A Hybrid Reinforcement Learning-MPC Approach for Distribution System Critical Load Restoration

Preprint

Abinet Tesfaye Eseye, Xiangyu Zhang, Bernard Knueven, Matthew Reynolds, Weijia Liu, and Wesley Jones

National Renewable Energy Laboratory

Presented at the 2022 IEEE Power & Energy Society General Meeting
Denver, Colorado
July 17-21, 2022

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-2C00-81440
February 2022

A Hybrid Reinforcement Learning-MPC Approach for Distribution System Critical Load Restoration

Preprint

Abinet Tesfaye Eseye, Xiangyu Zhang, Bernard Knueven, Matthew Reynolds, Weijia Liu, and Wesley Jones

National Renewable Energy Laboratory

Suggested Citation
Tesfaye Eseye, Abinet, Xiangyu Zhang, Bernard Knueven, Matthew Reynolds, Weijia Liu, and Wesley Jones. 2022. *A Hybrid Reinforcement Learning-MPC Approach for Distribution System Critical Load Restoration: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-2C00-81440. https://www.nrel.gov/docs/fy22osti/81440.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-2C00-81440
February 2022

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Electricity Delivery and Energy Reliability. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

A Hybrid Reinforcement Learning-MPC Approach for Distribution System Critical Load Restoration

Abinet Tesfaye Eseye, Xiangyu Zhang, Bernard Knueven, Matthew Reynolds, Weijia Liu, and Wesley Jones

***Abstract*— This paper proposes a hybrid control approach for distribution system critical load restoration, combining deep reinforcement learning (RL) and model predictive control (MPC) aiming at maximizing total restored load following an extreme event. RL determines a policy for quantifying operating reserve requirements, thereby hedging against uncertainty, while MPC models grid operations incorporating RL policy actions (i.e., reserve requirements), renewable (wind and solar) power predictions, and load demand forecasts. We formulate the reserve requirement determination problem as a sequential decision-making problem based on the Markov Decision Process (MDP) and design an RL learning environment based on the OpenAI Gym framework and MPC simulation. The RL agent reward and MPC objective function aim to maximize and monotonically increase total restored load and minimize load shedding and renewable power curtailment. The RL algorithm is trained offline using a historical forecast of renewable generation and load demand. The method is tested using a modified IEEE 13-bus distribution test feeder containing wind turbine, photovoltaic, microturbine, and battery. Case studies demonstrated that the proposed method outperforms other policies with static operating reserves.**

***Index Terms*— Distribution system, model predictive control, operating reserve, reinforcement learning, restoration.**

I. INTRODUCTION

As modern power systems host an increasing share of new generation from variable and uncertain renewable energy (VURE) sources like wind and solar, the need for adequate operating reserves has increased to hedge against the variability and uncertainty of these sources. Operating reserves are also vital for improved load restoration after extreme event-caused outages. They help power systems effectively schedule and dispatch generation and storage assets to restore prioritized critical loads and increase the total load restoration over a look-ahead control horizon (outage duration). When these reserves are not sufficient to keep the frequency stability of the power system, load shedding may become necessary.

The authors are with the U.S. National Renewable Energy Laboratory, Golden, CO 80401, USA. AbinetTesfaye.Eseye@nrel.gov.

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Electricity (OE) Advanced Grid Modeling (AGM) Program. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

This research was performed using computational resources sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory.

Critical load restoration is one of the power system automation technologies. Augmenting it with intelligent and robust operating reserve management technique can enable the system to restore more critical loads after an extreme event-caused disaster. References [1] and [2] propose classical optimization formulation and model predictive control (MPC) approaches for critical load restoration in distribution systems under extreme events. The latter considers the co-optimization of power and reserve products of distributed energy resources (DERs). However, although this work demonstrated the impact of reserves in improving the load restoration process, the reserve requirements were determined by a fixed ad hoc rule, so they may not be suitable when the system operating condition changes.

