

CityLearn Challenge 2023

Andrea Blushi

1 Introduzione

La CityLearn Challenge 2023 affronta la natura del controllo avanzato negli edifici, unendo le sfide legate alla progettazione di algoritmi di controllo, alla qualità delle previsioni e alla resilienza della rete elettrica. Grazie all'utilizzo della libreria CityLearn, i partecipanti mirano a massimizzare le prestazioni in termini di comfort e consumo energetico della rete.

L'obiettivo di questo documento è analizzare, attraverso il dataset della CityLearn Challenge 2023 - Phase 3 (dataset che coinvolge le tipologie di dispositivo interessate e possibili casistiche), le prestazioni dei diversi algoritmi e, in particolare, delle funzioni di reward, sia quelle proposte dai partecipanti che quelle da noi sviluppate. Questo lavoro rappresenta un'estensione della relazione principale ed verranno ignorate dettagli della sfida che non interessano lo sviluppo dei reward ed algoritmi.

2 Criteri di Valutazione

La competizione CityLearn 2023 si articola in due tracce principali:

- **Control Track** (traccia considerata): sviluppo di agenti intelligenti per la gestione energetica, finalizzati all'ottimizzazione delle prestazioni della rete elettrica;
- **Forecasting Track**: sviluppo di modelli di regressione avanzati per la previsione dei profili di consumo energetico a 48 ore.

Per quanto riguarda la Control Track, che rappresenta la nostra analisi, è essenziale comprendere la struttura di valutazione adottata per comprendere su quali parametri i partecipanti si sono concentrati. Il punteggio finale (*ScoreControl*) viene calcolato come combinazione ponderata di quattro componenti fondamentali:

- *ScoreControlComfort*: valuta il comfort termico degli occupanti;
- *ScoreControlEmissions*: misura le emissioni di carbonio prodotte;
- *ScoreControlGrid*: analizza le prestazioni della rete elettrica;
- *ScoreControlResilience*: quantifica la capacità di reazione durante interruzioni di corrente.

La formulazione matematica del punteggio è data da:

$$Score_{Control} = 0.3 \cdot Score_{Control}^{Comfort} + 0.1 \cdot Score_{Control}^{Emissions} + 0.3 \cdot Score_{Control}^{Grid} + 0.3 \cdot Score_{Control}^{Resilience}$$

Dove considerandole come Cost Function risultano le corrispettive medie:

$$Score_{Control}^{Comfort} = \text{discomfort_proportion}$$

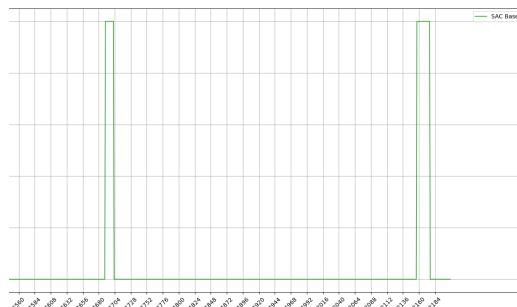
$$Score_{Control}^{Emissions} = \text{carbon_emissions_total}$$

$$Score_{Control}^{Grid} = \begin{cases} \text{ramping_average} \\ \text{daily_one_minus_load_factor_average} \\ \text{daily_peak_average} \\ \text{all_time_peak_average} \end{cases}$$

$$Score_{Control}^{Resilience} = \begin{cases} \text{one_minus_thermal_resilience_proportion} \\ \text{annual_normalized_unserved_energy_total} \end{cases}$$

3 Situazione generale del dataset

È importante sottolineare alcune casistiche rilevanti presenti nel dataset principale, al fine di analizzare come varia il comportamento del sistema in risposta a determinati eventi. Tra queste, spiccano le interruzioni di corrente, che si verificano in specifici timestep ben definiti. E' da specificare che ogni timestep corrisponde ad un'ora di simulazione.

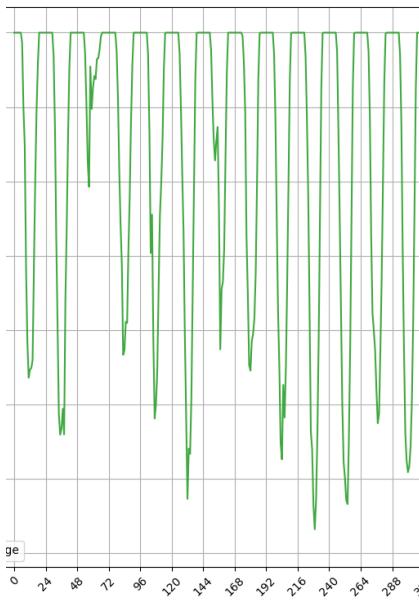


Un altro aspetto significativo riguarda il setpoint della temperatura, che tende a rimanere in modalità cooling (raffreddamento) attorno al 25°C. Il comfort termico viene definito da una banda di tolleranza pari a ±2°C, pertanto il sistema è considerato in comfort quando la temperatura interna si

mantiene (in caso di setpoint a 25) tra i 23°C e i 27°C. Sono comunque presenti leggere variazioni del setpoint nel tempo, come mostra il seguente grafico.



Un ulteriore aspetto da considerare è la generazione di energia solare. Il grafico seguente mostra l'andamento della potenza generata durante l'arco della giornata:



Si osserva che i valori diventano negativi durante il giorno. Questo comportamento è dovuto al fatto che il grafico rappresenta la potenza rilasciata dal sistema: un valore negativo indica energia generata e immessa nella rete o utilizzata localmente. Di conseguenza, il grafico va interpretato in maniera inversa: maggiore è il valore negativo, maggiore è la generazione solare. Si evidenzia dunque una produzione significativa di energia nelle ore diurne. Inoltre, si nota una diminuzione della generazione solare tra i timestep 48

e 72. Questo calo può essere attribuito a una giornata meno soleggiata, caratterizzata da nuvolosità o piogge, che riducono l'efficienza dei pannelli fotovoltaici. Tali condizioni meteorologiche sono confermate dall'analisi del file weather.csv, che fornisce informazioni dettagliate su irraggiamento solare, copertura nuvolosa e precipitazioni.

4 SAC Base

In questa sezione analizziamo le prestazioni dell'algoritmo SAC (Soft Actor-Critic) implementato da CityLearn, utilizzandolo con 10 episodi di addestramento e valutando solo le statistiche del primo edificio (Building 0), al fine di valutare l'impatto delle diverse funzioni di reward sulle metriche chiave del sistema. L'approccio base, che inizialmente effettua scelte casuali per esplorare lo spazio delle azioni, verrà esaminato per comprendere come evolve il suo comportamento al variare dei reward durante il ciclo di addestramento. Utilizzeremo un approccio incrementale per raggiungere un reward che soddisfi i nostri obiettivi. L'algoritmo SAC risulta una semplificazione rispetto agli algoritmi adottati dai vincitori, ma ci permette in questo caso di analizzare bene le problematiche di questo algoritmo e dei suoi reward.

4.0.1 Reward Base

Formula

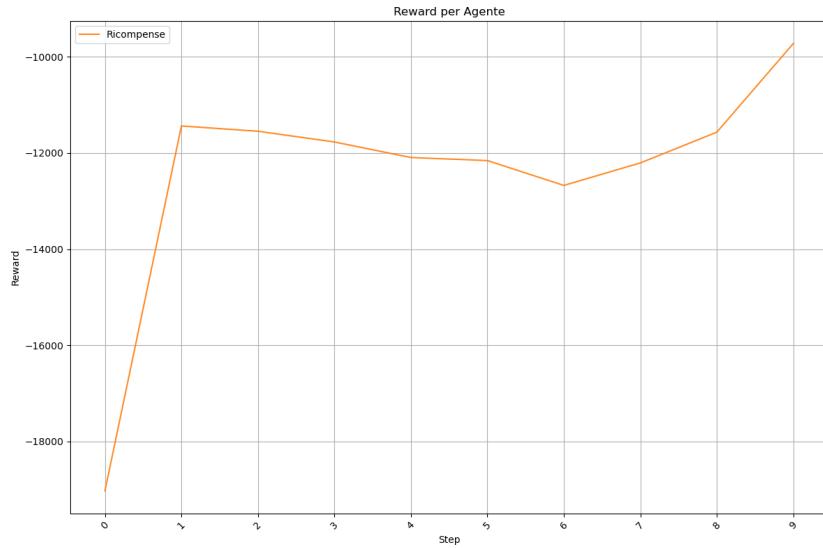
$$\min(-e, 0)$$

Statistiche generali Di seguito sono riportate le funzioni di costo coinvolte nella Reward Function di base:

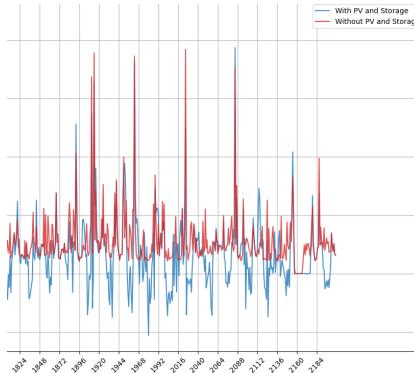
name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
cost_function							
all_time_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.834
annual_normalized_unserved_energy_total	0.017	0.015	0.020	0.012	0.026	0.019	0.018
carbon_emissions_total	0.586	0.846	0.682	0.634	0.403	0.500	0.609
daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.673
discomfort_proportion	0.840	0.420	0.794	0.855	0.990	0.822	0.787
electricity_consumption_total	0.590	0.850	0.679	0.634	0.404	0.501	0.610
one_minus_thermal_resilience_proportion	1.000	0.895	1.000	1.000	1.000	0.970	0.977
ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	0.939

Nel continuo del documento verranno analizzate nel dettaglio alcune caratteristiche, ma come si può osservare, emergono subito consumi relativamente bassi e un discomfort relativamente elevato, risultati che rispecchiano l'obiettivo della reward.

