



**Politecnico  
di Torino**

**Politecnico di Torino**

Corso di Laurea Magistrale  
in Ingegneria Edile  
Indirizzo Progetto e gestione  
A.a. 2020/2021  
Sessione di Laurea Luglio 2021

# **Reinforcement Learning control of electrochromic windows for enhancing energy efficiency and visual comfort**

Investigation of evolving control strategy for an adaptive  
transparent envelope component addressing conflicting goals

**Relatore:**

*Prof. Alfonso Capozzoli*

**Candidato:**

*Sandri Luca*

**Correlatori:**

*Ing. Silvio Brandi*

*Ing. Fabio Favoino*

*“Gli uomini passano, le idee restano.  
Restano le loro tensioni morali  
e continueranno a camminare sulle gambe di altri uomini”.*  
~ Giovanni Falcone

*“Men pass, ideas remain.  
Their moral tensions remain  
and they will continue to walk on the legs of other men”*  
~ Giovanni Falcone

## Abstract

Climate Adaptive Building Shells are envelope elements that have dynamic features instead of the traditional static ones. This target practice is relevant because of different environmental conditions affecting operative building loads that vary by locations, seasonal weather and future climate evolution, and because of the urgent need for decarbonisation. The large amount of time spent in indoor environment in modern days also requires attention for comfort, affecting people wellbeing. Smart windows can affect both thermal loads and lighting conditions, but active glazing technologies require a control strategy to be deployed. The control methods for building systems and smart windows in particular are Rule-Based or Model Predictive-based. The former mainly rely on a priori-knowledge and can only indirectly adapt on-field through rearranging boundaries of its condition statements. However, these adaptable ranges maintain fixed associated actions. The second can develop a sequence of optimized steps based on disturbances, even predicting them. This ability however comes with the need for a model, with possible drawbacks concerning computational time, calibration precision and development work. Some researchers have applied Reinforcement Learning in the building sector but mainly for service equipment.

The present work investigates the adaptability and suitability of Tabular Q-learning Reinforcement Learning to develop an evolving control policy for a smart window in order to address contrasting objectives. In particular, an electrochromic window capable to switch between 4 different states has been chosen as smart glazing technology. Different RL agents have been tested, introducing various designs of state space based on different combination of most implemented variables found in literature. In all tested configurations, incident solar radiation on window was provided to the controller as input variable. The objective of the controller was to minimize the energy consumption for key building services (namely, heating, cooling and lighting) while ensuring the visual comfort. Improvements in both fields can have positive economic effects related to reduced energy utilization in the operational phase of buildings and enhanced user wellbeing respectively. In order to evaluate the goodness of developed controllers, a Rule-Based Controller has been selected as baseline.

Results suggested that electrochromic windows have a much higher potential effect on lighting service than on heating. The most energy efficient design embedded input variables

representing the cumulative measures of external disturbances for the following 4 hours, that can be considered analogous to weather forecasts. This approach yielded to save an additional 3% of total energy compared to the reference RBC. When both energy consumption and visual comfort were embedded in the reward function, the most relying secondary input variables were occupancy and external temperature (when the measurement was categorized based on 3 discretization levels). The proposed controller was able to achieve between 64,7 and 67,5 % of Useful Daylight Index compared to 54,8% of the RBC strategy and still saving more energy, up to 1,76% less than RBC. Hence, Tabular Q-learning resulted as a good alternative to adaptive windows control strategies that want to address conflicting goals in the built environment.

# Summary

Abstract .....	i
List of tables .....	v
List of figures .....	vii
1. Introduction .....	1
1.1. Decarbonisation and the impacts of buildings.....	1
1.2. Adaptive Envelopes.....	4
1.2.1. Concept of the envelope .....	4
1.2.2. Static envelopes .....	4
1.2.3. Definition of adaptive envelopes.....	5
1.2.4. Potentials of adaptive envelopes.....	5
1.2.5. The transparent envelope.....	6
1.2.6. Smart windows .....	9
1.3. Control strategies.....	16
1.3.1. Building Automation Control Systems.....	16
1.3.2. Adaptive envelope components control .....	20
1.3.3. EC control.....	23
1.3.4. Reinforcement Learning.....	35
2. Methodology .....	42
2.1. Virtual environment.....	42
2.2. Indicators and benchmarking.....	45
2.2.1. Reference RBC and dummy controllers.....	47
2.3. Control problem formulation.....	48
2.3.1. Shared characteristics of investigated control method .....	48
2.3.2. Reward and Q-learning.....	49
2.3.3. Q-table: input variables and states.....	51
2.3.4. Action selection .....	53
3. Results and Discussion.....	57
3.1. Defining the control strategy core .....	60
3.2. Exploring additional input variables.....	65
3.3. Exploration of cumulative future disturbances as state variables.....	67
3.4. Combination of more additional parameters .....	73
3.5. Exploration of Q-table design for reward based on UDI only.....	75

3.6. Exploration of weight for UDI and input variables.....	78
4. Conclusion .....	83
Appendices.....	87
Bibliography .....	88

# List of tables

<b>TABLE 1.1</b> OPTIMAL VALUE OF DIFFERENT CONTROL PARAMETER FOR ECG WINDOW RBC BY LOCATIONS. FROM [55] .....	24
<b>TABLE 1.2</b> THERMO-OPTICAL PROPERTIES OF PVC-GLAZING TESTED BY [25] $T_{vis}$ : VISIBLE TRANSMISSION COEFFICIENT; G-VALUE: TOTAL SOLAR HEAT GAIN COEFFICIENT (-); $T_N$ : TRANSMISSIVITY (-).....	25
<b>TABLE 1.3</b> PROPERTIES OF THE SIMULATED GLAZING. FROM [58] .....	27
<b>TABLE 1.4</b> AVERAGE PERFORMANCE DIFFERENCE OF CONTROLLED EC GLAZING AGAINST STATIC ALTERNATIVE. VALUES AGGREGATED FROM [54] .....	28
<b>TABLE 1.5</b> MAIN CHARACTERISTICS OF REPORTED CASE STUDIES DEALING WITH SMART WINDOWS CONTROL (PT 1).....	33
<b>TABLE 1.6</b> MAIN CHARACTERISTICS OF REPORTED CASE STUDIES DEALING WITH SMART WINDOWS CONTROL (PT 2).....	34
<b>TABLE 1.7</b> EXAMPLES OF Q-TABLE IN A PARTICULAR TIME-STEP FROM THE PRESENT STUDY. STATES ARE HERE EXPRESSED AS A TUPLE WITH THE VALUE OF A 4-LEVEL INPUT AND ANOTHER BINARY INPUT. ....	37
<b>TABLE 2.1</b> OVERALL CHARACTERISTICS OF THE SIMULATED CONSTRUCTION IN COMPLIANCE WITH LOCAL REGULATION.....	43
<b>TABLE 2.2</b> SMART GLAZING PROPERTIES OF THE FOUR SELECTABLE STATES.....	44
<b>TABLE 2.3</b> DAYLIGHT GLARE COMFORT CLASSES AND RELATIVE DGP THRESHOLDS [56].....	47
<b>TABLE 2.4</b> CONDITIONS FOR PASSIVE RBC ACTION SELECTION RELYING ON INCIDENT SOLAR RADIATION ON THE FAÇADE .....	48
<b>TABLE 2.5</b> IDENTIFICATION OF DUMMY CONTROLLERS, HENCE AGENT KEEPING THE SAME STATE WINDOW THROUGH THE WHOLE YEAR.....	48
<b>TABLE 3.1</b> PERFORMANCES OF RBC AND DUMMY CONTROLLERS. ....	58
<b>TABLE 3.2</b> ENERGY PERFORMANCES OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT DISCOUNT FACTORS.....	60
<b>TABLE 3.3</b> IDENTIFICATION OF THE FOUR TESTS ABOUT TOTAL INCIDENT SOLAR RADIATION DISCRETIZATION. ....	62
<b>TABLE 3.4</b> PERFORMANCES OF RBC AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT TOTAL INCIDENT SOLAR RADIATION DISCRETIZATION.....	63
<b>TABLE 3.5</b> PERFORMANCES OF RBC AND THREE REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT EXTERNAL TEMPERATURE DISCRETIZATIONS.....	65
<b>TABLE 3.6</b> IDENTIFICATION OF THE FOUR TESTS ABOUT TOTAL INCIDENT SOLAR RADIATION FORECAST DISCRETIZATION.....	68
<b>TABLE 3.7</b> PERFORMANCES OF RBC AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT CUMULATIVE INCIDENT SOLAR RADIATION DISCRETIZATION.....	70
<b>TABLE 3.8</b> IDENTIFICATION OF THE FOUR TESTS ABOUT TOTAL INCIDENT SOLAR RADIATION FORECAST DISCRETIZATION.....	71
<b>TABLE 3.9</b> PERFORMANCES OF RBC AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT CUMULATIVE EXTERNAL TEMPERATURE DISCRETIZATION.....	72
<b>TABLE 3.10</b> PERFORMANCES AND IMPROVEMENTS COMPARED TO RBC OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT STATES SPACE BASED ON THREE INPUT VARIABLE EACH. ....	73

<b>TABLE 3.11</b> LEGEND OF STATES SPACE DESIGN FOR Q-TABLE OF IR-F+OCC.....	75
<b>TABLE 3.12</b> PERFORMANCES OF RBC AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT SECONDARY INPUT VARIABLE AND REWARD FUNCTION BASED ON UDI ONLY.....	77
<b>TABLE 3.13</b> PERFORMANCES OF RBC AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT $w_{UDI}$ AND OCCUPANCY OR EXTERNAL TEMPERATURE AS ADDITIONAL INPUT VARIABLE.....	81

# List of figures

<b>FIGURE 1.1</b> WARMING STRIPES ILLUSTRATING THE ANNUAL GLOBAL TEMPERATURES FROM 1850-2017, FROM [1]. THE COLOUR SCALE COVERS 1.35°C OF CHANGE IN GLOBAL TEMPERATURES.....	1
<b>FIGURE 1.2</b> THE WEIGHT OF BUILDINGS AND CONSTRUCTION SECTOR ON GLOBAL EMISSIONS AND FINAL ENERGY IN 2019, FROM [5].....	2
<b>FIGURE 1.3</b> RESIDENTIAL BUILDINGS IN EU CLASSIFIED BY CONSTRUCTION YEAR (2014) FROM [7].....	3
<b>FIGURE 1.4</b> PROGRESSION OF BUILDING ENVELOPES REFURBISHMENT FROM OLD STOCK TO FUTURE EFFICIENCY, FROM [6].....	3
<b>FIGURE 1.5</b> LIMITS OF THERMAL TRANSMITTANCE (U-VALUE) UNDERGOING ENERGY REFURBISHMENTS BETWEEN 2019 AND 2021, FROM ITALIAN DM MINIMUM REQUIREMENTS 26 JUNE 2015. IMAGE FROM THE COURSE "ENERGY AUDIT AND CERTIFICATION OF BUILDINGS", TAUGHT AT POLITECNICO DI TORINO IN 2018/19.....	7
<b>FIGURE 1.6</b> SOME EXAMPLES OF DIFFERENT GLAZING TECHNOLOGIES. FROM [24].....	10
<b>FIGURE 1.7</b> COMPOSITION OF AN ELECTROCHROMIC DEVICE. IMAGE FROM [29].....	12
<b>FIGURE 1.8</b> EXAMPLES OF GASOCHROMIC WINDOW USING ARGON TO DILUTE H <sub>2</sub> AND O <sub>2</sub> . FROM [24].....	13
<b>FIGURE 1.9</b> REPRESENTATION OF SUSPENDED PARTICLES STATES USED IN A GLAZING. FROM [24].....	14
<b>FIGURE 1.10</b> PERFORMANCE OF OPTIMAL WINDOWS AND ACTUAL GLAZING TECHNOLOGIES. FROM [24].....	15
<b>FIGURE 1.11</b> GENERAL REPRESENTATION OF A RECEDING HORIZON LOGIC, WITH THE DIFFERENT VARIABLES COMPUTED AT TIME-STEP K (ABOVE) AND K+1 (BELOW). IMAGE FROM [41].....	17
<b>FIGURE 1.12</b> CROSS-CASE COMPARISON OF IMPLEMENTED FAÇADE FUNCTIONS. ADAPTIVE: ADAPTS INDEPENDENTLY DUE TO ITS AUTOMATED CONTROL; FLEXIBLE: ADAPTABLE ONLY MANUALLY BY THE USER; FULFILLED: FUNCTION REACHED WITH STATIC APPLICATION. IMAGE FROM [39].....	22
<b>FIGURE 1.13</b> CLIMATE ZONE DISAGGREGATION OF THE U.S. REPORTING SIMULATED CLIMATE LOCATIONS. FROM [55].....	25
<b>FIGURE 1.14</b> SCHEME OF RBC DEVELOPED FROM USERS' INTERACTIONS ANALYSIS BY [49].....	29
<b>FIGURE 1.15</b> (LEFT) BLIND OPERATION DURING YEAR: (FROM TOP) EC WINDOWS, REFERENCE WINDOW. BLUE/GREEN SCALE REFERS TO SHADING OVER UPPER PANE ONLY. RED/YELLOW SCALE RELATES TO TIMES FOR WHICH SHADING OVER BOTH PANES. (RIGHT) EC PANE TRANSMITTANCE DURING YEAR: (FROM TOP) UPPER PANE; LOWER PANE. IMAGE FROM [59].....	30
<b>FIGURE 1.16</b> AGENT AND ENVIRONMENT INTERACTION.....	35
<b>FIGURE 1.17</b> SAVINGS PERFORMANCE BY THREE DIFFERENT CONTROLLERS SIMULATED CHRONOLOGICALLY (ABOVE) AND BACKWARDS (BELOW). FIGURE BY [44].....	36
<b>FIGURE 1.18</b> STRUCTURAL SCHEME OF THE NEURAL NETWORK USED FOR DEEP Q-LEARNING FROM [62].....	38
<b>FIGURE 1.19</b> E-GREEDY (WHOLE FIGURE) AND GREEDY (FRAMED PORTION) ACTION SELECTIONS.....	39
<b>FIGURE 2.1</b> OFFICE TEST ROOM MODEL (MEASUREMENTS IN CM) FROM [25].....	43
<b>FIGURE 2.2</b> HEAT MAPS OF INCIDENT SOLAR RADIATION AND EXTERNAL TEMPERATURE.....	44
<b>FIGURE 2.3</b> QUALITATIVE REPRESENTATION OF UDI BANDS.....	46
<b>FIGURE 2.4</b> SCALED UDI (LEFT) AND SE (RIGHT) SCORES FOR EACH TIME-STEP (X-AXIS) OF THE SAME RL CONTROL TEST.....	50
<b>FIGURE 2.5</b> EVOLUTION OF E THROUGH THE SIMULATION, BY TIME-STEP.....	54

<b>FIGURE 3.1</b> CARPET PLOT OF RBC REFERENCE ACTIONS .....	58
<b>FIGURE 3.2</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF DUMMY CONTROLLERS WITH DIFFERENT CONSTANT ACTION.....	59
<b>FIGURE 3.3</b> ENERGY PERFORMANCES OF SEVEN REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT DISCOUNT FACTORS. ....	61
<b>FIGURE 3.4</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT DISCOUNT FACTORS. ....	62
<b>FIGURE 3.5</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT TOTAL INCIDENT SOLAR RADIATION DISCRETIZATION, COMPARED TO RBC. ....	64
<b>FIGURE 3.6</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT TOTAL INCIDENT SOLAR RADIATION DISCRETIZATIONS.....	64
<b>FIGURE 3.7</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF SEVEN REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT DISCOUNT FACTORS COMPARED TO RBC TOTAL SITE ENERGY (DOTTED LINE).....	66
<b>FIGURE 3.8</b> CARPET PLOTS OF ACTIONS SELECTED THROUGH RL FOR DIFFERENT EXTERNAL TEMPERATURE DISCRETIZATION. ....	67
<b>FIGURE 3.9</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT CUMULATIVE TOTAL INCIDENT SOLAR RADIATION DISCRETIZATION, COMPARED TO RBC.....	68
<b>FIGURE 3.10</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT CUMULATIVE INCIDENT SOLAR RADIATION DISCRETIZATION. ....	69
<b>FIGURE 3.11</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT CUMULATIVE EXTERNAL TEMPERATURE DISCRETIZATION, COMPARED TO RBC .....	71
<b>FIGURE 3.12</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT CUMULATIVE EXTERNAL TEMPERATURE DISCRETIZATION. ....	72
<b>FIGURE 3.13</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS IR-F+ET (ON THE LEFT) AND IR-F+OCC (ON THE RIGHT).....	74
<b>FIGURE 3.14</b> Q-TABLE OF IR-F+OCC HEAT MAP.....	74
<b>FIGURE 3.15</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT SECONDARY INPUT VARIABLE AND REWARD FUNCTION BASED ON UDI ONLY, COMPARED TO RBC .....	76
<b>FIGURE 3.16</b> CARPET PLOTS OF ACTIONS SELECTED THROUGH RL FOR DIFFERENT DISCOUNT FACTORS. ....	77
<b>FIGURE 3.17</b> ENERGY PERFORMANCES DIFFERENCE BETWEEN RBC (BASELINE) AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT $W_{UDI}$ AND OCCUPANCY AS ADDITIONAL INPUT VARIABLE... 79	79
<b>FIGURE 3.18</b> ENERGY PERFORMANCES DIFFERENCE BETWEEN RBC (BASELINE) AND REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT $W_{UDI}$ AND EXTERNAL TEMPERATURE AS ADDITIONAL INPUT VARIABLE. ....	79
<b>FIGURE 3.19</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT $W_{UDI}$ AND OCCUPANCY OR EXTERNAL TEMPERATURE AS ADDITIONAL INPUT VARIABLE. ....	80
<b>FIGURE 3.20</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT $W_{UDI}$ AND OCCUPANCY AS ADDITIONAL INPUT VARIABLE. ....	81

<b>FIGURE 3.21</b> CARPET PLOTS OF ACTIONS SELECTED BY REINFORCEMENT LEARNING CONTROLLERS WITH DIFFERENT $W_{UDI}$ AND EXTERNAL TEMPERATURE AS ADDITIONAL INPUT VARIABLE. ....	82
<b>FIGURE 4.1</b> ENERGY PERFORMANCES AND IMPROVEMENTS BY SERVICE OF BEST REINFORCEMENT LEARNING CONTROLLERS DEVELOPED IN THE PRESENT STUDY .....	86

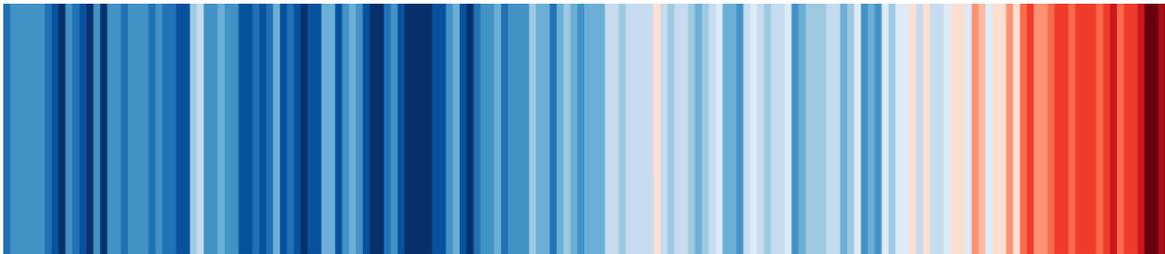
## Acronyms and symbols

a	action	PC	Photo-Chromic
AI	artificial intelligence	PCM	Phase Change Materials
ANN	Artificial Neural Network(s)	PECCs	Photoelectrochromic cells
BACS	Building Automation and Control Systems	PMV	Predicted Medium Vote
BEM	Building Energy Model	PPD	Predicted Percentage of Dissatisfied
C	Cooling	PV	Photovoltaics
CABS	Climate Adaptive Building Shells	Q	state/action value
CDD	Cooling degree day	R	Reward
COP	Coefficient Of Performance	RBC	Rule-based control
DGP	Discomfort Glare Probability	RH	Relative Humidity
DGU	Double Glass Unit	RHC	Receding Horizon Control
E, pi	Estimation based on policy pi	RL	Reinforcement Learning
EC	Electrochromic	s	state
Ep	Primary energy	SE	Site Energy
ET	external temperature	SHGC	Solar Heat Gain Coefficient
ETF	external temperature forecast	SPD	Suspended Particle
GC	Gasochromic	SSN	season
H	Heating	t	timestep
HDD	Heating degree day	TC	Thermo-Chromic
HVAC	Heating, Ventilation, Air-Conditioning	TGU	Triple Glass Unit
IR	overall incident solar radiation	TT	Thermo-Tropic
IRF	overall incident solar radiation forecast	Tv	visible transmittance
IT	internal temperature	UDI	Useful Daylight Illuminance index
L	Lighting	U-value	Overall heat transfer coefficient value
LCD	Liquid Crystal	WPI	work-plane illuminance
MPC	Model predictive control	WWR	Window-to-Wall Ratio
NIR	Near-infrared	$\alpha$	learning rate
OCC	Occupancy	$\gamma$	discount factor
		$\epsilon$	threshold for $\epsilon$ -greedy action selection

# 1. Introduction

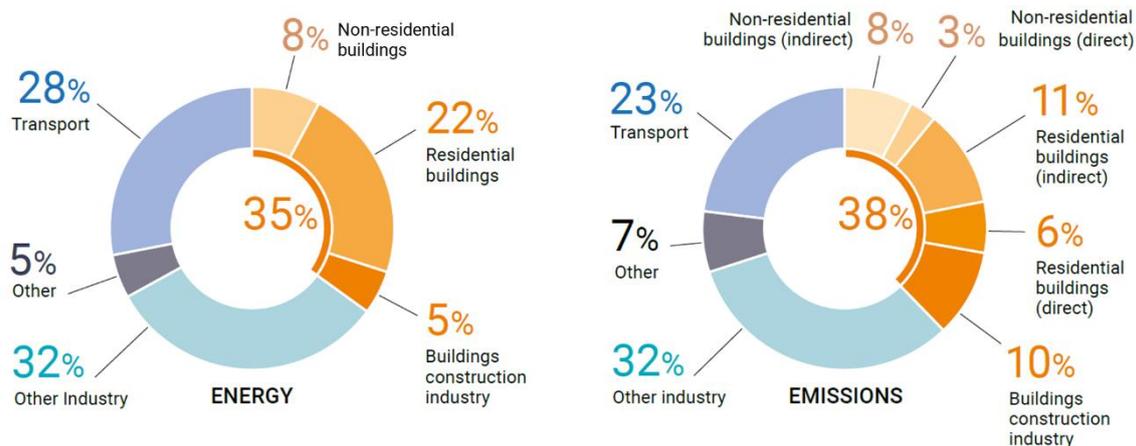
## 1.1. Decarbonisation and the impacts of buildings

Nowadays it is clear that energy savings are not simply a reduction of costs but also a tool to ease the anthropogenic emission of climate-altering gases into the atmosphere, an urgent and important goal because of the correlation with global warming. High-level temperatures and protracted heatwaves are becoming more frequent phenomena in many parts of the world. In the Northern Hemisphere, 29 countries shattered all-time historical measures of summer temperatures and Australia as well overcame its hottest record, as in 2019 the summer temperature was around 3°C warmer compared to the 1961-90 average [1]. A simple illustration of global warming is shown in Figure 1.1.



**Figure 1.1** *Warming stripes* illustrating the annual global temperatures from 1850-2017, from [2]. The colour scale covers 1.35°C of change in global temperatures

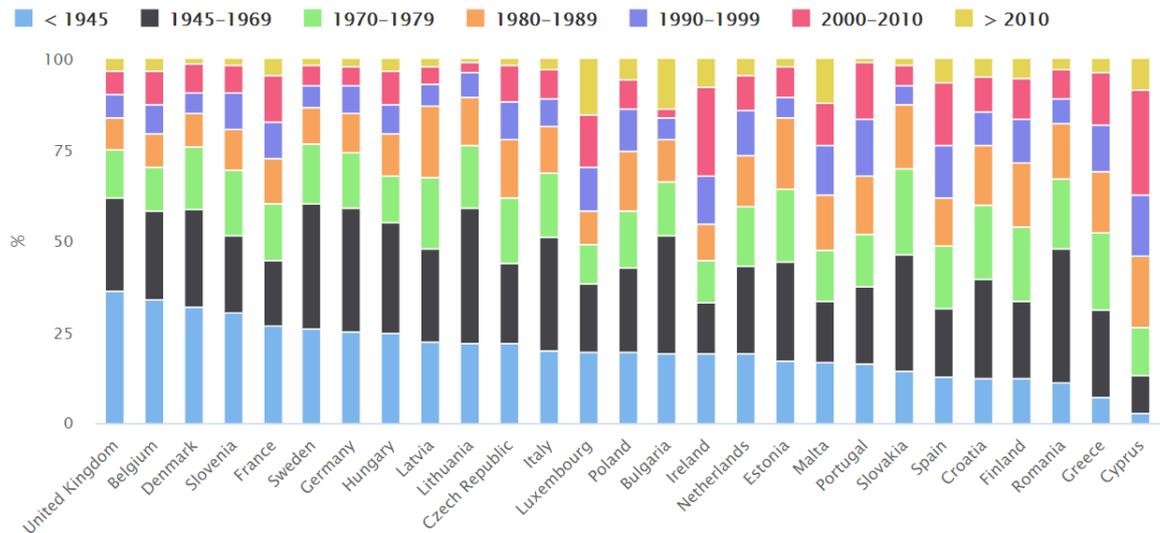
Climate change is a complex subject, often exemplified with CO<sub>2</sub> emissions and its relation with energy production. The buildings and construction sector combined are responsible for over one-third of global final energy consumption and nearly 40% of total direct and indirect CO<sub>2</sub> emissions [3]. Buildings spend massive amounts of energy worldwide to preserve indoor temperature. Especially the energy utilization of residential and commercial buildings in developed countries are growing more rapidly, observing the share varying from 20% to 40% yielding to energy use comparable to transportation and industrial sectors [4], and emissions seems not to decline. The latest Global status report for buildings and construction [5] found in the building construction and operation sector the largest contributor to the emissions and energy demand. Both in residential and non-residential buildings, power generation for commercial heat and electricity (labelled as indirect emissions in Figure 1.2) is the major emitting activity.



**Figure 1.2** The weight of buildings and construction sector on global emissions and final energy in 2019, from [5]

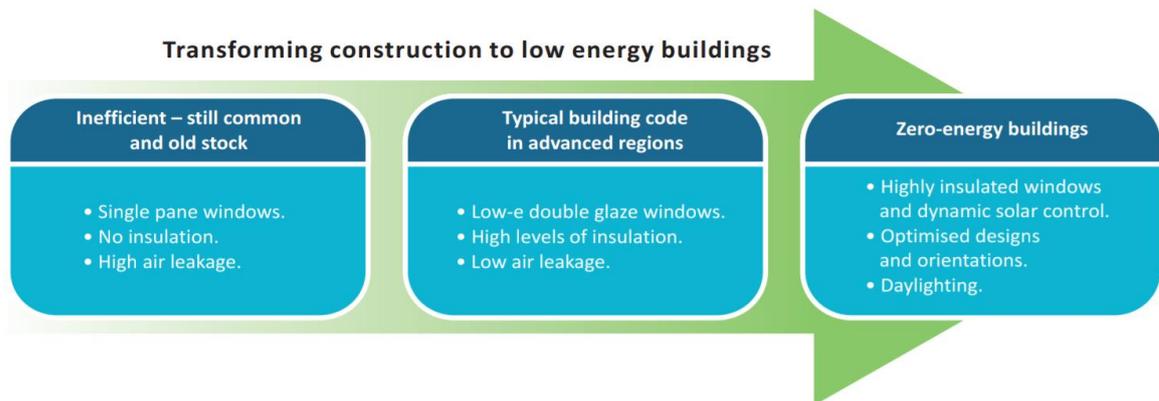
While electricity is consolidating in building services, since 2010 fossil fuel consumption grew as well at a marginal annual average growth rate of 0,7%. In the same time period, direct emissions from building registered a 5% rise. This wrong direction is also led in part by the construction of new facilities globally, in fact, between 2010 and 2019, the final energy use per m<sup>2</sup> has been declining continually by 0.5% to 1% meanwhile average annual floor area increased with a higher rate of 2,5% in the same period [1]. Hence, there is a need for effective refurbishment solutions which has to be simple to integrate with existing buildings.

Deep renovation of inefficient existing buildings is a key operation to try to constrain the temperature rise. Not only it leads to energy consumption decline, but to reduced capital costs as well, because holistic improvements in building envelopes and end-use equipment enable smaller heating and cooling systems need. According to the International Energy Agency [6], building retrofitting should rise from the 2014 value of 1% per year to a target rate of 2% per year. This is particularly true for building stock in continental northern hemisphere countries, where at least 75% of it is predicted to be still standing in 2050. For example, in the European Union the vast majority of residential buildings (which represent the majority of the floor area) has been built before 2000 [7]. This is significant because, in 50 years of EU policy to promote energy efficiency in buildings, the most important step came just in 2002 with the introduction of the Energy Performance of Buildings Directive [8].



**Figure 1.3** Residential buildings in EU classified by construction year (2014) from [7]

In the Technology Roadmap of 2014 [6], the IEA identified 3 stages of energy efficiency in buildings, visible in Figure 1.4. The higher level is characterized by greater passive design, enhanced by highly insulation also for windows and free contributions to the indoor thermal comfort. An innovative approach should be provided by advanced facades developed to reduce cooling loads and meanwhile harvesting natural daylight. In the document, five technologies are listed as R&D investments which can yield to major returns and one of them is “lower-cost automated dynamic shading and glazings”.



**Figure 1.4** Progression of building envelopes refurbishment from old stock to future efficiency, from [6]

## 1.2. *Adaptive Envelopes*

### 1.2.1. **Concept of the envelope**

The building envelope have several purposes. In addition to protect people from safety threats and shelter building elements themselves from damages such as moisture-related ones, the main goals deal with the conservation of comfortable and stable indoor conditions despite the outdoor climate. Temperature, lighting, and Indoor Air Quality (IAQ), that takes into account pollutants' concentration and relative humidity, are nowadays the principal requirements in developed countries for indoor environments.

As stated by Perino *et al.* [9], for centuries designers and engineers dedicated on creating a shield to ensure a separation effect. The goal was essentially to provide a bubble of artificial climate, making the indoor environment insensitive to its surroundings in concern to unwanted phenomena, such as heat flows and air infiltration. In the end, comfort was entrusted to the installation of mechanical and electrical systems to bear the users' needs for heating, ventilation, air-conditioning (HVAC) and artificial lighting.

Furthermore, in heating-dominated climates, the attention to cooling, artificial lighting and plug-loads is lower compared to the concern for keeping sufficient warm temperature during the cold season. This focus yielded to an energy conservation approach that brought to building elements oriented to minimize the heat losses and maximize the free gains. In the last decades, European and National standards set decreasing limits for these flows and year by year improvements tend to become marginal [9].

