

Control of a PCM ventilated facade using reinforcement learning techniques[☆]

Alvaro de Gracia^a, Cèsar Fernández^b, Albert Castell^c, Carles Mateu^b, Luisa F. Cabeza^{c,*}

^aCELiMIN, Universidad de Antofagasta, Antofagasta, Chile

^bComputer Science Department, University of Lleida, Lleida, Spain

^cGREA Innovació Concurrent, Universitat de Lleida, Lleida, Spain

Abstract

Artificial intelligence techniques have been successfully applied to control dynamic systems looking for an optimal control. Among those techniques, reinforcement learning has been shown as particularly effective at reducing the dimensionality of some real problems and solving control problems by learning from experience. The use of thermal energy storage active systems in the building sector is identified as suitable option to reduce their energy demand for heating and cooling. However, these systems might be expensive and require appropriate control strategies in order to improve the performance of the building. In this paper a ventilated facade with PCM is controlled using a reinforcement learning algorithm. The ventilated facade uses mechanical ventilation during nighttime to solidify the PCM and releases this cold stored to the inner environment during the peak demand period. It is crucial to decide correctly the schedule of charge and discharge process of the PCM according to the weather and indoor conditions. An experimentally validated numerical model is used to test the performance of the control algorithm under different weather conditions. Important improvements on the energy savings due to the use of control strategies were found and supported by the data under the different tested climatic conditions.

Keywords: Artificial intelligence, Control system, Phase change materials (PCM), Thermal energy storage (TES), Numerical modeling, Computational Sustainability

[☆]The work partially funded by the Spanish government (ENE2011-28269-C03-01, ENE2011-22722 and ULLE10-4E-1305). The authors would like to thank the Catalan Government for the quality accreditation given to their research group (2014-SGR-123). The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement # PIRSES-GA-2013-610692 (INNOSTORAGE). Alvaro de Gracia would like to thank Education Ministry of Chile for Grant PMI ANT1201.

*Corresponding author. Address: Edifici CREA, Pere de Cabrera s/n, 25001, Lleida, Spain. Tel: +34.973.00.35.77.

Email address: lcabeza@diei.udl.cat (Luisa F. Cabeza)

1. Introduction

In last decades the energy consumption of buildings has been increasing as a result of increasing comfort requirements. The contribution of buildings in the global energy consumption is very significant and thus, a lot of attention and efforts are focused on decreasing building energy requirements [8]. For such purpose, the European Directive 2010/31/EU [5] states ambitious objectives for 2020, such as reducing by 20% the global energy consumption and the greenhouse gas emissions. The use of renewable energies presents a huge potential since they contribute to reduce the greenhouse gas emissions while also reducing the pressure exerted to limited natural resources, such as fossil fuels, and the energy dependence on third countries.

However, several renewable energies present intermittent availability, which results in the necessity of conventional back-up systems, low efficiencies of renewable systems, and loss of potential energy savings. The use of thermal energy storage can solve the mismatch between supply and demand of renewable energies, helping to overcome such problems [2]. One of the most promising and widely studied thermal energy storage technologies is latent heat. It is based on the phase change (usually solid-liquid) of a material in order to store energy at an almost constant temperature and with high energy densities. In the design of such thermal energy storage systems (TES) it is critical to select the correct phase change material (PCM) based on its melting temperature and enthalpy. The phase change temperature range must be within the available energy source temperature and the required temperature level. This is of great importance when both temperature levels are close, since the system will be extremely sensitive to small changes.

This sensitivity has been demonstrated in a previous publication [4], where a ventilated double skin facade (VDSF) with PCM in its air chamber was experimentally tested with the aim to provide cooling during summer by taking advantage of night free cooling strategies and thermal storage. The PCM was integrated into the building structure and mainly used as an active system to store cold at night and to provide cooling during peak hours. However, the system was very sensitive to external ambient conditions and it was not possible to achieve net electrical energy savings due to excessive use of fans to charge and discharge the storage. Therefore, a suitable control system was required.

Although control systems are common in many building installations, the studied systems do not operate based only in real time inputs (such as weather conditions and internal temperature) but also based on forecasts (especially weather forecasts). In order to optimize the use of fans, the storage must be charged during the time period of lower external temperatures and the cold must be released during the peak demand period. Both periods depend on the weather forecast which will define both the available cold and the cooling demand. Therefore, these advanced control systems need to be able to re-adapt the operational schedule depending on different predicted boundary conditions.

In this paper, reinforcement learning techniques are used to control an innovative ventilated facade with PCM in its air chamber. Reinforcement learning has been applied in real world applications in a broad range of problems, from resource allocation to scheduling. As an example, [6, 7] show how approximate dynamic programming

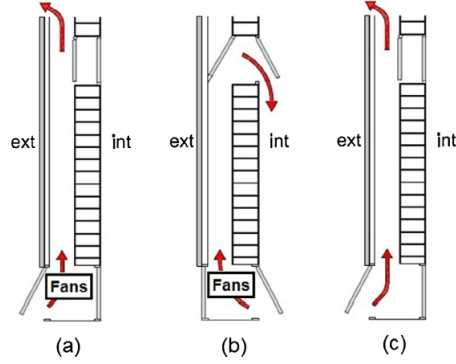


Figure 1: Modes of operation of the VDSF. (a) Charge process, (b) Discharge process, (c) Overheating prevention

techniques may help complex automated vacuum waste collection systems to derive more efficient operation policies, reducing their operation cost and energy requirements. In this work, the efficiency improvement when algorithms based on reinforcement learning ideas are applied to optimal control is evaluated by means of numerical simulation.

