Learning Control for Batch Thermal Sterilization of Canned Foods

S. Syafiie [a,*] F. Tadeo [a] M. Villafin [b] A. A. Alonso [b]

[a] *Department of Systems Engineering and Automatic Control, University of Valladolid, 47011 Valladolid, Spain, {syam|fernando}@autom.uva.es*

[b] *Process Engineering Group IIM-CSIC, Vigo, Spain, {marcosvm|antonio}@iim.csic.es*

* Corresponding author.

Learning Control for Batch Thermal Sterilization of Canned Foods

Abstract

A control technique based on Reinforcement Learning is proposed for the thermal sterilization of canned food. The proposed controller has the objective of ensuring a given degree of sterilization during Heating (by providing a minimum temperature inside the cans during a given time) and then a smooth Cooling, avoiding sudden pressure variations. For this three automatic control valves are manipulated by the controller: a valve that regulates the admission of steam during Heating, and a valve that regulate the admission of air, together with a bleeder valve, during Cooling. As dynamical models of this kind of processes are too complex and involve many uncertainties, controllers based on learning are proposed. Thus based on the control objectives and the constraints on input and output variables, the proposed controllers learn the most adequate control actions by looking up a certain matrix that contains the state-action mapping, starting from a preselected state-action space. This state-action matrix is constantly updated based on the performance obtained with the applied control actions. Experimental results at laboratory scale show the advantages of the proposed technique for this kind of processes.

Key words:
Intelligent Process Control, Sterilization Process, Food Process, Batch Process, Reinforcement Learning.

1 Introduction

The food industries are nowadays facing critical changes in response to consumers, which, in addition to health and safety awareness, demand an ever larger diversity of food products with high quality standards. On the other hand, these industries are in a permanent quest for new markets and population sectors not accessible before, which immediately translates into the search for more efficient processes, in order to gain market share (Bruin and Jongen, 2003).

This paper concentrates on the design of controllers for a specific process in the food industries, namely the so-called thermal processes for sterilization of canned foods (Lewis, 2006; Ramaswamy and Singh, 1997). These processes are very important for minimizing the activity of harmful microorganisms in food, thereby reducing health risks and increasing the durability of the products. For the problem at hand, the microorganism activities are reduced through thermal sterilization in pressurized retorts using steam. Unfortunately, thermal processing also produces the deterioration of the organoleptic properties of the food when conditions are not carefully controlled. For this reason, an appropriate control of the process is fundamental to guarantee the safety and quality of the products (Lewis, 2006; Ramaswamy and Singh, 1997).

Thus, the central objective of controllers for the sterilization process is the inactivation of microorganisms present in the foodstuff, while preserving as much as possible product quality, avoiding very quick variations in temperature and pressure and minimizing the operation time. For this, the sterilization process can be divided in three stages that use different control strategies: Venting, Heating and Cooling. Venting in normally carried out manually, so the stages of the process relevant from the point of view of controller design are Heating (where the main objective is to ensure a given degree of sterilization by ensuring a given temperature during a certain time by manipulating the entrance of steam in the retort), and Cooling (where the temperature is carefully decreased by replacing the steam with air).

The kinetics of thermal destruction of microorganisms or degradation of nutrients are usually assumed to follow pseudo-first-order kinetics (e.g. the TDT model) with an exponential-type temperature dependence (Balsa-Canto et al., 2002a,b). Such kinetics constitutes the basis to quantify the degree of sterilization, usually given in terms of *lethality* (in units of time), that defines the amount of time required to produce a certain decimal reduction. For details, the reader is referred to Ramaswamy and Singh (1997). Unfortunately, due to the complexity of the process, the variability of the products to be sterilized and the reduced number of sensors it is not feasible to derive models adequate for model-based controller design. To deal with this issue, this paper concentrates on the application of a control technique based on learning. More precisely, a Model-Free Learning Controller (MFLC) will be develop for this thermal sterilization processes. This MFLC is based on *Reinforcement Learning*, so it is an agent-based technique based on re-framing the problem of achieving process control objectives by learning through interaction with the process (see Figure 1), taking always into account the inherent constraints in input and output signals. The (*agent*) interacts with the rest of the process (also called *environment* in learning approaches): the agent selecting actions and the environment responding to those actions and presenting new situations to the agent. The environment also provides rewards, that are numerical values that the agent tries to maximize, as they give a measurement of performance (Sutton and Barto, 1998). More specifically, the agent and the environment interact at each of a sequence of discrete time step. At each time step, the agent receives some representation of the environment's state, and on that basis selects an action. The agent receives a numerical reward, and moves to a new state (Sutton and Barto, 1998). Thus, the reward function depends on the recent state, action and successor state: with time, the agent gathers more information and provides optimal actions for every visiting state.

Although Reinforcement Learning ideas seem promising, they were not developed for process control problems (Sutton and Barto, 1998; Bertsekas and Tsitsiklis, 1996), so in this paper the Model-Free Learning Control (MFLC) technique (Syafiie et al., 2007a; Syafiie et al., 2007b) is used to control the sterilization process. This MFLC is gives a feasible implementation of Reinforcement Learning for process control problems, by providing a precise but simple definition of symbolic states and actions, based on control objectives and the constraints on input and output variables. This methodology is complementary to other intelligent control approaches (such as Fuzzy Logic or Neural Networks), in the sense that initial values for the parameters of the MFLC algorithm can be derived from previous controllers. Starting from these initial parameters, using learning MFLC provides a simple methodology to improve the controller by interaction with the plant.

