Restoring Distribution System Under Renewable Uncertainty Using Reinforcement Learning

Xiangyu Zhang (xiangyu.zhang@nrel.gov), Abinet Tesfaye Eseye, Ben Knueven and Wesley Jones
Computational Science Center, NREL

IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids
November 11-13, 2020

Summary I

1. Why RL?

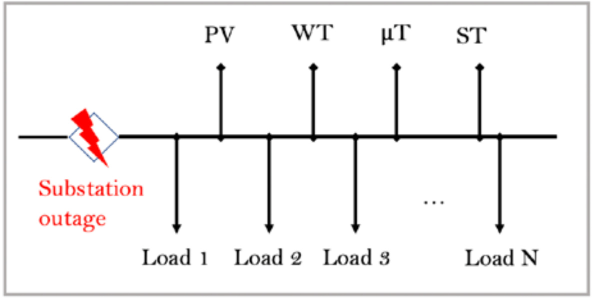
- Nature of the problems for power system control: *optimal*, *nonlinear*, *stochastic*, and *fast*.
  - ❑ RL is for optimal sequential decision making, which maximize an expected cumulative reward (certain objective).
  - ❑ Compared with optimization approaches (e.g., stochastic programing), RL can handle system nonlinearity and stochasticity more easily.
  - ❑ RL optimal control policy is trained offline through simulation* and it only requires policy evaluation during real-time control, which provides great action readiness (suitable for scenarios that needs fast response).
- Optimal control problems suitable for RL: sequential optimal control with *strong temporal dependency*. Not good for snapshot optimization such as solving OPF.

2. What's the problem?

- Distribution system load restoration with renewable and dispatchable DERs

$$\underset{\mathbf{P}^t, p_t^{\mu}, p_t^{\theta}, p_t^{\alpha} (t \in \mathcal{T})}{\text{maximize}} \quad \mathcal{C} = \sum_{t \in \mathcal{T}} \mathbf{H}^T \mathbf{P}_t - \epsilon \sum_{t \in \mathcal{T}} \mathbf{H}^T [\mathbf{P}_{t-1} - \mathbf{P}_t]^+ - \beta \mathbf{1}^T \mathbf{P}^{\alpha} \qquad (1)$$

  - ❑ Sequential optimal control with temporal dependency and uncertainty.
  - ❑ Support the grid operator to take a sequence of fast, correct and coordinated actions for fast system recovery.

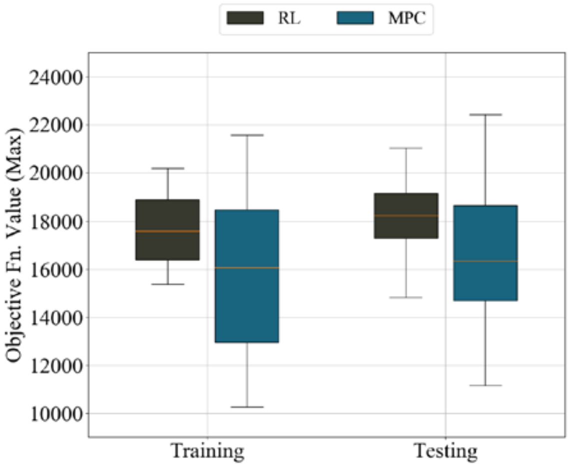


- For a proof of concept, in this paper, a single-bus distribution system is considered, from energy adequacy perspective.

Summary II

3. How did it work?

- Compared the performance of an RL controller and a deterministic MPC, given imperfect forecast of future renewable generation.
    - ❑ RL controller learned from experience that the imperfect renewable forecast cannot be fully trusted; the policy learned shows a more stable performance when compared with the MPC's performance.
    - ❑ RL controller's performance does not deteriorate under unseen testing scenarios. (Good for real-world applications)



4. What's next?

- Increase problem complexities:
    - ❑ System complexity: e.g., consider power flow (e.g., using OpenDSS for simulation) and other operational constraints.
    - ❑ Baseline complexity: e.g., use the state-of-the-art stochastic programming based controller as baseline.
    - ❑ Uncertainty complexity: e.g., consider uncertainty in upstream substation restoration time.
- Explore the techniques for selecting a proper training data (scenario diversity ←→ true distribution)

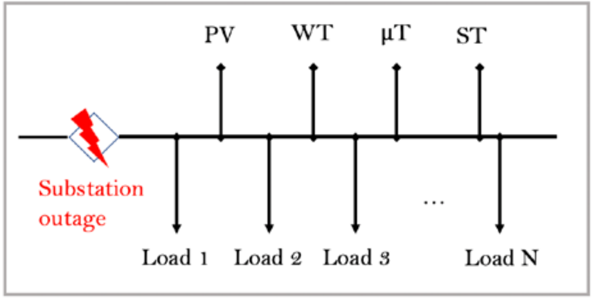
Background

Many power system control problems require *optimal* and sometimes *fast-response* control to the *nonlinear* system with consideration of *uncertainty*. Oftentimes, these problems are solved by optimization-based approaches (e.g., model predictive control or stochastic programming). This study investigates an alternative controller based on reinforcement learning (RL).