Reinforcement learning (RL)-based control methods have been applied to a number of control problems such as games, robotics, transportations, etc. In the area of power (distribution) systems, RL has been applied to voltage control [3], DER scheduling [4] and load restoration [5], [6]. However, these studies do not consider the notion of reserves from the DERs. Moreover, due to the RL difficulties of handling sophisticated operational constraints and the unavailability of open-source or commercial platforms for designing RL simulation environments to learn power system problems, some of these studies do not capture the complete underlying operating condition in distribution systems.

This paper proposes a hybrid control approach for determining operating reserve requirements by combing deep RL and MPC. In the proposed approach, the RL determines a policy to quantify the reserve requirements and the MPC models the grid operation incorporating the RL policy actions (operating reserves) and forecasts of renewable power and load demand. We formulated the reserve requirement determination problem as a sequential decision-making problem based on the Markov Decision Process (MDP) and designed an RL learning environment based on the OpenAI Gym framework and MPC simulation. The RL agent reward and MPC objective function aim to maximize and monotonically increase total restored load and minimize load shedding and renewable power curtailment. The method is applied on a modified IEEE 13-bus distribution test system containing wind, solar, microturbine, and energy storage battery, compared with other operating reserve determination methods.

The remaining sections of the paper are organized as follows. The model of the distribution grid operation is presented in Section II. Section III introduces the proposed hybrid controller framework. Experimental findings and conclusions are given in Sections IV and V, respectively.

II. Distribution Grid Operation Model

A. Optimal Resource Scheduling and Load Restoration

The critical (prioritized) load restoration process is modeled as a constrained optimization problem whose objective function and constraints are presented below.

1) *Objective Function*:

The objective function aims to maximize the total restored load and minimize load shedding and renewable power curtailment. It is formulated as follows.

$$\max_{DV} \sum_{i\in N}\sum_{t\in T}\omega_i L_{i,t}\Delta t - \psi\sum_{i\in N}\sum_{t\in T}\omega_i LS_{i,t}\Delta t - \sum_{i\in N}\sum_{t\in T}(\alpha P_{i,t}^{wt,curt} + \beta P_{i,t}^{pv,curt})\Delta t \quad (1)$$

where $DV = \{L_{i,t}, LS_{i,t}, P_{i,t}^{wt,cut}, Q_{i,t}^{wt}, P_{i,t}^{pv,cut}, Q_{i,t}^{pv}, P_{i,t}^{g}, Q_{i,t}^{g}, P_{i,t}^{es}, Q_{i,t}^{es}, SOC_{i,t}^{es}, a_{i,t}^{es}, b_{i,t}^{es}\}$ is the set of decision variables of the optimal scheduling and restoration (OSR) problem. $L_{i,t}$ is the restored load at node i and time t and given by $L_{i,t} = P_{i,t}^{l} + Q_{i,t}^{l}$ where $P_{i,t}^{l}$ and $Q_{i,t}^{l}$ are the active and reactive load components; $LS_{i,t}$ is the load shedding and given by $LS_{i,t} = \max((P_{i,t-1}^{l} - P_{i,t}^{l}), 0) + \max((Q_{i,t-1}^{l} - Q_{i,t}^{l}), 0)$; $P_{i,t}^{wt,curt}$ and $P_{i,t}^{pv,curt}$ are the wind and solar power curtailments; $Q_{i,t}^{wt}$ and $Q_{i,t}^{pv}$ are the reactive power injection/absorption by the wind and solar power electronic converters; $P_{i,t}^{g}$ and $Q_{i,t}^{g}$ are the active and reactive power output of the dispatchable generator; $P_{i,t}^{es}$ and $Q_{i,t}^{es}$ are the net active (charging/discharging) and reactive (injection/absorption) power of the energy storage; $SOC_{i,t}^{es}$ is the state of charge (SOC) of the storage; $a_{i,t}^{es}$ and $b_{i,t}^{es}$ are binary variables indicating the charging and discharging status of the storage; ω_i is the priority weight (criticalness level) of the load; ψ is the load shedding penalty; α and β are penalties for the wind and photovoltaic (PV) power curtailments; Δt is the length of the control periods; T is the control horizon; and N is the number of electrical nodes.

