

Article

The Potential of Control Models Based on Reinforcement Learning in the Operating of Solar Thermal Cooling Systems

Juan J. Diaz *  and José A. Fernández 

Departamento de Ingeniería Energética, Escuela Técnica Superior de Ingenieros Industriales,
Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain

* Correspondence: juanjose.diaz.carrillo@alumnos.upm.es; Tel.: +34-637320673

Abstract: The objective of this research work was to investigate the potential of control models based on reinforcement learning in the optimization of solar thermal cooling systems (STCS) operation through a case study. In this, the performance of the installation working with a traditional predictive control approach and with a reinforcement learning (RL)-based control approach was analyzed and compared using a specific realistic simulation tool. In order to achieve the proposed objective, a control system module based on the reinforcement learning approach with the capacity for interacting with the aforementioned realistic simulation tool was developed in Python. For the studied period and the STCS operating with a control system based on RL, the following was observed: a 35% reduction in consumption of auxiliary energy, a 17% reduction in the electrical consumption of the pump that feeds the absorption machine and more precise control in the generation of cooling energy regarding the installation working under a predictive control approach. Through the obtained results, the advantages and potential of control models based on RL for the controlling and regulation of solar thermal cooling systems were verified.

Keywords: solar energy; linear Fresnel collector; absorption cooling; hourly and parametric simulation; simulation tool; Q-learning; reinforcement learning; Python; EES



Citation: Diaz, J.J.; Fernández, J.A.

The Potential of Control Models
Based on Reinforcement Learning in
the Operating of Solar Thermal
Cooling Systems. *Processes* **2022**, *10*,
1649. [https://doi.org/10.3390/
pr10081649](https://doi.org/10.3390/pr10081649)

Academic Editor: Timothy Langrish

Received: 3 July 2022

Accepted: 16 August 2022

Published: 19 August 2022

Publisher's Note: MDPI stays neutral
with regard to jurisdictional claims in
published maps and institutional affili-
ations.



Copyright: © 2022 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the Creative Commons
Attribution (CC BY) license ([https://
creativecommons.org/licenses/by/
4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

It is well known that, at the present time, the thermal requirements of the diverse variety of buildings (residential, commercial and industrial) have been increasing with a growing trend and that the highest percentage of the electrical consumption of these buildings is precisely due to the systems installed to meet these demands.

For years, work has been carried out on the development and implementation of alternative technologies that allow the satisfaction of the increasing, growing thermal demand in the various sectors of this discipline such as air conditioning and commercial and industrial refrigeration. Alternative technologies that permit to discharge, mitigate and even minimize the growth of the saturated current electrical system, resulting in a decrease in the use of fossil fuels and its corresponding consequences such as greenhouse gases and global warming.

In order to respond to the aforementioned need, in recent years, research work has focused on the development of different types of equipment, systems and solutions, such as photovoltaic panels, flat and concentrating solar collectors, absorption systems and adsorption systems, among many others.

However, the implementation of these systems and solutions is no longer enough; now, control and regulation systems are becoming essential for designing and establishing strategies that optimize all those factors that intervene in the functioning and operation of new proposed systems and solutions, as well as in the environments that require these treatments. These control and regulation systems need to be oriented toward the optimization of the obtained green energy resources, reduction of the electrical consumption and the auxiliary energy requirements.

Traditionally, numerous and very different types of control methods and models have been used for the regulation of HVAC and refrigeration systems and equipment [1–4]. These control models can be classified into three large groups: traditional, rule-based models, model-based approaches and data-driven demand response-based control models.

Traditional, rule-based models have been extensively implemented in various types of mechanical and electrical installations [5–13]. The main advantages of this type of control model are found in its plain and intuitive structure, easy implementation, low initial cost, quick response and feedback controller [4,14,15]. However, the limitation in modulation capability, the small scalability, the constant need for operator manipulation, the lack of learning capability, the requirement of more complex rules with higher maintenance/update costs for the complex system and the non-optimal performance constitute its main disadvantages [4,14,15].

The model-based approach has been widely investigated in recent years [16–23]. This control approach, in accordance with [4,24], is considered so valuable for controlling real systems due to its high accuracy, anticipating behavior, capability to consider hard constraints, disturbance robustness, adaptation to shifting in performing conditions and flexibility related to using available explicit models. On the contrary, the high initial and installation costs, as well as the complexity in identifying the appropriate system model, constitute its most unfavorable aspects.

Despite the important advantages indicated above, the commercial and industrial application of the model-based approach has been limited, since its correct, stable and accurate performance depends largely on the use of very precise assumptions and customized models with an elevated grade of detail [25]. These conditions are impractical and very difficult to achieve given the complexity of the dynamics of thermal systems and of those factors involved in their operation, so it is common to observe this type of control system working with low convergence and unstable and deficient performance.

The overcoming of the drawbacks mentioned above and the generation of optimal control policies can be achieved through the implementation of the data-driven demand response-based control model due to its multiple favorable characteristics. Some of these favorable characteristics are the capability to find an optimal action policy without requiring a model of the system, the flexibility of the algorithm to learn from the interaction with the environment, lower computational costs, high accuracy and, above all, the ability to adapt to real-time measurements and its correct performance in dynamic and uncertain environments.

A data-driven demand response-based control model has its own limitations, such as the requirement for a large amount of data, which makes it difficult to train the model directly from the real system or environment. In these cases, a model of the system and a large number of simulations are required to achieve agent training.

Nowadays, the implementation of the data-driven demand response-based control model is favored due to:

- The availability and accessibility of the building control and automation data provided by sensors and control, measurement and automation devices;
- The simplicity of the processes of collection, analysis and management of the building control and automation data derived from the virtues of big data and the powerful computing equipment available in the current market.

The aforementioned virtues of the data-driven demand response-based control model and the favorable conditions for its utilization have recently aroused deep interest in the investigation and implementation of this control approach through the use of the reinforcement learning technique in various types of application.

The main fields in which this type of control approach has been investigated and implemented are those subdisciplines involved in the treatment and operation of buildings and facilities of all kinds, such as heating, ventilation and air conditioning, refrigeration, electricity, lighting, energy storage, batteries, domestic hot water, communications, sustainability and building energy management.

Building energy management is one of the areas in which the implementation of reinforcement learning has been most studied to date [26–31]. The optimization of the performance and operation of solar domestic hot water and heating systems using control models based on reinforcement learning has also been studied [32–36].

Likewise, various reinforcement learning techniques and algorithms have been implemented in HVAC systems or solutions in various types of building, all with the aim of finding the optimal control policy that allows the reduction of energy consumption and/or improves the comfort of the treated space [25,37–43].

Some other examples from the many disciplines and fields related to energy in which the potential of reinforcement learning in process optimization has been studied are: power and energy systems [44–49], urban energy management [50], storage energy [51] and electric distribution systems [14].

Despite the recently conducted studies focused on the use of control models based on machine learning, there is still much to explore, especially in terms of HVAC systems and solutions. At the present time, even with the well-known virtues of machine learning for controlling thermal systems, no detailed study has been conducted regarding the implementation of reinforcement learning to optimize the design and operation of solar thermal cooling systems.

Considering the aforementioned evidence, the present research work was performed with the objective of analyzing and studying the potential of control systems based on reinforcement learning in the optimization of the operation and functioning of solar thermal cooling systems driven by linear Fresnel collectors. The main contributions and innovative aspects of this study are the following:

- Implementation of a control system based on reinforcement learning in the regulation of a solar thermal cooling system;
- The modular integration of a simulation tool developed in EES with a reinforcement learning algorithm module developed in Python, simplifying and facilitating the agent training;
- Proposal of a conceptual scheme for the real implementation of the control of solar thermal cooling systems using reinforcement learning;
- Verification of the potential of using a control model based on reinforcement learning in the optimization of the operation of an absorption solar cooling system.

Through the performed simulations, it was possible to verify the advantages and potential of control models based on RL for controlling and the regulation of solar thermal cooling systems. It was observed that, for the studied period and with the solar thermal cooling system operating with a control system based on RL, there was a 35% reduction in consumption of auxiliary energy, a 17% reduction in electrical consumption by the pump that feeds the absorption machine and more precise control in the generation of cooling energy regarding the installation working under a predictive control approach.

2. Materials and Methods

2.1. General Description of the Study

The study consisted of evaluating and verifying the potential of a control model based on reinforcement learning in optimizing the operation of solar thermal cooling systems driven by linear Fresnel collectors (STCS_LFC). This was conducted through a practical case study in which the performance of a solar cooling installation operated by a control system based on RL and by a predictive control system was simulated and compared.

The objective of the study was limited to verifying the effect of reinforcement learning on improving system performance and not the maximum optimization that could be achieved with this technique. Therefore, only a few parameters of the many involved in its operation and on which its optimization depends were considered.

The optimization of the operation of the proposed STCS_LFC was evaluated by considering the auxiliary energy requirements, the electrical consumption of sub-circuit C and the satisfaction of the refrigeration demand as dependent variables and the conditions

of the storage tank leaving flow (mass/volumetric flow and temperature) as independent variables. That is, with the Q-learning control approach, the aim was to find the control policy that maximizes the use of solar energy in satisfying the cooling demand and, thereby, reduces the consumption of auxiliary energy and the electrical consumption of the pump of the sub-circuit that feeds the absorption machine.

The simulations of the STCS_LFC operating with a predictive control approach were carried out through the simulation tool [52]. To perform the simulations of the STCS_LFC controlled through the reinforcement learning approach, it was necessary to implement a programming module based on the Q-learning algorithm in operation mode, the agent of which had been previously trained.

2.2. Description of the Case Study

In the case study, the hourly simulation of a solar cooling installation with the configuration shown in Figure 1 was analyzed for a typical summer day (on 1 July) in the city of Riyadh. This solar cooling installation was dimensioned to satisfy the cooling demand of a typical set of electro-mechanical buildings for the maintenance of the medium-capacity railway installations located in Riyadh.

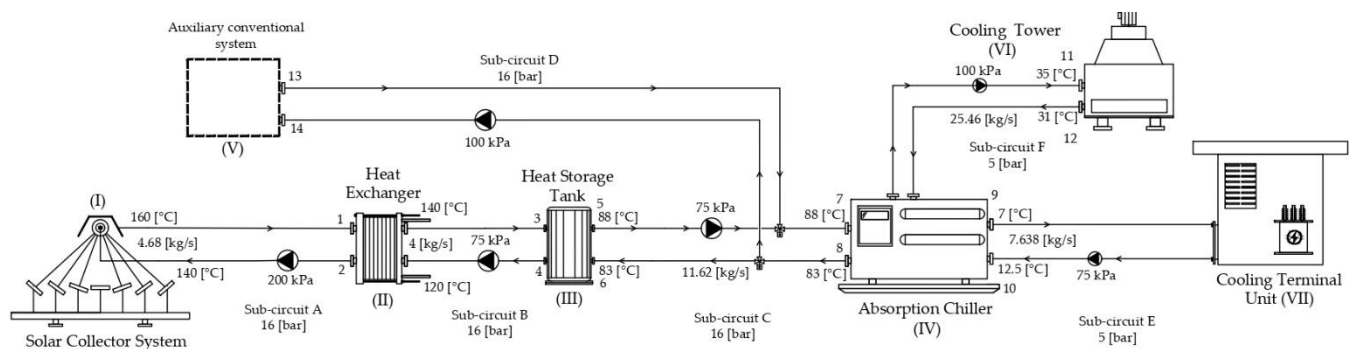


Figure 1. Schematic diagram of the proposed solar thermal cooling system.

The buildings are treated through an air-handling unit (AHU) fed with chilled water generated by the absorption machine. The AHU is arranged with a recirculation module and supplies the minimum requirements of outdoor air.

2.2.1. Building Characteristics

The properties of the buildings' construction materials (enclosures and glazing), as well as the design and use characteristics of the air-conditioning and lighting systems, were considered in compliance with the indications of [53] and in accordance with the climatic zone of the place where the study was carried out.