### 1.2.2. **Static envelopes**

In order to conserve energy, characteristics of the building such as shape, orientation, sun and wind exposure can be established during the project stage. Building components are designed based on thicknesses and composition of materials, and tightness, thus, to reduce heat losses by transmission and ventilation.

The commonly used building components are not intended to change. Meanwhile the indoor environment is oriented to stay constant, weather and seasonal climate provide a range of different conditions. Some of these changes are taken into account, such as the sun path and the thermal displacement. But many happens on an hourly or sub-hourly basis, like in the case of human behaviours that through preferences and schedules can make vary

internal gains and services' needs as well. As an example, a technology that can do poorly in both summer and winter is a static glazed façade. The transmission losses rise the heating demand, while the great solar gains increase the cooling need and the oversized transparent portion can cause glare discomfort.

### **1.2.3. Definition of adaptive envelopes**

Even if there are several terms used to describe this technological approach, 'adaptability' is appropriate as intended by Ferguson *et al.* [10] that refers to the ability to vary over time the physical characteristics of the system in order to address multiple requirements despite changing conditions. In the definition of Climate Adaptive Building Shells [11], authors find the ultimate goal in the overall building performance and specify that variations have to be also repeatable.

In practical terms, a building shelter component can act at a macro or micro scale. Having moving parts, or taking advantage of fluids' flows such as water or air, is a change at a macro scale and the term 'kinetic' well describe the process. Examples of that are smart shadings or the integration of natural ventilation both on external surfaces and inside the building. Many developers focused on the micro scale to switch the envelopes characteristics. Phase Change Materials (PCM) can store or release latent heat instead of simply reducing the transmission. A lot of research has been done about smart windows that mainly concern with changing light transmitting properties.

Why and when these changes occur depend on the type of control, that can be classified as extrinsic or intrinsic. Extrinsic, or active, solutions depend on other components that play the roles of sensors and processors. When a device computed the acquired variables, the responsive building envelope is instructed as an actuator for the chosen policy. Rather to rely on other elements, if the adaptation is caused directly by the environment, we have an intrinsic, or passive, technology and the components is, conceptually, both a sensor and an actuator. Usually, this approach can lead to a lower energy use and lower complexity to run.

### **1.2.4. Potentials of adaptive envelopes**

This ability of relevant features to fluctuate in a range, instead of delivering a fixed value, clear the way for more sophisticated management of heat fluxes, ventilation rates and

daylight harvesting, in addition to shifting energy in time as well. Furthermore, the building component envelopes doesn't need to be a static compromise between performances under different conditions but can adjust to best fit several trade-off situations such as: solar gains and overheating, daylight and glare, clean air and draught risk, solar shading and artificial lighting.

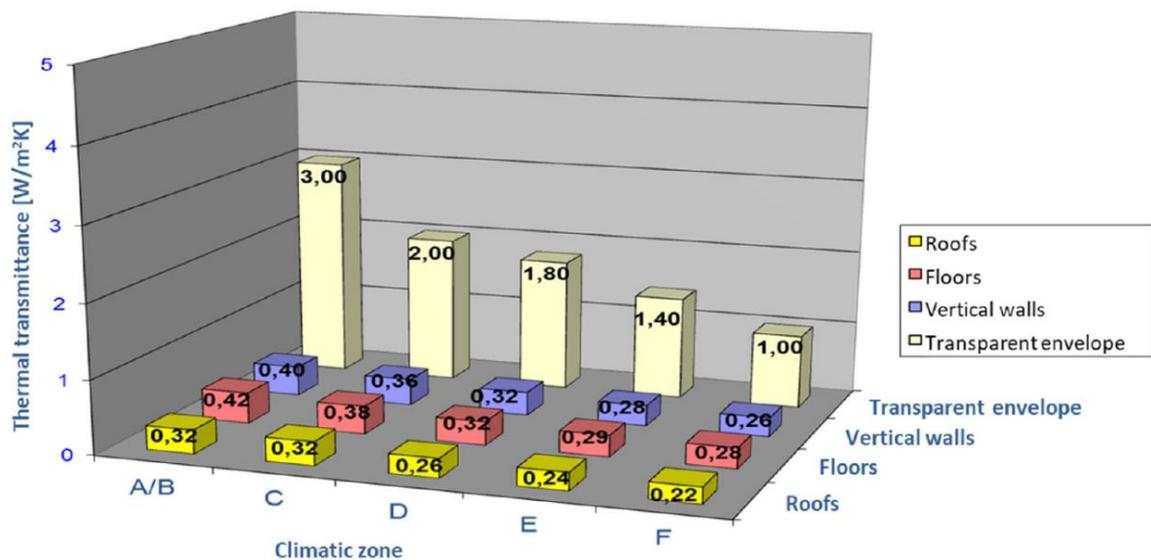
For example, as seen before, really low thermal transmittance can reduce the heating demand but increase the cooling demand. But bi-directional thermal diodes can both prevent and promote heat transfer through the wall construction [12]. The heat flux can be also lowered because of smart glazing like SMARTGlass [13]. A version of this full-scale prototype is composed by a PCM layer and a Thermo-Tropic (TT) one. While the PCM layer reduces the visible solar radiation entering the building to prevent both overheating and glare, TT layer passively contains the solar transmittance to better control the storage/release process of latent heat by the PCM layer. This configuration allowed Goia *et al.* to reduce the total heat flux in all conditions compared to a reference double glazed unit.

As in the SMARTGlass experience, where different switchable materials are combined to enhance their efficacy, adaptive components themselves can be combined to work together. Corgnati *et al* [14] tested a multifunctional facade module in Turin (45.08 N, 7.68E, 2894 HDD) and obtained a net positive specific daily energy flow entering the room even in the cold season. The ACTRESS prototype was designed as an Opaque module and a Transparent one. The performance was made possible using several technologies, and physical phenomena as well. PV panels creating a ventilated cavity, inlet/outlet grids, Vacuum Insulation Panel, PCM and electric heated carpets constituted the opaque portion, while the transparent one consisted in two high performant glazing systems.

### **1.2.5. The transparent envelope**

The transparent portion of a building envelope plays a key role. Because the light is fundamental for carrying out human activities, during the use of indoor environments we exploit artificial light when natural light is not enough. The presence of glazed surfaces therefore saves lighting energy [15], [16], but their benefits are not only those ones. In fact, daylight is an important factor in our biological processes, as well as promoting our health and comfort [17].

From an energy point of view, windows are a weak component. The U.S. Department of Energy stated that 25-30% of heating and cooling energy use in residential buildings, where the Window-to-Wall Ratio (WWR) is usually small, is caused by heat losses and heat gains through windows [18]. Fenestrations have in fact a much smaller insulation potential compared to opaque components, reason why, for example, Italian regulation sets much higher limits for windows' U-values than for the rest of the envelope, as shown in the Figure 1.5.



**Figure 1.5** Limits of thermal transmittance (U-value) undergoing energy refurbishments between 2019 and 2021, from Italian DM Minimum Requirements 26 June 2015. Image from the course "Energy Audit and Certification of Buildings", taught at Politecnico di Torino in 2018/19

In hot seasons, transparent components can become an issue as well because heat gains for square meter of an un-shaded glazing can be 100 times the heat transferred through an insulated wall [19].

The energy challenge is well known, but windows also affect the wellbeing and productivity of people. Existing standards concerning health and wellbeing of building occupants, like the international WELL Building Standard (WELL) and the European EN 17037, set minimum level of exposure time to daylight because it can strengthen users physical and psychological health [20]. Those factors are important even from an economic point of view, at least in developed countries: although energy savings can reduce building operation spending, the costs of operation and maintenance are smaller than costs for

salaries of occupants [21]. There is still a debate on how productivity is possibly altered by temperature and lighting. Overheating can lead to a 12% drop in thinking performance and 26% in typing, and this only with a Predicted Medium Vote value, which measures thermal comfort, of + 0.5 (slightly warm) [21]. An extensive recent review of lighting impacts on productivity [22] observed that most of the analysed studies found a positive correlation between illuminance and performance. In particular, a research carried on more than 21000 American elementary school students showed a 21% improvement in learning rates for children using classrooms with more daylight than the environment of the control group [23].

Once the role of transparent component is clarified, the main characteristics of glazing for common use in construction are identified. The higher the heating load, the more important it is that the overall heat transfer coefficient is low. As showed, natural light is generally to be maximized and the visible light transmittance is obviously positively related to this resource. However, a high value can also bring disadvantages, such as glare discomfort and increased cooling load, thus its optimal level is related to boundary conditions, which are not usually static as described above. Ultimately, Solar Heat Gain Coefficient (SHGC) indicates the amount of solar energy transmitted through the window as heat. The design of this specific characteristic greatly depends on the climate and should be low to reduce cooling load and high to favour heat gains in the cold season [24]. In order to address both energy efficiency and visual comfort, designers choosing *static windows* (intended as “with not changeable features in time”) have to balance the mentioned qualities and find the best compromise fitting their specific scenario. This implies limitations in pursuing optimal performance. Moreover, as we have seen, the conditions are not always predictable, therefore the compromise will be greatly affected by uncertainty.

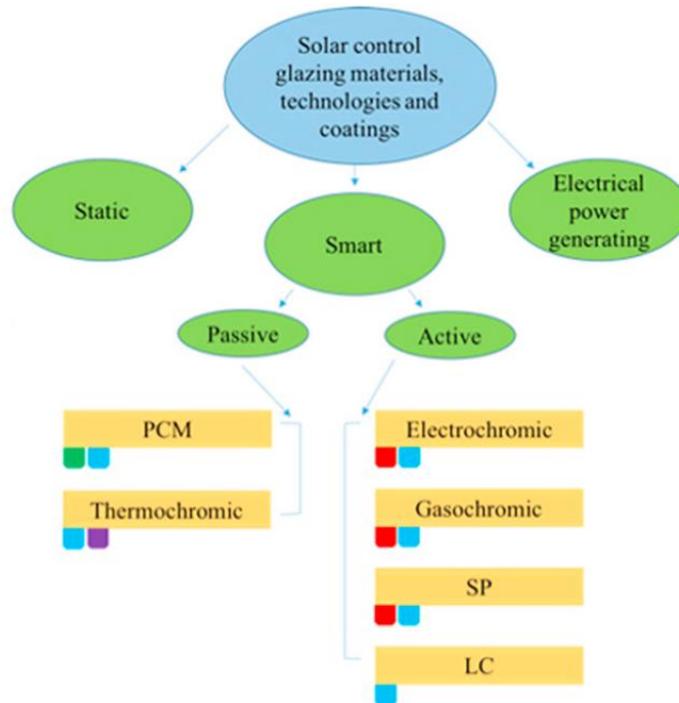
For the sake of brevity, the different technologies for static windows will not be described. Though, it is appropriate to cite that there are several of them, from more common like low emitting coating and multiple panes glazing to more particular like reflective and aerogel equipped. These products, although they are better performing than simple glazing, still have the same limit of not being able to adapt to conflicting needs [25]. Before to dive into adaptable glazing, it is also worth mentioning windows with integrated photovoltaic cells. They, in addition to standard glazing features, can produce electricity depending on weather conditions. This can reduce the use of energy from the grid, in all

seasons, but reduce the daylight available indoor, thus potentially increasing artificial lighting demand and compromising benefits related to natural lighting. In conclusion, movable solution like smart shadings can prevent heat gains and glare discomfort, but their adaptability is not as accurate as advanced glazing technologies.

### **1.2.6. Smart windows**

In addition to static technologies and to glazing components integrated with photovoltaic cells, alternative options for transparent envelope lies in the so-called *smart windows*, which are a building envelope measure needing development to become cost-effective [6]. In general, switchable windows can interact in both thermal physical domain and the optical one. The adaptable characteristics can vary the conduction/convection/radiation/storage of thermal energy and/or influence occupants' visual perception by means of the transparency/colour of the building component. Thanks to their distinctive peculiarity, they can both facilitate the preservation of the users' comfort (both thermal and visual) and contribute to the achievement of decarbonization targets. These adaptive building elements are categorized by Rezaei *et al.* [24] in passive (intrinsic control) and active (extrinsic control). Electrochromic (EC), Liquid Crystal (LCD), and Suspended Particle (SPD) devices respond to external control signals. They have a functional layer modulated by electricity (through injection/removal of electrons, or magnetic field variation) that vary their optical properties. Gasochromic (GC) technology have the same type of control, but the transition is driven by gas injection. On the contrary, passive technologies are directly sensitive to internal energy variations, depending on surface temperature or incident radiation. These variations of the functional layer trigger the switchable physical properties of the device. These technologies include Thermo-Chromic (TC), Thermo-Tropic (TT), Phase Changing Materials (PCMs) and Photo-Chromic (PC) glazing.

The description of several smart windows is carried on exposing their distinctive characteristic, explaining their composition and consequently the functioning. Lastly, some remarkable performances are reported. After this detailed overviews, a cross-evaluation is proposed.



**Figure 1.6** Some examples of different glazing technologies. From [24]

### Thermochromic devices

A Thermochromic material is able to change its optical properties depending on its temperature. This feature is proper of some transition metal compounds that go through not only micro-structural transformations caused by temperature variations, but also qualitative modifications of their electronic characteristics led by this transformation. The change occurs at well-defined temperatures, called “critical”. Therefore, the nature and influence on the transmitted solar radiation lay in the choice of the material. The most common thermochromic material is  $\text{VO}_2$ , which is relatively infrared transparent when at temperature below the critic one, whereas it mainly reflects infrared light when above the limit. The Vanadium Oxide is located in the interlayer of a laminated glass.

Giovannini *et al.* tested a TC glass in Turin with the intention of evaluating its effectiveness for different objectives. Although the office equipped with this technology achieved a 20.6% reduction in primary energy for heating, cooling need even increased (+23.2%). The impact on lighting consumption was almost nil (+0,6%), but compared to a selective glazing the TC window provided more daylight suitable to perform a visual task [26].

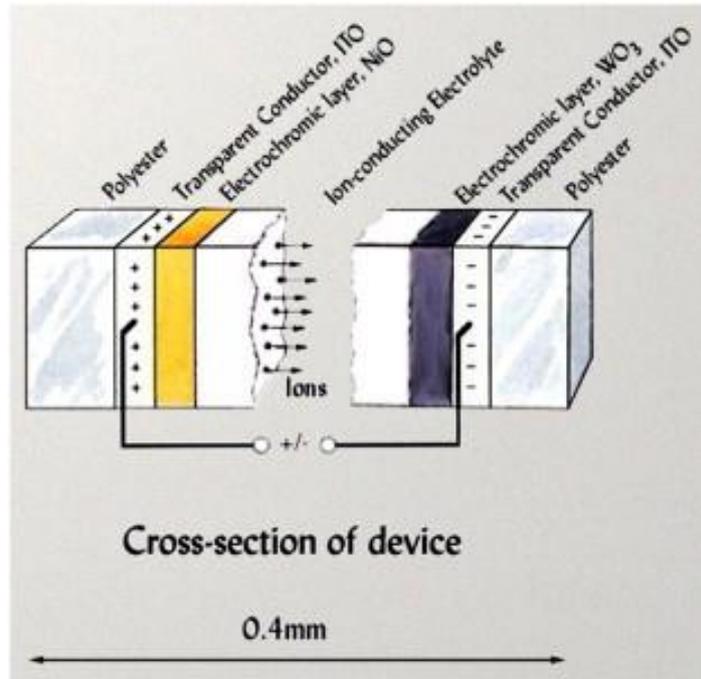
### Phase Change Materials devices

Phase Change Materials (PCMs) are special composite that store/dissipate heat through transition between solid and liquid states. They are becoming of general interest in the construction sector [27] because of their ability to increase the thermal inertia of components (notoriously low characteristic in glazing), hence smoothing out temperature fluctuations and shifting peak loads. PCMs find application in fenestration as filling of the interlayer between window glass panes.

A 2015 review reported that none completely transparent PCMs have been commercially used for windows [27]. Therefore, nevertheless they transmit a good amount of visible light, PCM devices are not suitable where visual contact through glazing is desired. Furthermore, because every PCM have a specific melting point, selecting the optimal one for the scenario of interest is crucial [28]. A full-scale PCM glazing prototype tested in a temperate sub-continental climate by Goia *et al.* [13] showed the significant improvement in cooling energy savings (more than 50% reduction of summer energy gains).

### Electrochromic devices

An Electrochromic material is able to change its optical properties in respond to an electric signal. This category of glazing is reported as built with organic or inorganic materials [24]. In the first case, molecules of the material vary their colour through reduction/oxidation. In inorganic material, the functioning is based on metal oxides. Most details have been found on this last technology. For the operation of an electrochromic device, several layers are necessary. An example is given in Figure 1.7. In addition to the two transparent outer substrates, which can be in glass or polyester, the intermediate functional part is composed of several films. These usually are an electrochromic film (typically based on  $WO_3$ ) and a counter ion electrode (typically based on NiO) separated by a purely ion conductor and in contact with two transparent conductors at the ends.



**Figure 1.7** Composition of an electrochromic device. Image from [29]

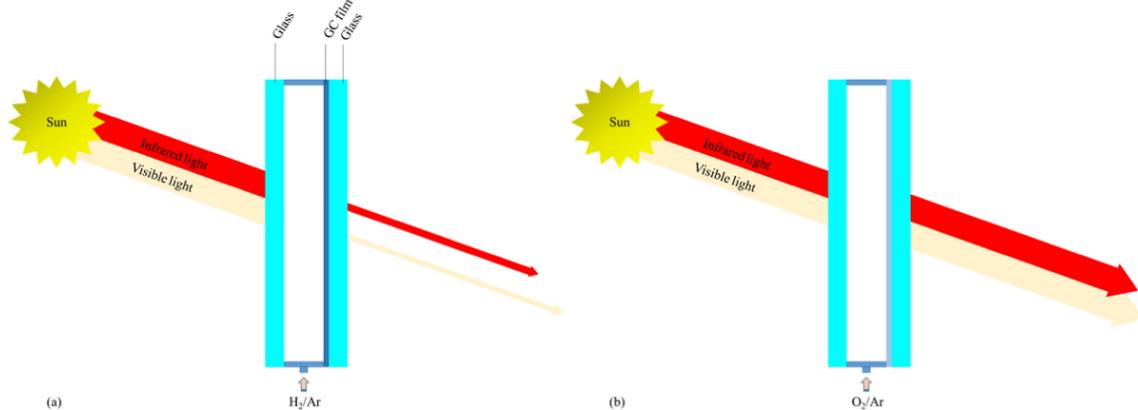
The two external conductors are actually coated with the relative metals and connected to a circuit. When a positive voltage is applied, charge is transferred between the EC material and the ion storage material and it lead to the optical transparency variation. The original transparency can be re-established by reversing the voltage polarity. These transitions can require a time of several minutes or even in the range of seconds and they can afford a high modulation of the visible transmittance (up to 60%) [24].

A study by Clear et al. [30] showed a significant potential energy saving for cooling (19-26% annual reduction of the peak cooling load) when the control was oriented to minimize solar heat gain, whereas, if optimised for visual comfort, the fenestration could yield to a 48-67% reduction of energy for lighting. The comparison is referred to best static windows (in 2006) and it was conducted on a forty-three employees office. Benefits in diminished glare and lower reflections in screens were also reported.

### Gasochromic devices

The transition of optical properties in gasochromic devices is driven by gas injection. The gasochromic material (tungsten oxide is generally used) is coated with a catalyst layer and located between the glass panes. The catalyst divides  $H_2$  into  $H^+$  ions, therefore the volume of hydrogen in the interlayer define the transparency level. When the hydrogen concentration rise, caused by introduction of diluted hydrogen gas, transparency drops. On

the contrary, the injection of diluted oxygen gas restores the normal characteristics. A scheme of the composition and functioning can be seen in Figure 1.8.



**Figure 1.8** Examples of gasochromic window using Argon to dilute H<sub>2</sub> and O<sub>2</sub>. From [24]

Regarding the adaptation potential, a study conducted in laboratory [31] showed that GC windows transparency can vary between 0,77 and 0,06. Feng *et al.* [32] investigated the energy benefits through a simulation of a commercial office building in Shanghai. After the sampling of thermal and optical features of a GC device, they compared the performance with several commercial technologies. The HVAC (heating, ventilation, and air conditioning) load of the virtual GC window against a double glass unit was 11,5% lower.

### Liquid Crystal devices and Suspended Particles devices

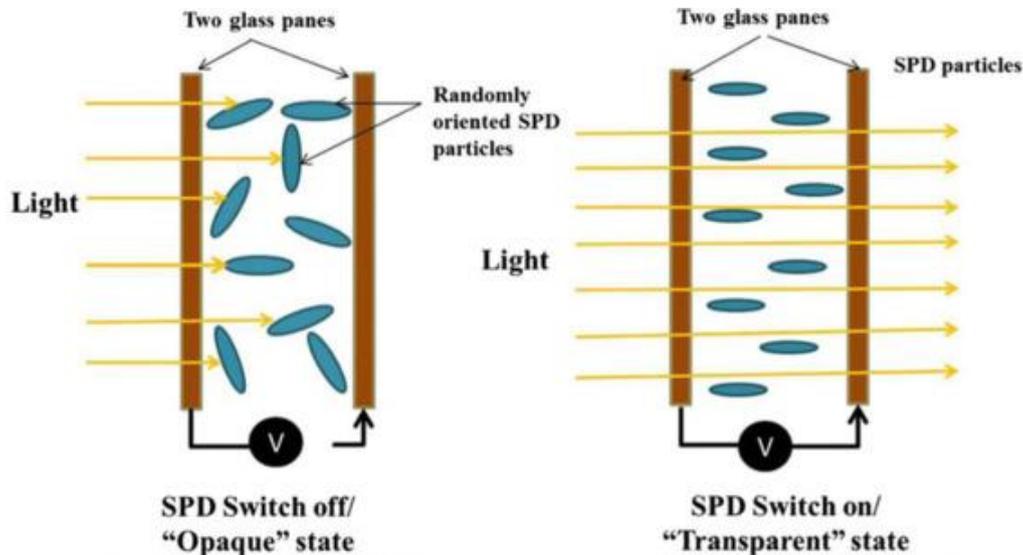
Suspended Particles devices and Liquid Crystal devices are both electroactive devices and were originally developed for displays. This means that they need a continuous power supply to maintain the maximum transparency, but also that they are relatively faster than other adaptive transparent devices. Therefore, they present an approach further than EC and GC technologies.

LC-based windows have two conductive electrodes and between them liquid crystal molecules are collocated. These molecules, in absence of an electric field, scatter light and make the window appears white and translucent.

SP devices are instead composed by 3-5 layers, with at the extremity two transparent conductors and a gel or an organic fluid between them where adsorbing dipole particles (usually polyhalide) are suspended. In the off state, particles remain random and absorb

light, thus decreasing the light transmission of the device. The transition time is about 100–200 ms only.

In both technologies, the application of an AC voltage aligns active elements so that the light is undisturbed. Figure 1.9 shows a scheme of the two states in a SP glazing. SP can have different transmittance ranges that typically are 0.79–0.49 and 0.50–0.04.



**Figure 1.9** Representation of suspended particles states used in a glazing. From [24]

The principal drawback of these technologies is the continuous power supply needed, resulting in an energy consumption from 5 to up to 20 W/m<sup>2</sup> for a LC device [33] and in SP devices requiring more energy than EC windows as well [24].

### Photovoltachromic devices

An interesting solution is represented by the application of Photoelectrochromic cells (PECCs) as glazing integration [34]. PECCs are generally photoelectrochemical cells composed by a redox electrolyte that divide two electrodes. This transparent functional layer not only can range its coloration, but also generate the energy needed for the transition. This process, depending on the available irradiance, can therefore prevent the use of external voltage.

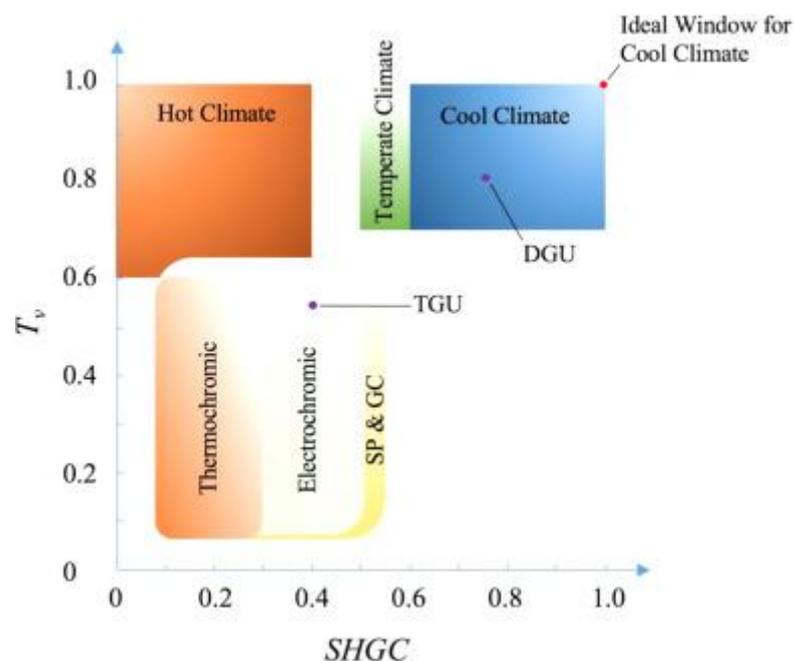
Favoino *et al.* [25] simulated different control approaches exploiting this technology and compared several performances to conventional static windows. The reduction of energy use has been marginal (maximum of 12%) and no significant improvements in visual comfort has been reached.

## Comparison of Windows Technologies

At the beginning of this section, the different goals of fenestrations and then the role of their properties have been covered. The goal was to investigate the state-of-the-art to identify an archetypal smart window suitable for pursuing distinct purposes, from energy savings in the different building services and visual comfort as well. Therefore, a good range of adaptability and the possibility to submit several control policies are desired.

Most of the cited glazing systems have a solar modulation power and glare control, but devices based on PCM or PECCs materials have low visibility that is, in this case, a major drawback. TC glazing also provide low visibility, but their main limit is the type of activation that not only is typically passive, but also depend on a hard to regulate property as temperature. LC-based windows cannot modulate SHGC, thus, their application greatly excludes cooling energy savings.

Figure 1.10 shows the climate-based optimal ranges for visible transmittance ( $T_v$ ) and Solar Heat Gain Coefficient (SHGC), and the potential of some of the principal described technologies. DGU stands for Double Glass Unit, meanwhile TGU stands for Triple Glass Unit. Electrochromic, Gasochromic and Suspended Particle devices are the smart windows most adaptable for different climatic conditions [33].



**Figure 1.10** Performance of optimal windows and actual glazing technologies. From [24]

EC glazing have more limited modulation levels compared to GC and SP glazing. These last two are also faster, with transition in SP occurring almost instantly, but changes in weather conditions usually happen in the range of minutes. Even though energy spent by EC devices depend on the number of switches that are performed, the continuous energy supply needed by SP devices lead most probably to higher consumption, making theoretically harder to obtain energy savings. GC windows need special equipment to function that can discourage their implementation.

In the end, EC technology seems the most effective opportunity. Its potential ability to pursue contrasting objectives has also been described and furthermore, in future, spectrally tunable EC could become commercially available as well. This version of EC devices is able to independently modulate visible and NIR light, thus control SHGC without compromising natural lighting [24].

### *1.3. Control strategies*

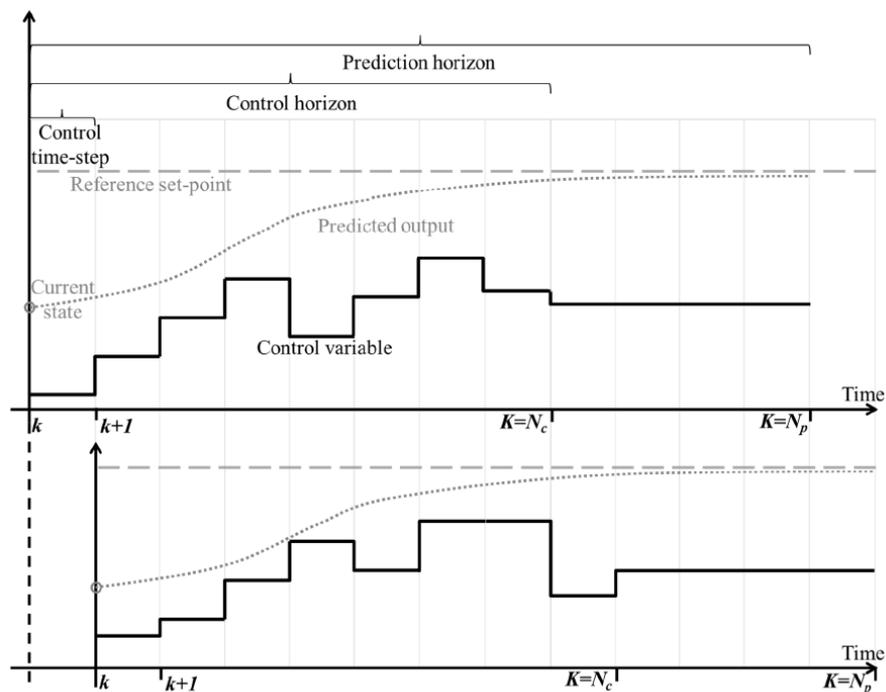
#### **1.3.1. Building Automation Control Systems**

##### *BACS wide-spreading and challenges*

If the building envelope has until now had energy conservation as its main purpose, the technical systems has focused on increasing efficiency. This is made possible by both new generation and emission technologies, but also thanks to intelligent controls aimed at minimizing waste. Building Automation and Control Systems (BACS) have been developed to save energy while providing internal comfort. [35]. Dounis *et al.* [36] conducted a review because the HVAC systems energy is highly related to the control strategy applied to them, although the concepts governing the control of HVAC systems are similar to those of other types of processes. BACS are also called *multiple-input and multiple-output* (MIMO) systems [37], this complexity make possible many difficulties: nonlinear dynamics, time-varying disturbances (such as weather), time-varying system dynamics and set-points (ruled, for example, by occupancy and seasons), different services with possible countering policies [38]. In fact, even though the computerization of physical processes has already reached high standards in fields such as Industry 4.0, the automation of envelope components still requires research in a recent state-of-the-art case study [39].

## Common control logics

Control is normally applied to temperatures and flow rates for the heating/cooling fluid or binary states such as on/off for lights. Rule-based control (RBC) is the most adopted control logic in practice [40]. These logics are sets of condition statements (sometimes only one), centred on environmental boundary set-point(s), usually static and pre-determined by their developers relying on experience [36]. Another common method is Model predictive-based control (MPC) that looks for the optimal values of the controlled variables computing objective functions. The peculiarity of this approach is the prediction of disturbances (intended as uncontrolled inputs) based on mathematical models that synthetically represent the involved physical system. The simulation of sequences of actions can be carried out to apply only the first step of an optimal sequence and then calculating a new sequence for the next time-step. The replanning approach is defined as Receding Horizon Control (RHC). An example can be found in the Figure 1.11.