Per ogni reward analizzato verrà mostrato come la somma dei reward per ciascun periodo vari fino a raggiungere l'episodio deterministico in cui si registra il picco; si può osservare come in pochi episodi si ottengano risultati prossimi alla convergenza. Il grafico mostra chiaramente come, dopo pochi episodi, la somma dei reward migliori sensibilmente, avvicinandosi gradualmente alla convergenza, ma non riuscendo a migliorare decisamente le prestazioni.



Comportamento dei consumi Il primo grafico che andremo ad analizzare riguarda i consumi generali. Data la sua complessità, verrà trattato solo in casi specifici. Lo consideriamo in questo caso per comprenderne la rappresentazione.



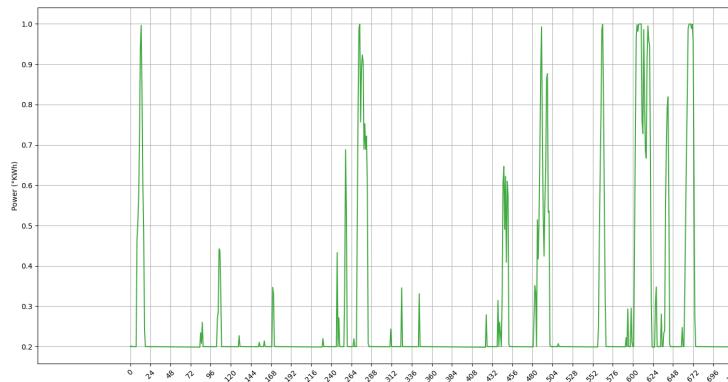
In blu è rappresentato il `net_electricity_consumption`, ovvero i consumi totali dei vari dispositivi. In rosso, invece, abbiamo il `net_electricity_consumption_without_storage`.

che corrisponde al precedente sottraendo l'energia utilizzata o restituita dalle batterie e di quella generata dai pannelli solari. Si può osservare come i consumi rossi siano generalmente più elevati, poiché viene ignorata l'autosufficienza dell'edificio, mentre nel caso del blu è possibile anche scendere sotto lo zero, restituendo energia alla rete.

Durante il blackout attorno allo step 2160, entrambi i grafici crollano a zero per un breve periodo. Una volta tornato il giorno, i consumi rossi aumentano mentre quelli blu rimangono bassi, indicando che durante il blackout il consumo è stato possibile solo grazie alla luce solare.

Infine, si nota che quando i picchi blu superino quelli rossi, si può ipotizzare un tentativo di immagazzinare energia che non è stata generata. Ciò è supportato anche dal grafico delle batterie. In generale, si osservano forti variazioni di corrente, che giustificano il ramping registrato e i picchi particolarmente elevati. Tuttavia, confrontando questo grafico con quelli successivi, si può notare come i consumi si riducano mediamente, risultando tra i più bassi.

Comportamento delle batterie Associato ai consumi, è interessante analizzare come l'agente utilizzi le batterie per ridurre i consumi. Di seguito viene mostrato un grafico dello stato di carica delle batterie.



Non essendo l'agente influenzato da reward legati all'uso delle batterie, si osservano accumuli casuali e sporadici, che raramente raggiungono i picchi. Nei reward successivi si vedrà quanto il loro utilizzo diventi più regolare se stimolato da una funzione di ricompensa adeguata.

È importante notare che, nonostante l'incapacità di apprendere un comportamento ottimale, l'agente tende comunque ad accumulare energia solo durante il giorno, comprendendo implicitamente che caricare di sera sia inefficiente ed uno spreco, in quanto non vi è generazione solare. Nonostante ciò,

l'utilizzo complessivo rimane inefficiente: ad esempio, durante un blackout, il sistema si è ritrovato con le batterie completamente scariche.

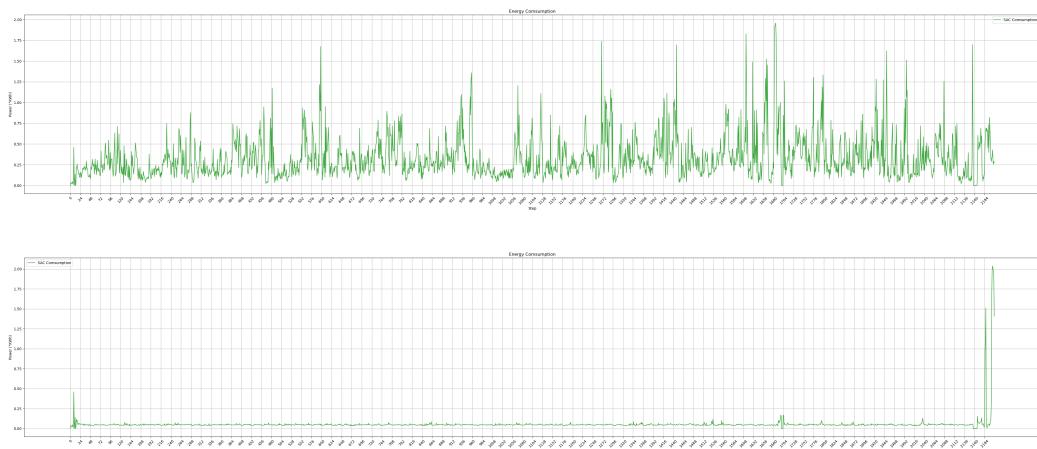
Comportamento dei dispositivi I consumi sono strettamente legati all'utilizzo dei dispositivi e, se l'agente non è incentivato a regolare il discomfort, è naturale pensare che l'energia utilizzata dai dispositivi di climatizzazione sia nulla, portando il discomfort al massimo. Nella realtà, però, non è così.

Infatti, CityLearn blocca l'agente nei primi 12 step, ignorando ogni sua azione sui dispositivi di climatizzazione e fornendo direttamente il carico ideale. Successivamente, il controllo viene restituito all'agente. Questo comportamento serve a fornire un input di addestramento su come utilizzare i dispositivi, ma in questo caso tende a confondere l'agente, che non riconosce tale fase come un totale spreco energetico.

Nonostante ciò, confrontando con altri algoritmi, si osserva un utilizzo ridotto dei dispositivi di raffreddamento, e tale tendenza diventa più evidente all'aumentare dei periodi di addestramento.

Con lo scopo di valutare, è stata introdotta un'ulteriore penalizzazione nei reward in caso di utilizzo dei dispositivi di raffreddamento: si può notare una drastica riduzione dell'uso, pur mantenendo una leggera instabilità, dovuta alla stimolazione iniziale nei primi 12 step.

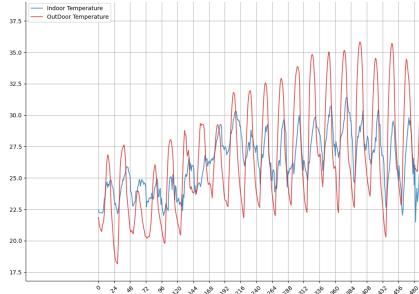
Di seguito viene riportato il confronto:



Inoltre, non si osserva un utilizzo coerente dei dispositivi di raffreddamento, il cui comportamento appare casuale. Ad esempio, allo step 432, si registra un picco di consumo a mezzanotte, nonostante un setpoint di 22°C, una temperatura esterna di 20°C e una temperatura interna di 23°C, condizioni in cui l'intervento del dispositivo sarebbe poco giustificato.

