Energy Management of a Residential Heating System through Deep Reinforcement Learning

Silvio Brandi [1], Davide Coraci [1], Davide Borello [1], Alfonso Capozzoli [1]

www.baeda.polito.it

[1] BAEDA Laboratory, DENERG, Politecnico di Torino, Italy

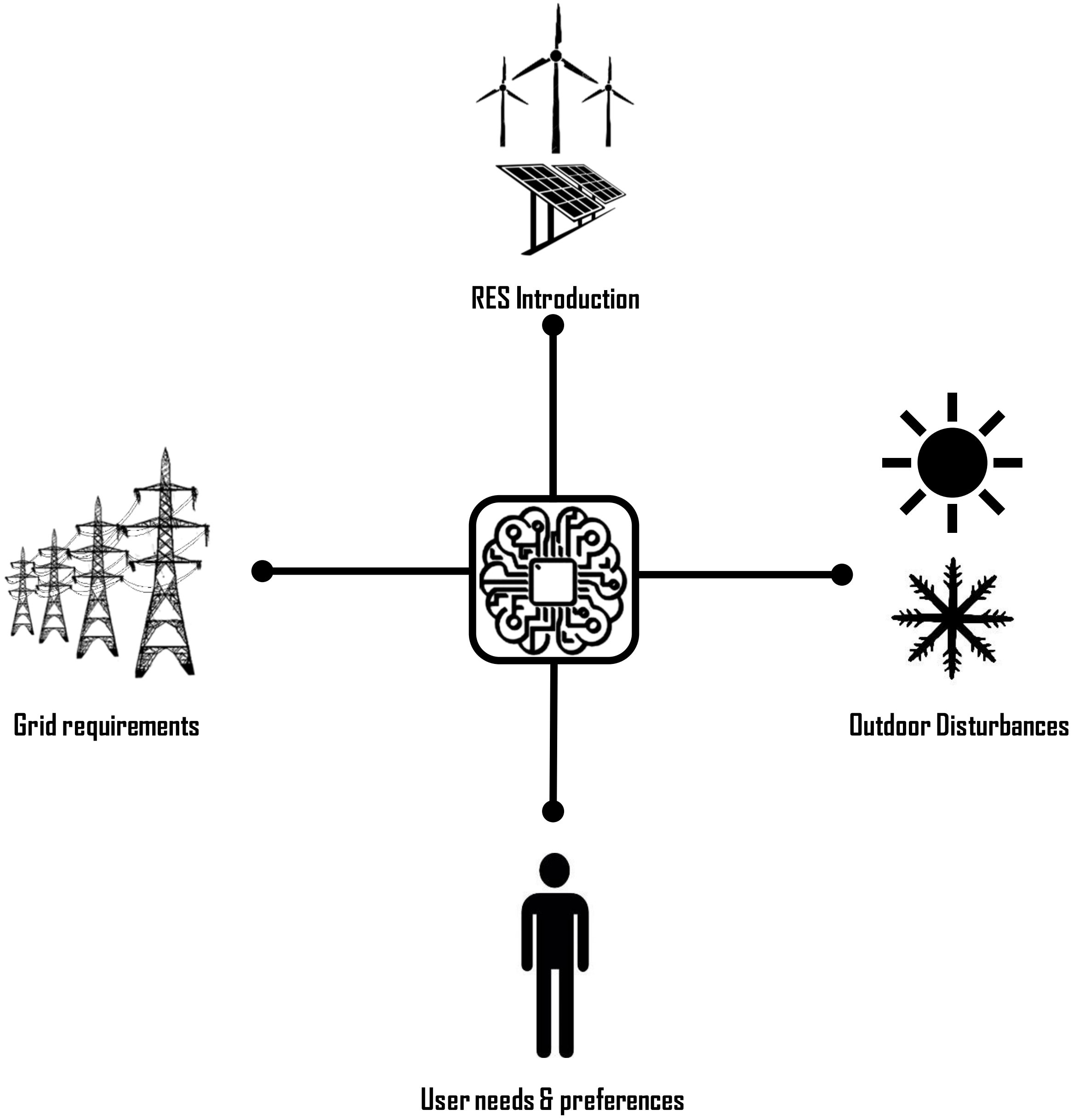
SEB-21, 15-17 September 2021

Problem Statement

1. The increasing penetration of HVAC systems, the introduction of RES and storage has changed the framework of building energy managing → Energy Flexibility

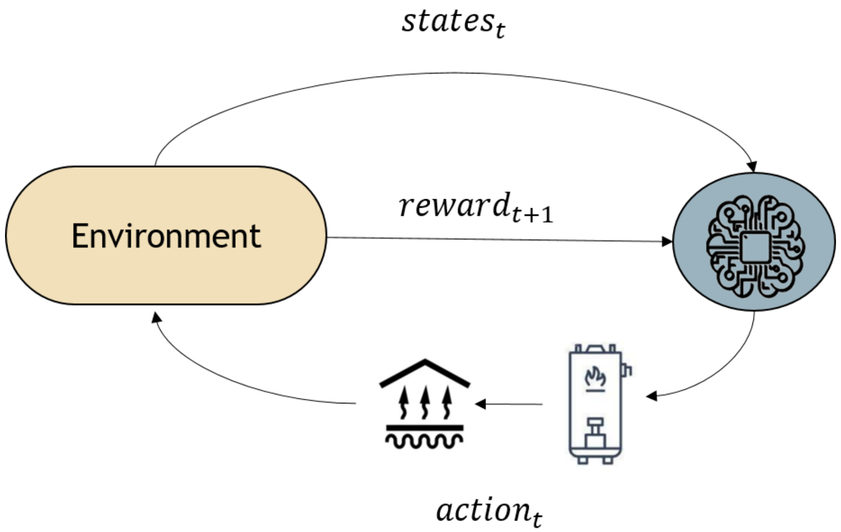
2. Classic control strategies (i.e., ON/OFF or PID) are inadequate to adapt to continually changing of users preferences, grid requirements and disturbances → Adaptive control

