific details. Different reasonable actions will be defined under different energy emergency situations and corresponding grades will be given. Suitable weight coefficients will be considered for the metrics in (3) towards updating the Performance Score. It is of interest to combine current metrics with economics effects, such as economic unit commitment. We plan to report those works in near future.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Award Number ECCS-1916142 and the U.S. Department of Energy (DOE) under award DE-OE0000895.

REFERENCES

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

[2] A. B. Birchfield, T. J. Overbye, and K. R. Davis, "Educational applications of large synthetic power grids," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 765–772, 2018.

[3] T. Xu, A. B. Birchfield, and T. J. Overbye, "Modeling, tuning, and validating system dynamics in synthetic electric grids," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6501–6509, 2018.

[4] Z. Mao, H. Huang, and K. Davis, "W4ips: A web-based interactive power system simulation environment for power system security analysis," *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 2020.

[5] T. J. Overbye, D. A. Wiegmann, A. M. Rich, and Y. Sun, "Human factors aspects of power system voltage contour visualizations," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 76–82, 2003.

[6] M. Panteli and D. S. Kirschen, "Situation awareness in power systems: Theory, challenges and applications," *Electric Power Systems Research*, vol. 122, pp. 140–151, 2015.

[7] F. L. Greitzer, A. Schur, M. Paget, and R. T. Guttromson, "A sensemaking perspective on situation awareness in power grid operations," in *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*. IEEE, 2008, pp. 1–6.

[8] J. H. Obradovich, "Understanding cognitive and collaborative work: Observations in an electric transmission operations control center," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 55, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2011, pp. 247–251.

[9] T. J. Overbye, Z. Mao, A. Birchfield, J. D. Weber, and M. Davis, “An interactive, stand-alone and multi-user power system simulator for the pmu time frame,” in *2019 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2019, pp. 1–6.

[10] B. Leonardi and V. Ajjarapu, “Investigation of various generator reactive power reserve (grpr) definitions for online voltage stability/security assessment,” in *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*. IEEE, 2008, pp. 1–7.

[11] E. Vugrin, A. Castillo, and C. Silva-Monroy, “Resilience metrics for the electric power system: A performance-based approach,” *Report: SAND2017-1493*, 2017.

[12] W.-L. Hu, C. Rivetta, E. MacDonald, and D. P. Chassin, “Optimal operator training reference models for human-in-the-loop systems,” in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[13] M. Bronzini, S. Bruno, M. De Benedictis, S. Lamonaca, M. La Scala, G. Rotondo, and U. Stecchi, “Operator training simulator for power systems: Training evaluation methodologies based on fuzzy logic,” in *2010 IEEE International Symposium on Industrial Electronics*. IEEE, 2010, pp. 2035–2040.

[14] X. Xu, H. Zhang, C. Li, Y. Liu, W. Li, and V. Terzija, “Optimization of the event-driven emergency load-shedding considering transient security and stability constraints,” *IEEE Transactions on Power Systems*, pp. 2581–2592, 2016.

[15] J. Wang, X. Wang, and Y. Wu, “Operating reserve model in the power market,” *IEEE Transactions on Power systems*, vol. 20, no. 1, pp. 223–229, 2005.

[16] “Electricity generation capacity”. https://www.eia.gov/tools/faqs/faq.php?id=101&t=3.

[17] “Reliability Guideline: Reactive Power Planning”. https://www.nerc.com/comm/PC_Reliability_Guidelines_DL/Reliability%20Guideline%20-%20Reactive%20Power%20Planning.pdf.

[18] Y. Liu, T. Xu, and T. J. Overbye, “Locational dependence of inertia’s impacts on critical clearing time,” in *2018 North American Power Symposium (NAPS)*. IEEE, 2018, pp. 1–6.

[19] A. Kumar, S. Srivastava, and S. Singh, “Congestion management in competitive power market: A bibliographical survey,” *Electric Power Systems Research*, vol. 76, pp. 153–164, 2005.

[20] “Transmission Constraints and Congestion in the Western and Eastern Interconnections”. https://energy.gov/sites/prod/files/2014/02/f7/TransConstraintsCongestion-01-23-2014%20.pdf.

[21] “Electric Grid Test Case Repository”. [Online]. Available: https://electricgrids.engr.tamu.edu/

[22] B. L. Thayer, Z. Mao, Y. Liu, K. Davis, and T. J. Overbye, “Easy simauto (esa): A python package that simplifies interacting with powerworld simulator,” *Journal of Open Source Software*, vol. 5, no. 50, p. 2289, 2020. [Online]. Available: https://doi.org/10.21105/joss.02289

[23] T. Xu, Y. Liu, and T. J. Overbye, “Metric development for evaluating inertia’s locational impacts on system primary frequency response,” in *2018 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2018, pp. 1–6.