The minimum outdoor air requirements were based on [54], according to the type of space or building to be treated. The internal thermal loads were calculated by applying a ratio of 230 W/m² derived from previously estimated mechanical and electrical losses considered for this type of buildings. The detailed analysis performed to obtain the ratio for the internal thermal loads is outside the scope of this work.

2.2.2. Refrigeration System

The flow diagram of the proposed solar cooling thermal system in this study is shown in Figure 1 [52]. This configuration is characterized by operating under an indirect coupling control mode [55] similar to that of commissioned installations, such as those in [56–58] among others.

The main elements of this system are: a field of concentrating linear Fresnel collectors (I), a stratified hot water storage tank (III), a single-effect absorption chiller using the pair LiBr–H₂O (IV) and a cooling tower (VI). The complementary elements of the proposed system are: a heat exchanger (II), a conventional auxiliary thermal energy supply subsystem

(V) and various water distribution elements consisting of pipes, pumping groups and control valves that interconnect the aforementioned subsystems.

The main considered aspects for the operation of the proposed solar cooling system in this study are mentioned below:

- The linear Fresnel collectors are arranged horizontally and with a north–south orientation;
- The heat transfer fluid in each of the sub-circuits is pressurized water;
- The storage tank has stratification and is arranged in a vertical position.

In the schematic diagram in Figure 1, the design conditions of the proposed solar thermal cooling system for the case study are presented, and Tables 1–4 show the technical and sizing characteristics of its components.

Table 1. Technical information of the linear Fresnel collector from manufacturer datasheet.

Parameter	Unit	Value
Manufacturer	-	Industrial Solar
Model	-	LF-11
Module width	m	7.5
Module length	m	4.06
Aperture area per module	m ²	22
Receiver height	m	4
Number of modules	ut	32

Table 2. Design characteristics of the cooling tower.

KaV/L	L/G Ratio	Approach	Range
0.2936	1.2	12	4

Table 3. Technical information of the absorption chiller from manufacturer datasheet.

Parameter ¹	Unit	Value
Manufacturer	-	Yazaki
Model	-	WFC-1C100
UA _a	kW/K	100.5
UA _c	kW/K	49.81
UA _g	kW/K	157.9
UA _e	kW/K	44.1
η _{Hx}	%	65
$\dot{m}_{solution_a}$	Kg/s	2.217

¹ The values of the parameters: product of the overall heat transfer coefficient and the heat exchanger area at the absorber, generator, condenser and evaporator, respectively (UA_a, UA_g, UA_c, UA_e); the efficiency of the heat exchanger of the absorption machine (η_{Hx}) and the mass flow rate of the solution at the outlet of the absorber in the refrigeration cycle ($\dot{m}_{solution_a}$) were estimated through an iterative process using the mathematical model of the absorption machine in conjunction with the technical information provided by the manufacturer in the datasheet.

Table 4. Design characteristics of the storage tank.

Parameter	Unit	Value
Volume	L	15,000
Diameter	m	2.285
Height	m	3.657
Thickness tank	mm	10
Thickness insulation	mm	50
Thermal conductivity of the tank	W/mK	16.3
Thermal conductivity of the insulation	W/mK	0.05

The pressure losses of the pumps were previously estimated through a general analysis of each of the sub-circuits, considering the pipes, hydraulic components and accessories present, as shown in Figure 1.

The operating parameters (range, the ratio between the water mass flow rate (L) and the air mass flow rate (G) and the approach) of the cooling tower were defined considering the recommendations of [59] for air-conditioning applications, and they are shown in Table 2. The value of the tower characteristic (KaV/L) was calculated using the mathematical model of the cooling tower. For this calculation, the values of the operating parameters and the ambient design conditions for cooling were considered for the hottest month of the year.

2.2.3. Climatic Conditions

The studied building was considered to be located in Riyadh, Saudi Arabia. The geographic information and the annual cooling design conditions were defined in accordance with the information provided by the ASHRAE climatic conditions database [60] corresponding to the year 2017. In addition, the climatic conditions regarding the annual hourly profile of the direct normal irradiation, dry bulb temperature, relative humidity, atmospheric pressure and wind speed were obtained from the database of the software Design Builder [61] and from the software System Advisor Model (SAM) [62].

2.2.4. Study Period

The study was conducted for a typical summer day in the city of Riyadh, Saudi Arabia. The selected day was 1 July. The climatic conditions for the studied day were defined according to Section 2.2.3.

2.2.5. Software

The software Design Builder [61] was used for the modeling of the studied buildings and for the estimation of the hourly cooling demand.

The simulation and analysis of the solar cooling systems driven by linear Fresnel collectors were accomplished using the realistic simulation tool developed in the research work [52]. Detailed information about this simulation tool is provided in Section 2.4.

2.3. Work Sequence of the Proposed Study

The work sequence followed to conduct the study on the potential of reinforcement learning in optimizing the operation of solar thermal cooling systems driven by linear Fresnel collectors is presented in Figure 2, where: DNI: direct normal irradiance (W/m^2); and STCS_LFC-QL: the simulation module developed in EES that integrates the STCS_LFC tool [52] and the Q-learning module developed in Python software.

2.4. Simulation Tool for Analysis of Solar Thermal Cooling Systems Driven by Linear Fresnel Collectors

The simulation tool [52] was used to perform the simulations of the solar thermal cooling system driven by linear Fresnel collectors studied in the practical case of this research work. This simulation tool was used, on the one hand, individually with a predictive control approach and, on the other hand, in conjunction with the Q-learning module described in Section 2.5 in agent training mode and later in operation mode.

The STCS_LFC simulation tool [52] was developed using the EES software [63]. It consists of a main program that integrates and interconnects: (1) subroutines containing the governing equations for the system components of a solar cooling thermal installation (submodules), (2) input data and (3) a set of control statements. The resolution of the mathematical model is carried out through a parametric table that interacts with the main program window by means of specific programming statements, as well as with variables that serve as a bridge among them. These bridge variables allow the exchange

of information among the parametric table, the control procedure and the base/main mathematical model in the main window.

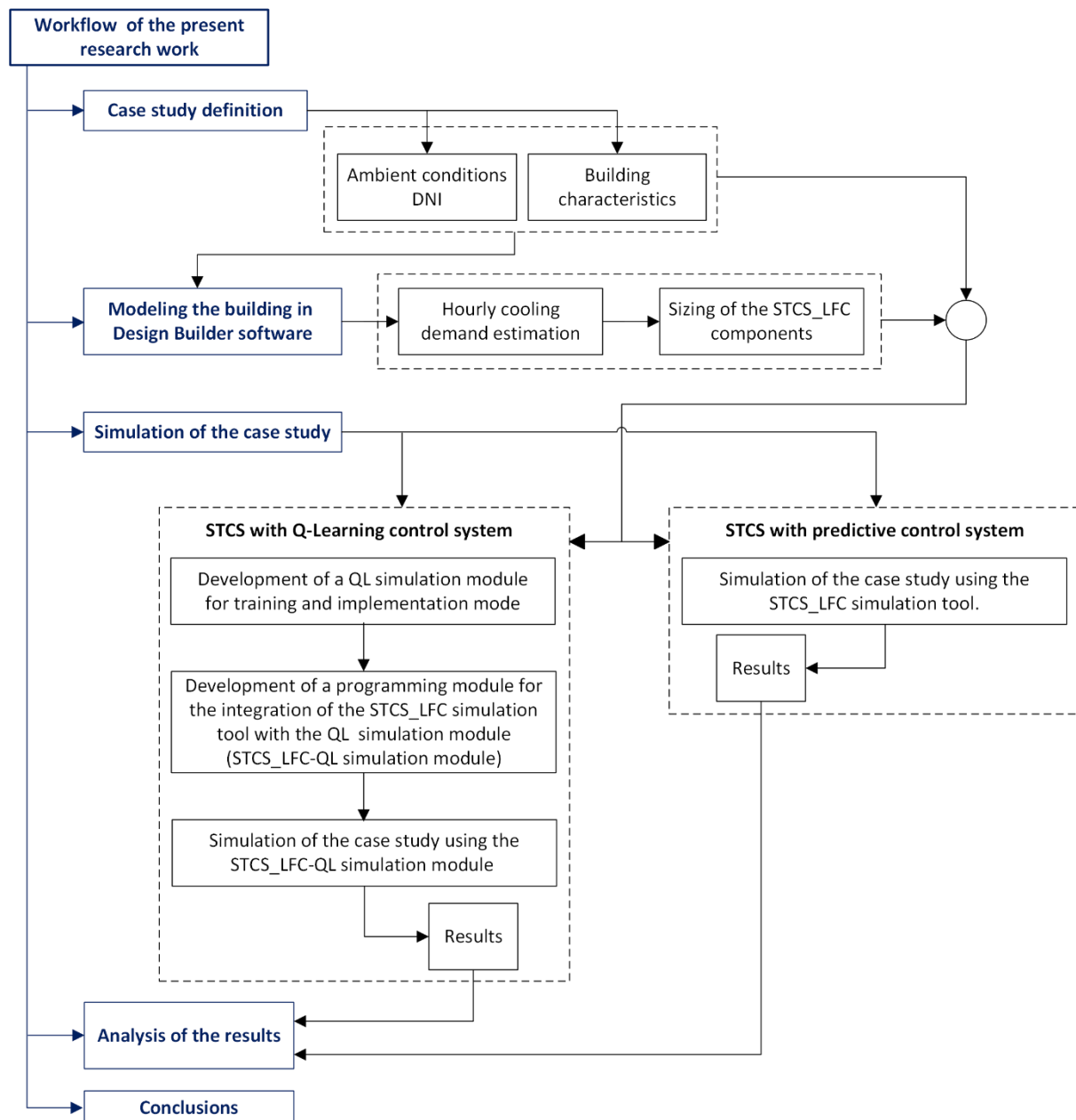


Figure 2. Workflow of the proposed study.

The integrated mathematical model of the simulation tool [52] considers the ambient conditions, the thermal loads of the building, the dimensioning data of each of the components of the system and the simultaneous interaction among them to conduct a realistic, simple and precise analysis. The structure diagram of the developed mathematical model is presented in Figure 3.

In accordance with [52], the correct implementation and accuracy of the mathematical model used for the development of the simulation tool were verified through a code and calculation verification. This verification was conducted through an evaluation by comparison among the numerical solutions obtained from the mathematical model in the assessment related to the known correct answers of a proposed case study or experimental results.

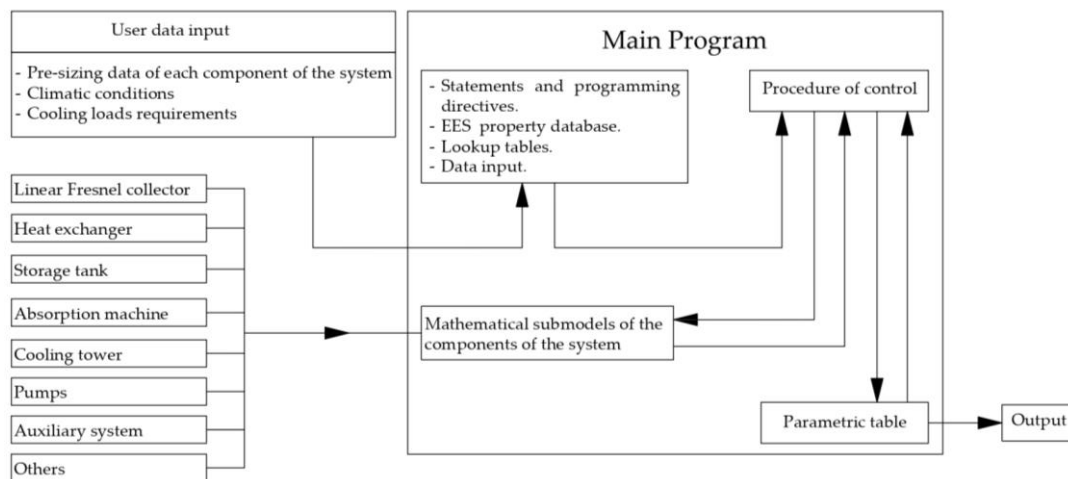


Figure 3. Structure diagram of the developed simulation tool in EES.

