Deep Reinforcement Learning with Local Interpretability for Transparent Microgrid Resilience Energy Management

M. Nejati, M. de Oliveira and F. Guéniat
florimond.gueniat@bcu.ac.uk

3rd July 2025

Birmingham City Univ., College of Engineering

Why Resilience is Fundamental for Microgrids

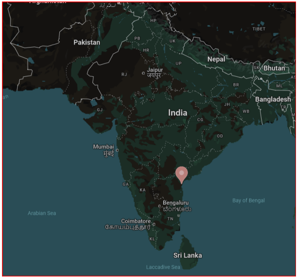
- **Isolated communities** : Limited/no grid access → Self-reliance essential
- **Natural disasters** : Cyclones damage transmission infrastructure
- **Energy-water-food nexus** : Power outages cascade to vital services

Understanding decision made by a system is necessary for adoption.

Context

A real micro grid scenario[1], during an High Impact scenario (cyclone Layla).

- Actual loads from a village Ongole (India)
- Actual wind and solar data from the location of Ongole, including the Layla cyclone



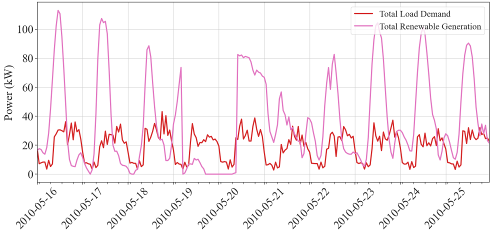
The design and sizing done using HOMER pro.

1. Amiri, Mohammad Hossein Nejati, et al. *Sustainable Cities and Society*, 2025

Context

A real micro grid scenario [1], during an High Impact scenario (cyclone Layla).

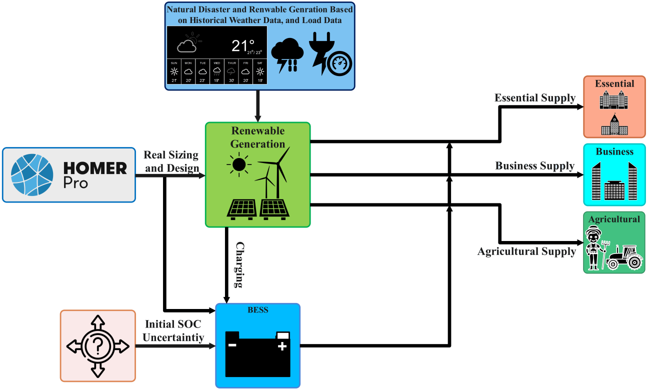
- Actual loads from a village Ongole (India)
- Actual wind and solar data from the location of Ongole, including the Layla cyclone



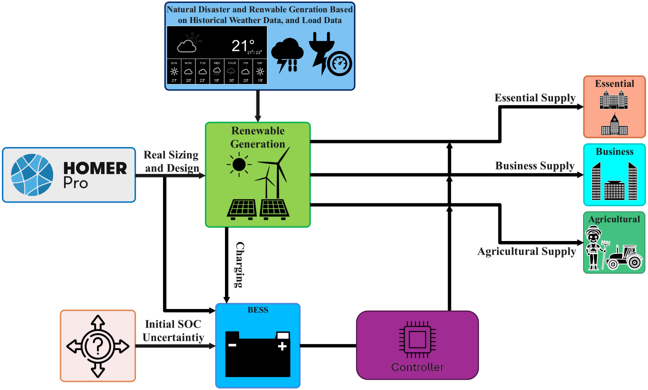
The design and sizing done using HOMER pro.