The rest of this article is structured as follows: First the background and scope are stated in Section 2. A short presentation of the thermal sterilization process is given in Section 3. The proposed technique to control the sterilization process by using Model-Free Learning Control (MFLC) is given in Section 4. The MFLC application for controlling a sterilization process at laboratory scale is discussed in Section 5. Finally, some conclusions are given in Section 6.

2 Background and scope

In industrial sterilization processes for canned food the most common controllers are still PID. For example in Mulvaney et al. (1990), a Proportional Integral (PI) controller was developed for this process. A study using a combination of the linearizing-transformation of differential geometry and the quality-control of Q-PID/Q-PI was presented by Alonso et al. (1993), whereas a PID-type controller with parameters selected using Internal Model Control (IMC) was reported by Alonso et al. (1997, 1998). It was found that PID controllers work well during Heating as long as the plant is operated in small neighborhoods of the constant-heating temperature around the tuning region; unfortunately, frequently the controllers have to be retuned to operate in other conditions (for example, when the type and amount of cans change) , which is cumbersome.

Advanced control strategies have also been proposed for this process, such as the online correction of the lethality value reported by Teixeira and Tucker (1997). In Kuma et al. (2001), an algorithm based on three control modes was presented, but no specific proposal was given on how to regulate the steam, water, drain, air and bleeder valves. An optimal control problem with state and control constraints governed by a nonlinear heat equation was proposed by Kleis and Sachs (1999). The discretized optimal control was expressed as a large-scale continuous optimization, which can be solved using sequential quadratic programming. However, the proposed algorithm was mathematically complicated. A closed-loop optimal receding horizon controller (RHC) incorporating model uncertainty was designed and studied by Chalabi et al. (1999), where a non-gradient method was used to solve the corresponding nonlinear optimization problem. Unfortunately, this kind of controllers requires that all the states of the system to be measurable, which is impractical. Since all these advanced controllers are difficult to design and need a precise mathematical model of the process, the most frequent control technique in industry is still, therefore, a manual supervision of PID controllers.

To deal with problems of batch to batch variations and the complexity of the models for control, techniques based on learning would be adequate as they adapt to the specific situation at hand through the result of previous experiences. Techniques based on Reinforcement Learning have been selected, as they provide a rigorous methodology for learning without detailed mathematical models of the controlled plant, using a simple algorithm suitable for real-time implementation (Sutton and Barto, 1998).

In particular the MFLC approach, previously proposed by some of the authors (Syafiie et al., 2007a; Syafiie et al., 2007b), will be used to control the thermal processing, as it corresponds to a feasible implementation of Reinforcement Learning algorithms (Sutton and Barto, 1998) for Process Control. This technique is used because it is simple and does not need a precise *a priori* model of the process, but incorporates basic knowledge of the process behavior (information from output range, control limitations, loop interactions, etc). Thus, in MFLC controllers the control objective is expressed as the optimization of a desired performance index by learning to apply appropriate control actions through interaction with the plant. In particular, the MFLC approach proposed here is based on Q-learning (Sutton and Barto, 1998; Bertsekas and Tsitsiklis, 1996). However, the idea can be easily augmented to improve learning speed by applying other methodologies in literature, such as lazy learning (Atkenson et al., 1997a,b), near optimal closed-loop control (Ernst, 2003) and q-iteration with CMACS (Timmer and Riedmiller, 2007).

We must point out that, although for simplicity, and in order to represent industrial practice, the problem at hand is represented as a sequence of two dynamical systems (during Heating a single-input single-output system, and during Cooling a two-input single-output system), if needed the proposed approach can be extended to more complex multiple-input multiple-output systems using the ideas of Riedmiller (1997).

3 Batch Thermal Sterilization Process

The thermal sterilization processes for prepackaged food can be carried out in continuous or batch units. This article concentrates on learning to control the thermal sterilization process in batch units, as it is the most frequent approach in the industry, and the one that can make better use of a learning approach.

3.1 Process Description

The sterilization process is assumed to be carried out in batch steam retorts as depicted in Figure 2. A typical operation cycle involves several stages, which in this paper are assumed to be the following:

- *Venting*: In this initial stage, steam is introduced in the retort to eliminate the air, so heat transmission is more efficient during Heating. At this stage, bleeder and drain valves are open. When the pressure in the retort, P_r, matches that corresponding to saturated steam, P_s, at that temperature, there is only steam in the retort, so Heating can start.
- *Heating*: The objective of this central stage is that the temperature inside the retort is at the level required, for enough time to reach the desired microbiological lethality. At time t the lethality $F(t)$ is defined as follows:

$$F(t) = \int_0^t 10^{\frac{T(\mathrm{t})-T_{ref}}{z_{ref}}} d\mathrm{t} \tag{1}$$

  where z_ref and T_{ref} are parameters that depend on the container and the product, which are obtained experimentally, and $T(\mathrm{t})$ is the temperature at the critical point (the point inside the product with lowest temperature), (see Ramaswamy and Singh (1997); Alonso et al. (1997)). This lethality is affected by small variations in the temperature, so automatic control is required during this cycle.
- *Cooling*: Once the Heating period concludes, the product is cooled with water down to room temperature. At the same time, air is injected into the retort to avoid sudden pressure drops that could result in the bursting of the product containers. Pressure control during this stage is especially important for glass containers or conduction heated-type products where the existence of sharp temperature gradients between the inside and the outside of the product induces high differential pressure (Alonso et al., 1997, 1998).

4 MFLC Technique

The Model-Free Learning Control technique (MFLC) that is proposed here for batch sterilization processes is a control technique, based on Reinforcement Learning (Sutton and Barto, 1998; Bertsekas and Tsitsiklis, 1996), which gives a feasible implementation of automatic learning in process control problems, by providing a precise definition of symbolic states and actions, based on control objectives and the constraints on input and output variables. It has been presented in detail by the some of the authors in Syafiie et al. (2007a); Syafiie et al. (2007b), so only the main ideas are given here.

