ISES International Solar Energy Society

Conference Proceedings

EuroSun 2014

Aix-les-Bains (France), 16 - 19 September 2014

Performance of a coupled Reinforced Learning-Fuzzy Control approach to the Control of a Solar Domestic Hot Water system

Anton Soppelsa[1], Davide Bettoni[1], Roberto Fedrizzi[1]

[1] Institute for Renewable Energy, EURAC Research, Viale Druso 1, 39100, Bolzano-Bozen (Italy)

Abstract

The paper discusses the results of a comparative analysis of the performance of different control strategies applied to a reference solar DHW system. Three classes of control strategies have been considered: so-called naïve control strategies based on the on-off control of the solar collectors pump using temperatures difference, solar irradiation or both; fuzzy-based control strategies; and a reinforced learning-based strategy coupling a Q-learning algorithm to a fuzzy controller. The performance figures used in the analysis are the seasonal performance factor at the primary side of the circuit (SPF_{coll}), the seasonal performance factor of the whole DHW preparation process (SPF_{DHW}) and the number of times the circulation pump is switched on and off (N_{ON-OFF}). The analysis, carried out numerically, has been performed using the TRNSYS simulation software coupled to a LabVIEW implementation of the controllers. The analysis suggests that controllers able to find a nearly optimal policy without requiring prior modelling of the system can be implemented using a reinforced learning algorithm and supports the fact that well designed control strategies can increase significantly the performance of such systems.

Key Words: Reinforced Learning, Fuzzy Control, Solar Thermal, Solar Domestic Hot Water, TRNSYS, LabVIEW

1. Introduction

Despite its potential for growth, the market of solar thermal system in Europe is nowadays seeing a stagnation period. Industries in the field are trying to improve their products by either reducing the production cost of their components or selling advanced products granting better performance. Moreover, most of the small scale domestic solar system are still installed without any particular attention to their efficient control. Then it comes not as a surprise that much attention is paid to the opportunities offered by control optimization and to the tools that allows it. Currently, the optimization of the solar thermal system control is based on computationally intensive and time-consuming simulations carried out with specialized software tools. Following the recent application of soft computing techniques in the field of building automation (Dalamagkidis, 2007) we started a research line to investigate the applicability of these techniques to solar thermal system, with the long term aim of developing self-optimizing controllers able to increase their overall performance and reduce the simulation efforts spent for their development. In this paper we show the results of a simulations campaign carried out to compare the performance of a reference solar domestic hot water system (SDHW) controlled by naïve control strategies, a fuzzy logic based control strategy and a coupled reinforced learning-fuzzy logic based control strategy.

Fuzzy logic (FL) is a rule-based decision making method used for expert system and process control. FL is based on the fuzzy sets theory, a set theory where membership is a matter of degree, and deals with variables assuming linguistic values (such as “COLD”, “MILD”, “HOT”) instead of numbers. It has been successfully applied to several areas of science and technology, in particular to system control (Zadeh, 1965, 1968; Klir, 1995). The key elements of a Fuzzy Controller (FC) are a set of “if-then” rules (knowledge-base), an internal logic processor (inference mechanism), and two other components called fuzzifier and de-fuzzifier (among the

Selection and/or peer review under responsibility of EuroSun2014 Scientific Committee
doi:10.18086/eurosun.2014.03.26 Available at http://proceedings.ises.org

vast literature describing fuzzy control see, for example, Zilouchian and Jamshidi, 2001; Passino and Yurkovich, 1998). In our implementation of the FC we considered the so called Mamdami fuzzy inference system, characterized by a linguistic variable in the "then" clause of the rules base. The cooperation of fuzzifier, knowledge base, inference mechanism and defuzzifier results in a non-linear static map between the input and output of the controller. Fuzzy controllers show the advantage of being much more easily set up and tuned by people without a controls theory background than the classical controllers.

Reinforcement Learning (RL) is a machine learning (ML) methodology developed in the 1979 (Sutton and Barto, 1998) inspired by the research in the neurobiology field. It is based on the interaction between an agent and its surrounding environment (Fig. 1). The interesting discussion about what is concretely meant by "agent" would lead us too far from the scope of this paper and is avoided. Here it suffices to think at the agent as a component that can: a) choose what action to perform next on the environment; b) perform that action; c) sense the modifications caused by that action, including an evaluative feedback signal, called reinforcement. The reinforcement signal can be seen as a reward (positive) or a punishment (negative) received from the environment as a consequence of its actions. The learning process of the agent is influenced by the environment through the rewards. We note that, in general, this agent-environment configuration is identical to that of feedback-controlled systems where the dualism consist of components controller-process. The difference between the two paradigms is that in the former, the agent is expected to self-learn how to behave at best in the environment while in the latter, the controller is expected to drive the process as prescribed by the controller designer, a very different perspective. In the context of the RL theory, a policy is the set of rules followed by the agent to determine what action to perform at each time step. The objective of the whole learning process is to find the optimal policy to perform whatever task the agent is supposed to carry out. Optimality is defined with respect to the maximization of the cumulative reward.