2) *Constraints*:

I) *Restored Loads Feasible Range:*

$$\begin{aligned} &0 \le P_{i,t}^{l} \le P_{i,t}^{l,dem}\ ;\ 0 \le Q_{i,t}^{l} \le Q_{i,t}^{l,dem} \\ &Q_{i,t}^{l} = \left(\frac{Q_{i,t}^{l,dem}}{P_{i,t}^{l,dem}}\right) P_{i,t}^{l} \end{aligned} \quad (2)$$

where $P_{i,t}^{l,dem}$ and $Q_{i,t}^{l,dem}$ are the forecasted active and reactive load power demands before the extreme event.

II) *Dispatchable DERs Output Power Feasible Range:*

$$\begin{aligned} &0 \le P_{i,t}^{g} \le P_{i}^{g,max}\ ;\ 0 \le Q_{i,t}^{g} \le Q_{i}^{g,max} \\ &Q_{i}^{g,max} = \sqrt{(S_{i}^{g,max})^2 - (P_{i}^{g,max})^2} \end{aligned} \quad (3)$$

where $P_{i}^{g,max}$, $Q_{i}^{g,max}$ and $S_{i}^{g,max}$ are, respectively, the maximum active, reactive, and apparent power of the dispatchable generator.

III) *Fuel Usage of Fuel-Fired Dispatchable DERs:*

$$\sum_{t\in T} P_{i,t}^{g}\Delta t \le E_{i,max}^{g,p}\ ;\ \sum_{t\in T} Q_{i,t}^{g}\Delta t \le E_{i,max}^{g,q} \quad (4)$$

where $E_{i,max}^{g,p}$ and $E_{i,max}^{g,q}$ are, respectively, the maximum allowable active and reactive energy productions associated with the fuel consumption of fuel-fired generators.

IV) *Storage Power Limits and Complementary Operation:*

$$\begin{aligned} &0 \le P_{i,t}^{es,c} \le a_{i,t}P_{i}^{es,max}\ ;\ 0 \le P_{i,t}^{es,d} \le b_{i,t}P_{i}^{es,max} \\ &0 \le Q_{i,t}^{es,a} \le a_{i,t}Q_{i}^{es,max}\ ;\ 0 \le Q_{i,t}^{es,i} \le b_{i,t}Q_{i}^{es,max} \\ &a_{i,t} + b_{i,t} \le 1 \\ &P_{i,t}^{es} = P_{i,t}^{es,d} - P_{i,t}^{es,c}\ ;\ Q_{i,t}^{es} = Q_{i,t}^{es,i} - Q_{i,t}^{es,a} \end{aligned} \quad (5)$$

where $P_{i,t}^{es,c}$, $P_{i,t}^{es,d}$, $Q_{i,t}^{es,a}$ and $Q_{i,t}^{es,i}$ are, respectively, the charging power, discharging power, absorbed and injected reactive power; $P_{i}^{es,max}$ and $Q_{i}^{es,max}$ are the maximum active and reactive power capacity of the storages; and $a_{i,t}$ and $b_{i,t}$ equal to 1 when the storages are charging (or absorbing reactive power) and discharging (or injecting reactive power), respectively, and 0 otherwise.

V) *Storage SOC Feasible Range and Dynamics:*

$$SOC_{i,min}^{es} \le SOC_{i,t}^{es} \le SOC_{i,max}^{es} \quad (6)$$

$$SOC_{i,t}^{es} = SOC_{i,t-1}^{es} + \left(\frac{\eta_i^{es,c}P_{i,t}^{es,c}}{C_i^{es}} - \frac{P_{i,t}^{es,d}}{\eta_i^{es,d}C_i^{es}}\right)\Delta t \quad (7)$$

where $SOC_{i,min}^{es}$ and $SOC_{i,max}^{es}$ are the minimum and maximum SOCs of the storage; $\eta_i^{es,c}$ and $\eta_i^{es,d}$ are the charging and discharging efficiencies of the storage; and C_i^{es} is the rated holding capacity of the storage.