The mean deviation between the experimental data and the calculated numerical solution from the model for the main components of the solar cooling installation are indicated below:

- Mean convergence error calculated for the linear Fresnel collector, <3.87%;
- Mean convergence error calculated for the storage tank, <1.10%;
- Mean convergence error calculated for the absorption chiller, <0.48%;
- Mean convergence error calculated for the cooling tower model, <1%.

With the results indicated above, the accuracy of the simulation tool is demonstrated, and its use in this research work is justified.

2.5. Interaction between the Solar Thermal Cooling System Simulation Tool and the Q-Learning Module

Figures 4 and 5 show the flowcharts that illustrate the interaction between the Q-learning module developed in Python and the solar thermal cooling system simulation module developed in EES for the training and implementation phases, respectively.

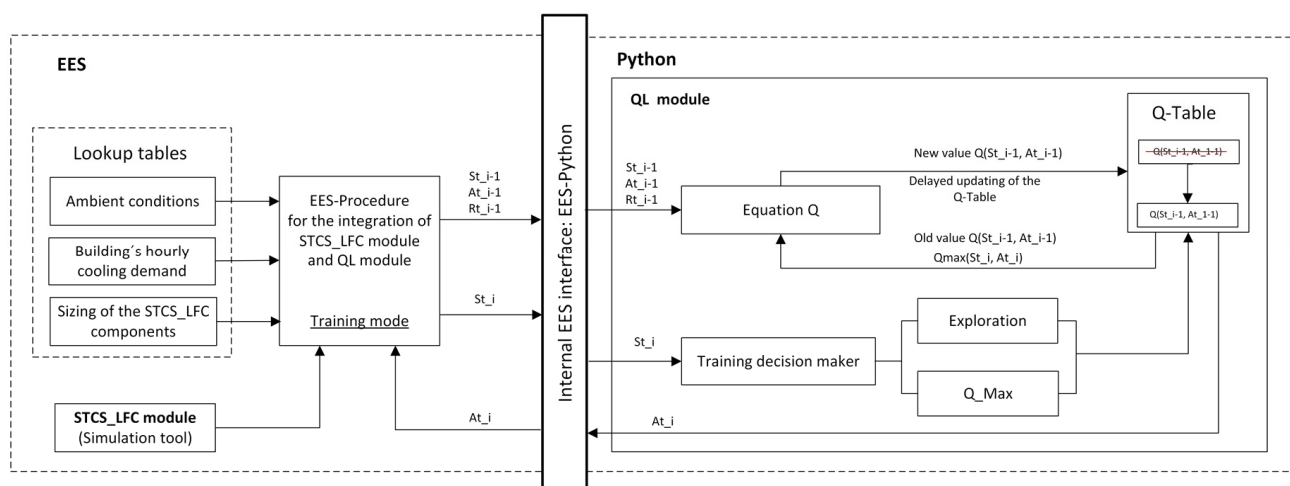


Figure 4. Flowchart of the interaction of the developed modules in EES and Python for the case study of the present research work in training mode.

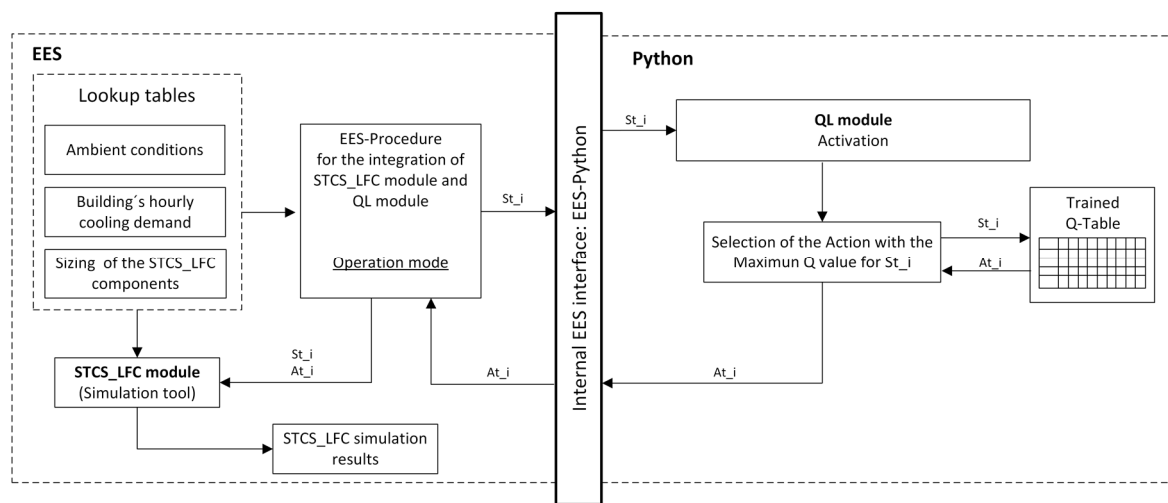


Figure 5. Flowchart of the interaction of the developed modules in EES and Python for the case study of the present research work in operation mode.

The combination of these two pieces of software aimed to facilitate and optimize the implementation of the Q-learning technique, since each of the involved modules was developed in the most appropriate programming languages for its resolution. In addition, in this way, each of them was independently assigned a specific and complementary activity within the installation control process, thus avoiding programming problems when trying to use a programming language to resolve issues for which they are not fully prepared or when its operation brings more problems than benefits.

For example, with the EES software it is not possible to program a sequence that allows an array to be updated in a simple and optimal way; however, in Python, this is something that is very simple, practical and safe to do. The same thing happens if an attempt is made to program a complex thermodynamic system in Python, which does not have the necessary tools and functions to allow for this to be carried out correctly, simply and, in general, without convergence problems, but is possible with EES. The interaction between EES and Python in the implementation of Q-learning constitutes one of the most innovative aspects of this research work.

On the other hand, it is important to mention that, since there were no real measurements of the performance of the proposed solar cooling thermal system to be studied, the idea was to use the tool [52] to realistically simulate the thermal behavior of the installation and, subsequently, use the obtained results for the learning of the agent during the training phase. As shown in Figure 4, during the training phase, the Q-learning module indicated the actions on the STCS_LFC based on the training criteria (exploration or reward optimization) which were implemented directly in the installation through the simulation tool [52], obtaining specific feedback for this action. This procedure (realistic live training) is advantageous in the following cases:

- Cases like this in which there is not a real installation in operation from which to obtain measurements and data on its behavior;
- Case studies in which there is only an unreliable and incomplete set of data and measurements on the real behavior of an installation in which interpolations are usually used to cover ranges for which information is not available;
- Cases in which the starting initial information is obtained from simulating the system in general and unrealistic conditions due to the limitations of the used software, for which the obtained feedback presents serious information gaps, so it is necessary to resort again to interpolations and other data-handling techniques.

In Figures 4 and 5, the following variables are considered:

$S_{t,i}$: state in the time step i of the study period;

A_{t-i} : action for the state in the time step i of the study period;
 R_{t-i} : reward for state in the time step i of the study period;
 S_{t-i-1} : previous state to the time step i for the study period;
 A_{t-i-1} : action for the previous state to the time step i of the study period;
 R_{t-i-1} : reward for the previous state to the time step i of the study period.

The optimization of the policy was achieved through the Bellman equation, also known as the equation Q, which is essentially used for updating the Q-values in the Q-table for each episode, as illustrated in Figure 4. The Bellman equation is shown in Equation (1), the variables of which are indicated below:

$Q_{old}(S_t, A_t)$: Q-value of taking an action (A_t) at the state (S_t);
 $Q_{new}(S_t, A_t)$: updated Q-value of taking an action (A_t) at the state (S_t);
 $Q_{max}(S'_t, A_t)$: Q-value of the resulting next state (S'_t) taking the optimal action (A'_t);
 α : learning rate ($0 < \alpha \leq 1$);
 γ : discount factor ($0 < \gamma \leq 1$);
 R_t : reward for the current time step.

$$\underbrace{Q_{new}(S_t, A_t)}_{\text{New value}} \leftarrow (1 - \alpha) \cdot \underbrace{Q(S_t, A_t)}_{\text{Old value}} + \alpha \cdot \left[R_t + \gamma \cdot \underbrace{Q_{max}(S'_t, A_t)}_{\text{estimate of optimal future value}} \right] \quad (1)$$

During the training phase, the agent faces the dilemma of exploring new states while maximizing the global reward at the same time, which is known as the exploration vs. exploitation trade-off. In order to achieve a balance between exploration and exploitation, first, hyperparameter values (α , γ) are intentionally chosen that permit a deep exploration of states and actions and, subsequently, chosen values are more focused on the learning process and on obtaining long-term rewards. Finally, the agent has enough information to make the best decision in the future.

The best trade-off points between training time (number of episodes) and the optimal reward obtained are reached when a convergent pattern to the maximum reward value that can be obtained is observed in the calculated reward values.

2.6. Q-Learning Module for the STCS_LFC Operation Control

The developed Q-learning module for the operating control of the solar thermal cooling system of the proposed case study in the present work was developed using the programming software Python, version 3.10. Table 5 details the used Python modules and their implemented functions in the code of the developed algorithm.

Table 5. Python modules and their implemented functions in the programming code of the developed Q-learning module.

Modules	Functions
Built-in functions	Int()
	Float()
	Print()
	len()
	Format()
Random	random.choice()
Syst	sys.stdin.readline()
	sys.exit()

Table 5. Cont.

Modules	Functions
Numpy	np.load()
	np.random.uniform()
	np.nanargmax
	np.nanmax()
	np.save()
	np.zeros()
	np.array()
	np.set_printoptions()

2.6.1. Parameters of the Q-Learning Algorithm Applied to the Practical Case Study

- States

In this practical case study, each state was represented by a vector of length 6 composed of the variables: time, DNI, ambient temperature and relative humidity, hot water storage tank average temperature and cooling demand of the building. A total of 732 possible states that could occur in this case study were identified, and they are shown in Table 6.

Table 6. Defined states for the case study of this research work.

States	Time	DNI	Ambient Temperature	Relative Humidity	Cooling Demand	Average Tank Temperature
	h	W/m ²	°C	%	kW	°C
1–61	7	256	29	20	0	80–140
62–122	8	521	32	15	290	80–140
123–183	9	658	34,6	14	276	80–140
184–244	10	781	36	9	283	80–140
245–305	11	769	38	8	293	80–140
306–366	12	772	39	10	300	80–140
367–427	13	772	40	10	305	80–140
428–488	14	768	41	10	310	80–140
489–549	15	774	41,6	8	310	80–140
550–610	16	609	42	8	315	80–140
611–671	17	437	41	7	315	80–140
672–732	18	150	40	8	0	80–140

- Actions

The set of actions defined for the training of the agent are shown in Table 7, and they were represented by the variation of the volumetric flow of the hot water entering the generator of the absorption machine. The variation of this flow was performed with volumetric flow steps between 0.000125 m³/s and 0.0025 m³/s.

Table 7. Defined actions for the case study of this research work.

Actions	Hot Water Flow to Be Supplied at the Generator
	m ³ /s
1	0.008125
2	0.00825
3	0.0085
4	0.00875
5	0.009
6	0.00925
7	0.0095
8	0.00975

Table 7. Cont.

Actions	Hot Water Flow to Be Supplied at the Generator
	m ³ /s
9	0.01
10	0.01025
11	0.0105
12	0.01075
13	0.011
14	0.011125
15	0.01125

The actions were defined according to the operating range of the absorption machine dimensioned to satisfy the cooling demand of the analyzed building considering the indications of the manufacturer's datasheet.

