Planning for Resilient Power Distribution Systems using Risk-Based Quantification and Q-Learning

Shuva Paul and Anamika Dubey
School of Electrical Engineering and Computer Science
Washington State University

Shiva Poudel
Electricity Infrastructure and Buildings
Pacific Northwest National Laboratory

2021 IEEE Power & Energy Society General Meeting (PESGM) | 978-1-6654-0507-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/PESGM46819.2021.9638213

***Abstract*—Grid hardening is one of the most effective approaches to improving the power distribution systems' resilience against extreme events. Unfortunately, hardening and upgrading the entire system is prohibitively expensive, mostly to protect against high-impact low-probability (HILP) events. This paper adopts a reinforcement learning (RL) algorithm to effectively search for the optimal hardening strategy to enhance power distribution systems' resilience – the resilience is quantified using a risk-based metric for the loss of load probability. The proposed *Q-learning* algorithm identifies the sequential optimal actions for grid hardening that minimizes the Conditional Value at Risk ($CVaR$) for the loss of load for a given budget. A case study is presented using the IEEE 123-bus test feeder to demonstrate the proposed approach's effectiveness in optimally allocating the budget-limited resources in resilient distribution system planning.**

***Index Terms*—Resource planning, resilience, grid hardening, machine learning, power distribution systems.**

I. Introduction

The growing frequency and duration of extreme weather events significantly increase the power grid's propensity for extended disruptions. Power distribution systems are especially prone to extended outages, given their radial topology with limited visibility and controllability–around 90% of the power outage incidents are related to the distribution systems in the United States [1]. The staggering cost of power system outages and their impacts on personal safety demands expedited incorporation of resilience in the aging and stressed power distribution systems towards extreme weather events [2]. Thus, it is of growing concern to minimize the impacts of such catastrophic events with effective resilience-enhancing strategies.

Recently, the resilience oriented design of the power distribution system is of growing research interest. The efforts are focused on how to harden/upgrade the system given the budget constraints optimally. A few popular methods include the optimization-based decision support tool and robust modeling for designing/upgrading the distribution network [3]–[5], and graph-theoretic formulation for resilient distribution grid topology designs [6]. While promising, the traditional distribution grid planning methods pose limitations due to an inadequate characterization of HILP events and/or overly simplified representation of power systems operational models [7]. Incorporating HILP events requires a risk-based characterization of resilience [8]. Unfortunately, incorporating such risk-based metrics in a model-based optimization setting for planning poses modeling and computational challenges [9]. This calls for a resilience planning framework that uses the risk-based characterization of HILP events' impacts in identifying optimal strategies to enhance distribution grid resilience.

In recent years, to address some of the limitations of the model-based techniques, machine learning (ML) algorithms have gained popularity with strategic decision-making in the power grid [10]. For example, ML tools help respond to the HILP events by generating accurate component outages and load curtailment forecasts [11]. Recent work also focuses on generating synthetic and realistic power grid data to analyze different events and their impacts on the distribution systems [12]. In [13], the authors analyzed the power grid's resilience utilizing different machine learning techniques. Likewise, *Q-learning* has been used to identify worst impact zones in the power distribution systems [14], and vulnerability assessment of the power transmission system using different machine learning and game theory techniques [15]. The related literature, however, is sparse on resilient distribution grid planning using ML techniques.

Inspired by the need to improve the distribution system resilience, our primary focus of this study is to identify the optimal resilience enhancement strategy with a risk-based resilience quantification. We focus on the specific distribution grid problem hardening to wind storms by undergrounding the distribution lines toward this goal. Here, we adopted a risk-averse approach to resource planning in the distribution grid that includes adequate models to characterize HILP events and incorporates advanced operations [8]. To manage the resulting problem formulation's computational tractability, we develop a framework based on the *Q-learning* approach to generate the optimal decisions for line hardening for a given budget constraint. Resilience is quantified using risk-based metrics, value-at-risk (VaR), and conditional value-at-risk ($CVaR$), as proposed in our previous articles [8]. The planning problem is modeled as a Markov decision process (MDP) with the sequential line upgrade actions to minimize the resulting $CVaR$. We validate the proposed approach using simulation on the IEEE 123-bus test system. While other RL approaches can be used, *Q-learning* employed in this work was successful in obtaining strategic line hardening decisions. Further research

This work was supported by the National Science Foundation (NSF) under Grant #1944142.