VI) *Renewable Power Curtailment Feasible Range:*

$$0 \le P_{i,t}^{wt,cut} \le P_{i,t}^{wt}\ ;\ 0 \le P_{i,t}^{pv,cut} \le P_{i,t}^{pv} \quad (8)$$

where $P_{i,t}^{wt}$ and $P_{i,t}^{pv}$ are the wind and PV power forecasts.

VII) *Power Electronic Converters Operation:*

$$-\sqrt{(S_{der})^2 - (P_{der}^{max})^2} \le Q_{der,t} \le \sqrt{(S_{der})^2 - (P_{der}^{max})^2} \quad (9)$$

where S_{der} is the rated apparent power of the DER (wind, PV or storage) converter; P_{der}^{max} is the rated active power of the DER; and $Q_{der,t}$ is the injected or absorbed reactive power by the DER converter.

VIII) *Power Flow Constraints:*

$$P_{ij,t} = P_{j,t}^{l} - \left(P_{j,t}^{g} + P_{j,t}^{wt} - P_{j,t}^{wt,cut} + P_{j,t}^{pv} - P_{j,t}^{pv,cut} - P_{j,t}^{es}\right) + \sum_{k\in N} A_{jk}P_{jk,t}\ ;\ \forall t \in T\ ,\ \forall j \in N\ ,\ i = r(j) \quad (10)$$

$$Q_{ij,t} = Q_{j,t}^{l} - \left(Q_{j,t}^{g} + Q_{j,t}^{wt} + Q_{j,t}^{pv} + Q_{j,t}^{es}\right) + \sum_{k\in N} A_{jk}Q_{jk,t}\ ;\ \forall t \in T\ ,\ \forall j \in N\ ,\ i = r(j) \quad (11)$$

$$V_{j,t} = V_{i,t} - 2(r_{ij}P_{ij,t} + x_{ij}Q_{ij,t}); \forall t \in T, \forall j \in N, i = r(j) \quad (12)$$

where $P_{ij,t}$ and $Q_{ij,t}$ are the active and reactive power flows from node i to node j; r_{ij} and x_{ij} are the resistance and reactance of the distribution line connecting nodes i and j;

and A is an adjacency matrix that expresses the distribution network topology and its element A_{ij} is set to 1 if node i is the parent of node j and 0 otherwise; and $V_{i,t}$ and $V_{j,t}$ are the squares of the voltages at nodes i and j.

IX) *Nodal Voltage Bounds:*

$$v_{min}^2 \leq V_{i,t} \leq v_{max}^2 \tag{13}$$

where v_{min} and v_{max} are, respectively, the allowable minimum and maximum magnitudes of the nodal voltages.

B. Operating Reserve Requirements

In this paper, the distribution system operating reserve service is provided by two dispatchable DERs, a microturbine (MT) generator, and an energy storage (ES) battery. The formulations governing the relationship between the power outputs of these DERs and the operating reserve requirements are given below.

$$SOC_{i,t_f}^{es} \geq \mu_{t_i}^{es} \; ; \; E_{i,t_f}^{g} \geq \nu_{t_i}^{g} \tag{14}$$

where SOC_{i,t_f}^{es} and E_{i,t_f}^{g} are, respectively, the end-of-control-horizon ES SOC and MT fuel levels; $\mu_{t_i}^{es}$ and $\nu_{t_i}^{g}$ are, respectively, the end-of-horizon ES SOC and MT fuel reserve requirements; and t_i and t_f are, respectively, the initial and final time periods of the control horizon $[t_i, t_f]$.

We propose a deep RL approach to determine the reserve requirements μ_{t_i} and ν_{t_i} to maximize and monotonically increase an aggregate load restoration. We incorporate these reserve requirement equations into the OSR problem formulated in $(1)-(13)$, and run the MPC simulation for the outage (restoration) period.

III. Proposed Hybrid Controller

Here we present the proposed RL-MPC hybrid control approach, including the basics of RL and MPC, while making practical connections to the OSR problem and operating reserve requirements. To solve $(1)-(14)$ using deep RL, first we need to reformulate the OSR problem as an MDP.