- Reward

The reward was the opposite value of the sum of the penalties stipulated for the auxiliary energy consumption, the electrical consumption of the pump that supplies hot water to the absorption machine and the precision in the generation of the required cooling energy at each time step for a certain state and for a specific selected action.

The total reward of an episode was calculated using the Equation (2), considering the rewards for auxiliary energy consumption (R_A), electrical consumption of the pump that supplies hot water to the absorption machine (R_B) and the precision in the generation of the required cooling energy (R_C) for each time step. These values were estimated independently using Equations (3)–(5), respectively.

Equations (3)–(5) were obtained through an iterative process and defined in order to guarantee a significant gradient that maximizes the target value to be achieved (R_A , R_B and R_C) considering the maximum and minimum values that the proposed installation can reach in terms of auxiliary energy consumption, electrical consumption and cooling capacity.

The weight of each type of reward (R_A , R_B and R_C) in Equation (2) was selected and customized by the authors, prioritizing low auxiliary energy consumption over low electrical consumption and precision in the generation of the required cooling demand.

$$Total\ reward_{episode} = \sum_{i=1}^n [R_A + R_B + R_C] \quad (2)$$

$$R_A = 0.95 \cdot e^{-0.02944 \cdot Percentage_{Q_{Auxiliar\ energy}i}} \quad (3)$$

$$R_B = -0.0025 \cdot \left(Percentage_{P_{electric\ of\ pump\ Ci}} \right) + 0.25 \quad (4)$$

$$R_C = \left(-4.46 \times 10^{-2} + 4.88 \times 10^{-3} \cdot \left(Percentage_{Q_{e_i}} \right) - 2.38 \times 10^{-5} \cdot \left(Percentage_{Q_{e_i}}^2 \right) \right) \quad (5)$$

$$Percentage_{Q_{Auxiliar\ energy}i} = \left[\left(\frac{Q_{Auxiliar\ energy}i}{Q_{generatori}} \right) \cdot 100 \right] \quad (6)$$

$$Percentage_{P_{electric\ of\ pump\ Ci}} = \left[\left(\frac{P_{electric\ of\ pump\ Ci}}{P_{max\ electric\ of\ pump\ C}} \right) \cdot 100 \right] \quad (7)$$

$$Percentage_{Q_{e_i}} = \left[\left(\frac{Q_{e_i}}{Q_{cooling\ demand_i}} \right) \cdot 100 \right] \quad (8)$$

In Equations (6)–(8), the following variables are considered:

$Percentage_{Q_{Auxiliar\ energy}i}$: percentage of the total required heat transfer rate at the generator that is supplied by the auxiliary system in the i time step of the study day, [%];

$Q_{Auxiliary\ energy\ i}$: heat transfer rate supply by the auxiliary system in the i time step of the study day, [kW];

$Q_{generator\ i}$: required heat transfer rate in the generator of the absorption machine in the i time step of the study day, [kW];

$Percentage_{p_{electric\ of\ pump\ C\ i}}$: percentage of the consumption of the pump of the sub-circuit C regarding its maximum consumption in the i time step of the study day, [%];

$P_{electric\ of\ pump\ C\ i}$: electric consumption of the pump of the sub-circuit C in the i time step of the study day, [kW];

$P_{max\ electric\ of\ pump\ C}$: design maximum electric consumption of the pump of the sub-circuit C [kW];

$Percentage_{Q_{e\ i}}$: percentage of the generated cooling by the absorption machine regarding the cooling demand in the i time step of the study day, [%];

$Q_{e\ i}$: generated cooling by the absorption machine in the i time step of the study day, [kW];

$Q_{cooling\ demand\ i}$: cooling demand in the i time step of the study day, [kW].

2.6.2. QL Module Algorithm

The algorithm, in Python, of the developed Q-learning module for the case study of this research work and its detailed flowchart are presented in the Appendix A and in Figure 6, respectively. In this algorithm, the updating of the Q-value was performed on delay.

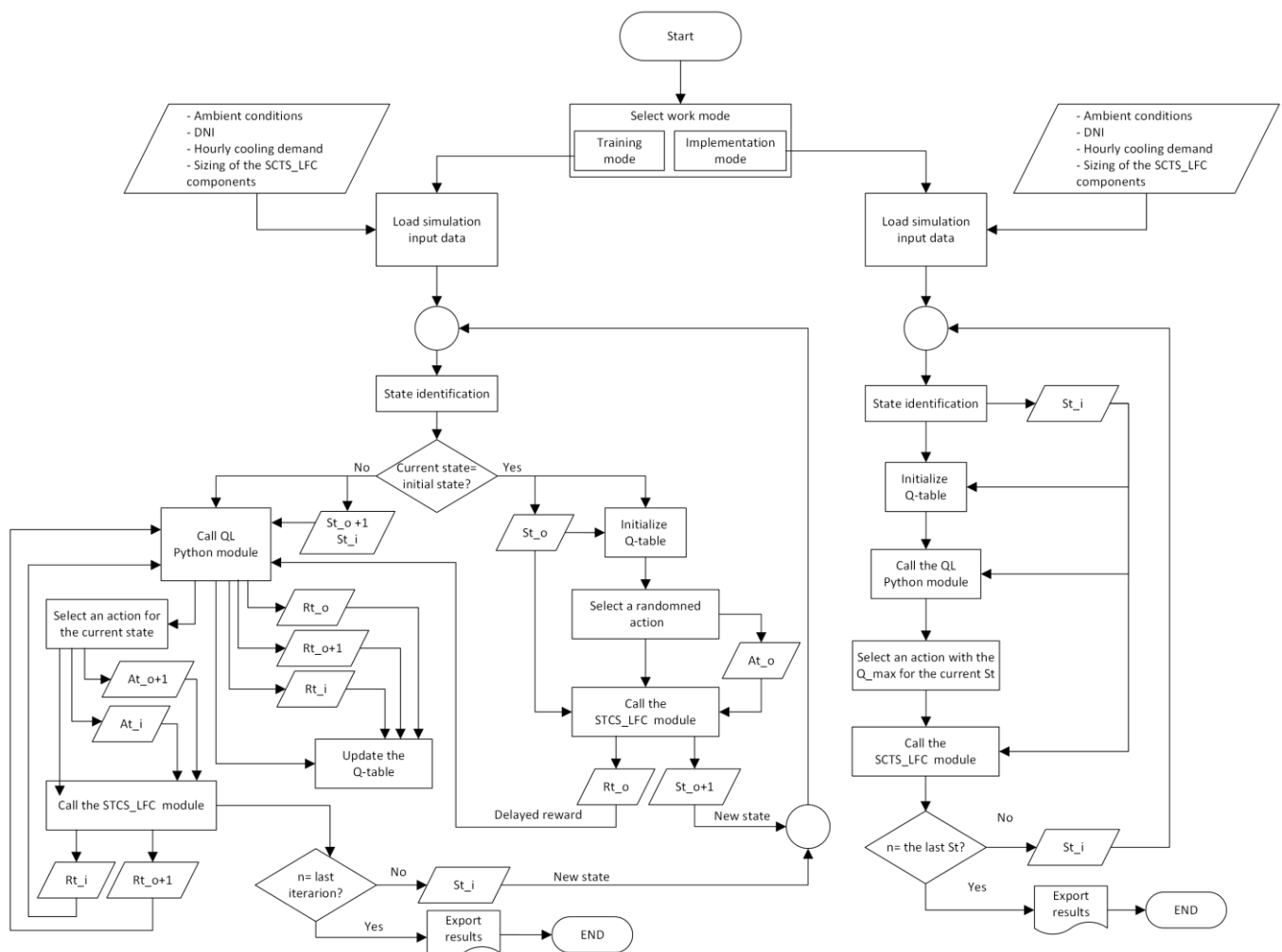


Figure 6. Detailed Q-learning algorithm flowchart for the case study of the present research work.

In Figure 6 the following variables are considered:

S_{t_0} : initial state of the study period;
 A_{t_0} : action in the initial state of the study period;
 R_{t_0} : reward for the applied action in the initial state of the study period;
 S_{t_0+1} : next state to the initial;
 A_{t_0+1} : action for the next state to the initial;
 R_{t_0+1} : reward for the next state to the initial;
 S_{t_i} : state in the time step i of the study period;
 A_{t_i} : action in the time step i of the study period;
 R_{t_i} : reward for state in the time step i of the study period.

3. Results

3.1. Case Study Results

This section analyzes the obtained results from the simulations carried out on the solar cooling system of the proposed case study. This installation was simulated operating under a predictive control system and under a reinforcement learning control approach.

The reinforcement learning control approach was studied considering a trained agent for different numbers of iterations or episodes (1, 10, 100, 1000 and 1500) for the defined study period.

Figure 7 illustrates the obtained total reward for each episode performed during the training of the agent Q. The total reward represented the sum of the obtained rewards at every time step during each episode, calculated using the Equation (2).

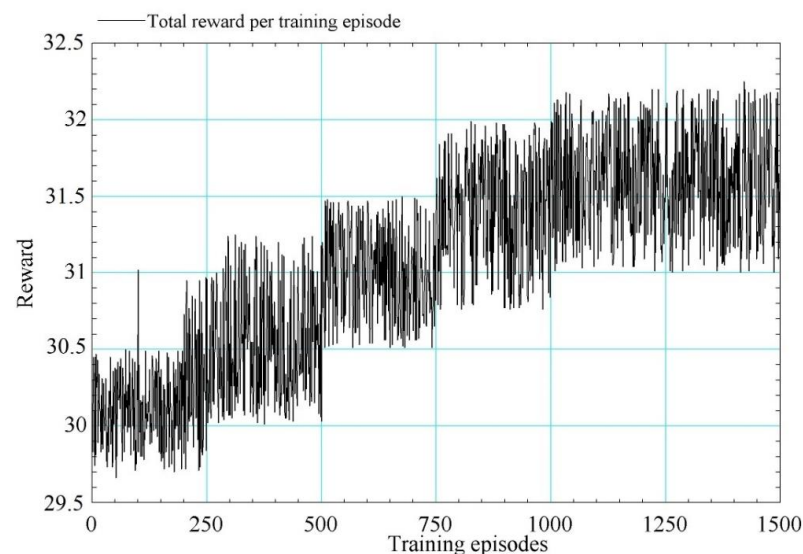


Figure 7. Total reward per episode during the training phase of the agent Q.

The value of the total reward in the first episode was 29.83, and its value kept oscillating between 29.66 and 31 during the following 250 episodes. After these episodes, the values of the total reward began to increase progressively until they converged towards a value of 32 in the last 500 episodes, approximately. The minimum and maximum total reward reached were 29.66 and 32.25 in episodes 52 and 1421, respectively, which means a difference of 8.73%.

The pattern of the curve shown in Figure 7 indicates the learning process of the Q agent. During the first 655 iterations, the observed behavior was oscillating and represented a phase marked more by the exploration of the agent than by the search for the optimal sequence of actions that maximize the reward. During this interval of episodes, the authors intentionally chose values of the hyperparameters (α , γ) that allowed the agent to perform a deep exploration of the states and the proposed actions.

From approximately episode 655, a convergent pattern was observed towards a total reward value of 32, which indicates the learning of agent Q. During this interval, the authors chose values of hyperparameters (α , γ) more focused on the learning process (incorporation of knowledge acquired in previous episodes) and on obtaining a long-term reward.

The training of Q agent in the present study was completed in a total of 1500 iterations with an acceptable level of convergence, taking a time of approximately 9.5 h on a computer with the following characteristics: Intel (R) Core (TM) i7-9750H processor unit and 32 GB in RAM memory.

Figures 8–10 show the comparative evolution of the behavior of the simulated solar thermal cooling system as a function of the number of training episodes performed.

