Available online at www.sciencedirect.com

SCIENCE DIRECT®

Energy and Buildings 38 (2006) 148–161

ENERGY and BUILDINGS

www.elsevier.com/locate/enbuild

Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory Part 2: Results and analysis

Simeng Liu [a,*], Gregor P. Henze [b]

[a] Department of Architectural Engineering, University of Nebraska-Lincoln, 1110 South 67th Street, PKI 243, Omaha, NE 68182-0681, USA
[b] Department of Architectural Engineering, University of Nebraska-Lincoln, 1110 South 67th Street, PKI 203D, Omaha, NE 68182-0681, USA

Received 31 March 2005; received in revised form 28 May 2005; accepted 3 June 2005

Abstract

This paper is the second part of a two-part investigation of a novel approach to optimally control commercial building passive and active thermal storage inventory. The proposed building control approach is based on simulated reinforcement learning, which is a hybrid control scheme that combines features of model-based optimal control and model-free learning control. An experimental study was carried out to analyze the performance of a hybrid controller installed in a full-scale laboratory facility. The first paper introduced the theoretical foundation of this investigation including the fundamental theory of reinforcement learning control. This companion paper presents a discussion and analysis of the experiment results. The results confirm the feasibility of the proposed control approach. Operating cost savings were attained with the proposed control approach compared with conventional building control; however, the savings are lower than for the case of model-based predictive optimal control As for the case of model-based predictive control, the performance of the hybrid controller is largely affected by the quality of the training model, and extensive real-time learning is required for the learning controller to eliminate any false cues it receives during the initial training period. Nevertheless, compared with standard reinforcement learning, the proposed hybrid controller is much more readily implemented in a commercial building.

Keywords: Load shifting; Thermal Energy Storage (TES); Optimal control; Learning control; Reinforcement learning

1. Introduction

As the first part of the report of this research, the companion paper [1] has presented a brief introduction to the general background of this project. The fundamental theory of classic reinforcement learning and its variation, simulated reinforcement learning has been introduced. The hybrid learning controller was developed based on the architecture of simulated reinforcement learning. In order to validate the feasibility and evaluate the performance of the hybrid control approach, an experiment was conducted at a full-scale laboratory facility called Energy Resource Station (ERS) at the Iowa Energy Center in Ankeny, IA. A detailed discussion and analysis of the experiment and its results are presented in the following sections.

2. Description of experimental study

2.1. Introduction to the experimental facility

The experiment was carried out in the Energy Resource Station, operated by the Iowa Energy Center (IEC). The ERS is a unique demonstration and test facility, where laboratory-testing capabilities are combined with real building characteristics. The ERS is capable of simultaneously testing two full-scale commercial building systems side-by-side with identical thermal loading. The ERS building, a

DOI of original article: 10.1016/j.enbuild.2005.06.002.

* Corresponding author.

E-mail addresses: sliu@mail.unomaha.edu (S. Liu), ghenze@unl.edu (G.P. Henze).

URL: http://www.ae.unomaha.edu/ghenze

0378-7788/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2005.06.001

single-story structure with a concrete slab-on-grade, has a height of 4.6 m and a total floor area of 855 m^2. The building is divided into a general area (office space, service rooms, media center, two classrooms, etc.), and two sets of identical test rooms, labeled A and B, adjacent to the general area. The eight test rooms are organized in pairs with three sets of zones having one exterior wall (east, south, and west) and one set that is internal. Fig. 1 presents a layout of the ERS including the four sets of identical test rooms used for the experiment.

The test facility has a central heating plant, consisting of a natural gas-fired boiler, and a cooling plant with three nominal 35 kW air-cooled chillers that operate in both chilled-water and ice-making modes. The chilled-water loop is filled with 22% propylene glycol water solution. In addition, the building includes a 440 kW h internal melt ice-on-tube thermal energy storage tank as well as pumps and auxiliary equipment needed to provide cooling. Hence, several modes of operation between these sources of cooling are possible in order to supply chilled-water to the air-handling units (AHUs).

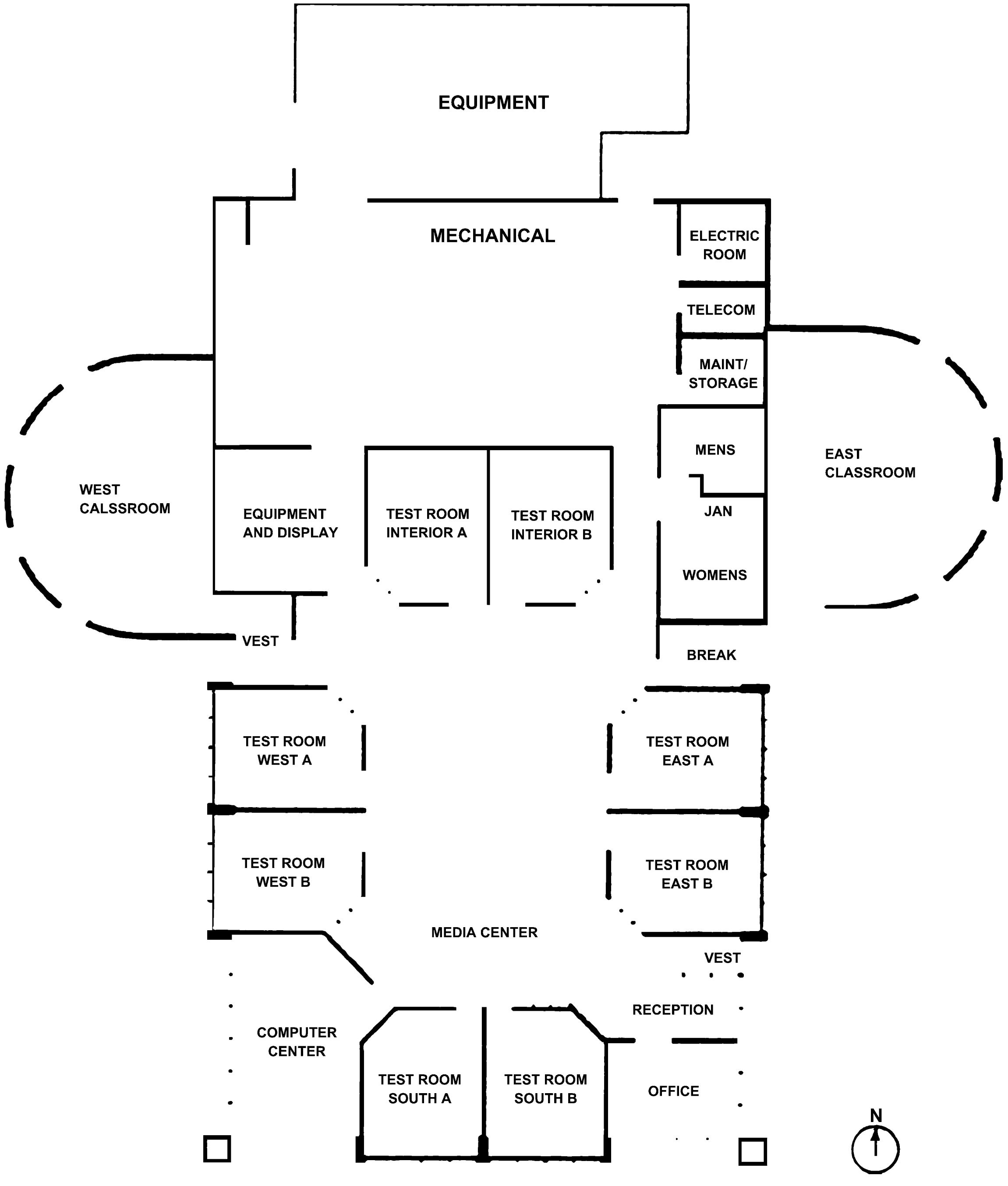

Fig. 1. Layout of the test facility at the Energy Resource Station (ERS), Ankeny, IA.

A primary–secondary flow arrangement, with dedicated constant-volume chiller pumps and secondary variable-flow distribution pumps in the AHU loop using variable-frequency drive (VFD) control, is installed in the ERS. Secondary HVAC systems include three AHUs that condition the building; test rooms A and B are served by two identical single-duct variable air volume (VAV) units with reheat AHU systems A and B, and the general area is served by a similar but larger AHU-1. An overhead air-distribution system utilizing pressure-independent VAV boxes supplies air to each test room using hydronic or three-stage electrical resistance reheat. Finally, there is an on-site weather station that measures outdoor air dry-bulb temperature, relative humidity, wind speed and direction, atmospheric pressure, total normal incidence solar flux, and global horizontal solar flux.