3. Model-based control strategies (e.g., Model Predictive Control (MPC)) were explored, showing excellent ability in improving comfort conditions and energy efficiency in buildings. However, their application is limited since requires the definition of an accurate model of the environment to be controlled → Model-free control strategies



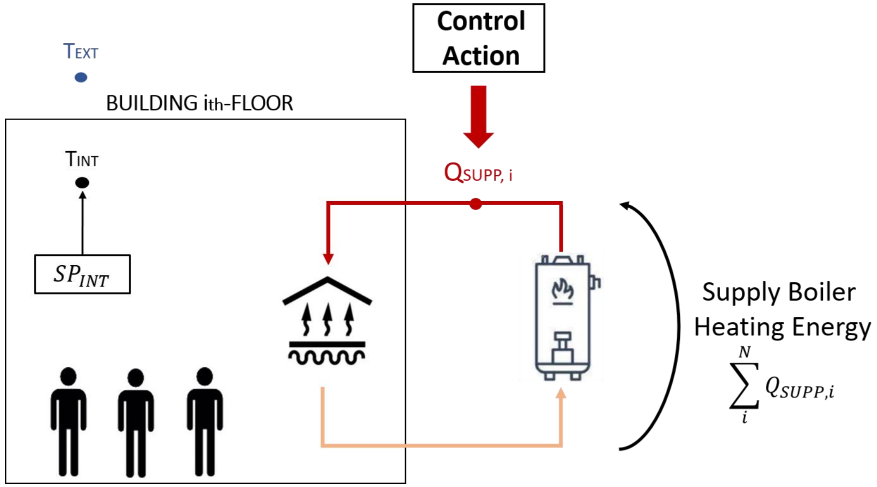
Paper Contribution

Application of an adaptive and model-free control strategy (Reinforcement Learning) to control the supply heating power per each floor of a residential building, enhancing comfort condition and energy efficiency for a residential building, with a probabilistic model that simulates the windows' opening/closing.



Evaluation of DRL agent adaptability properties with respect to:

- Different weather conditions.
- Indoor comfort requirements.
- Structural conditions (i.e., doubled thermal mass and installation of high-performance windows) .



Methods: Reinforcement Learning

It is a **learning technique** that aims at realizing control agents.

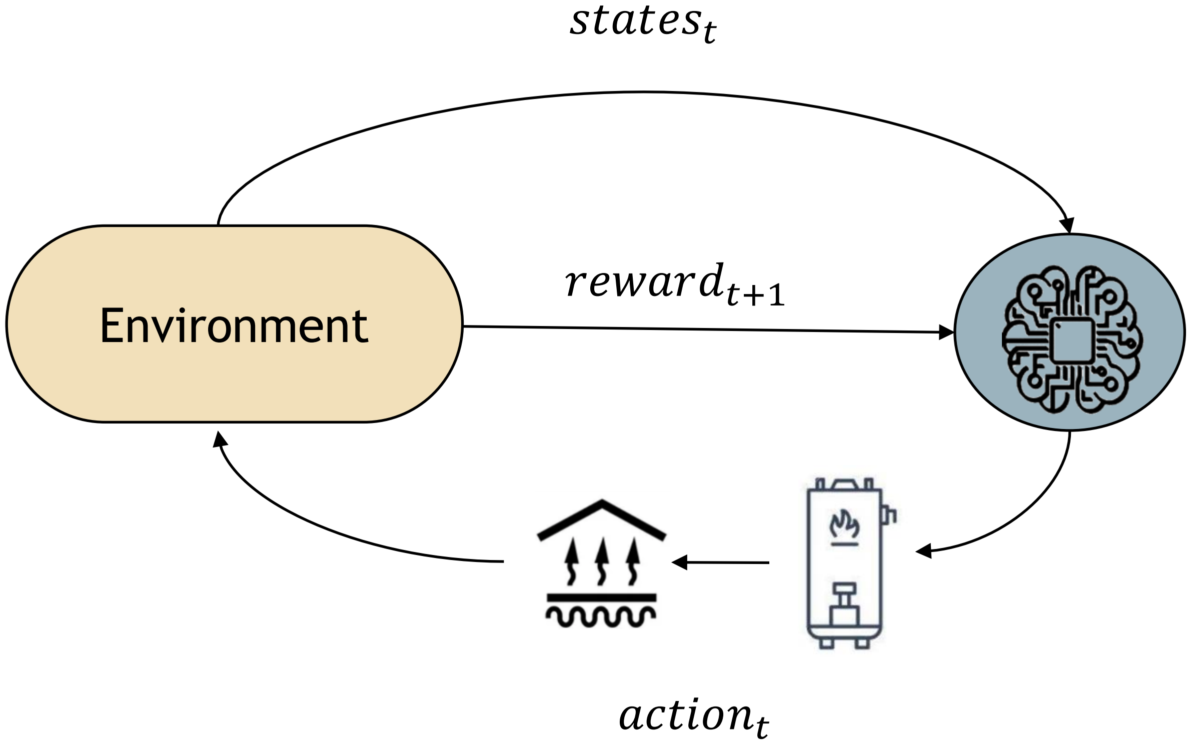
RL agent is trained through trial-and-error interaction with the environment to learn an **optimal control policy π** that maximizes the objective function, called **reward function.**

Two functions are used to define the problem and show the expected return of the control policy:

- State-value function $v_\pi(s)$
- Action-value function $q_\pi(s,a)$

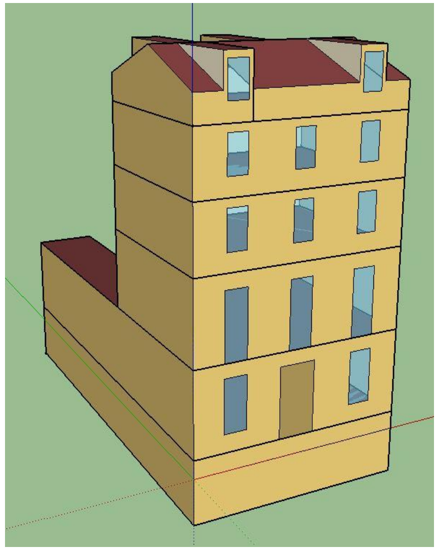
Soft Actor-Critic (SAC) is an **off-policy** control algorithm which allow the use of **continuous action and state space.**

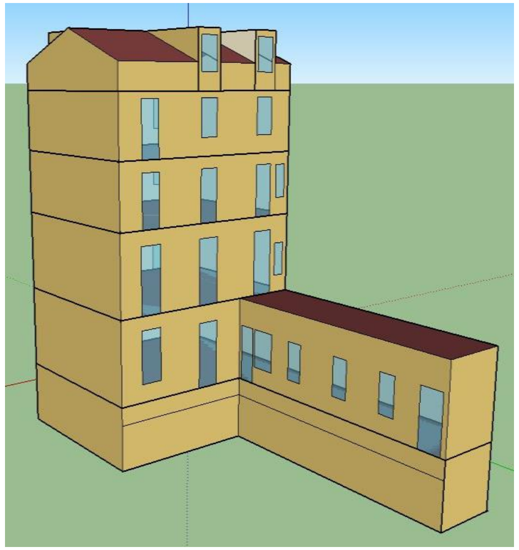
State-value and action-value functions are approximated with two different neural networks: the **Actor** and the **Critic.**



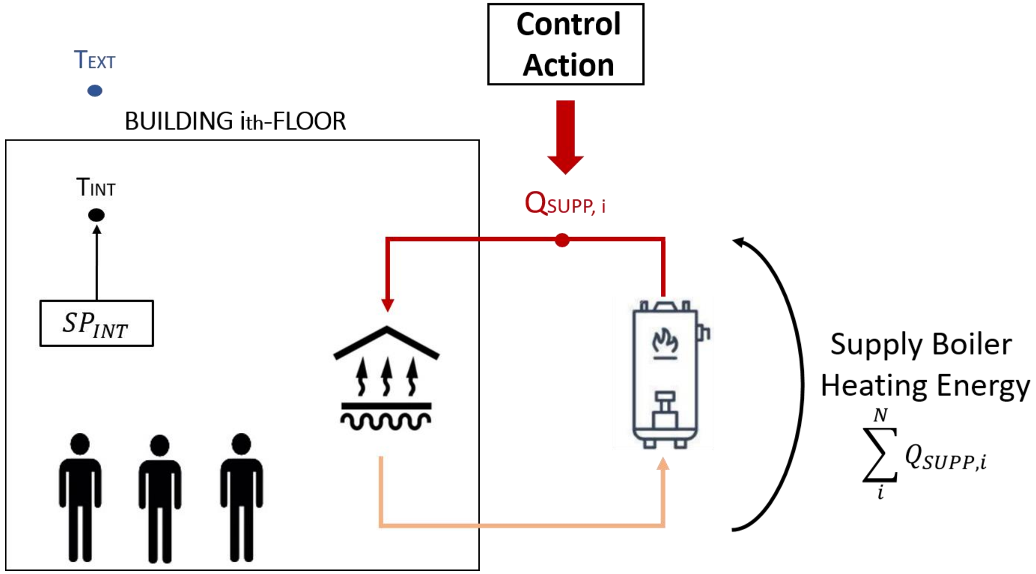
Case Study

The residential building is located in Turin, Italy, and consists of five heated floors plus the basement. Each floor corresponds to a thermal zone.





The building is heated by a **radiant floor heating system** and the heating energy is provided by a natural gas-fired boiler.



Windows' opening behaviour models

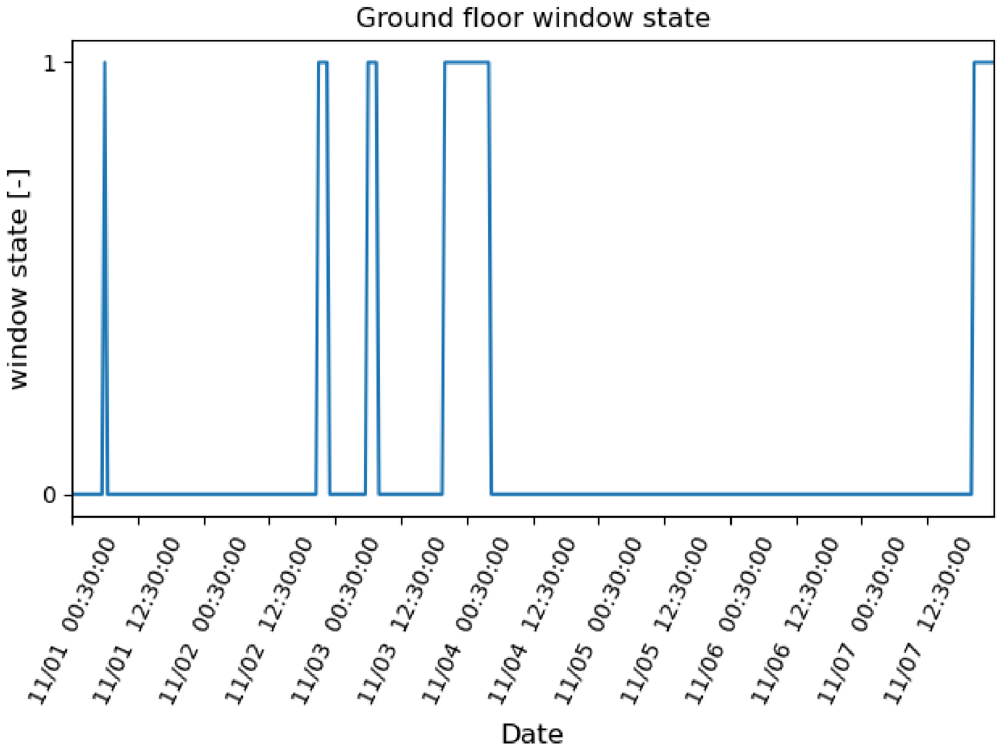
To better represent the **occupants' behaviour**, and in particular the **windows opening and closing**, four different model were tested