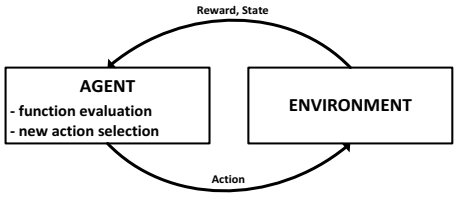

Fig. 1: Reinforcement learning scheme.

We applied a variant of a particular RL algorithm called Q-Learning (Watkins, 1989). This algorithm belongs to class of unsupervised, model-free RL methods. Here the agent does not have any prior knowledge or model of the system characteristics from which it could estimate the next possible state. Essentially, the agent does not know what are the effects on the environment of a certain action and chooses the next action on the basis of the cumulative effect of the actions performed in the past.

A sound and self-consistent recall the fuzzy control and Q-learning theories and the description of the FC and Q-learning method implementations, which in the Q-learning case required some adaptations with respect to the implementation found in the textbooks to give satisfying results, are topics outside the scope of this paper, which will focus on the simulation results obtained simulating a mathematical model of the reference system. Apart of this brief introduction, the paper is organised in 3 sections. In section 2 are described the reference system and the simulation set-up as long as the assessed control strategies and the performance figures used to evaluate them. In section number 3 the results of the simulations are shown and discussed. Finally, in section number 4, the conclusions of this work are drawn.

2. Reference system, performance figures and control strategies

The background material of this paper is divided further in three sub-sections. The first one is dedicated to the description of the solar system used as a test-bed for the new controllers. The second one describes what performance figures have been chosen to assess the quality of the control. The third one describes the controllers in the arena.

2.1 Reference system

The solar system considered for the assessment is a medium-size solar system for domestic hot water (DHW)

with 12.15 m^2 collector field, 1 m^3 storage and a 40 kW electrical backup. The system has been designed to match a DHW demand of about 50 l/day/person for 12 persons and its layout is presented in Fig. 2. The thermal storage consists of a container of water with two internal heat exchangers, one for the solar primary circuit and one for the domestic water circuit. A storage model including a stratification device has been selected, because this type of storages promote the stratification of the internal temperature with the result of increasing the heat exchange process between the stored water and the heat exchanger for DHW. The configuration with two heat exchangers is also commonly used to avoid the problem of legionella. In the primary circuit a mixture of water and propylene glycol (30% in volume) has been used as it is customary, in the central and northern regions of Europe, to make use of glycolised water in order to avoid freezing during winter seasons. A collectors area has been selected using standard design rules of thumbs of solar thermal systems for DHW. The collectors are faced to the south direction and have a slope of 30° on horizontal. The DHW request profile has been computed using DHWcalc (Jordan and Vajen, 2000, 2001) considering a four-family house with the afore mentioned consumption rate. The electrical backup systems has been included in order to satisfy the energy demand in case not enough solar energy is harvested and stored in the tank. The backup system allows to reach a DHW set temperature of 40°C.

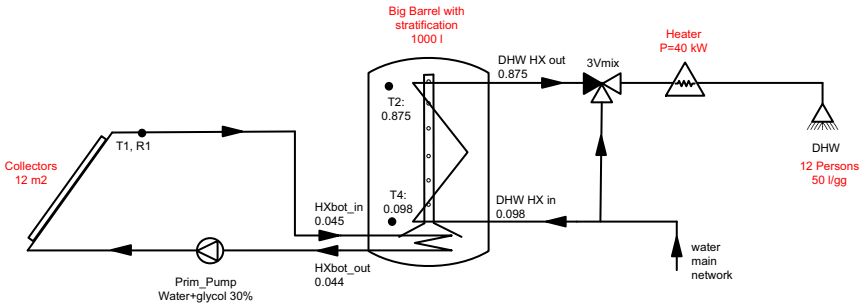

Fig. 2: Diagram of the SDHW system used as a test-bed for the control strategies assessment.

The TRNSYS program (Klein et al., 2006) has been chosen as simulation platform and the model of the system has been developed using its component library. In particular the model of the thermal storage (Drück, H., 2006) had already been validated previously using monitoring data from a SolarCombi+ installation (D'Antoni et al., 2011) and following the procedure reported in (D'Antoni M. et al., 2012). Finally, the source of the climatic data used to perform the simulations is the Meteonorm database. The choice of employing such a simple solar system has been made for two reasons: the wide application potential of such a system and the fact that working with a well-studied system, whose behaviour and optimal control strategies are well known, makes the assessment of the new controllers more intuitive.