2. Methodology

2.1. Prototype and operating principle description

The ventilated facade studied has been implemented in an experimental house-like cubicle in order to measure its thermal behavior. The cubicle is built with alveolar brick walls, and the ventilated facade with PCM is located in the south wall. The ventilated facade is composed of an external layer of glass covered with insulating panels that are used as the outer opaque skin to avoid solar radiation during summer. These panels are held by an insulated metallic structure. The air cavity is 15 cm thick and 112 PCM panels (RT21 from Rubitherm) divide it into 14 airflow channels.

Moreover, the system is equipped with six gates distributed at inlet (bottom) and outlet (top) of the cavity and three fans located at the bottom, so different airflow paths can be programmed, as well as mechanical or natural ventilation conditions (Fig.1). Air can flow from indoors/outdoors to indoors/outdoors. Detailed description of the system can be found in de Gracia et al.[4].

The operating principle of this innovative ventilated facade during the cooling period is based on using the low temperatures at night (below the congealing temperature of the PCM) to solidify completely the PCM, and store cold to provide a cooling supply during the daytime when required by the demand.

During the charge process (Fig.1(a)) air is pumped from outdoors using the fans in order to intensify the convective heat exchange. The three fans require 55 W in total at minimum speed; hence it is critical to reduce as much as possible their use while ensuring complete solidification of the PCM. The VDSF is used as a cold storage

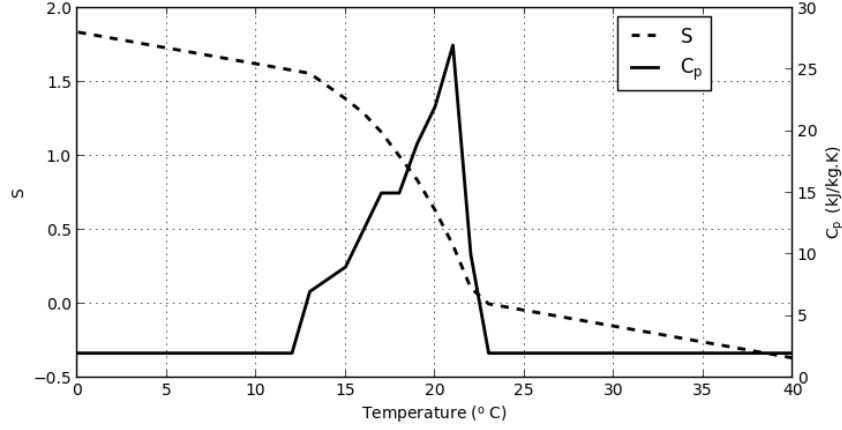


Figure 2: S evolution vs Temperature

system before a cooling supply to the inner environment is needed. Then, instead of using conventional HVAC systems which implies the electrical energy consumption, the discharge process starts (Fig.1(b)). Once the PCM has been melted the facade is opened outdoors and natural convection avoids overheating inside the VDSF (Fig.1(c)).

2.2. Numerical model

A simple model based on the assumption that all the PCM included in the facade is at the same temperature (isothermal model from here on) is used to simulate the thermal performance of the VDSF with PCM under different weather conditions and control strategies. The isothermal model aims to feed the reinforcement learning model in order to determine the best control strategies. The huge amount of cases and scenarios that should be evaluated, in order to achieve an optimal control strategy under the different weather conditions, makes crucial the use of a simple numerical tool without any iterative process. In this way, it is possible to reduce dramatically the required time and computational resources and hence deal with this huge number of case studies. Standard numerical methods based on iterative processes, such as finite volume or finite element methods could not be applied to create the reinforcement learning model.

In the simple isothermal model, the phase change is modeled as a variation of the heat capacity depending on the average temperature of the PCM. A parameter called S is defined to determine how far the material is to full solidification. The value of the parameter S depends directly on the average temperature of the PCM. It has a value of 0 at the beginning of the solidification process (23°C) and a value of 1 for fully solidified PCM at 18°C. Moreover, as negatives values of S indicate liquid state, the lower is the negative value, the further is the PCM from the melting region. On the other hand, values below 0 indicate that the PCM is in liquid region, the higher is this number, the further from the phase change range. According to the data provided by the manufacturer [10], a latent heat of 93 kJ/kg was considered within this temperature

Parameter	RT21	Units
k	0.2	$\frac{W}{m \cdot K}$
C_p	2	$\frac{kJ}{kg \cdot K}$
L	93	$\frac{kJ}{kg}$
ρ_l	880	$\frac{kg}{m^3}$
ρ_s	770	$\frac{kg}{m^3}$
ΔT_{PCM}	18-23	$^{\circ}C$

Table 1: Thermo-physical properties of the used materials

range of phase change (from 18°C to 23°C). Table 1 provides the used thermo-physical properties, where k is the thermal conductivity, C_p is the heat capacity, L is the latent heat, ρ_l and ρ_s is the density in liquid and solid state respectively and ΔT_{PC} is the melting range. Out of this temperature range, the S value is calculated according to the ratio between the sensible heat to/from the phase change boundaries and the latent heat of the phase change range (Fig.2). Moreover, the charge and discharge processes are assumed to behave as an internal flow through a constant temperature surface. Hence, the temperature at the outlet of the channel at an instant ' t ' can be calculated from the Eq.1:

$$T_t^{outlet} = T_t^{PCM} - (T_t^{PCM} - T_t^{inlet}) \cdot e^{\frac{-h \cdot A_{conv}}{m \cdot C_p}}. \quad (1)$$

Once the outlet temperature is calculated, the power of charge or discharge of the PCM can be obtained by an energy balance between the inlet and the outlet of the air flow at the ' t ' instant. In order to avoid any iterative method that will result in a high computational cost, the isothermal model makes use of an explicit scheme to discretize the time. Hence, the amount of energy stored or released by the PCM can be obtained by integrating the power of charge or discharge through the time step. Once the energy that the PCM has exchanged with the air is determined, the state of the PCM at the instant ' $t + 1$ ' can be calculated as expressed a follows:

$$S_{t+1} = S_t + \frac{Q_{t+1}}{m_{PCM} \cdot L} \quad (2)$$

being Q_{t+1} the amount of released/absorbed energy from t to $t + 1$.