**Figure 1.11** General representation of a receding horizon logic, with the different variables computed at time-step  $k$  (above) and  $k+1$  (below). Image from [41]

## Common inputs

In order to carry on a well-informed development of a control strategy, it is important to investigate not only the algorithms but also the involved variables. For this reason, an exploration of the inputs is carried out for general building systems and, in the next section, for CABS as well. Serale *et al.* [41] stated that the most significant disturbances taken into account to predict and calibrate the model are correlated to climate, occupancy and energy distribution through the grid. This last focus is mainly useful to shift peak loads (hence, make the electricity distribution more sustainable) or simply utilize energy in low-priced periods. Occupancy and climate values are more transversally applied [38]. Occupancy behaviour can be represented by direct binary states (occupied/unoccupied) or correlated information such as scheduled set-points for the controlled and adjacent zones. Many variables describe the environmental disorders: indoor and outdoor temperature (also processed as difference between the two), incident solar radiation (related to both visual and thermal comfort), other internal gains like the ones related to occupant activities and equipment and lighting use, Relative Humidity (RH), and wind velocity. Few studies embed occupants' comfort in the control strategies and for most of the cited studies by Park *et al.* [42], occupants' thermal comfort is simplified with a temperature range, rather than be represented by multiple parameters employed by thermal comfort researchers.

## Constraints' role

Many measured states and calculated variables can be framed as constraints, matching how they correlate to their equivalent of the real system. When the problem constraints are inflexible and it is compulsory that they are fulfilled, they are classified as hard constraints, although they are defined as soft constraints when they represent an adaptable limit and it is not rigorously needed their achievement. Normally, soft constraints are created using a slack variable that is able to shift the limit in a specified range, with a consequent penalty in the cost function. It is reported a comprehensive example represented by a RHC objective function based on a glare discomfort indicator (in this case DGP), in addition to energy consumption indicators (SE) [25]. This action selection was oriented to find the optimal smart window state (in terms of g-value and visible transmittance) taking into account visual comfort (Z component) only when discomfort threshold is exceeded:

$$\min \begin{cases} \text{if } DGP \leq 0.35, f(X) = NSE = SE - E_{PV} \left[ \frac{\text{kW h}}{\text{m}^2 \text{ y}} \right] \\ \text{if } DGP > 0.35, f(X) = NSE + Z = NSE + \frac{\sum_{i=1}^t DGP_i}{\text{Glare A}} \left[ \frac{\text{kW h}}{\text{m}^2 \text{ y}} \right] \\ X(t) = (g\text{-value}(t) [-], \tau_{vis}(t)[-]) \end{cases} \quad (1.1)$$

### Model function

Once identified the domain of the problem, understanding how possible boundaries can be combined is suitable. Control strategies can be divided in model-based and model-free. First, the conceptual roles of models in model-based methods is presented. Complete approaches present a model formulated as inter-dependent calculations aiming to represent the thermodynamical behaviour of the building, for example with R-C networks for the physical building. Furthermore, MPCs need at least an additional model to compute the forecast of disturbances. Finally, a possibly high-accuracy model of the building can be simulated to design the controller characteristics before to implement it on field. This last type of model can alternatively be the only one constructed to test, in the case of RBC, or train, in the case of Reinforcement Learning (RL). In this case the controller can be categorized as model-free because it doesn't utilize a system reproduction during application. In short, Reinforcement Learning is a particular machine learning technic where the artificial intelligence (AI) picks an option, get a reward based on the effects of the choice and, trying to maximize the reward, reinforce positive selection, usually outlining a policy.

### Model-based and model-free control

Despite the wide-spreading of model-based methods, limits of this approach were found in studies about BACS proposing model-free methods [43], [44] and model-based method [45] as well. These indications, focused on an actual application of the model-based strategy, regard the effort to build the model and its performance. First, the dependence on a priori knowledge is exposed. In fact, the model used on-field by the controller is to be determined by historical operational data or regression of field test data. This process requires an effort in terms of human labour, and sufficiently precise and numerous sensors as well. Both aspects are time-consuming and stricter as finer is the representation of calibrated model and/or forecasted disturbances. The second limit lies in the risk of model

uncertainty. Zhu *et al.* [45] exposed that there are two types of model error leading to uncertainties: model parameter error and model structure error. Model parameter error concerns the same concept of the importance of good training dataset in order to get an unbiased AI. Regarding building models, the major risk lies in fault historical data [45]. Model structure error refers to the simplification of a complex system like the built environment. The difficulties in representing it have been exposed at the beginning of this chapter.

A circumscribed comparison between the two methods and a local feedback control as well has been conducted by Qiu *et al.* [44] who, aiming to find an optimal policy to control the cooling water system, assessed a method based on reinforcement learning and a model-based one. The energy conservation performance of the model-based controller was better than the model-free one (14% energy savings against 11%), that, in turn, was better than the local feedback one (which reached 7 %) considered more basic. The former could precisely predict the outcome energy consumption affected by the two control actions thanks to more equipment models. Another identified reason for its better performance is the objective function focused on maximising the COP at each time-step. Though, the advantages of the implemented reinforcement learning are the evolving of its policy and the independency by a priori knowledge (with the possible drawback of building its own experience from scratch). In fact, this type of controller continues to learn and its energy saving rate could reach 12% during the second applied cooling season.

### **1.3.2. Adaptive envelope components control**

#### **Control dependency**

After the general overview of consolidated control strategies applied on management of building systems, an additional investigation is conducted about the adaptable envelope components. The chance to control building boundaries is particularly attractive because of the broadly exposed potential of CABS in tackling multiple goals, even if addressing conflicting targets make the research for control strategies more challenging. In fact, a control strategy improving energy savings in a service, can sometimes decrease the same performance for other loads or thermal and visual comfort. In the well-known review conducted in 2013 [11], CABS are considered a concept still to explore because of the relatively limited amount (44) of documented cases in the study. For example, a simulated smart window addressing multiple performances resulted in a broad range of possible

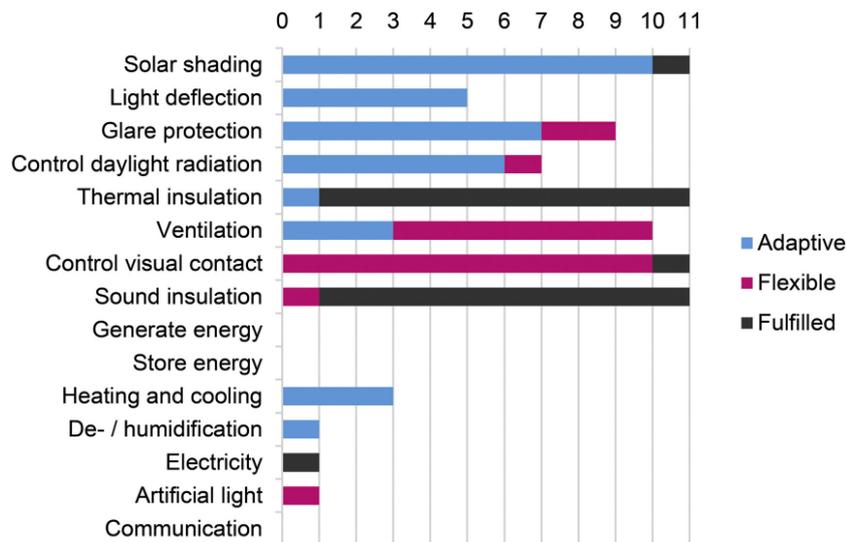
energy savings (+10% to -20%), daylight availability (40% variation) and thermal comfort (70% variation) [46]. The variations are due to different simple control strategy simulations and, hence, underline the high dependency from the control approach.

#### MPC methods for smart windows

The state-of-the-art investigation by Loonen *et al.* [11] highlighted that MPC supplied with weather forecasts, and optimal control (intended as based on minimize a cost function over a specified time period) are the most promising approaches. Concerning a photovoltachromic switchable glazing, a comparison between RHC e RBC showed a significant increased energy saving performance with the former method [25]. The model-based controller was built by Favoino *et al.* including the prediction of the disturbance on the energy balance caused by entering solar radiation. The simulations run with different temperate climates and the MPC potential resulted higher in heating dominated climates.

#### Common functions and adaptability

Despite the wide range of theoretical actions performable by CABS, Böke *et al.* [39] found that, in practice, more automation has been implemented regarding solar radiation phenomena. The conclusions of their recent study (2020) are to be referred carefully because of the restricted sample concerning only one country, namely Germany. However, putting aside the functions reachable with static elements, it is possible to see in Figure 1.12 that many described implemented façades yielded automatic adaptability in solar shading, glare protection, daylight availability control and light deflection. Most of these goals are reachable with EC management as described in the next section. It is also worth noting that, regarding these four tasks, glare control required relatively more intervention by users.



**Figure 1.12** Cross-case comparison of implemented façade functions. Adaptive: adapts independently due to its automated control: Flexible: adaptable only manually by the user; Fulfilled: function reached with static application. Image from [39]

### Users, an important aspect for adaptability

Concerning user interactions, it is valuable to underline that user behaviour is an important factor that modifies energy demand [47], [48]. A wide investigation on users' interventions in an electrochromic window control [49] suggested that the occupants' ability to correct or bypass the controller's actions is useful for intelligent control of adaptive components. This is justified by the authors not only by the disturbance represented by users, but also by the complexity of visual comfort. In fact, the simulation of the user-based control strategy yielded to annual energy lighting savings about 48%, outperforming the occupancy-independent method (35%) based on daylight only. Also, Cano *et al.* [50] has shown that considering user behaviour could yield to great percentage of savings in a smart building. Moreover, a fascinating control strategy has been developed by Cheng *et al.* [51] for taking into account preferences and complains by users. Occupants interacted through their satisfaction votes with the Q-learning (a RL method) based lighting and blind controller that was able to decrease excessive energy consumption while preserving visual comfort. This user centred approach was then investigated also by Park *et al.* [52] implementing a RL algorithm too but applied only on lighting control.

### 1.3.3. EC control

From the previous sections, it is possible to conclude that MPC and optimal control have a major role in CABS control compared with their application in BACS, where RBC is still relevant. These control strategies rely on model(s), and this approach has advantages and drawbacks. Relevant variables have been presented in order to understand the possible inputs for an EC controller. Lastly, an investigation about practical possible applications of CABS demonstrated how smart windows can have a significant role in designing adaptive envelopes and what effects are achievable by switchable building component actions. Once identified this useful base of controls in buildings, a narrowed research has been conducted about EC control.

#### Control dependency

First, also for EC windows the range of possible savings depending on the control method is broad. This has been explored through several simulations applying different control strategies, from the simpler one (binary action according either to the outdoor solar radiation or to the indoor light level) to the finer one (fuzzy controller based on 2-year data collection of users control associated with different meteorological conditions) [53]. Energy use differences resulted in substantial deviations of 24% for heating, 39% for cooling, and 20% and 63% for winter and summer lighting, respectively.

#### Common inputs for smart windows

According to a diverse overview by Favoino *et al.* [25], the most applied input variables in RBC of smart glazing are illuminance (on work-plane and external), vertical solar irradiance, presence of occupants, occurrence of heating/cooling loads, internal and external air temperature, seasonal services set-points. Illuminance monitoring is particularly important for smart windows control because of the role of transparent building envelope dealing with the complex provision of visual comfort. The work-plane illuminance (WPI) and glare risk are often used as control variables and are also been indirectly evaluated through a fine sensor unit embedded in an EC glazing, called EPD, with a vision of outdoor space [54]. In this study, the measurement was therefore more than a singular value but the luminance distribution depending on the sky background, sun, clouds, and landscape bodies. The generated map was then connected to an interior

geometric model. Weather components can vary in the order of minutes, influencing both energy demand and visual comfort performance. For this reason, CABS seeking daylight harvesting and solar shading to improve those aspects should be able to act in order of minutes to tackle those disturbances [11].

### Orientation and climate dependency

Hoon Lee *et al.* [55] simulated the heating and cooling annual energy performances of a benchmark building model to evaluate differences between six locations with different climates and between the four cardinal orientations. The thermo-visual properties of the 2 virtual states of the windows were 0.50 and 0.10 for SHGC, meanwhile for  $T_{vis}$  were 0.72 and 0.10. Hence, 0.50 – 0.72 was the clearest action.

Control strategies were simple RBCs with a singular control statement acting as threshold. The authors selected for this purpose outdoor air temperature (OAT), room air temperature (RAT), global horizontal irradiance (GHI), and solar radiation incident on windows (SRW). Between those, OAT resulted as the better performing for most window orientations and locations, but no particular value outperformed others. The optimal value for RAT is more often 24°C or 26°C, while 22°C is the inferior boundary for significant energy savings, except in Miami where the higher is the value (up to 30°C) lower are the heating and cooling consumption. The application of SRW as singular condition resulted in energy savings only in the three warmer climate. In Los Angeles and Baltimore optimal value was about 370-400 W/m<sup>2</sup>. Precise values identified by the authors are reported in Table 1.1, only for south orientation.

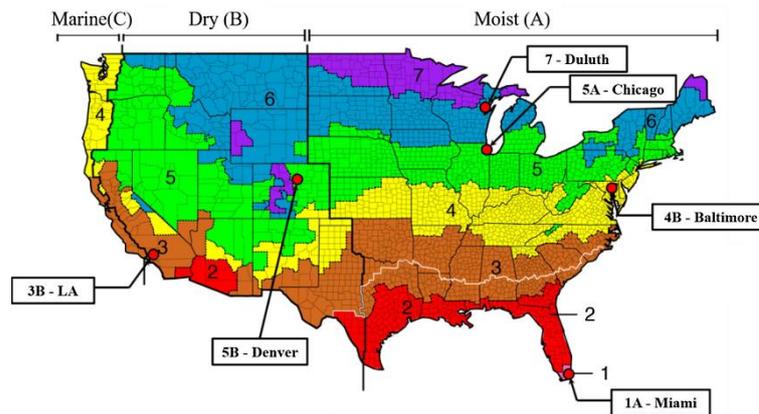
**Table 1.1** Optimal value of different control parameter for ECG window RBC by locations.

From [55]

	<b>MIA</b>	<b>LAX</b>	<b>BAL</b>	<b>CHI</b>	<b>DEN</b>	<b>DLH</b>
<b>OAT (°C)</b>	16.06	18.25	18.56	19.19	18.56	18
<b>RAT (°C)</b>	22	23.94	23.94	23.94	25.94	25.63
<b>SRW (W/ m<sup>2</sup>)</b>	2.0	397.2	400	419.6	401.4	415.3
<b>GHI (W/ m<sup>2</sup>)</b>	16.8	454.3	515.4	605.4	771.9	741.8

The west orientation yielded to the highest energy-saving potential (2.64–8.43%) depending on location, followed by similar performances by the south (1.59–6.01%) and

east (2.17–7.39%). As expected, north orientation led to stunted performances (1.70–2.03%). When smart glazings were simulated with the three best performing orientations, annual energy savings compared with the base case ranged from 13.46% (Miami) to 5.23% (Duluth), meanwhile locations with intermediate savings were, from the highest to lowest: Los Angeles, Baltimore, Chicago and Denver. Readings these results associated with the climate locations, visible in Figure 1.13, suggests that simulated smart windows, with basic strategies controllers, can achieve greater thermal energy savings for generally lower latitudes.



**Figure 1.13** Climate zone disaggregation of the U.S. reporting simulated climate locations.  
From [55]

### Climate-based control comparison

Favoino *et al.* [25] simulated different control strategies performed in 3 different climates (Rome, London and Sydney). The action selection was based on 4 states of the smart window embedded with photoelectrochromic cells (already described in the section about switchable glazing technologies). The properties of the glazing are reported in Table 1.2.

**Table 1.2** Thermo-optical properties of PVC-Glazing tested by [25]

$\tau_{vis}$ : visible transmission coefficient; g-value: total solar heat gain coefficient (-);  $\tau_n$ :

transmissivity (-)			
$\tau_{vis}$ (-)	g-value (-)	$\tau_n$ (-)	States
0.595	0.508	0.6486	1
0.446	0.396	0.4863	2
0.341	0.325	0.3719	3
0.238	0.238	0.2596	4

Only the control strategies useful for the present research are described and renamed for this purpose. All of them concerned incident solar radiation and some of them used a glare discomfort constrain. This constrain was oriented to guarantee a Discomfort Glare Probability (DGP) [56] lower than 0.35, whenever possible, representing the highest level of acceptability. *Basic RBC* acted passively in proportion of perpendicular solar radiation. Implementing the DGP constrain too, they obtained a *Glare RBC*. *Optimal Glare RBC* performed an optimisation problem each hour to minimize the sum of heating, cooling and lighting loads, taking into account the DGP constrain as well. Lastly, *Glare RHC* was based on minimizing the total energy load over a daily period while preserving visual comfort as described above.

Taking static operation (always clear or always darkened) of the same window as baseline, active control policies yielded between 2% and 12% energy reduction, depending on the climate. This was possible even lowering glare risk, but distinctions between control strategies and climates are necessary. In fact, the above-described exploration of control policies for smart windows demonstrated that, in a cooling dominated climate, energy savings and visual comfort (measured by the authors also in terms of Useful Daylight Illuminance index (UDI) [57]) are likely to be positively correlated. On the contrary, *Glare RHC* and *Optimal Glare RBC* implementation highlighted that introduction of glare constraints in a heating dominated climate (i.e. London climate) enhance visual comfort (about 8% higher UDI and DGP) but typically limits the total energy saving potential. Simpler control strategies like *Basic RBC* and *Glare RBC* yielded to lower energy use where heating and cooling loads are similar (namely, in Rome) but had the opposite result in London and Sidney, where the same condition is not met.

#### Controller based on seasonal users' behaviour

Another test of control strategies was carried out by Assimakopoulos *et al.* [53] through building energy simulations over 90-days periods each for winter and summer season. The 2 virtual actions corresponded to the properties of the electrochromic glazing validated against experimental data. These properties, registered in a connected paper [58], are reported in Table 1.3.

**Table 1.3** Properties of the simulated glazing. From [58]

	U- value (W/m <sup>2</sup> K)	Solar heat gain coefficient, g-value (-)	Visible transmittance, T <sub>vis</sub> (-)	Solar transmittance, T <sub>sol</sub> (-)
Bleached state	1.42	0.36	0.5	0.3
Coloured state	1.42	0.22	0.32	0.17

The control methods relevant for the present investigation are the following. Strategy 3: On–off control based on threshold above which the glazing is set at the coloured state and vice versa. Respectively 3a, 3b and 3c controller, had as threshold 500 lux or 400 lux for indoor lighting, or 350 W/m<sup>2</sup> for solar radiation. Strategy 6: Fuzzy controller was tuned applying a backpropagation algorithm founded on data gathered through two experiments during which an EC window was manually controlled by occupants. From the winter experiment, 6a strategy was developed, meanwhile the 6b originated from the summer one. Each fuzzy controller was then applied to both winter and summer simulations.

The best overall performance (heating, cooling and lighting consumption) was accomplished by strategy 3a, and secondly by strategy 6a. If lighting load is excluded, Strategy 6b led to the highest savings, but with similar results (variances smaller than 2%) by the binary controllers (3a, 3b and 3c). Therefore, a control policy relying on seasonal variation could have an advantage on simple methods (on–off). Though, it is worth noting that performances in balanced seasons (namely, spring and autumn) were not investigated.

### Commercial product

The following experiment is cited because a commercial EC glazing, hence available glazing properties, was tested [54]. The window had two sections (superior and inferior one) with four available actions each: 1% (full tint), 6% (medium tint), 18% (light tint), 60% (clear) of luminous transmittance. The switch from the clear to full tint state took between 10-30 min according to weather. The EPD sensor already described was implemented to get a luminance map of the inside. In fact, luminous transmittance was the variable to be maximised, with lower section having priority over the upper one to favour the view of the outside, meanwhile WPI and DGP were input variables to be preserved. The WPI was constrained to be between 500 lux (daylight provision requirement according

to EN 12464-1) and 2000 lux (avoidance of excessive daylight and solar heat gain). DGP was measured from two points and constrained, as in the previous exposed study, to stay below 0.35.

The action selection was applied every 15 min with additional buffer time and was performed in this way:

1. WPI is calculated for 16 EC tint combinations,
2. The ones not fulfilling the WPI constrain are discarded and, if no combination meets the condition, clear state is selected (assuming that daylight availability is poor).
3. Next, the luminous map is generated and the corresponding DGP is estimated in two points.
4. Finally, from the combinations satisfying DGP values, the one including the clearest state in the lower window section is selected.

Performances of a low-e double glazing with 62% visible light transmittance were contemporary measured in an adjacent room. Testbed was in Berkeley, CA, all windows were south-oriented and experiment was conducted in October. Energy savings were not the goals of the experiment, in fact, only mean visual comfort improvements (aggregated by the writer) are presented in Table 1.4, divided by weather description from the analysed paper.

**Table 1.4** Average performance difference of controlled EC glazing against static alternative.

Values aggregated from [54]

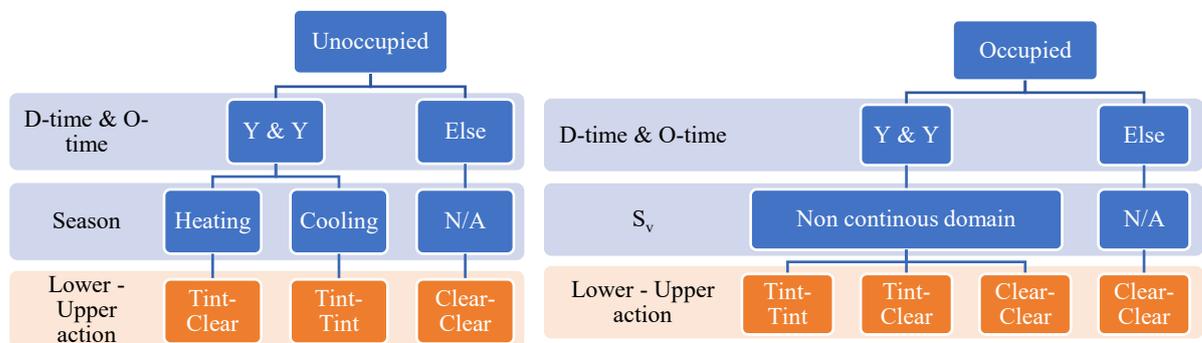
	WPI (%)	DGP (%)
<b>3-DAYS “CLEAR SKY”</b>	74	89
<b>2-DAYS “THIN CLOUDS”</b>	60	78
<b>2-DAYS “PARTLY CLOUDY”</b>	57	67

### RBC for visual comfort

After a 15-month collection of data about users’ interaction with an automatic 2 EC glazing window (up and bottom section), Lee *et al.* [49] tested an RBC during the last 6 months. During recording and experimenting period, interior manual venetian blinds and automatic dimming of lighting were present. The test room was a conference room; hence, overall lighting consumption were highly dependent on occupancy. Whole window properties for the two combinable actions were: SHGC = 0.39 or 0.08,  $T_v$  = 0.50 or 0.03.

The input variables were presence/absence of occupants (Occ), cooling or heating season (Ssn), occurrence of sun above the horizon (D-time), time correspondence with EC operational period constrained by authors between 9:00 and 19:00 (O-time) and, lastly, exterior vertical light level ( $S_v$ ).

The control strategy was rule-based and is summarized as follows: when unoccupied during daytime, the smaller upper section was set to clear or tinted to reduce heating and cooling consumptions, respectively, and the lower one was tinted to prevent discomfort glare in the instant when someone would enter the room. Whole window was clear when D-time or O-time condition was not met. Ssn was not taken into account if the room was occupied, but  $S_v$  instead. A schematic representation is available in Figure 1.14.



**Figure 1.14** Scheme of RBC developed from users' interactions analysis by [49]

Concerning users' interaction on at least one of the two sections, 65% of them occurred within the first 5 min of occupancy, half of which were just anticipation of the automated policy. The total 43 overrides took place in 4% of the 328 meetings. This can infer that the implemented control policy well preserved visual comfort. West-facing window tested in Washington DC had a high WWR (40%). This architectural factor favoured, according to the authors, the consistent influence of the EC control. Thanks to EC control, annual energy savings simulated without occupancy-based lighting through EnergyPlus were 48%. For the specific services, namely heating, cooling, fan and lighting, energy savings were, respectively, 63%, 34%, 52% and 49%.

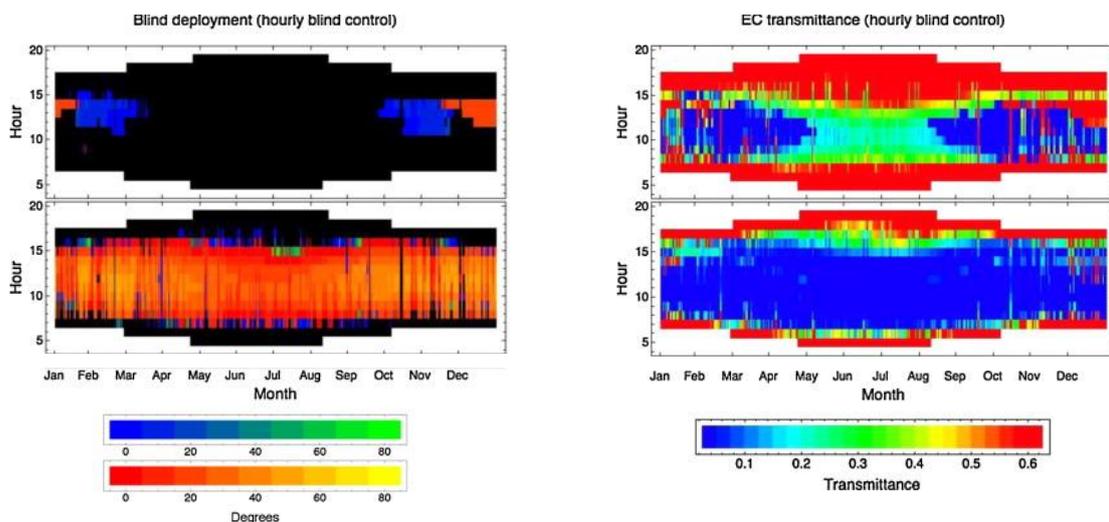
### Restricting blinds needs

In a simulation study focused on lighting energy savings [59], 2 EC panes transmittance was linearly interpolated between 0.05 and 0.60 and this give the opportunity to investigate

a dense action domain, as visible in Figure 1.15. The optimisation calculation was based on rendered images. Luminance ratios between visual task and surrounding background, and vertical illuminance at the occupant's eye constrain EC control. A control variable for blinds operation was the direct visibility of sun's orb by the occupant.

After the deployment of venetian blinds to avoid direct sun visibility, additional lowering of the shading device was actuated only if the action performed by optimization of window transmittance was leading to visual discomfort. Transmittance values, possibly different for the two panes, were selected through least-squares optimization with WPI as target parameter and in compliance with vertical illuminance and luminance ratios constrains.

Before to present observations useful for the present study, some disclaimers are to be taken into account: the electric lighting system was dimmable, and the south-facing window was almost high and large like the room. In addition, transmittance of the reference window was 0.60, and simulations were located in Berkeley/Oakland, California.



**Figure 1.15** (Left) Blind operation during year: (from top) EC windows, reference window. Blue/green scale refers to shading over upper pane only. Red/yellow scale relates to times for which shading over both panes. (Right) EC pane transmittance during year: (from top) upper pane; lower pane. Image from [59]

As it is possible to see from Figure 1.15, double-pane EC glazing controlled hourly required marginal application of blinds, on the contrary of reference glass where integral shading obstructed external view for most of the time, limiting the already cited benefits of this visibility aspect. The baseline to evaluate annual lighting energy use was compared to

reference window with blinds operated once per day. The simulated smart window yielded to 48% lighting energy savings.

#### Cross comparison of reported control strategies for smart windows

In Table 1.5 and Table 1.6, a synthesis of case studies main features is provided. Most of the reviewed researches about smart glazing control focused on energy consumption for lighting. Studies regarding energy savings commonly neglected visual comfort (Favoino and Chang are exceptions in the available group), meanwhile some papers concern visual comfort only, usually employing DGP as parameter. DGP however focus on glare risk only, not representing when the natural light is enough to avoid the artificial one. The shared use of lighting energy and/or visual comfort indexes suggests that the research field oriented smart windows adaptability to visual effects instead of a comprehensive multi-objective functionality. Accordingly, light measurements are always present as inputs. WPI is a very direct measurement of lighting conditions which can indicate both exceeding and insufficient illuminance, but this approach can require several sensors for few smart windows, depending on the amount of desks that the glazing can affect. External parameters instead can be measured by few devices for several transparent components, depending on the façade dimensions. Solar radiation incident on window is the most used input of this type.

Regarding the possible actions selected by controllers, visible transmittance of adaptable glazing range between 0,6 and 0,03 for simulated transparent elements, however, the only commercial electrochromic window (Wu) is able to reduce it even to 0,01. These values set the acceptable action range for the present work on EC glazing control. The majority of studies implemented the adaptable glazing to the sun-oriented façade. However, the extensive investigation of Hoon Lee *et al.* [55] suggested that east and west orientations have greater energy savings potential.

Lastly, a climate classification appeared to miss in different studies. Based on Köppen climate classification, cities have been labelled accordingly and from this reading is clear that the majority of tests in the present restricted sample have been conducted in Temperate climates (labelled “C”) and/or with warm or hot summer (labelled “b” or “a”, respectively). Again, most promising energy performances potential is not the most investigated, reading this climate trend with the observed correlation between warmer climates and energy

savings based on Hoon Lee *et al.* research. To conclude, a sun-oriented smart window in a temperate climate is judged as a representative case study because of its more average effect and for being a common framework that then allow comparison and connection between different researches.