4.1 MFLC Architecture

The MFLC architecture is represented in Figure 3: as with most Reinforcement Learning algorithms, it is based on describing the system in terms of symbolic states, so the controller learns how good the application of a given action in a given state is, by applying the action to the system and then checking the quality of the response. The evaluation of the effect of each action is done by estimating the expected return mathematically, storing the values of this return (which measure the quality of the response) in the so-called Q-matrix (discussed in section 4.2).

The MFLC is based on a precise selection of states, actions and control signals (discussed in sections 4.3 and 4.4), with the objective of representing typical problems in process control and being easily understood by the final user. The operation of the algorithm, represented in Figure 3 is based on, first, the selection of the agent of one action from those available in the current state, using the "Policy". Then, the action is converted to a control signal in the "Calculation U" block. Then, based on the measured output, the "Situation" block estimates the next state and the corresponding reward. From this reward, the so-called Q-value is updated in the "Critic" block, which reflects the adequacy of the action applied. As time goes by, actions are selected by the agent, and learning is carried out by checking the quality of the response: Actions that drive the system into the goal state are considered to be good, so its Q-value is increased. On the other hand, actions that do not drive the system into the goal state are punished.

4.2 Q-matrix

Mathematically, the objective in MFLC is to maximize the expected return (Sutton and Barto, 1998) taking into account the control and state constraints.

A central part of the learning algorithm is the estimation of this expected return. For this, the state-action value function, $Q(s, a)$, is used, as it contains the expected return, when starting from the state s, the agent applies the action a, and thereafter follows the policy π:

$$Q^{\pi}(s,a) = E_{\pi}\{R_t|s_t = s, a_t = a\} = E_{\pi}\left\{\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}|s_t = s, a_t = a\right\}. \quad (2)$$

This function is stored in a matrix $Q(s_t, a_t)$, the Q-matrix. At each sampling time, these Q-values are calculated by taking into account the current and future benefits: when action a_t has been selected and applied to the plant, the system moves to a new state, s_{t+1}, and receives a reinforcement signal, r_{t+1} (which evaluates the quality of the response), so the Q-matrix is updated as follows:

$$Q(s_t, a_t) \leftarrow (1-\alpha)Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{b \in A_{s_{t+1}}} Q(s_{t+1}, b)] \quad (3)$$

where:

- The learning rate, $\alpha \in (0, 1]$, is a tuning parameter that can be used to optimize the speed of learning (a large learning rate makes learning faster, but might induce oscillations). It is required for computation of expectation in the form of an iterative averaging.
- The discount factor, $\gamma \in (0, 1]$, is used as a factor to weight the effect more heavily in the near future: If γ is small, the agent learns to behave only for short-term reward; the closer γ is to 1 the greater the weight assigned to long-term reinforcements.
- $A_{s_{t+1}}$ is the finite set of possible actions in the new state.

4.3 State Representation

A central issue in all Reinforcement Learning algorithms is the definition of the states, which are symbolic and represent the "distance" to the goal. In MFLC, the states are defined based on the control objective and the constraints on the control signal and the states, as follows: the control objective is considered to be to maintain the desired output inside the band $r - d$ and $r + d$, as shown in Figure 4. The width of this band is defined based on the tolerance of the system (which depends on measurement noise, disturbances and the specifications). This band is defined as the *goal band*, and corresponds to the *goal state*, where the agent should drive the system and ensures that it remains there (it is now assumed, without loss of generality, that is exactly in the middle of the working range). To describe the rest of the symbolic states, it is considered that the agent has h states from the goal state to the maximum positive or minimum negative error of the system, f (Selecting h is a trade-off: this number must be large enough to describe all the different responses of the process, but small enough to reduce computational time and the size of the Q-matrix). The "span" of each state can be calculated as follows:

$$c = \frac{f-d}{h}. \tag{4}$$

Thus, the positive bound parameter can be presented as:

$$\omega_i = d + (i-1)c, i \in [1, ..., h] \tag{5}$$

(For negative errors, the bound parameter is trivial by changing signs). Thus, the vector of symbolic states can be presented as follows:

$$g_j = \begin{cases} e - \omega_j \text{ if } e \leq \omega_j; \\ \omega_j - e \text{ else,} \end{cases} j \in [1, ..., 2h+1] \tag{6}$$

where e is the tracking error. The symbolic current state, s_t, is just:

$$s_t = \arg\max_j(g_j). \tag{7}$$

4.4 Action Representation

In the single-input single-output version of MFLC, the control signal $u_t \in \mathbb{R}$ is calculated by varying the previous control signal in a magnitude calculated from the difference of the numerical values of the selected optimal action, $a_t \in \mathbb{N}$, with respect to the *wait action*, a_w (action corresponding to maintaining the previous control signal). That is:

$$u_t = u_{t-1} + k(a_w - a_t), \tag{8}$$

where k is the tuning parameter. This gives a PI-like structure, which simplifies initialization and tuning for the end user. At each state there is only a finite set of possible actions (see Figure 5). These actions are selected based on the systems description: in particular, from the limitations on the minimum and maximum variations of the control signal, as follows: Let the control variations be bounded as follows:

$$\underline{\Delta u} \leq |\Delta u| \leq \overline{\Delta u}, \tag{9}$$

where $\underline{\Delta u}$ and $\overline{\Delta u}$ are known bounds. The number of total actions needed to satisfy the constraints can be calculated as follows:

$$N_a = 2h\left(round\left(\frac{\overline{\Delta u} - \underline{\Delta u}}{kh}\right)\right) + 1, \tag{10}$$

where the round-up function is used. From (8), (9) and (10), the value corresponding to the wait action a_w, can be calculated as follows:

$$a_w = \frac{N_a + 1}{2}. \tag{11}$$

If there is no overlapping, the number of actions in each state can be calculated being $n_a = \frac{N_a - 1}{2h}$. However, to increase the number of available actions and represent nonlinear action-to-space relations (important in process control), a degree of overlapping must be included (see Figure 5). Of course, at each state, not all the actions are available: Each state has a subset of actions. For example, during Heating, if the tracking error for temperature is very small, the only actions available are those that increase to correct the temperature. Thus, the number of actions in each state is $n_a^\beta = n_a(1+\beta)$, where β is a parameter that gives the degree of overlapping with neighboring states (always selected such that n_a^β is integer). Then, the available actions for every state go from a_p^j to a_b^j (except in the goal state, where there is only the wait action). The idea is presented in Figure 5 and developed in Syafiie et al. (2008). Those available actions can be calculated as

$$\begin{aligned} a_p^j &= a_p^{j-1} + (j-1)v, \\ a_b^j &= a_p^j + n_a^\beta - 1, \end{aligned} \tag{12}$$

where $v = \beta \frac{n_a^\beta}{h}$ and a_p^{j-1} is the first action in the state j calculated as

$$a_p^{j-1} = \begin{cases} 1, & \text{if } j = 1 \\ 2a_w - a_b^{j-2}, & \text{if } j = h+2 \end{cases}. \tag{13}$$

The strategy for selecting one action from those available ones is through *exploration* and *exploitation* policies. The agent explores those available actions to know the optimal value function by executing trial actions, following the ε-greedy policy (Sutton and Barto, 1998). This means that the action which has the maximum Q-value will be selected with $1-\varepsilon$ probability and the rest will explore trial actions selected from those available in the state.

5 Thermal Control of Prepackaged Food

This section explains the application of MFLC ideas for batch thermal sterilization. The first part of this section discusses the control strategy, followed by a discussion on the selection of the parameters of the controllers for the Heating and Cooling stages of these sterilization processes.

As discussed in Section 3, there are three crucial steps in controlling the sterilization process: Venting, Heating and Cooling.

The proposed control strategy for these cycles is shown in Figure 6. As the venting stage can be controlled using a simple technique (keeping bleeder and drain valves fully open until the pressure inside the retort P_r reaches the steam pressure P_s), the control application therefore concentrates on Heating and Cooling. The use of MFLC for Heating and Cooling is now presented.

5.1 Heating Control Strategy

During Heating, the control objective is to maintain the temperature inside the goal band by manipulating the steam valve. To evacuate the condensed water from the retort, the drain valve is open. Also, the bleeder valve is slightly open.

Mathematically, during Heating, the objective is to maintain the retort temperature within a tolerance of $\pm 2.0^o$C with respect to the provided reference. Thus, the goal band is $r - 2.0$ to $r + 2.0$. The output range is considered to be ± 4.0 ^{o}C with respect to the reference. Thus, from these numbers and following the ideas presented in Section 4, there are 21 symbolic states, where state #11 corresponds to the goal state. The actions are then defined based on the possible control variations: the signal must vary within the following bounds:

$$0.0001 \leq |\Delta u| \leq 0.008. \tag{14}$$

Thus, the Q matrix size is 1601×21, where the wait action is action #801 (this matrix will be denoted Q_H). The tuning parameter is selected to be $k = 10^{-5}$, based on the control constraints. To include some nonlinearity, a small overlap is considered, with the number of actions in every symbolic state to be 158. Therefore, in state #1 the actions are $\#1, \cdots, \#158$, in state #2 the actions are $\#71, \cdots, \#228$, and so on, following (12). The controller parameters are summarized in Table 1.

The objective of the control task is to maintain the process in the goal state, or return it to the goal state if there has been any disturbance or change of reference. To achieve this, maximum reward is introduced for actions causing the process error to be smaller than the previous one. Actions that move the system away from the goal band are punished. Therefore, the reward is given as:

$$R_t = \begin{cases} 1.0 & \text{if } |e_t| \leq |e_{t-1}|, \\ -1.0 & \text{otherwise.} \end{cases} \tag{15}$$

Of course, more complex reward functions could be selected, but this particular reward function has been selected following the ideas in Smart (2002), which recommends not indicating a detailed path for the agent to achieve the goal, but only the goal, as the path assumed to be the most adequate might not really be the best (learning takes care of finding the most adequate approach). Thus, this gives an approach parallel to the Mayer-type objective functions in Optimal Control (Stryk and Bulirsch (1992)), with the trajectory constrained by the limited number of actions available in each state.

Heating finishes when the desired lethality time t_l is reached (where t_l is evaluated from (1). That is, denoting by t_v the starting time of the Heating, the agent switches from Heating control to Cooling control when $t >= t_v + t_l$.

5.2 Cooling Control Strategy

The state-action space has been discussed in detail for the Heating stage in Section 5.1. In the Cooling stage, the objective of the controller design is to avoid sudden pressure drops by regulating air and bleeder valves. The air valve is used to increase or maintain pressure, while the bleeder valve is used to reduce the pressure inside the retort. Avoiding sudden pressure drops is aimed at avoiding food container bursts. On the other hand, the food containers are cooled down to room temperature. This is achieved by flowing water into the retort. In this stage, the water stream is set with a fixed stream. When the retort temperature is reached, the water flow is cut off. To avoid large disturbances at the beginning of the Cooling stage, the steam present in the retort is gradually eliminated. However, the drain valve is kept open.

