B : Charge
D_s : Self Discharge Rate
C : Charge Rate
D : Discharge Rate
ν : Roundtrip Efficiency
$T_t = T_{t-1} + \frac{1}{r_a}(T_{a,t} - T_{t-1}) + \frac{1}{r_b}(T_{b,t} - T_{t-1}) + \frac{1}{r_h} L_{TCL} a_{tcl,t} + q$
$T_{b,t} = T_{b,t-1} + \frac{1}{r_b}(T_{t-1} - T_{b,t-1})$
T : Indoor Temperature
T_a : Outdoor Temperature
T_b : Building Temperature
L_{TCL} : Nominal Heating Power
q : Unintentional Heat Gain
$a_{ess} \in [-1, 1]$
$a_{fdr} \in [-1, 1]^H$
$a_{tcl} \in [-1, 1]$
$a_* = 0$
a_t
s_t, a_t, r_t, s_{t+1}
s_t, s_{t+1}
$\pi(s_t), \pi(s_{t+1}), H(a, s)$
$min(Q_1, Q_2)$
$Q(s, a)$
$a_* = Uniform(-1, 1)$

Table 1: Accumulated rewards split by origin per setting. Evaluation was done after every epoch and the best result is shown.

Experiment	Algorithm	ESS	FDR	TCL	Discomfort	Total	Saved
Initial Setting	Idle	0.00	-1207.53	-5099.46	-6144.05	15604.1	
	Threshold	523.19	-1012.63	-5033.79	-1159.65	21372.26	
	PPO	-23.78	-1263.76	-5498.38	-1951.5	19317.72	
	SAC	85.66	-1150.97	-5421.83	-3207.93	18360.07	
Experiment	Algorithm	ESS	FDR	TCL	Discomfort	Total	Saved
Stacked Carbon Intensity	Idle	0.00	-1207.53	-5099.46	-6144.05	15604.1	
	Threshold	523.19	-1012.63	-5033.79	-1159.65	21372.26	
	PPO	0.00	-1225.71	-5696.45	-1354.04	19778.94	
	SAC	124.05	-1190.63	-5438.74	-1464.48	20085.34	
Experiment	Algorithm	ESS	FDR	TCL	Discomfort	Total	Saved
Eased Exploration	Idle	0.00	-1207.53	-5099.46	-6144.05	15604.1	
	Threshold	523.19	-1012.90	-5033.79	-1159.65	21371.99	
	PPO	-1	-1	-1	-1		
	SAC	657.56	-879.83	-13175.57	-4388.83	10268.47	
Experiment	Algorithm	ESS	FDR	TCL	Discomfort	Total	
Terminal Reward	Idle	0.00	-1207.53	-5099.46	-6144.05	15604.1	
	Threshold	523.19	-1012.63	-5033.79	-1159.65	21372.26	
	PPO	-16.92	-1244.29	-5995.35	-2038.94	18759.64	
	SAC	-9.08	-1122.2	-8500.46	-4916.73	13506.67	

Table 2: Accumulated rewards split by origin per setting. Evaluation was done after every epoch and the best result is shown.

	Idle	Random	Threshold	PPO	SAC
ESS	0.00	-66.23	**523.19**	-23.78	85.66
FDR	-1207.53	-1215.82	**-1012.63**	-1263.76	-1150.97
TCL	-5099.46	-15043.22	**-5033.79**	-5498.38	-5421.83
	-6144.05	-8212.51	**-1159.65**	-1951.5	-3207.93
Saved Carbon	Base	-10018.28	**783.76**	-478.93	-180.15
	Idle	Random	Threshold	PPO	SAC
ESS	0.00	-66.23	**523.19**	0.00	124.05
FDR	-1207.53	-1215.82	**-1012.63**	-1225.71	-1190.63
TCL	-5099.46	-15043.22	**-5033.79**	-5696.45	-5438.74
	-6144.05	-8212.51	**-1159.65**	-1354.04	-1464.48
Saved Carbon	Base	-10018.28	**783.76**	-615.17	-198.33
	Idle	Threshold	PPO	SAC	
ESS	0.00	523.19	233.85	**657.56**	
FDR	-1207.53	-1012.90	-1085.54	**-879.83**	
TCL	-5099.46	**-5033.79**	-5644.43	-13175.57	
	-6144.05	**-1159.65**	-2205.01	-4388.83	
Saved Carbon	Base	**783.49**	-189.13	-7090.85	
	Idle	Random	Threshold	PPO	SAC
ESS	0.00	-66.23	**523.19**	-16.92	-9.08
FDR	-1207.53	-1215.82	**-1012.63**	-1244.29	-1122.2
TCL	-5099.46	-15043.22	**-5033.79**	-5995.35	-8500.46
	-6144.05	-8212.51	**-1159.65**	-2038.94	-4916.73
Saved Carbon	Base	-10018.28	**783.76**	-949.57	-3324.75
	Ensemble	Threshold	PPO	SAC	
ESS	657.56				
FDR	-879.83				
TCL	-5033.79				
	-1159.65				
Saved Carbon	1050.37				