In order to ease future implementations on industrial control hardware, the development of the advanced control strategies based on fuzzy logic and reinforcement learning has been made using the LabVIEW programming language. The controllers have been interfaced to TRNSYS as custom-made types. Interfacing the two programs is straightforward, although on some machines we experienced negative interactions between the LabVIEW Runtime Engine and the TRNSYS executable, preventing the communication between the two programs. The cause of this problem has been fully understood and a workaround has been developed for the machine set-ups showing the problem. The development of a more elegant solution requires tinkering with the TRNSYS or LabVIEW Runtime Engine source code, something that certainly falls well beyond the possibilities of the average user of the two programs.

2.2. Performance metrics

The assessment of the control strategies is performed considering three different performance figures calculated from a set of basic quantities readily available within the simulation environment. The basic instantaneous values include: the collectors heat extracted from the panels ($\dot{Q}_{coll}$), the solar global irradiance incident on the collectors plane (I_G), the electrical power used by the primary pump (W_{el_pump}), the electrical power consumed by the backup (W_{el_backup}) and the total DHW demand ($\dot{Q}_{DHW}$). The selected performance figures are the seasonal performance factor calculated at the primary side of the circuit (SPF_{coll}), the seasonal performance factor of the whole DHW preparation process (SPF_{DHW}) and the total number of on-off cycles performed by the circulation pump ($N_{ON\text{-}OFF}$). All these performance figures must be calculated over a reference period of time. In this paper we show results from the yearly- and monthly-based analyses. Additional figures,

like the collector efficiency (η_{coll}), the gross solar yield (GSY) and the thermal energy lost at the primary circuit (Q_{loss}) have been calculated to improve the understanding of the results of the comparison by giving to the reader the widest possible perspective on the problem. The exact definition of these quantities is given in the following.

The seasonal performance factor "SPF_{coll}" of the primary circuit is defined as the ratio between the thermal energy collected by the system over a reference period of time and the electrical energy consumed by the circulation pump of the solar circuit in the same period of time (eq. 1). As an index it is a measurement of the effectiveness of the solar system from the point of view of the energy generation.

$$SPF_{coll} = \frac{\int_{month,year} \dot{Q}_{coll} dt}{\int_{month,year} W_{el,pump} dt} \quad \text{(eq. 1)}$$

Aside to the seasonal performance factor of the collector, the seasonal performance factor of whole system "SPF_{DHW}" is computed. This is defined as the ratio between the total DHW energy demand and the total electrical energy employed for satisfy this demand over the same period of time (eq. 2). The total electrical demand is composed by the consumption of the auxiliary electrical heater and the electrical energy consumed by the circulation pump of solar circuit ($W_{el,DHW} = W_{el,pump} + W_{el,backup}$).

$$SPF_{DHW} = \frac{\int_{month,year} \dot{Q}_{DHW} dt}{\int_{month,year} W_{el,DHW} dt} \quad \text{(eq. 2)}$$

The collectors efficiency is the ratio between the yearly energy collected by the solar system and the energy that hit the collectors, estimated using the global irradiance on the collectors plane and the gross panels surface area (A_{coll}).

$$\eta_{coll} = \frac{\int_{month,year} \dot{Q}_{coll} dt}{\int_{month,year} I_G \cdot A_{coll} dt} \quad \text{(eq. 3)}$$

Gross solar Yield "GSY": gives the energy captured from solar field per unit of collectors area. This parameter represents how much energy the collectors are able to extract over the time period.

$$GSY = \frac{\int_{month,year} \dot{Q}_{coll} dt}{A_{coll}} \quad \text{(eq. 4)}$$

The global radiation on the collectors plane is simply given by:

$$GR30° = \int_{month,year} I_G dt \quad \text{(eq. 5)}$$

where I_G is the global irradiance on the collectors plane. Finally, Q_{loss} is the heat lost at the primary circuit when $\dot{Q}_{coll}$ is negative.

$$Q_{loss} = \frac{1}{2} \int_{month,year} \left|\dot{Q}_{coll}\right| - \dot{Q}_{coll} dt \quad \text{(eq. 6)}$$

This inconvenient situation may happen during the initial phases of the system start-up or in the evenings when the difference between the inlet and outlet temperature of the internal heat exchanger is negative.