To select the structure of the Q-matrix for this stage, denoted now Q_C, a similar strategy as in Section 5.1 is used. Since there are two control signals (the Air and Bleeder valves), this Q_C-matrix is designed with three dimensions (one state for each combination of two actions): The matrix represents the space of error in the pressure to the air-valve-action and the bleeder-valve-action.

The control parameters for the Cooling state are shown in Table 2. Even though the same controller gain, k, is used in the design of the air and bleeder action spaces, the gain, can, however be tuned separately in implementation.

6 Results and Discussion

This section discusses the application of the proposed MFLC controller for controlling thermal canned food sterilization in a laboratory plant, placed at the Maritime Research Center, Vigo, Spain. The agent-based MFLC is initialized by training using a virtual plant (simulation). Then, online application is implemented at the laboratory-scale autoclave.

6.1 Plant Description

A schematic of the batch retort unit used for testing the algorithms developed in this paper is presented in Figure 2. The vessel, built in steel, has an approximate weight of 150 kg, and dimensions of approximately 1m of length and 60 cm of diameter. To record the evolution of the relevant variables during processing, three PT100, eight thermocouples and a pressure sensor are located inside the vessel. A computer system is used to gather and analyse real time data. Process Control is carried out using Labview, with an external module WebDAQ that connects the PT100 and pressure sensors to the controller by means of an Ethernet port, and an ADAM that connect the thermocouples. A NiDAQ card is used to actuate the valves, that are Siemens PV90 (DN15)-flat seat, with nominal linear characteristics.

6.2 Initial Training of the Agent

The detailed model of the thermal canned-food process using a retort proposed in Alonso et al. (1997) was used to train the Q_H and Q_C matrices. The model, based on nonlinear dynamic equations, was numerically written and solved in Ecosimpro® simulation language (Ecosimpro, 1999), with training done for various learning stages. The main reasons for using a virtual plant for initial training are the reduction of costs and the prevention of damage to the products during learning for extreme situations. If a simulation were not available, the Q_H and Q_C matrices can be initialized adapting values from similar processes or using values from previous controllers.

The temperature and pressure responses of the first training stage using the Q_H-matrix are shown in Figures 7 and 8. During Heating, the control objective is to maintain a given pre-selected time-temperature profile so as to ensure the appropriate lethality by manipulating the steam valve. Note that the pressure does not need to be controlled during this stage, since the steam is saturated and no air is present in the retort after venting.

After the lethality time t_l is satisfied, the system enters the Cooling stage. In this stage, the temperature is not controlled. In other words, there is no valve regulation rule for controlling the temperature. So that the canned food reaches a cool temperature (approximately ambient temperature), water is passed into the retort at a fixed rate. The water valve is then gradually opened up to 30%. The valve opening in this position is to avoid flooding inside the retort and to provide enough water for cooling. In this Cooling stage, the objective of the controller is switched to control the pressure (see Figure 7b). To avoid sudden pressure drops, the air valve is initially fully open. At the same time, the bleeder valve is totally closed, to avoid losing air inside the retort. Both air and bleeder valves are regulated according to the pressure measured inside the retort. The last pressure reading of the Heating stage is used as an initial pressure set point. From this initial reference, the pressure reference is gradually reduced by 500 Pa if the system is inside the goal state and/or above 10^5 Pa. This value can be changed according to the resistance of the container material. After some training stages, the Q_C-matrix is used in the online implementation.

The agent is also trained for some environment changes, such as changes in the temperature of reference (Figure 9). The learning control is able to track the set point changes and correct the error. Finally, the responses are inside the desired region.

6.3 Application on the laboratory process

The online implementation of MFLC for controlling temperature and pressure of the canned food process is discussed in this section. As mentioned above, the feedback signals are the average temperature in the basket and the average pressure. Temperature responses during the Heating stage are shown in Figure 10a, and the pressures inside the retort are plotted in Figure 10b. The control signal is depicted in Figure 11: only the steam valve position is plotted, as the other valves remain constant. It can be seen that the steam valve works within the range from 0 to 20% opening. Therefore, the control signal is bounded within the desired range. In this application, the steam flow is equipped with a relief valve to reduce the pressure. Hence, the maximum pressure of the steam entering the retort is always about 2 atm.

In summary, adequate temperature control for the Heating process was obtained in the laboratory plant. From the laboratory application, the proposed learning control is able to track the temperature and keep it inside the desired bound (Figure 10 a) during the Heating stage. Also, the controller is able to regulate the system for setpoint changes, while the temperature remains within the desired bounds. The controller output for the setpoint regulation is presented in Figure 11. The controller manipulates the steam valve smoothly, with a control signal suitable for the regulation of the motorized valves.

After a relatively short time (approximately 7 minutes for settling time), the controller can bring the system to be and remain inside the desired bound, with only a small overshoot. The performances of the proposed controller are summarized in Table 3.

7 Conclusions

A procedure for automatic control of the sterilization process in canned food industry has been presented, based on the use of controllers based on learning. More precisely, a controller is proposed to manipulate the steam valve during Heating, using the Model-Free Learning Control (MFLC) strategy, followed by another MFLC controller to regulate the air and drain valves during Cooling. The results of the application of the methodology in a plant at laboratory scale show that the proposed controllers make it possible to maintain the temperature and pressure of the sterilization process within specifications, allowing the safe consumption of the food.

Acknowledgement

This work has been funded by project DPI2007-66718-C04-02. The authors are thankful to Prof. Ernesto Martinez, Dr. Carlos Vilas and Dr. Miriam Rodriguez for many helpful discussions.