Situazioni simili si verificano anche in altri step, dove si notano picchi di utilizzo durante le ore notturne, momento in cui il fabbisogno reale sarebbe minimo.

In aggiunta, il grafico delle temperature mostra un andamento molto irregolare e instabile, suggerendo una difficoltà da parte dell'agente nel mantenere un controllo termico costante. Le variazioni repentine e spesso incoerenti tra interno ed esterno evidenziano una gestione poco efficace della climatizzazione.

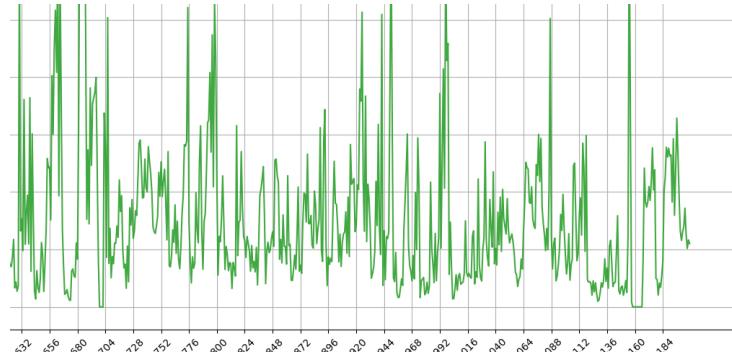


Un altro comportamento interessante da analizzare è quello durante i blackout. Un agente, idealmente, dovrebbe scegliere se soddisfare il fabbisogno dei *non_shiftable_load* (come ad esempio i frigoriferi) oppure dei dispositivi di raffreddamento, bilanciando l'ottimizzazione delle diverse Cost Function.

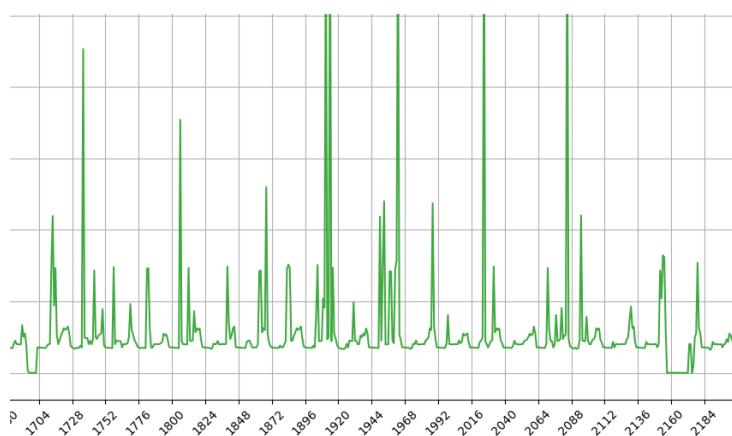
Nel dataset considerato sono presenti due blackout principali: attorno allo step 1680 e allo step 2160. Nel primo caso si osserva un picco iniziale di consumo, seguito da un collasso e da una ripresa piuttosto rapida. Nel secondo caso, invece, il sistema rimane vicino allo zero per un periodo più lungo, per poi riattivarsi con l'arrivo della produzione solare. Questo suggerisce che nel primo blackout l'agente ha tentato di consumare energia immediatamente, anche di notte, mentre nel secondo ha atteso la generazione solare del giorno. Questo rafforza l'idea che il comportamento sia ancora casuale.

Per quanto riguarda i *non_shiftable_load*, si preferisce non soddisfare il fabbisogno dei dispositivi essenziali, indicando una mancanza di priorità esplicita nel reward.

Va sottolineato che la gestione dei *non_shiftable_load* risulta quasi invariata tra i vari reward testati: le uniche differenze evidenti emergono durante i blackout. Questo porta alla considerazione che, nella progettazione del reward, è necessario per lo sviluppatore definire con chiarezza le priorità tra i diversi gruppi di dispositivi.

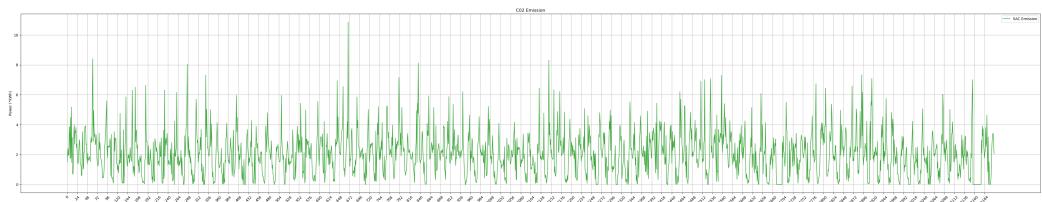


Cooling Device



Non-Shiftable Load

Comportamento delle emissioni Nella maggior parte dei casi, le emissioni seguono strettamente il comportamento dei consumi. Nonostante ciò, per completezza, si riporta di seguito un'immagine che conferma quanto detto.



Considerato il Reward Base e le caratteristiche generali dell'ambiente, l'analisi d'ora in poi si concentrerà esclusivamente sulle componenti che sono influenzate direttamente dal reward. In tale modo si possono riconoscere delle componenti per la costruzione di un reward più efficiente

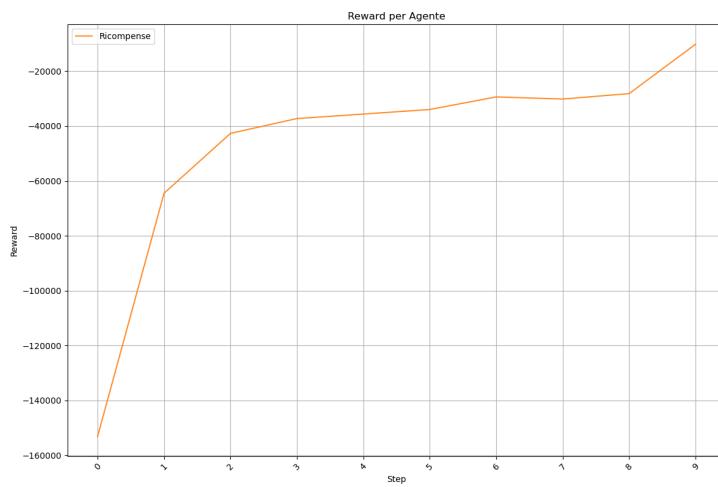
4.1 Comfort Reward

Formula

$$\begin{cases}
 -|T_{in} - T_{stp}|^3 & \text{if } T_{in} < (T_{stp} - T_b) \text{ and cooling} \\
 -|T_{in} - T_{stp}|^2 & \text{if } T_{in} < (T_{stp} - T_b) \text{ and heating} \\
 -|T_{in} - T_{stp}| & \text{if } (T_{stp} - T_b) \leq T_{in} < T_{stp} \text{ and cooling} \\
 0 & \text{if } (T_{stp} - T_b) \leq T_{in} < T_{stp} \text{ and heating} \\
 0 & \text{if } T_{stp} \leq T_{in} \leq (T_{stp} + T_b) \text{ and cooling} \\
 -|T_{in} - T_{stp}| & \text{if } T_{stp} \leq T_{in} \leq (T_{stp} + T_b) \text{ and heating} \\
 -|T_{in} - T_{stp}|^2 & \text{if } (T_{stp} + T_b) < T_{in} \text{ and cooling} \\
 -|T_{in} - T_{stp}|^3 & \text{otherwise}
 \end{cases}$$

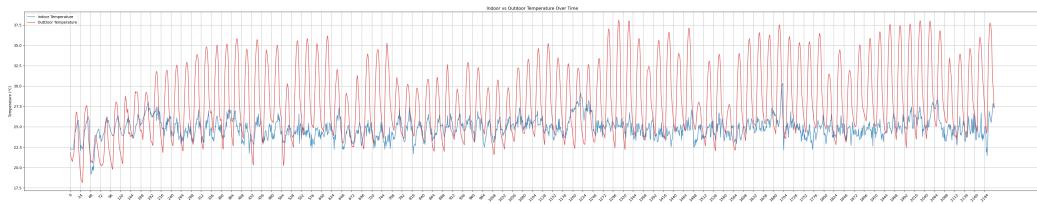
Statistiche generali Come si può ben immaginare, ignorando del tutto i consumi e considerando esclusivamente il comfort, le statistiche pendono esclusivamente da un lato, dimostrando consumi eccessivi ma un ottimo comfort. Gli unici due parametri che ne giovano quindi sono *discomfort_proportion* e la *one_minus_thermal_resilience_proportion*.

	name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
cost_function								
all_time_peak_average		NaN	NaN	NaN	NaN	NaN	NaN	0.880
annual_normalized_unserved_energy_total	0.020	0.016	0.014	0.018	0.019	0.014	0.017	
carbon_emissions_total	0.842	1.084	0.966	0.848	1.036	0.813	0.932	
daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.930	
discomfort_proportion	0.084	0.036	0.022	0.026	0.135	0.117	0.070	
electricity_consumption_total	0.840	1.083	0.966	0.851	1.036	0.810	0.931	
one_minus_thermal_resilience_proportion	0.364	0.789	0.429	0.515	0.870	0.636	0.600	
ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	1.035	



L'andamento dei reward, risulta molto simile a quello della Reward Base Function, dimostrando però una maggiore stabilità nella crescita.