2.3. Control strategies

2.3.1 Traditional control

As introduced in section 1, the traditional method to control SDHW systems is based on the temperature difference between collectors outlet and thermal storage. More evolved systems make use of the solar irradiance to switch on the pump of the primary circuit (primary pump), as reported in (Furbo and Shah, 1996). In this case the control signal of the pump can also be function of the radiation. The new control strategies assessed in this work, are compared against the four naïve control strategies listed below.

A) Control of the temperature difference between the collectors outlet and the thermal storage (TD) using an hysteresis between 2-7 °C (the pump, running at constant speed, is switched on and off accordingly to the hysteresis output);

B) Control of the primary pump using an hysteresis on the irradiance between 100 and 150 W/m^2 (the pump, running at constant speed, is switched on and off accordingly to the hysteresis output);

C) A combination of A and B where the above TD control is activated via the hysteresis on the solar irradiation;

D) A variant of the C strategy where a mass flow modulation proportional to the global irradiation (linear modulation with maximum at 600 W/m2) is applied whenever required by both the irradiance and TD hystereses.

Moreover, in order to make the simulation as realistic as possible, in all cases presented above the detection of storage overheating and collector stagnation was added to the model along with the controls managing them.

2.3.2 Fuzzy control

Following the idea of increasing the control performance of the system, as from the on/off cycles of the pump and as from the performance point of view, a further analysis on the control strategy "B" (only on the irradiation) has been made designing a single-input single-output (SISO) fuzzy logic controller (F). The global irradiation on the panels plane is the controller input, while the pump command is its output. The three membership functions shown in Fig. 3 have been used to capture the linguistic terms of "LOW", "MEDIUM" and "HIGH" irradiance. The output membership functions have been defined in the similar way, with three triangular membership functions, equally distributed on the interval between 0 and 1, representing the normalized pump command. The defuzzification step is performed using the centre-of-area method. Three rules are defined as knowledge base, relating each one of the three input membership functions to a corresponding output membership function (low radiation with low speed, medium radiation with medium speed and high radiation with high speed). Two different families of FL controllers have been implemented and tested. The first family is obtained by varying the value of the X_{max} parameter (corresponding to the maximum of the "HIGH" membership functions) as shown in Fig. 3. The second family is obtained by varying the "MEDIUM" membership functions. For the sake of brevity we will consider only the first case in the following. The trans-characteristics of the 7 members of the controllers family resulting from the variation of the "HIGH" membership function are shown in Fig. 3, where X=1 corresponds to an irradiance of 1200 W/m^2.

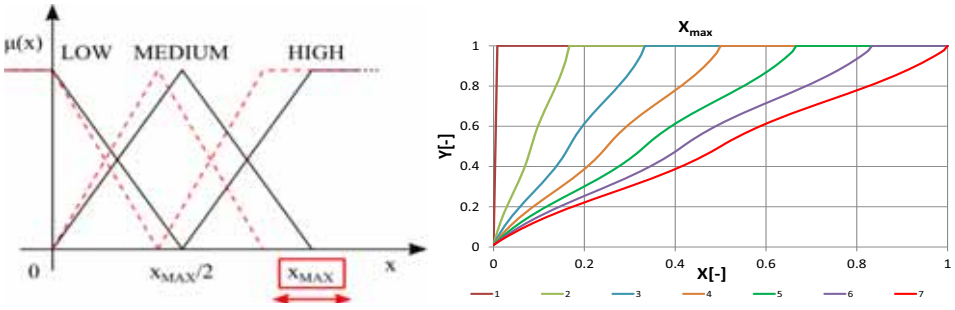

Fig. 3: Definition of the input membership functions of the FC.

2.3.1 Q-learning

The third and more advanced control strategy involves the coupling of the above fuzzy controller to the Q-learning algorithm (Q). Q-learning agents have the ability to learn an optimal policy without requiring a model of the plant they control, which is why we have considered them in the first place. In our case the environment consists of a discretisation of the continuous domains used to represent the solar irradiation and the storage temperature. The so-called reward function is computed as a linear combination of the collectors efficiency and the ratio between the heat supplied by the backup system over the integral of global irradiation on the

collectors plane. More precisely, the reward function used in the learning process is the following:

$$r = \frac{w_1 \cdot \int_{\Delta t} \dot{Q}_{coll} dt - w_2 \cdot \int_{\Delta t} W_{el,pump} dt}{\int_{\Delta t} I_G \cdot A_{coll} dt} \quad \text{(eq. 7)}$$