Table 1.5 Main characteristics of reported case studies dealing with smart windows control (pt 1)

Constrains' limits	States limits	State	Objective function / Goal(s)	Controller type	Study
∖	3 configurations: (a) 500; (b) 400; (c) 350	3 configurations: (a), (b) Lux; (c) Solar radiation	Ep (H, C, L)	RBC	<b>Assimakopoulos (2007)</b>
∖	Occ = yes, no; Ssn = cooling, heating; Schedule = both day time & working hour, else; S(v) =	Occupance; Season; Schedule; Exterior vertical light level	Ep (L)	User-based RBC	<b>Lee (2012)</b>
Avoid sun visibility; 2 Luminance ratios	∖	Luminance map	WPI optimization	Opt RBC	<b>Fernandes (2013)</b>
∖	∖	tailored comfort index, number of lights on & blind slat angle	Reward = 0,8 x Tailored comfort index + 0,2 x Ep(C, L)	User-based RL	<b>Cheng (2016)</b>
DGP < 0.35	∖	Solar Radiation	Ep (H, C, L) & DGP (if exceeding)	RBC / Opt RBC / RHC	<b>Favoino (2016)</b>
WPI: 500-2000; DGP < 0.35	∖	WPI	T <sub>v</sub> at lower pane	Opt RBC	<b>Wu (2019)</b>
∖	Range, Significant: [OAT] -20/ 40, 18; [RAT] 18/ 32, 24; [SRW] 0/ 1000, 400; [GHI] 0/ 2000, 450	Outdoor air temperature, Room air temperature, Solar radiation incident on window, Global horizontal irradiance	Ep (H, C)	Opt RBC	<b>Hoon Lee (2020)</b>
∖	∖	wet bulb temperature & system cooling load	system COP	RL	<b>Qiu (2020)</b>

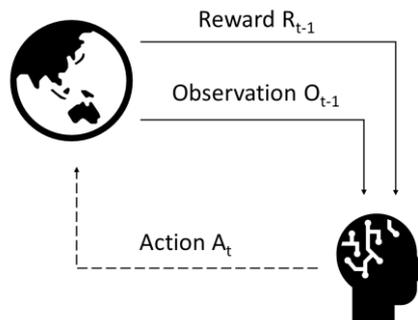
Table 1.6 Main characteristics of reported case studies dealing with smart windows control (pt 2)

Savings	WWR	Climate	Period	Orientation	Action	Study
Performances similar to fuzzy control based on users behaviour	\	\	3x2 months (summer and winter)	South	2 actions - SHGC & t(vis): (0.36, 0.5); (0.22, 0.32)	<b>Assimakopoulos (2007)</b>
Great savings, influenced by large WWR (H=63%, C=34%, Fan=52%, L=49%)	40%	Washington DC (Cfa)	6 months	West	2 actions for 2 panes - SHGC & t(vis): (0.39, 0.5); (0.08, 0.03)	<b>Lee (2012)</b>
Almost avoided blind's need; L=48% influenced by large WWR	80-90 %	Berkeley (Csb)	Year	South	blinds & infinite actions for 2 panes - t(vis): 0.05-0.6	<b>Fernandes (2013)</b>
Visual comfort: 94% "high score"; Energy savings: beaten by manual and integrated control	\	Beijing (Dwa)	10 days	East	number of lights on & blind slat angle	<b>Cheng (2016)</b>
Energy & visual comfort competitive (applying objective function)	60%	Sydney (Cfa), Rome (Csa), London (Cfb)	Solar year	Sun-oriented (depending on Hemisphere)	4 actions - g-value & t(vis): (0.508,0.595); (0.396,0.446); (0.325,0.341);	<b>Favoino (2016)</b>
\	\	Csb (Berkeley)	October	South	4 actions - commercial-based t(vis): 0.01; 0.06; 0.18; 0.6	<b>Wu (2019)</b>
Occurrence of comfort: WPI - 83%; DGP - 95%	33%	Miami (Aw/As), L.A. (Csb), Baltimore (Cwa), Chicago (Dfa),	Year	4	2 actions - SHGC & t(vis): (0.50, 0.72); (0.10, 0.10)	<b>Hoon Lee (2020)</b>
Savings: RL = 11% MPC = 14%	\	(metro station)	3 months (cooling season)	\	frequencies of fans & pumps	<b>Qiu (2020)</b>

### 1.3.4. Reinforcement Learning

#### Peculiarity and use of Reinforcement Learning

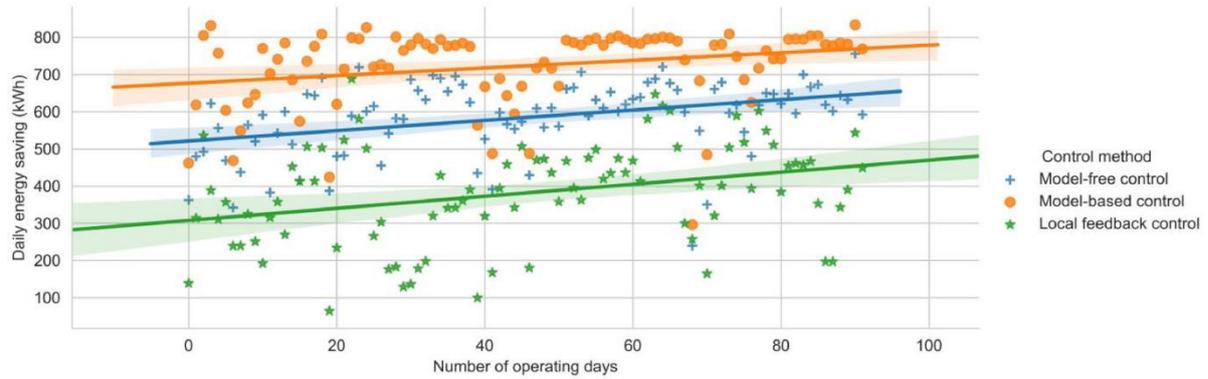
Reinforcement Learning is a particular machine learning method where the reward is the only feedback, there is no supervisor. The agent receives a delayed reward for its action and previous actions have an influence on the reward as well, because they affect the environment. The agent's purpose is to achieve a control policy that increases the cumulative reward. At each time-step, the agent receives a reward, that enhance an instantaneous or delayed policy update. From the environment an observation comes as well, based on which the controller selects an action. The action will affect the environment and then new observation and reward will come. A visualisation of this workflow is Figure 1.16.



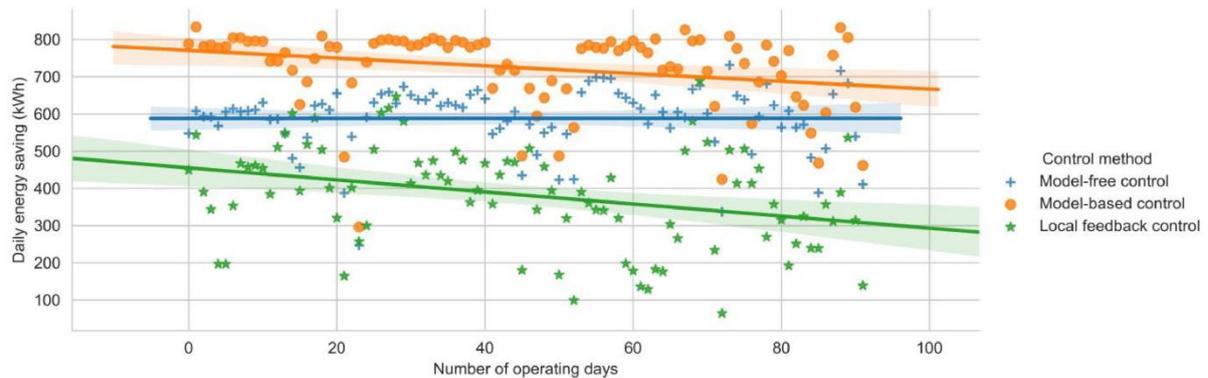
**Figure 1.16** Agent and Environment interaction

This control policy evolves in time to improve performances. On the contrary, deterministic methods based on models mainly independent of time sequence can have saving performances anchored to the input data, without previous experience advantage. This was visible in the learning effectiveness evaluation by Qiu *et al.* [44] taking into account Figure 1.17, where only the Deep RL run did not result in an inverted saving trend when simulated backwards. Another controllers' comparison for HVAC application was about both Tabular Q-learning and Batch Q-learning with Memory Replay (a RL based on ANN) for PV/T, and, after a three-years simulation, all methods achieved a 10% improvement over the RBC baseline [37]. The learning process made also possible self-adaptation to local environment and input variations. Lastly, a statistical comparison between a RL controller and a basic RBC applied to a radiant heating system showed that the former had more than 95% possibility to saves 16.6% more heating energy when

deployed [60]. For the building environment, RBC and MPC are validated control methods, but the reported case studies suggests that RL can be applied to building environments (at least, HVAC systems).



(a). Forward direction simulation (from 19<sup>th</sup> June to 18<sup>th</sup> September)



(b). Reverse direction simulation (from 18<sup>th</sup> September to 19<sup>th</sup> June)

**Figure 1.17** Savings performance by three different controllers simulated chronologically (above) and backwards (below). Figure by [44]

Main components of an RL control can be policy, value function and model. Depending on which and how many components are used, the agent is categorized in different ways, namely value based, policy based, or actor critic, model free or model based.

### Value function and Reward

Value function (1) evaluates future rewards, and its role is to present the likely goodness of the next possible states. Based on these values, the controller applies an action selection.

$$v_{\pi}(s) = E_{\pi} [ R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | S_t = s ] \quad (1.2)$$

The reward is a scalar value to be maximized by the controller and its formulation is a central aspect of control design, especially in multi-objective problems. In fact, in a D-RL control investigation for radiant heating system [43], authors identified local sub-optimal solutions and broad performance variance between solutions with similar state values. This has been justified by the reward design that consisted in two contrasting contribution: Predicted Percentage of Dissatisfied (PPD) and the gap between temperature set-point and Internal Air Temperature (IAT) (to reduce heating demand). Contrasting performance can be both energy-related only and comfort-based. An example is the objective function for an MPC, which have a role similar to the reward in RL. In any case, each contribution to the reward can be associated with a weight in order to shift the focus of the controller.

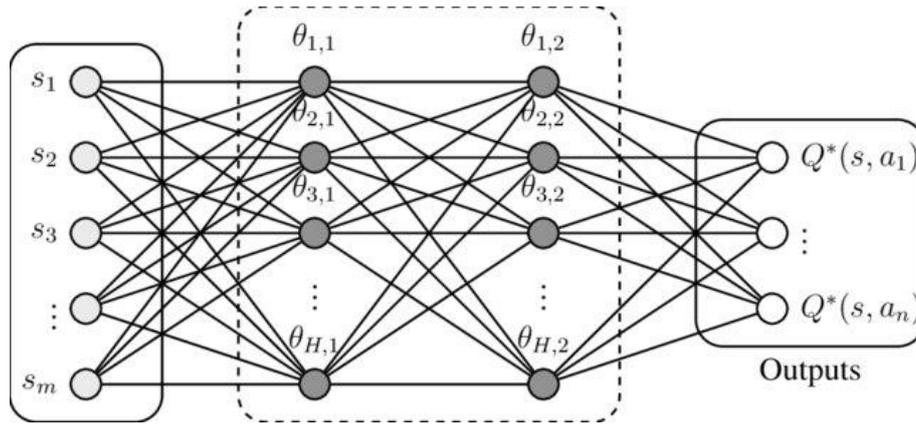
### Q-table and states

The most extensively employed model-free reinforcement learning method is Q-learning, due to its simplicity [61]. In this approach the state/action value is called Q-value and each state-action pair have its own. This state-action space can be represented by a table that store a Q-value in each cell, hence called Tabular Q-learning (TQ-RL), or by Artificial Neural Network(s) (ANN) where weights of the net are upgraded instead of Q-values, hence called Deep Q-learning or Double Q-learning depending on the employment of one or two ANNs. A representation of the former can be seen in Table 1.7, meanwhile Figure 1.18 depicts an ANN for RL.

**Table 1.7** Examples of Q-table in a particular time-step from the present study.

States are here expressed as a tuple with the value of a 4-level input and another binary input.

		Actions			
		0	1	2	3
States	(0,0)	-0.049	-0.01	-0.045	-0.044
	(0,1)	-0.001	-0.006	-0.007	-0.006
	(1,0)	-0.025	-0.005	-0.024	-0.025
	(1,1)	-0.003	-0.007	-0.005	-0.0
	(2,0)	-0.02	-0.02	-0.007	-0.019
	(2,1)	-0.003	-0.005	-0.001	-0.006
	(3,0)	-0.024	-0.015	-0.021	-0.02
	(3,1)	-0.002	-0.028	-0.032	-0.033



**Figure 1.18** Structural scheme of the neural network used for Deep Q-learning from [62]

Because of the nature of ANN, using this approach make more difficult to interpret the resulting policy, meanwhile the Q-table combined with the action selection is equivalent to a set of condition statements in RBC. One major difference between the two representations is the state space: the input variables of an ANN can be any value from a continuous state space. On the contrary, because of the finite number of rows, Tabular Q-learning need discrete inputs, hence continuous domains are to be divided in ranges in order to be represented by the level to which they belong. The amount and values of limits between levels are a control formulation task. Consequently, states of the Q-table are to be carefully selected because of the so-called *curse of dimensionality*. In fact, more states are present, more storage and time are required for the computation but, above all, less chances are given to every state-action pair to be explored. Hence, most relevant states are to prioritize.

### Action selection

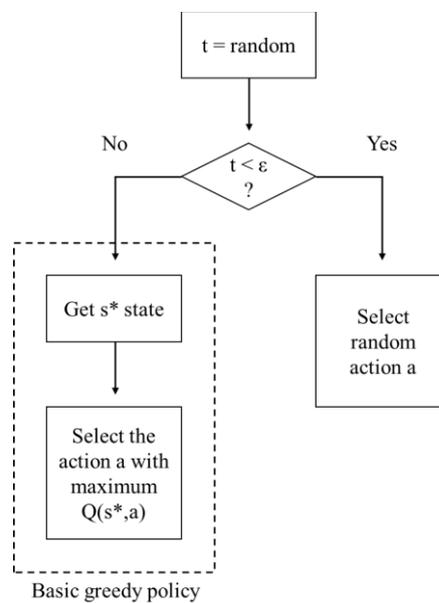
It may be safer to sacrifice instant reward to look for long-term benefits. This can be done with different action selection strategies and hyperparameters tuning. A policy it is a map from state to action and it can be deterministic if directly link the observation to the choice, or stochastic if the received state determines the actions' probabilities of being selected instead.

$$\text{Deterministic policy:} \quad a = \pi(s) \quad (1.3)$$

$$\text{Stochastic policy:} \quad \pi(a|s) = P[A_t = a|S_t = s] \quad (1.4)$$

Reinforcement learning is like trial-and-error learning where action-state values are updated when they occurred. This conceptually yields to the exploration-exploitation dilemma, because the more the agent exploit actions with higher estimated reward, major is the risk to get stuck in local minima and miss a better option for the same state. Therefore, exploration is useful to collect more information about the environment and upgrade the estimation, but usually lead to minor rewards in the short-term.

The greedy policy is the more basic method: based on the received state, the agent looks at the rewards' estimates for each possible action and simply select the action with the highest value. In tabular Q-learning, for example, it receives the state, it reads the correspondent row, find the maximum value and then the associated action. This greatly restrict exploration. The  $\epsilon$ -greedy policy and the Boltzmann action-selection are the most common approaches. The  $\epsilon$ -greedy method is similar to the greedy one but with a previous step: the agent selects a random number between 0 and 1 and then compare it to a threshold: the  $\epsilon$  parameter. If the random number is higher than  $\epsilon$ , then greedy policy is applied, meanwhile a random action is selected if the contrary is true. Thus, at each timestep, there is a probability ( $\epsilon$ ) for an explorative action to be selected. The risk is to make a choice despite its poor expected state/action value because no distinction is taken into account between actions. A representation of these policies is illustrated in Figure 1.19.



**Figure 1.19**  $\epsilon$ -greedy (whole figure) and greedy (framed portion) action selections

A method where an action is more likely picked according to its estimated Q-value is Boltzmann softmax action selection, where the current best action has most chance to be selected (allowing exploitation) and the other options are ranked based on their estimates (reducing the risk of inefficient explorative performance). These probabilities are led by  $\tau$ , called temperature, that can be lowered in time to shift towards a greedier selection.

$$P_t(a) = \frac{\exp\left(\frac{q_t(a)}{\tau}\right)}{\sum_{i=0}^n \exp\left(\frac{q_t(i)}{\tau}\right)} \quad (1.5)$$

Because of the evolution of estimates, Boltzmann action-selection is able to enhance greedy policy as gap between best and second-best action increase. In order to have a diminishing exploration also with  $\epsilon$ -greedy, Qiu *et al.* [44] gave to the greedy action a probability to be chosen increasing at each step, thanks to an addition addend. Excluded this addend, all actions had a probability proportional to their Q-value. The  $\epsilon$ -greedy selection combined with probability calculation is reported in (1.6).

$$\pi(a|s) = \begin{cases} \frac{10 \times \frac{q}{p}}{1 + 10 \times \frac{q}{p}} + \frac{Q(s, a)}{\sum Q(s, a)} \times \frac{1}{1 + 10 \times \frac{q}{p}} & \text{if } a = \arg \max_a Q(s, a) \\ \frac{Q(s, a)}{\sum Q(s, a)} \times \frac{1}{1 + 10 \times \frac{q}{p}} & \text{if } a \neq \arg \max_a Q(s, a) \end{cases} \quad (1.6)$$

### Q-value updating

In model based RL, the state/action value can be predicted from the received states. In model free RL, algorithms revise these values through Bellman equation (2). Beginning from an arbitrary initial state-action values, Bellman equation is iteratively computed. Hence, the initial Q-values should be far from possible final converged values. It is good practice to set them high so that they maintain an high value (hence, are more explored) until are sufficiently lowered during learning phase. In order to give more weight to rewards close in time,  $\gamma \in [0,1)$  is introduced as a discount factor, where low  $\gamma$  means more short-term strategy. The learning rate  $\alpha \in (0,1]$  represent the overriding rate of the current estimate ( $Q_{s,current}$ ) substituted by new one ( $Q_{s,new}$ ), where higher  $\alpha$  usually yields to faster convergence but imply the risk of oscillating estimates. With  $\alpha = 1$ , state/action value would be only the sum between received reward for the action selected based on the previous state

$s$  ( $R_s$ ) and the maximum estimate for the current state  $s'$  ( $Q_{s',\max}$ ) based on which the agent is going to pick an option.

A formulation of Bellman equation is presented here:

$$Q_{s,\text{new}} = (1 - \alpha) \times Q_{s,\text{current}} + \alpha \times (R_s + \gamma \times Q_{s',\max}) \quad (1.7)$$

### Advantages and challenges

A peculiarity of RL that can be both an advantage than a drawback is its learning from scratch that enhance the development of a policy more unbiased by developer. In fact, initial state/action values are commonly more optimistic than possible rewards to let the agent select often an option until the values' updating make the option less desirable. Even though, as in all machine learning techniques, the input variables selection can influence the training. The agent improves its policy step-by-step, this means that it needs an initial period to converge to a good policy and, at the same time, that tends continuously towards an optimal strategy, in the domain of the formulated problem. This continuous learning capability can enhance adaptation to users' behaviours and preferences if the agent is provided with occupant feedback [52].

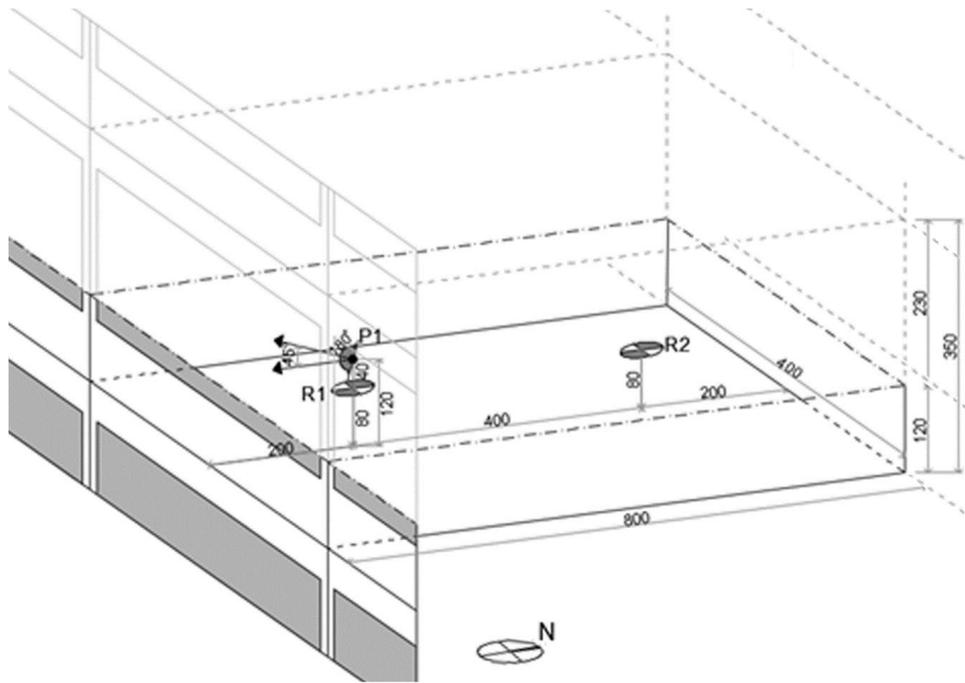
Timing then is an important factor, especially in non-stationary environments. Because of the auto-determination of a policy, RL can be useful to be directly implemented on field and having a fitted control for the specific environment, but in this case, convergence is not only necessary, but it has to occur in reasonable time, in fact some HVAC researches choose to run an offline simulation to the train the agent [43]. To speed convergence, a short time-step can be preferred, so that the agent gains more experience at the same time. But the response of the environment has to be taken into account as well, because if actions can greatly affect it, as in a cooling water systems control [44], rapid sequence of action selection can lead to oscillating behaviour. Lastly, for an environment like a building, that have disturbances according to month and seasons, it can be useful to build different action-space representation for different period, hence giving the opportunity even for different specific control policies. Yang *et al.* [37] tried to implement twelve autonomous learning processes for each month of the year and then combine the “experiences” before the operational phase.

## 2. Methodology

### 2.1. *Virtual environment*

Building Energy Model (BEM) has been widely applied in support to the building design phase, as well as for HVAC optimal control. In order to perform the energy simulation, the implemented thermal model is replicated with EnergyPlus version 8.3.3. EnergyPlus model can be considered certified as the software undergoes two key types of validation tests [63]. The integration of a software like EnergyPlus is utilized for several tasks: controlling the building systems to preserve the set indoor comfort requirements; changing the thermo-optical properties of the adaptive window during simulation runtime according to a defined control strategy; enabling the update of the thermal history of the Q-table in RL according to prior selected actions done by the EC window in the states and rewards framework for the agent. In particular, the “Construction State” permits to switch the glazing thermo-optical properties during the simulation. Making use of this functionality to simulate an EMS actuator for a smart window was validated against experimental records and validated models [64]. Concerning optimal simulation only, Tian *et al.* [65] suggested that EnergyPlus daylight module has a lower accuracy than Radiance in predicting the visual environment when venetian blinds are controlled under different sky conditions. The study also validated Daysim 4.0, a Radiance based simulation software, that received validation by Reinhart and Walkenhorst *et al.* [66] as well. The simulations are carried out through PyCharm, an Integrated Development Environment (IDE) employed for the Python language.

The model characteristics associated with the present case study are the same used by Favoino *et al.* [25] for an investigation on RHC. A schematic representation is available in Figure 2.1. The simulated test case building is an enclosed room with a South-oriented external wall 3.5 m high and 4 m wide, while the depth is 8 m.



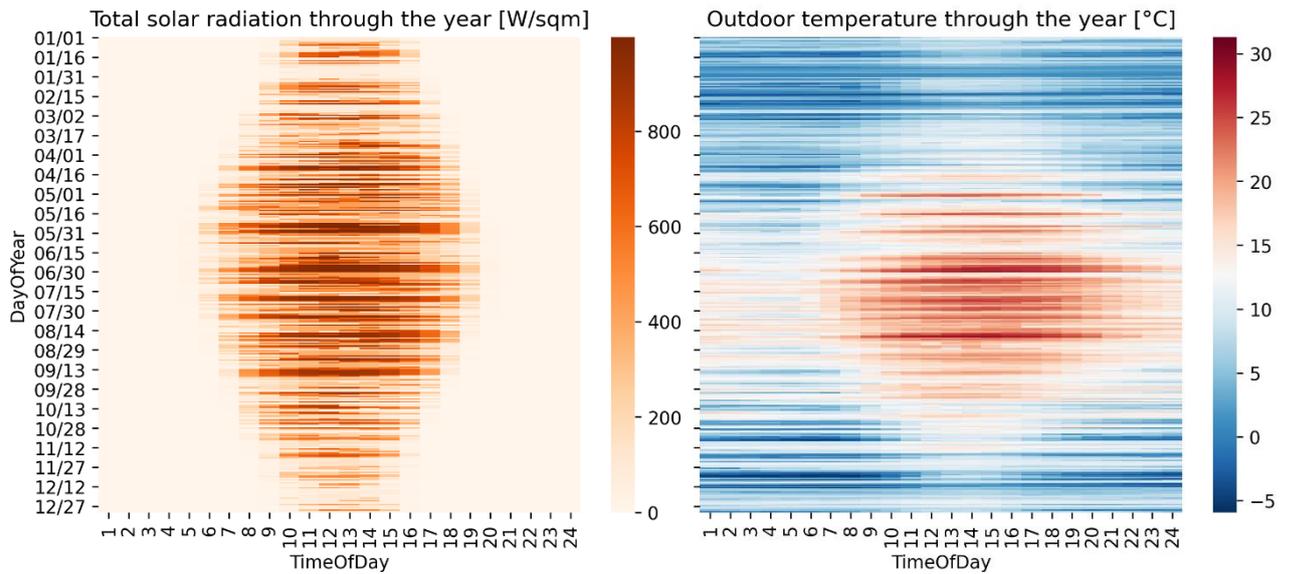
**Figure 2.1** Office test room model (measurements in cm) from [25]

The opaque portion of the envelope is a conventional curtain wall, while the horizontal partitions are formed by concrete slabs (Table 2.1).

Table 2.1 Overall characteristics of the simulated construction in compliance with local regulation.

<b>Construction</b>	<b>Unit</b>	<b>Curtain wall</b>	<b>Concrete slab</b>
Internal thermal capacity	[kJ/m <sup>2</sup> K]	21.7	67.8
External thermal capacity	[kJ/m <sup>2</sup> K]	23.2	29.3
Superficial mass	[kg/m <sup>2</sup> ]	54	675
Time lag	[hrs]	1.63	10.61

The location is London, hence, in a heating dominated climate, which have HDD (18°C of baseline) of 2953 and CDD (10°C of baseline) of 926, labeled Cfb according to Köppen-Geiger classification. An heat map of both Incident solar radiation and External temperature through the simulation period is depicted in Figure 2.2. These have not been used in the design process.



**Figure 2.2** Heat maps of incident solar radiation and external temperature

Opposite to the external side, there is a partition separating the environment from a corridor with the same properties of the test room. All other internal boundaries, namely the floor, the ceiling and lateral partitions, overlook an identical office each. Benchmark test office implements a building façade characterized by thermal and optical properties fulfilling the minimum requirements of local regulations [67]–[69]. Therefore, thermal transmittance is  $0.27 \text{ W/m}^2\text{K}$  for opaque wall. Other regulated characteristics are reported in Table 2.1. Materials’ reflectivity implemented for the different building elements are: 0.5 for walls and partitions, 0.2 for floor and for the external ground, and 0.8 for ceiling.

Because of the focus on EC devices as adaptive building element to control, window’s properties are separately reported. Its geometry set a Window-to-Wall-Ratio (WWR) of 60%. The glazing, like the opaque portion of the envelope, is compliant to the minimum requirements according to the national standards, hence  $2 \text{ W/m}^2\text{K}$  of thermal transmittance. Table 2.2 reports instead the optical properties’ configuration for the 4 possible actions of all the implemented controllers.

**Table 2.2** Smart glazing properties of the four selectable states.

State	$\tau_{\text{vis}}$ [-]	g-value [-]
3	0.595	0.508
2	0.446	0.396
1	0.341	0.325
0	0.238	0.238

Both EC glazing control and the building systems affect typical building services related to indoor comfort, namely air temperature and lighting. Regarding indoor temperature, 20°C and 26°C are set-points for heating and cooling respectively, meanwhile nocturnal set-back are 12 °C and 40 °C in turn. For illuminance instead, threshold is 320 lux and is preserved by dimmable artificial lighting when daylight is not sufficient. This value has been suggested by Mardaljevic *et al.* [70] as minimum illumination level and, in the present study, is measured at desk high level (0,8 m) in two points: R1 and R2, at 1,5 and 3,5 m far from the curtain wall respectively (Figure 2.1).

The variance of occupation, and lighting and equipment use is taken into account through peak loads and schedules defined following ASHRAE standard 90.1 [71]. While the office is occupied by 2 people, power density of lighting and equipment are, respectively, 12,75 W/m<sup>2</sup> and 13,45 W/m<sup>2</sup>. Comfortable indoor air temperature is supplied of heating and cooling thanks to a reversible heat pump with average seasonal COPs, 3,5 in winter and 2,5 in summer. Thus, all these services are provided with electrical appliances, therefore the energy consumed by the building is considered a performance indicator without a conversion to primary energy.

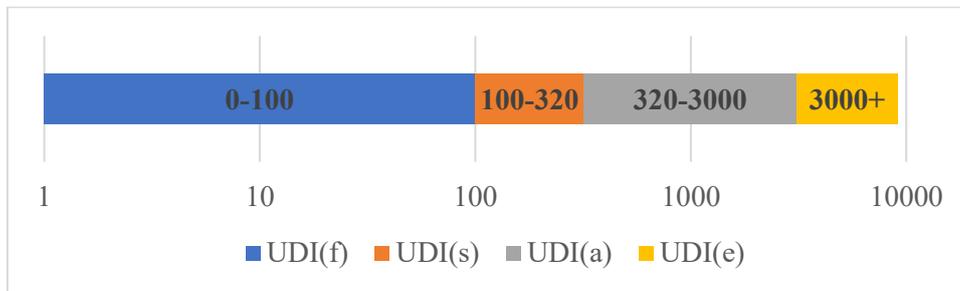
## 2.2. *Indicators and benchmarking*

The aim of this study is to explore the dual adaptability of smart glazing, in particular electrochromic windows, and reinforcement learning control to achieve energy savings, while maintaining or even enhancing visual comfort. These performance goals have been found associated with several research about smart windows, but further investigation on reinforcement learning control could yield to optimize the balance between contrasting objectives. Therefore, three key indicators are used to evaluate the control designs: an energy indicator, an indicator for daylight availability and, lastly, a glare indicator.

Because all services are exclusively provided with electricity, conversion to primary energy is avoided to be independent from national contexts. Hence, energy is measured as the specific amount delivered to the building, neglecting consumptions from site to source. Individual values are available for heating, cooling and lighting. Their sum give the total specific yearly Site Energy (SE) used for comparing the results:

$$SE = E_{heating} + E_{cooling} + E_{lighting} \left[ \frac{kW h}{m^2 y} \right] \quad (2.1)$$

Useful Daylight Illuminance index (UDI) [57], is an indicator of the annual percentage of time when, in a certain portion of space, the illuminance lies between specific illuminance limits, resulting from human comfort research. These limits divide four levels of illuminance (Figure 2.3), in which the two central ones enhance to reduce both artificial lighting consumption and occupant discomfort. In particular, UDI<sub>a</sub> represents autonomy, while UDI<sub>s</sub> indicates a supplementary contribution that still needs artificial lighting. On the contrary, the bottom and the top levels lack of an overall positive effect on visual comfort. The former (UDI<sub>f</sub>) indicates a negligible contribution from daylight, hence making artificial lighting the only reliable source, meanwhile the latter (UDI<sub>e</sub>) indicates an excessive illuminance that raises the probability for glare occurrence [72].