Each model proposes two equations based on **logistic regression** to predict the state (open/close) of windows

$$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 * x_1 + \alpha_2 x_2$$

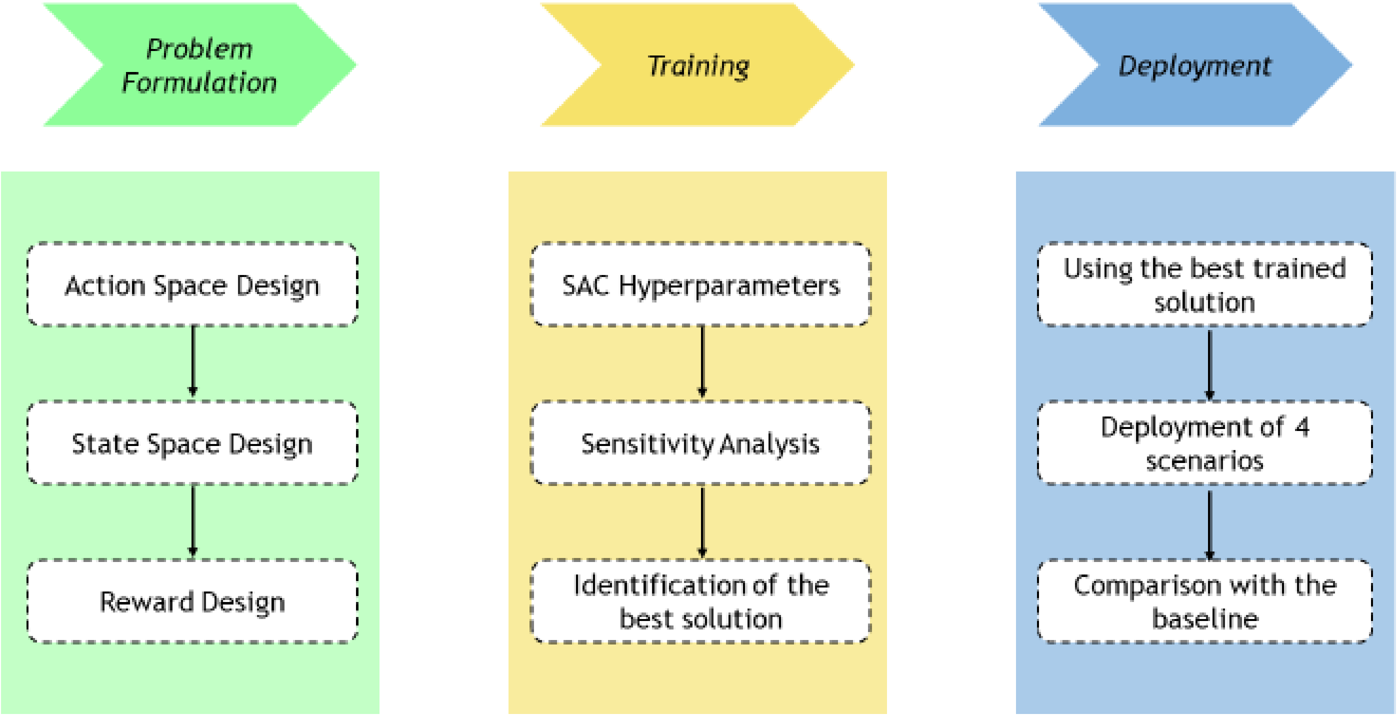
The model proposed by Anderson et al. was selected after a qualitative analysis and it was implement in the building's energy model.

It adopts as predictive variables the indoor temperature and relative humidity, CO_2 concentration, outdoor temperature and relative humidity, wind speed and solar radiation.