Comportamento delle temperature In questo reward, risulta molto interessante valutare l'andamento delle temperature in quanto propone un comportamento da considerarsi ideale. Infatti a differenza di prima, le variazioni di temperatura interna si assestano molto vicine al setpoint, rimanendo poco sensibili alle temperature esterne, presentando comunque un adattamento da giorno a notte. Inoltre si può notare come sia abbastanza insensibile ai blackout, ignorando però il crollo verso la fine del blackout allo step 1680, in quanto dovuto alla fine di autonomia elettrica.

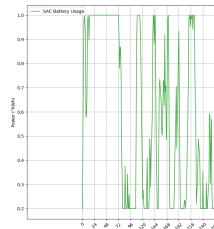


Comportamento dei consumi Ovviamente, una gestione ideale delle temperature comporta un costo. L'uso indiscriminato dei dispositivi di raffreddamento fa lievitare i consumi, spostando il grafico dei consumi verso l'alto rispetto al reward osservato in precedenza. Inoltre, si nota un aumento significativo del consumo dei dispositivi di raffreddamento, che però seguono uno schema più regolare. Ad esempio, inizialmente, le temperature esterne più basse richiedono un minor utilizzo dei dispositivi.

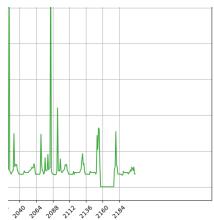


Comportamento delle batterie Poiché le batterie, come i consumi, sono totalmente ignorate e non influenzano il comportamento del comfort, il loro utilizzo risulta puramente casuale. Si può ipotizzare che l'agente comprenda che, durante i blackout, la possibilità di immagazzinare energia consenta comunque l'uso dei dispositivi di raffreddamento. Infatti, nei due periodi si osserva che viene consumata del tutto la batteria, mantenendo un utilizzo del dispositivo (sfruttando anche l'energia solare).

A sostegno della tesi della casualità, si osserva come l'agente conservi energia per tre giorni senza mai rilasciarla, risultando così inutile. Inoltre, si nota come tenda ad immagazzinare energia in modo eccessivo, aumentando così gli sprechi.



Comportamento dei Non-shiftable load Un problema significativo di questo reward è l'uso dei non-shiftable load. Sebbene, durante il normale andamento, il loro utilizzo rimanga invariato, durante il blackout vengono completamente spenti per favorire l'uso dei dispositivi di raffreddamento. Ciò, nella vita reale, potrebbe causare delle problematiche, ignorando le necessità degli occupanti.



Comportamento dei consumi Una caratteristica fondamentale di questo reward, rispetto ad altri, è che penalizza anche l'utilizzo inutile dell'energia. Per chiarire, si consideri il caso del dispositivo di raffreddamento: se la temperatura interna scende al di sotto del limite inferiore del setpoint, significa che il sistema di raffreddamento è stato utilizzato eccessivamente, e questo comportamento viene giustamente penalizzato. A supporto di ciò, l'errore in questa condizione viene elevato a un esponente pari a tre nella formula,

aumentando l'entità della penalità. Oltre a questa penalità, è presente anche una penalità nel caso l'agente superi il setpoint (nel caso di raffreddamento), raffreddando eccessivamente pur rimanendo all'interno della banda di comfort. Questo comportamento consente al sistema di mantenersi nella parte superiore della banda di comfort, permettendo un minor utilizzo del dispositivo di raffreddamento. Di conseguenza, nella valutazione dell'algoritmo, si osserva una riduzione dei consumi rispetto al baseline (di circa il 10%), in quanto quest'ultimo non effettua questa ottimizzazione.

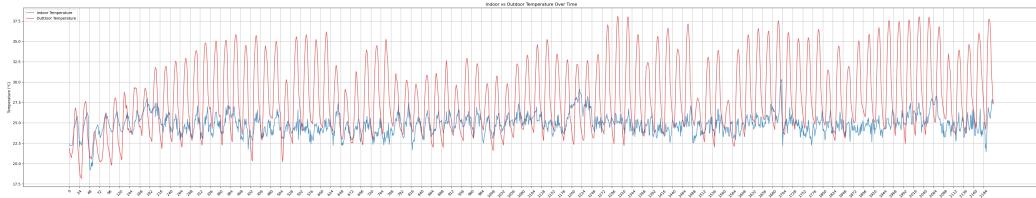
Andando ad aumentare il peso dell'errore, modificando la formula in questo modo:

$$\begin{cases} -3|T_{in} - T_{stp}|^3 & \text{if } T_{in} < (T_{stp} - T_b) \text{ and cooling} \\ -|T_{in} - T_{stp}|^2 & \text{if } T_{in} < (T_{stp} - T_b) \text{ and heating} \\ -|T_{in} - T_{stp}| & \text{if } (T_{stp} - T_b) \leq T_{in} < T_{stp} \text{ and cooling} \\ 0 & \text{if } (T_{stp} - T_b) \leq T_{in} < T_{stp} \text{ and heating} \\ 0 & \text{if } T_{stp} \leq T_{in} \leq (T_{stp} + T_b) \text{ and cooling} \\ -|T_{in} - T_{stp}| & \text{if } T_{stp} \leq T_{in} \leq (T_{stp} + T_b) \text{ and heating} \\ -|T_{in} - T_{stp}|^2 & \text{if } (T_{stp} + T_b) < T_{in} \text{ and cooling} \\ -3|T_{in} - T_{stp}|^3 & \text{otherwise} \end{cases}$$

Cioè, moltiplicando l'errore per un coefficiente, si può osservare una riduzione dei consumi, pagata però con un aumento del discomfort. Questo avviene perché l'agente, spinto da una eccessiva penalizzazione, sviluppa una sorta di "timore" nell'utilizzo del dispositivo di raffreddamento, portandolo così a non intervenire abbastanza, specialmente nei casi in cui la temperatura interna supera il limite superiore del comfort band. Di conseguenza, la formula originariamente sviluppata da CityLearn risulta particolarmente ben bilanciata, riuscendo a ottimizzare il comfort.

Di seguito tabelle e grafici riguardanti a questa modifica:

name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
cost_function							
all_time_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.880
annual_normalized_unserved_energy_total	0.020	0.016	0.014	0.018	0.019	0.014	0.017
carbon_emissions_total	0.842	1.084	0.966	0.848	1.036	0.813	0.932
daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.930
discomfort_proportion	0.084	0.036	0.022	0.026	0.135	0.117	0.070
electricity_consumption_total	0.840	1.083	0.966	0.851	1.036	0.810	0.931
one_minus_thermal_resilience_proportion	0.364	0.789	0.429	0.515	0.870	0.636	0.600
ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	1.035



4.2 Solar Penalty Reward

Formula

$$\sum_{i=0}^n -\left(\left(1 + \frac{e}{|e|} \times \text{storage}^{SoC} \right) \times |e| \right)$$

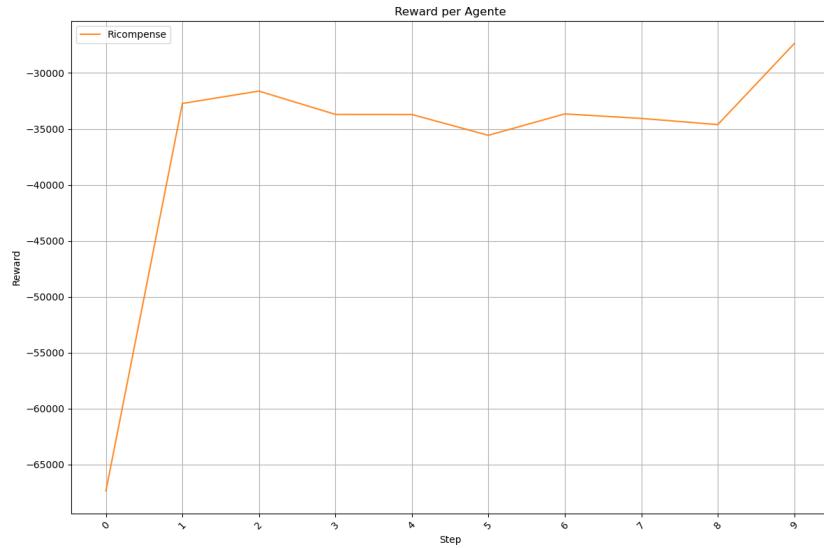
Statistiche generali Il Solar Penalty Reward è progettato per incentivare comportamenti che minimizzano il consumo di energia elettrica dalla rete e massimizzano l'uso dell'energia solare. Risulta un'evoluzione del reward di base aggiungendo le seguenti caratteristiche importanti:

- C'è una distinzione, e conseguente premio, tra consumo proveniente dalla rete elettrica e rilascio di energia;
- Si tiene in considerazione lo stato di carica delle batterie, modulando il reward su di esso;
- Aumenta la penalità quando l'edificio consuma energia e il sistema di accumulo non è completamente carico, oppure quando l'edificio produce energia in rete e il sistema di accumulo non è completamente scarico.