**Figure 2.3** Qualitative representation of UDI bands

Therefore, the sum of the central positively comfort-correlated hourly levels only is computed to represent the percentage of time in which good natural lighting is occurred:

$$UDI = UDI_s + UDI_a = \frac{\sum_{i=1}^{8760} t(100 \leq E_H \leq 3000)}{\sum_{i=1}^{8760} i} [\%] \quad (2.2)$$

Even though UDI take into account also discomfort for glare, to specifically address this phenomenon Discomfort Glare Probability (DGP) is adopted in addition. It is an indicator for the probability that direct solar radiation and/or high contrast in the field of view produce visual discomfort [56]. The DGP shows the predicted share of people with visual dissatisfaction resulting in a value between 0 and 1. It is measured at each time-step with the following equation:

$$DGP = 5.87 \times 10^{-5} E_v + 9.18 \times 10^{-2} \log \left( 1 + \sum_i \frac{L_{s,i}^2 \omega_{s,i}}{E_v^{1.87} P_i^2} \right) [-] \quad (2.3)$$

where  $E_v$  is vertical illumination and  $P_i$  is the position index.

As well as UDI, DGP consent to categorize visual performances in levels [72]. Each of the four classes indicates the occurrence of DGP being between different thresholds for at least 95% of the referred period, namely an hour. The different levels and limits are reported in Table 2.3.

**Table 2.3** Daylight glare comfort classes and relative DGP thresholds [56].

Daylight Glare Comfort Class	DGP Threshold
Imperceptible glare	$0.00 \leq DGP < 0.35$
Perceptible glare	$0.35 \leq DGP < 0.40$
Disturbing glare	$0.40 \leq DGP < 0.45$
Intolerable glare	$0.45 \leq DGP$

For this study only the best class (Class A) incidence contributes to represent an indicator of low risk for glare discomfort. In fact, a DGP of less than 0,35 is associated with imperceptible glare.

$$\text{Glare} = \frac{\sum_{i=1}^{8760} t(DGP \geq 0.35)}{\sum_{i=1}^{8760} i} \Bigg|_{occupied} [\%] \quad (2.4)$$

### 2.2.1. Reference RBC and dummy controllers

The described indicators are employed to compare the different tested RL controllers against a basic rule-based controller as a benchmark. Its control strategy is analogue to a controller implemented by Favoino *et al.* [73] investigating control strategies for smart windows on the same test bed. Once received the incident solar radiation, the agent evaluates in which pre-set range the state lies and select the associated actions. Incident solar radiation ranges are the following:

**Table 2.4** Conditions for passive RBC action selection relying on incident solar radiation on the façade

<b>Range [W/m<sup>2</sup>]</b>	[0,100]	(100,250]	(250,700]	>700
<b>State</b>	3	2	1	0

In order to evaluate the potential benefits and drawbacks of each glazing state in the present scenario, four independent runs are carried out keeping in each of them a different constant action through the whole year. In analogy with machine learning field, they are called in this study *dummy controllers*. In particular, the letter “D” and the correlated state number identify uniquely each simulation, as clarified in the following table:

**Table 2.5** Identification of dummy controllers, hence agent keeping the same state window through the whole year

<b>Controller code</b>	D-0	D-1	D-2	D-3
<b>Constant state</b>	0	1	2	3

## 2.3. Control problem formulation

### 2.3.1. Shared characteristics of investigated control method

The present study explores a variety of Tabular Q-learning reinforcement learning control strategies, each of which is different from the others in at least one of the following design elements: hyper-parameters, states’ variables, states’ discretization, reward. In this section, explored differences of each element and their encodings are presented in order to make results and discussion more readable and synthetic.

Before to dive into explored design possibilities, the simulation time-step is specified. A short time-step is desired because of the control method, Tabular Q-learning, that demand to sufficiently explore all distinct state-action points to refine the policy. In fact, shorter time-step means more Q-value updates in the same period. But improvements in EC glazing switching time are still under development, many on-field experiments showed that a 15-minute adaptation is feasible. Furthermore, as described in previous chapters, the building environment can be affected by numerous disturbances, several of which can occur in few

minutes, especially weather. In particular, the present problem is based on a dynamic phenomenon like sunlight. However, effects on heating and cooling loads are typically slow and a reward based on a brief time-step could misrepresent the goodness of the agent's choice. For these reasons, every simulation run with a hourly time-step.

### 2.3.2. Reward and Q-learning

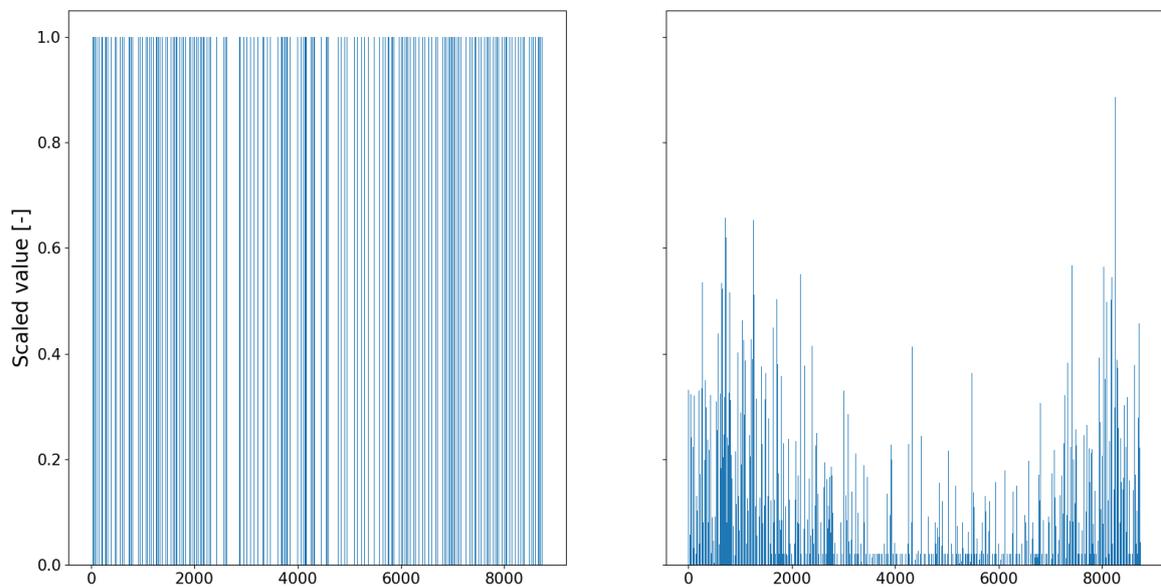
Firstly, the implemented rewards are presented. Since the goals of the control strategy investigation are both energy savings and visual comfort, the same measured parameters used to compute performance evaluation have been tested as factors contributing to the reward function, namely Site Energy, Useful Daylight Index and Discomfort Glare Probability. Once identified the best controller configuration concerning SE only, UDI and DGP are tested as stand-alone rewards with different learning rates to ensure the usability of the index in the present control framework. The tests with DGP had hardly readable effects, therefore the concerning reward is not presented. This still allows to take into account glare thanks to UDI classes, because excluding the highest one the agent is not incentivized to trespass the theoretical visual discomfort limits. In fact, UDI score excludes  $UDI_e$ , other than  $UDI_f$ . Two reward functions are hence reported:  $eR$  focus on energy savings only, while  $euR$  is a linear combination of both SE and UDI scores.  $e-R$  is simply the subtraction of delivered energy for all services (heating, cooling and lighting) to try to minimize overall consumption despite agent's actions can have contrasting effects on distinct service demands.

$$R_e = -SE_t = -E_{heating,t} - E_{cooling,t} - E_{lighting,t} [-] \quad (2.5)$$

Since this reward is never positive, initial Q-values are reset to 0 at the beginning of each simulation. The reason for this is the Bellman equation that allow Q-values' updates. Because each new state-action value can be influenced by all the previous ones and the next one as well, a low initial Q-value can detain the updates at lower scores than their theoretic optimum. This empirically low score of the state-action point can, because of both action selection policies favouring high scores, bring the related action to be explored less often despite its potential benefits.

To build a reward function relaying on both measured energy consumption and visual comfort index ( $euR$ ), the difference between scores' variance is to be addressed. Both SE

and UDI are then scaled between 0 and 1. For UDI, the process is straightforward because at each time step only one of the classes get a 0.25 score. Hence, UDI score was the sum of the two comfort classes (0 or 0.25) multiplied by a scalar factor of 4. Regarding SE, min-max scaling is applied. Minimum is obviously 0, but maximum has not a theoretical limit. A rough research for the maximum possible SE for this specific case study has been carried out applying the already presented dummy controllers. Comparing all the SE scores at each time-step for each of these four simulations, D-3 had the higher SE equals to 0.0289. Therefore, in the euR framework, SE of each time-step is divided by 0.03. An illustration of UDI and SE scaled scores over a whole simulation is presented in Figure 2.4.



**Figure 2.4** Scaled UDI (left) and SE (right) scores for each time-step (x-axis) of the same RL control test.

Finally, the two scaled reward components are weighted using  $w_{\text{UDI}} \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ , a factor focused on UDI, meanwhile SE was multiplied by the resulting difference between 1 and  $w_{\text{UDI}}$ .

$$R_{e,u} = -SE_{t,\text{scaled}} \times (1 - w_{\text{UDI}}) + \text{UDI}_{t,\text{scaled}} \times w_{\text{UDI}} \quad (2.6)$$

$$SE_{t,\text{scaled}} = \frac{SE_{t,\text{scaled}}}{0.03} \quad (2.7a)$$

$$\text{UDI}_{t,\text{scaled}} = (\text{UDI}_{a,t} + \text{UDI}_{s,t}) \times 4 \quad (2.8b)$$

The same reasons for the initial Q-value in *eR* tests are applied to *euR*. In particular, because of the positive domain of UDI score, a positive value is to be selected. Furthermore, an appropriate initial score should be far from possible rewards but not too far. This is a trade-off between a slow correction of the initial value (more distance means more selection of the state-action point to converge) and a biased selection in the learning phase. Thus, because the positive UDI score is influenced by its weight, Q-values at the beginning of a *euR* simulation are set equal to 2 times UDI weight.

An ultimate specification is needed about the implemented Bellman equation. The roles of learning rate ( $\alpha$ ) and discount factor ( $\gamma$ ) have already been discussed in 1.3.4. Because of the variety of scenarios in which the EC window acts, *curse of dimensionality* is a risk in this research. Furthermore, the simulations cover a yearly period, hence presenting a changing environment at which the agent have to adapt sufficiently fast. For these two reasons,  $\alpha$  is set to 1. Regarding the discount factor, initial tests with *eR* and state space suggested a low  $\gamma$  value for higher energy savings. This could be related to the already cited high-disturbed situation, for which a switch in optical properties of the EC window slightly affect the complex building environment. Thus,  $\gamma$  is set to 0,5 for all simulation. Further investigation of these hyper-parameters has not been conducted to maintain a low degree of variables' difference between the numerous tested control methods.

### **2.3.3. Q-table: input variables and states**

The reward functions return a singular value, based on which the correlated state-action value get updated. The present study explores the Tabular Q-learning method for reinforcement learning, thus the state-action space is represented by a table, called Q-table, where each row corresponds to a state and columns refer to actions. Actions are always the four possible state presented in Table 2.2. States, instead, are a central design element to be explore in this research. One reason for many following simplifications in the states design is the already described curse of dimensionality that affects the present RL approach. Five input variables returned by the environment and two indirect variables have been tested. The direct environmental variables are overall incident solar radiation (IR), internal (IT) and external (ET) temperature, occupancy (OCC), season (SSN). The two additional variables represent the forecast for overall incident solar radiation (IRF) and external (ETF) temperature.

Only OCC and SSN are tackled as binary state in this study. In particular, OCC is 1 if any number of people is in the room for any amount of time since the previous time-step, 0 otherwise. OCC could have had more levels (for example, test bed office has two workstations), but potentially highly dynamic occupants' behaviour could bring to more heterogeneous exploration and, furthermore, we believe the agent acting as the room is occupied (without differentiate diverse occupancy degree) even if the office is quickly visited is a good approach because its action is always delayed compared with the received inputs. The other binary state is SSN which can be 0 for the cold season or 1 for the hot season. The boundaries between the two main seasons are the first time-steps of 20<sup>th</sup> March and 23<sup>rd</sup> September, chosen because they are equinoxes. It is worth noting that this qualitative information could be symbolized by other factors. For example, we can expect that the indoor air temperature, influenced by seasonal set-points, represent a correlation with the period of the year. Another observation is that, even if these binary variables can have only two levels, when they are combined with other measures in states, each additional 2-level input double the Q-table size, conceptually providing two distinct Q-tables for the two scenarios.

Unlike OCC and SSN, all other input variables are approached as part of continuous domains to be discretized in levels. The disadvantages of a large number of levels have already been portrayed. Because of those disadvantages, a task of the present control design is to find a compromise between a fast and accurate convergence and a precise control policy identification that enhance visual comfort and energy savings. Keeping low the number of levels make finding significant limits essential.

Overall incident solar radiation (IR) is the total amount of diffuse and direct incident solar radiation on the external façade. Because of its common use in smart windows control, this is the first input variables that have been explored. Different Q-tables have been tested taking into account IR only, dividing the measures in 4 up to 6 levels, with different limits. From the literature review, it is possible to note that in Baltimore, where the climate is similar to the London one (Cwa and Cfb Köppen class respectively), an optimal value to be implement as singular threshold for an RBC was about 370-400 W/m<sup>2</sup>. Other warmer climates yielded to a maximum threshold of 415.3 W/m<sup>2</sup>. Favoino et al. [25] reported threshold from several studies: 95 to 315 W/m<sup>2</sup>; 100 to 850 W/m<sup>2</sup>; 150 W/ m<sup>2</sup>; 189, 315 or 630 W/ m<sup>2</sup> depending on the WWR; 200 W/ m<sup>2</sup>; 250 W/ m<sup>2</sup>; 350 W/ m<sup>2</sup>. For this reasons,

the bottom level has never been narrower than 0 up to 100 W/m<sup>2</sup>. With intermediate limits at 150, 200, and 250 W/ m<sup>2</sup>, the minimum threshold for the top level is tested at 300 and 500 W/ m<sup>2</sup>. The wider gap between the latter values is chosen because the top state can represent when the daylight availability is excessive for visual comfort, hence desired only to save heating thermal energy. Once identified the best discretisation for IR, all the other input variables presented below are tested in different combinations always including IR.

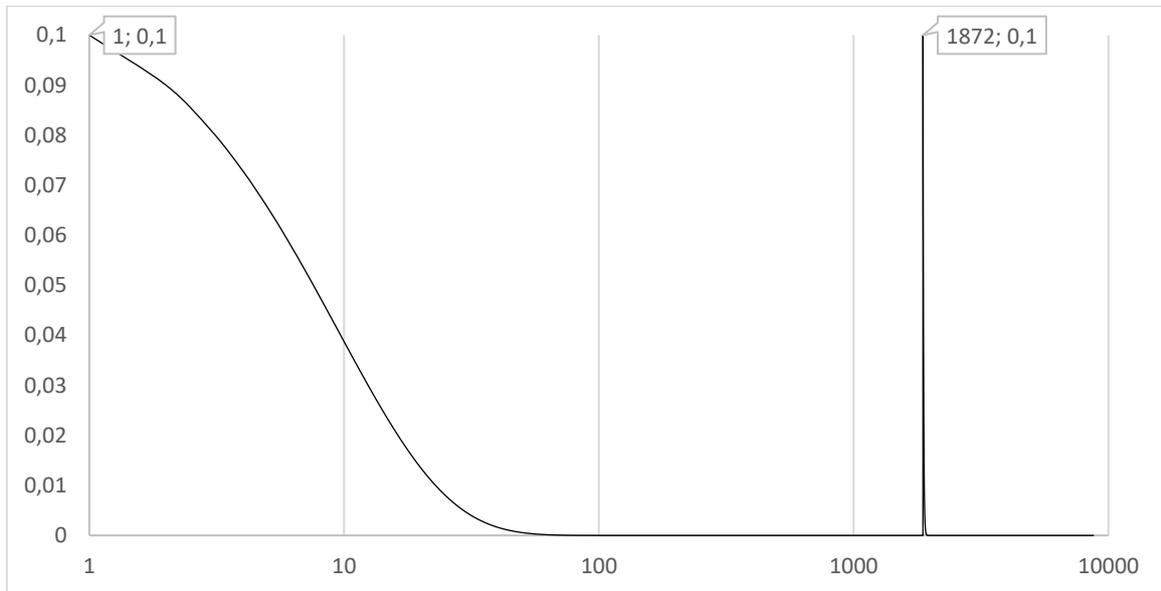
Internal temperature (IT) is explored framing its role as describing the season or the lack/occurrence of comfort temperature bands. Limits of 18°C and 26°C are then implemented, testing both or one at time, in combination with IR. External temperature (ET) is explored framing its role as disturbance of the internal thermal comfort. Therefore, limits corresponding to setpoints are used, namely 20°C and 26°C, and an additional limit of 10°C is introduced to differentiate between “cool” and “cold” temperature, even if in a simple way.

In the literature review of this study, many control strategies for both HVAC system and adaptive envelope rely on MPC. This control method can predict disturbances and use this information to compute an optimal policy. The present RL agent is not able to do so, but a connection with online database of forecast weather could provide similar additional input variables. In the simulation framework, relying on a climate database to reproduce external environment conditions, it is possible to access these parameters with virtual precision. Hence, the following input parameters are based on data to be considered as better than actual forecast available on field. The two prediction of incident solar radiation (IR-F) and outdoor temperature (ET-F) want to be a measure for the imminent future disturbances and potentially represent the beginning and ending of the day. They are implemented through the sum of their relative next 4 hours values. Possible limits to identify their levels are four times the limits tested for the corresponding environmental input.

#### **2.3.4. Action selection**

Finally, after the introduction of the Q-table design, the action selection method is presented. Between the three most diffuse methods presented in 1.3.4, greedy is excluded because the high disturbance level of the present problem demands exploration.  $\epsilon$ -greedy instead has been highly tested applying the framework described in Figure 1.19. In order

to reduce in time the degree of exploration carried out by the agent, the initial value of  $\epsilon = 0.1$  is constantly decrease by a factor of 0.1. This reduction is interrupted at the beginning of the hot season, when  $\epsilon$  is re-initialized at 0.1. Once the hot season end,  $\epsilon$  restart from the last value used in the cold season and continue to decrease of 10 % until the end of the simulated year.



**Figure 2.5** Evolution of  $\epsilon$  through the simulation, by time-step

This approach wants to increase flexibility of the learning process for the upcoming potentially opposite environment. Softmax has also been investigated but never found an optimal policy. Its exploration behaviour has been limited down to a temperature  $\tau = 0.005$  with a decaying factor of 0.1 at each time-step but all tested agent still switched the window state almost continuously.

Lastly, the simulation process is described in the following pseudocode in order to provide a comprehensive overview of the relation between the Reinforcement Learning elements, as identified in this section:

```

INITIALIZE hyper-parameters and settings
SET Q-table design (variables to be read and limits between levels)
INITIALIZE Q-table entries
INITIALIZE  $\epsilon_{hot-season}$  and  $\epsilon_{cold-season}$  equals to 0,1
START EnergyPlus (First hour of 1st January)

```

```

SET first state as previous state
REPEAT :
    GET observation from environment (input variables and indicators)
    COMPUTE Reward
    READ last state
    READ last action
    UPDATE Q-value corresponding to last state-last action entry
    IF timestep is between hot season limits THEN:
        SET  $\varepsilon$  equal to  $\varepsilon_{hot-season}$ 
        SET  $\varepsilon_{hot-season}$  equal to  $\varepsilon - 10\%$ 
    ELSE
        SET  $\varepsilon$  equal to  $\varepsilon_{cold-season}$ 
        SET  $\varepsilon_{cold-season}$  equal to  $\varepsilon - 10\%$ 
    SET  $p$  as random number between 0 and 1
    IF  $p < \varepsilon$  THEN
        SELECT random action
    ELSE
        SELECT the action with the maximum Q-value, between those
        corresponding to last state
    SET window properties in EnergyPlus according to selected action
    SET previous state equal to last state
UNTIL First hour of next 1st January

```

```

INITIALIZE hyper-parameters and settings
SET Q-table design (variables to be read and limits between levels)
INITIALIZE Q-table entries
INITIALIZE  $\varepsilon_{hot-season}$  and  $\varepsilon_{cold-season}$  equals to 0,1
START EnergyPlus (First hour of 1st January)
SET first state as previous state
REPEAT :
    GET observation from environment (input variables and indicators)
    COMPUTE Reward
    READ last state
    READ last action
    UPDATE Q-value corresponding to last state-last action entry
    IF timestep is between hot season THEN:
        SET  $\varepsilon$  equal to  $\varepsilon_{hot-season}$ 
        SET  $\varepsilon_{hot-season}$  equal to  $\varepsilon - 10\%$ 
    ELSE
        SET  $\varepsilon$  equal to  $\varepsilon_{cold-season}$ 
        SET  $\varepsilon_{cold-season}$  equal to  $\varepsilon - 10\%$ 
    SET  $p$  as random number between 0 and 1
    IF  $p < \varepsilon$  THEN
        SELECT random action
    ELSE
        SELECT the action with the maximum Q-value, between those
        corresponding to last state
    SET window properties in EnergyPlus according to selected action
    SET previous state equal to last state
UNTIL First hour of next 1st January

```

### 3. Results and Discussion

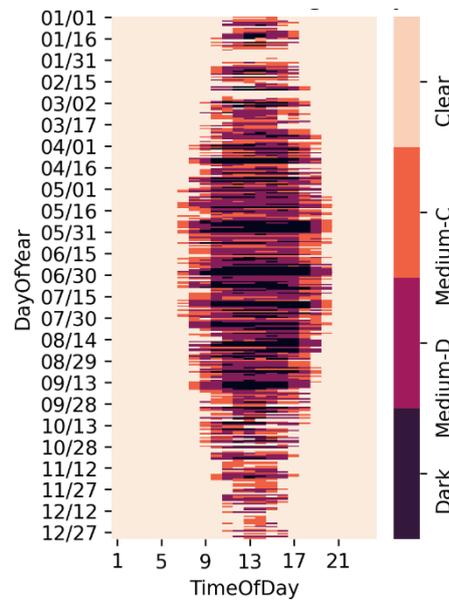
The aim of this study is to compare applicability of a Reinforcement learning agent in managing electrochromic glazing states against a consolidated control method, Rule based control. For all conducted RL tests, the results of the already described RBC are employed as baseline to compute the percentage of difference. All these ratios indicate a better performance when positive, hence a lower score on SE or Glare results in a correlated positive percentage, on the contrary, a lower UDI results in a negative percentage of difference.

Various illustrations are used to present the behaviour and performances of RL trials. Q-table is a fundamental element of a Tabular RL. The entries are presented in a colour scale based on their state-action values. This qualitative representation allows to track usefulness of state levels, in fact it is possible to evaluate for each state if there is a clear better action or if the 4 Q-values are analogous, and to weigh soft or hard distinctions between similar states, thus if they result in the same policy or which of them is more explored. Another description of the agent's behaviour is the carpet plot of its actions through the whole yearly simulated period. This gives an impression of the number of switches performed by the controller and the convergence speed. Both representations can show also seasonal control strategies, in fact some images report Q-tables depicted in different moments of the year.

Because the goal is to research a good control method, comparing performances of different simulations is essential. Therefore, two slightly different charts are used, depending on the relevant focus. When the energy focus is more important, a simple bar chart depicting absolute SE for each service is presented with an additional series to indicate the total SE. In order to consider visual comfort as well, series for UDI and Glares indicators are added. Both visualisations allow to spot strengths and weaknesses for different objectives in each compared controller. Performances for both energy consumption and visual comfort are always reported in any case, with the exception of prototype tests on the discount factor.

The results of reference controller and dummy controllers are presented here. Figure 3.1 is the carpet plot depicting RBC policy, which is directly correlated to 4 IR ranges, hence it also provides a classification of the IR through the simulation period. A 4-level

colour scale provide the representation for each of the 4 actions available to the controller. From top to bottom (days) and from left to right (time) is the chronologic order of the simulation. It is possible to see how in the central period of the year, during working hours, RBC heavily rely on the two darker actions. Carpet plots of dummy controllers are not reported because they are of a single colour each.



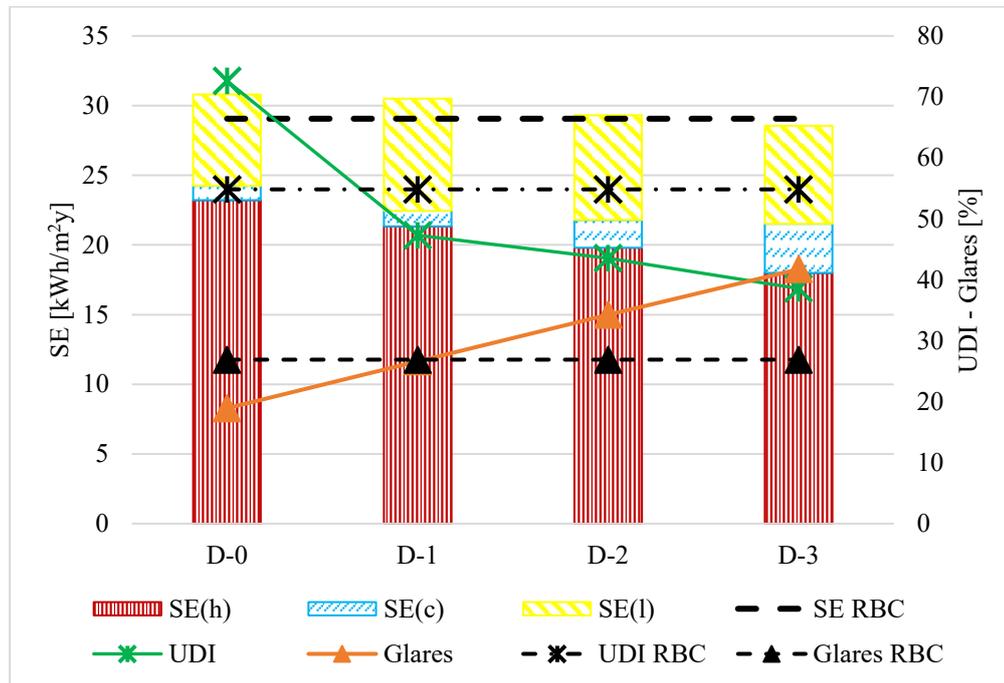
**Figure 3.1** Carpet plot of RBC reference actions

Performances in absolute values of these basic control strategies are reported in Table 3.1.

**Table 3.1** Performances of RBC and dummy controllers.

Baselines	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glares [%]
<b>RBC</b>	20,55	1,16	7,35	29,06	54,83	26,9
<b>D-0</b>	23,21	1,04	6,56	30,8	72,61	18,94
<b>D-1</b>	21,34	1,12	8,04	30,5	47,29	26,61
<b>D-2</b>	19,8	2	7,5	29,31	43,54	34,26
<b>D-3</b>	17,98	3,55	7,03	28,56	38,57	41,73

The joint evaluation of results is framed in Figure 3.2, which shows the cumulative energy consumption divided by service, as well as the UDI and Glares scores.



**Figure 3.2** Energy performances and improvements by service of dummy controllers with different constant action.

Mono-action results provide empirical limits for each performance addressable with a unique action, and strengths and weaknesses derived from their application. As expected, darker states result in more cooling energy savings and for heating the contrary is true. The ranking for total SE is the same that the one for SE(h), even with the poor performance of D-3 in cooling. This was predictable because of the heating-dominated climate of London, which brings to annual heating energy consumption being 10 times the cooling one. But clearest states greatly lack in visual comfort, resulting worse than RBC in all the indicators implemented in the present study. The D-3 simulation even provide almost twice the glare occurrence compared to RBC. Hence, it is clear how much energy savings and visual comfort can be contrasting goals in the case under analysis. Lastly, RBC performances are similar to keeping the EC window always in the medium-dark states, with the exception that the former provide more useful daylight as it is shown by gaps in UDI and SE(l)

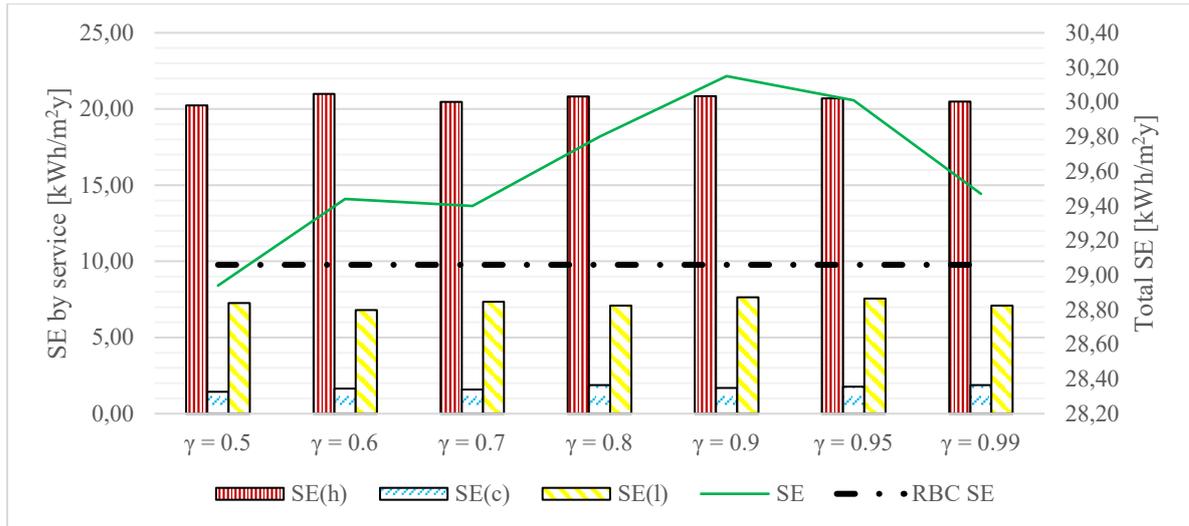
### 3.1. Defining the control strategy core

Before to investigate the design of states, the first tests are oriented to identify a structural parameter: the discount factor  $\gamma$ . Between the available input variables received from the simulated environment, IR and ET are two of the information most applied by reported studies on adaptive transparent envelopes. Hence a simple state space is built employing these measures and setting their levels with a constant step (with the exception of 26°C for ET because of its correlation with indoor temperature setpoint). In particular, for these RL agents, SR limits are 0, 200, 400, 600, 800 W/m<sup>2</sup>, while ET levels have 0, 5, 10, 15, 20, 26, 30 °C as thresholds. Seven different  $\gamma$  are tested: 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 0.99. The former is named  $\gamma_5$ , the second  $\gamma_6$ , and the others with the same approach, according to their significant digits. Results of these seven simulations are reported in Table 3.2.