Source: Anderson et al., 2011

Framework



Problem Formulation

Action Space Design

State Space Design

Reward Design

Training

SAC Hyperparameters

Sensitivity Analysis

Identification of the best solution

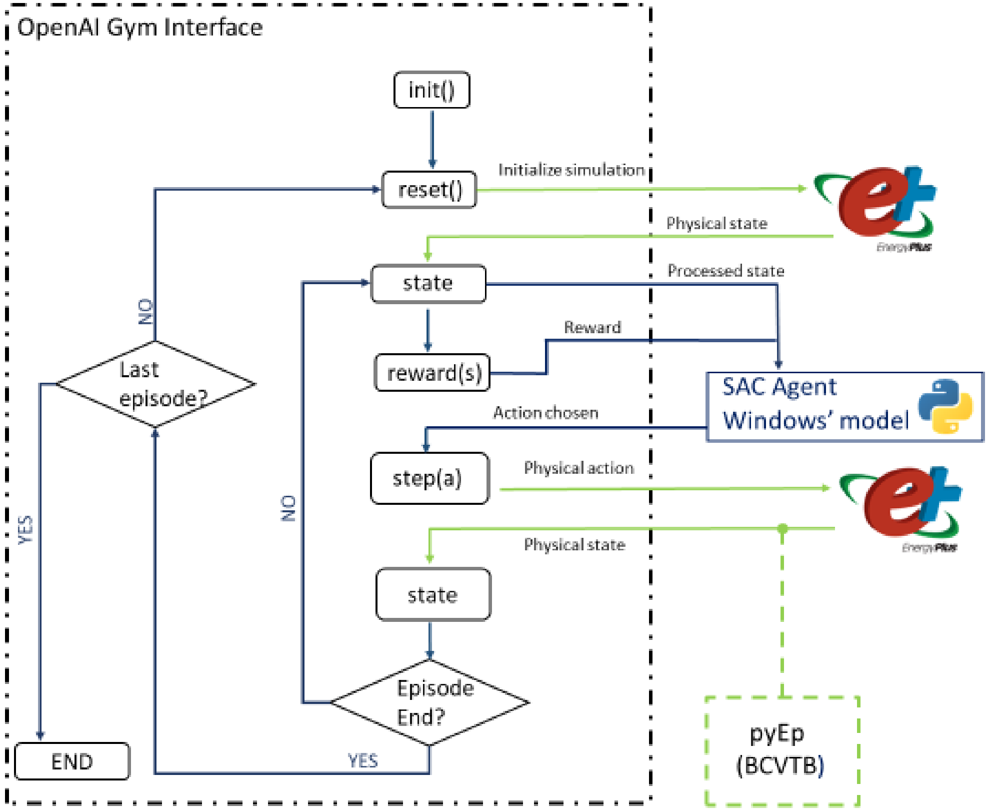
Deployment

Using the best trained solution

Deployment of 4 scenarios

Comparison with the baseline

Co-simulation environment combining **EnergyPlus** and **Python** through **BCVTB.**



Control Strategies – Baseline Control Logic

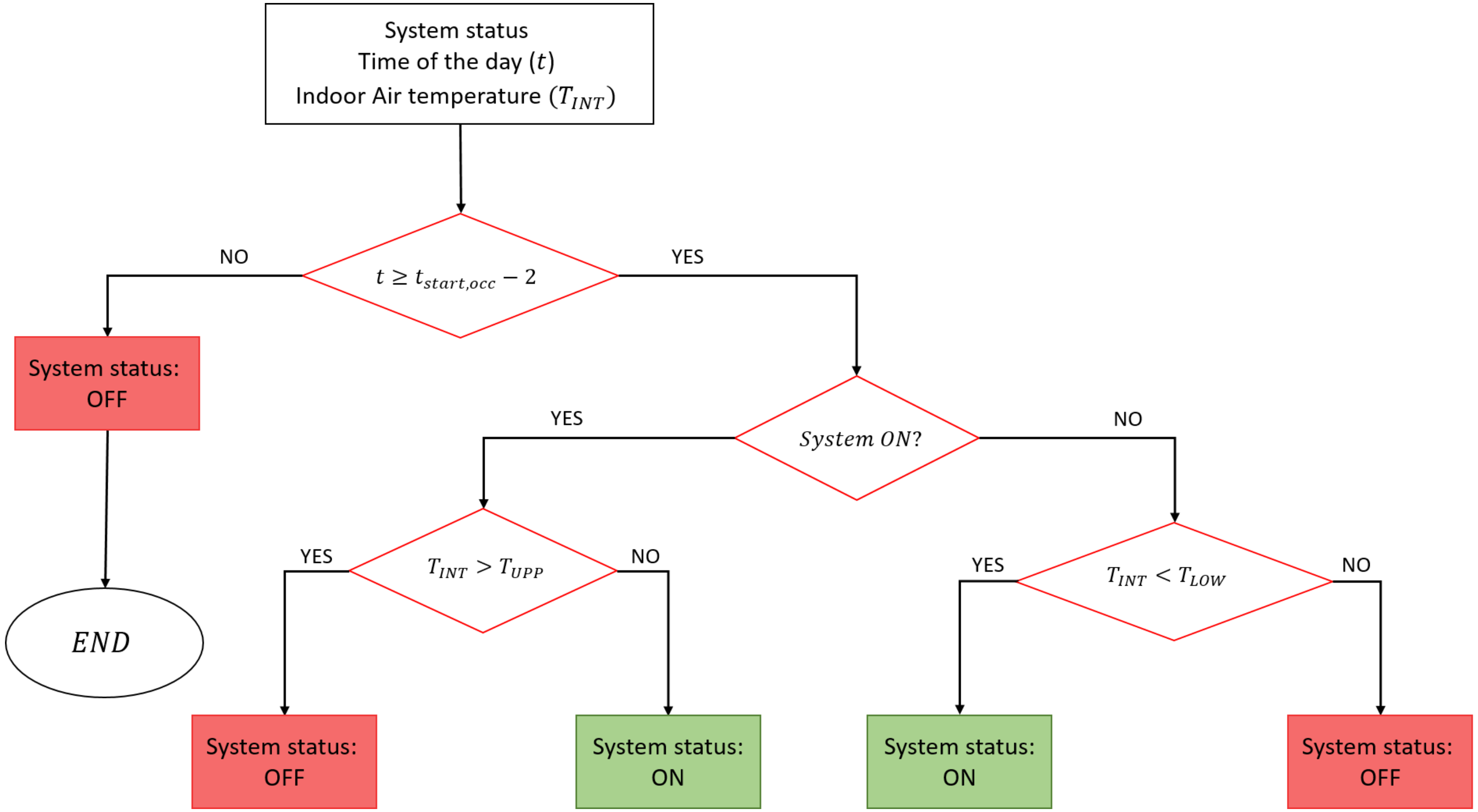
The baseline control logic is a **combination** of **rule-based** and **climatic-based** for the control of the supply power

The system is switched on two hours before the arrival of people;

If the indoor temperature is larger than 21 °C, the system is switched off

If the indoor temperature is less than 19 °C, the system is switched on

The system is switched off when occupants leave the thermal zone



Control Strategies – Design of DRL Controller

The goal of the designed controller is to **optimize** both the **energy consumption** and the **control of the indoor temperature** during the occupancy period. The setpoint was set equal to 20 °C and the temperature acceptability ranges between 19 °C and 21 °C.