The calculation of the state parameter allows determining the temperature of the PCM at the new instant of calculation ' $t + 1$ ' by interpolation in its definition. With the temperature of PCM at the new instant, Eq.1 can be applied to determine the outlet temperature at the new instant, and so on. When the system is in storage mode all the gates are closed and no air circulation exists. The PCM is not charged neither discharged by the air flux but it gains or losses energy depending on the outer climatic conditions. In this period, heat losses or gains to the outer environment are taken into account.

The isothermal model has been validated experimentally; more details about the model and the assumptions made can be found in de Gracia et al.[3].

2.3. Reinforcement Learning

Design controllers to minimize a given measure on dynamic systems over time is traditionally known as optimal control. Dynamic programming [1] solves optimal control problems through the well known Bellman equations. Even though dynamic programming is the only feasible way to solve optimal control from a general perspective, it suffers from what so called *dimensionality curse*, that is, an exponential growth of computational requirements with the number of variables. To overcome such a problem, many methods have been devised, by looking for smart searching strategies that avoid the complete search of dynamic programming algorithms. An example of such methods is reinforcement learning [12], being an effective technique to solve control problems by searching the correct answer by successive approximations.

Typically, in a reinforcement learning paradigm, the system to be modeled is represented by its state at a given time, $s_t \in \mathcal{S}$, being \mathcal{S} the set of states. Time $t \in T$ is discrete, being T the set of time slots when actions are taken, and usually, time slots are equally spaced (T_s). At each state, one can take an action from a possible set of actions, $a \in \mathcal{A}_{s_t}$. As a consequence of action a on a state s_t , the system produces a reward, $r_t(s_t, a) \in \mathbb{R}$, and leaves the system at a new state s_{t+1} .

In this particular case, the system state at time t is determined only by the PCM temperature (T_t^{PCM}). As $T_t^{PCM} \in \mathbb{R}$, later it will be shown how to avoid an infinite number of states by using gradient descent methods. The set of actions \mathcal{A}_{s_t} consists in non activate the ventilators or activate them at a set of determined speeds. The reward $R_t(s_t, a)$ is computed as the amount of thermal energy obtained from the PCM minus the electrical energy consumed by the ventilators, as follows:

$$r_t(s_t, a) = \left(\frac{\dot{Q}_t}{f} - E_p(a) \right) \cdot T_s, \quad (3)$$

where \dot{Q}_t is the cooling power supplied at t , f is a thermal-electrical equivalence factor (usually 3), $E_p(a)$ is the electrical power consumption of the ventilators operating at the speed determined by action a and T_s is the time slot duration. Finally, state s_{t+1} is determined by from Eq.1 and Eq. 2.

Under this considerations, the model is deterministic, in other words, a perfect foresight is always expected. Effectively, given a certain state, the future state is defined by the action over this state and the external temperature conditions. There is no uncertainty, and one can employ the classical Bellman formulation to solve it. For non-stochastic problems and discrete time, it may be expressed as:

$$\max \left\{ \sum_{t \in T} \gamma^t r_t(s_t, a) \right\}, \quad (4)$$

where γ is a discount factor. Eq. 4 clearly indicates the goal of maximizing the system rewards along a period of time by finding the optimal actions. Bellman optimality

equations can be derived from Eq. 4 describing the system in terms of the state values ($v_t(s_t)$) as:

$$v_t(s_t) = \max_{a \in A_{s_t}} \{r_t(s_t, a) + \gamma v_t(s_{t+1})\}, \quad (5)$$

allowing an iterative resolution of Eq. 4. Frequently, solving a set of system of equations as in Eq. 5 may not be suitable for large problems because the number of states tends to grow exponentially with the number of the variables. Monte Carlo methods (MC) overcome this problem by sampling only those states that are involved in the *experience*. Effectively, Monte Carlo methods learn from simulated experiences by averaging sample returns. During an episode – an episode is a particular simulation that spans over all the time T , in this case, a day – MC updates state values at each time slot. This simulation is repeated with as many episodes as needed until convergence.

Even though those methods are not complete, their efficiency in terms of speed and guaranteed convergence, as well as the quality of the obtained solutions, make them powerful tools to solve large problems. Usually the obtained solutions are quasi-optimal, but close enough to the optimal solution, being useful for certain application requirements and easy to obtain in terms of computation time.

For a fixed policy, i.e. the set of actions depending on the state are previously determined, MC methods update system values visiting them through a set of iterations until the system values converge to stationary values, according to a rule as the following:

$$v_t(s_t) \leftarrow v_t(s_t) + \alpha (g_t - v_t(s_t)), \quad (6)$$

being $g_t = \sum_k \gamma^k r_{t+k+1}$ the expected total reward from state s_{t+1} . At this point, a new family of learning algorithms emerge, called *temporal difference learning* (TD). The simplest TD method, known as $TD(0)$, updates $v_t(s_t)$ values at each point of the iteration, without waiting the end of an episode, according to:

$$v_t(s_t) \leftarrow v_t(s_t) + \alpha (r_{t+1} + \gamma v_{t+1}(s_{t+1}) - v_t(s_t)). \quad (7)$$

$TD(0)$ methods can be used to control, i.e. determining the best policy, expressing Eq. 7 in terms of state-action pair values rather than state values:

$$q_t(s_t, a_t) \leftarrow q_t(s_t, a_t) + \alpha (r_{t+1} + \gamma q_{t+1}(s_{t+1}, a_{t+1}) - q_t(s_t, a_t)). \quad (8)$$

As this updating rule employs the sequence $(s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1})$, it is called *Sarsa*. Sarsa algorithms are the base for the learning algorithm employed in this study.