**Table 3.2** Energy performances of reinforcement learning controllers with different discount factors.

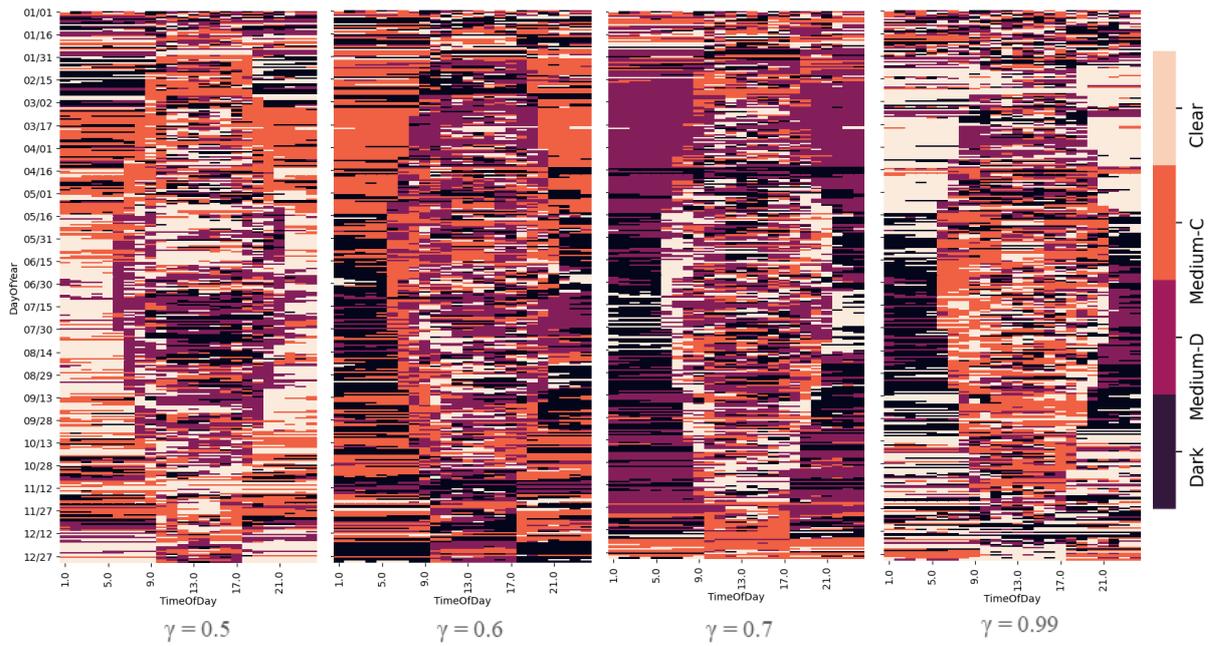
TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]
$\gamma = 0.5$	20,23	1,44	7,27	28,94
$\gamma = 0.6$	21,00	1,64	6,80	29,44
$\gamma = 0.7$	20,46	1,59	7,35	29,40
$\gamma = 0.8$	20,83	1,88	7,10	29,80
$\gamma = 0.9$	20,84	1,68	7,63	30,15
$\gamma = 0.95$	20,69	1,77	7,55	30,01
$\gamma = 0.99$	20,49	1,88	7,09	29,47

Similarly to dummy controllers, the lowest heating SE corresponds to the more energy efficient simulation, in this case the one with  $\gamma = 0.5$ . Despite the good performance on heating,  $\gamma_5$  have a medium SE(h) that places it in fourth place, while it even gets the second-best result in cooling energy savings. Thanks to these achievements, RL agents with  $\gamma$  equals to 0.5 is the only one outperforming the RBC (as it is shown in Figure 3.3, reporting energy-related scores only).



**Figure 3.3** Energy performances of seven reinforcement learning controllers with different discount factors.

Carpet plots of the four most energy efficient algorithms only are reported. From the heterogeneity and boundaries of action selection depicted in Figure 3.4, it is possible to conclude that, despite good overall energy performance, with higher discount factors, in this case 0.7 and 0.99, the RL controller struggle to converge for the first two months approximately.  $\gamma_5$  and  $\gamma_6$  have a similar issue for about one month, but, despite the clearly different control policy, they could be valid alternatives. As anticipated in the methodology section, at the end 0.5 is maintained for all successive tests.



**Figure 3.4** Carpet plots of actions selected by reinforcement learning controllers with different discount factors.

The second principal design element is the total incident solar radiation space on which the RL agent will rely. This disturbance has been selected for its frequent application in case studies similar to the present one. The goal, as for each input variable, is to identify the number of levels in which the measure’s domain can be divided, as well as their boundaries. In order to do so, four different discretization are embedded in as many separate simulations. In Table 3.3 is possible to read the coding of these tests and the limits set for each of them. The names are attributed based on the number of levels and the inferior limit of the higher tier. This choice is justified by the already discussed *curse of dimensionality* linked to the size of the Q-table and the relevance of the brightest state for the various objectives influenced by the EC control.

**Table 3.3** Identification of the four tests about total incident solar radiation discretization.

Code	Total incident solar radiation limits [W/m <sup>2</sup> ]					
	100	150	200	250	300	500
4-500		X		X		X
4-300		X		X	X	
5-500	X		X		X	X
5-300	X	X		X	X	

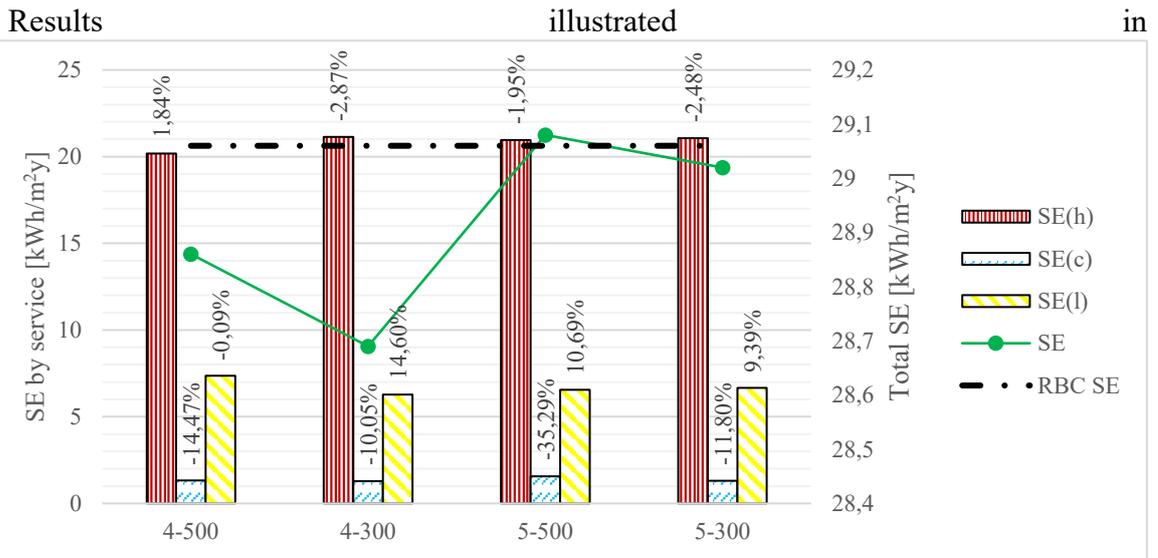
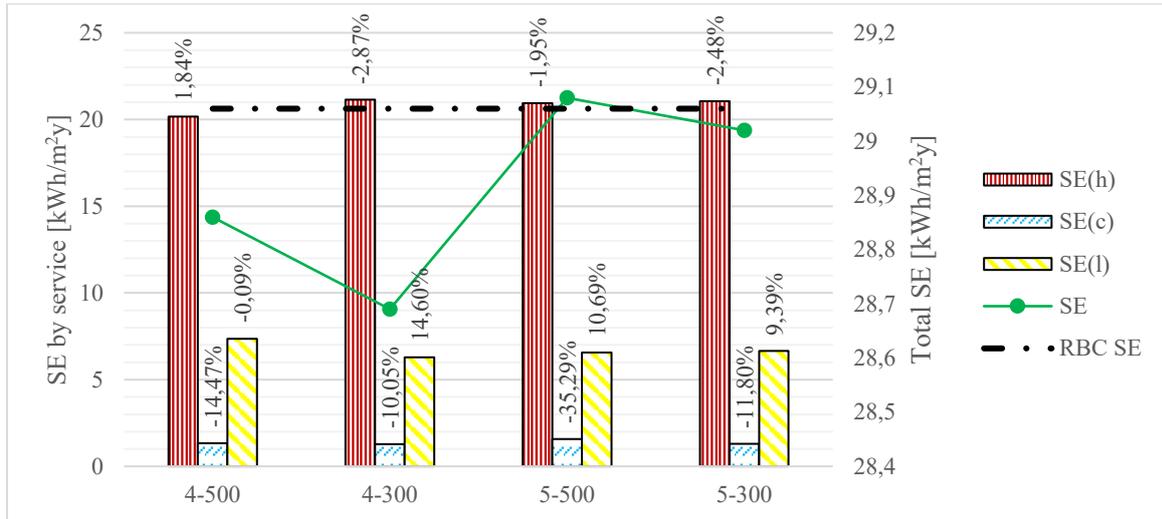


Figure 3.5 shows that agents with  $300 \text{ W/m}^2$  as IR highest limit save more SE than correspondent agents with the same number of levels. 4-500 is the only controller without  $300 \text{ W/m}^2$  as limit and have clearly worst performances in SE(l) and UDI (see also Table 3.4).

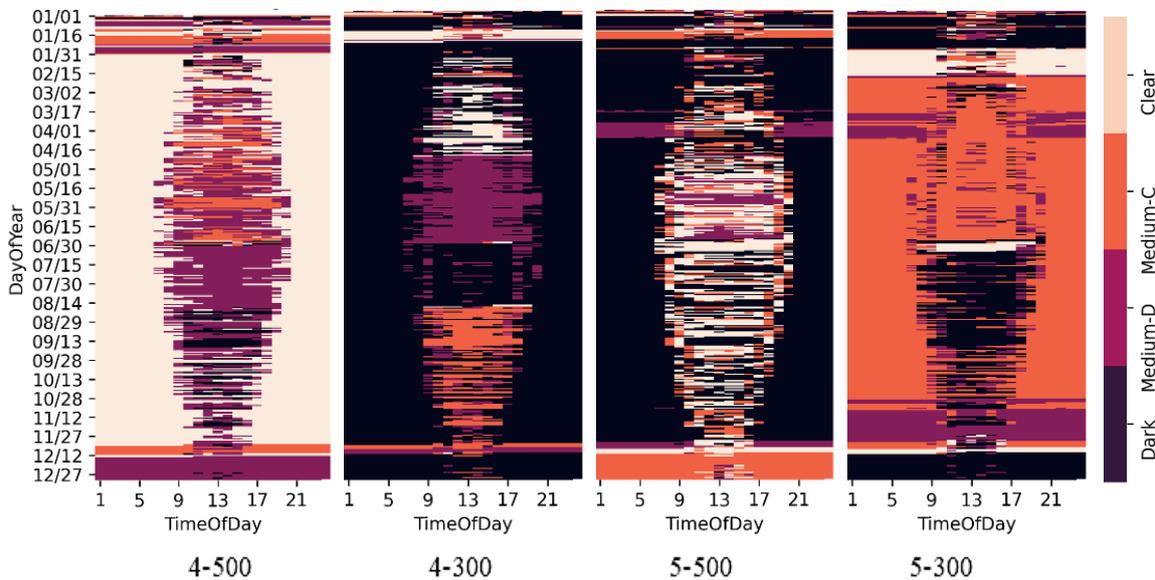
**Table 3.4** Performances of RBC and reinforcement learning controllers with different total incident solar radiation discretization.

TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glares [%]
<b>RBC</b>	20,55	1,16	7,35	29,06	54,83	26,9
<b>4-500</b>	20,17	1,33	7,36	28,86	50,58	34,55
<b>4-300</b>	21,14	1,28	6,28	28,69	57,65	32,97
<b>5-500</b>	20,95	1,57	6,56	29,08	55,46	35,08
<b>5-300</b>	21,06	1,3	6,66	29,02	59,1	33,42



**Figure 3.5** Energy performances and improvements by service of reinforcement learning controllers with different total incident solar radiation discretization, compared to RBC.

Furthermore, its carpet plot (Figure 3.6) shows a restrict application of the darkest window's state. These observations suggest that 300 is a more significant value for IR compared to 500, enhancing to better regulate excessive incoming daylight. 4-300 and 5-300 make larger use of darker states and this provide higher cooling energy savings than other agents, even if still less energy efficient than RBC. This suggests that those controllers could be more adaptable than the others with 500 as threshold for the upper state.



**Figure 3.6** Carpet plots of actions selected by reinforcement learning controllers with different total incident solar radiation discretizations.

In particular, 4-300 achieve the best trade-off between lighting and cooling energy savings that allows to have the lowest SE indicator even with the poorest heating energy performance (overall SE is 1,26% less than RBC). For all these reasons, and also in order to maintain a low number of states predicting the addition of other input variables, the following limits for SR are set for all the following tests: 150, 250 and 300.

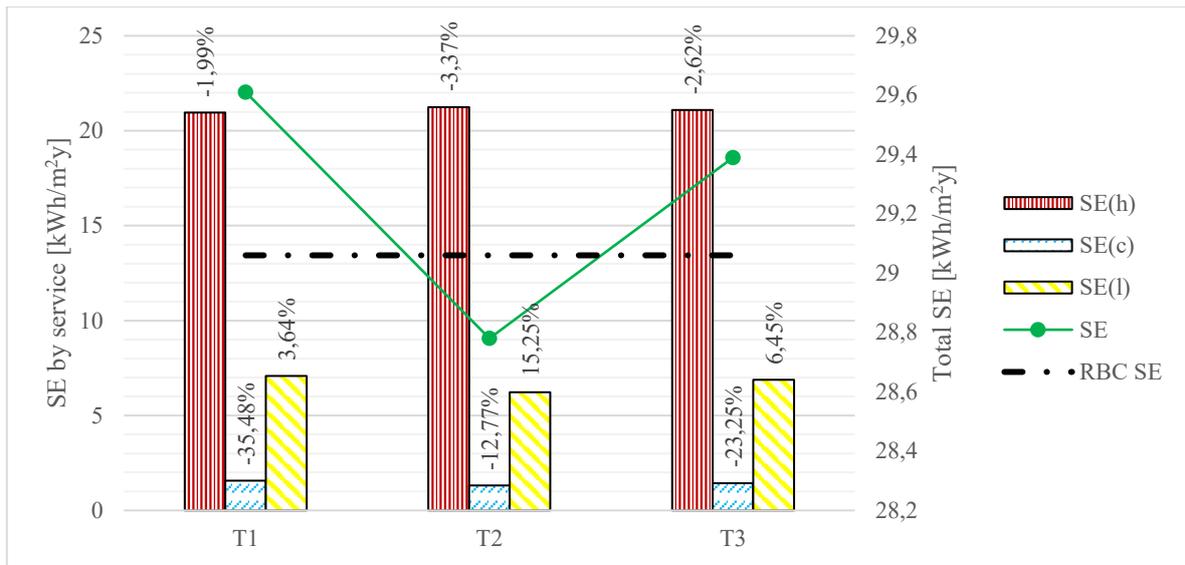
### 3.2. Exploring additional input variables

Once identified the fundamental elements of the RL agent's design, other tests are carried out in order to improve the achieved control method. The second disturbance to be presented is external air temperature. Three simulations with as many different numbers of levels provide the results registered in Table 3.5.

**Table 3.5** Performances of RBC and three reinforcement learning controllers with different external temperature discretizations.

TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glare [%]
<b>RBC</b>	20,55	1,16	7,35	29,06	54,83	26,9
<b>T1</b>	20,96	1,57	7,08	29,61	48,97	31,44
<b>T2</b>	21,24	1,31	6,23	28,78	52,82	30,17
<b>T3</b>	21,09	1,43	6,88	29,39	49,22	30,59

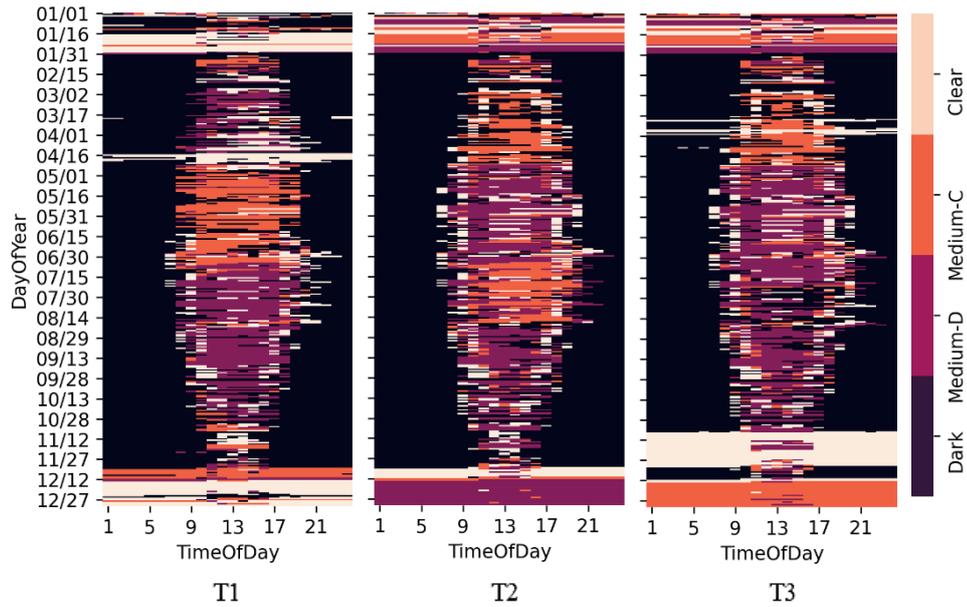
*T1* make use of only one threshold, that is 20°C. In *T2* 26°C is implemented as well as second limit and, finally, an additional tier is introduced to build *T3*, dividing the inferior range in below and above 10°C. Site Energy consumption of the three agents and the RBC baseline are presented Figure 3.7.



**Figure 3.7** Energy performances and improvements by service of seven reinforcement learning controllers with different discount factors compared to RBC total Site Energy (dotted line)

The introduction of ET generally improves lighting consumption, as suggested by the positive percentage of difference achieved by all the controllers, but it has a negative effect on heating load. Despite this effect, *T2* is able to save more overall energy than RBC reference, even if just 0,96%, that is less than the *4-300* agent based in IR only. This better performance of *T2* could be yielded by a faster adaptability, in fact, compared to the others two agents, it has the higher savings both in cooling and lighting (where it saves more than 15% more energy than RBC).

Lastly, the action selection through the year of the three controllers tested with External Temperature is depicted as heat map in Figure 3.8. As in many other cases present in this document, a general observation is proposed: towards the end of autumn, agents tend to drastically rearrange their control policy.



**Figure 3.8** Carpet plots of actions selected through RL for different external temperature discretization.

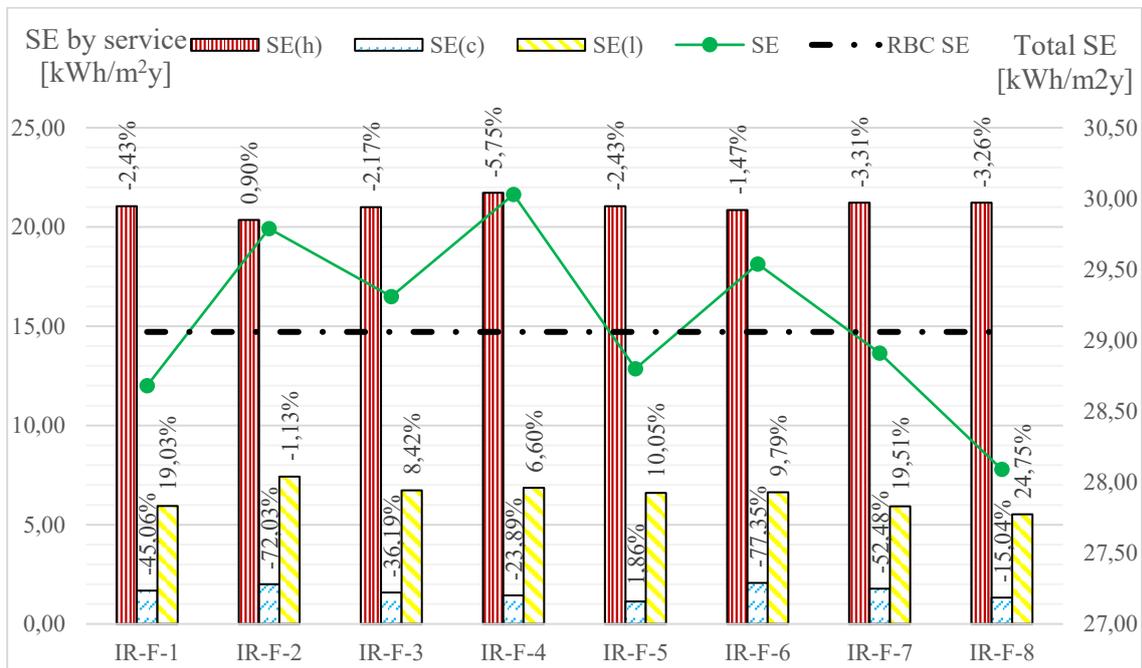
### 3.3. *Exploration of cumulative future disturbances as state variables*

In order to investigate further potential savings by a RL agent, forecast information are embedded in the states space. At each time step, the agent received the sum of the next 4 hours of IR or ET. Similarly to previous tests, discretization of incident solar radiation *forecast* (IR-F) and external temperature *forecast* (ET-F) is studied by varying the numbers of levels, but tested values are not designed to mimic a-priori knowledge but to create a correlation between the in-time variable and its forecast. In practice, limits value for IR-F are always 4 times some of the consolidated thresholds for IR, or 4 times the limits tested for ET. The coding of the different *IR-F* agents is synthetized in Table 3.6, with the exception of the last approach where the limit value of 600 is substituted by 400.

**Table 3.6** Identification of the four tests about total incident solar radiation forecast discretization.

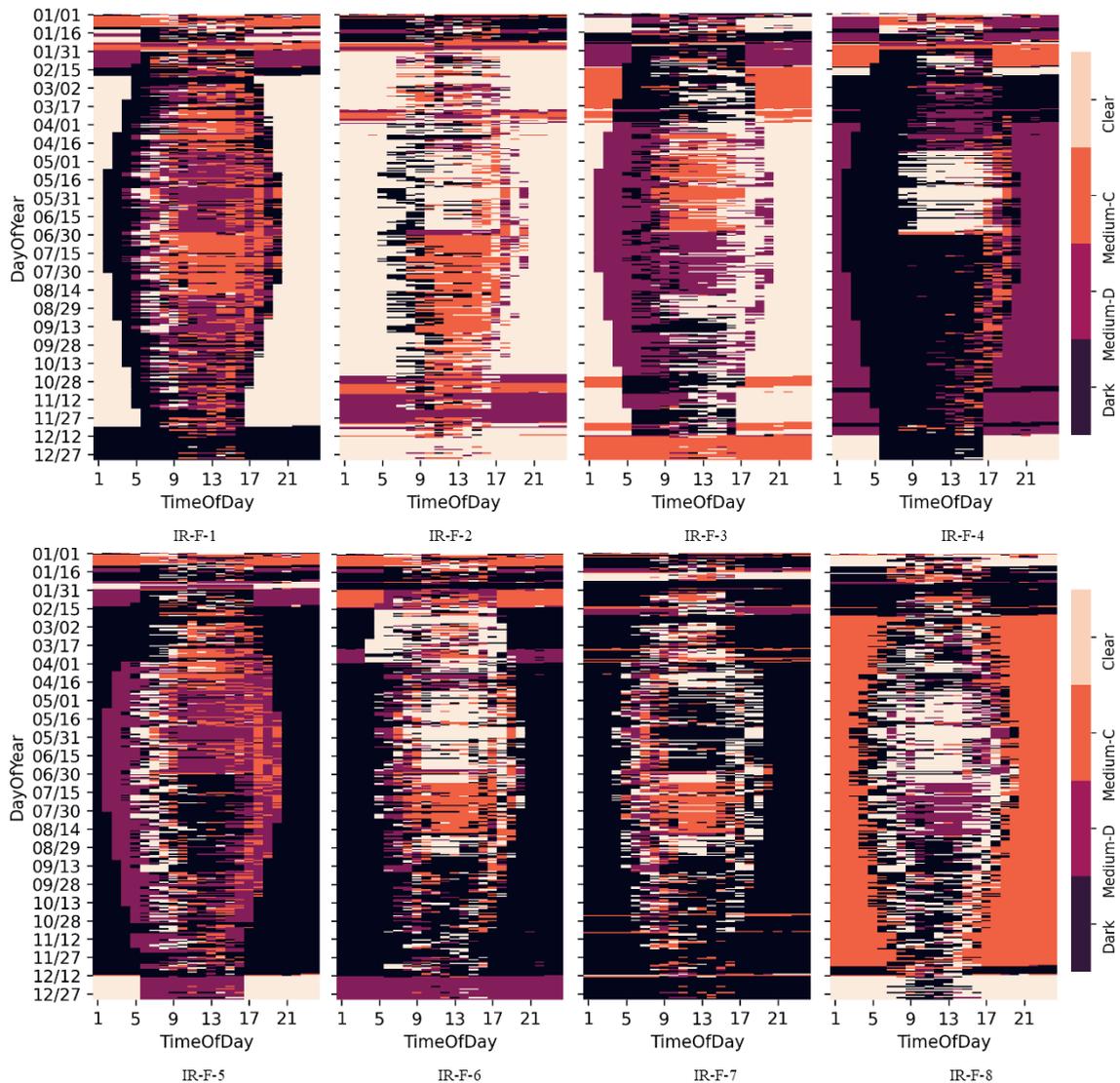
Code	Cumulative Total incident solar radiation limits of next 4 hours [W/m <sup>2</sup> ]			
	0	600	1000	1200
IR-F-1	X	X		X
IR-F-2		X		X
IR-F-3	X			X
IR-F-4	X	X		
IR-F-5	X	X	X	
IR-F-6	X	X	X	X
IR-F-7		X	X	X
IR-F-8		<b>400</b>	X	X

In general, the introduction of this input variable led to good lighting performances, in fact Figure 3.9 shows lower consumption in this service for all the simulation except for IR-F-2, probably due to a poor states-space design.



**Figure 3.9** Energy performances and improvements by service of reinforcement learning controllers with different cumulative total incident solar radiation discretization, compared to RBC

Looking at the carpet plots of all agents in fact (Figure 3.10), it is evident how the limit at  $0 \text{ W/m}^2$  yield to a clear change in the policy at the beginning and at the end of the day. This is not valid only for test *IR-F-6* which, despite having a differentiation between positive and negative future cumulative IR, does not present a boundary in the action selection. This could be due to an excessive number of levels (*IR-F-6* is the only 20-states agent tested with IR-F) that makes the next maximum action value to vary more often (see (1.7)).



**Figure 3.10** Carpet plots of actions selected by reinforcement learning controllers with different cumulative incident solar radiation discretization.

This weak policy learning is correlated with poor energy savings, where the only indicator lower than reference is SE(l) which is however an average result in for particular Q-table approach, as reported in Table 3.7. Hence, these observations reinforce the idea that more information does not mean better choices by a Tabular Q-learning agent.

**Table 3.7** Performances of RBC and reinforcement learning controllers with different cumulative incident solar radiation discretization.

TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glare [%]
<b>RBC</b>	20,55	1,16	7,35	29,06	54,83	26,9
<b>IR-F-1</b>	21,05	1,68	5,95	28,68	49,80	31,33
<b>IR-F-2</b>	20,36	2,00	7,43	29,79	48,59	33,56
<b>IR-F-3</b>	21,00	1,58	6,73	29,31	50,87	33,65
<b>IR-F-4</b>	21,73	1,44	6,86	30,03	59,12	27,33
<b>IR-F-5</b>	21,05	1,14	6,61	28,80	60,76	27,53
<b>IR-F-6</b>	20,85	2,06	6,63	29,54	53,13	33,34
<b>IR-F-7</b>	21,23	1,77	5,92	28,91	59,06	28,40
<b>IR-F-8</b>	21,22	1,33	5,53	28,09	56,95	30,23

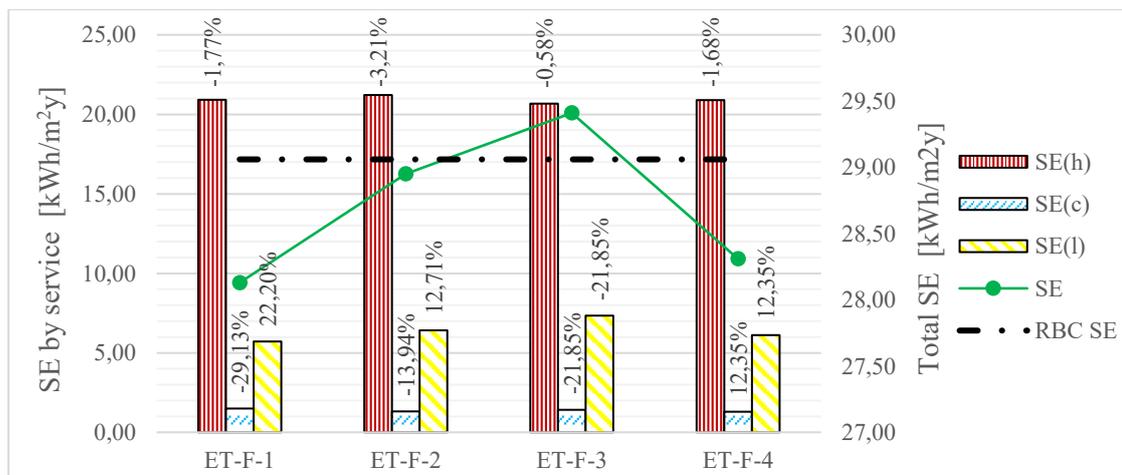
Another import level to maintain is the one above 1000 or 1200, in fact the *IR-F-4* carpet plot suggests that the agent ends up selecting the darker states in most cases when IR-F is above 600 W/m<sup>2</sup> as last limit. But 600 appears to be a significant value as well, because in the *IR-F-3* test the controller have difficulty to converge in a policy for the period between mid-August and mid-September: its carpet plot shows a large application of the clearest state to return to the darkest one after and this could bring to a poor cooling management and a higher glare risk as well. Lastly, a 4-states discretization for IR-F is investigated getting rid of the threshold of 0. Because implementing 600, 1000 and 1200 as limits yielded to good performances, the minor threshold is shifted towards zero selecting 400 instead (that is 4 times 100, one of the tested limits for IR). This last test with the cumulative IR achieved a 3,35% reduction in overall energy consumption compared to the reference RBC and the best lighting performance between the IR-F simulations. The other agent to keep into account is *IR-F-5* because, despite having just a 0,89% improvement (intended as described at the beginning of this chapter) in SE, it is the only one able to perform better than the baseline in consumption for cooling.