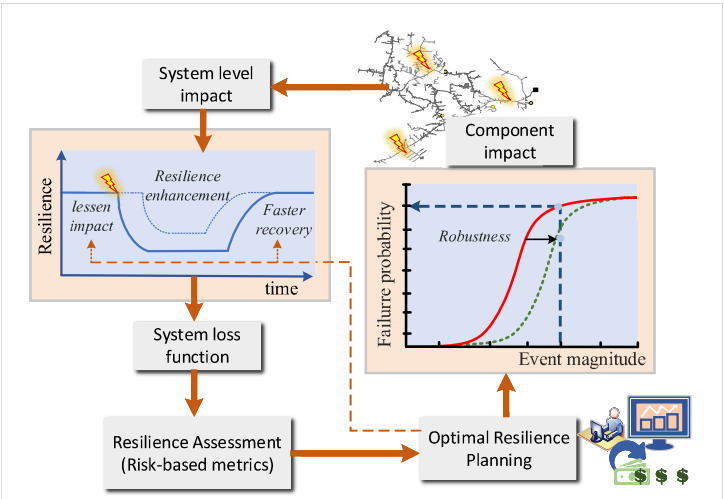

Fig. 1. Optimal planing for enhancing resilience to weather events via line hardening using fragility curves.

is needed to validate the scalability of the proposed algorithm for larger systems.

II. Problem Description

Investments that augment distribution grid resilience include system hardening and operational measures such as deploying new sensing and control technologies and distributed energy resources for faster recovery [2]. Hardening measures refer to topology and structural changes to make the network less susceptible to damage under extreme events. In contrast, operational measures refer to smart actions to deal with the emergency as it unfolds effectively. Grid hardening denoted as infrastructure reinforcement actions, is one of the most effective methods to protect systems against extreme weather events. Various grid hardening strategies include overhead structure reinforcement, vegetation management, undergrounding, and integrating black-start resources.

Fig. 1 shows the procedure of resilience enhancement using the concept of fragility curves. These component-level fragility curves can be used to model the impacts of hurricanes or other high-wind events on power system components. By hardening, components are made robust to higher intensities of weather events by reducing the probability of wind-induced damages. In essence, hardening modifies the fragility curve for the distribution system components and reduces system loss. This probabilistic loss can be measured using a conceptual resilience curve as the event progresses. The system tries to avoid, react, and/or recover from such an event using the predefined risk-based metrics. In a long-term planning problem, the sequential hardening of the system is more relevant. As the budget becomes available, the simulation-based framework can assess the system's resilience for a probabilistic event and decide on line hardening. Such a process is beneficial for utilities as planning for extreme events is an incremental process. It is essential to learn from past events to plan efficiently for the future.

Although system hardening could reduce component failures and restoration efforts, hardening and modernizing the entire system is prohibitively expensive. Hence, it is imperative to allocate budget-limited resources effectively. As a result, the problem of optimally hardening the given network for a given budget constraint is of interest.

III. Methodology

A probabilistic formulation of the risk-based resilience measurement framework for a distribution system can characterize the HILP events and their impacts on the distribution network. The adoption of *Q-learning* framework takes the candidate lines for hardening, measures the resilience based on the risk-based metric, and through the learning process, identifies the optimal line for hardening from the set of candidate lines. In this section, we discuss the approach to measure the resilience of the system using risk-based metrics. Next, we detail the RL-based approach to identify the optimal line hardening decisions.

A. Risk-based metrics

The planning problem's main goal is to optimally allocate the available resources to minimize the highest-impact events' risk. This requires a mechanism to quantify the risks associated with the highest impact events for a given resource allocation. In this work, we use the framework based on Monte-Carlo simulation [8] to evaluate the impacts of HILP events (e.g., wind storms) on distribution system performance and quantify the risks posed by such events on the system's resilience. The risk-based quantitative measures that are adopted in this work are: Value-at-risk (VaR) and conditional-value-at risk ($CVaR$) [8].