The gap between updating at each iteration point as in Eq. 8 or wait at the episode's end as in Eq. 6 is filled $TD(\lambda)$ algorithms. Those algorithms not only compute the reward at different iteration points, but also weight the iteration points by a factor $0 \leq \lambda \leq 1$. Consequently, the expected gain in a $TD(\lambda)$ algorithm, known as *lambda*-return, is computed as:

$$g_t = (1 - \lambda) \sum_{n=1} \lambda^{n-1} g_t^{(n)},$$

being $g_t^{(n)}$ the n -step predictions:

$$g_t^{(n)} = r_{t+1} + \gamma r_{t+2} + \dots + \gamma^{n-1} r_{t+n} + v_{t+n}(s_{t+n}).$$

Using this return, $\lambda = 1$ defines a Monte Carlo method as in Eq. 6 meanwhile $\lambda = 0$ reduces it to a $TD(0)$. λ -return methods can reduce the updating error but add some difficulties to its implementation. In this sense comes to help the backward view of $TD(\lambda)$ methods. Instead of making updates ahead of the time sequence, updates are redefined in the backward direction, obtaining incremental and simpler implementations from a computational point of view. This backward view methods introduce a new variable, $Z_t(s)$, for each state that indicates the *eligibility* of this state in the update. At each state (s_t), this eligibility variable is updated as:

$$Z_t(s) = \begin{cases} \gamma\lambda Z_{t-1}(s) & s \neq s_t; \\ \gamma\lambda Z_{t-1}(s) + 1 & s = s_t. \end{cases}$$

Therefore, updates result:

$$v_t(s) \leftarrow v_t(s) + \alpha\delta_t Z_t(s), \quad \forall s \in \mathcal{S},$$

where $\delta_t = r_{t+1} + \gamma v_{t+1}(s_{t+1}) - v_t(s_t)$. See [11, 13] for further details. The *Sarsa*(λ) algorithm implemented in this study is detailed in Alg. 1. Finally, only remains to

Algorithm 1: *Sarsa*(λ) algorithm from [12].

```

Set  $q_t(s, a) = 0$  for all  $t \in T, s \in \mathcal{S}, a \in \mathcal{A}_s$ 
for each episode do
     $Z_t(s, a) = 0$  for all  $t \in T, s \in \mathcal{S}, a \in \mathcal{A}_s$ 
    for each  $t \in T$  do
        Initialize  $s$  and  $a$ 
        For action  $a$ , observe its return  $r$  and next state  $s'$ 
        Select next action  $a'$  from  $s'$  according a  $\epsilon$ -greedy1 policy of  $q_{t+1}(s', a')$ 
         $\delta \leftarrow r + \gamma q_{t+1}(s', a') - q_t(s, a)$ 
         $Z_t(s, a) \leftarrow Z_t(s, a) + 1$ 
        for all  $s \in \mathcal{S}, a \in \mathcal{A}_s$  do
             $q_t(s, a) \leftarrow q_t(s, a) + \alpha\delta Z_t(s, a)$ 
             $Z_t(s, a) \leftarrow \gamma\lambda Z_t(s, a)$ 
         $s \leftarrow s'$ 
         $a \leftarrow a'$ 

```

explain how one dealt with the continuous nature of the state space. Linear gradient-descent functions allow to approximate state values, as well as state-action pair values, as a linear combination of the system features. Assuming that the system has n features ($x_i(s)$) that describe its state s , such as temperatures, PCM mass , . . . , one can approximate its state value as a linear combination with a given weight w_i for each feature, such as $v(s) = \sum_{i=1}^n w_i x_i(s) = \mathbf{w}^\top \mathbf{x}$. The features may be constructed in different ways. This approach used binary features and tile coding. It means that a

¹ ϵ -greedy policy choose the best action with probability $(1 - \epsilon)$ and a random action with probability ϵ , avoiding this way getting stuck in local minimum.

feature range of values is partitioned in some tiles, splitting each tile the continuous space into small binary domains. In the experiments, 4 randomly spaced tiles are chosen, being each tile split into 4 domains. As only one parameter (PCM temperature) defines the system state, taking slot times of 10 minutes gives a state space dimension of $4 \times 4 \times 144 = 2,304$ possible states.

2.4. Experimentation benchmark

The energy performance of the system, as well as the required control strategy, depends strongly on the weather conditions. The Köppen-Geiger Climate Classification [9] was used to select representative locations according to the weather. This classification divides the main climate in A: equatorial, B: arid, C: warm temperate, D: snow and E: polar. Moreover, it defines the level of precipitation in W: desert, S: steppe, f: fully humid, s: summer dry, w: winter dry, m: monsoonal. Finally, it provides details about temperature as h: hot arid, k: cold arid, a: hot summer, b: warm summer, c: cool summer, d: extremely continental, F: polar frost, T: polar tundra. The combination of the previous definitions gives overall information about the weather conditions at each location. All the possible climates were analyzed in this study, except those that do not need any cooling supply during the whole year (E: polar and mostly all D: snow main climates).

For each location, outdoor temperature profiles were obtained during all the summer days, i.e. from June 21st to September 22nd for north hemisphere locations and from December 21st to March 20th for south hemisphere. Since temperature data set is sampled hourly [14], it was interpolated linearly at the sample rate required by the algorithm. Therefore, an interval time between decisions of 10 minutes was taken. Such an interval is a good trade-off between accuracy and state space dimensionality.

As mentioned in subsection 2.3, the optimization goal is to maximize the net electrical energy savings due to the use of the VDSF with PCM. Hence, not only the energy benefits from the cooling discharge have to be taken into account, but the electrical energy consumed by the fans both during charge and discharge process, as well. In order to compare the cooling energy to the electrical energy consumed by the fans, a value of 3 was used for the thermal/electrical energy ratio (f in Eq. 3), simulating a standard air conditioning equipment with COP=3.