References

A. A. Alonso, R. I. Perez-Martin, N. V. Shukla, and P. B. Deshpande, On-line quality control of non-linear batch systems: application to the thermal processing of canned foods, *Journal of Food Engineering*, Vol. 19, pp. 275-289, 1993.

A. A. Alonso, J. R. Banga and R. P. Martin, A Complete Dynamic Model for the Thermal Processing of Bioproducts in Batch Units and its Application to Controller Design, *Chemical Engineering Science*, vol. 52, no. 8, 1997, pp. 1307-1322.

A. A. Alonso, J. R. Banga and R. P. Martin, Modeling and Adaptive Control for Batch Sterilization, *Computers and Chemical Engineering*, vol. 22, no. 3, 1998, pp. 445-458.

C. G. Atkenson, A. W. Moore and S. Schaal, Locally Weighted Learning, *Artificial Intelligence Review*, Vol. 11, pp. 11–73, 1997a

C. G. Atkenson, A. W. Moore and S. Schaal, Locally Weighted Learning for Control, *Artificial Intelligence Review*, Vol. 11, pp. 75–113, 1997b

E. Balsa-Canto, A. A. Alonso, J. R. Banga, A novel, efficient and reliable method for thermal process design and optimization. Part I: theory, *Journal of Food Engineering*, Vol. 52, pp. 227 -234, 2002a.

E. Balsa-Canto, A. A. Alonso, J. R. Banga, A novel, efficient and reliable method for thermal process design and optimization. Part II: Application, *Journal of Food Engineering*, Vol. 52, pp. 235 -247, 2002b.

D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, Massachusetts, 1996.

S. Bruin, Th. R. G. Jongen, Food Process Engineering: the last 25 years and challenges ahead, *Comprehensive Reviews in Food Science and Food Safety*, Vol. 2, 2003, pp. 42-81.

Z. S. Chalabi, L. G. van Willigenburg and G. van Straten, Robust Optimal Receding Horizon Control of the Thermal Sterilization of Canned Foods, *Journal of Food Engineering*, vol. 40, 1999, pp. 207-218.

EA Int (1999), *EcosimPro User Manual*, Available from: www.ecosimpro.com (accessed 17 June 2010).

Damien Ernst, *Near optimal closed-loop control. Application to electric power systems*, PhD thesis at University of Liège, Belgium, 2003.

D. Holdsworth and R. Simpson, *Thermal Processing of Packaged Foods*, Springer Verlag, 2007.

D. Kleis and E. W. Sachs, Optimal Control of the Sterilization of Prepackaged Food, *SIAM Journal on Optimization*, vol. 10, no. 4, 1999, pp. 1180 - 1195.

M. A. Kumar, M. N. Ramesh and S. Nagaraja Rao, Retrofitting of a Vertical Retort for On-line Control of the Sterilization Process, *Journal of Food Engineering*, vol. 47, 2001, pp. 89 - 96.

M. J. Lewis, Thermal Processing, in J.G. Brennan, *Food Processing Handbook*, Wiley, 2006, pp. 33-70.

S. J. Mulvaney, S. S. H. Rizvi, and C. R. Johnson, Dynamic modelling and computer control of a retort for thermal processing, *Journal of Food Engineering*, Vol. 11, pp. 273 - 289, 1990

H. S. Ramaswamy and R. P. Singh, Sterilization Process Engineering, in K. J. Valentas, E. Rotstein, R. P. Singh, *Handbook of Food Engineering Practice*, CRC Press, New York, 1997.

Martin Riedmiller, Learning Control for Continuous MIMO Dynamical Systems Using Neuro Dynamic Programming, in proceeding of *Third European Workshop on Reinforcement Learning*, Rennes, France, October 13-14, 1997.

William Donald Smart, *Making Reinforcement Learning Work on Real Robots*, PhD thesis at Brown University, 2002.

O. von Stryk and R. Bulirsch, Direct and Indirect Methods for Trajectory Optimization, Annals of Operations Research, vol. 37, 1992, pp. 357-373.

R. S. Sutton and A. G. Barto, *Reinforcement Learning: an Introduction*, The MIT Press, Cambridge, MA, 1998.

S. Syafiie, F. Tadeo and E. Martinez, Learning to Control pH Processes at Multiple Time Scales: Performance Assessment in a Laboratory Plant, *Chemical Product and Process Modeling*, vol. 2, no. 1, 2007a, article no 7.

S. Syafiie, F. Tadeo and E. Martinez, Model-Free Learning Control of Neutralization Process Using Reinforcement Learning, *Engineering Applications of Artificial Intelligence*, Vol. 20, pp. 767 – 782, 2007b.

S. Syafiie, Garcia M., Vilas C., Alonso A., Martinez E., Tadeo F., Intelligent Control Based on Reinforcement Learning of Batch Thermal Steriliza- tion of Canned Foods, in *Proceedings of the 17th IFAC World Congress*, Seoul, South Korea, July 2008.

A. A. Teixeira and G. S. Tucker, On-line Retorts Control in Thermal Sterilization of Canned Foods, *Food Control*, vol. 8, no. 1, 1997, pp. 13 - 20.

S. Timmer and M. Riedmiller, Fitted Q Iteration with CMACs, in *Proceedings of the International Symposium on Approximate Dynamic Programming and Reinforcement Learning* (ADPRL), Honolulu, USA, April 2007.