VaR calculates the maximum loss expected over a given period and given a specified degree of confidence. VaR refers to the lowest amount ζ such that with probability α, the loss will not exceed ζ. In the case of resilience quantification, the decrease in resilience or system loss function, $U(I)$, is measured to quantify the VaR metric. The probability of system loss, $U(I)$, when impacted by event I not exceeding a threshold ζ is given by (1).

$$\psi(\zeta) = \int_{U(I)\leq\zeta} p(I)dI \tag{1}$$

where, ψ is the cumulative distribution function for the loss which determines the behavior of random event, I. By definition, with respect to a specified probability level α in $(0, 1)$, VaR_α is given by (2).

$$VaR_\alpha = \min\{\zeta \in \mathbb{R} : \psi(\zeta) \geq \alpha\} \tag{2}$$

Similarly, the $CVaR$ metric is defined to calculate the expected system loss due to probabilistic threat events, conditioned on the events being HILP, i.e., due to the top 1-α% of highest impact events (See Fig. 2). In our case, the metric finds the expected resilience loss in MWh, and is mathematically represented in (3).

$$CVaR_\alpha = (1-\alpha)^{-1} \int_{U(I)\geq VaR_\alpha} U(I)\ p(I)\ dI. \tag{3}$$

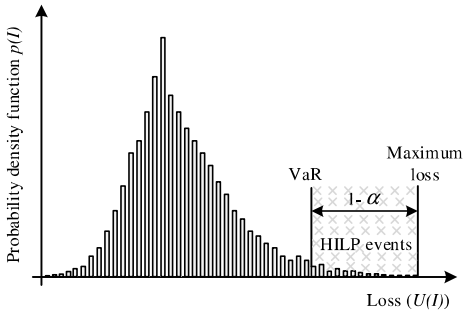

Fig. 2. VaR and $CVaR$ assessment for a probabilistic weather event. HILP events identified as the top $(1-\alpha)\%$ high impact disruptions.

B. Q- Learning

Q-learning is a Reinforcement Learning (RL) algorithm and one of the semisupervised machine learning algorithms. Typically an agent interacts with the environment by executing actions. In return, the environment provides feedback to the learning agent based on the action evaluation, which we call rewards. An agent aims to maximize the reward by taking optimal actions. A typical learning framework involves an environment, states ($s \in S$), corresponding actions ($a \in A$), and rewards (r). The environment is generally formed as Markov Decision Process (MDP). The applications of RL includes autonomous vehicle, robotics, navigation, etc.

An agent in *Q-learning* follows the Bellman equation to update the cumulative rewards for its corresponding state and action.

$$Q = \sum_{t=1}^{N} \gamma^{t-1} r_t \left(s_t, a_t\right) \tag{4}$$

Here Q represents the quality of state, S. t represents the time steps and ranges from $\{1, 2, 3, \ldots, N\}$, γ represents the discount factor and ranges from 0 to 1, r represents the reward at state, s due to the execution of action, a. Value of γ close to 0 ensures the learning process is focused on short term reward, and the value of γ close to 1 helps the learning agent focus on long term rewards.

$$a_t^* = \arg\max_{a_j \in A_t} Q\left(s_t, a_j\right) \tag{5}$$

Here, a_t represents the optimal action at time step t which maximizes the Q value of that state. The generic Bellman equation can be expressed as follows:

$$\begin{aligned} Q\left(s_t, a_t\right) \leftarrow & (1-\alpha) Q\left(s_t, a_t\right) \\ & + \alpha \left\{ r_{t+1}\left(s_t, a_t\right) + \gamma \max_a Q\left(s_{t+1}, a\right) \right\} \end{aligned} \tag{6}$$

where, α represents the learning rate. The equation above can simplified as follows:

$$Q(s_t, a_t) \leftarrow r(s_t, a_t) + \gamma \max_a Q(s_{t+1}, a) \tag{7}$$

There is an exploration-exploitation trade-off, which helps to take random action during exploration and greedy action during exploitation. The value of ϵ balances this exploration-exploitation trade-off during the learning process.

IV. Proposed Q-Learning Framework

The proposed *Q-learning* framework for resilience planning enables strategic decision making on behalf of a power grid agent. The components of this RL based framework includes state ($s \in S$), action ($a \in A$), reward (R). The states are generally represented as follows: $S = \{s_1, s_2, s_3, \ldots, s_3\}$. Here, the *Q-learning* framework is developed as a one-shot process. And the set of actions A can be generally represented as follows: $A = \{a_1, a_2, a_3, \ldots, a_N\}$. The states are the power system states with the selection of line hardening actions. The set of actions includes the tentative set of lines. In this study, we randomly select 10 lines, and the learning agent aims to select an optimal line for hardening actions to make the grid resilient battling the HILP events. The learning framework is developed in Python, and the power system is conducting operations in MATLAB. The learning framework sends an action execution command with the action index to the MATLAB. The MATLAB executes the action, calculates the $CVaR_\alpha$, and returns the value to the Python. The learning framework receives the returned value and considers it as the reward. For the experiment's ease, we consider 10 random lines as candidate lines to be selected for hardening during the HILP events. For the event, we consider wind-related events. The set of candidate lines are: $A = \{1, 2, 3, \ldots, 10\}$. The initial state of the system is considered as the steady-state condition of the system. The agent changes action strategies by selecting lines from the action set, A. To conduct the learning process, the agent explores and exploits about 500 episodes. The learning aims to find the optimal line that maximizes the value of $CVaR_\alpha$ so that the learned line can be used for line hardening.