With this reward structure the controller is rewarded proportionally to the efficient use of the collectors and punished proportionally to the inefficient use of the primary pump. The choice of using this reward form has been made with the aim of learning energy efficiency using an as simple and intuitive method as possible to calculate it. With this regard, the instantaneous SPF_{coll} would be even more intuitive but early simulations showed that its use leads to worse performance (in terms of SPF_{coll}) than those obtainable with the above choice with the weights w_1 and w_2 reported in Tab. 1. The application of the Q-learning method requires to identify the environment states relevant to the application. This identification is critical because the designer has to find a compromise between two opposite needs: the one of describing the environment at best, calling for more variables, finer discretisation and therefore a high number of states, and the one of maintaining the problem tractable, calling for a reduced number of states. No reference was found in literature regarding the more convenient definition of states for thermal systems nor regarding how to perform optimal discretisation. We opted for considering two state variables: the global irradiance on the collectors plane and temperature of the water in the storage. The resulting state space has been discretised by subdividing the irradiance admissible interval (from 0 to 1000 $W/m^{2)}$ in 11 levels and the temperature admissible interval from 10 to 90 °C in 9 levels. Regarding the actions performed by the agent, as the Q-learning algorithm is applied on top of the FC, they consist in the selection of one of the FC shown in Fig. 3 for the next iteration. Our implementation of the Q-learning algorithm makes use of an iteration, or observation, period (Δt) longer than the control period, which was set to 1 minute. A subset of the parameters influencing the Q-learning algorithm are summarized in Tab. 1.

Tab. 1: A selection of the parameters describing the Q-learning algorithm.

Parameter	Units	Value
Number of states N_s	[-]	99
Number of actions N_a	[-]	7
Observation period Δt	[min]	5
w_1	[-]	1
w_2	[-]	370

3. Results

3.1. Naïve control strategies

The comparison of different control strategies has been made using the performance figures introduced in section 2.3. Starting from the, Tab. 1 shows the results of the yearly analysis of the four control strategies explained before while monthly data are shown in Fig. 4 (SPF_{coll}) and Fig. 5 (N_{ON-OFF}). The low value of the SPF_{coll} in case B is mainly due to the need of keeping the activation threshold rather low (to avoid stagnation) which result in higher energetic costs and also higher heat losses. The best performance of SPF_{coll} is related to the case D where the temperature difference control, activation threshold and pump speed modulation is adopted. From the system performance point of view, the best performance SPF_{DHW} are achieved when the losses are decreased (control on DT) with higher levels of temperature in the storage and less usage of electrical backup (case C). In this case, however, the number of on-off cycles is higher, according with the monthly profile reported in Fig. 5. An increase of the temperature hysteresis (from 2-7°C to 2-14°C) used in the controllers (A), (C) and (D) results in a decreased number of on/off cycles but reduces, at the same time, the performance. The number of on-off cycles, however, remains high, not less than 5200 for all the examined cases.

Tab. 2: Comparison between the four naïve control strategies - yearly data

Case	η_{coll}	SPF_{coll}	on/off pump	GSY	$GR_{30°}$	W_{el_pump}	DHW_{demand}	SPF_{DHW}	W_{el_Backup}	Q_{loss}
	[-]	[-]	[-]	[kWh/m^2]		[kWh]		[-]	[kWh]	
A	0.41	**301**	47426	602	1470	*24.3*	7660	**5.4**	*1400*	95
B	0.40	**159**	371	583		*44.6*		**4.7**	*1500*	215
C	0.40	**267**	11046	635		*28.9*		**5.5**	*1360*	56
D	0.41	**319**	7606	606		*23.1*		**5.3**	*1430*	52

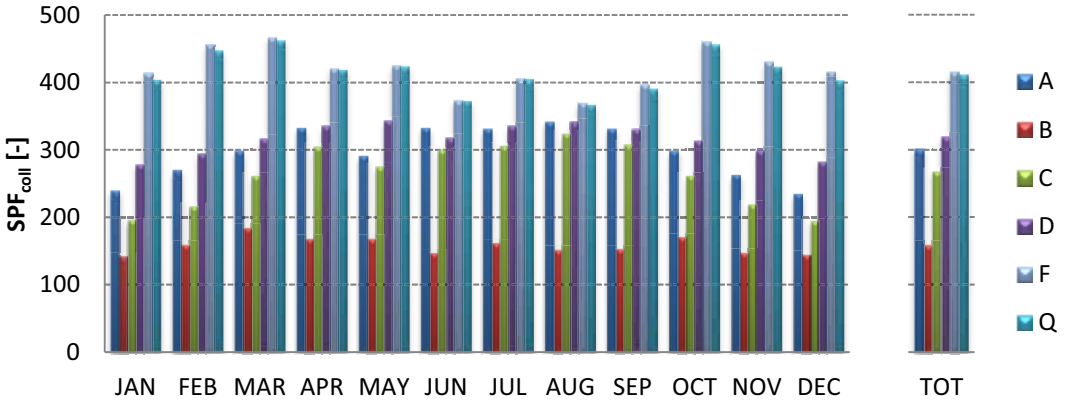

Fig. 4: Monthly SPF_{coll} profile of the analysed control strategies.