	RL	Optimization-based
Real-time computation	Light, involving only policy evaluation. Optimal control can be generated instantly.	Heavy, requires solving optimization problem within control intervals.
Handling nonlinearity	Able to learn control policies for non-linear systems.	Better if system is linearized.
Handling stochasticity	Able to use raw historical data, learn distribution implicitly.	Treat as deterministic problem/scenario-based stochastic programming/…
Training requirement	Require, can be computationally intensive, offline.	Not required.

In this study, the advantages of using RL as an alternative for solving a power system optimal control problem will be explored. A distribution system load restoration problem is presented here.

Load Restoration Problem V1 (Single Bus Case)



Objective:

During the upstream substation downtime, by leveraging renewable generations and properly controlling dispatchable generators accordingly, the control objective is to *maximize the prioritized load pick-up* with the consideration of the *penalty for repeated load shedding and renewable curtailment.*

Assumptions:

1. The network/power flow constraints are not considered.

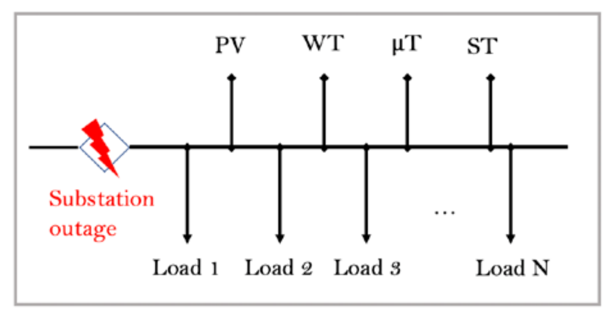
2. Fuel for micro-turbine and the initial storage for battery are limited, and these two dispatchable resources alone are not sufficient to restore the system.

3. Only imperfect forecast for renewable generation is available.

4. The demand from each critical load ($p^i, \forall\, i \in \mathcal{L}$) is assumed to be time-invariant over the control horizon, and it can be partially restored.

5. The length of the restoration control horizon/upstream repair time is deterministic and known in advance (e.g., 6 hours.) and the control interval is five minutes.

Load Restoration Problem V1 (Single Bus Case)



Notations:

PV: ρ
Wind: ω
Storage: θ
Micro Turbine: μ
Curtailment: α

$\mathcal{R} = \{\rho, \omega\}$ (PV, Wind)

$\mathcal{G} = \{\theta, \mu\}$ (Storage, Micro Turbine)

$$[x^1, x^2, \dots, x^N]^+ = [(x^1)^+, (x^2)^+, \dots, (x^N)^+]$$

$$(x^i)^+ = \max(0, x^i)$$

$$\mathbf{H} = [\eta^1, \eta^2, \dots, \eta^N]^T \in \mathbb{R}^N$$

$$\mathbf{P}_t = [p_t^1, p_t^2, \dots, p_t^N]^T \in \mathbb{R}^N$$

Reward term for prioritized load restoration.

Penalty term for shedding previously restored load.

Penalty term for renewable curtailment.

$$\underset{\mathbf{P}^t, p_t^\mu, p_t^\theta, p_t^\alpha (t\in\mathcal{T})}{\text{maximize}} \quad \mathcal{C} = \underbrace{\sum_{t\in\mathcal{T}} \mathbf{H}^T \mathbf{P}_t}_{\text{Reward term for prioritized load restoration}} - \underbrace{\epsilon \sum_{t\in\mathcal{T}} \mathbf{H}^T [\mathbf{P}_{t-1} - \mathbf{P}_t]^+}_{\text{Penalty term for shedding previously restored load}} - \underbrace{\beta \mathbf{1}^T \mathbf{P}^\alpha}_{\text{Penalty term for renewable curtailment}} \qquad (1)$$