Concerning the use of the cumulative sum of External Temperatures, the coding of different *ET-F* agents is synthesized in Table 3.8.

**Table 3.8** Identification of the four tests about total incident solar radiation forecast discretization.

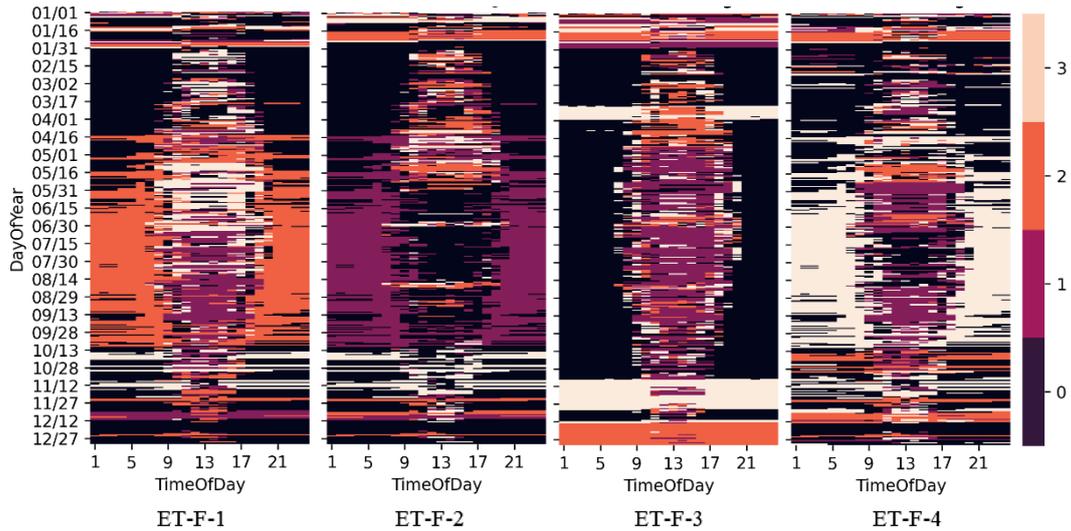
Code	Cumulative Total incident solar radiation limits of next 4 hours [W/m <sup>2</sup> ]			
	0	40	80	104
ET -F-1		X	X	
ET -F-2		X	X	X
ET -F-3			X	X
ET-F-4	X	X	X	

Looking at Figure 3.11, they seems to rely on lighting savings to reduce SE, as well as controllers provided with IR forecasts. However, they perform better in heating consumption, even if reference RBC still reduce heat gains more than all of them.



**Figure 3.11** Energy performances and improvements by service of reinforcement learning controllers with different cumulative external temperature discretization, compared to RBC

*ET-F-1* achieved the minor SE and SE(l) as well, in fact its carpet plot (Figure 3.12) appears clearer than others, especially in the central part of the day at the end of spring, when there is a good sunlight availability but neither heating nor cooling loads have a major impact on temperature control.



**Figure 3.12** Carpet plots of actions selected by reinforcement learning controllers with different cumulative external temperature discretization.

This suggests that knowing if the next hours external temperature will be on average below/above 10 it's a significant information. In fact the worst performing agent lacks of 40 as threshold. Table 3.9 reports 29,41 kWh/ m<sup>2</sup>y as SE score for *ET-F-3*.

**Table 3.9** Performances of RBC and reinforcement learning controllers with different cumulative external temperature discretization.

TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glares [%]
<b>RBC</b>	20,55	1,16	7,35	29,06	54,83	26,9
<b>ET-F-1</b>	20,91	1,50	5,72	28,13	51,14	32,11
<b>ET-F-2</b>	21,21	1,32	6,42	28,95	60,24	24,42
<b>ET-F-3</b>	20,67	1,41	7,33	29,41	47,9	30,9
<b>ET-F-4</b>	20,9	1,3	6,11	28,31	55,19	28,51

The 4-times-26 limits led to second-rate savings and is not even able to yield to cooling loads lower than similar agents, hence, it is judged as negligible. Finally, the representation of really cold temperatures provided by 0 as limit, enhance the controller to select the clearest action in time-steps when the agent without this additional threshold (*ET-F-1*) selects instead the medium-clear EC window state. This finer selection seems to appear also during “hot-season” days, when *ET-F-1* relies mainly on two actions out of four.

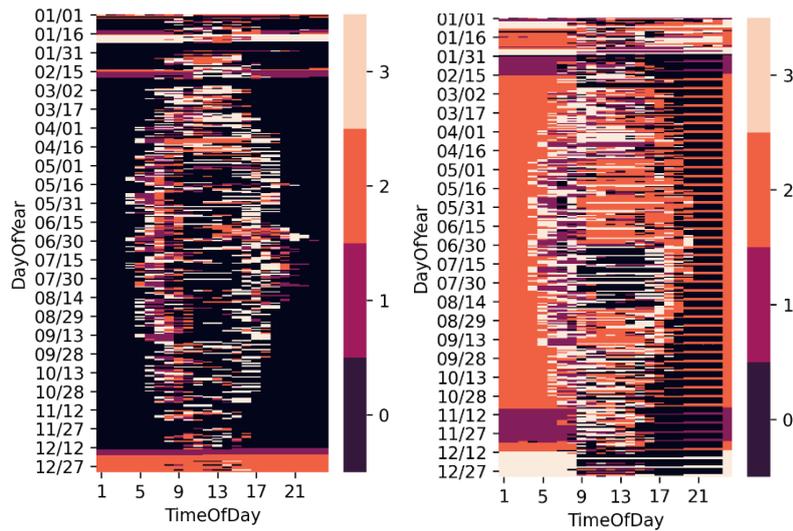
### 3.4. Combination of more additional parameters

Some combinations of illustrated Q-table designs have been explored non-systematically. Hence, here are reported simply the two most interesting agents: *IR-F+ET* and *IR-F+Occ*. Both of them still maintain the levels established for IR. The former has 600, 1000 and 1200 as limit values for IR-F, while the second presents 400, 1000 and 1200. The ET thresholds are 20 and 26 °C, *IR-F+Occ* is designed with an occupied/unoccupied variable instead of external temperature levels. In Table 3.10 are reported improvements in the lighting management over 12% and this reinforce the theory that the IR forecast enhance lighting energy savings.

**Table 3.10** Performances and improvements compared to RBC of reinforcement learning controllers with different states space based on three input variable each.

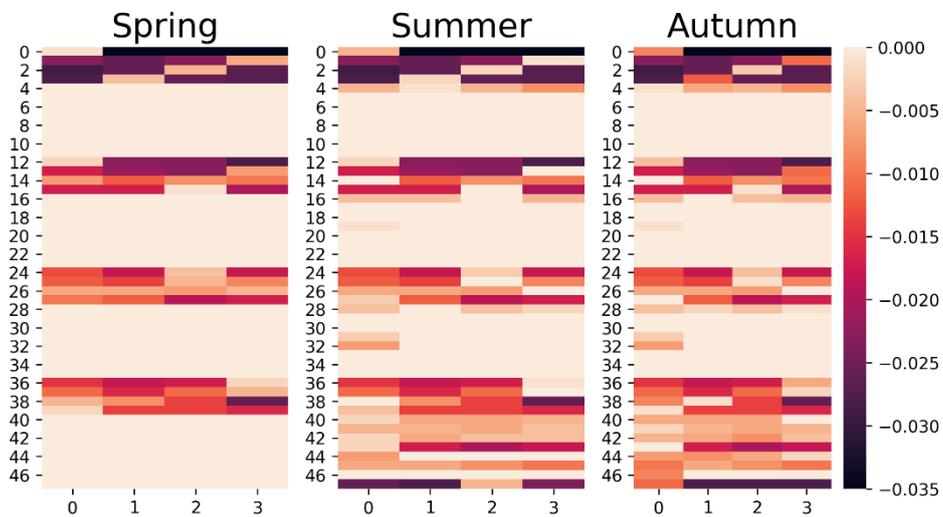
TEST	Measure	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glare [%]
RBC	Absolute	20,55	1,16	7,35	29,06	54,83	26,90
IR- F+ ET	Absolute	21,09	1,22	6,45	28,76	61,14	26,23
	Improvement	-2,63%	-4,98%	12,21%	1,03%	11,51%	2,49%
IR- F+ Occ	Absolute	20,34	1,53	6,45	28,33	53,44	31,04
	Improvement	1,02%	32,25%	12,19%	2,52%	-2,54%	15,38%

*IR-F+ET* make wide use of the darkest state, mostly in the central period of the day, as showed by the relative carpet plot in Figure 3.13. This could be the cause of an improvement in visual comfort compared to reference controller, even though it was not a goal of the RL agent.



**Figure 3.13** Carpet plots of actions selected by reinforcement learning controllers IR-F+ET (on the left) and IR-F+Occ (on the right).

*IR-F+Occ* instead achieves better overall energy performance, thanks to an improvement in heating consumption other than in lighting. Also, its Q-tables illustration (Figure 3.14) for three different simulation steps is reported to make two observations. In order to make this figure more readable, a legend for the state index is provided in Table 3.11.



**Figure 3.14** Q-table of IR-F+Occ heat map.

**Table 3.11** Legend of states space design for Q-table of IR-F+Occ

State	IR min	IR max	ET min	ET max	IR-F min	IR-F max
0	-10	150	-10	20	-0.1	600
1	-10	150	-10	20	600.0	1000
2	-10	150	-10	20	1000.0	1200
3	-10	150	-10	20	1200.0	4000
4	-10	150	20	26	-0.1	600
..	-	-	-	-	-	-
8	-10	150	26	40	-0.1	600
..	-	-	-	-	-	-
12	150	250	-10	20	-0.1	600
..	-	-	-	-	-	-
24	250	300	-10	20	-0.1	600
..	-	-	-	-	-	-
36	300	1000	-10	20	-0.1	600
..	-	-	-	-	-	-
47	300	1000	26	40	1200.0	4000

Firstly, this agent has a broad state space (46 states, resulted from the combination of three different input variable discretization). Secondly, the sporadic or null experience of some states (see the blank regions corresponding, for example, to cold outdoor air temperature occurring with high radiations) could have influenced action selection and prevented a clear policy. This is suggested by the behaviour at the beginning and at the end of days reported in its actions' map depicted above.

### *3.5. Exploration of Q-table design for reward based on UDI only*

In addition to energy savings, this study wants to investigate the adaptability of Reinforcement Learning to deal with conflicting goals, such as solar gains and glare protection in winter or daylight harvesting and cooling loads reduction in summer. UDI is used as synthetic index for visual comfort, with the difficulty that both excessive and deficient indoor illuminance bring a negative reward to the agent. The research for better

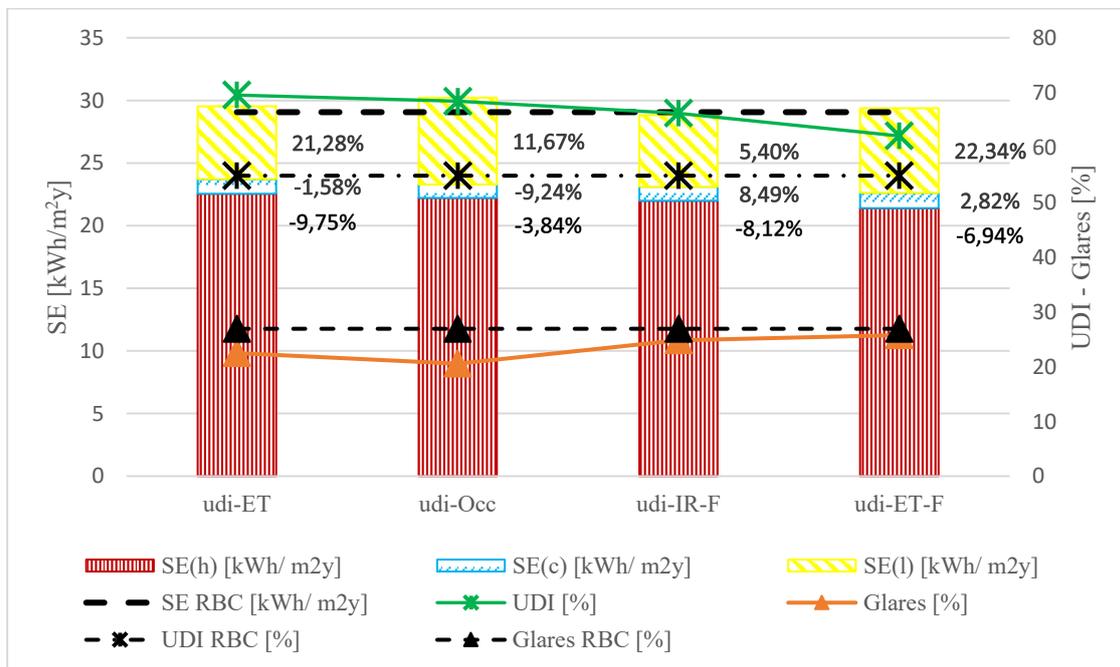
energy performance employed various combinations for each of the different input variable. The same process is not repeated with the energy-comfort reward, but identified best discretization are exploited instead. Five different simulations are carried out with a reward concerning visual comfort only, thus they are called *udi-R*.

$$R_{UDI} = UDI_t \quad (3.1)$$

All the *udi-R* agents' Q-tables include the 4 IR tiers separated by 150, 250 and 300 W/m<sup>2</sup>. Differences are in the secondary input variable. These state designs, together with the associated code, are identified as follow:

- udi-ET: External temperature limits at 20 and 26 °C
- udi-Occ: Presence/absence variable
- udi-IR-F: Incident solar radiation forecast limits at 400, 1000 and 1200 W/m<sup>2</sup>
- udi-ET-F: External temperature forecast limits at 0, 40 and 80 °C

Figure 3.15 shows that all the tests achieved an UDI improvement compared to reference controller and as expected, this increases in natural lighting yields to lighting energy savings, even if not proportional.



**Figure 3.15** Energy performances and improvements by service of reinforcement learning controllers with different secondary input variable and reward function based on UDI only, compared to RBC

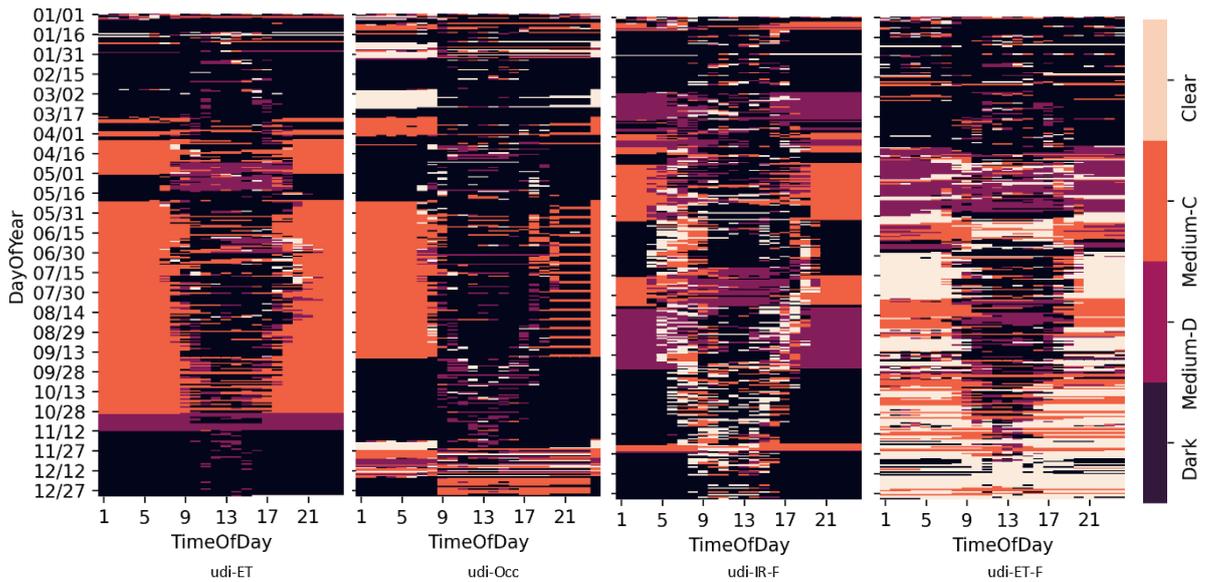
This lack of correlation between UDI and lighting consumption is probably due to the employment of dimmable lights and the computation of UDI indicator, which include the supplementary but not sufficient light class (UDI<sub>s</sub>), as already described in Figure 2.3.

Table 3.12 registers how *udi-ET* and *udi-Occ* almost hits 70% of UDI, that is a value near to *D-0* performance (72,61%) which has the best visual comfort between the *dummy* controllers.

**Table 3.12** Performances of RBC and reinforcement learning controllers with different secondary input variable and reward function based on UDI only.

TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glare [%]
RBC	20,55	1,16	7,35	29,06	54,83	26,90
udi-ET	22,55	1,18	5,79	29,52	69,54	22,47
udi-Occ	22,22	1,06	6,95	30,23	68,40	20,53
udi-IR-F	21,98	1,13	5,71	28,81	66,23	24,78
udi-ET-F	21,41	1,22	6,74	29,37	62,08	25,74

The good performance in visual comfort in fact, as visible from carpet plots (Figure 3.16), is justified by agents extensively application of darkest EC states except for *udi-ET-F*, which has the worst visual performance and also seems to not converge. In general, the two simulations based on forecasts tend to apply more times the clearest state.

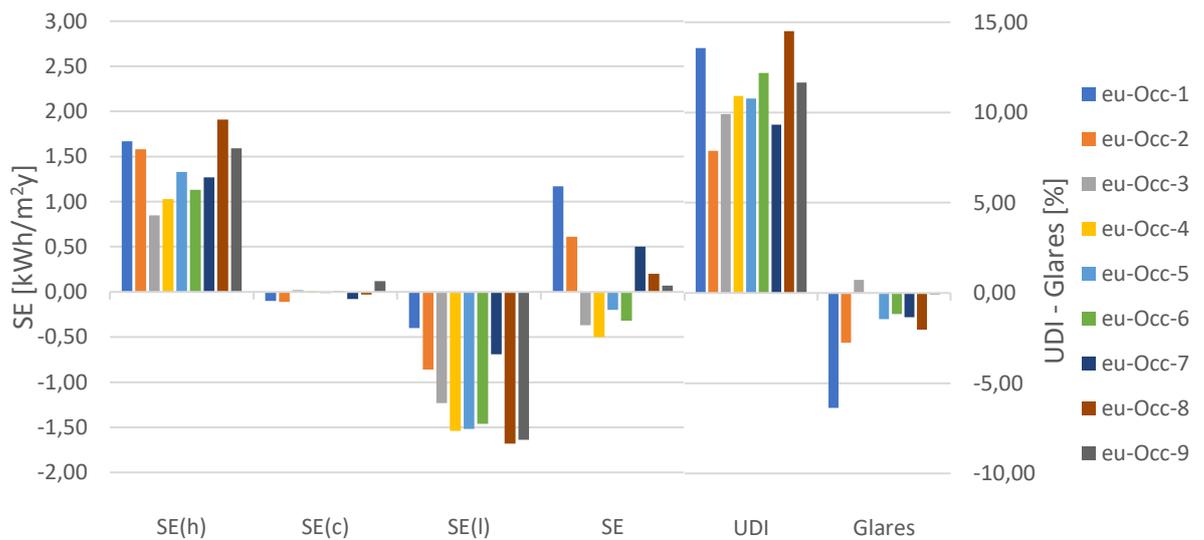


**Figure 3.16** Carpet plots of actions selected through RL for different discount factors.

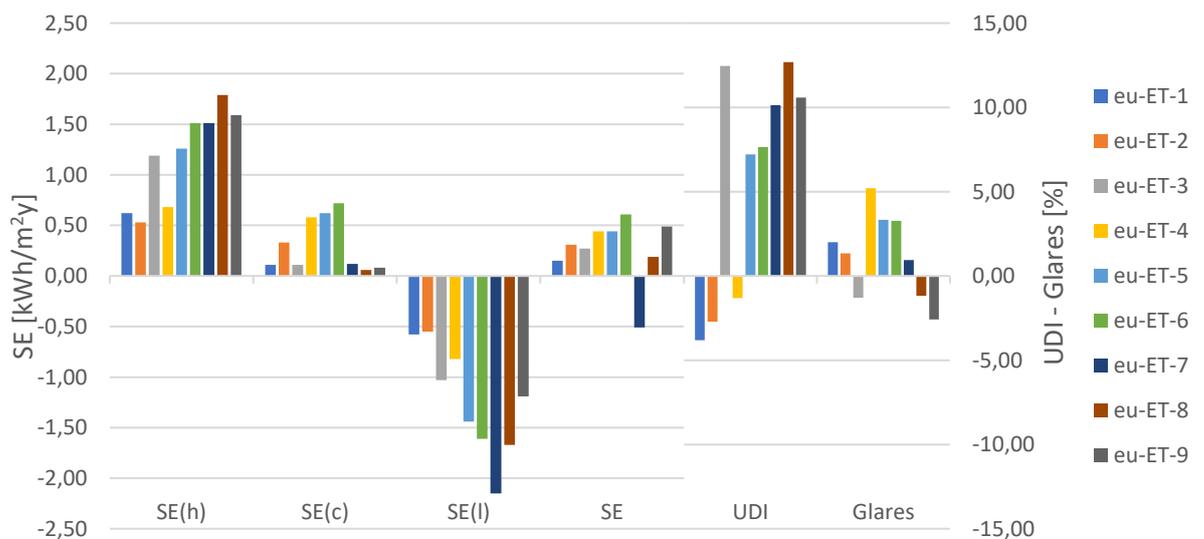
*Glare*s indicator appears in Figure 3.15 to be specular to UDI, indicating that the latter is a good overall parameter to target visual comfort. Between these *udi-R* agents (not having any energy parameter in the reward), only *udi-IR-F* is also able to reduce consumptions more than the present RBC baseline, meanwhile ranking third in visual comfort. It also shows, unlike other tests, to select the more transparent state at the beginning and at the end of the day, suggesting a policy related to daylight intensity over time.

### 3.6. *Exploration of weight for UDI and input variables*

In order to have a two-component reward is suitable to balance their influences. As reported in 2.3.2, this study employs a weight for energy and visual comfort parameters. Since UDI is a parameter linked to occupancy, an initial analysis on the weight to be associated to the UDI component of the reward is carried out by combining IR and occupancy. A second analysis focus on external temperature as additional input variable, because it is the better performing in visual comfort when this is the only component of the reward function. These simulations are coded as follow: having an energy-comfort reward, their code begin with *eu* (“u” is for UDI); the second component indicates the exclusive input variable, that can be *Occ* or *ET*; finally, the design weight is represented, adding for example 5 if  $w_{UDI}$  is set to 0,5. In the graphs reporting the absolute differences between the baseline and the simulations (Figure 3.17 and Figure 3.18), it is possible to see the performance of agents with 8 different weights ( $w_{UDI}$ ), evenly distributed between 0.1 and 0.9, for each of the two states space design.



**Figure 3.17** Energy performances difference between RBC (baseline) and reinforcement learning controllers with different  $w_{UDI}$  and occupancy as additional input variable.

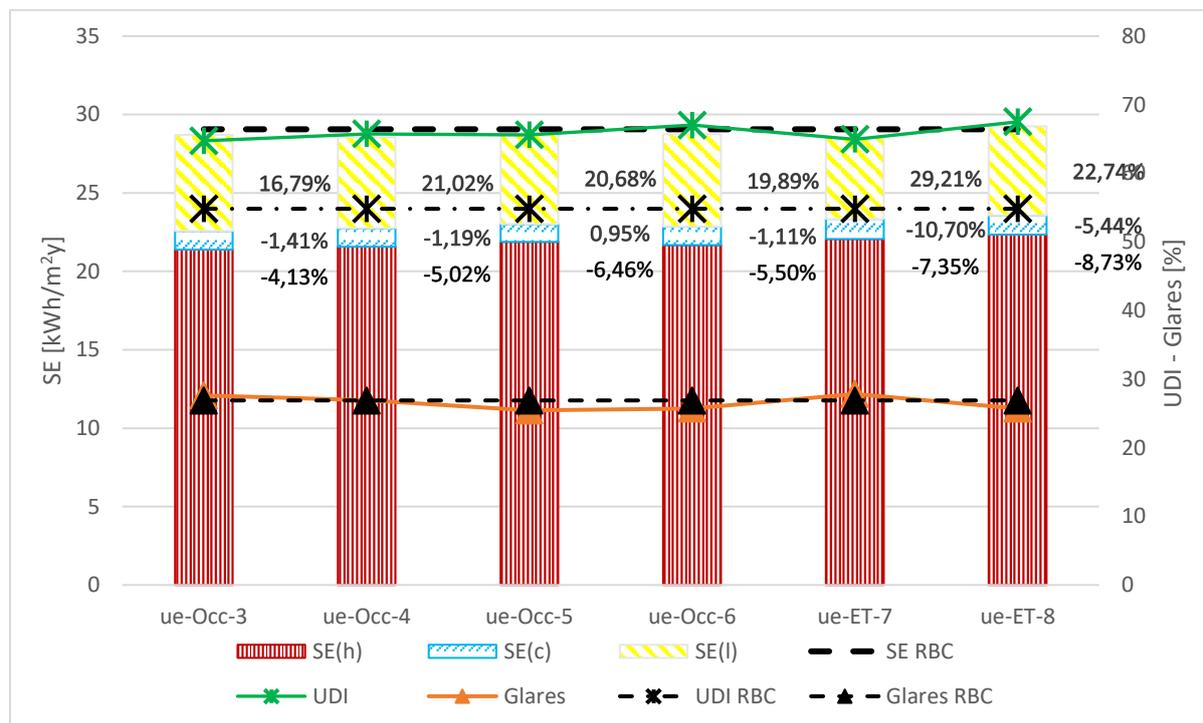


**Figure 3.18** Energy performances difference between RBC (baseline) and reinforcement learning controllers with different  $w_{UDI}$  and external temperature as additional input variable.

Only 3 out of 18 agents do not achieve more useful daylight than RBC, all of which employ ET and a low  $w_{UDI}$ . The real challenge is to perform better at the same time also in energy savings. The pattern that correlate UDI and lighting savings is repeated and heating consumptions are always higher than reference. External temperature appears to be more reliable than occupancy in order to reduce artificial light, but this higher incoming radiation yields poor cooling performance, as expected, while heating consumption does not suggest

the same energy conservation. Effects on glare risk incidence (not measured in the reward function) are diverse, with *euR-Occ* agents showing a more solid reduction behaviour. The better performing RL agents are highlighted for further analysis: these provide more energy savings than RBC except for *eu-ET-8* which, despite its 0,19 kWh/m<sup>2</sup>y SE gap, have an excellent UDI performance (additional 12,7% compared to baseline).

The comparison of selected agents, visible in Figure 3.19, illustrates narrow differences between the six of them and this suggests that achieved performances are towards the optimum for the present study's framework.



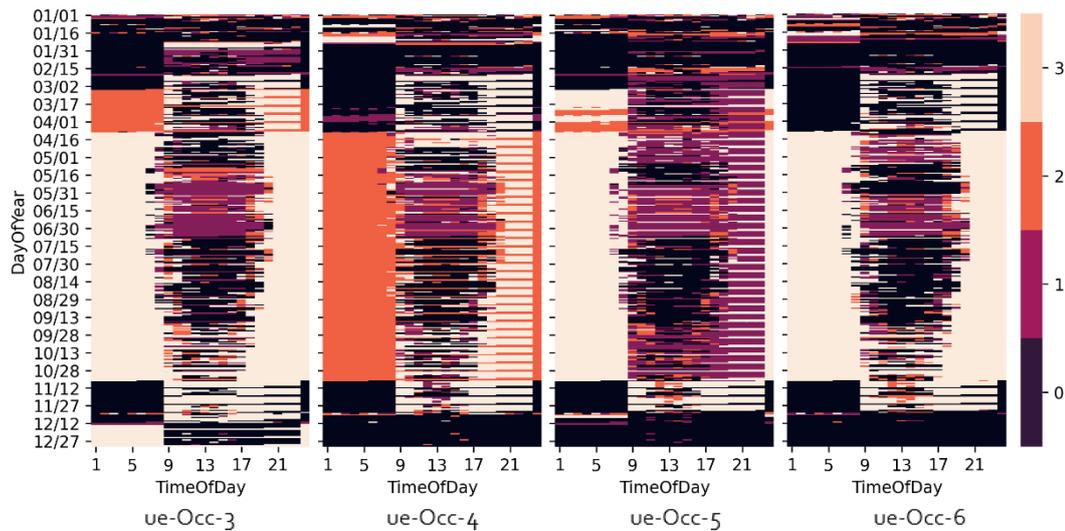
**Figure 3.19** Energy performances and improvements by service of reinforcement learning controllers with different  $w_{UDI}$  and occupancy or external temperature as additional input variable.

Regarding the energy-visual comfort compromise, these tests (selected on a SE savings base) always reach 64% of UDI, providing at least an additional 9,9% compared to RBC performance. However, between those, only *eu-Occ-3*, *eu-Occ-4* and *eu-ET-7* are also able to reduce the Glares indicator. Comparing the results provided in Table 3.13 to baseline, the higher energy improvements are found in *ue-ET-7* (-0,51 kWh/m<sup>2</sup>y) and *ue-Occ-4* (-0,50 kWh/m<sup>2</sup>y). But the latter provides better lighting conditions, with 0,78 % more UDI than *ue-ET-7* and almost the same glare occurrence than RBC instead of a 1% increase found in *ue-ET-7*, that is still comparable to a non-affecting solution.

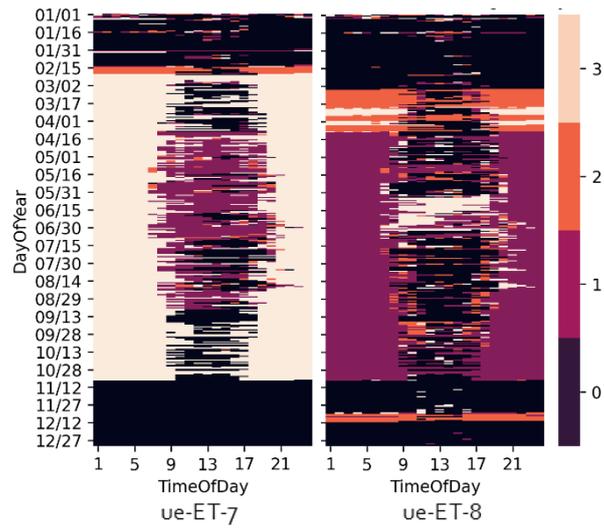
**Table 3.13** Performances of RBC and reinforcement learning controllers with different  $w_{UDI}$  and occupancy or external temperature as additional input variable.