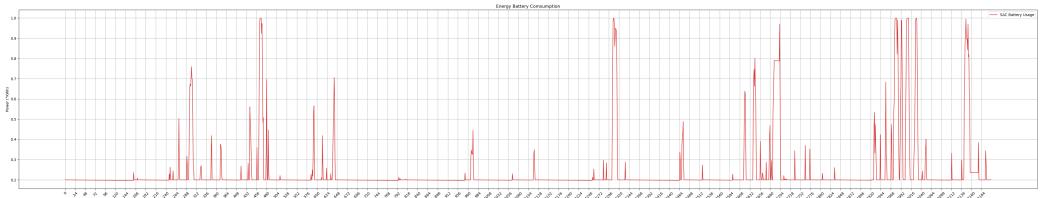
Attraverso questo reward si riescono ad ottenere delle statistiche riguardanti il consumo, ed il consumo durante blackout, molto simili rispetto ai reward base.

	name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
	cost_function							
	all_time_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.830
	annual_normalized_unserved_energy_total	0.009	0.005	0.011	0.014	0.011	0.014	0.011
	carbon_emissions_total	0.647	0.771	0.688	0.658	0.501	0.528	0.632
	daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.695
	discomfort_proportion	0.694	0.431	0.715	0.654	0.964	0.757	0.702
	electricity_consumption_total	0.653	0.768	0.688	0.659	0.508	0.525	0.633
	one_minus_thermal_resilience_proportion	0.879	0.842	1.000	0.879	0.957	0.970	0.921
	ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	0.914

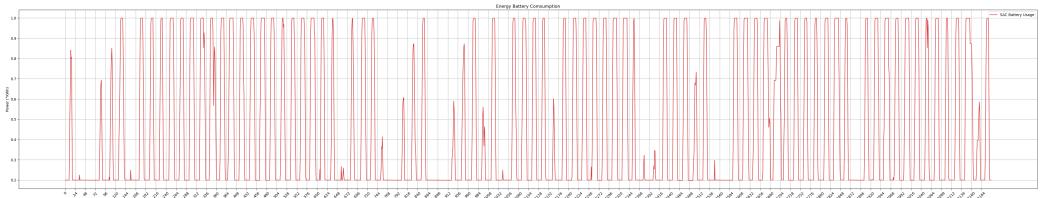
L’andamento dei reward, in questo caso, risulta più simile a quello del reward base, mostrando una maggiore instabilità. Da ciò si può trarre una considerazione piuttosto intuitiva: i reward legati ai consumi, a causa dell’instabilità intrinseca della rete energetica, sono più sensibili alle sollecitazioni e tendono a manifestare comportamenti meno regolari rispetto a quelli focalizzati esclusivamente sul comfort.



Comportamento delle batterie Per quanto riguarda gli altri comportamenti, il Solar Penalty mostra un andamento molto simile al Reward base, come ci si poteva aspettare. La particolarità emerge nell’utilizzo delle batterie: sebbene la formula teorica sembri ben bilanciata, essa risulta fortemente influenzata dal coefficiente legato al consumo. Questo coefficiente tende a penalizzare l’autoconsumo eccessivamente, limitando di fatto l’incentivo all’utilizzo della generazione solare immagazzinata. Inoltre, il focus sullo stato di carica fa sì che l’agente abbia ”paura” di caricare o scaricare troppo in certi momenti, portandolo ad evitare l’uso attivo della batteria. Si osserva infatti che, stimolando l’uso delle batterie con il parametro solar generation (aggiungendo al reward $\text{reward}+ = \text{solar}_{\text{generation}} * 4 * \text{storage}_{\text{soc}}$), l’agente viene maggiormente incentivato a sfruttare la batteria in modo più attivo e regolare, però si percepisce una perdita delle altre prestazioni, in quanto più spinto ad ignorare la produzione di energia. Nota positiva, che non sono più presenti momenti di ricarica durante ore notturne, o per lo meno sono ridotte.



Senza solar generation



Con solar generation

4.3 Solar Penalty And Comfort Reward

Formula

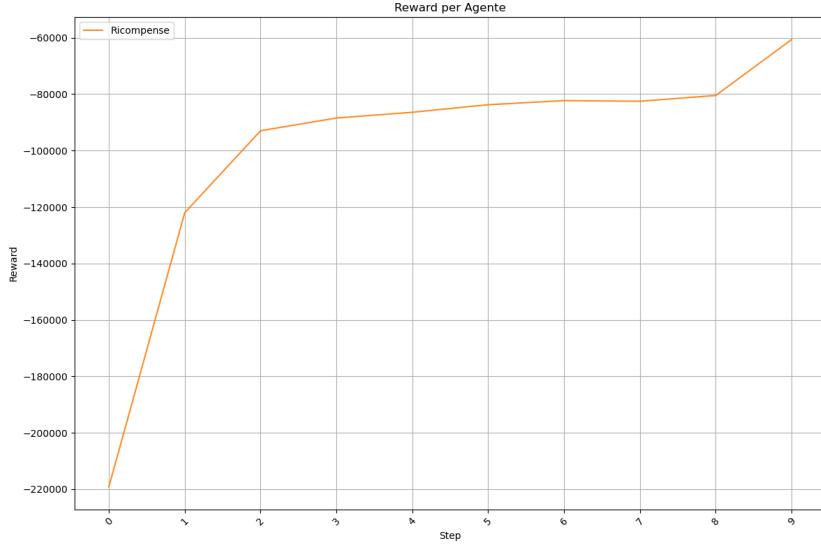
$$\text{Reward}_{\text{totale}} = \alpha \cdot \text{Reward}_{\text{solar}} + \beta \cdot \text{Reward}_{\text{comfort}}$$

Statistiche generali Finora sono state ottenute soluzioni che agivano esclusivamente o sul discomfort o sul consumo. Questo reward, invece, implementa direttamente la combinazione del Solar Penalty Reward e del Comfort Reward descritti in precedenza, unificandoli in un'unica formula a cui vengono assegnati, di default, pesi uguali. In questo modo, viene data priorità a entrambi gli aspetti in modo bilanciato.

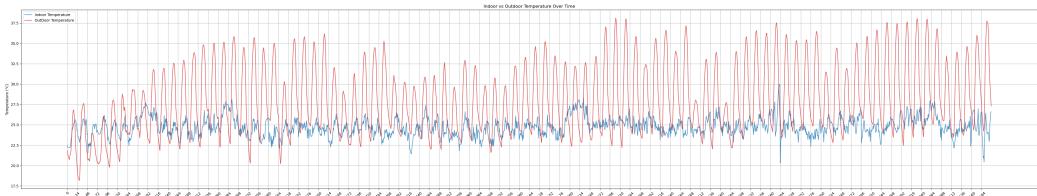
	name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
	cost_function							
	all_time_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.951
annual_normalized_unserved_energy_total	0.022	0.008	0.011	0.014	0.015	0.013	0.014	
carbon_emissions_total	0.831	1.056	0.929	0.844	0.994	0.764	0.903	
daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.876	
discomfort_proportion	0.110	0.035	0.039	0.035	0.157	0.155	0.089	
electricity_consumption_total	0.830	1.056	0.929	0.846	0.995	0.766	0.904	
one_minus_thermal_resilience_proportion	0.545	0.895	0.286	0.485	0.913	0.939	0.677	
ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	0.999	

Si può notare come, nonostante venga assegnato un peso uguale tra consumi e comfort, la situazione tenda a favorire maggiormente il comfort. Questo evidenzia quanto sia fondamentale non esagerare con la pesatura,

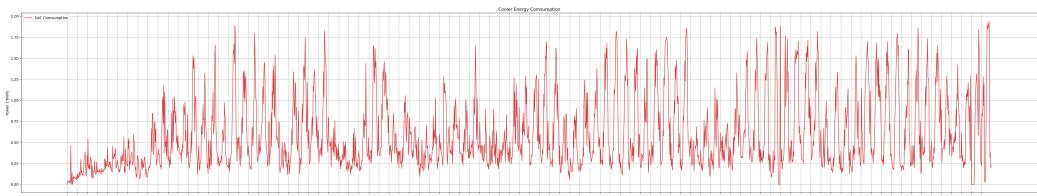
poiché potrebbe portare a una soluzione sbilanciata. Tale comportamento si osserva chiaramente anche nei reward ottenuti, che seguono più da vicino l'andamento del comfort rispetto ai consumi. Per ottenere un equilibrio risulta fondamentale l'assegnazione dei rispettivi pesi prima nella combinazione.