As we mentioned earlier, the *Q-learning* agent utilizes exploration-exploitation trade-off to balance between random action and greedy/optimal action selection. The agent selects an action with a ϵ-greedy policy. In ϵ-greedy policy, an agent selects the optimal action that gives the maximum reward

Algorithm 1: Q-learning for optimal line hardening

Input : Fragility curve, Set of lines ($\mathcal{A}$), Budget
Initialize the action counter, probability, and Q-table;
if *Number of executed action* $\leq$ *Budget* **then**
  for *Maximum number of episodes* **do**
    if *Prob* $> \epsilon$ **then**
      a $\leftarrow$ Rand ($\mathcal{A}$);
    else
      a $\leftarrow$ Greedy ($\mathcal{A}$);
    end
    Update the action counter and probability, $\mathcal{P}$;
    Execute the action;
    Calculate $CVaR_\alpha$ and assign the reward;
    Update *Q-table*;
  end
else
  Terminate the simulation;
end
Output: Optimal selection of line for hardening.

TABLE I
CSFR AND THEIR ASSOCIATED PROBABILITIES.

Actions	Cum. Sum of Future Rewards	Probability
1	**−9180.56**	**0.7200**
2	−9380.99	0.0824
3	−9357.11	0.0861
4	−9212.15	0.0904
5	−9310.78	0.0904
6	−9385.20	0.0819
7	−9403.18	0.0802
8	−9410.01	0.0789
9	−9309.34	0.0859
10	−9471.44	0.0859

with $1-\epsilon$ probability and a random action with ϵ probability. Algorithm 1 represents the workflow of the adoption of *Q-learning* while selecting the optimal action for line hardening. The traditional *Q-learning* solves optimal decision-making by maximizing the cumulative sum of future rewards (CSFR). However, we want to minimize the expected long term discounted reward. So the reward function is formulated as follows:

$$R = -CVaR_\alpha = -(1-\alpha)^{-1} \int_{U(I)\geq VaR_\alpha} U(I)\ p(I)\ dI \quad (8)$$

The action selection probability, $P(s,a)$ can be expressed as below:

$$\mathrm{P}(s,a) \leftarrow \frac{C(s,a)}{\sum_{a\in A} C(s,a)} \quad (9)$$

where $C(s,a)$ represents the frequency of state s visited by the agent while taking action $a \in A$. The probability is measured based on the frequency of that specific state-action pairs visited.

V. SIMULATION RESULTS AND DISCUSSIONS

The modified IEEE 123-bus test feeder [8] is selected as the test case in this study. We randomly select 10 lines as the learning agent's target set to choose as the optimal action for line hardening. Note that the set of lines to harden in practice would be selected based on some metrics. A commonly used way of selecting lines is to upgrade previously damaged facilities or perform targeted hardening based on experiences. The selection can be made using the fragility curves for different intensities of weather event [16] or using the assets information and ranking the lines based on some predefined vulnerability index or topological metrics [17].

During the simulation, we have selected the lines one by one. This sequential decision-making approach is adopted because the budget to harden the overall system is not always available at once. However, if the budget is available for multiple lines, the selected candidate can be more than one. On the other hand, as the budget becomes available, the approach can be re-run, and the additional candidate lines can be selected from the set. As described in the previous section, the learning agent aims to select the optimal action minimizing the CSFR with the help of *Q-learning*. In the first case study, we assume the agent has a budget of only one line hardening out of 10 lines. We conduct the learning to select the optimal action for line hardening. The fragility curve

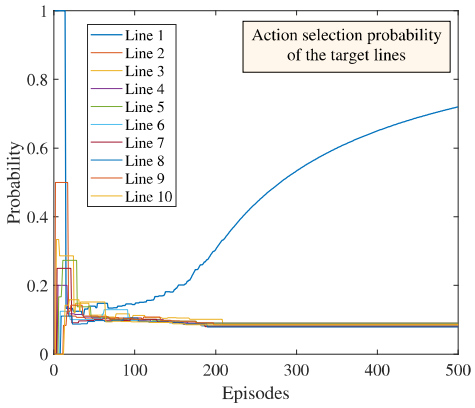

Fig. 3. Probability update for optimal line selection for hardening.

represents the failure probability of the distribution system components associated with the wind speed.