1. Amiri, Mohammad Hossein Nejati, et al. *Sustainable Cities and Society*, 2025

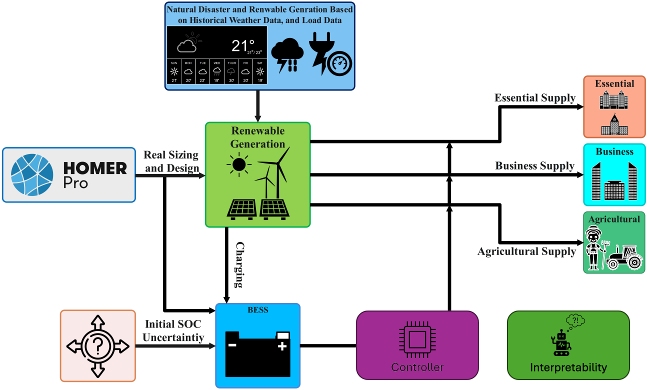
Three Stages design



Three Stages design

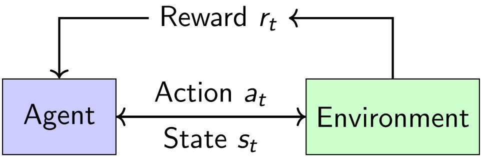


Three Stages design



Deep Reinforcement Learning Fundamentals

- **Agent** : Controller making decisions
- **Environment** : Smart grid (solar, wind, battery, loads)
- **State** s_t : Current observations (generation, demand, SOC)
- **Action** a_t : Control decisions (Battery dis/charge/idle, load priority)
- **Reward** r_t : Performance metric (loads satisfaction + resiliency)



DRL for Isolated Smart Grid Management

System Components

- Renewable generation (solar/wind)
- Battery storage
- Critical loads

DRL Objectives

- Maximize load satisfaction
- Increase resiliency

Objectices means that the rewards needs to capture :

- Load priorities
- Battery dis/charge
- Resiliency

Reward shaping : Reliability

The main aim is to minimize the negative power imbalances P_{sh}.
No imbalance $\rightsquigarrow$ Rewards maximized :

$$r_t = \left(1 - \frac{7P_{\text{sh},1,t} + 2P_{\text{sh},2,t} + 1P_{\text{sh},3,t}}{7L_{1,t} + 2L_{2,t} + 1L_{3,t}}\right).$$

It is normalized by loads L.

Reward shaping : Resilience

The resilience[2] is captured at the end :

$$r_{\text{final}} = r_{\text{episode}} + \text{RI}.$$

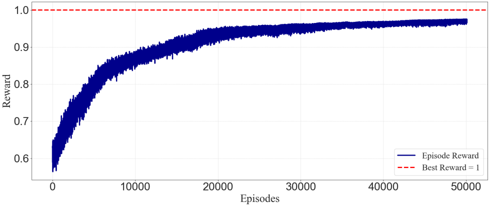
RI captures the resiliency index $RI \in [0, 1]$:

$$\text{RI} = 1 - \frac{7\sum_{t=1}^{T} P_{\text{sh},1,t} + 2\sum_{t=1}^{T} P_{\text{sh},2,t} + 1\sum_{t=1}^{T} P_{\text{sh},3,t}}{7\sum_{t=1}^{T} L_{1,t} + 2\sum_{t=1}^{T} L_{2,t} + 1\sum_{t=1}^{T} L_{3,t}}.$$

2. MHN Amiri, F Guéniat *Sustainable Cities and Society*, 2024

Training

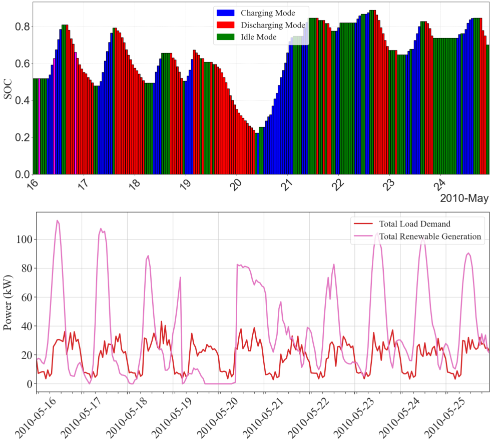
Training done using PPO algorithm and StableBaselines3 [3] library.



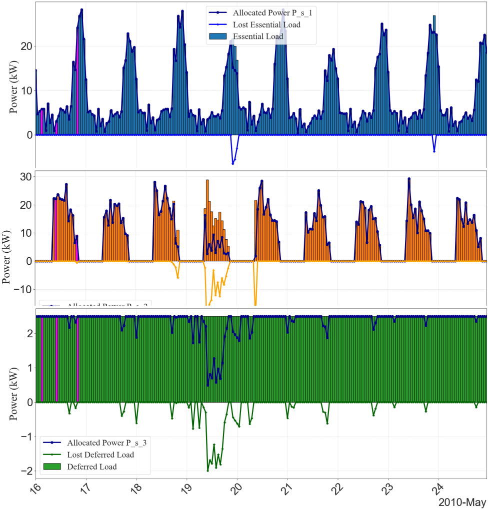
3. https://stable-baselines3.readthedocs.io/

Action : Evolution of SoC

SoC acts as expected.