Therefore, it is important that the control system could select the most appropriate timing for the charge process according to the weather prediction. In this study, the charge period might span from 0:00 to 8:00. Moreover, discharge periods may start at different times. In this study the discharge process was programmed to start at 10:00, 12:00 and 14:00. Between both periods, fans remain stopped.

Furthermore, the electric fans can operate at different speeds, from 0 (stopped) to 5 (higher speed), with important differences in terms of volumetric flow rate as well as electrical power consumption. Our experimentation has not shown any evidence of better performance, in energy consumption terms, when using more than fan speed 1. More explicitly, best results are obtained when using only speeds 0 and 1. Sensible higher electrical power requirements for fan speed above 1 explain this fact. Being so, we only show results for a fan speed policy of stopped (speed 0) or activated (speed 1).

Moreover, the adequate amount of PCM used in the VDSF might depend also on the weather conditions, the performance of the system and the potential energy benefits

that the system can provide. In this paper it is also explored how a possible increase in the PCM mass could affect the net energy savings achieved by the system in all the studied climatic conditions. The reinforcement learning model was used to determine the net energy savings if using 224 (x2) and 336 (x3) panels of PCM instead of the current 112 panels.

Finally, as the objective of this paper is to establish an efficient schedule of the fans activation based on temperature forecasts, it was checked how weather forecast accuracy impacts in the studied system performance. In this sense, temperature predictions and real measurements were captured from a couple of meteorological stations so predicted and real performances could be compared. The two selected meteorological stations correspond to Arenys de Mar (41° 34' 41" N, 2° 33' 18" E) and Lleida (41° 36' 4" N, 0° 37' 21" E, which were identified by Köppen-Geiger classification as Csa and Cfb, respectively.

3. Results

3.1. Energy performance around the world

As example of the obtained data from the reinforcement learning model, a typical behavior of temperatures and energy for a given day is depicted in Fig. 3. The upper plot shows the evolution of the PCM and outlet temperatures depending on external temperatures according to the isothermal model, once a given policy is determined for fans activity. As can be observed, the automatic scheduler decides to activate the charge phase mostly when the external temperature achieves its minimum. This can be considered an optimal behavior, even though small activation periods are produced at early hours. This occurs because the learning strategy does not complete all the possible search space, returning a near optimal solution, not really far from optimal. The same effect explains the small deactivation period during the discharging phase, which starts as soon as allowed at 10:00 AM. Moreover, the lower plot shows the energy balance of our system, i.e. thermal energy delivered (normalized to electrical energy) minus electrical energy consumed by the fans. In this example, the final balance is positive, achieving a net electrical gain of 1.79 MJ in a day.

As it was previously stated in subsection 2.4, the optimal operational schedule was calculated for all the cities involved in the study and for all the summer days. The results of the obtained summer average net electrical energy savings for each city are shown in the color map Fig.4. The results demonstrate that the system is suitable to operate under continental and snow climate regions while it is difficult to achieve important net savings in arid and equatorial areas. This coincides with prior numerical studies using SP-22 instead of RT-21 [14]. The results also confirm that the use of the PCM without hysteresis (such as RT-21) increases the number of possible suitable cities in which the system can provide net savings. Furthermore, there is an increment in the net energy savings in comparison to this prior study. This increment is justified because of the better performance of RT-21 and because of the use of the control system based on the reinforcement learning model to regulate the use of fans, especially during the charge process.

Moreover, Table 2 shows the summer average net electrical energy savings achieved and its standard deviation when the system is controlled by the reinforcement learning

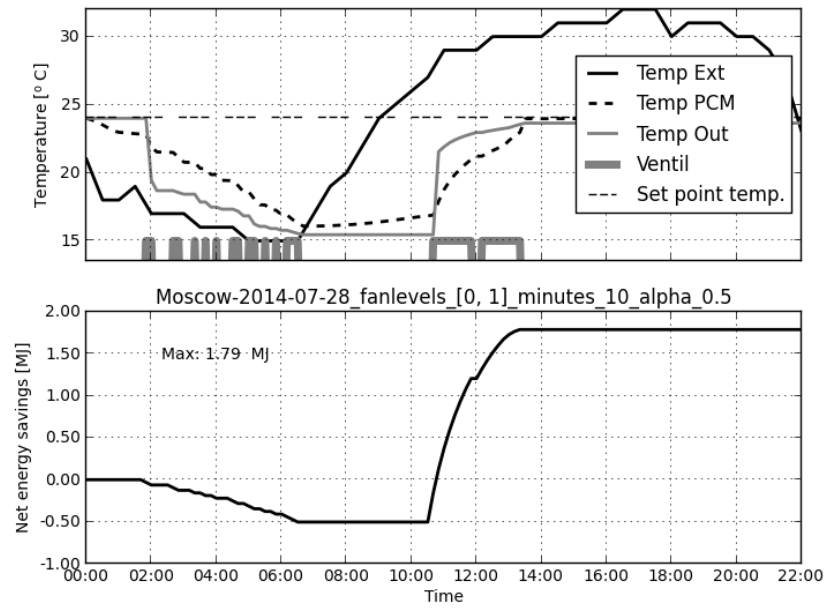


Figure 3: Evolution of temperatures and net electrical energy savings for Moscow 28/07/2014