The **action-space** is continuous, and it includes the supply power values of each floor.

Actions	Constraints [kW]
$A_{ground\ fl.}$	$0 \leq SP_{ground\ fl.} \leq 11$
$A_{first\ fl.}$	$0 \leq SP_{first\ fl.} \leq 6.5$
$A_{second\ fl.}$	$0 \leq SP_{second\ fl.} \leq 5.0$
$A_{third\ fl.}$	$0 \leq SP_{third\ fl.} \leq 5.0$
$A_{fourth\ fl.}$	$0 \leq SP_{fourth\ fl.} \leq 6.5$

The **state-space** is composed of 26 adaptive variables

Hour of the Day
Day of the Week

Outdoor Temperature
Direct Solar Radiation

Time to Occupancy
Start/End

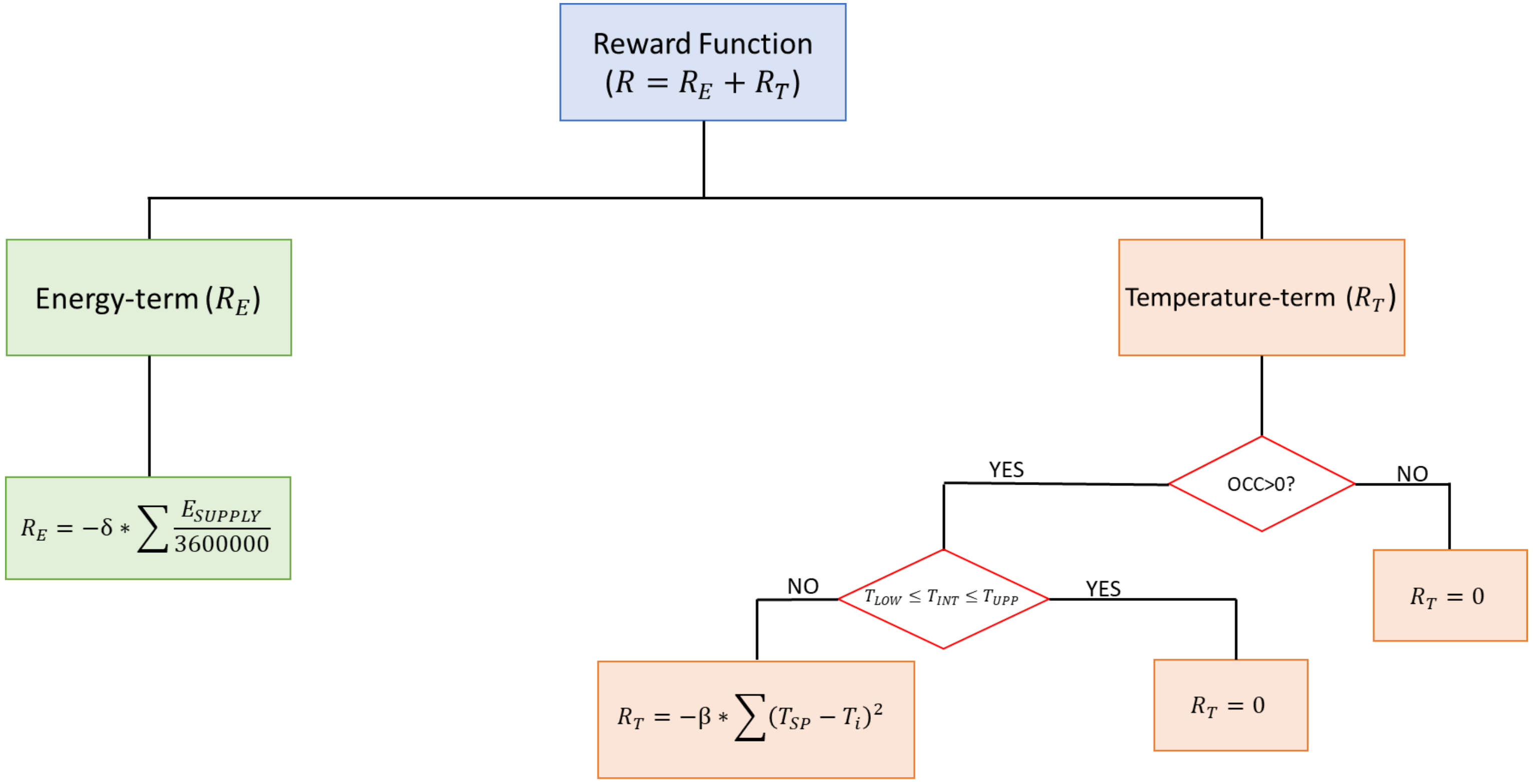
ΔT $T_{sp} - T_{zone,i}$
ΔT $T_{sp} - T_{zone,i}$ 1 hour lag
ΔT $T_{sp} - T_{zone,i}$ 2 hours lag
ΔT $T_{sp} - T_{zone,i}$ 4 hours lag

Control Strategies – Design of DRL Controller

The **reward** that the agent receives after having taken actions at each control time step depends on two competing values:

- **energy-related** term
- **temperature-related** terms.

The coefficient **δ** and **β** are introduced to weight the importance of the two terms of the reward function.



Reward Function ($R = R_E + R_T$)

Energy-term (R_E)

$R_E = -\delta * \sum \frac{E_{SUPPLY}}{3600000}$

Temperature-term (R_T)

OCC>0?