Q-learning based simulation for optimal line hardening is presented in Table I. Table I represents the *Q-Table* of the learning process. The first column of the table represents the possible actions, i.e., the set of candidate lines to harden. The second column represents the resulting converged values of the cumulative future reward function ($-CVaR_\alpha$) for $\alpha = 0.95$. The last column represents the action selection probability of the lines as an optimal decision. From the table, we can observe that the selecting line results in a maximized reward for the RL agent and thus a maximum reduction in $CVaR_\alpha$ for loss-of-load probability. Note that the $CVaR_\alpha$ for the base case, without any line upgrades for α = 0.95, was 1147 MWhr. Next, we further elaborate on the working of the proposed algorithm. Figure 3 represents the probability update for all the lines to be selected as an optimal decision by the agent. The initial oscillations are due to random action selection during the exploration stage of the learning. For the proposed algorithm, the value of ϵ is gradually reduced from 0.9 to a low value. The initial value of ϵ = 0.9 represents that, initially, there is a 90% probability of exploration (random action selection) and 10% probability of exploitation (greedy action selection).

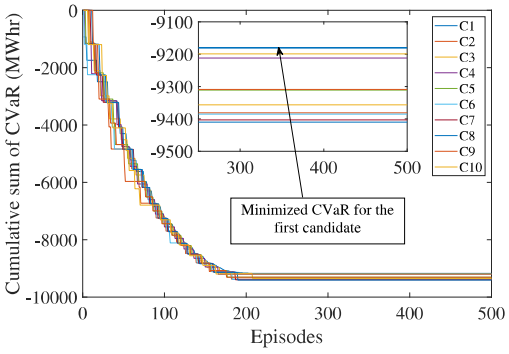

Fig. 4. CSFR for first candidate selection.

Fig. 4 shows that after enough exploration and exploitation, the agent converges to the maximum cumulative future rewards (cumulative $CVaR_\alpha$) by selecting optimal action for line hardening. Since, from Table I, it is found that line 1 accumulates the maximum value of the CSFR, line 1 is the optimal choice as the first candidate for line hardening.

TABLE II
CSFR AND THEIR ASSOCIATED PROBABILITIES.

Actions	Cum. Sum of Future Rewards	Probability
$[1, 2]$	-7321.28	0.1818
$[1, 3]$	-7428.17	0.0845
$[1, 4]$	-7382.40	0.0867
$[1, 5]$	-7372.99	0.1055
$[1, 6]$	-7462.17	0.0756
$[1, 7]$	$\mathbf{-7320.00}$	$\mathbf{0.6660}$
$[1, 8]$	-7432.40	0.0821
$[1, 9]$	-7421.90	0.0800
$[1, 10]$	-7381.16	0.0837

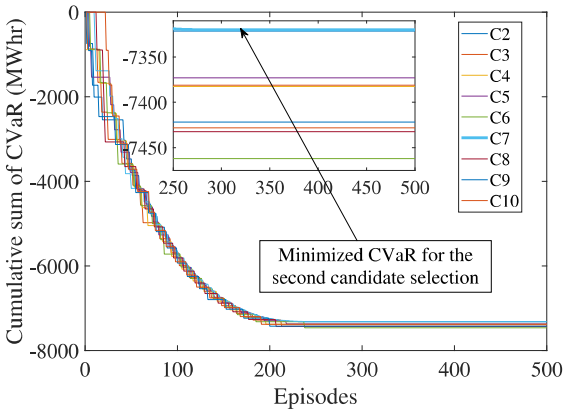

Fig. 5. CSFR for second candidate selection.

Now, let us assume the budget increases from one-line hardening to two-line hardening. Since line 1 has been selected from the tentative lines, this line will be removed from the set of potential lines for the next selection. Following the same process, we continue looking for the second optimal action for line hardening out of 9 lines (excluding line 1). Table II shows that line 7 has the maximum CSFR while selecting the optimal line for hardening as the second candidate, and Figure 5 shows the convergence of the learning curve for choosing the second line for hardening.