TESTS	SE(h) [kWh/ m <sup>2</sup> y]	SE(c) [kWh/ m <sup>2</sup> y]	SE(l) [kWh/ m <sup>2</sup> y]	SE [kWh/ m <sup>2</sup> y]	UDI [%]	Glare [%]
RBC	20,55	1,16	7,35	29,06	54,83	26,90
ue-Occ-3	21,40	1,18	6,12	28,69	64,74	27,62
ue-Occ-4	21,58	1,17	5,81	28,56	65,74	26,92
ue-Occ-5	21,88	1,15	5,83	28,86	65,61	25,45
ue-Occ-6	21,68	1,17	5,89	28,74	67,02	25,74
ue-ET-7	22,06	1,28	5,20	28,55	64,96	27,84
ue-ET-8	22,34	1,22	5,68	29,25	67,51	25,72

This results can be related to an apparent better balance between the two middle EC window states found in *ue-Occ-4*, compared to a rare selection of the medium-clear state visible in the *ue-ET-7*'s carpet plot (Figure 3.20 and Figure 3.21).



**Figure 3.20** Carpet plots of actions selected by reinforcement learning controllers with different  $w_{UDI}$  and occupancy as additional input variable.



**Figure 3.21** Carpet plots of actions selected by reinforcement learning controllers with different  $w_{UDI}$  and external temperature as additional input variable.

## 4. Conclusion

The present work investigated the adaptability and suitability of Reinforcement Learning to develop an evolving control policy for a smart window in order to address conflicting goals. In particular, an electrochromic window have been chosen as smart glazing technology and 4 states of whom represented the 4 available actions for the controllers. These target practice is part of an adaptive approach for the building elements, which is relevant because of different environmental conditions affecting operative building loads that vary by locations, seasonal weather and future climate evolution, and because of the urgent need for decarbonisation. The large amount of time spent indoor in modern days also requires attention for comfort, affecting people wellbeing.

Simulations run always with the same location, namely London, but the analysed period was not only a season but a whole year. The trained controllers were developed in a simulation environment connecting EnergyPlus to Python. Diverse RL agents have been tested, proposing designs of state space based on different combination of commonly applied measurements found in literature. This state spaces was embedded in an action-state table updating its entries at each time-step, hence framing the control strategy as Tabular Q-learning. The exploration of different Q-table proposals concerned different amount of levels in which input variables were discretized and the variation of the boundaries of these ranges. The action-selection method has been  $\epsilon$ -greedy, with decreasing  $\epsilon$  over time.

This study took into account the energy consumption for key building services (namely, heating, cooling and lighting) and visual comfort as well. Improvements in both fields can have positive economic effects related to, respectively, reduced energy utilization in the operational phase of buildings and enhanced users wellbeing. In order to evaluate the goodness of developed controllers, a Rule-Based Controller has been utilized as baseline. Carpet plots representing the actions selected through the year have been widely employed to evaluate control behaviour and as a convergence indicator in addition.

Main general results are:

- The simple design exploiting incident solar radiation only already showed potential energy savings higher than reference RBC. As concluded by Loonen *et al.* [11], weather forecasts can enhance control strategies. Between the several tests conducted for the present

work, the most energy efficient agents have been *IR-F-8* and *ET-F-1*, which saved 3,35% (the former) and 3,20% (the latter) more Site Energy compared to reference RBC . Both of them were provided with cumulative amount of weather disturbances in the following 4 hours, respectively, incident solar radiation on window and external temperature.

- Energy savings improvements were reached focusing on lighting despite heating service was the most demanding system. In addition, the pattern that correlate UDI and lighting savings is repeated. In these cases, heating consumptions are usually higher than reference. This suggests that electrochromic windows have a much higher potential effect on lighting service than on heating. It is difficult to express the same limit for cooling loads management because the simulations in this study are carried out in a heating dominated climate, labelled “Cfb” according to Köppen-Geiger scheme. However, heating and cooling loads not yield discomfort, because they result only in higher spending to maintain set-points. On the contrary, artificial lighting and glare are conditions which aggravates users’ wellbeing and, for the former, energy demand increase is also true.

- The introduction of a weighted UDI component in the reward function brought to a major application of the clearest EC window state, indicating that this approach allows the controller to understand the positive effect of this policy when employed in low daylight availability hours. The same behaviour was missing in agents having total Site Energy only as feedback from the environment, except for some controllers supplied with “forecast” input variables which exploited clear states few hours at the beginning and end of the day, but not in the night. Agents provided with IR and Occupancy appeared to be more suitable when the reward function concern total energy and UDI, achieving an excellent improvement in the latter parameter and still reducing energy consumption more than a basic control method. Hence, Tabular Q-learning is judged as a good alternative to control strategies that want to address conflicting goals in a dynamic and disturbed building environment.

- The same state space design also yielded excellent visual performance, reaching almost 70% of UDI when this parameter was the only one embedded in the reward function. The same thing is true for RL controller based on IR and external temperature.

- Lastly, curse of dimensionality is a factor to take into account in the development of a Tabular Q-learning algorithm. From the development of this work, this issue appears to

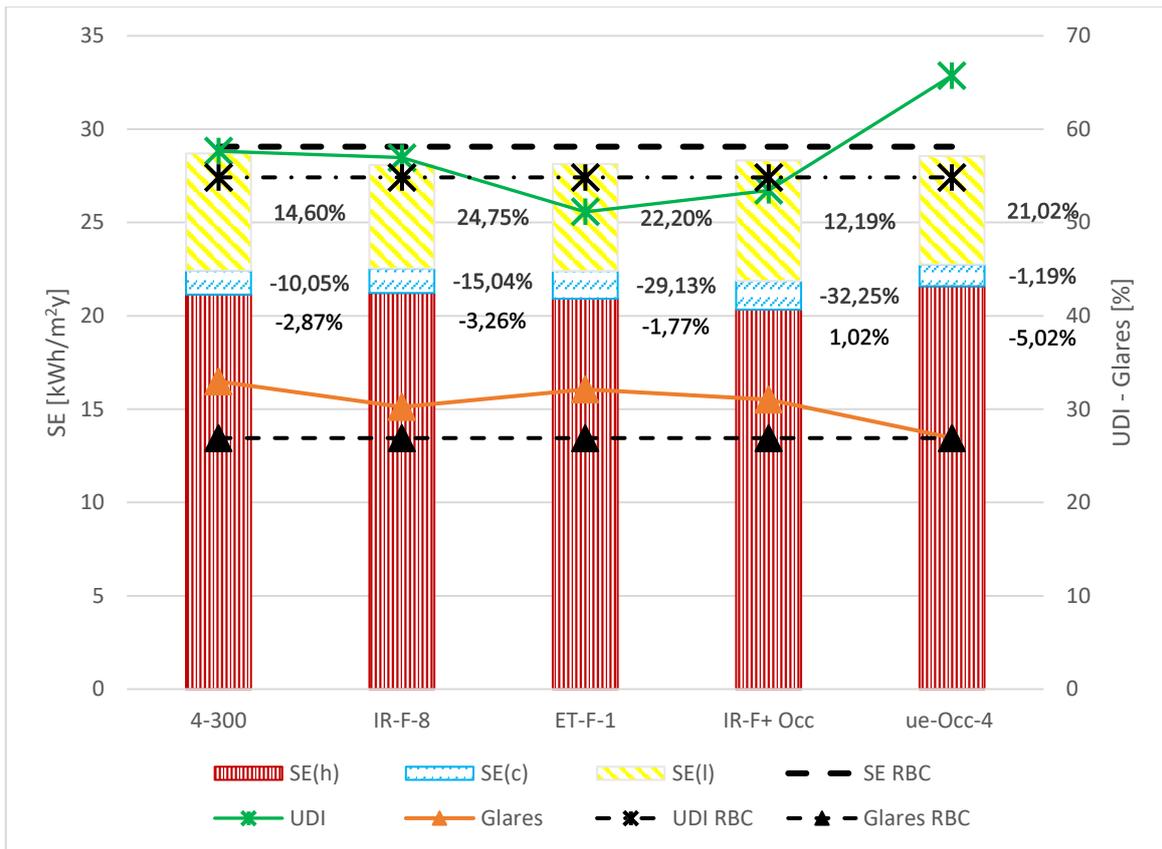
depend on number of levels and input variable at the same time. IR-F-6 is the only 20-states agent tested with IR-F and appeared to be affected by a large state space yielding to weak convergence. On the contrary, IR-F+Occ has a broad state space (46 states, resulted from the combination of three different input variable discretization) resulting in weak convergence at beginning and end of day but however in good energy performance.

Limits and useful future developments are:

- Most of the produced carpet plots of this work show a re-calibration of the control policy when the hot season ends. This risk of losing track of acquired knowledge useful for each season may limit agent's performances. Some reported studies made use of seasonal control strategies. In the tabular reinforcement learning framework, it could be possible to address different periods with different Q-tables. However, the design of interaction between them and effects on convergence speed should be analysed. These sensitive issue depend also by the RL method applied, therefore more complex methods of reinforcement learning could be investigated, from using Soft-Max action selection to Q-values estimation through Neural Network.

- Other environmental conditions such as climates should be cross-evaluated. Furthermore, because of the focus on lighting service developed by the machine learning algorithm in order to reduce energy consumption and dimmable lighting system embedded in the present framework, a controller dealing with binary artificial lights should be investigated to mimic the most common system in the building residential sector.

- A deeper analysis on agents addressing visual comfort and energy savings at the same time could additionally improve RL controller's behaviour in this CABS field. Moreover, it is worth repeating that users' disturbances and preferences, and visual comfort complexity demand for on-field experimentation, even taking into account users interaction with controller. This is suggested because a comparison across best performing agents from different Q-table design (Figure 4.1) shows how *ue-Occ-4* (having the two-components reward function) achieved much higher UDI providing lower energy savings than agents not provided with visual comfort indicator, even if still higher than reference control strategy. For this reason, an evaluation on the necessary compromise between high energy efficient controller and users' productivity correlate to visual comfort is recommended.



**Figure 4.1** Energy performances and improvements by service of best reinforcement learning controllers developed in the present study

## Appendices

Classification of major climatic types according to the modified Köppen-Geiger scheme

1st	2nd	3rd
A (Tropical)	f (Rainforest)	
	m (Monsoon)	
	w (Savanna, Dry winter)	
	s (Savanna, Dry summer)	
B (Arid)	W (Desert)	
	S (Steppe)	
		h (Hot)
		k (Cold)
C (Temperate)	w (Dry winter)	
	f (No dry season)	
	s (Dry summer)	
		a (Hot summer)
		b (Warm summer)
		c (Cold summer)
D (Continental)	w (Dry winter)	
	f (No dry season)	
	s (Dry summer)	
		a (Hot summer)
		b (Warm summer)
E (Polar)	T (Tundra)	
	F (Eternal frost (ice cap))	

## Bibliography

- [1] “Tracking Buildings 2020 – Analysis - IEA.” <https://www.iea.org/reports/tracking-buildings-2020> (accessed Jun. 24, 2021).
- [2] “Warming stripes | Climate Lab Book.” <https://www.climate-lab-book.ac.uk/2018/warming-stripes/> (accessed Jun. 26, 2021).
- [3] “Buildings – Topics - IEA.” <https://www.iea.org/topics/buildings#key-findings> (accessed Jun. 24, 2021).
- [4] L. Pérez-Lombard, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, Jan. 2008, doi: 10.1016/j.enbuild.2007.03.007.
- [5] United Nations Environment Programme, “2020 Global Status Report for Buildings and Construction - Towards a zero-emissions, efficient and resilient buildings and construction sector,” *Global Status Report*, 2020.
- [6] International Energy Agency (IEA), “Technology Roadmap. Energy efficient building envelopes,” *Oecd*, p. 68, 2013, [Online]. Available: <https://www.iea.org/publications/freepublications/publication/TechnologyRoadmapEnergyEfficientBuildingEnvelopes.pdf>
- [7] “EU Buildings Factsheets | Energy.” [https://ec.europa.eu/energy/eu-buildings-factsheets\\_en](https://ec.europa.eu/energy/eu-buildings-factsheets_en) (accessed Jun. 26, 2021).
- [8] M. Economidou, V. Todeschi, P. Bertoldi, D. D’Agostino, P. Zangheri, and L. Castellazzi, “Review of 50 years of EU energy efficiency policies for buildings,” *Energy and Buildings*, vol. 225. Elsevier Ltd, p. 110322, Oct. 15, 2020. doi: 10.1016/j.enbuild.2020.110322.
- [9] M. Perino and V. Serra, “Switching from static to adaptable and dynamic building envelopes: A paradigm shift for the energy efficiency in buildings,” *Journal of Facade Design and Engineering*, vol. 3, no. 2, pp. 143–163, 2015, doi: 10.3233/fde-150039.
- [10] S. Ferguson, K. Lewis, A. Siddiqi, and O. L. de Weck, “Flexible and reconfigurable systems: Nomenclature and review,” *2007 Proceedings of the ASME International*

- Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2007*, vol. 6 PART A, no. May 2014, pp. 249–263, 2008, doi: 10.1115/DETC2007-35745.
- [11] R. C. G. M. Loonen, M. Trčka, D. Cóstola, and J. L. M. Hensen, “Climate adaptive building shells: State-of-the-art and future challenges,” *Renewable and Sustainable Energy Reviews*, vol. 25. Pergamon, pp. 483–493, Sep. 01, 2013. doi: 10.1016/j.rser.2013.04.016.
- [12] W. Chun, S. J. Oh, H. Han, J. T. Kim, and K. Chen, “Overview of several novel thermodiode designs and their application in buildings.” [Online]. Available: <https://academic.oup.com/ijlct/article/3/2/83/693636>
- [13] F. Goia, M. Perino, and V. Serra, “Experimental analysis of the energy performance of a full-scale PCM glazing prototype,” *Solar Energy*, vol. 100, pp. 217–233, Feb. 2014, doi: 10.1016/j.solener.2013.12.002.
- [14] S. P. Corgnati, M. Perino, and V. Serra, “Experimental assessment of the performance of an active transparent façade during actual operating conditions,” *Solar Energy*, vol. 81, no. 8, pp. 993–1013, Aug. 2007, doi: 10.1016/j.solener.2006.12.004.
- [15] P. Greenup, J. M. Bell, and I. Moore, “Importance of interior daylight distribution in buildings on overall energy performance,” *Renewable Energy*, vol. 22, no. 1, pp. 45–52, Jan. 2001, doi: 10.1016/S0960-1481(00)00027-6.
- [16] M. B. C. Aries and G. R. Newsham, “Effect of daylight saving time on lighting energy use: A literature review,” *Energy Policy*, vol. 36, no. 6, pp. 1858–1866, Jun. 2008, doi: 10.1016/j.enpol.2007.05.021.
- [17] P. Boyce, C. Hunter, and O. Howlett, “The benefits of daylight through windows,” no. January 2003, pp. 1–88, 2003.
- [18] E. Saver, “Update Existing Windows to Improve Efficiency”, [Online]. Available: <https://www.energy.gov/energysaver/design/windows-doors-and-skylights/update-or-replace-windows>

- [19] K. M. Al-Obaidi, M. Ismail, and A. M. A. Rahman, "A review of skylight glazing materials in architectural designs for a better indoor environment," *Modern Applied Science*, vol. 8, no. 1, pp. 68–82, 2014, doi: 10.5539/mas.v8n1p68.
- [20] P. Tregenza and J. Mardaljevic, "Daylighting buildings: Standards and the needs of the designer," *Lighting Research and Technology*, vol. 50, no. 1, pp. 63–79, 2018, doi: 10.1177/1477153517740611.
- [21] R. Kosonen and F. Tan, "Assessment of productivity loss in air-conditioned buildings using PMV index," *Energy and Buildings*, vol. 36, no. 10 SPEC. ISS., pp. 987–993, Oct. 2004, doi: 10.1016/j.enbuild.2004.06.021.
- [22] I. Konstantzos, S. A. Sadeghi, M. Kim, J. Xiong, and A. Tzempelikos, "The effect of lighting environment on task performance in buildings – A review," *Energy and Buildings*, vol. 226. Elsevier Ltd, p. 110394, Nov. 01, 2020. doi: 10.1016/j.enbuild.2020.110394.
- [23] L. Hescong, "Daylighting in Schools: An Investigation into the Relationship between Daylighting and Human Performance. Detailed Report," *Hmg-R-9803*, no. October, p. 140, 1999, doi: 10.13140/RG.2.2.31498.31683.
- [24] S. D. Rezaei, S. Shannigrahi, and S. Ramakrishna, "A review of conventional, advanced, and smart glazing technologies and materials for improving indoor environment," *Solar Energy Materials and Solar Cells*, vol. 159, pp. 26–51, Jan. 2017, doi: 10.1016/j.solmat.2016.08.026.
- [25] F. Favoino, F. Fiorito, A. Cannavale, G. Ranzi, and M. Overend, "Optimal control and performance of photovoltachromic switchable glazing for building integration in temperate climates," *Applied Energy*, vol. 178, pp. 943–961, Sep. 2016, doi: 10.1016/j.apenergy.2016.06.107.
- [26] L. Giovannini, F. Favoino, V. Serra, and M. Zinzi, "Thermo-chromic glazing in buildings: A novel methodological framework for a multi-objective performance evaluation," in *Energy Procedia*, Feb. 2019, vol. 158, pp. 4115–4122. doi: 10.1016/j.egypro.2019.01.822.
- [27] S. E. Kalnæs and B. P. Jelle, "Phase change materials and products for building applications: A state-of-the-art review and future research opportunities," *Energy*

- and Buildings*, vol. 94. Elsevier Ltd, pp. 150–176, May 01, 2015. doi: 10.1016/j.enbuild.2015.02.023.
- [28] F. Goia, M. Perino, and V. Serra, “Improving thermal comfort conditions by means of PCM glazing systems,” *Energy and Buildings*, vol. 60, pp. 442–452, May 2013, doi: 10.1016/j.enbuild.2013.01.029.
- [29] C. G. Granqvist, S. Green, G. A. Niklasson, N. R. Mlyuka, S. von Kræmer, and P. Georén, “Advances in chromogenic materials and devices,” *Thin Solid Films*, vol. 518, no. 11, pp. 3046–3053, Mar. 2010, doi: 10.1016/j.tsf.2009.08.058.
- [30] R. D. Clear, V. Inkarojrit, and E. S. Lee, “Subject responses to electrochromic windows,” *Energy and Buildings*, vol. 38, no. 7, pp. 758–779, 2006, doi: 10.1016/j.enbuild.2006.03.011.
- [31] H. Li, S. Yu, and X. Han, “Fabrication of CuO hierarchical flower-like structures with biomimetic superamphiphobic, self-cleaning and corrosion resistance properties,” *Chemical Engineering Journal*, vol. 283, pp. 1443–1454, Jan. 2016, doi: 10.1016/j.cej.2015.08.112.
- [32] W. Feng, L. Zou, G. Gao, G. Wu, J. Shen, and W. Li, “Gasochromic smart window: Optical and thermal properties, energy simulation and feasibility analysis,” *Solar Energy Materials and Solar Cells*, vol. 144, pp. 316–323, Jan. 2016, doi: 10.1016/j.solmat.2015.09.029.
- [33] R. Baetens, B. P. Jelle, and A. Gustavsen, “Properties, requirements and possibilities of smart windows for dynamic daylight and solar energy control in buildings: A state-of-the-art review,” *Solar Energy Materials and Solar Cells*, vol. 94, no. 2. North-Holland, pp. 87–105, Feb. 01, 2010. doi: 10.1016/j.solmat.2009.08.021.
- [34] C. Bechinger, S. Ferrere, A. Zaban, J. Sprague, and B. A. Gregg, “Photoelectrochromic windows and displays,” *Nature*, vol. 383, no. 6601, pp. 608–610, 1996, doi: 10.1038/383608a0.
- [35] P. Domingues, P. Carreira, R. Vieira, and W. Kastner, “Building automation systems: Concepts and technology review,” *Computer Standards and Interfaces*, vol. 45. Elsevier B.V., pp. 1–12, Mar. 01, 2016. doi: 10.1016/j.csi.2015.11.005.

- [36] A. I. Dounis and C. Caraiscos, "Advanced control systems engineering for energy and comfort management in a building environment-A review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6–7. Pergamon, pp. 1246–1261, Aug. 01, 2009. doi: 10.1016/j.rser.2008.09.015.
- [37] L. Yang, Z. Nagy, P. Goffin, and A. Schlueter, "Reinforcement learning for optimal control of low exergy buildings," *Applied Energy*, vol. 156, pp. 577–586, 2015, doi: 10.1016/j.apenergy.2015.07.050.
- [38] A. Afram and F. Janabi-Sharifi, "Theory and applications of HVAC control systems - A review of model predictive control (MPC)," *Building and Environment*, vol. 72, pp. 343–355, 2014, doi: 10.1016/j.buildenv.2013.11.016.
- [39] J. Böke, U. Knaack, and M. Hemmerling, "Automated adaptive façade functions in practice - Case studies on office buildings," *Automation in Construction*, vol. 113, p. 103113, May 2020, doi: 10.1016/j.autcon.2020.103113.
- [40] F. Oldewurtel *et al.*, "Use of model predictive control and weather forecasts for energy efficient building climate control," *Energy and Buildings*, vol. 45, pp. 15–27, Feb. 2012, doi: 10.1016/j.enbuild.2011.09.022.
- [41] G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, and A. Bemporad, "Model Predictive Control (MPC) for enhancing building and HVAC system energy efficiency: Problem formulation, applications and opportunities," *Energies*, vol. 11, no. 3, 2018, doi: 10.3390/en11030631.
- [42] J. Y. Park and Z. Nagy, "Comprehensive analysis of the relationship between thermal comfort and building control research - A data-driven literature review," *Renewable and Sustainable Energy Reviews*, vol. 82. Elsevier Ltd, pp. 2664–2679, Feb. 01, 2018. doi: 10.1016/j.rser.2017.09.102.
- [43] Z. Zhang, A. Chong, Y. Pan, C. Zhang, and K. P. Lam, "Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning," *Energy and Buildings*, vol. 199, pp. 472–490, 2019, doi: 10.1016/j.enbuild.2019.07.029.
- [44] S. Qiu, Z. Li, Z. Li, J. Li, S. Long, and X. Li, "Model-free control method based on reinforcement learning for building cooling water systems: Validation by measured

- data-based simulation,” *Energy and Buildings*, vol. 218, 2020, doi: 10.1016/j.enbuild.2020.110055.
- [45] N. Zhu, K. Shan, S. Wang, and Y. Sun, “An optimal control strategy with enhanced robustness for air-conditioning systems considering model and measurement uncertainties,” *Energy and Buildings*, vol. 67, pp. 540–550, Dec. 2013, doi: 10.1016/j.enbuild.2013.08.050.
- [46] R. C. G. M. Loonen, S. Singaravel, M. Trčka, D. Cóstola, and J. L. M. Hensen, “Simulation-based support for product development of innovative building envelope components,” *Automation in Construction*, vol. 45, pp. 86–95, Sep. 2014, doi: 10.1016/j.autcon.2014.05.008.
- [47] T. A. Nguyen and M. Aiello, “Energy intelligent buildings based on user activity: A survey,” *Energy and Buildings*, vol. 56. Elsevier, pp. 244–257, Jan. 01, 2013. doi: 10.1016/j.enbuild.2012.09.005.
- [48] H. Allcott and S. Mullainathan, “Behavior and energy policy,” *Science*, vol. 327, no. 5970, pp. 1204–1205, 2010, doi: 10.1126/science.1180775.
- [49] E. S. Lee, E. S. Claybaugh, and M. Lafrance, “End user impacts of automated electrochromic windows in a pilot retrofit application,” *Energy and Buildings*, vol. 47, pp. 267–284, Apr. 2012, doi: 10.1016/j.enbuild.2011.12.003.
- [50] M. V. Cano, M. Zamora-Izquierdo, and A. Skarmeta, “User-centric smart buildings for energy sustainable smart cities,” *Transactions on Emerging Telecommunications Technologies*, vol. 25, Jan. 2014, doi: 10.1002/ett.2771.
- [51] Z. Cheng, Q. Zhao, F. Wang, Y. Jiang, L. Xia, and J. Ding, “Satisfaction based Q-learning for integrated lighting and blind control,” *Energy and Buildings*, vol. 127, pp. 43–55, 2016, doi: 10.1016/j.enbuild.2016.05.067.
- [52] J. Y. Park, T. Dougherty, H. Fritz, and Z. Nagy, “LightLearn: An adaptive and occupant centered controller for lighting based on reinforcement learning,” *Building and Environment*, vol. 147, pp. 397–414, Jan. 2019, doi: 10.1016/j.buildenv.2018.10.028.
- [53] M. N. Assimakopoulos, A. Tsangrassoulis, M. Santamouris, and G. Guarracino, “Comparing the energy performance of an electrochromic window under various

- control strategies,” *Building and Environment*, vol. 42, no. 8, pp. 2829–2834, Aug. 2007, doi: 10.1016/j.buildenv.2006.04.004.
- [54] Y. Wu, T. Wang, E. S. Lee, J. H. Kämpf, and J. L. Scartezzini, “Split-pane electrochromic window control based on an embedded photometric device with real-time daylighting computing,” *Building and Environment*, vol. 161, p. 106229, Aug. 2019, doi: 10.1016/j.buildenv.2019.106229.
- [55] J. Hoon Lee, J. Jeong, and Y. Tae Chae, “Optimal control parameter for electrochromic glazing operation in commercial buildings under different climatic conditions,” *Applied Energy*, vol. 260, p. 114338, Feb. 2020, doi: 10.1016/j.apenergy.2019.114338.
- [56] J. Wienold and J. Christoffersen, “Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras,” *Energy and Buildings*, vol. 38, no. 7, pp. 743–757, Jul. 2006, doi: 10.1016/j.enbuild.2006.03.017.
- [57] A. Nabil and J. Mardaljevic, “Useful daylight illuminances: A replacement for daylight factors,” *Energy and Buildings*, vol. 38, no. 7, pp. 905–913, Jul. 2006, doi: 10.1016/j.enbuild.2006.03.013.
- [58] M. N. Assimakopoulos, A. Tsangrassoulis, G. Guarracino, and M. Santamouris, “Integrated energetic approach for a controllable electrochromic device,” in *Energy and Buildings*, May 2004, vol. 36, no. 5, pp. 415–422. doi: 10.1016/j.enbuild.2004.01.040.
- [59] L. L. Fernandes, E. S. Lee, and G. Ward, “Lighting energy savings potential of split-pane electrochromic windows controlled for daylighting with visual comfort,” *Energy and Buildings*, vol. 61, pp. 8–20, Jun. 2013, doi: 10.1016/j.enbuild.2012.10.057.
- [60] Z. Zhang and K. P. Lam, “Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system,” *BuildSys 2018 - Proceedings of the 5th Conference on Systems for Built Environments*, no. November, pp. 148–157, 2018, doi: 10.1145/3276774.3276775.

- [61] C. J. C. H. Watkins and P. Dayan, “Technical Note: Q-Learning,” *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992, doi: 10.1023/A:1022676722315.
- [62] F. B. Mismar, J. Choi, and B. L. Evans, “A Framework for Automated Cellular Network Tuning with Reinforcement Learning,” *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7152–7167, 2019, doi: 10.1109/TCOMM.2019.2926715.
- [63] “ANSI, ASHRAE. 140. Standard method of test for the evaluation of building energy analysis computer programs. Atlanta (GA), USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2011. p. 272.”.
- [64] “Favoino F, Cascone Y, Bianco L, Goia F, Zinzi M, Overend M, et al. Simulating switchable glazing with energyplus: an empirical validation and calibration of a thermotropic glazing model. In: Proceedings of building simulation 2015. Hyderabad, India.”.
- [65] C. Tian, T. Chen, and T. ming Chung, “Experimental and simulating examination of computer tools, Radlink and DOE2, for daylighting and energy simulation with venetian blinds,” *Applied Energy*, vol. 124, pp. 130–139, Jul. 2014, doi: 10.1016/j.apenergy.2014.03.002.
- [66] C. F. Reinhart and O. Walkenhorst, “Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds,” *Energy and Buildings*, vol. 33, no. 7, pp. 683–697, Sep. 2001, doi: 10.1016/S0378-7788(01)00058-5.
- [67] UNI/TS 11300-1, “Prestazioni energetiche degli edifici – Parte 1: Determinazione del fabbisogno di energia termica dell’edificio per la climatizzazione estiva ed invernale.” 2014.
- [68] “CIBSE Approved Document L2A: conservation of fuel and power in new buildings other than dwellings, 2013 edition with 2016 amendments, ISBN 978 1 85946 512 7”.
- [69] “The Australian Building Codes Board, NCC 2016. Volume one. Building Code of Australia. Class 2 to class 9 buildings, Section J – Energy Efficiency; 2016.”.
- [70] “Mardaljevic J, Andersen M, Roy N, Christoffersen J. Day lighting metrics: is there a relation between useful daylight illuminance and daylight glare probability? In:

Proceedings of the building simulation and optimization conference (BSO12), Loughborough.”.

- [71] “ANSI/ASHRAE/IES Standard 90.1. Energy standard for buildings except low-rise residential buildings; 2013.”.
- [72] J. Mardaljevic *et al.*, “DAYLIGHTING METRICS : IS THERE A RELATION BETWEEN USEFUL DAYLIGHT ILLUMINANCE AND DAYLIGHT GLARE PROBABILITY? VELUX A / S , Adalsvej,” *Proceedings of the Building Simulation and Optimization Conference BSO12*, 2012, [Online]. Available: <https://infoscience.epfl.ch/record/179939?ln=en>
- [73] M. S. Piscitelli, S. Brandi, G. Gennaro, A. Capozzoli, F. Favoino, and V. Serra, “Advanced Control Strategies For The Modulation of Solar Radiation In Buildings : MPC- enhanced Rule-based Control Technology Energy Building Environment research group , Department of Energy , Politecnico di Torino , Italy,” pp. 869–876, 2019.

## Acknowledgements

My complete gratitude to all scientists who, since the 60s, worked on researches about global warming, even though for decades the chances of communicating and being financed were against them. Their efforts made possible for me to be aware and to find a goal.

An affectionate thanks to the centenarian scout movement, from which I learnt so much that I'm not sure I would be here, where I'm glad to be, if I haven't met them. I'll be still in dept for some years at least.

A sincere thanks to Silvio Brandi who gave me the needed attention and support to carry on this challenge, and to Professor Alfonso Capozzoli for being both frank and understanding. A special thanks to Fabio Favoino for its extensive research work in the field of this document and for the fundamental support he gave me. Discussions with you, your points of view and knowledge have been essential and greatly appreciated.

Last but not least, my parents Anna and Franco had a role so important that I cannot completely understand it. I have been lucky for all the opportunities they managed to give me. I thank also my sister Serena for the positive discussions she's capable of and my friend Gianluca to be source of curiosity and sympathy.