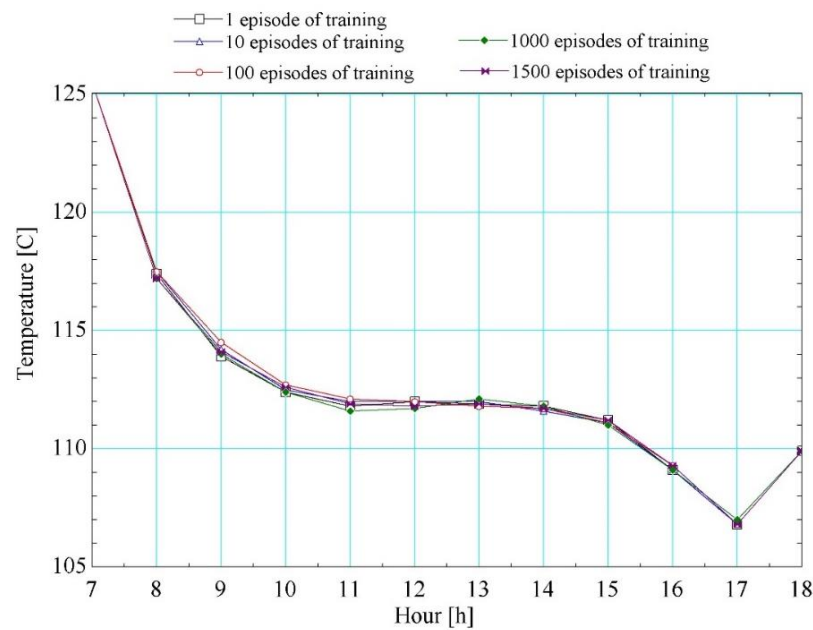


Figure 8. Hourly average tank temperature for different number of episodes.

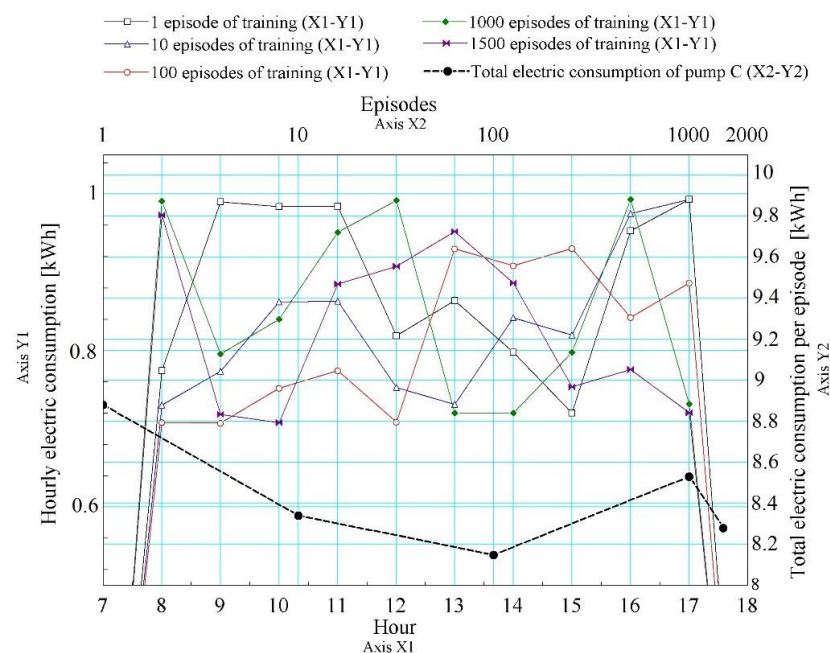


Figure 9. Electric consumption of the pump in sub-circuit C for different numbers of episodes.

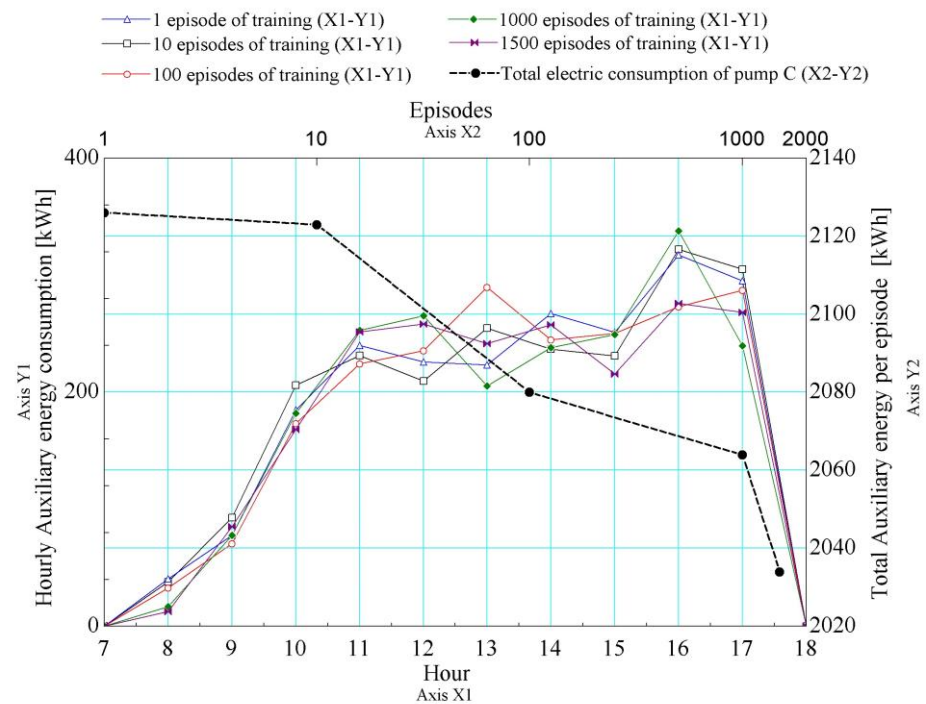


Figure 10. Auxiliary energy requirements for different numbers of episodes.

Figure 8 illustrates the hourly average tank temperature as a function of agent training episodes resulting from the simulations of the proposed cooling installation using the reinforcement-learning-based control system. A decreasing and very regular pattern in terms of the trajectory of the resulting curves for each number of the studied episodes was observed, with slight variations in the value of the obtained temperature for each hour of the studied day. This convergent behavior indicates the limitations or restrictions of the simulated solar cooling system regarding the solar thermal energy absorbed and stored compared to the requirements with which the RL algorithm has to deal to optimize it and thereby to reduce the contribution of auxiliary energy.

A decrease in the total electrical consumption of the pump that recirculates water in sub-circuit C of the system shown in Figure 1 at the end of the training was observed, as shown in Figure 9, which reflects the learning process of the agent and the proper selection and functioning of the proposed reward equation. The achieved reduction in electrical consumption between the beginning and the end of the training was 7.25%.

The influence of the number of training episodes on the auxiliary energy requirement of the solar thermal cooling system of the present case study is presented in Figure 10. In this, a reduction in the auxiliary energy consumption was observed with the increment in the number of training episodes of the agent, which demonstrates, once again, the learning process of the agent and the proper selection and functioning of the proposed reward equation. The optimization in the auxiliary energy consumption reached within the training process was 4.52%.

The performance of the solar thermal cooling system of the present case study operating with a predictive control model and with a reinforcement learning control model was analyzed using the obtained results from the simulations, which are represented in Figures 11–15.

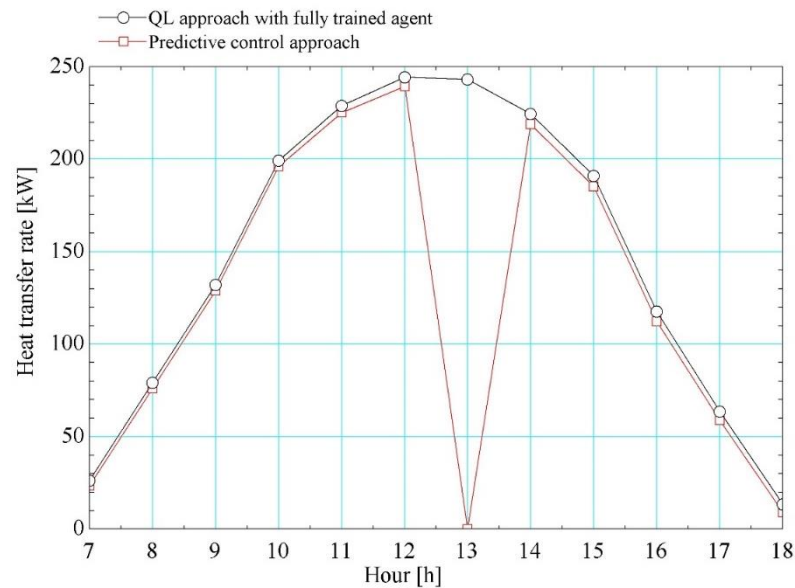


Figure 11. Total net absorbed solar heat transfer rate with each implemented control approach.

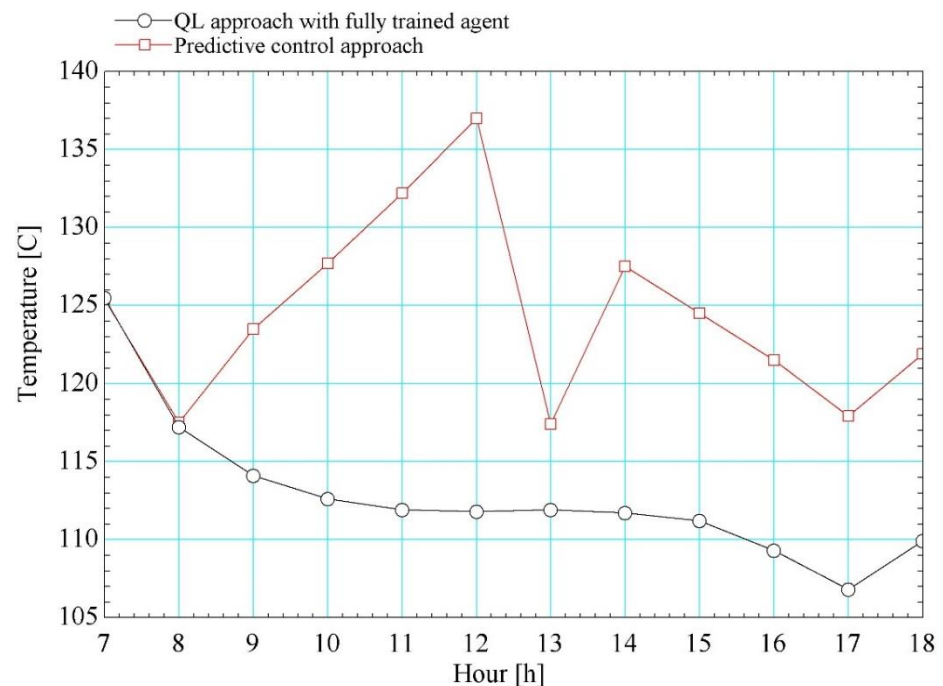


Figure 12. Hourly average tank temperature with each implemented control approach.

In Figure 11, the hourly total net absorbed solar heat transfer rate by the collector is represented for each studied control approach. In this, it can be seen that when the installation operated with the control system based on RL, the solar collection system worked from 7 a.m. to 6 p.m. without interruptions. However, when the installation was operated by the predictive control system, a deactivation of the solar collectors was observed after 13 h.

The deactivation at 13 h occurred because the predictive control system during the first hours of the day was not able to find the optimal combination of water flow and temperature required by the generator of the absorption machine to satisfy cooling demand, deciding to store part of the absorbed solar energy and use auxiliary energy to drive the absorption machine. This energy storage resulted in a progressive increase in the

temperature of the tank until the maximum setpoint temperature was reached, causing the deactivation of the solar collectors, as can be seen in Figure 12.