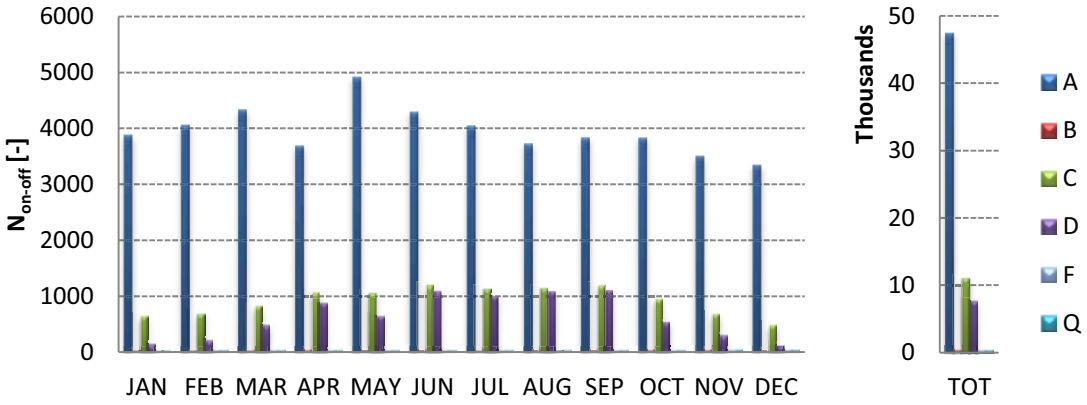

Fig. 5: Monthly on/off cycles profile of the analysed control strategies.

3.2. Fuzzy logic control

In Tab. 3, the yearly results of the family of FCs obtained by varying the meaning of "HIGH" irradiation are reported and compared. The first column shows the action names corresponding to the use of each FC while the second column recalls the irradiance corresponding to the maximum controller output. This analysis has been conducted in order to understand the effect of different fuzzy controllers and to have an idea of what to expect from the Q-learning controller.

Increased values of the collectors SPF (from 150 to 415) correspond to increased values of the maximum irradiance. The SPF of the whole system however shows much little variation and tops when the "HIGH" membership function has it maximum at 600 W/m^2. After this point, highest SPF_{coll} values are obtained at the expense of lower SPF_{DHW}. This happens because the pump is driven at a lower speed, the total amount of collected energy is lower and the need to resort to the backup is more frequent. A similar analysis has been

conducted by varying the "MEDIUM" membership function. Also in this case the data (not reported here for the sake of brevity) confirm this trend, although an even more extreme increase of the collector SPF, reaching a value of 691, is obtained when the "MEDIUM" membership function is shifted to higher values of irradiance. Also this performance peak in terms of SPF_{coll} correspond to the poorest performance from the point of view of SPF_{DHW}, dropping to 4.2.

Tab. 3: Comparison of performance using different meanings of "HIGH" radiation - yearly data

	"HIGH" irradiance	**η_{coll}**	**SPF_{coll}**	**on/off pump**	**GSY**	**$GR_{30°}$**	***W_{el_pump}***	**DHW_{demand}**	**SPF_{DHW}**	***W_{el_Backup}***	**Q_{loss}**
	[W/m2]	[-]	[-]	[-]	[kWh/m2]		[kWh]		[-]	[kWh]	
a1	0	0.39	**114**	367	574		*61.1*		**4.9**	*1500*	547
a2	200	0.40	**150**	399	581		*46.9*		**5.1**	*1470*	458
a3	400	0.40	**188**	414	584		*37.8*		**5.1**	*1450*	387
a4	600	0.40	**231**	406	585	1470	*30.8*	7660	**5.1**	*1460*	326
a5	800	0.40	**287**	420	584		*24.7*		**5.1**	*1480*	271
a6	1000	0.40	**352**	428	583		*20.1*		**5.0**	*1510*	243
a7	1200	0.40	**415**	413	580		*17.0*		**4.9**	*1540*	212

Again this situation happens because the primary pump is run at minimum speed most of the time, thermal energy is collected very efficiently but not enough energy is absorbed (GSY drops from about 580 to 558). Comparing with data reported in Tab. 2, the effect of the fuzzy controller is clear. A performance increase ranging from 30% to nearly 4-fold in terms of SPF_{coll} with respect of the naïve control cases are obtained. The fuzzy controller greatly reduces the primary pump on-off cycles to nearly 1 a day while increasing the collectors SPF. The other side of the medal is represented by the system SPF, which decreases about 7% with respect of what we consider the best overall naïve control, case D.