The ERS is not a particularly good candidate for the use of building thermal mass as documented by Braun [2] due to three reasons:

- It is a lightweight single-story structure with a high exterior surface area-to-volume ratio.
- Significant thermal coupling with the ground, the ambient environment and the zones adjacent to the test rooms is present.
- The test zones are not equipped with a representative amount of furniture, and the floor is carpeted, which reduces thermal coupling to the massive structure.

2.2. Hybrid control phase I: simulated learning

A training model of the ERS facility was developed in the Matlab/Simulink computing environment to model the dynamic thermal response of the building and energy consumption of the HVAC system. The model was calibrated using the experimental data previously obtained in the model-based predictive optimal control experiment.

The Q-learning algorithm was applied to train the learning controller. The configuration of state and action space, and Q-table, representing the evaluative information on the state–action pairs, was set up as depicted in Table 1.

As previous analysis by [3,4] revealed, the training procedure is strongly affected by the selection of learning parameters α and γ. A literature review shows that there is no general rule that can be applied to identify the optimal learning parameters. However, previous parametric analyses can provide a rule-of-thumb to initiate the learning and training processes.

Table 1
State and action space configuration for multi-task scenario

	Variable name	Dimension	Value range
State space	Building modes	3, 6	1–3, 1–6
	State-of-charge	10	0.0–1.0
Action space	Global zone air Temperature setpoint	10	20–24 °C on-peak 15–30 °C off-peak
	TES charging–discharging rate	20	$\mu_{\min} - \mu_{\max}$

Table 2
State and action space configuration for the simulated learning phase

Simulation cases	γ	α	Action-selection algorithm		Training period (day)
			ε-Greedy	Softmax, τ	
1	0.65	0.15	0.10	–	4000–6000
2	0.65	0.05	–	0.8	3000
3	0.65	$0.05 \rightarrow 0$	–	0.6	3000

Time constraints did not allow us to carry out an extensive parametric analysis to identify the optimal learning parameters. The Q-table was initiated with zero for entries, and then trained sequentially by different sets of learning parameters. Three simulation cases primarily contributed to the final formation of the Q-table. Table 2 lists the selected learning parameters.

As Table 2 shows, the controller starts with a higher learning rate. The selection of the ε-greedy exploration algorithm can make the controller explore more evenly among all the available actions. The second simulation then starts with the previously trained Q-table, the learning rate decreases and the softmax algorithm is applied with a higher temperature. The third case is similar to the second case; however, a dynamic learning rate, which exponentially decays with training time, and a lower temperature are applied.

2.3. Hybrid control phase II: implemented learning

With the experience or knowledge from the simulated training phase, the controller was implemented at the ERS test facility. Control programs were developed for the learning controller to govern the operation of the test facility by the reinforcement learning algorithm. The control sequence of the learning controller can be seen in Fig. 2. All of the control programs were developed in Matlab, and the overall structure is depicted in Fig. 2.

As shown in Fig. 2, a main supervisory control program initiated the learning parameters and governed the control

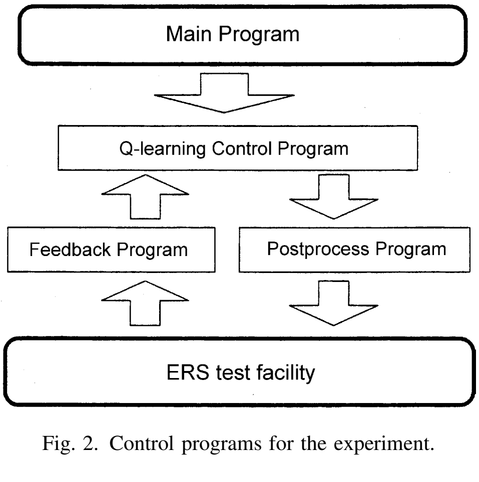

Fig. 2. Control programs for the experiment.

sequence. The control commands were issued hourly. The following three major routines were called sequentially each hour:

(1) *Reinforcement learning control*: At the beginning of each hour, the control action was selected by the learning agent. It used the softmax exploration algorithm to select an action according to the current knowledge base of the learning controller. The knowledge is contained in the form of a state–action pair lookup table, which was previously trained.
(2) *Post-processing program*: The control strategy was then interpreted by a post-processing program, which was responsible for: (a) setting up the communication channel with the ERS and (b) sending the post-processed command to the BAS.
(3) *Building feedback*: This program monitored the power consumption of the HVAC system including the chillers, pumps, and AHU fans. At the end of each hour, the monitored power consumption was summarized and the current hourly cost was then calculated. This information was used to update the knowledge base of the learning controller.

3. Analysis of experimental results

The experimentation was carried out from September 6 to 16, 2004 covering two experimental cases. In the first case, the learning controller operated hourly, and in the second case, the controller ran in 30 min intervals. It was expected that the learning controller would evolve faster with higher sampling frequency. However, due to physical limitations of the cooling plant, operation at a 30 min frequency caused the plant to have problems for executing the action selected by the controller. Transitions between different operating modes made it difficult to operate stably within a 30 min period. As a result, the action selected by the controller could not be executed fully, and the state transition was not accomplished as intended. On the other hand, this effect did not significantly influence the overall performance of the hourly case, since this frequency was acceptable for plant operation. For this reason, the discussion of experiment mainly focuses on the analysis of data for the hourly case. It should be noted that the limited experimentation time did not allow for the effect of learning control to be clearly observed, but it still allows for the evaluation of reinforcement learning control in terms of feasibility, robustness, and adaptiveness.

3.1. Experimental data analysis

The experiments carried out can be considered as the second phase of the hybrid control scheme, i.e., the implemented learning phase. In this phase, the learning parameter settings were different from the ones used in the simulated learning phase due to the following reasons:

(1) The learning controller had been trained in the simulated learning phase. Even though the controller was not expected to work truly optimally, it would have found the fundamentally correct action patterns through training by the simulator. As the parametric analysis showed, the learning rate should decrease over time. As a consequence, smaller learning rate is expected to be applied in the implemented learning phase.
(2) In the ERS field implementation, the objective of the implemented learning phase was not only to guide the learning controller to find the optimal policy, but also to let the learning controller take control of the overall building to save operating costs. To meet this objective, the learning controller was encouraged to select the action that would bring the highest cost savings most of the time. This is referred to as the greedy policy in our previous discussion. The greedy policy is achieved either by using lower ε in ε-greedy exploration method, or by choosing lower temperature τ value in the softmax method. Both methods encourage the controller to take the greedy policy most of the time but to periodically select exploratory actions.

The learning parameters for the simulated learning phase were shown in Table 2. Table 3 compares the learning parameters applied in the experiment and simulation training for the last case, which refers to case 3 in Table 2.

As shown in Table 3, both the learning rate, α, and softmax exploration parameter, τ, are reduced compared with the value used in the simulated learning phase. By doing so, the learning controller is expected to prefer the greedy policy during most of the experiment time. The first experiment case ran from midnight September 6 to midnight 12, 2004. Fig. 3 depicts the ambient conditions during the experiment.

Fig. 4 depicts the profiles of the zone air temperature setpoint, one of the two control actions of the learning controller, and the measured room air temperature, which is taken as the hourly average value of all eight test rooms.