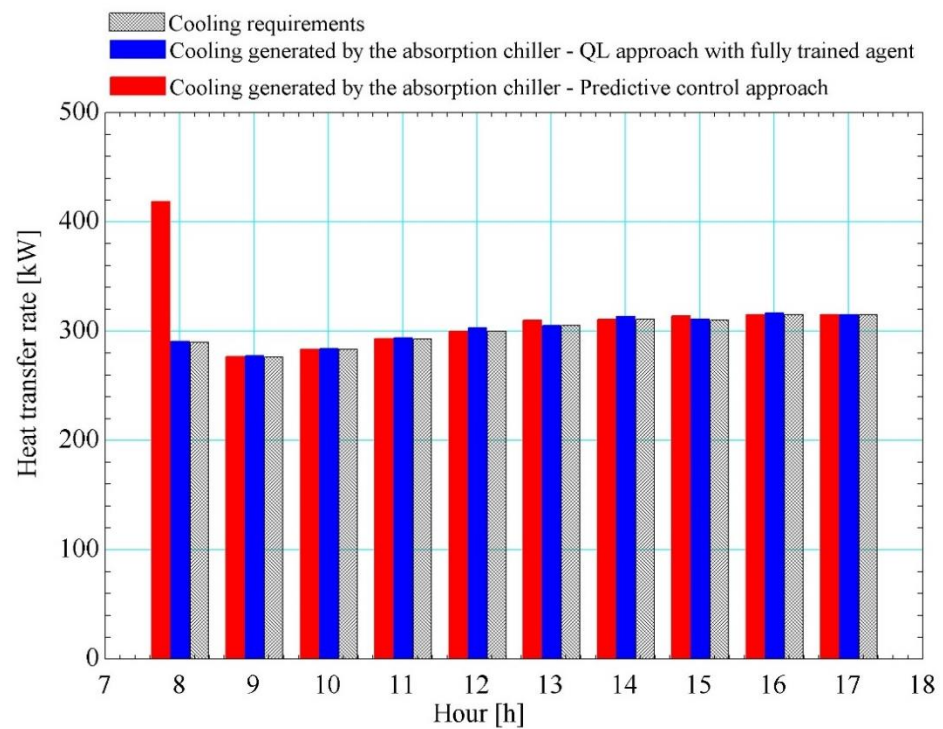


Figure 13. Cooling generated with each type of implemented control approach.

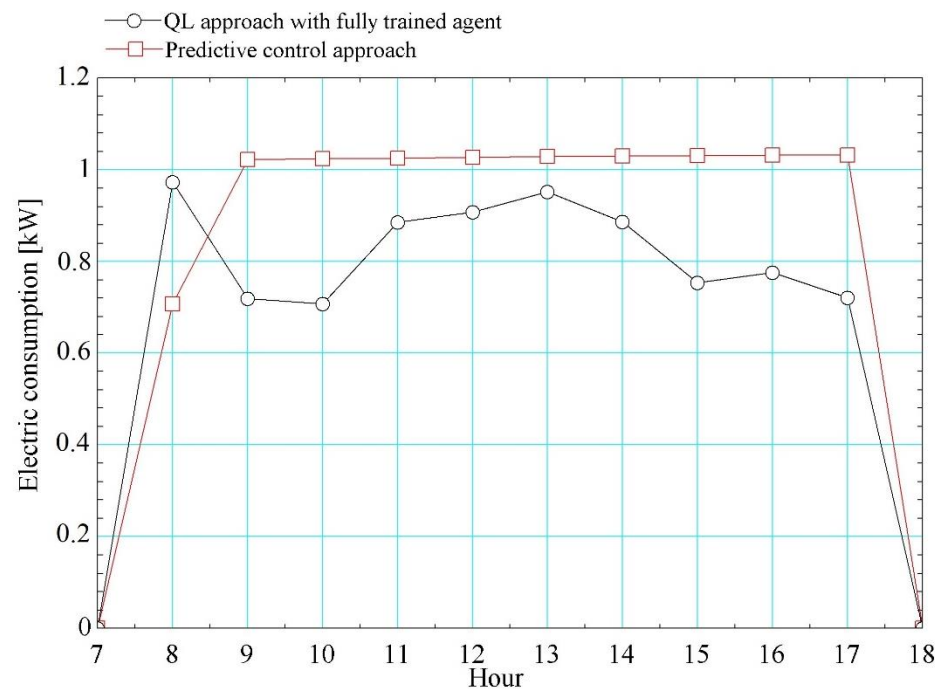


Figure 14. Electric consumption of the pump in sub-circuit C for the different implemented control approaches.

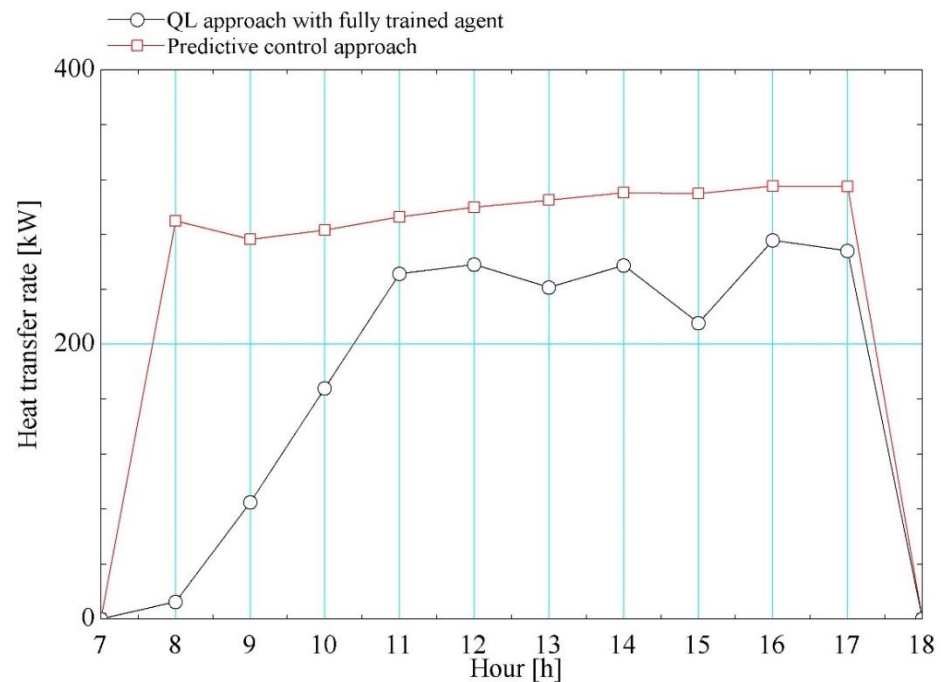


Figure 15. Hourly auxiliary energy requirements for the different implemented control approaches.

The RL control approach is capable of finding the optimum quantity and temperature of the water leaving the hot water storage tank to be supplied to the absorption machine for satisfying the cooling requirements. The foregoing remark is reflected in Figure 12 with the decreasing of the tank temperature, which enabled the solar collection system to remain operational during the hours of availability of solar energy.

The capability of the control system based on RL to regulate the solar cooling system in a more optimal way allowed the absorption of 19.72% more solar energy compared to the solar energy absorbed when the installation operated with the predictive control system.

Figure 13 illustrates comparatively the hourly cooling generation of the absorption machine operating with each type of control approach proposed in this study. The graph shows a greater precision of the control system based on QL compared to the predictive control system in the satisfaction of the cooling demand of the system. At 8 o'clock, a deviation in the predictive control system led to an overproduction of cooling, increasing the amount of thermal energy that must be provided, which, in this case, was auxiliary energy, as shown in Figure 15.

The electrical consumption of the pump in the sub-circuit C when the STCS of the present case study operated with the control system based on RL was less than when the operation was controlled with the predictive control, as illustrated in Figure 14. The reduction of the consumption of the pump of this sub-circuit achieved by the RL control approach with respect to the consumption with the predictive control system reached 1.68 kWh, that is, 16.86% lower.

Figure 15 shows the hourly heat transfer rate supplied by the auxiliary system to satisfy the cooling demand of the building. In this, it can be observed that the thermal requirements supplied by the auxiliary system were lower for every hour of the studied day when the installation operated with the control system based on RL due to the ability of this control approach to regulate in a more optimal way the operation of the installation.

The high consumption of auxiliary energy observed when the solar cooling installation operated with the predictive control system was due to its inability to find, during the first hours of the studied day, the optimal combination of water flow and temperature to supply the generator of the absorption machine according to the cooling demand, generally resulting in a higher consumption of auxiliary energy.

The achieved optimization in the auxiliary energy consumption through the use of the RL control approach compared to the predictive control system was 1080 kWh, that is, a reduction of 34.68%.

3.2. Defiant Aspects of Control Models Based on RL for Operating Solar Thermal Cooling Systems

During the realization of this research work, some defiant aspects and drawbacks were identified in the applicability of control models based on RL for the operation and functioning of solar thermal cooling systems, which are described below:

- Need for a large amount of data for agent training: To perform agent training, a large amount of data is required, which can be obtained directly from the environment and the installed system. In general, this amount of data is not available, so modeling of the installation and the environment to be treated is commonly used. If this were the case, a large number of simulations would be required to adequately complete the agent training;
- Complexity in the definition of the reward equation and the hyperparameters: The definition of the reward equation is not a trivial task; it requires significant knowledge about the operation of the system, the mathematical models and computer programs to be used. The effectiveness of the RL model is conditioned to the appropriate choice of states, actions, rewards and the hyperparameters (α , γ), and it also requires a high number of simulations of episodes that allow the correct learning of the agent. The choice of the hyperparameters (α , γ) and the reward equation must be adjusted according to the obtained results and according to the observed level of convergence. Each solar cooling installation is different; they have their own characteristics in terms of configuration, capacity, residual energy availability, climatic conditions, etc. Therefore, the reward equation and the value of the hyperparameters have to be recalculated and particularized. The foregoing results in a high number of hours of simulations, analysis and operating adjustments, which obviously have an economic impact;
- Complexity in the manipulation of data and interrelationships between variables: One of the main drawbacks or challenges in the implementation of RL models in the regulation of solar thermal cooling systems is the complex programming and data manipulation that are necessary due to the high number of involved variables in their operation, as well as for the complex interaction between them. One way of dealing with the aforementioned complexity in programming and data management is through the combined and integrated use of two programs, one specifically designed to perform the simulations of thermal systems and another for the data handling and complex programming. This procedure was successfully used in this research work, allowing great practicality, ease, time savings and efficiency during the agent training stage as well as in the operation of the installation with the control system based on RL (with the agent already trained).

Despite the defiant aspects and drawbacks mentioned above regarding the applicability of RL-based control models, the obtained results demonstrate their enormous potential in the regulation and operation of this type of system.

3.3. Control Scheme for Solar Thermal Cooling System Based on RL for Real Implementation

In Figure 16, a schematic diagram of a control model based on Q-learning for real implementation is presented at a conceptual level. Depending on the complexity of the system, the number of states, actions and variables to be controlled, it may be necessary to use a more powerful algorithm such as one for deep Q-learning.