3.3 Q-learning method coupled to FC

The results of the simulation of the Q-learning algorithm spanning a period of 5 years are reported in Tab. 4. Over the time, the control algorithm learns what is the best FC in the family obtained by varying the meaning of "HIGH" irradiation. As a result of this process, the annual SPF_{coll} value increases, reaching a value similar to that obtained with the FC number 7. In fact, the same or even a better performance with respect of the best FC case reported in Tab. 3 was expected, because in principle the Q-learning method has the freedom to choose different FC in different states. Apparently, this is not the case (or if it is the method is not able to find it) and the reinforcement drives the learning process towards a uniform policy (the one that prescribes to use FC 7 in every occasion). Apart from its performance figures, one aspect that in the Authors' opinion nicely capture how the algorithm converges to the solution is the SPF_{coll} time evolution. In particular, the relative error of the monthly SPF at the beginning of the learning compared with the SPF at the end of the learning, which is reported in Fig. 7 for the first 15 months.

The first-learning curve shows that the algorithm converges relatively easy if the environmental conditions stays more or less the same (in February the error drops to 25% from nearly 40% of January). From the second half of winter and spring the system reaches continuously new states and the algorithm needs to explore them before exploitation can occur. By the summer time the controller is ready to take benefit from what it has learnt so far and the performance in these months is almost at top. Finally, at the end of the autumn when the winter time approaches, other unexplored conditions arise that did not happen at the beginning of the learning and the relative error of SPF_{coll} rises again, albeit reaching only about 13% of what recorded at the beginning of the learning. In order to increase the learning speed of the controller we introduced the possibility of embedding per-calculated values in the algorithm. The idea is to provide an easily computable estimation of the information the Q-learning algorithm would store internally at the end of the learning period and use it to provide controllers with such information already pre-programmed. This is straightforward and boils down to

initialising a matrix with the right numbers.

Tab. 4: Q-learning applied to the FLC on the high radiation and medium radiation levels

Case	year	η_{coll}	SPF_{coll}	on/off pump	GSY	$GR_{30°}$	W_{el_pump}	DHW_{demand}	SPF_{DHW}	W_{el_Backup}	Q_{loss}
		[-]	[-]	[-]	[kWh/m^2]		[kWh]		[-]	[kWh]	
"HIGH" irradiance	1st	0.40	**370**	446	582		*19.1*		**5.0**	*1520*	225
	2nd	0.40	**409**	425	581		*17.3*		**4.9**	*1540*	217
	3rd	0.40	**408**	422	581	1470	*17.3*	7660	**4.9**	*1540*	216
	4th	0.40	**410**	420	581		*17.2*		**4.9**	*1540*	216
	5th	0.40	**411**	422	581		*17.2*		**4.9**	*1540*	215

Fig. 7 shows what happens if the Q-learning algorithm is started with a fraction (β) of the knowledge accumulated over a five year period. Four different values of β are used in this example: 0, 0.2, 0.5, 0.8. This parameter represents the fraction of the final converged matrix that is used to initialize the algorithm, β = 0 corresponding to learning from scratch starting from a null matrix. As expected, the higher the β parameter the better the learning performance.

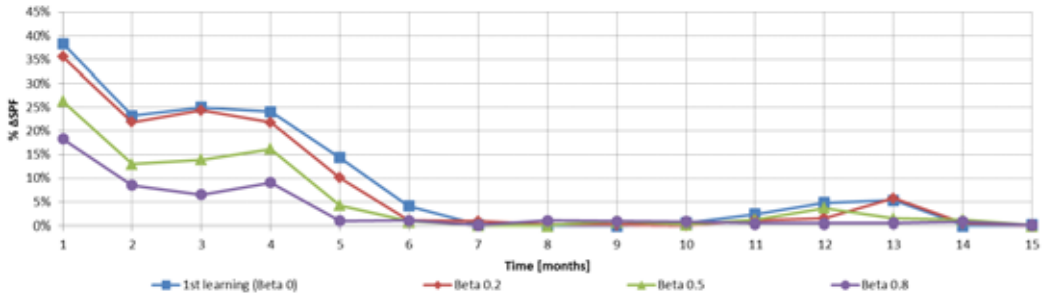

Fig. 7: Monthly SPF_{coll} relative error (convergence values after five years) comparing the first learning and the second learning phase with different β 0 (blue), 0.2 (red), 0.5 (green), 0.8 (purple).

Different analyses have been conducted changing internal parameters of the Q-learning algorithm. Many of them showed to have little influence on the overall performance of the control. Instead, using an higher number of states, reducing the influence of explorative actions or changing the observation period Δt has a strong impact on the learning speed, the time required by the algorithm to achieve the best results. As the algorithm need to perform a minimum of explorative actions per state in order to understand what are the best actions to perform, the bigger this minimum, the longer the observation time or the bigger the state space, the longer becomes the learning. The results presented in this paper have been produced with an observation period of 5 minutes, 5 times the simulation time step.