YES

NO

$R_T = 0$

$T_{LOW} \le T_{INT} \le T_{UPP}$

NO

YES

$R_T = -\beta * \sum (T_{SP} - T_i)^2$

$R_T = 0$

δ weight of the energy-related term

β weight of the temperature-related term

Training phase

Hyperparameters influence the **behaviour and performance** of the DRL agent → it was performed a **sensitivity analysis on hyperparameters**

The KPIs considered to assess DRL performance:

1. Energy saving (with reference to baseline).
2. Cumulative sum of temperature violations during the occupancy hours

$$\sum_{i=0}^{t_{end}} T_{VIOLATION,i} \quad [°C]$$

A temperature violation occurs when, during the presence of occupants, the temperature is not within the acceptability range [-1, 1] of the 20°C set-point. **It is estimated in this way:**

1. If $T_{int} < T_{LOW}$:
   $T_{VIOLATION} = T_{LOW} - T_{int}$
2. If $T_{int} > T_{UPP}$:
   $T_{VIOLATION} = T_{UPP} - T_{int}$

Hyperparameters	Value
DNN architecture	3 layers
Episode length	61 days (2928 control step)
Buffer Size	11520
Discount factor	0.9; 0.95; 0.99
Learning rate	0.001; 0.0005; 0.0001
β	1; 5; 10
δ	0.1; 0.01
Batch size	128; 256; 512
Neurons per hidden layer	128; 256; 512
Number of episodes	10; 25

Deployment phase

In the deployment phase the agent's **adaptability** to change in the environment is tested.

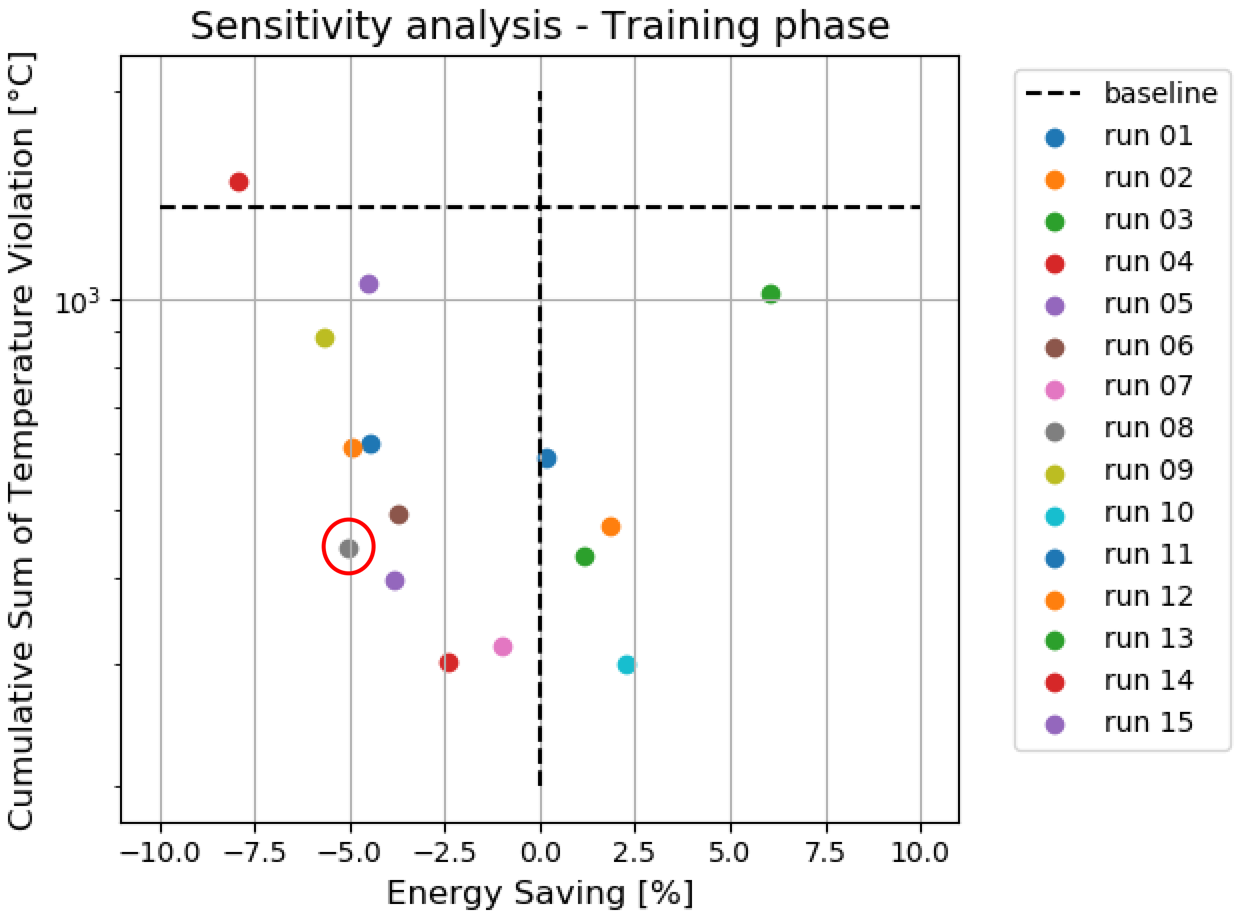
The agent is tested in four different scenarios, and the simulation lasts two months, from 1st January to 28th February.

- S1 - **weather**: the control environment does not change, the aim is to evaluate the adaptability of the control agent to different weather conditions;
- S2 – **indoor temperature**: the goal of this test is to evaluate the adaptability of the SAC controller in satisfying different temperature requirements. The zone setpoint temperature was set equal to 21 °C, consequently, the new acceptability range was between 20 and 22 °C.
- S3 - **windows**: in this scenario, the agent's adaptability was tested improving the energy performance of the transparent building envelope.
- S4 – **internal mass**: the internal mass was increased to rise the thermal inertia of the building.

Results – Training phase

The configuration chosen is the eighth, which presents the best **trade-off** between energy saving and reduction of cumulative sum of temperature violations.