From Fig. 4, we can see that the actual room temperature follows the setpoint profile in general, which implies the action was well executed by the plant As shown in Fig. 3, the weather during the test was not a typical late summer condition, but cooler than expected, especially, from September 7 to 9. For this reason, the room temperature did not float up much during the off-peak period even though the controller chose a setpoint of 30 °C. The setpoint at the on-peak period, which is also the occupied period, was

Table 3
Learning parameters for the experiment

Hybrid control phase	γ	α	Action-selection algorithm	
			ε-Greedy	Softmax, τ
Simulated learning	0.65	0.05 → 0	–	0.6
Implemented learning	0.65	0.01	–	0.1

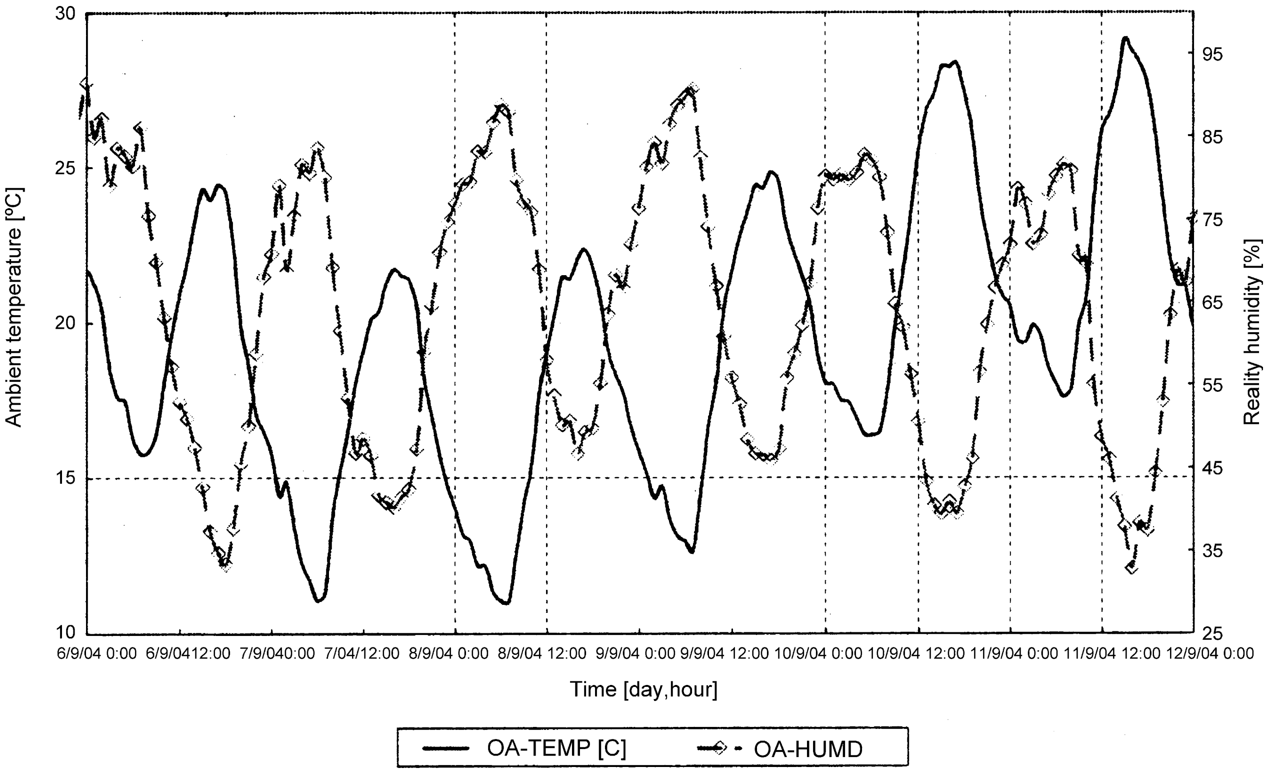

Fig. 3. Ambient conditions during first experiment.

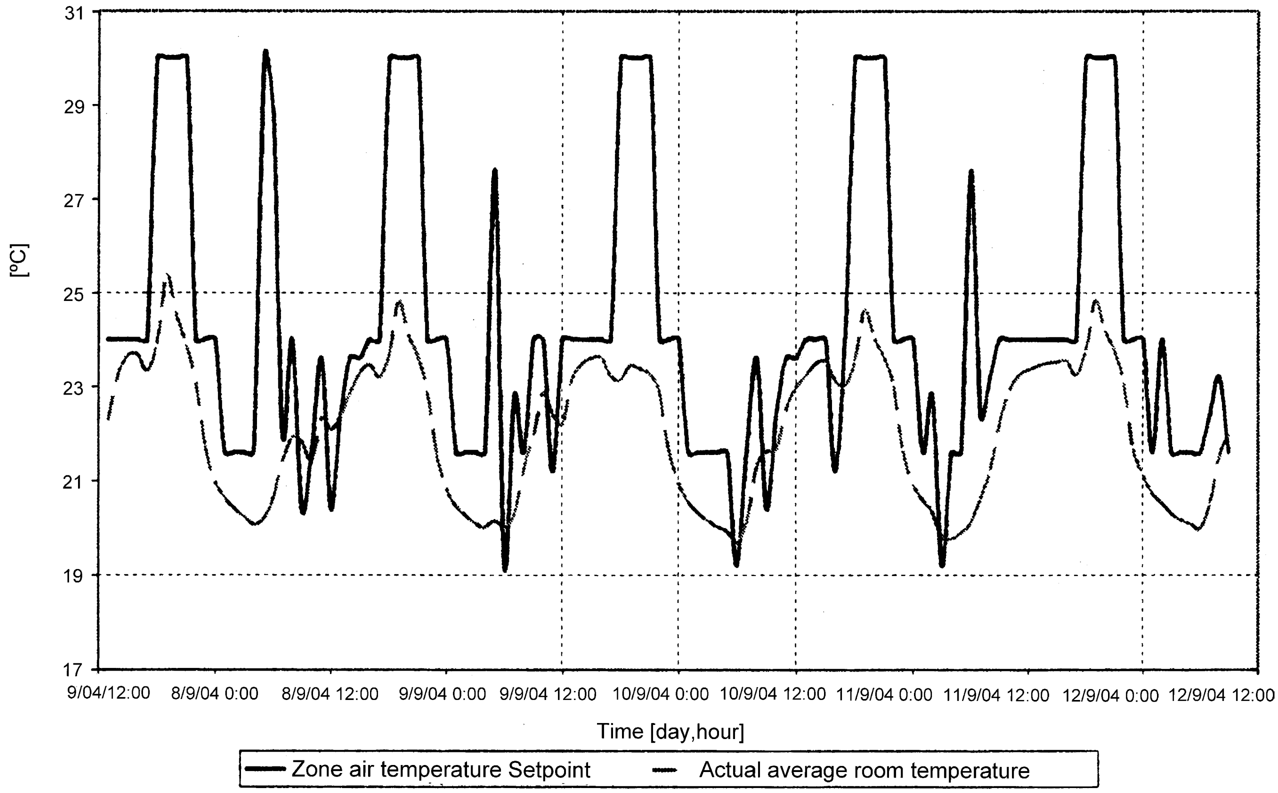

Fig. 4. Test room temperature profiles.

around the upper bound of the feasible range, which is 24 °C. The controller set the setpoint as low as 19–22 °C from midnight until the onset of the on-peak period. This shows that the controller found the benefit of utilizing the passive thermal storage by precooling the building with the knowledge from the simulation training. As mentioned earlier, the ERS test facility is not a good candidate for demonstrating the effect of passive thermal storage, which was also revealed by the experiment with model-based predictive optimal control the previous year [5]. As observed in the simulated learning phase, even though the controller had realized the merit of precooling, it was still hard for the controller to identify the truly optimal action. For this reason, the setpoint profile was not smooth during the early-morning off-peak period. After the on-peak period but before midnight, the controller decided to let the setpoint float because precooling would not be effective this many hours before the next on-peak period.

Fig. 5 depicts the action profiles of the TES system and the resultant state-of-charge. It can be observed that the TES is cyclically charged in the off-peak period and discharged during the on-peak period as expected. It should be pointed out that there are modifications that were made to the configuration of the action space during the experiment. The upper bound $u_{\max}$ of the charge rate during the on-peak period and the lower bound of the discharge rate $u_{\min}$ during the off-peak period were set to zero. By doing so, action for the TES was restricted to charging in the off-peak period, and discharging during the on-peak period. The modification was made with two considerations in mind. One is that such operation is almost common sense for TES operation, and can be considered as a priori knowledge for the controller, before it is implemented in the real application. In the previously conducted simulation analysis, the learning controller did find the charging–discharging cyclical operation pattern given enough training time. The other reason is relative to the plant operation in the ERS. Switching modes between charging and discharging involves complicated valve operations in the piping system. Immediately after switching modes, the cooling coil load may not be met due to the fact that chilled-water temperature cannot immediately be maintained at the desired setpoint. Since the learning controller is operated hourly, the original action space would allow the plant to oscillate between charging and discharging, especially when the plant is not operating according to the greedy policy but is exploring the benefit of unseen actions. This could lead to the room temperature not being maintained, and potentially damage the system because of the frequent changes in the operation mode. Furthermore, the learning controller in the implemented learning phase is supposed to be trained with previous experience. It would not make sense to allow the controller to make ice in the occupied and on-peak period when we know this is not a good choice.