subject to

$\sum_{g\in\mathcal{G}} p_t^g + \sum_{r\in\mathcal{R}} \hat{p}_t^r - p_t^\alpha = \mathbf{1}^T \mathbf{P}_t$	(2)	Power balance
$\mathbf{0} \le \mathbf{P}_t \le \mathbf{P}$	(3)	Load feasibility
$\underline{p^\mu} \le p_t^\mu \le \overline{p^\mu}$	(4)	Micro-turbine power feasibility
$\sum_{t\in\mathcal{T}} p_t^\mu \cdot \tau \le E^\mu$	(5)	Micro-turbine energy availability
$-p^{\theta,ch} \le p_t^\theta \le p^{\theta,dis}$	(6)	Storage power feasibility
$S_{t+1}^\theta = \mathcal{F}(S_t^\theta, p_t^\theta)$	(7)	Storage state equation
$\underline{S^\theta} \le S_t^\theta \le \overline{S^\theta}$	(8)	Storage SOC feasibility
$S_0^\theta = s_0$	(9)	Storage SOC initial value

RL Formulation and Learning

❖ RL Markov decision process (MDP) formulation

State Space:

$$\mathbf{s}_t = \Big[\underbrace{\widehat{\boldsymbol{P}}^{\rho}_t, \widehat{\boldsymbol{P}}^{\omega}_t}_{\text{Renewable prediction (imperfect) for the next hour}}, \underbrace{S^{\theta}_t, \dot{E}^{\mu}_t, \frac{\mathbf{1}^{\mathrm{T}}\mathbf{P}_t}{\mathbf{1}^{\mathrm{T}}\mathbf{P}}}_{\text{Current system status}}, \underbrace{t}_{\text{Current control step}}\Big] \in \mathbb{R}^{24}$$

Renewable prediction (imperfect) for the next hour

Current system status

Current control step

Action Space:

$$\mathbf{a}_t = [p^{\mu}_t, p^{\theta}_t, p^{\alpha}_t] \in \mathbb{R}^3$$

Reward structure:

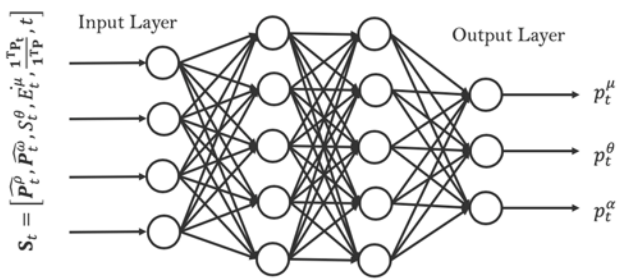
$$\mathrm{r}_t = \mathbf{H}^T\mathbf{P}_t - \mathbf{H}^T[\mathbf{P}_{t-1} - \mathbf{P}_t]^+ - \beta p^{\alpha}_t$$

❖ RL Objective

Train an optimal control policy that maximize the control rewards:

$$\pi^* = \underset{\pi_{\boldsymbol{w}} \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_{\pi}[\sum_{t\in\mathcal{T}} \gamma^t \mathrm{r}_t]$$

where $\pi_{\boldsymbol{w}}$ is a parameterized control policy ($\mathbf{a}_t = \pi_{\boldsymbol{w}}(\boldsymbol{s}_t)$), which is usually instantiated by a neural network in deep RL.



A policy gradient algorithm uses gradient ascent for policy training:

$$\boldsymbol{w}^{t+1} = \boldsymbol{w}^t + \alpha\widehat{\nabla}_{\boldsymbol{w}} J(\boldsymbol{w}) = \boldsymbol{w}^t + \alpha\widehat{\nabla}\mathbb{E}_{\pi_{\boldsymbol{w}}}[\sum_{t\in\mathcal{T}} \gamma^t \mathrm{r}_t]$$

Controller Evaluation

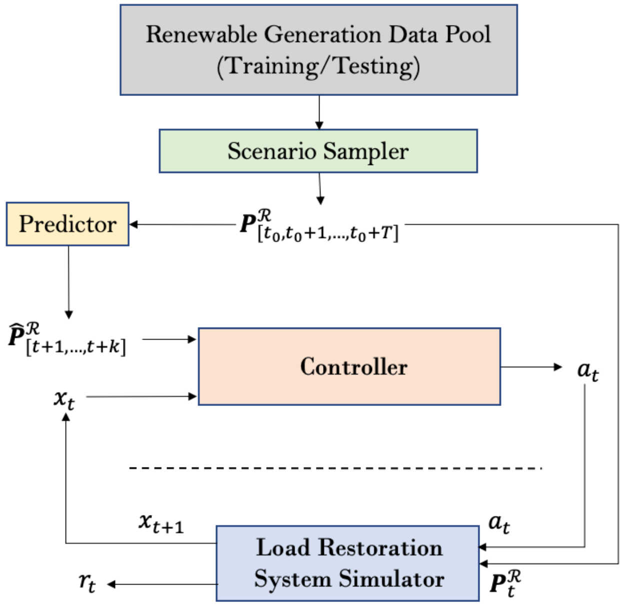
❖ Baseline controller

Baseline: a deterministic model predictive control (MPC) based controller based on mixed integer linear programming (MILP) is used.