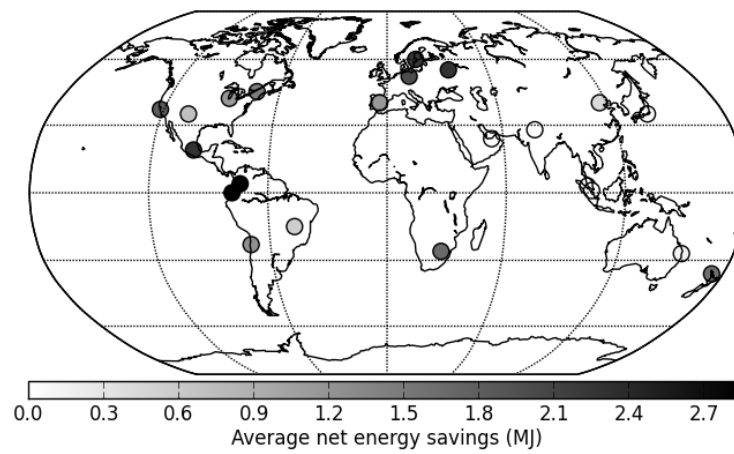


Figure 4: Summer average net electrical energy savings in each studied city

model (automatic) and when is controlled manually, programming the charge process from 06:00 to 08:00 and from 04:00 to 08:00. It can be seen, that the improvement is robust and constant in all the cities, increasing the average savings in the cities in which the climate makes the use of the VDSF suitable and avoiding negative net savings in cities in which is not. Within this context, Fig. 5 shows the comparison of the net electrical energy savings achieved when using control based on reinforcement learning (automatic) versus manually scheduled charge period (from 06:00 to 08:00). This figure is useful to demonstrate that the improvement is constant through all the summer period and to highlight how the automatic control system is beneficial to manage different control requirements. On one side, Montreal shows a climate in which the system has to operate almost during all summer but it has short periods with no possible energy benefits in which the control has to stop the fans and wait for better weather conditions; here the automatic control maximizes the benefits. In addition, Chicago shows a similar pattern during the first two months of summer (short periods of non-benefits between periods of 2 MJ/day of net savings), but has a different profile during the last summer month, in which the fans have to be stopped during the first two weeks, and active during the last two weeks, achieving net savings of 3 MJ/day. The automatic control demonstrated its capability to adapt to these control requirements. Moreover, Quito shows a climate in which the VDSF can operate during all summer, while in Singapore it cannot provide any benefit. The use of automatic control increases the benefits during all the period in the case of Quito, while it prevents waste of energy in the case of Singapore.

3.2. Discharge starting times and PCM mass

As it was previously described, once the night charge process has finished, the system starts the storage period, in which it is subjected to heat gains from the outer environment. The start of the discharge process is driven by the cooling demand of the building and might influence strongly the benefits extracted from the VDSF, especially if the storage media (PCM) is not well insulated from the outdoor environment. In order to evaluate if the discharge schedule is critical for the performance of the system, Fig. 6 compares the average net electrical savings achieved during summer in case that discharge is needed at 10:00, at 12:00 or at 14:00. It can be seen that even though the performance is reduced if the storage period is increased, this reduction is very small (around 0.2 MJ/day between discharging at 10 a.m. or at 14 a.m.), which demonstrates that if the outer envelope is well insulated the system can store effectively the PCM solidified until cooling is demanded.

In addition, Fig.7 presents the ratio between the net electrical energy savings provided by the system when using double (x2) or triple (x3) amount of PCM by the net savings using the standard amount of PCM. Only locations with significant net energy savings provided when using standard amount of PCM are considered for the comparison (net savings above 0.5 MJ/day). It can be seen that increasing the amount of PCM increases the net energy savings from the system in all cases by a factor higher than the factor in which the PCM has been increased (x2 or x3). This is very significant in cities in which the performance of the system is weak, such as Albuquerque or Brasilia.

Table 2: Summer average net electrical energy savings and standard deviation for automatic and manual scheduling

City	Energy / Std. Dev. [MJ]		
	Automatic	Manual	
		Charge period	
		6:00-8:00	4:00-8:00
Abu Dhabi	0.00 / 0.00	-0.35 / 0.01	-0.73 / 0.01
Albuquerque	0.71 / 0.60	0.46 / 0.64	0.36 / 0.74
Beijing	0.43 / 0.68	-0.01 / 0.60	-0.26 / 0.71
Berlin	2.00 / 0.77	1.76 / 0.87	1.49 / 0.77
Bogota	2.84 / 0.07	2.66 / 0.12	2.32 / 0.08
Chicago	1.16 / 0.95	0.76 / 0.97	0.62 / 0.10
Kuala Lumpur	0.00 / 0.00	-0.35 / 0.00	-0.72 / 0.00
Madrid	1.14 / 0.73	1.14 / 0.73	1.02 / 0.83
Mexico DF	2.28 / 0.21	2.21 / 0.24	2.08 / 0.17
Montreal	1.45 / 0.93	1.22 / 0.94	1.10 / 0.93
Moscow	2.30 / 0.77	1.99 / 0.91	1.81 / 0.76
New Delhi	0.00 / 0.00	-0.35 / 0.00	-0.72 / 0.00
San Francisco	1.81 / 0.44	1.55 / 0.44	1.53 / 0.41
Singapore	0.00 / 0.00	-0.35 / 0.00	-0.72 / 0.00
Stockholm	2.23 / 0.77	1.75 / 1.06	1.47 / 0.95
Tokyo	0.07 / 0.16	-0.23 / 0.24	-0.58 / 0.31
Antofagasta	1.23 / 0.46	1.16 / 0.45	1.13 / 0.49
Auckland	1.60 / 0.69	1.49 / 0.66	1.43 / 0.64
Brasilia	0.55 / 0.38	0.34 / 0.35	0.21 / 0.42
Brisbane	0.15 / 0.25	-0.27 / 0.16	-0.57 / 0.28
Johannesburg	1.67 / 0.47	1.32 / 0.50	1.32 / 0.47
Quito	2.86 / 0.07	2.80 / 0.05	2.43 / 0.04