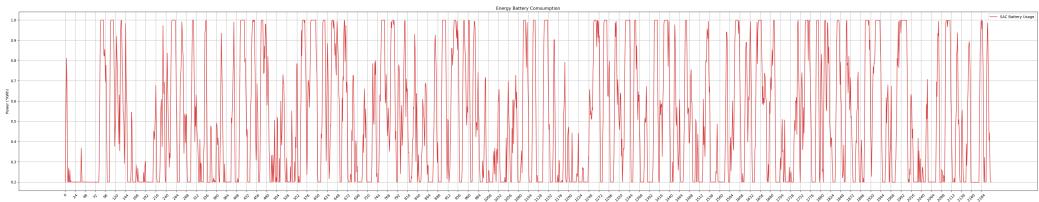
Comportamento delle temperature Come nel caso del comfort, le temperature tendono a essere più stabili e vicine al setpoint, sebbene si osservi una leggera instabilità.



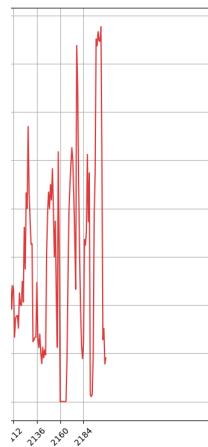
Questo comportamento comporta un maggiore utilizzo dei dispositivi durante la giornata, generando dei picchi di consumo che aumentano i consumi generali. Inoltre, tali picchi influenzano direttamente il ramping, un valore che deve essere attentamente monitorato per garantire la stabilità della rete.



Comportamento delle batterie Risulta interessante osservare come, con l'aumentare dei consumi dovuti ai dispositivi di raffreddamento, l'agente sia maggiormente stimolato a utilizzare le batterie in modo più regolare, accumulando energia durante il giorno e rilasciandola di notte. Questo comportamento porta a un utilizzo delle batterie, contribuendo anche alla riduzione delle emissioni di carbonio. Tuttavia, le batterie non riescono a risolvere la problematica del ramping, poiché vengono incentivate unicamente alla riduzione complessiva dei consumi, senza considerare direttamente la variabilità istantanea della potenza richiesta, accentuando così le problematiche dovute al dispositivo di raffreddamento. Inoltre, si può notare come, in alcuni casi, la batteria tenda ad accumulare energia durante la notte, causando così un aumento complessivo dei consumi. Ciò può essere dovuto al peso ridotto del reward.

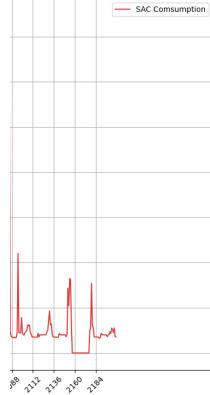


Comportamento durante i blackout Un utilizzo più efficiente delle batterie comporta un miglioramento significativo della flessibilità durante i blackout. In particolare, si registrano valori di discomfort sensibilmente inferiori, poiché i dispositivi di raffreddamento riescono a rimanere operativi per buona parte della durata del blackout, garantendo così condizioni termiche più stabili e confortevoli per gli occupanti. Inoltre si riesce a migliorare la *annual normalized unserved energy*.



Cooling device

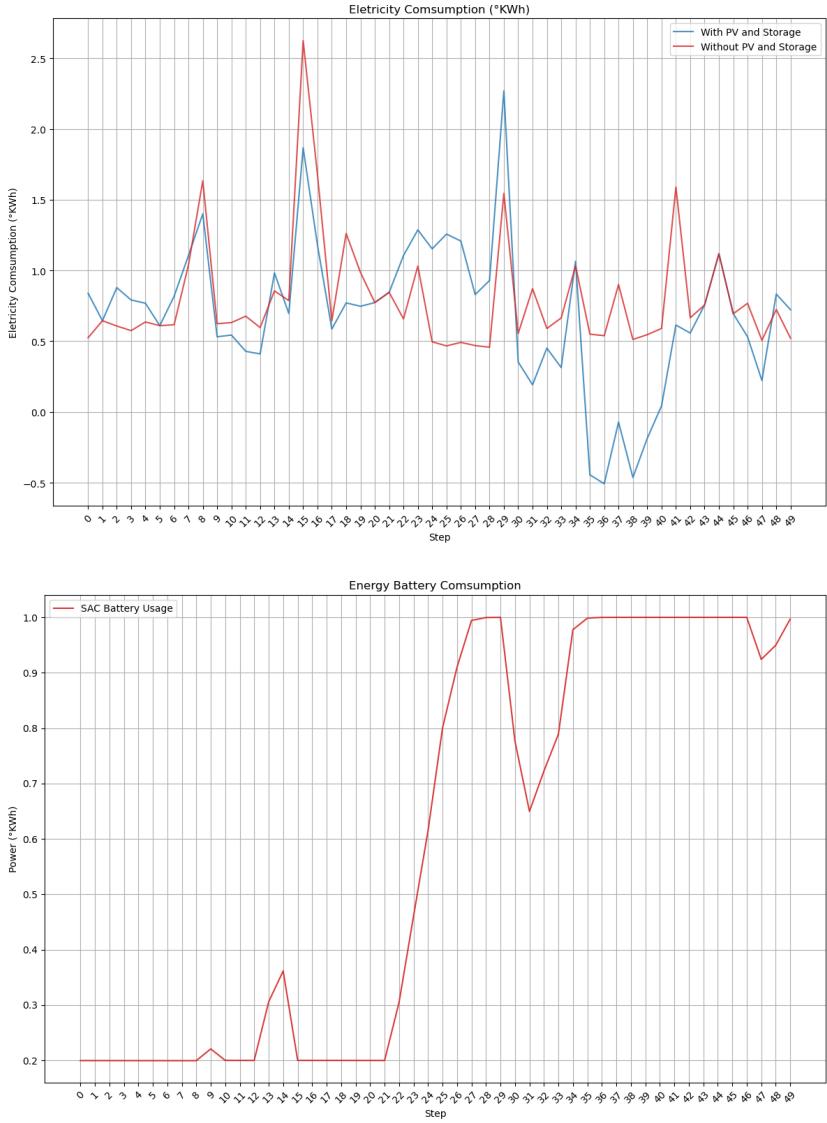
Purtroppo, non essendo considerati i non-shiftable load nel reward, durante i blackout vengono totalmente spenti causando così il loro completo inutilizzo. Risulta importante perciò trovare un modo per considerarli.



Non-shiftable loads

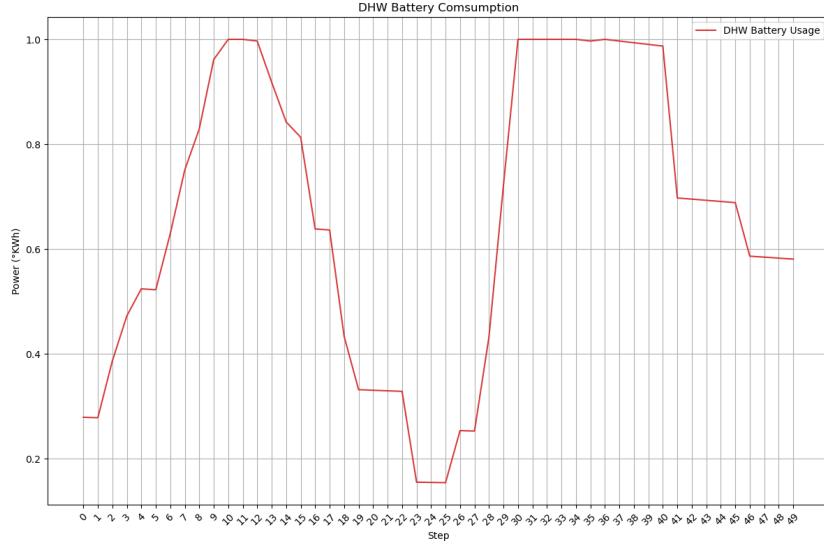
Considerazioni approfondite su 48 ore Tra i reward valutati, questo rappresenta il miglior compromesso tra comfort e consumi. Per questo motivo, verranno selezionate 48 ore della simulazione per consentire un'analisi giornaliera approfondita e valutare concretamente l'efficienza di SAC con questo reward.