Reliability and Resiliency



Resiliency Index after the cyclone :

$$RI > 0.97$$

Battery life : 15.11 years

The Need for Interpretability

The Problem

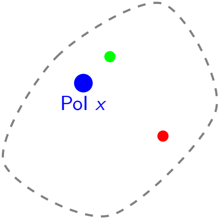
- Complex agents are uninterpretable
- We need to understand why decisions are made
- Especially critical for stakeholders

Which features specific to this decision were most important ?

LIME Solution

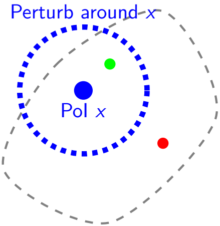
- Explains specific decisions
- Works for any model
- Provides intuitive local explanations

How LIME Works



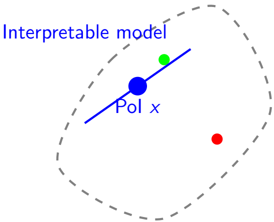
1. Chose Point of Interest

How LIME Works



1. Chose Point of Interest
2. Perturb around that point

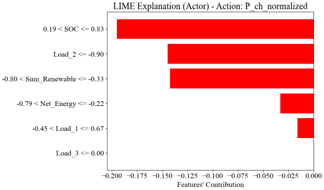
How LIME Works



1. Chose Point of Interest
2. Perturb around that point
3. Construct a linear model.

   Weights allow to understand how features affect output

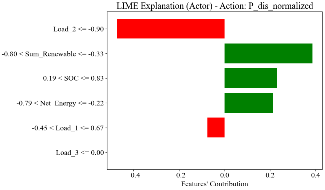
Explaining IDLE mode at h4



For the charging action : all features discourage charging.

$\rightsquigarrow$ no encouragement for charging.

Explaining IDLE mode at h4

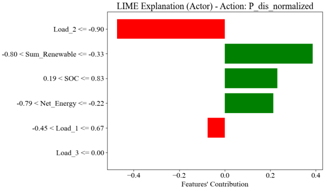


For the discharging action :

- SOC, renewable generation, and net energy encourage discharging
- L_2 strongly discourage discharging

$\rightsquigarrow$ no clear encouragement for discharging.

Explaining IDLE mode at h4



For the discharging action :

- SOC, renewable generation, and net energy encourage discharging
- L_2 strongly discourage discharging

$\rightsquigarrow$ no clear encouragement for discharging.

This balance of influences leads the actor to select the idle action.

Key outputs

- New DRL management system inherently resilient to unpredicted events
- Performances enforced with a new reward shaping including Resiliency Index $\Leftrightarrow RI \approx 0.97$
- Fully interpretable framework : Each decisions can be justified

Cumulative Rewards

Cumulative Reward (Return)

Agent's *long-term* objective :

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

- $\gamma \in [0, 1)$: Discount factor (prioritizes near-term rewards)
- r_t : Immediate reward at time t

DRL Optimization Goal

Learn policy $\pi(a|s)$ (aka "rules") that maximizes expected return :

$$\max_{\pi} \mathbb{E}\left[R_t \mid s_t, \pi\right]$$

Implemented via :

- Deep neural networks approximating :
  - Policy $\pi(a|s)$ (*actor*)
  - Value function $V^{\pi}(s) = \mathbb{E}_{\pi}[R_t \mid s_t = s]$ (*critic*)
- Parameter updates via gradient ascent :

$$\nabla_\theta J(\theta) \propto \mathbb{E}\left[\nabla_\theta \log \pi_\theta(a|s) \cdot R_t\right]$$

DRL agents *learn by experience* to maximize cumulative future rewards through function approximation.

LIME Mathematics and Advantages

Mathematical Formulation

$$\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

s.t. g simple enough

- f : Complex model
- g : Interpretable model
- π_x : Proximity measure
- $\Omega(g)$: Model complexity

Key Advantages

- Model-agnostic
- Human-interpretable
- Local fidelity
- Sample instability
- No global perspective