Fig. 1. Agent-environment interaction

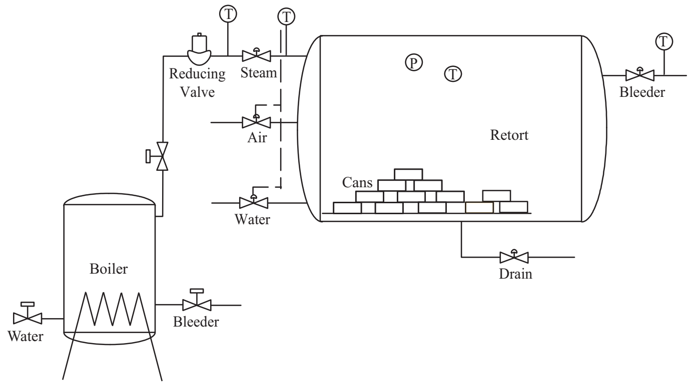


Fig. 2. Schematic of batch sterilization for controller design

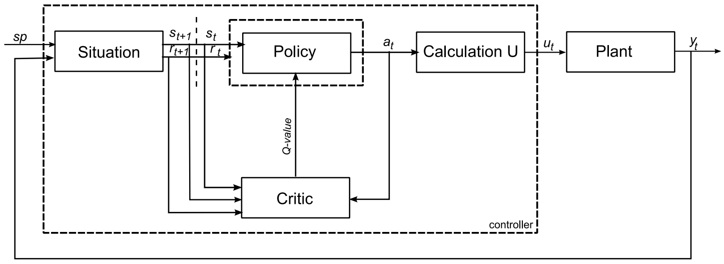


Fig. 3. MFLC architecture

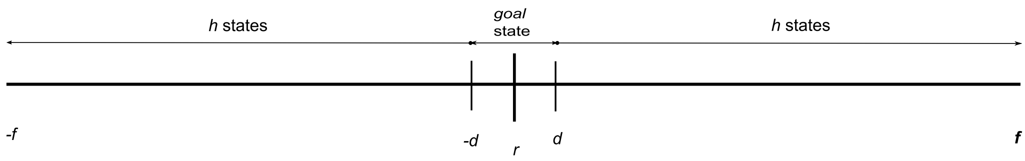


Fig. 4. Definition of the symbolic states in MFLC

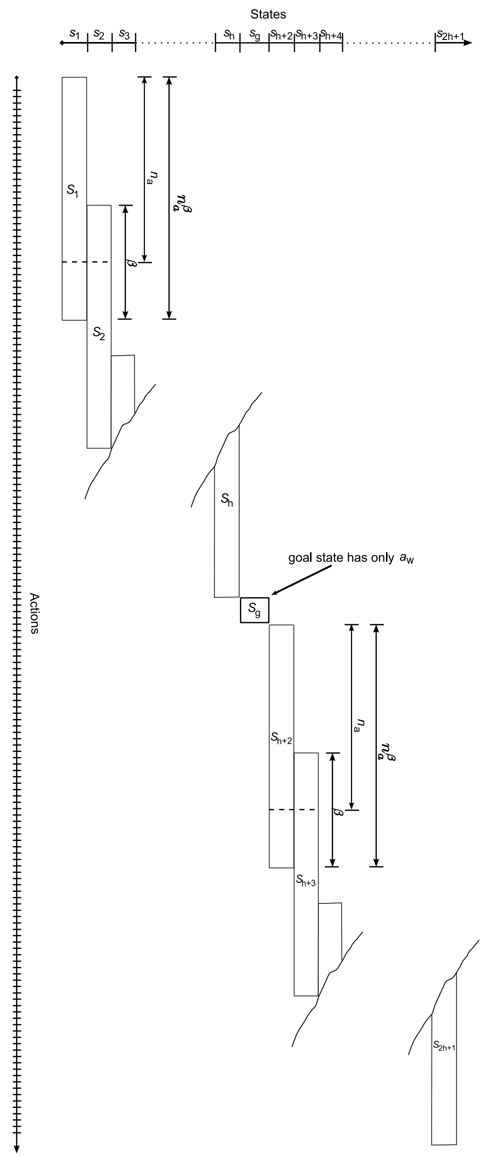