Fig. 5 also indicates that the TES is only partially utilized. The state-of-charge is only charged to <50%, and the main

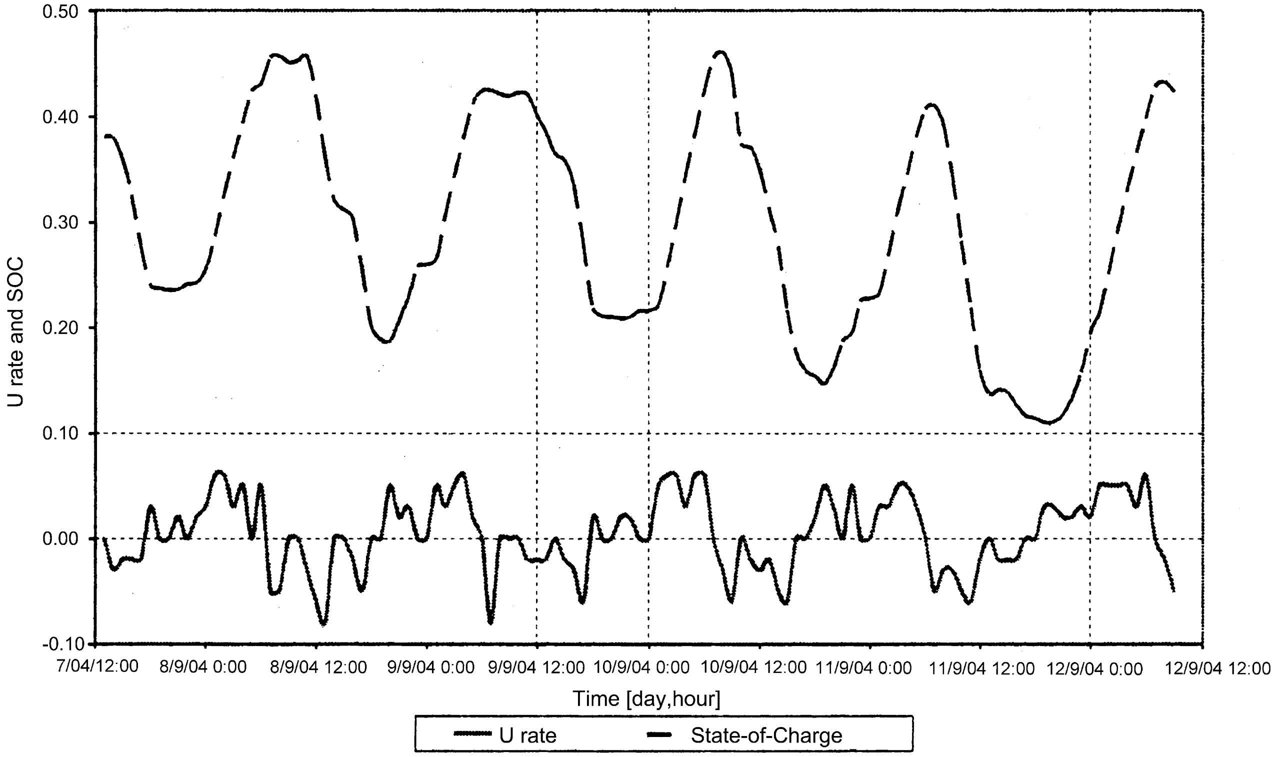

Fig. 5. TES action and state-of-charge profiles.

chiller still needs to be turned on to meet the load even though TES inventory is available. This indicates that the learning controller did not fully discover the benefit of utilizing the TES system. However, manually calculating the cumulative charging rate of the TES shows that the state-of-charge should be higher than the measured value. In fact, this phenomenon was also observed in the predictive optimal control experiment in 2003, where approximately, 13% charging load was not deposited into the TES due to heat loss. The same effect was found in the hybrid control approach experiment. Besides, heat loss, there is another reason that contributes to the ineffectiveness in the TES charging mode. In the off-peak period, there are two operational modes most often invoked by the learning controller: charging the TES and dormancy of the HVAC plant. According to ERS historical data, the TES can be charged up to 70% overnight if there are no interruptions. However, the controller may switch the plant operation mode between charging and dormant because the learning controller runs hourly. During the transition mode, the chilled-water cannot immediately reach the desired setpoint, which is about −5 °C. As a result, the chiller needs to be cooled down to effectively charge the TES every time the mode is changed. This effect cannot be neglected when there are many operating mode changes.

Even though the learning controller did not fully utilize the TES due to the reasons provided above, the control strategy is still reasonable because the controller did find the right control action pattern. Comparison of the entries of the Q-table before and after the experiment shows that the 5-day operation is not enough to change the values much because of the limited test time and the low learning rate.

3.2. Calibration of the training model

Another important concern is whether the model used in the simulated learning phase matches the actual environment. Even though the calibration procedure was carried out before the experiment, it is still necessary to evaluate the accuracy of the training model as it determines the pre-trained experience of the controller. Fig. 6 compares the measured cooling load during the test days and the simulated load that uses the actual weather condition, recorded zone air temperature setpoint, and TES charging–discharging rate during the experiment.

It can be seen that there are deviations between the two profiles, but the simulated cooling load fundamentally captures the trend of the measured data, and the agreement is considered acceptable. However, in the simulated learning phase, the TES is modeled as a simple ideal thermal storage system. The ineffective TES charging mode and the heat loss due to the poor insulation of the ice tank had not been considered, and as a result, drastic deviations were found between the measured TES inventory, and the simulated one. In addition, the actual discharging rate was higher than the simulated value due to heat loss through the ice tank insulation.

The simulated learning phase serves in the role of a teacher who is responsible for offering the student, which in

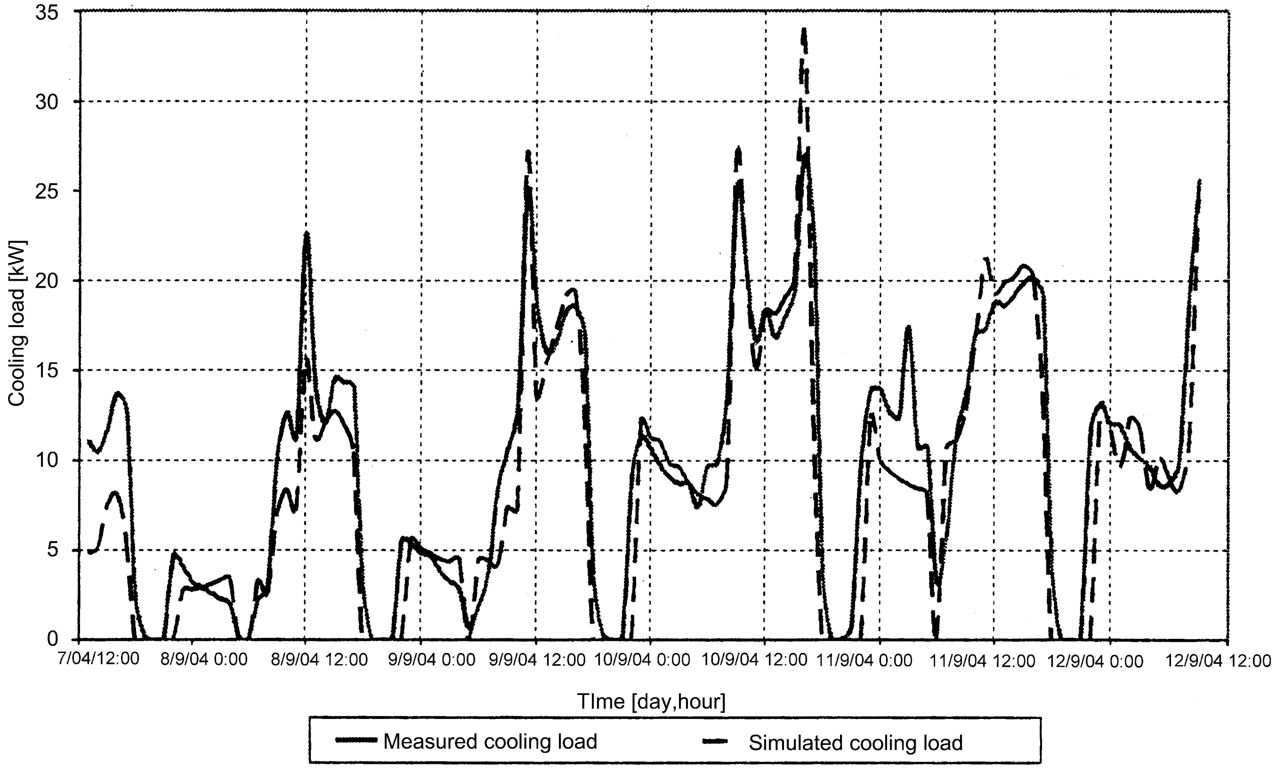

Fig. 6. Cooling load profiles of measured and simulated data.

our case, is the learning controller, fundamental process knowledge before the student is exposed to the actual environment. An obvious question is how the mismatch of the training model affects this knowledge and consequent control actions. In order to answer this question, the training model needs to be calibrated, and the simulated learning phase should be repeated using the refined model. The model will also be used to evaluate the performance of the learning controller by comparing it with other control strategies. It is not possible to carry out the comparison of these control strategies in the actual facility due to lack of available testing time and inability to replicate exact weather conditions. Therefore, the training model should be calibrated as close as possible to the actual HVAC plant.