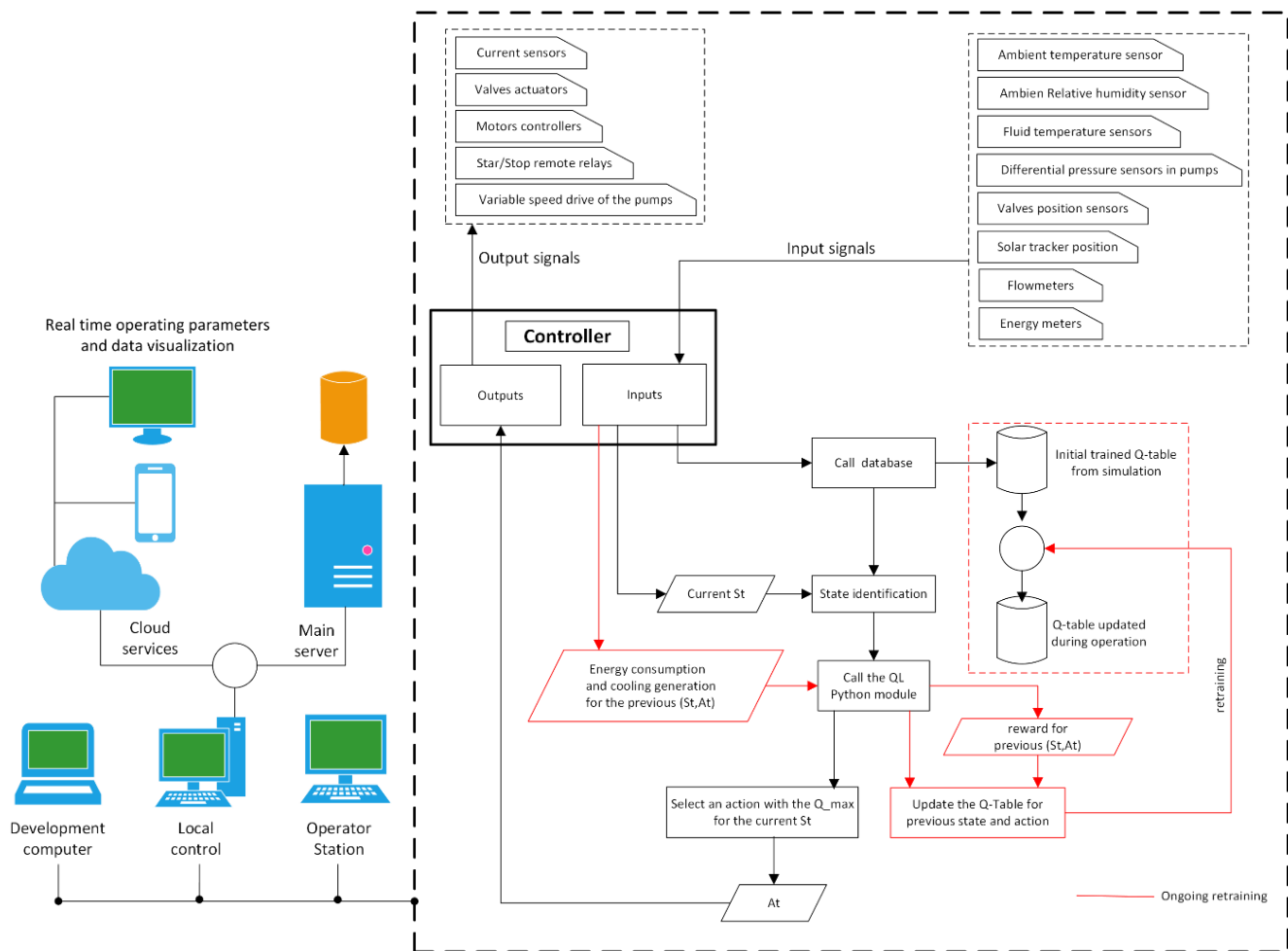


Figure 16. Conceptual control scheme for STCS_LFR based on RL for real implementation.

4. Conclusions

A control model based on the reinforcement learning technique was implemented for the operation of the solar thermal cooling system driven by linear Fresnel collectors proposed for the case study of this research work.

The implementation of this control model was carried out through the interaction of a QL module developed in Python and a simulation module of a SCTS_LFC developed in EES. This interaction was achieved through a programming procedure developed in EES that served as a communication gateway between both modules.

In the absence of experimental data on the studied building, the HVAC system and the solar cooling installation, the agent was trained using models. The solar thermal cooling system driven by linear Fresnel collectors was simulated through the tool [52], and the modeling of the building and the HVAC system was carried out using the software Design Builder. Depending on the required degree of detail, the model of a building and its corresponding HVAC system may require a time of 40 to 100 h.

The obtained results from the simulations carried out demonstrate the advantages of the control system based on reinforcement learning compared to the predictive control approach. For the proposed case study, a reduction of 35% in the auxiliary energy requirements and 17% in the electrical consumption of the pump that feeds the absorption machine was observed when the solar cooling installation was operated with a control system based on RL.

Although, in the present research work, for simplicity, only some variables of the SCTS_LFC were considered to be controlled from the developed RL approach, the potential

of this control approach for the controlling and operating of this type of thermal system was verified, and, thereby, demonstrated that reinforcement learning can open a new door in energy optimization not only for solar thermal cooling systems but also in other types of thermodynamic systems with sequential processes.

The effectiveness of the RL model is conditioned to the appropriate choice of states, actions, rewards and the hyperparameters (α , γ), and it also requires a high number of simulations of episodes that allow the correct learning of the agent.

The choice of the hyperparameters (α , γ) must be adjusted according to the obtained results and according to the observed level of convergence. In this research work, one thousand and five hundred episodes were performed for the agent training. Firstly, hyperparameter values were chosen that allowed a deep exploration by the agent, and, then, more importance was given to achieving greater long-term rewards.

The time required for agent training was 9.5 h using a computer with a processor Intel (R) Core (TM) i7-9750H and 32 GB in RAM memory.

Finally, a control scheme based on RL for solar thermal cooling systems driven by linear Fresnel collectors was proposed at a conceptual level, in which all the variables of the various components of the system that could be controlled for the optimization of its operation were considered. Future research could analyze the scenario described above, although it would be necessary to use more powerful QL algorithms that allow the handling of a greater number of variables, states and actions.

Author Contributions: All authors participated equally in the conceptualization, investigation, writing—original draft preparation and in the writing—review and editing. Conceptualization, J.J.D. and J.A.F.; Formal analysis, J.J.D. and J.A.F.; Investigation, J.J.D.; Methodology, J.J.D. and J.A.F.; Writing—original draft, J.J.D. and J.A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The algorithm of the Q-learning module developed in Python for the case study of the present research works is shown below:

```

“Q-learning module with EES connection”
import math
import sys
import string
import numpy as np
import random

SR=sys.stdin.readline()
mode=int(sys.stdin.readline())

if (mode==−1)
    SR=“Call python(‘SCTS_QL_module.pyw’, Previous_State, Previous_Action, Previous_Reward, Alpha, Gamma, epsilon, j, State: Next_action)”
    if (mode==−2): #return units of the inputs
        SR=‘unitless, unitless, unitless, unitless, unitless, unitless, unitless, unitless ’
    if (mode==−3): #return units of the outputs
        SR=‘unitless’
    if (mode<0):
        print(SR)
        sys.exit(0)#return

Previous_State=float(sys.stdin.readline())
Previous_Action=float(sys.stdin.readline())

```

```

Previous_Reward=float(sys.stdin.readline())
State=float(sys.stdin.readline())

Alpha=float(sys.stdin.readline())
Gamma=float(sys.stdin.readline())
epsilon=float(sys.stdin.readline())
j=float(sys.stdin.readline())

q=np.load("Table Q_SCTS.npy")

States=q[0:len(q)]
Actions=q[:,len(q)]

try:
    Current_state=int((State))
    P_St=int((Previous_State))
    P_Act=int((Previous_Action))
    q_previous_sa=q[P_St,P_Act]
if (j = 1):
    Random_action=random.choice([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15])
    Next_action=Random_action
    q_current_sa=q[Current_state,Next_action]
    qmax_Current_state=0
else:
    Training_choice=np.random.uniform(0,1)

    if (Training_choice <= epsilon):
        Random_action=random.choice([0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15])
        Next_action=Random_action
        q_current_sa=q[Current_state,Next_action]
        qmax_Current_state=0
    else:
        Next_action=np.nanargmax(q[Current_state])
        qmax_Current_state=np.nanmax(q[Current_state])
        q_current_sa=qmax_Current_state
q_new=((1-Alpha)*(q_previous_sa))+Alpha*(Previous_Reward+(Gamma*q_current_sa))
q[P_St,P_Act]=q_new
np.save("Table Q_SCTS",q)

SR=""
mode=0 #return mode=0 if there are no errors

except:
    SR='Unexpected error occurred in the Python program.'
    mode=1
    Next_action=-999

```

References

1. Santos, A. *Control Strategies and Algorithms for Obtaining Energy Flexibility in Buildings Energy in Buildings and Communities Programme Annex 67 Energy Flexible Buildings*; Danish Technological Institute: Taastrup, Denmark, 2019.
2. Belic, F.; Hocenski, Z.; Sliskovic, D. HVAC control methods—A review. In Proceedings of the 2015 19th International Conference on System Theory, Control and Computing (ICSTCC) 2015, Cheile Gradistei, Romania, 14–16 October 2015.
3. Mirinejad, H.; Sadati, H.; Maryam, G.; Hamid, T. Control Techniques in Heating, Ventilating and Air Conditioning (HVAC) Systems 1. *J. Comput. Sci.* **2008**, *4*, 777–783. [[CrossRef](#)]
4. Behrooz, F.; Mariun, N.; Marhaban, M.; Mohd Radzi, M.A.; Ramli, A. Review of Control Techniques for HVAC Systems—Nonlinearity Approaches Based on Fuzzy Cognitive Maps. *Energies* **2018**, *11*, 495. [[CrossRef](#)]
5. Maldonado, D.; Luján, M.; Rosiek, S.; Batlles, J.; Ushak, S. Exergy analysis of a solar heating and cooling system that uses phase change materials. *PCMSOL Project. Acta Nova* **2019**, *9*, 299–328.