$$\underset{\mathbf{P}^t, p_t^\mu, p_t^\theta, p_t^\alpha (t \in \mathcal{T})}{\text{maximize}} \quad \mathcal{C} = \sum_{t \in \mathcal{T}} \mathbf{H}^T \mathbf{P}_t - \epsilon \sum_{t \in \mathcal{T}} \mathbf{H}^T [\mathbf{P}_{t-1} - \mathbf{P}_t]^+ - \beta \mathbf{1}^T \mathbf{P}^\alpha \qquad (1)$$

- Only imperfect forecast for renewable generations are available and are updated every time step.
- Optimization problems are solved at each control interval with reduce planning horizon.

❖ Experimenting framework



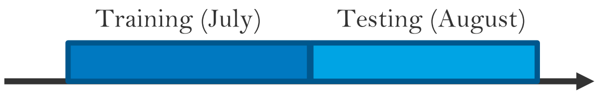
Case Study (Experiment Settings)

❖ Parameters

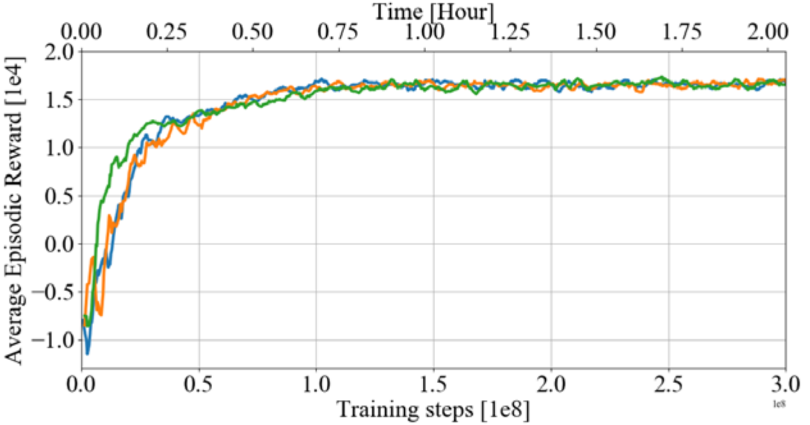
TABLE I
PARAMETERS USED FOR CASE STUDY

Var	Value	Var	Value
H	[1.0, 1.0, 0.9, 0.85, 0.8, 0.65, 0.45, 0.4, 0.3, 0.3]	$[\underline{p}^{\mu}, \overline{p}^{\mu}]$	[0, 300]
P	[33, 34, 8.5, 85, 60, 60, 58, 115, 64, 85]	E^{μ}	1000
$\mathcal{T}$	[1, 2, ..., 72]	$(p^{\theta,dis}, p^{\theta,ch})$	[200, 200]
$\mathcal{L}$	[1, 2, ..., 10]	s_0	720
PV	[0, 300]	$(\underline{S}^{\theta}, \overline{S}^{\theta})$	[160, 800]
Wind	[0, 150]	τ	1/12
ϵ	100	β	0.2