Hyperparameters	Value
DNN Architecture	3 layers
Discount factor	0.9
Learning rate	0.0001
β	1
δ	0.1
Batch size	256
Neurons per hidden Layer	256
Number of episodes	25



Results - Deployment phase

- Energy savings is calculated for i-th scenario with respect to the respective i-th baseline as follows:

$$E_{savings,i} = 100 * \frac{E_{scenario,i} - E_{baseline,i}}{E_{baseline,i}} \; [\%]$$

- Energy savings reached are low, with a maximum in the range of 5% for the S4 scenario

- The cumulative sum of temperature violations differences with baseline is calculated for i-th scenario as follows:

$$[\Delta\Sigma T_{VIOL}]_i = \Sigma_{scen,i} T_{VIOL,j} - \Sigma_{baseline,i} T_{VIOL,j} [°C]$$

- During the deployment period the cumulative sum of temperature violations ranged between 400 and 600 °C, with a peak for the scenario S3.

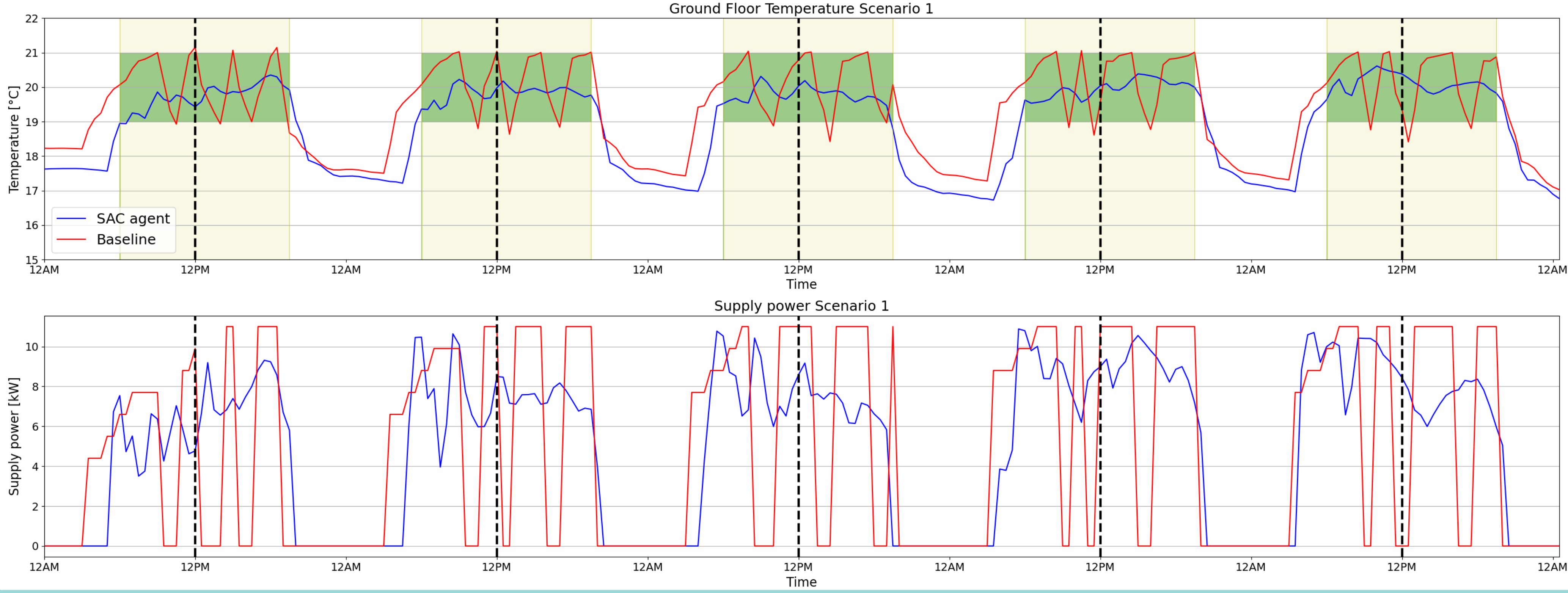
ENERGY HEATING CONSUMPTION			
Scenario	DRL Logic [MWh]	Baseline Logic [MWh]	Energy Saving [%]
1	20.6	21.2	-2.8
2	22.6	23.6	-4.2
3	20.0	20.3	-1.5
4	20.8	22.0	-5.5

CUMULATIVE SUM OF TEMPERATURE VIOLATION			
Scenario	DRL Logic [°C]	Baseline Logic [°C]	$\Delta\Sigma T_{VIOL}$ [°C]
1	592.2	1447.8	-855.6
2	395.3	1158.6	-763.3
3	603.2	1553.1	-949.2
4	505.6	508.9	-3.3

Results - Deployment phase

In scenario S1, the DRL agent allows to **reduce the temperature violations and energy supplied** through an optimal management of the pre-heating phase.

In particular, the developed agent switches-ON the heating system later than the baseline, reducing the corresponding energy supplied and ensuring that indoor comfort requirements are met.



Conclusions and Future works

Conclusions

- The main result achieved by the SAC control agent is the significant reduction of the cumulative sum of temperature violations in all scenarios, obtaining at the same time an energy saving.

- The SAC control agent presents good **adaptability** in:
  a) Outdoor weather conditions;
  b) Indoor comfort requirements;
  c) Building envelope change;

Future Works

- Introduce comfort parameters, such as Predicted Percentage of Dissatisfied (PPD) and the Predicted Mean Vote (PMV), in the reward function.

- Compare the performance of SAC with model-based solution such as MPC. A comparison in terms of performance, computational cost and modelling effort could be interesting.

TEBE
TECHNOLOGY ENERGY BUILDING ENVIRONMENT

Politecnico di Torino
1859
Department of Energy
"G.Ferraris"

BAEDA Lab
Building
Automation
Energy
Data
Analytics

BAEDA Lab website: www.baeda.polito.it

E-mail: silvio.brandi@polito.it
Or: davide.coraci@polito.it
Or: alfonso.capozzoli@polito.it

THANK YOU FOR PAYING ATTENTION!