The model was calibrated through system identification [6]. Two calibration procedures were carried out consecutively. The first one aimed at minimizing the deviation of the state-of-charge profiles between simulated and measured values. Two correction factors were introduced to reflect these factors contributing to the loss of state-of-charge. One correction factor is a discount factor $F_1 \in (0, 1]$ that was applied in the charging process. In the model-based predictive optimal control experiment [5], about 12% charging load was lost due to heat losses, which implies that $F_1 = 0.88$. This value was used as the initial value in our calibration procedure. The second factor is an amplification parameter $F_2 > 1$, which was applied in the discharging process. It is not necessarily the reciprocal of F_1 because other factors may contribute to the ineffectiveness of heat transfer as explained in the last section. The second calibration is intended to improve the match of the HVAC plant power profiles, and the calibrated parameters are the equipment efficiencies for pumps, fans, and chillers. Table 4 lists the initial and calibrated parameter values for the calibration procedures.

Table 4
Calibration of training model

Parameters	F_1	F_2
Calibration of state-of-charge (SOC)		
Initial value	1.00	1.00
Calibrated value	0.62	1.28

Parameters	η_{fan}	η_{pump}	COP_{chw}	COP_{pre}	COP_{ice}
Calibration of power data					
Initial value	0.85	0.85	2.1	3.4	2.4
Calibrated value	0.65	0.8	2.4	2.9	2.8

In Table 4, η_{fan} and η_{pump} stand for the motor efficiency of the circulation fans and pumps; COP_{chw} and COP_{ice} indicate the COP for the main chiller in chilled-water making mode and ice-making mode. COP_{pre} is the efficiency of the additional precooling chiller that is responsible for the precooling load when the main chiller is not available. Figs. 7 and 8 present the state-of-charge profiles and power consumption profiles of the plant with the calibrated model.

It can be seen from both Figs. 7 and 8 that better agreement between the measured and simulated data was

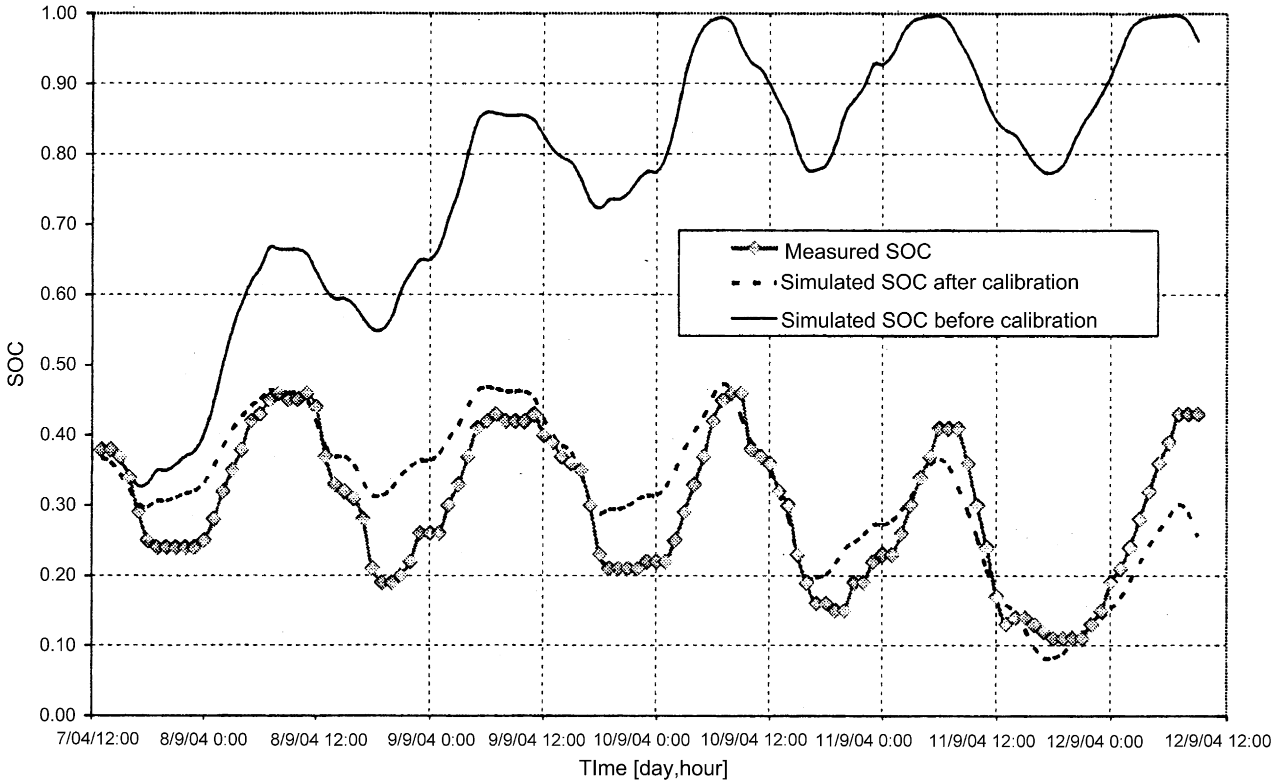

Fig. 7. State-of-charge profiles of the calibrated model.

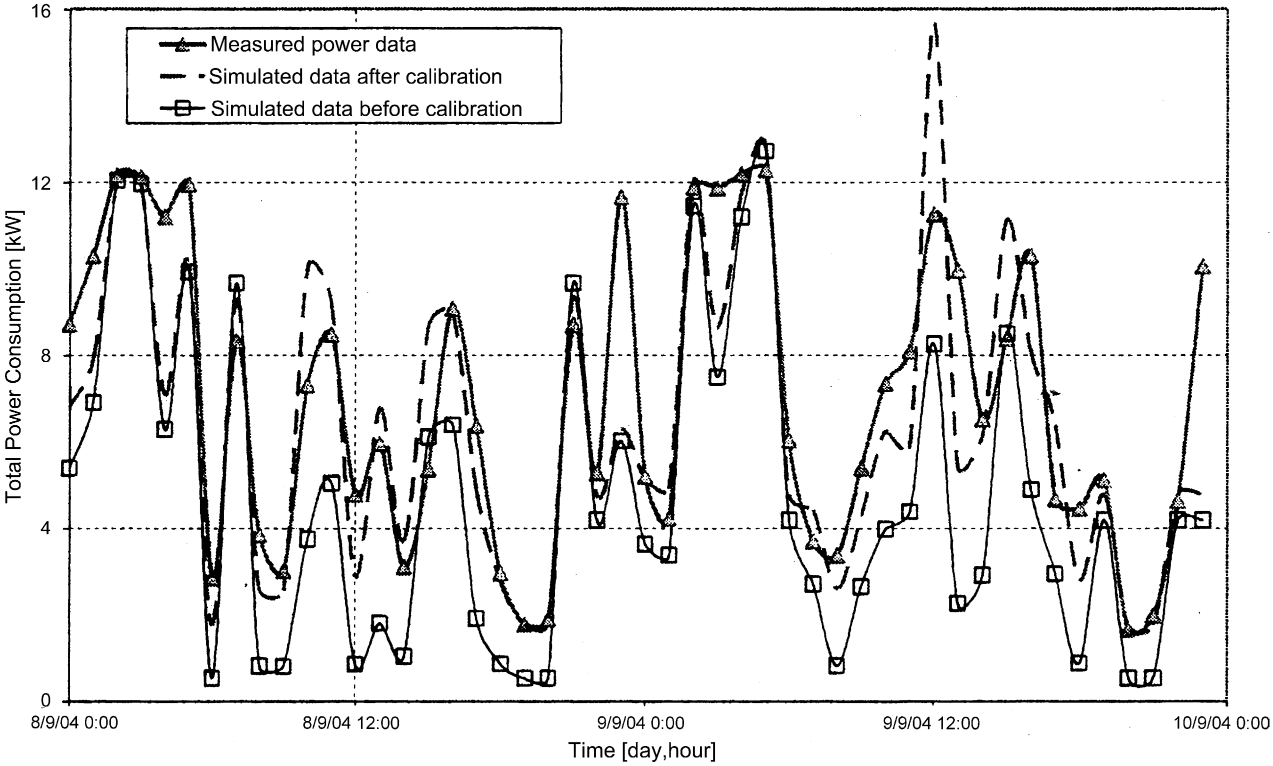

Fig. 8. Power profiles of the calibrated model.

achieved through the calibration process. The discount factor F_1 is lower than the value calculated in the experiment of the model-based predictive optimal control approach. The power-data deviation is mainly attributed to the incorrect initial efficiency values for fans and pumps, which are higher than the calibrated values.