4. Conclusions

This analysis of the fuzzy and Q-learning base controllers applied to a simple SDHW system suggest that the use modern soft-computing techniques in the field of solar thermal system can bring important practical advantages. The first one is increased efficiency in the energy collection, in our case SPF_{coll} goes from 315 of case A to 370 of case Q after the first year and 409 after the second year. The second one is increased lifetime of the primary pump. Certainly, the higher the number on-off cycles underwent by an electrical device the higher the probability of breakage. With this regard, avoiding DT control greatly reduce N_{ON-OFF} from about 130 times a day to about 1.3 times a day and controllers (F) and (Q) obtain this result while increasing the performance on SPF_{coll} and paying a limited penalty on SPF_{DHW}. Finally, self-learning controllers are attractive because they can be easily modified to adapt to the plant "aging". When the aging modifies substantially the behaviour of the system this characteristic is valuable, although this was not the case of the SDHW considered

in this study.

The performance shown by the fuzzy controller and the Q-learning algorithm with regard to the SPF_{DHW} figure deserves a special comment. In both cases, the figure turns out to be about 11% less than with the best naïve controller (C). This happens because the chosen reward function does not punish the controller for the backup use, making the controller somewhat blind to that aspect. Moreover, in our case study the maximization of the collection efficiency does not imply the maximization of the overall system efficiency, because SPF_{DHW} is totally dominated by the total heat supplied by the solar system and scarcely influenced by the efficiency of how it is collected (the energy spent for running the primary pump is in the worst case less than 5% of the energy consumed by the back-up system).

However, a careful analysis of the power flows, reveals that most of the degradation is due to the fact that heat is lost during heat transfers when DT is negative. Now, the (B) and (F) controls can't help with this regard because they are totally independent from DT. The (Q) control cannot help either: although operations with negative DT are inefficient and punished, it is based on a family of controllers which cannot perform the right action to overcome the problem (avoid switching on the pump if DT is negative). This strongly suggests to include DT in the Q-learning by augmenting its state. These facts are not regarded as limitations of the result of this work, which aimed at assessing the applicability of the Q-learning method on solar thermal systems and showing its potential, not at developing an optimal control for a system for which this was already known.

5. References

Blank, N.T. and Ertel, W, 2011. Introduction to artificial intelligence, Springer

Dalamagkidis K. et al., 2007. Reinforcement learning for energy conservation and comfort in buildings, Building and Environment, vol. 42, no. 7, pp. 2686–2698.

D'Antoni M. et al., 2011. Parametric analysis of a novel Solar Combi+ configuration for commercialization. In Proc.: 4th International Air-Conditioning, Larnaka, Cyprus.

D'Antoni M. et al., 2012. Validation of the numerical model of a turnkey Solar Combi+ system. Solar Heating and Cooling Conference, San Francisco, CA

Drück, H., 2006. Trnsys Type 340. Multiport Store- Model, version 1.99F. ITW Stuttgart University, Germany.

Furbo S. and Shah L. J., 1996. Optimum Solar Collector Fluid Flow Rates, Proceedings of Eurosun 1996

Jordan U. and Vajen K., 2000. Influence of the DHW profile on the Fractional Energy Savings: A Case Study of a Solar Combi-System. Solar Energy Vol.69, pp. 197-208.

Jordan U. and Vajen K., 2001. Realistic Domestic Hot-Water Profiles in Different Time Scales. FB. Physik, FG. Solar, Universität Marburg, D-35032 Marburg.

Klir G. J. and Yuan, B., 1995. Fuzzy sets and fuzzy logic: theory and applications, Prentice Hall PTR

Klein S.A. et al., 2006. Trnsys 17 A transient simulation program. Solar Energy Laboratory, University of Wisconsin, Madison.

Passino K. M. and Yurkovich S., 1998. Fuzzy control. Menlo Park, Calif.: Addison-Wesley.

Sutton R. S. and Barto A. G., 1998. Reinforcement Learning: An introduction. MIT Press.

Watkins, C. J. C. H., 1989. Learning from delayed rewards, PhD Thesis, University of Cambridge, England

Zadeh L. A., 1965. Fuzzy sets, Information and control, vol. 8, pp. 338–353.

Zadeh L. A., 1968. Fuzzy algorithms, Information and control, vol. 12, pp. 94–102.

Zilouchian A. and Jamshidi M., 2001. Intelligent control systems using soft computing methodologies. Boca Raton FL: CRC Press.