Finally, in Table III, we compare the $CVaR_\alpha$ before and after hardening the optimal lines selected by the *Q-learning* agent. The value of $CVaR_\alpha$ reduces after the hardening of the first and second candidate lines compared to the base case (no line hardening). Similarly, if the budget allows, the learning agent can select an additional line for hardening, further increasing the system's resilience.

TABLE III
$CVaR_\alpha$ CALCULATION BEFORE AND AFTER LINE HARDENING

	Base Case	First Candidate	Second Candidate
$CVaR_\alpha$ (MWhr)	1147	918	732

VI. CONCLUSION

We propose a *Q-learning* approach to plan resilient power distribution systems by identifying the optimal line hardening actions for a given budget constraint. The proposed approach optimizes the allocated budget to optimally harden the line to improve a risk-based resilience metric characterizing HILP events' impacts. Here, we harden the distribution system against high wind speed events by undergrounding the lines. The proposed *Q-learning*-based algorithm can effectively search for optimal hardening actions by minimizing the conditional value at risk ($CVaR$) for the loss of load. Further, a sequential upgrade/hardening plan is generated upon relaxing the budget constraint. Finally, we demonstrate the proposed approach's applicability using numerical experiments on the IEEE 123-bus test system. It is shown that the proposed *Q-learning* algorithm converges and results in optimal risk-averse hardening decisions to increases the distribution grid resilience.

REFERENCES

[1] President's Council of Economic Advisers and the U.S. Department of Energy's Office of Electricity and Energy Reliability, "Economic benefits of increasing electric grid resilience to weather outages," Aug 2013. [Online] Available: http://energy.gov/downloads/economic-benefits-increasing-electric-grid-resilience-weather-outages.

[2] E. National Academies of Sciences and Medicine, *Enhancing the Resilience of the Nation039;s Electricity System*. Washington, DC: The National Academies Press, 2017.

[3] W. Yuan, J. Wang, F. Qiu, C. Chen, C. Kang, and B. Zeng, "Robust optimization-based resilient distribution network planning against natural disasters," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2817–2826, 2016.

[4] E. Yamangil, R. Bent, and S. Backhaus, "Resilient upgrade of electrical distribution grids," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[5] S. Ma, B. Chen, and Z. Wang, "Resilience enhancement strategy for distribution systems under extreme weather events," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1442–1451, 2018.

[6] A. Dubey and S. Santoso, "Availability-based distribution circuit design for shipboard power system," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1599–1608, 2015.

[7] C. A. MacKenzie and C. W. Zobel, "Allocating resources to enhance resilience, with application to superstorm sandy and an electric utility," *Risk Analysis*, vol. 36, no. 4, pp. 847–862, 2016.

[8] S. Poudel, A. Dubey, and A. Bose, "Risk-based probabilistic quantification of power distribution system operational resilience," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3506–3517, 2020.

[9] A. Chaudhuri, M. Norton, and B. Kramer, "Risk-based design optimization via probability of failure, conditional value-at-risk, and buffered probability of failure," in *AIAA Scitech 2020 Forum*, p. 2130, 2020.

[10] T & D World, *Transmission System Operator Uses AI to Reduce Costs*, (accessed Nov. 3, 2020). Available at: https://www.tdworld.com/test-and-measurement/article/21119870/transmission-system-operator-uses-artificial-intelligence-to-reduce-costs.

[11] R. Eskandarpour and A. Khodaei, "Leveraging accuracy-uncertainty tradeoff in svm to achieve highly accurate outage predictions," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1139–1141, 2018.

[12] S. Soltan, A. Loh, and G. Zussman, "A learning-based method for generating synthetic power grids," *IEEE Systems Journal*, vol. 13, no. 1, pp. 625–634, 2019.

[13] R. Nateghi, "Multi-dimensional infrastructure resilience modeling: An application to hurricane-prone electric power distribution systems," *IEEE Access*, vol. 6, pp. 13478–13489, 2018.

[14] S. Paul and F. Ding, "Identification of worst impact zones for power grids during extreme weather events using q-learning," in *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, 2020.

[15] S. Paul, Z. Ni, and C. Mu, "A learning-based solution for an adversarial repeated game in cyber-physical power systems," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2019.

[16] M. Panteli, P. Mancarella, D. N. Trakas, E. Kyriakides, and N. D. Hatziargyriou, "Metrics and quantification of operational and infrastructure resilience in power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4732–4742, 2017.

[17] S. Chanda and A. K. Srivastava, "Defining and enabling resiliency of electric distribution systems with multiple microgrids," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2859–2868, 2016.