The COP value of the new chiller is lower than previously assumed, and the calibration confirmed the idiosyncratic fact that the main chiller operation in the ice-making mode was more energy efficient than in the chilled-water mode, which was also revealed in [5].

3.3. Performance evaluation

With the calibrated model, the performance of the controller can be evaluated from the perspective of cost savings. The hybrid controller is compared with the following five control strategies:

- *Base case*: This case stands for the simplest but most costly scenario, in which neither active nor passive thermal storage is considered. The zone air temperature is controlled by nighttime setback control.
- *Storage-priority with night setback*: This case stands for the situation that utilizes the active thermal storage only using storage-priority strategy. Standard chiller-priority is not discussed because the main chiller is sized to meet the entire load. The zone air temperature is controlled by nighttime setback control, like the base case.
- *Optimal control of passive thermal storage only*: In this case, no active TES system is considered, but the building is controlled with optimized zone air temperature setpoints that are found by model-based optimization.
- *Optimal control of passive thermal storage with storage-priority*: This case considers both thermal storage media. The building is controlled with the optimized zone air temperature setpoints, which are found by model-based optimization, and storage-priority is used to control the active thermal storage.
- *Model-based optimization of active and passive thermal storage*: By using the recorded actual weather condition and the calibrated model, model-based optimization can be carried out assuming no mismatch in modeling and weather prediction. This is considered to be the true optimal control strategy, and is supposed to have the maximum cost savings among all cases.

Table 5 summarizes the plant operating cost and savings for the investigated cases. Case 6 uses the measured data from 13:00 p.m. on September 7 to 9:00 a.m. on September 12, 2004, which is the execution period of the first experiment hybrid control. All other cases are carried out using the calibrated model, and the recorded actual weather conditions are used for the same time period. As mentioned earlier, the ERS test facility is not a good candidate for demonstrating the effect of passive thermal storage. In our analysis, utilizing the passive thermal storage only yields 6.7% cost savings. Storage-priority control is able to meet all

Table 5
Comparison of costs and savings of hybrid control strategy with other cases

No.	Case	Cost ($)	Saving (%)
1	Base case	102	–
2	Active thermal storage only (storage-priority control)	92	9.8
3	Passive thermal storage only (optimized T_{sp})	95.1	6.7
4	Active and passive thermal storage (storage-priority control with optimized T_{sp})	86.6	15
5	Active and passive thermal storage (model-based optimization)	68	34
6	Measured hybrid learning control	93.5	8.3
7	Simulated hybrid learning control	91.9	9.9

the on-peak cooling load only with the TES, the savings are reduced by excessive charging and the ice-making COP is not particularly favorable. As expected, case 5 of model-based optimization offers the highest savings by accurately and judiciously using both active and passive thermal storage media. However, these savings can only can be achieved when the model and prediction is perfect. Case 4 also provides better cost savings compared with utilizing either passive or active thermal storage only, but the TES is still overcharged. It can be seen that even though the model is calibrated, the simulated operating costs (case 7) are still not identical with the measured value. Both achieve cost savings between the best cases (cases 4 and 5) and the worst cases (cases 1 and 2). They provide unimpressive but reasonable cost savings because both active and passive thermal storage inventory have been only partially utilized.

3.4. Refined simulation analysis

The performance analysis in the last section shows that the hybrid controller did bring about cost savings compared with the base case. However, the simulation studies, where different control strategies were applied to the calibrated model also revealed that there is a greater savings potential that the learning controller did not fully realize. Since limited experimental time did not allow the controller to explore further, a simulation study was carried out using the calibrated model in order to further analyze the hybrid control approach. The simulated learning phase was

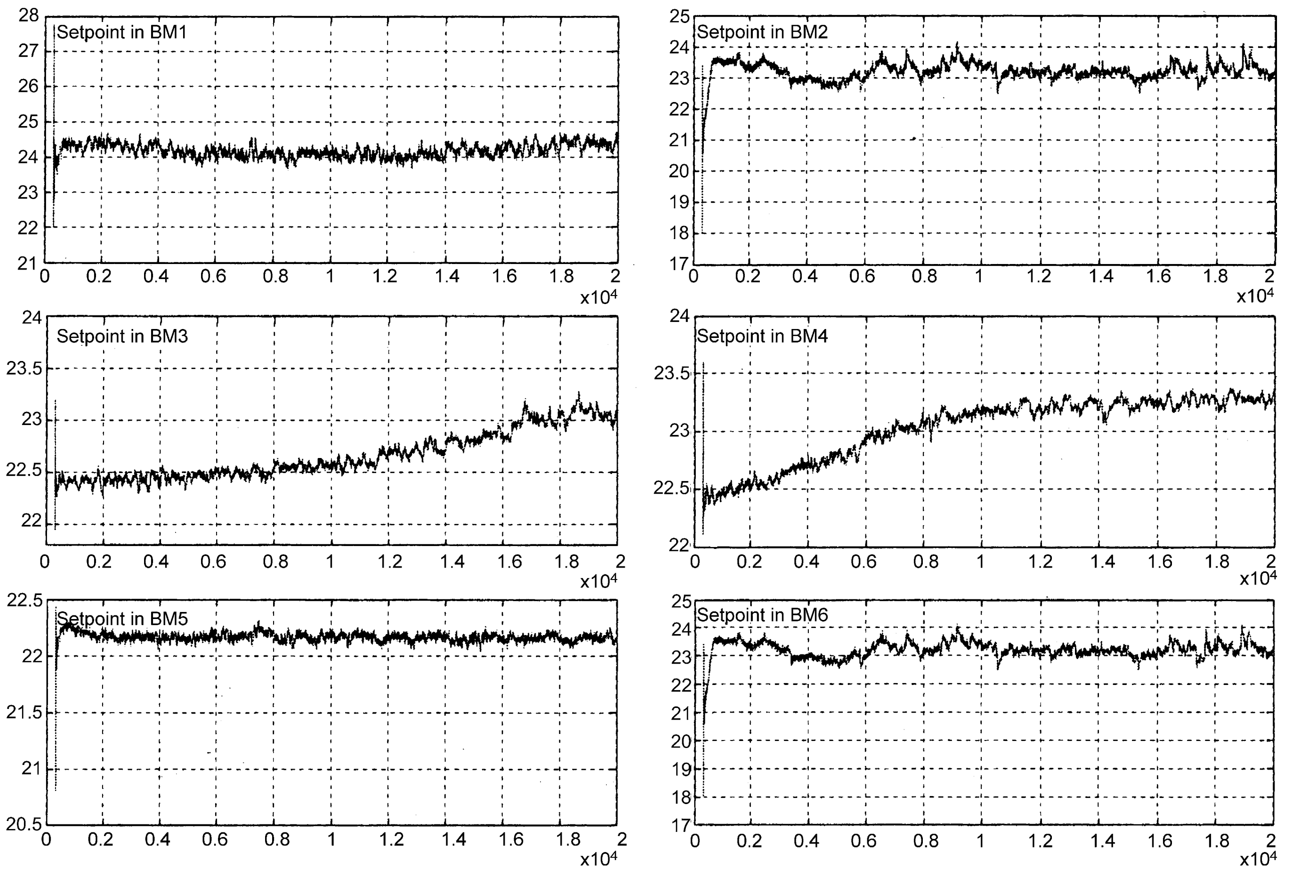

Fig. 9. Learning of T_{sp} with calibrated training model.

Table 6
Comparison of optimization case 1 with experiment

Case	Cumulative charge rate	Cumulative discharge rate	Cost ($)	Savings (%)
Actual experiment	1.82	−1.32	93.5	8.3
Simulated experiment	1.93	−1.47	91.9	9.9
Simulated experiment with calibrated model	2.38	−1.85	89.6	12.1

repeated to re-train the learning controller with the calibrated training model. The controller was then implemented in the simulation environment to repeat the experiment for the same weather data and plant conditions, e.g., initial state-of-charge of the TES. The objective of the simulation study was to see if the controller performed better or differently than with the uncalibrated training model.