A. Markov Decision Process (MDP)

At each time step t, an agent obtains a state s_t from the state space S, and chooses an action a_t from the action space A based on its policy $\pi(a_t|s_t)$. Consequently, the system transitions to the next state $s_{t+1} \sim \rho(s_{t+1}|s_t, a_t)$ and the agent receives an immediate scalar reward r_t. The elements of the MDP are described as follows.

1) *State:*

The state vector $s_t \in S$ is used to describe the distribution grid and OSR problem at time t and it is defined below:

$$\begin{aligned} s_t = \Big[& \mathbf{P}_{[t,t+T]}^{wt}, \mathbf{P}_{[t,t+T]}^{pv}, \mathbf{P}_t^{g}, \mathbf{P}_t^{es}, \mathbf{P}_t^{rl}, \mathbf{Q}_t^{rl}, \\ & \mathbf{E}_t^{g}, \mathbf{E}_{t_f}^{g}, \mathbf{SOC}_t^{es}, \mathbf{SOC}_{t_f}^{es}, t \Big] \end{aligned} \tag{15}$$

where $\mathbf{P}_{[t,t+T]}^{wt}$ and $\mathbf{P}_{[t,t+T]}^{pv}$ are, respectively, the vectors of wind and solar power forecasts at time t for future T look-ahead periods; $\mathbf{P}_t^{g}$ and $\mathbf{P}_t^{es}$ are, respectively, the real-time vectors of MT output power and ES net power; $\mathbf{P}_t^{rl}$ and $\mathbf{Q}_t^{rl}$ are, respectively, the real-time vectors of restored load active and reactive components; $\mathbf{E}_t^{g}$ and $\mathbf{SOC}_t^{es}$ are, respectively, the real-time vectors of MT remaining fuel and ES SOC; $\mathbf{E}_{t_f}^{g}$ and $\mathbf{SOC}_{t_f}^{es}$ are, respectively, the vectors of end-of-horizon MT remaining fuel and ES SOC; and t is the control step.

2) *Action:*

The action vector $a_t \in A$ at each time step t consists of the required reserve requirements from the dispatchable DERs, and it is defined as follows:

$$a_t = [\boldsymbol{\mu}_i, \boldsymbol{\nu}_i] \tag{16}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\nu}_i$ are, respectively, the vector of end-of-horizon MT fuel and ES SOC reserve requirements.

3) *Reward Function:*

The reward function is formulated based on the OSR objective function defined in (1). It is the load restoration reward but also penalizes load shedding and renewable curtailment, and is expressed as follows:

$$\begin{aligned} r = & \sum_{i \in N} \omega_i L_i - \psi \sum_{i \in N} \omega_i LS_i \\ & - \sum_{i \in N} (\alpha P_i^{wt,cut} + \beta P_i^{pv,cut}) \end{aligned} \tag{17}$$

To obtain the value of L_i, LS_i, $P_i^{wt,cut}$ and $P_i^{pv,cut}$ the OSR problem $(1)-(14)$ is solved by setting the values of $\mu_{t_i}^{es}$ and $\nu_{t_i}^{g}$ equal to the action a_t and the system state to what is given in s_t. Thus, computing $r(s_t, a_t)$ requires solving $(1)-(14)$ for the control horizon T based on the MPC simulation, where only the first step decisions are applied and the rest are discarded. The hybrid RL-MPC controller targets to search for the optimal operating action a_t at the present state s_t such that the expected aggregate reward of all the future states is maximized, as formulated below:

$$\max_{a_t \in A} \mathbb{E}\left[\sum_t \gamma^t [r(s_t, a_t)] \right] \tag{18}$$

where $\gamma \in [0, 1]$ is a discount factor expressing the importance of the present reward with respect to the future rewards. We solve the expected cumulative reward (18) using the popular RL algorithm known as Proximal Policy Optimization (PPO) proposed in [7].

B. Simulation Environment

To train the RL agent, it is necessary to develop an environment E that mimics the distribution grid operation. This environment consists of three modules: (i) Distribution system model, which describes the steady-date operation of the system (OSR problem); (ii) MPC simulator, which generates DER schedules and load restoration; (iii) Data generator, which provides the RL agent training data such as scenarios of wind and solar power forecasts, load demand, and grid topology information. The RL agent and environment interaction is illustrated in Fig. 1.