❖ Exogenous data splitting



❖ Learning curve

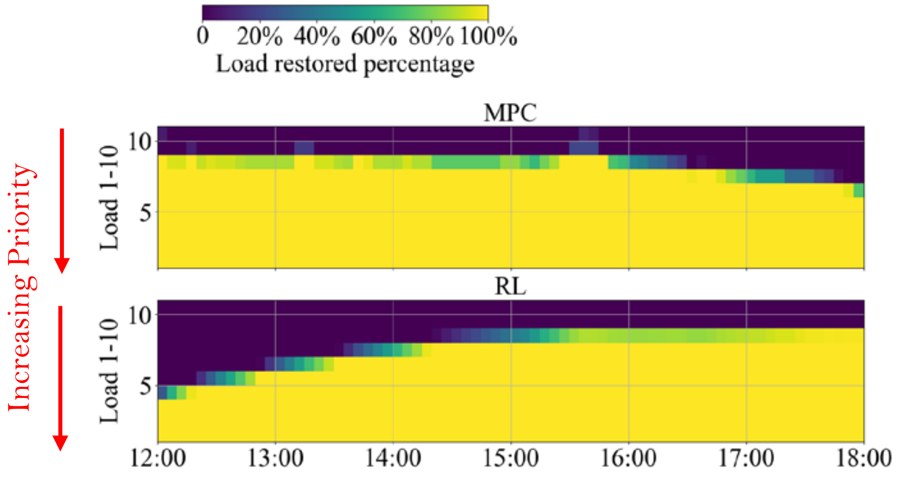


- Using an RL algorithm based on evolution strategies*.
- Training was conducted on ten computing nodes of the NREL high-performance computing system.
- Policy converged in one hour, around140 million steps of experience. (Yes, ES-RL is known to have low sample efficiency, but the wall-time training efficiency is okay.)

* Salimans, T., Ho, J., Chen, X., Sidor, S., & Sutskever, I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*.

Case Study
(Results for the single-bus system)

❖ Single Scenario



- Load 1 (Highest priority) ⟶ Load 10 (Lowest priority)
- RL controller starts with less load restored, but gradually pick-up more in a monotonic manner.
- MPC starts high, but Load 8 receives intermittent service and finally fully shed together with Load 7.

❖ Multiple Scenarios

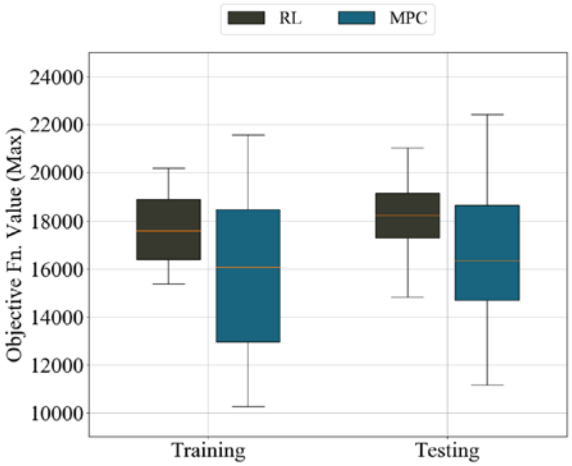
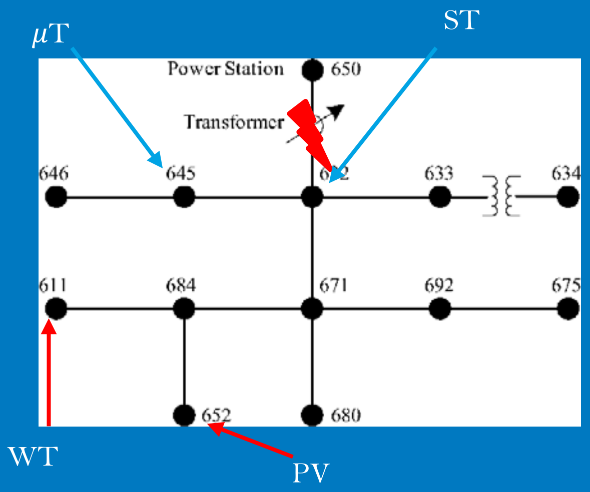


TABLE II
AVERAGE OBJECTIVE FN. VALUES OVER 25 SCENARIOS IN EACH CASE

Controller	Scenarios	
	Training	Testing
MPC	15811.77	16647.85
RLC	17633.46	18044.50

- In general, RLC can achieve a higher reward when compared with the objective function value of the deterministic MPC. RLC shows a relatively more stable behavior (less variance over different scenarios).

Future work (Network Constrained Case)



Goal:

Using DERs for distribution system load restoration after a substation outage.

Specific Objective:

During the upstream substation downtime, maximizing the prioritized load pick-up with the consideration of the penalty for repeated load shedding and renewable curtailment.

$$\underset{\mathbf{P}^t, p_t^{\mu}, p_t^{\theta}, p_t^{\alpha} (t \in \mathcal{T})}{\text{maximize}} \quad (1) \; + \text{other operational penalty (e.g., voltage deviation, line limit)}$$

$$\text{subject to} \quad (2) - (9)$$

+ power flow network constraints, both active and reactive power.

Thank you!
Questions?

www.nrel.gov

NREL/PR-2C00-78103

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 Improving Distribution System Resiliency via Deep Reinforcement Learning Project funded by the U.S. Department of Energy Office of Electricity Advance Grid Modeling 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.

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.

NREL Transforming ENERGY