Simulation cases have been carried out with different learning parameter settings. It is interesting to note that in most cases, the learning controller behaves differently after using the calibrated model. In general, the learning controller tends to use the active TES system more than the passive thermal storage inventory. Figs. 9 and 10 depict the learning process of a typical case.

In Fig. 9, the setpoints in the off-peak period (building modes 1, 2, 5, and 6) remained around 24 °C, but did not go down as had previously occurred. On the other hand, the TES charging rate increased in building modes 1, 2, and 6, and the discharging rate increased also in the on-peak period. This implies that the learning controller recognized that the TES needed to be commanded with higher values of charging and discharging activity due to the introduced discount factor, which represents the heat loss of the TES system. Applying the re-trained Q-table, the experiment was repeated in the simulation environment with the calibrated model.

Figs. 11 and 12 compare the control actions of the learning controller in the repeated simulation and the actual experiment The T_{sp} profiles shows that less precooling occurred in the repeated simulation, which confirms the inference of Fig. 9. In Fig. 12, the repeated simulation shows that more active storage activity had been commanded. Table 6 compares the cumulative TES activity in the refined simulation study with the measured data and previous simulation.

It can be clearly seen that the active TES system is more extensively utilized in terms of cumulative charging–discharging rates. This can be explained as follows: first, through calibration, it was found that during charging of the

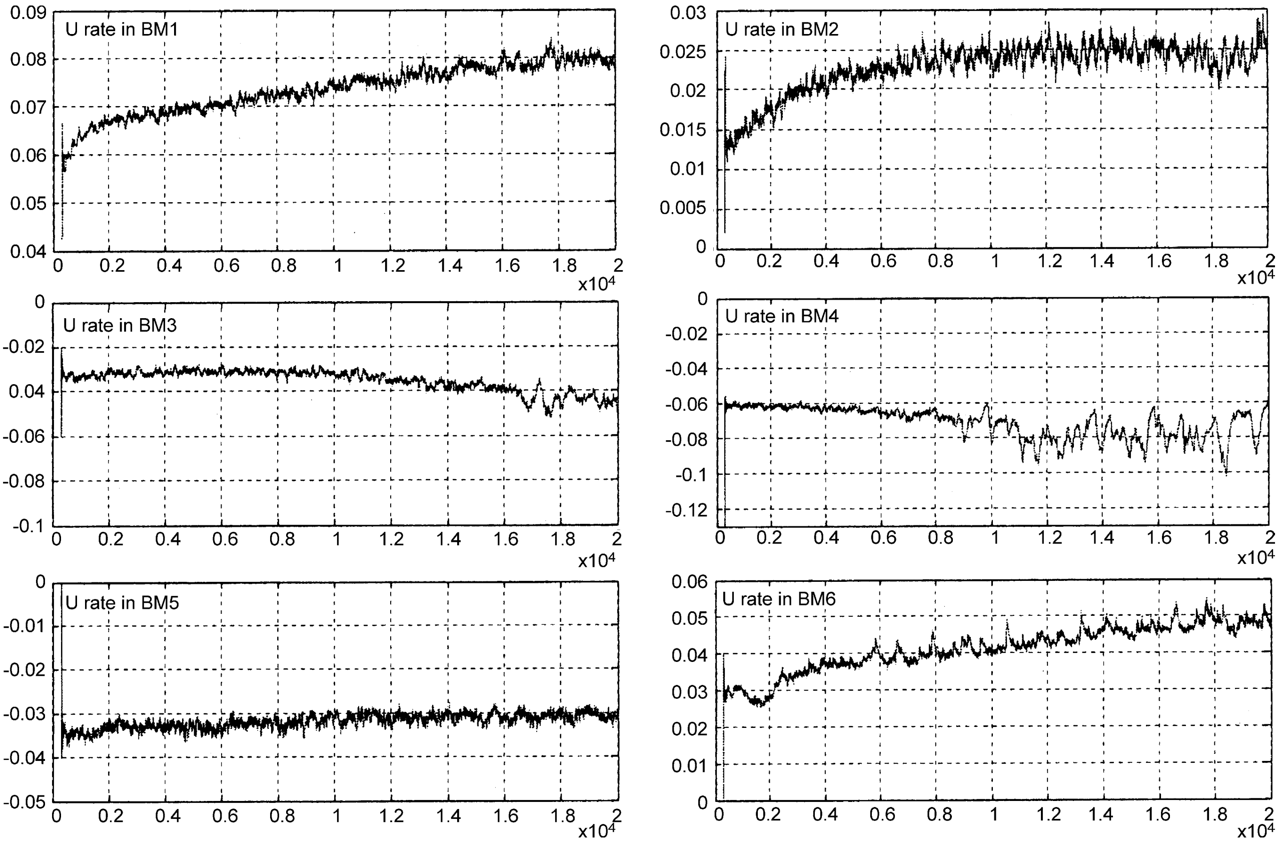

Fig. 10. Learning of u-rate with calibrating training model.

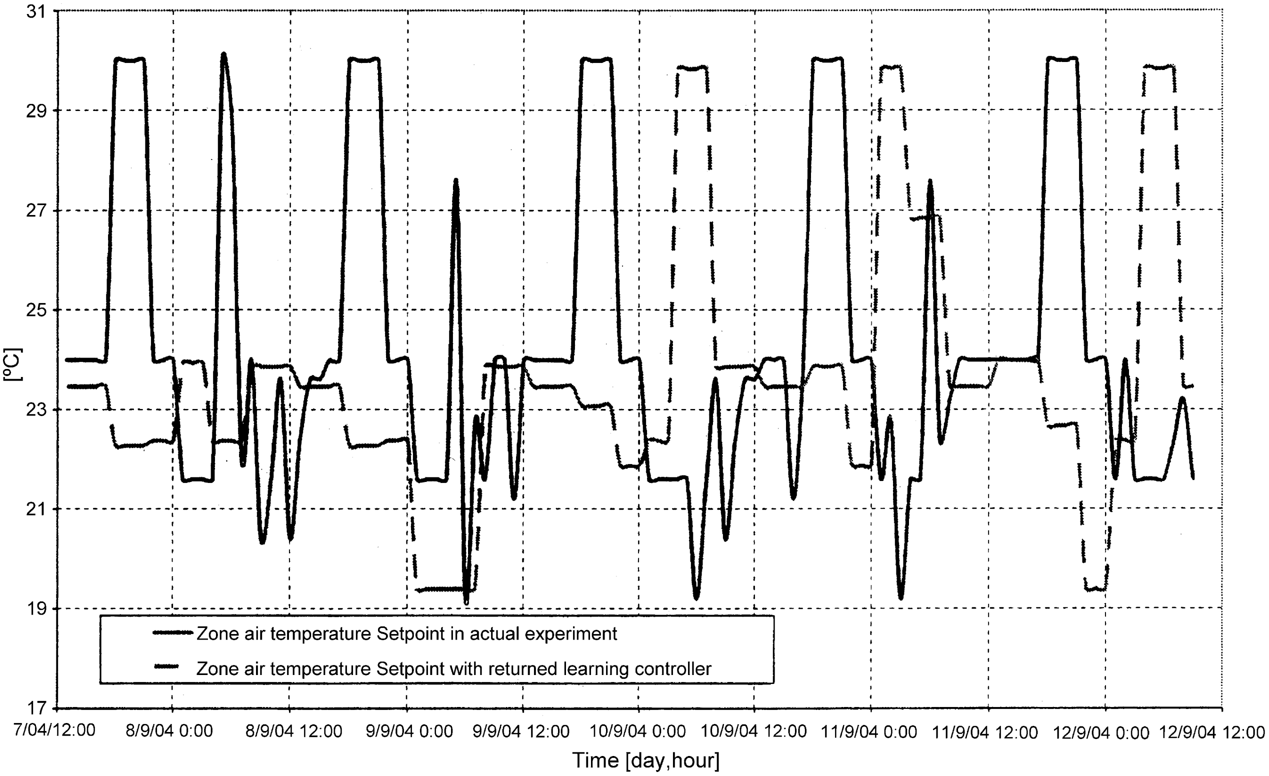

Fig. 11. T_{sp} with re-trained learning controller.