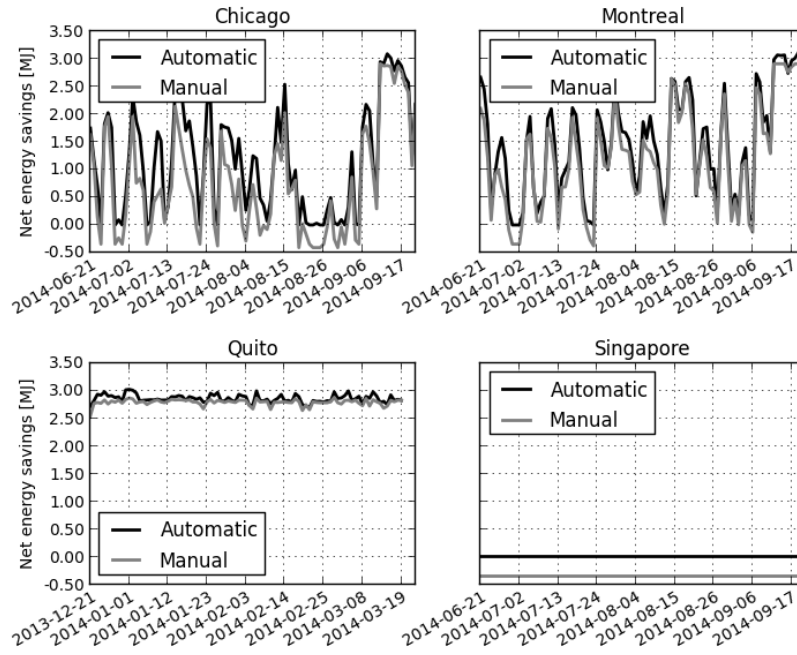


Figure 5: Net electrical energy savings comparison when using control based on reinforcement learning (automatic) versus manually scheduled charge period (from 06:00 to 08:00)

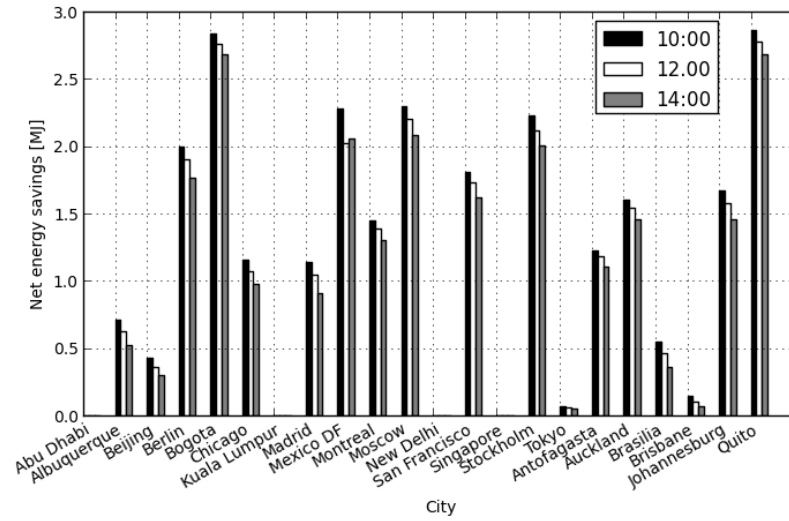


Figure 6: Net electrical energy savings comparison when discharging at different schedule

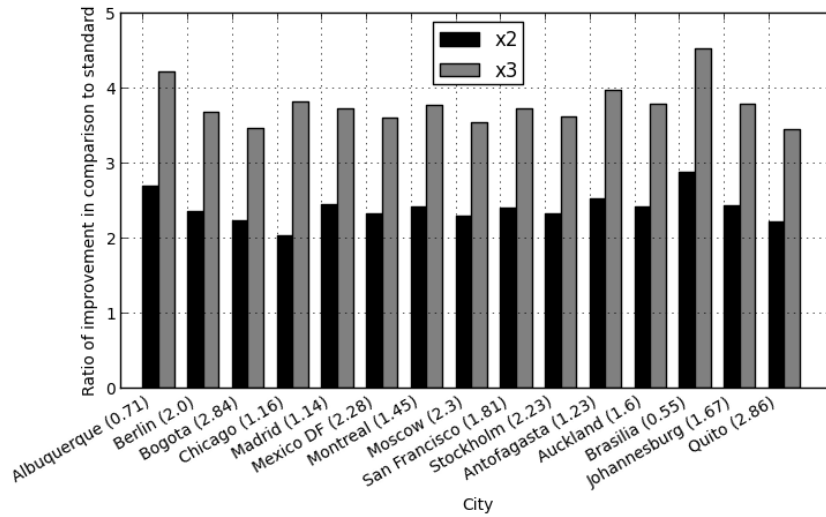


Figure 7: Ratio of increase of net electrical energy saving when using double or triple amount of PCM

3.3. *The effect of accurate forecasts*

In this section the performance of the system for scheduling based on predicted weather data versus real data is compared. The fans schedule is computed and decided according to the short term meteorological forecast obtained at 20:00 of the previous day. Then, the performance of the system was simulated according to this forecast predictions and to the real temperatures obtained the day after. This test is used to evaluate how sensitive the schedule defined by the reinforcement learning model is to the variations between the expected weather conditions and the real ones. The forecasted and real weather data were registered during two weeks at two different locations. The average performance of the system during these two weeks was slightly deviated in both places when comparing what was expected from the predictions and what really happened. In one hand, in Arenys de Mar (Csa), while the forecast predicted 1.79 MJ of net energy savings, only 1.48 MJ could be obtained (reduction of 17.3%). On the other hand, in Lleida (Cfb) the real net energy savings were increased to 2.33 MJ in comparison to the 2.14 MJ predicted (increase of 8.88%). As it was expected, the accuracy in the weather forecast is an important parameter in order to decide an optimal operational schedule of the system. As an example, Fig. 8 shows one day of operation in Lleida (Cfb). In this case, even the measured temperatures at night were lower than the predicted, they follow a similar evolution in time, hence the decision of whether activating the fans or not were mostly the same, and therefore slightly higher net energy savings were achieved.