The controller can be deployed for practical real-time distribution grid operation after the training is completed. Moreover, the controller can still pursue its learning and therefore adapt to new operating conditions and potential misrepresentations of the environment by calibrating its

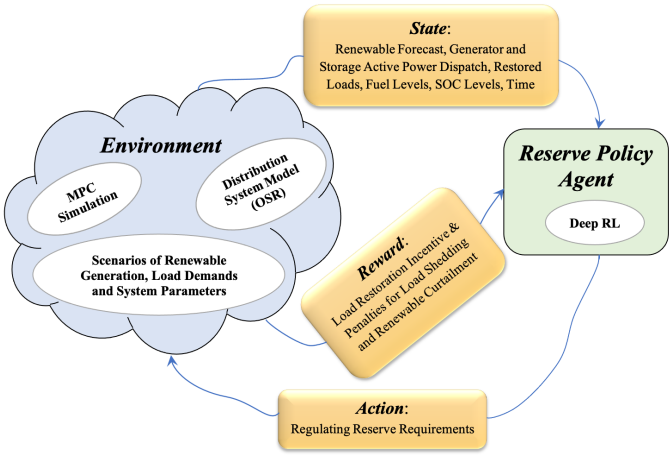

Fig. 1. Hybrid RL-MPC controller learning framework.

parameters via online feedback. This can also be used to adjust the model to time-varying future conditions of the system.

IV. Case Study and Simulation Results

The proposed RL-MPC hybrid controller is applied on a modified IEEE 13-bus test feeder, where loads are considered three-phase balanced, and four three-phase DERs comprising wind (WT), PV, MT, and ES are integrated to the feeder, as shown in Fig. 2 [2].

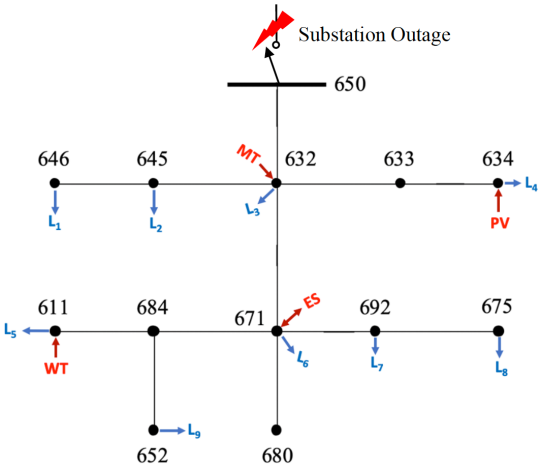

Fig. 2. Test system: modified IEEE 13-bus feeder with DERs.

The system is hit by an extreme event near the substation bus ('650') and automatically switch to an islanded operation. The DERs are then coordinated and managed through the proposed controller to restore and supply the system loads ($L_1 - L_9$) for 6 hours of outage period with 5 minutes of control step. The wind and solar power forecast data are generated by the method referred in [2]. The system data is provided in Table I.

The proposed RL-MPC simulation environment is developed according to OpenAI Gym [8] and the RL algorithm is implemented using RLlib Ray [9] library. The Xpressmp [10] solver is used to solve the MPC OSR problem. The whole simulation is performed using high performance computing (HPC) system in parallel across multiple CPU cores. The agent achieved convergence at a reward value of $40 * 10^3$ after about 202 episodes. The performance of the devised RL-MPC-based dynamic operating reserve policy (RP1) is investigated and compared against four MPC-based fixed operating reserve policies given in Table II:

TABLE I
System Parameters

Parameter	Unit	Value
N, ψ		13, 100
$T, \Delta t$	Hour	6, 5/60
t		[1, 2, ..., 72]
ω	\$/kWh	[1.0, 1.0, 0.9, 0.85, 0.8, 0.65, 0.45, 0.4, 0.3]
α, β	\$/kWh	0.2, 0.2
$P_t^{l,dem}, \forall t$	kW	[115, 85, 49.75, 200, 85, 199.75, 85, 324, 64]
$Q_t^{l,dem}, \forall t$	kW	[66, 52, 29, 115, 40, 109, 45, 141, 43]
$P^{g,max}, S^g$	kW, kVA	400, 500
$E_{max}^{g,p}, E_{max}^{g,q}$	kWh, kvarh	1000, 750
$E_{initial}^{g,p}, E_{initial}^{g,q}$	kWh, kvarh	1000, 750
$P^{es,max}, S^{es}$	kW, kVA	200, 250
$P^{wt,max}, S^{wt}$	kW, kVA	400, 500
$P^{pv,max}, S^{pv}$	kW, kVA	300, 375
$C^{es}, SOC_{initial}^{es}$	kWh	800, 160
$SOC_{min}^{es}, SOC_{max}^{es}$	%	20, 100
$\eta^{es,c}, \eta^{es,d}$	%	95, 90
v_{min}, v_{max}	pu	0.95, 1.05

TABLE II
MPC Fixed Reserve Policies

Reserve Policy (RP)	MT Reserve[kWh]	ES Reserve [kWh]
RP2	200	0.0
RP3	0.0	240
RP4	500	0.0
RP5	500	240

The performance comparison is presented based on 20 scenarios representing different outage beginning times and renewable (wind and solar) profiles. Statistics describing performances are visualized in Fig. 3.

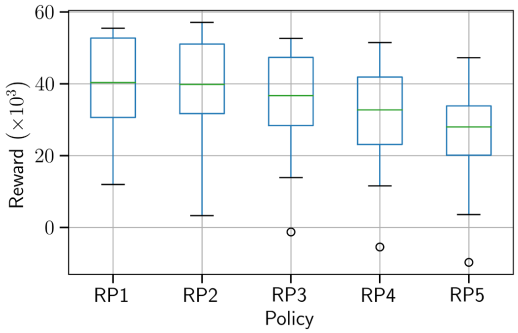

Fig. 3. Box and whiskers plot comparing rewards for scenarios 1-20 using policies RP1-5.

As shown in Fig. 3, RP1 not only produces the highest me-

dian reward, but also the highest worst-case reward over the 20 scenarios. Furthermore, the sample variance of rewards is lowest for RP1 at 168.7, compared to the next lowest sample variance of 195.1 for RP3. These observations suggest that the RL-MPC dynamic reserve policy is more robust than the fixed reserve policies.

For further performance comparison, we derive virtual best and worst reserve policies based on the value of the reward at each scenario by the policies (RP1 - RP5). A virtual best policy is a policy that contains the maximum values of the rewards by the policies for each scenario, while a virtual worst policy is a policy that contains the minimum values of the rewards. The performance comparison of RP1 against the virtual best and worst policies is presented in Fig. 4, which shows that rewards for using the RL-MPC dynamic reserve policy are equal to the best virtual policy reward values from RP1-RP5 in most of the scenarios (80%). Even when the proposed policy reward values are lower than the best policy, on average they are only 9.6% worse than the virtual best policy values. This validates the benefit of proposed learning-based dynamic reserve policy determination to improve load restoration, and hence resilience, for the distribution system.

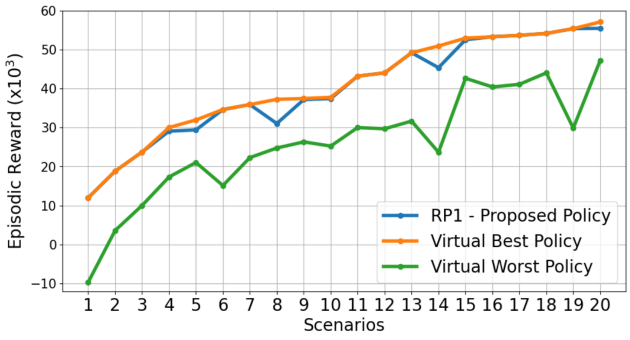

Fig. 4. Performance comparison of the proposed reserve policy and virtual policies over scenarios.