6. Psimopoulos, E.; Bee, E.; Luthander, R.; Bales, C. Smart Control Strategy for PV and Heat Pump System Utilizing Thermal and Electrical Storage and Forecast Services. In Proceedings of the ISES Solar World Congress 2017 and the IEA SHC International conference on Solar Heating and Cooling for Buildings and Industry, Abu Dhabi, UAE, 29 October–2 November 2017.
7. Liu, H.; Zabinsky, Z.B.; Kohn, W. Rule-based control system design for smart grids. In Proceedings of the IEEE PES General Meeting, Minneapolis, MN, USA, 25–29 July 2010; pp. 1–5.
8. Han, J.; Li, X.; Tang, T. Energy Management Using a Rule-Based Control Strategy of Marine Current Power System with Energy Storage System. *J. Mar. Sci. Eng.* **2021**, *9*, 669. [\[CrossRef\]](#)
9. Kanwar, A.; Hidalgo Rodriguez, D.; Von Appen, J.; Braun, M. A Comparative Study of Optimization- and Rule-Based Control for Microgrid Operation. In Proceedings of the Power and Energy Student Summit(PESS), Dortmund, Germany, 13–14 January 2015. [\[CrossRef\]](#)
10. Altes-Buch, Q.; Orosz, M.; Quoilin, S.; Lemort, V. Rule-based control and optimization of a hybrid solar microgrid for rural electrification and heat supply in sub-Saharan Africa. In Proceedings of the 30th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS), San Diego, CA, USA, 2–6 July 2017.
11. Dorokhova, M.; Ballif, C.; Wyrsh, N. Rule-Based Scheduling of Air Conditioning Using Occupancy Forecasting. *Energy AI* **2020**, *2*, 100022. [\[CrossRef\]](#)
12. Clauß, J.; Stinner, S.; Sartori, I.; Georges, L. Predictive rule-based control to activate the energy flexibility of Norwegian residential buildings: Case of an air-source heat pump and direct electric heating. *Appl. Energy* **2019**, *237*, 500–518. [\[CrossRef\]](#)
13. Pinamonti, M.; Prada, A.; Baggio, P. Rule-Based Control Strategy to Increase Photovoltaic Self-Consumption of a Modulating Heat Pump Using Water Storages and Building Mass Activation. *Energies* **2020**, *13*, 6282. [\[CrossRef\]](#)
14. Solberg, V.U. Reinforcement Learning for Grid Control in an Electric Distribution System. Master 's Thesis, Norwegian University of Life Sciences, AS, Norway, 13 May 2019.
15. Casini, M. Chapter 10—Building automation systems. In *Woodhead Publishing Series in Civil and Structural Engineering*; Casini, M.B.T.-C., Ed.; Woodhead Publishing: Cambridge, UK, 2022; pp. 525–581. ISBN 978-0-12-821797-9.
16. Ferreira, P.M.; Silva, S.M.; Ruano, A.E. Model based predictive control of HVAC systems for human thermal comfort and energy consumption minimisation. *IFAC Proc. Vol.* **2012**, *45*, 236–241. [\[CrossRef\]](#)
17. Komareji, M.; Stoustrup, J.; Rasmussen, H.; Bidstrup, N.; Svendsen, P.; Nielsen, F. Optimal model-based control in HVAC systems. In Proceedings of the 2008 American Control Conference, Seattle, WA, USA, 11–13 June 2008; pp. 1443–1448.
18. Xia, L.; Ma, Z.; Kokogiannakis, G.; Wang, S.; Gong, X. A model-based optimal control strategy for ground source heat pump systems with integrated solar photovoltaic thermal collectors. *Appl. Energy* **2018**, *228*, 1399–1412. [\[CrossRef\]](#)
19. Ferhatbegovic, T.; Zucker, G.; Palensky, P. Model Based Predictive Control for a Solar-Thermal System. In Proceedings of the 10th IEEE AFRICON, Livingstone, Zambia, 13–15 September 2011.
20. Unterberger, V.; Muschick, D.; Gölles, M. Model-Based Control Strategies for an Efficient Integration of Solar Thermal Plants Into District Heating Grids. In Proceedings of the ISES Solar World Conference 2017 and the IEA SHC Solar Heating and Cooling Conference for Buildings and Industry 2017, Abu Dhabi, UAE, 29 October–2 November 2017.
21. Maasoumy, M.; Pinto, A.; Sangiovanni-Vincentelli, A. Model-Based Hierarchical Optimal Control Design for HVAC Systems. In Proceedings of the ASME 2011 Dynamic Systems and Control Conference and Bath/ASME Symposium on Fluid Power and Motion Control, DSCC 2011, Arlington, VA, USA, 2 November–31 October 2011. [\[CrossRef\]](#)
22. Ahmed, O. Model-Based Control of Laboratory HVAC Systems. Doctoral Thesis, University of Wisconsin, Madison, WI, USA, 1996.
23. Kardos, T.; Kutasi, D.N. Modelling and Model-Based Control of an HVAC System. *Műszaki Tudományos Közlemények* **2019**, *10*, 25–30. [\[CrossRef\]](#)
24. Schwenzer, M.; Ay, M.; Bergs, T.; Abel, D. Review on model predictive control: An engineering perspective. *Int. J. Adv. Manuf. Technol.* **2021**, *117*, 1327–1349. [\[CrossRef\]](#)
25. Gao, G.; Li, J.; Wen, Y. Energy-Efficient Thermal Comfort Control in Smart Buildings via Deep Reinforcement Learning. *arXiv* **2019**, arXiv:1901.04693.
26. Zhang, H.; Seal, S.; Wu, D.; Bouffard, F.; Boulet, B. Data-driven Model Predictive and Reinforcement Learning-Based Control for Building Energy Management: A Survey. *IEEE Access* **2022**, *10*, 27853–27862. [\[CrossRef\]](#)
27. Han, M.; Zhang, X.; Xu, L.; May, R.; Pan, S.; Wu, J. *A Review of Reinforcement Learning Methodologies on Control Systems for Building Energy*; Höskolan Dalarna: Borlänge, Sweden, 2018.
28. Wang, Z.; Hong, T. Reinforcement learning for building controls: The opportunities and challenges. *Appl. Energy* **2020**, *269*, 115036. [\[CrossRef\]](#)
29. Dermardiros, V.; Bucking, S.; Athienitis, A. Simplified Building Controls Environment with a Reinforcement Learning Application. In Proceedings of the 16 th IBPSA International Conference and Exhibition, Rome, Italy, 2–4 September 2019; pp. 956–964.
30. Ojand, K.; Dagdougui, H. Q-Learning-Based Model Predictive Control for Energy Management in Residential Aggregator. *IEEE Trans. Autom. Sci. Eng.* **2021**, *19*, 70–81. [\[CrossRef\]](#)
31. Yu, L.; Qin, S.; Zhang, M.; Shen, C.; Jiang, T.; Guan, X. A Review of Deep Reinforcement Learning for Smart Building Energy Management. *IEEE Internet Things J.* **2021**, *8*, 12046–12063. [\[CrossRef\]](#)
32. Bettoni, D.; Soppelsa, A.; Fedrizzi, R.; del Toro Matamoros, R.M. Analysis and Adaptation of Q-Learning Algorithm to Expert Controls of a Solar Domestic Hot Water System. *Appl. Syst. Innov.* **2019**, *2*, 15. [\[CrossRef\]](#)

33. Peirelinck, T.; Ruelens, F.; Decnoninck, G. Using Reinforcement Learning for Optimizing Heat Pump Control in a Building Model in Modelica. In Proceedings of the 2018 IEEE International Energy Conference (ENERGYCON), Limassol, Cyprus, 3–7 June 2018.
34. Zhang, Z.; Lam, K. Practical Implementation and Evaluation of Deep Reinforcement Learning Control for a Radiant Heating System. In Proceedings of the 5th Conference on Systems for Built Environments, Shenzhen China, 7–8 November 2018.
35. Overgaard, A.; Kallesøe, C.; Bendtsen, J.; Nielsen, B. Mixing Loop Control using Reinforcement Learning. *E3S Web Conf.* **2019**, *111*, 05013. [\[CrossRef\]](#)
36. Wystrcil, D.; Kalz, D. Model-Based Optimization of Control Strategies for Low-Exergy Space Heating Systems Using an Environmental Heat Source. In Proceedings of the 13th Conference of International Building Performance Simulation Association, Chambéry, France, 26–28 August 2013.
37. Chen, B.; Cai, Z.; Bergés, M. Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy. In Proceedings of the 6th ACM International Conference, New York, NY, USA, 13–14 November 2019. ISBN 978-1-4503-7005-9.
38. Ahn, K.U.; Park, C.S. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. *Sci. Technol. Built Environ.* **2020**, *26*, 61–74. [\[CrossRef\]](#)
39. Wei, T.; Wang, Y.; Zhu, Q. Deep Reinforcement Learning for Building HVAC Control. In Proceedings of the 2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC), Austin, TX, USA, 18–22 June 2017.
40. Le, D.; Yingbo, L.; Wang, R.; Tan, R.; Wong, Y.; Wen, Y. Control of Air Free-Cooled Data Centers in Tropics via Deep Reinforcement Learning. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities and transportation, New York, NY, USA, 13–14 November 2019. ISBN 978-1-4503-7005-9.
41. Zhang, Z.; Chong, A.; Pan, Y.; Zhang, C.; Lu, S.; Lam, K. A Deep Reinforcement Learning Approach to Using Whole Building Energy Model For HVAC Optimal Control. In Proceedings of the 2018 ASHRAE/IBPSA-USA Building Performance Analysis Conference and SimBuild, Chicago, IL, USA, 26–28 September 2018.
42. Biemann, M.; Scheller, F.; Liu, X.; Huang, L. Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control. *Appl. Energy* **2021**, *298*, 117164. [\[CrossRef\]](#)
43. Chen, Y.; Norford, L.; Samuelson, H.; Malkawi, A. Optimal Control of HVAC and Window Systems for Natural Ventilation Through Reinforcement Learning. *Energy Build.* **2018**, *169*, 195–205. [\[CrossRef\]](#)
44. Perera, A.T.D.; Wickramasinghe, U.; Nik, V.; Scartezzini, J.-L. Introducing reinforcement learning to the energy system design process. *Appl. Energy* **2020**, *262*, 114580. [\[CrossRef\]](#)
45. Chen, X.; Qu, G.; Tang, Y.; Low, S.; Li, N. Reinforcement Learning for Decision-Making and Control in Power Systems: Tutorial, Review, and Vision. *arXiv* **2021**, arXiv:2102.01168.
46. Zhang, Z.; Zhang, D.; Qiu, R.C. Deep reinforcement learning for power system applications: An overview. *CSEE J. Power Energy Syst.* **2020**, *6*, 213–225. [\[CrossRef\]](#)
47. Perera, A.T.D.; Kamalaruban, P. Applications of reinforcement learning in energy systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [\[CrossRef\]](#)
48. Rioual, Y.; Laurent, J.; Diguët, J.-P. Reinforcement-Learning Approach Guidelines for Energy Management. *J. Low Power Electron.* **2019**, *15*, 283–293. [\[CrossRef\]](#)
49. Kohne, T.; Ranzau, H.; Panten, N.; Weigold, M. Comparative study of algorithms for optimized control of industrial energy supply systems. *Energy Inform.* **2020**, *3*, 12. [\[CrossRef\]](#)
50. Vázquez-Canteli, J.; Dey, S.; Henze, G.; Nagy, Z. CityLearn: Standardizing Research in Multi-Agent Reinforcement Learning for Demand Response and Urban Energy Management. *arXiv* **2020**, arXiv:2012.10504.
51. Zsembinski, G.; Fernández, C.; Vérez, D.; Cabeza, L.F. Deep Learning Optimal Control for a Complex Hybrid Energy Storage System. *Buildings* **2021**, *11*, 194. [\[CrossRef\]](#)
52. Diaz, J.; Fernández, J. Realistic Simulation Tool for practical Analysis of Solar Cooling Thermal Systems driven by Linear Fresnel Collectors. *Rev. Ing. UC* **2021**, *28*, 360–377. [\[CrossRef\]](#)
53. *Standard 90.1-2019; (SI Edition)—Energy Standard for Buildings Except Low-Rise Residential Buildings* (ANSI Approved; IES Co-Sponsored). American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE): Atlanta, GA, USA, 2019.
54. ANSI/ASHRAE. *Standard 62.1-2019; Ventilation for Acceptable Indoor Air Quality*. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE): Atlanta, GA, USA, 2019.
55. Kohlenbach, P. Solar Cooling with Absorption Chillers: Control Strategies and Transient Chiller Performance. Doctoral Thesis, Technical University of Berlin, Berlin, Germany, 13 January 2006.
56. Industrial Solar GmbH. FIFA World Cup Solar Cooled Demonstration Stadium. Available online: http://ship-plants.info/solar-thermal-plants/316-fifa-world-cup-solar-cooled-demonstration-stadium-qatar?unit_operation=12 (accessed on 1 July 2022).
57. Industrial Solar GmbH. Solar Cooling for Data Center. Available online: http://ship-plants.info/solar-thermal-plants/311-solar-cooling-for-data-center-south-africa?unit_operation=12 (accessed on 1 July 2022).
58. Kohlenbach, P.; Jakob, U. *Solar Cooling: The Earthscan Expert Guide to Solar Cooling Systems*; Routledge: London, UK, 2014; ISBN 9781317963981.
59. Chowdhury, B.; Islam, M.; Begum, F.; Parvez, A. Design and Performance Analysis of a Cooling Tower in Sulfuric Acid Plant. *J. Chem. Eng.* **2010**, *23*. [\[CrossRef\]](#)

-
60. ASHRAE. ASHRAE Climatic Design Conditions 2009/2013/2017. Available online: <http://ashrae-meteo.info/v2.0/> (accessed on 1 July 2022).
 61. *DesignBuilder, Version 5.5.0.012*; Design Builder Software Limited: Gloucester, UK, 2013.
 62. Blair, N.; Dobos, A.P.; Freeman, J.; Neises, T.; Wagner, M.; Ferguson, T.; Gilman, P.; Janzou, S. *System Advisor Model, Sam 2014.1. 14: General Description*; NREL Rep. No. TP-6A20-61019; National Renewable Energy Laboratory: Golden, CO, USA, 2014; Volume 13. [[CrossRef](#)]
 63. Klein, S.A.; Nellis, G. EES, Engineering Equation Solver. Available online: <http://www.fchart.com/ees/mastering-ees.php> (accessed on 1 July 2022).