Analizzando l'andamento dei consumi, è possibile individuare periodi di inefficienza semplicemente osservando il grafico. Sebbene non sia immediato comprenderne le conseguenze, si nota che in alcuni punti il grafico blu (che rappresenta i consumi considerando l'autosufficienza) supera quello rosso (che indica i consumi totali dei dispositivi). Questo suggerisce che l'agente ha accumulato energia anche in momenti in cui la casa non era autosufficiente, causando un aumento dei picchi di prelievo dalla rete. Si evidenzia inoltre che questa inefficienza avviene durante la notte, dove sappiamo che non avviene alcun tipo di produzione di energia.



Si osserva come l'accumulo eccessivo nella batteria abbia contribuito all'inefficienza rilevata. In particolare, si nota un comportamento poco ottimale dell'agente: durante un momento di produzione solare (step 12), la batteria viene prima caricata e poi nuovamente svuotata. Questo indica che l'agente non è ancora sufficientemente addestrato a gestire in modo ottimale l'accumulo energetico, tendendo a esagerare e generando consumi evitabili. Infatti percependo produzione è spronato ad accumulare energia, ma non moderandosi tende a superare la produzione e portare ad un consumo delle reti. Ciò inoltre contribuisce a non permettere di apprendere il modo ottimale l'utilizzo. Inoltre, l'accumulo inizia sistematicamente uno step dopo rispetto alla generazione solare. Questo suggerisce che l'agente non ha capacità predittive, e che

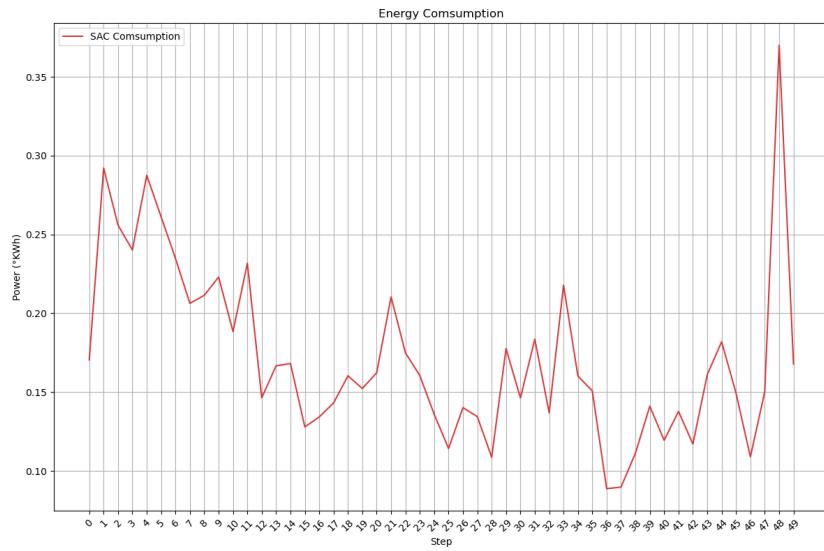
potrebbe trarre beneficio dall'integrazione di un modello di previsione. Si segnala infine un rilascio consistente allo step 47 che, sebbene eccessivo, risulta efficace per ridurre i consumi notturni. In generale, tuttavia, l'accumulo appare irregolare e poco coerente con le esigenze energetiche effettive.

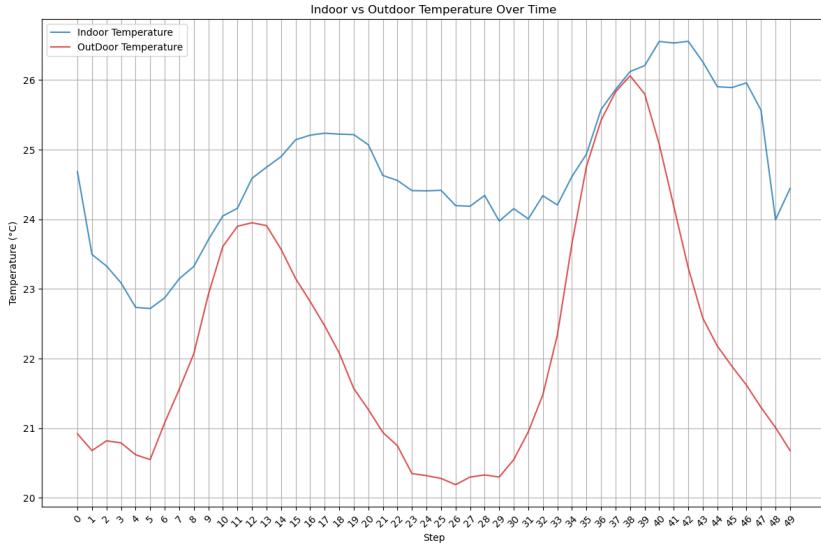


L'accumulatore DHW mostra un comportamento più regolare rispetto alla batteria elettrica, ma ciò non si traduce necessariamente in maggiore efficienza. Si osserva, ad esempio, che durante il primo giorno l'accumulo avviene principalmente nelle ore diurne, riducendo così il margine disponibile per sfruttare la produzione solare tramite la batteria elettrica. In questo scenario, sarebbe stato più opportuno accumulare prima energia nella batteria e successivamente soddisfare i bisogni del DHW. Inoltre, si nota un rilascio serale da parte dell'accumulatore DHW, seguito da un nuovo accumulo intorno allo step 26, corrispondente alle 2 di notte. Come suggerito anche dal vincitore della CityLearn Challenge, una strategia ottimale prevederebbe l'accumulo del DHW durante le ore notturne, in modo da alleggerire il carico sulla rete, oppure solamente nei casi in cui la batteria elettrica sia già completamente carica. L'attuale strategia, invece, tende ad aumentare i picchi di carico e i consumi prelevati dalla rete. Adottando l'approccio proposto dal vincitore della challenge, si potrebbero migliorare queste criticità e ottenere un comportamento energetico più efficiente. Inoltre c'è da considerare che durante una fase di blackout il suo comportamento rimane invariato, rubando disponibilità energia a dispositivi più essenziali. Con l'attuale algoritmo SAC ciò non risulta ancora risolvibile.

Per quanto riguarda la gestione della temperatura, il grafico della temperatura interna appare relativamente stabile intorno al setpoint, rimanendo

sempre all'interno della banda consentita. Le variazioni osservate sono dovute esclusivamente ai cambiamenti del setpoint, che causano le oscillazioni nel grafico. È evidente come queste variazioni di setpoint comportino bruschi cambiamenti nei consumi dei dispositivi di raffreddamento (come allo step 47), generando picchi e fenomeni di ramping. Un approccio più graduale nell'adattamento del setpoint potrebbe comportare un leggero aumento della discomfort proportion, ma permetterebbe una significativa riduzione di picchi e ramping nei consumi. Va comunque sottolineato che si tratta di una scelta soggettiva, legata alle preferenze dell'utente. Si riscontrano tuttavia alcune imprecisioni nel comportamento dell'agente. Ad esempio, dallo step 27 si nota un incremento graduale del setpoint, mentre i consumi mostrano forti oscillazioni, invece di seguire in modo coerente l'andamento del setpoint stesso, comportando consumi aggiuntivi evitabili. Questo comportamento si riscontra anche in altri momenti della simulazione.





Nel complesso, sebbene l'agente mostri prestazioni relativamente buone, sono presenti alcune imprecisioni che potrebbero essere facilmente corrette. Va inoltre considerato che i risultati positivi ottenuti potrebbero essere il frutto di un overfitting sul contesto specifico della simulazione, il che rischierebbe di compromettere l'efficacia dell'agente in scenari reali o meno controllati.

4.4 Comfort and Consumption Reduction Reward

Il Comfort and Consumption Reduction Reward è una funzione di reward ispirata al modello ideato dal primo classificato della CityLearn Post Challenge 2023.

Formula

$$R = -0.3 \cdot |\Delta T| - 0.1 \cdot s - 0.15 \cdot |r|$$

- $\Delta T = \begin{cases} 3 \cdot (T_{\text{indoor}} - T_{\text{setpoint}}), & \text{se } T_{\text{indoor}} < T_{\text{setpoint}} \\ T_{\text{indoor}} - T_{\text{setpoint}}, & \text{altrimenti} \end{cases}$

E' una versione semplificata del comfort reward.