4. Conclusions

The incorporation of TES active systems in the building design requires smart control techniques to enhance their performance and to make them attractive to architects and engineers, as well as to final users. A reinforcement learning algorithm is implemented to control the operational schedule of a ventilated facade with PCM in its air chamber. This algorithm uses a simple isothermal model to determine when is the optimal schedule to charge the PCM, and hence maximize the energy benefits from its use.

The Köppen-Geiger climate classification [9] was used to select and test the performance of this control system in one city for each climate region. It was demonstrated that the improvement due to the use of the automatic control is constant through all the summer period and all the studied climatic conditions. Furthermore, the paper evaluated how the starting time of the discharge and the amount of PCM influence in the performance of the whole system. On one hand, it was observed that the performance of the system is slightly affected by the starting time of the discharge (varying from 10:00 to 14:00) which reflects a high efficient storage sequence with low heat losses/gains from the outdoor environment. On the other hand, the amount of PCM is a critical parameter and has to be determined correctly according to the climate region. Increasing the amount of PCM by a factor of two or three enhances the net electrical energy savings achieved by the system by a higher ratio than this increased factor.

Finally, since the control system will be based on the weather forecast, the influence on the accuracy of this forecast was explored. It was observed that even though the performance of the system is very sensitive to an accurate weather forecast, the variations

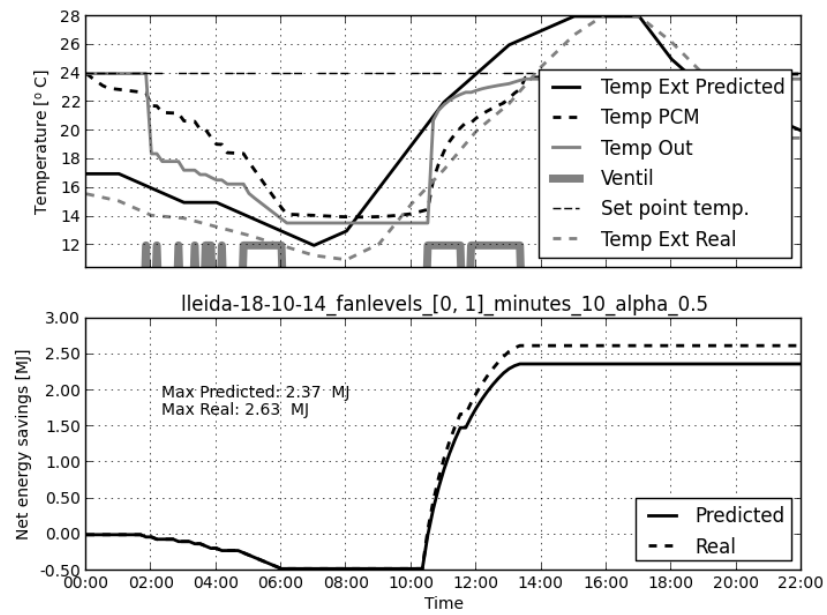


Figure 8: Performance comparison with real and predicted data

of the achieved net electrical energy savings between predicted and real weather data were lower than 18%.

Nomenclature

A_{conv}	Area air-PCM heat transfer	$[m^2]$
C_p	Heat capacity	$\left[\frac{J}{g \cdot K}\right]$
h	Heat transfer coefficient	$\left[\frac{W}{m^2 \cdot K}\right]$
L	Enthalpy of fusion	$\left[\frac{J}{g}\right]$
\dot{m}	Air mass flow rate	$\left[\frac{kg}{s}\right]$
m_{PCM}	Mass of PCM	$[kg]$
Q_{t+1}	Energy absorbed or released	$[J]$
S	PCM state	$[-]$
T^{inlet}	Inlet temperature	$[K]$
T^{PCM}	Average PCM temperature	$[K]$
T^{outlet}	Outlet temperature	$[K]$

References

- [1] Bellman, R., 1957. Dynamic Programming, 1st Edition. Princeton University Press, Princeton, NJ, USA.
- [2] Cabeza, L. F., 2014. Advances in Thermal Energy Storage Systems. Methods and Applications. Woodhead Publishing.
- [3] de Gracia, A., Castell, A., Fernández, C., Cabeza, L. F., 2015. A simple model to predict the thermal performance of a ventilated facade with PCM. Submitted to Energy and Building.
- [4] de Gracia, A., Navarro, L., Castell, A., Ruiz-Pardo, A., Álvarez, S., Cabeza, L. F., 2013. Thermal analysis of a ventilated facade with PCM for cooling applications. Energy and Buildings 65 (0), 508 – 515.
- [5] European Parliament and Council of the European Union, 2010. Directive 2010/31/EU. <http://www.epbd-ca.eu>.
- [6] Fernández, C., Manyà, F., Mateu, C., Sole-Mauri, F., 2014. Modeling energy consumption in automated vacuum waste collection systems. Environmental Modelling & Software 56 (0), 63 – 73.
- [7] Fernández, C., Mateu, C., Manyà, F., Sole-Mauri, F., 2015. Approximate dynamic programming for automated vacuum waste collection systems. Environmental Modelling & Software 67, 128 – 137.

- [8] IEA, 2012. Energy technology perspectives 2012. Pathways to a clean energy system.
URL dx.doi.org/10.1787/energy_tech-2012-en
- [9] Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15 (3), 259 – 263.
- [10] Rubitherm Technologies GmbH, 2014. www.rubitherm.de.
- [11] Sutton, R., 1988. Learning to predict by the methods of temporal differences. *Machine Learning* 3 (1), 9–44.
- [12] Sutton, R. S., Barto, A. G., 1998. *Introduction to Reinforcement Learning*, 1st Edition. MIT Press, Cambridge, MA, USA.
- [13] Watkins, C. J. C. H., May 1989. Learning from delayed rewards. Ph.D. thesis, King’s College, Cambridge, UK.
URL http://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf
- [14] Weather Underground, 2014. www.wunderground.com/weather/api/d/docs.