Figure 5 depicts the power dispatch and aggregate restored load, along with the dynamic reserves, by the proposed RL-MPC hybrid controller for a single scenario (Scenario 17 in Fig. 4) where the sub-station power was down at 12 a.m. As shown in Fig. 5, the devised controller manages the DERs interactively to monotonically increase the total restored load. The proposed controller is able to restore more loads for this scenario, compared to the benchmarks, as it can dynamically determine the required reserves based on the present and future operating condition of the system as opposed to the one-size-fits-all fixed reserves, which fail (or not effective) when the operating condition of the system changes.

V. CONCLUSIONS

This paper devised and demonstrated a hybrid RL-MPC controller to improve the load restoration process in a distribution system under an extreme event. The hybrid controller has benefited from the strengths of both controllers where the RL learned MPC parameters and the MPC modeled sophisticated operational constraints of the distribution system, which could be hard for the RL to handle alone. We applied the devised method to the IEEE 13-bus distribution test feeder with wind, PV, microturbine, and battery. The proposed RL-MPC dynamic reserve policy outperformed the other four MPC-based fixed reserve policies, in most tested scenarios, with respect to reward criteria for restoring more loads and shedding less. The present findings will continue in the next phase of our research, in which the hybrid RL-MPC controller will be explored to solve more relevant power system problems that are too difficult to be solved effectively by RL or MPC alone.

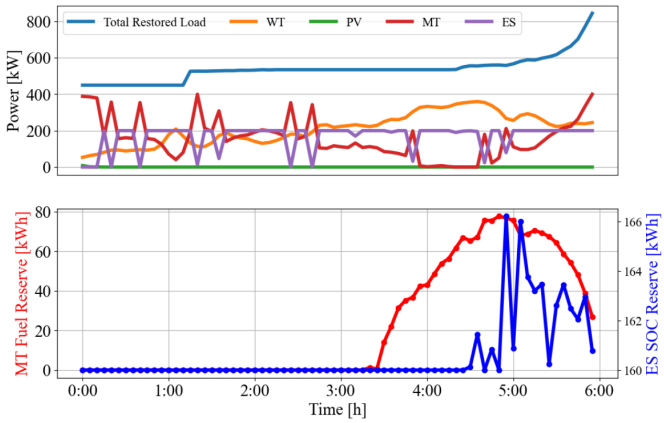

Fig. 5. Power dispatch and aggregate restored load.

REFERENCES

[1] Z. Wang, C. Shen, Y. Xu, F. Liu, X. Wu, and C. C. Liu, "Risk-limiting load restoration for resilience enhancement with intermittent energy resources," IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 2507–2522, May 2019.

[2] A.T. Eseye, B. Knueven, X. Zhang, M. Reynolds, and W. Jones, "Resilient Operation of Power Distribution Systems Using MPC-based Critical Service Restoration," 2021 IEEE GreenTech, April 2021.

[3] J.F. Toubeau , B. B. Zad , M. Hupez, Z. D. Grève, and F. Vallée, "Deep Reinforcement Learning-Based Voltage Control to Deal with Model Uncertainties in Distribution Networks," Energies, 13(15), Aug. 2020.

[4] M. M. Hosseini and M. Parvania, "Resilient Operation of Distribution Grids Using Deep Reinforcement Learning," IEEE Trans. on Ind. Info., 2021 (Accepted for publication).

[5] X. Zhang, A.T. Eseye, B. Knueven, W. Jones, "Restoring Distribution System Under Renewable Uncertainty Using Reinforcement Learning," in 2020 IEEE SmartGridComm, Nov. 2020.

[6] X. Zhang, A.T Eseye, M. Reynolds, B. Knueven and W. Jones, "Restoring Critical Loads in Resilient Distribution Systems Using a Curriculum Learned Controller," in IEEE PES GM. July 2021.

[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv: 1707.06347v2, Aug. 2017.

[8] OpenAI Gym. Available at: https://gym.openai.com/. Accessed on Oct 20, 2021.

[9] RLlib Ray. Available at : https://docs.ray.io/en/latest/rllib.html. Accessed on Oct 20, 2021.

[10] Xpressmp Solver. Available at: https://www.fico.com/en/products/fico-xpress-solver. Accessed on Oct 10, 2021.