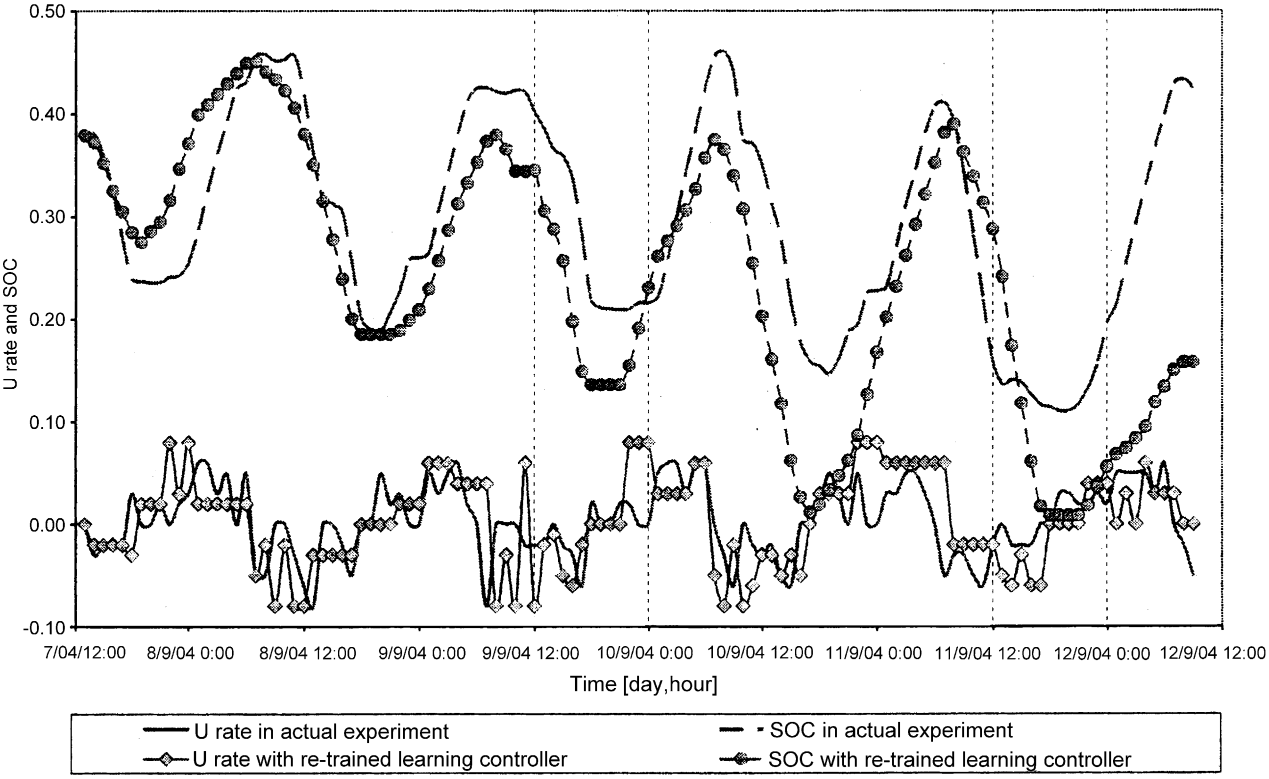

Fig. 12. u-Rate with re-trained learning controller.

active TES system, only 62% ($F_1 = 0.62$) of the ice-making chiller load leads to state-of-charge changes. Therefore, for the same change in TES inventory, the charge rate has to be $1/F_1$ times what it would be for an ideal lossless TES system, leading to substantially higher charge rates. Conversely, during discharging for the same contribution to the cooling load, the TES tank has to be discharged by $F_2 = 1.28$ times the value of a perfect TES system, i.e, the tank is depleted 28% faster, and consequently, the discharge rates are higher. Secondly, the COP value of the main chiller in ice-making mode COP_{ice} proved to be higher than initially assumed. As a consequence, the learning controller realized that charging active TES inventory at night was less costly, and thus, more TES inventory was used.

On the other hand, the utilization of passive thermal storage was not clearly observed in the refined simulation study as shown in Fig. 11, compared with the uncalibrated simulation. This is due to several factors. First, the building itself is relatively lightweight and the possible load shifting effect was small. Second, the learning parameters previously found for the uncalibrated training model were no longer effective in the refined simulation study, and modified learning parameters had not been found for the calibrated training model. Third, the COP value of the precooling chiller was lower than the initially assumed value. This reduces the potential benefits of precooling and makes it harder to be discovered by the learning controller. Overall, the learning controller in the calibrated training model achieves higher savings by utilizing active thermal storage inventory more extensively as shown in Table 6.

4. Conclusions

A hybrid control approach that is based on simulated reinforcement learning has been introduced in this paper. The hybrid approach was validated by an experiment carried out in the Energy Resource Station Laboratory building in Ankeny, IA, and an evaluation was made by analyzing the experimental data regarding the following aspects.

4.1. Feasibility

The performance analysis demonstrated that the hybrid control approach can provide reliable control utilizing both active and passive thermal storage inventories. Data analysis showed that the actions selected by the controller were well interpreted and executed. Previous trained knowledge guided the controller into the "right zone" to govern both storage media. The controller's activity was controlled by setting the appropriate learning parameters. The greedy policy will be taken most of the time in order to save operating costs. At the same time, by properly denning the action space, the controller will be kept from violating any constraints in thermal comfort and plant operation when it explores in search of a better control policy.

4.2. Advantages

The performance of the hybrid control approach was compared with a variety of control strategies. The hybrid control approach achieved 8.3% cost savings over the base case using the measured data. It outperformed both control strategies that use passive thermal storage only, but was inferior to storage-priority control and other control strategies utilizing both storage medias. However, the simulated data hint at better savings, and it is reasonable to believe that the hybrid controller is better than any control strategy that only uses one thermal storage media. Since the thermal storage inventories were only partially used, the hybrid controller cannot compete with either optimal control of passive thermal storage and storage-priority control of active thermal storage. However, the model-based predictive optimal control is assumed to be the true optimum because of no mismatch in modeling and prediction, which is hardly achievable in reality.

The hybrid control approach was designed based on simulated reinforcement learning, and substantial improvements were made to overcome the shortcomings of slowness that are associated with standard reinforcement learning. The controller was trained by a simulator first in the simulated learning phase, which greatly reduced the amount of time required in the standard reinforcement learning scenarios to acquire the same knowledge. During the implemented learning phase, the controller will first adopt the near-optimal policy learned in the simulation training, but due to the learning feature of reinforcement learning, it will improve the performance of the controller continuously. In summary, the hybrid control approach enjoys the advantages of the model-based approach in the simulated learning phase, and the advantages of the model-free approach in the implemented learning phase because there is no modeling and prediction required and because it is adaptive. Furthermore, the experiment shows that the procedure of implementing the controller is easier compared with model-based predictive optimal control. The core of the learning controller is a lookup table, the *Q*-table, and the remaining supervisory control program is less complex compared with a model that imitates the whole building. It would be possible for control systems manufacturers to implement the learning controller without much effort.

4.3. Disadvantages

As previously discussed, the quality of the simulator, or training model, is a key factor that determines the fidelity of the pre-trained knowledge of the learning controller. Deviations in the model may lead the controller to only find suboptimal policies. The controller will have difficulty in finding the optimal policy due to the fact that the learning rate is set to comparatively low values in the implemented learning phase in order to make the controller act greedily most of the time. Because the implemented learning phase takes on the form of standard reinforcement learning, it suffers from the same disadvantages of pure reinforcement learning. The learning parameters and the curse of dimensionality of the state and action space also impact the controller's ability to find the optimal policy.

References

[1] S. Liu, G.P. Henze, Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory, Part 1: theoretical foundation, Energy and Buildings, 38 (2006) 142–147.

[2] J. Braun, Load control using building thermal mass, Journal of Solar Energy Engineering 125 (3) (2003) 292–301.

[3] S. Liu, G.P. Henze, Reinforcement learning control for building active and passive thermal storage inventory, in: Proceedings of SimBuild: 2004, Boulder, CO, 2004.

[4] S. Liu, G.P. Henze, Reinforcement learning control for building active and passive thermal storage inventory, in: Proceedings of the 2005 International Solar Energy Conference, Orlando, FL, 2005.

[5] G.R. Henze, D. Kalz, S. Liu, C. Felsmann, Experimental analysis of model-based predictive optimal control for active and passive building thermal storage inventory, International Journal of HVAC&R Research 11 (2) (2005) 189–214.

[6] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Englewood Cliffs, NJ, 1987.