- $r = |C_t - C_{t-1}|$

E' il ramping.

$$\bullet \quad s = \begin{cases} L_{\text{non-shiftable}} - C_t, & \text{se power_outage} = 1 \\ 0, & \text{altrimenti} \end{cases}$$

Gestisce i non-shiftable loads.

Statistiche generali È interessante notare come l'approccio cambi: infatti, il consumo non viene più calcolato direttamente, ma premiato indirettamente attraverso il ramping. Inoltre, vi è un'implementazione che coinvolge i shiftable loads durante un periodo di blackout.

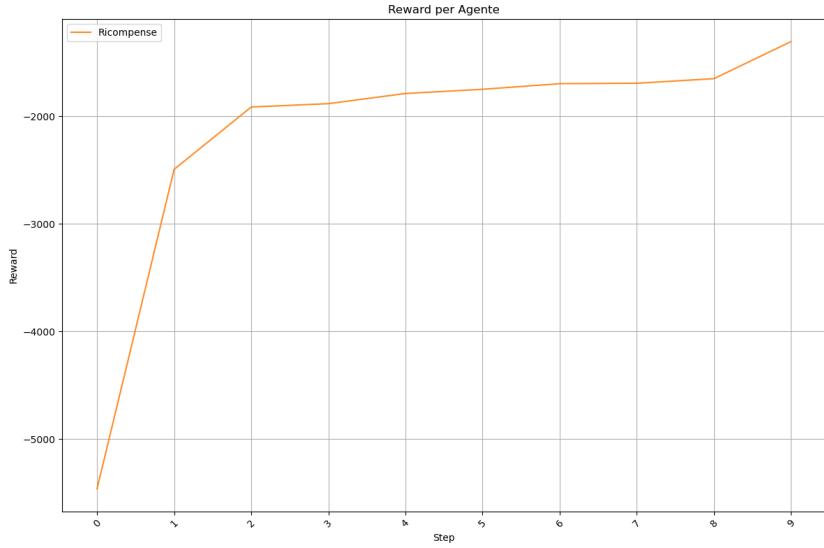
Si può osservare come, a differenza di prima, vengano preferiti i consumi al discomfort. Tuttavia, questo risulta essere un effetto fittizio, causato dal peso assegnato nel reward.

In questa analisi, ci concentriamo principalmente sull'effetto che tale reward ha sui shiftable loads, e come utilizzare il ramping può influenzare la rete.

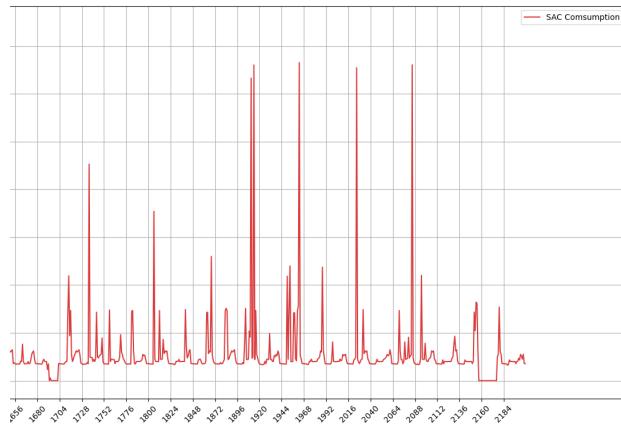
È da evidenziare che il discomfort viene calcolato in modo più basilare rispetto a prima, il che potrebbe ridurre la sensibilità del reward alle variazioni più sottili nelle condizioni di comfort, portandoci a considerare il reward precedente per utilizzare il discomfort.

	name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
cost_function								
all_time_peak_average		NaN	NaN	NaN	NaN	NaN	NaN	0.850
annual_normalized_unserved_energy_total	0.020	0.012	0.015	0.018	0.015	0.019	0.017	
carbon_emissions_total	0.767	0.942	0.853	0.778	0.900	0.711	0.825	
daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.853	
discomfort_proportion	0.381	0.109	0.108	0.099	0.516	0.351	0.261	
electricity_consumption_total	0.767	0.943	0.853	0.780	0.900	0.710	0.825	
one_minus_thermal_resilience_proportion	0.606	0.263	0.905	0.515	0.870	0.818	0.663	
ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	0.946	

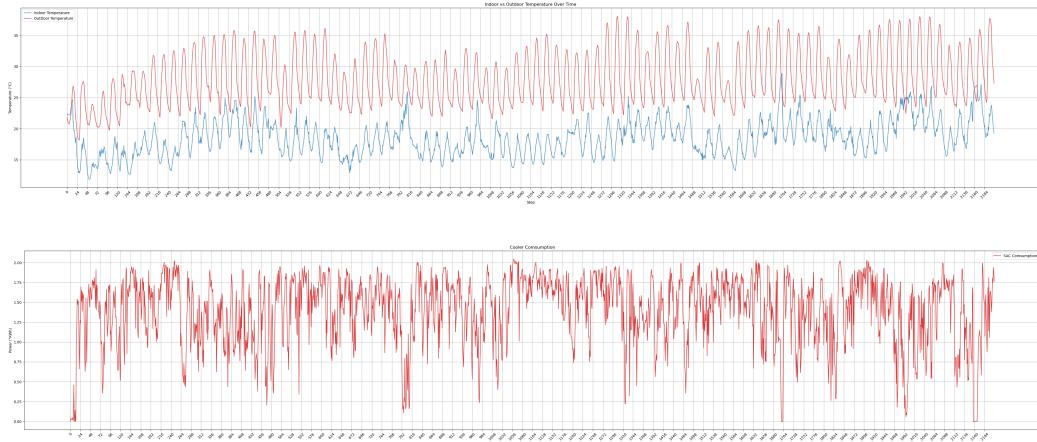
Il reward evidenzia un andamento più stabile rispetto al reward base, suggerendo una maggiore coerenza nel comportamento dell'agente.



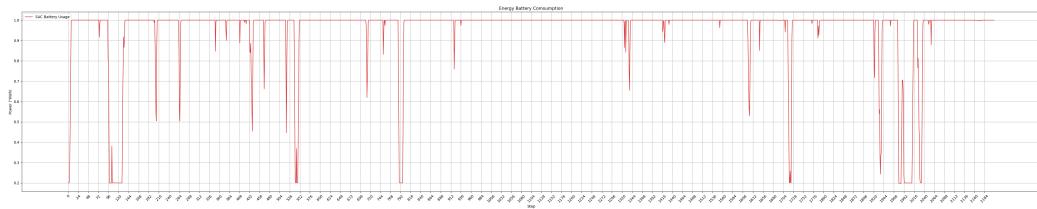
Comportamenti dei non-shiftable loads Benché siano considerati nel reward, il risultato appare poco efficiente. Probabilmente a causa della rarità delle interruzioni di corrente, che non possono essere rappresentati nella generalità che ricerca la policy. Perciò risulta complicato poter gestire questa problematica attraverso i reward, in quanto sono situazioni poco influenti nella policy generale.



il dislivello alzando l'asticella totale. Ciò, se non regolato dal discomfort, non solo comporta un incremento dei consumi, ma anche un aumento del discomfort, dovuto a un eccessivo impiego dei dispositivi.



L'utilizzo delle batterie rimane totalmente inefficiente.



Inoltre, il guadagno derivante dalla riduzione del ramping risulta minimo rispetto agli obiettivi del problema. Risulta quindi più efficiente incentivare l'agente a ridurre i consumi generali.

name	Building_1	Building_2	Building_3	Building_4	Building_5	Building_6	District
cost_function							
all_time_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	1.047
annual_normalized_unserved_energy_total	0.005	0.006	0.008	0.003	0.005	0.008	0.006
carbon_emissions_total	1.078	1.217	1.086	1.057	0.876	1.036	1.058
daily_peak_average	NaN	NaN	NaN	NaN	NaN	NaN	0.956
discomfort_proportion	0.631	0.438	0.366	0.613	0.624	0.494	0.528
electricity_consumption_total	1.087	1.207	1.084	1.066	0.882	1.029	1.059
one_minus_thermal_resilience_proportion	0.636	0.474	0.905	0.788	0.826	0.939	0.761
ramping_average	NaN	NaN	NaN	NaN	NaN	NaN	1.115

Riferimenti

- Migliore CityLearn Post Challenge 2023: <https://gitlab.aicrowd.com/dhchen/citylearn-2023-starter-kit/-/issues/9>
- AICrowd CityLearn Control Track: <https://www.aicrowd.com/challenges/neurips-2023-citylearn-challenge/problems/control-track-citylearn-challenge>