Fig. 5. State-Action space of Q-matrix

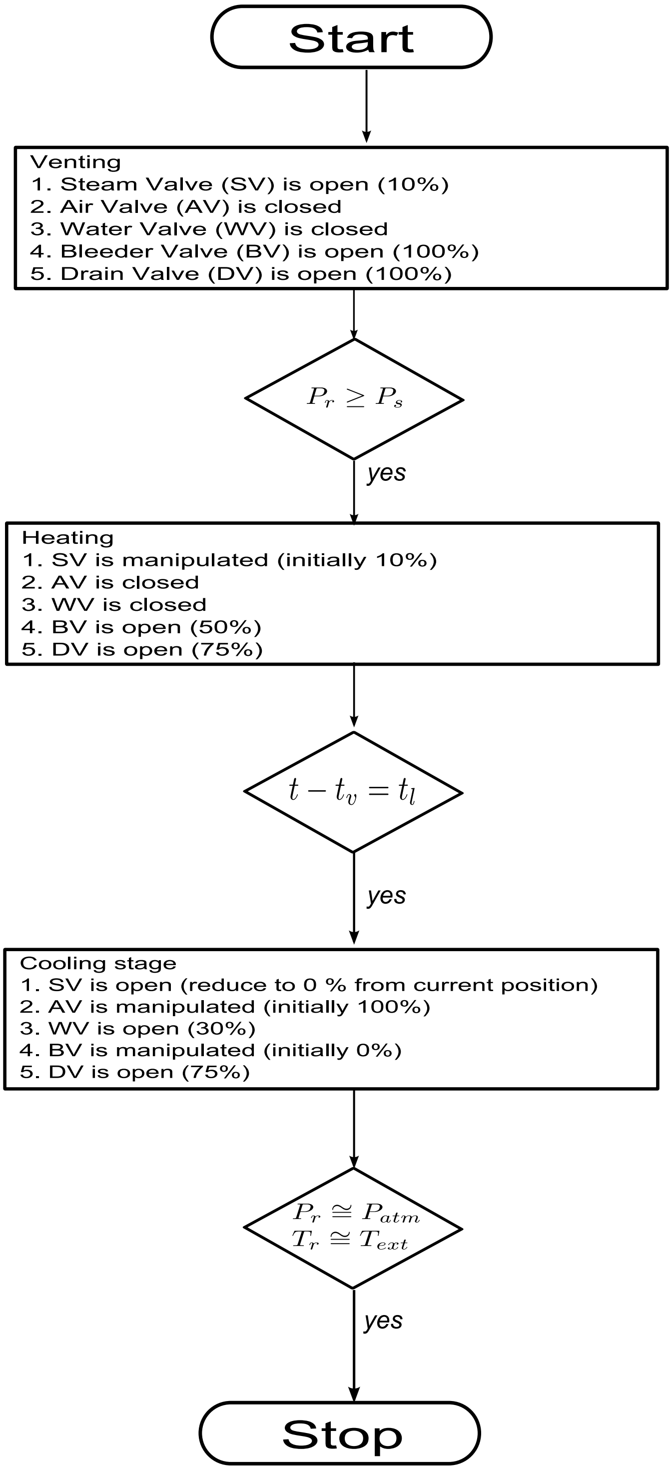
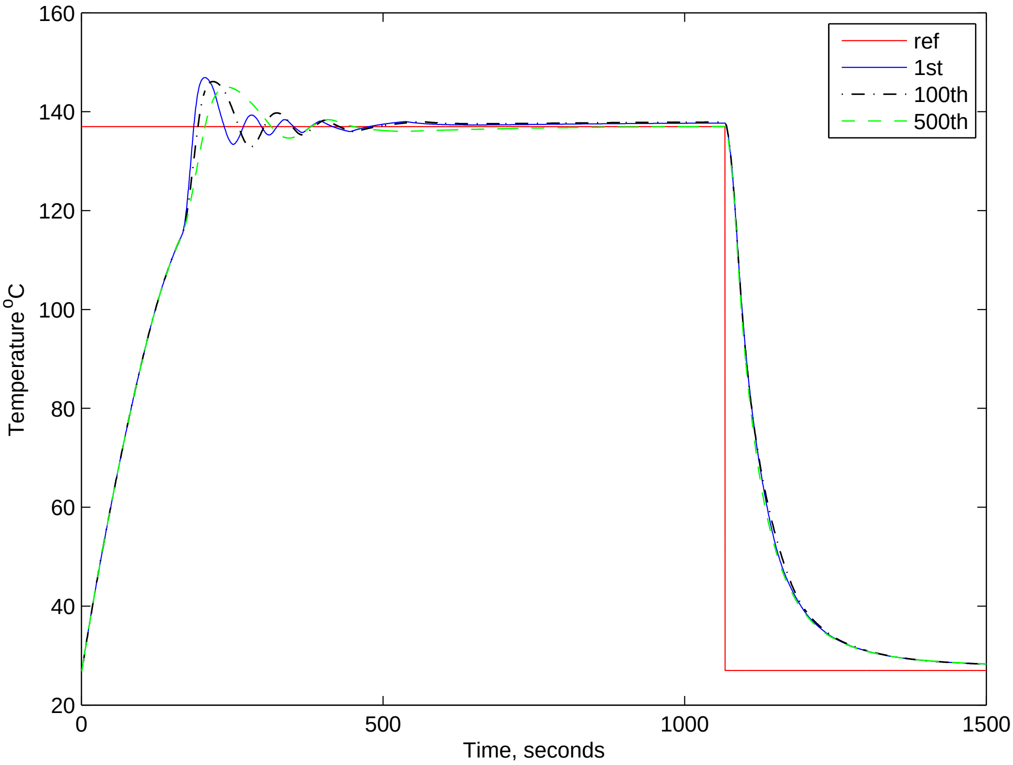
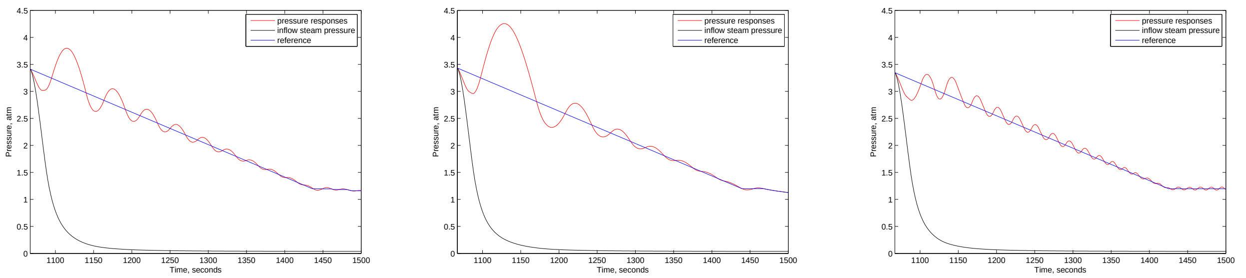


Fig. 6. Control logic implementation: P_r, P_s and P_{atm} are retort, steam and external pressures, t_v is the starting time of Heating, t_l is lethality time, T_r and T_{ext} are retort and ambient temperatures.

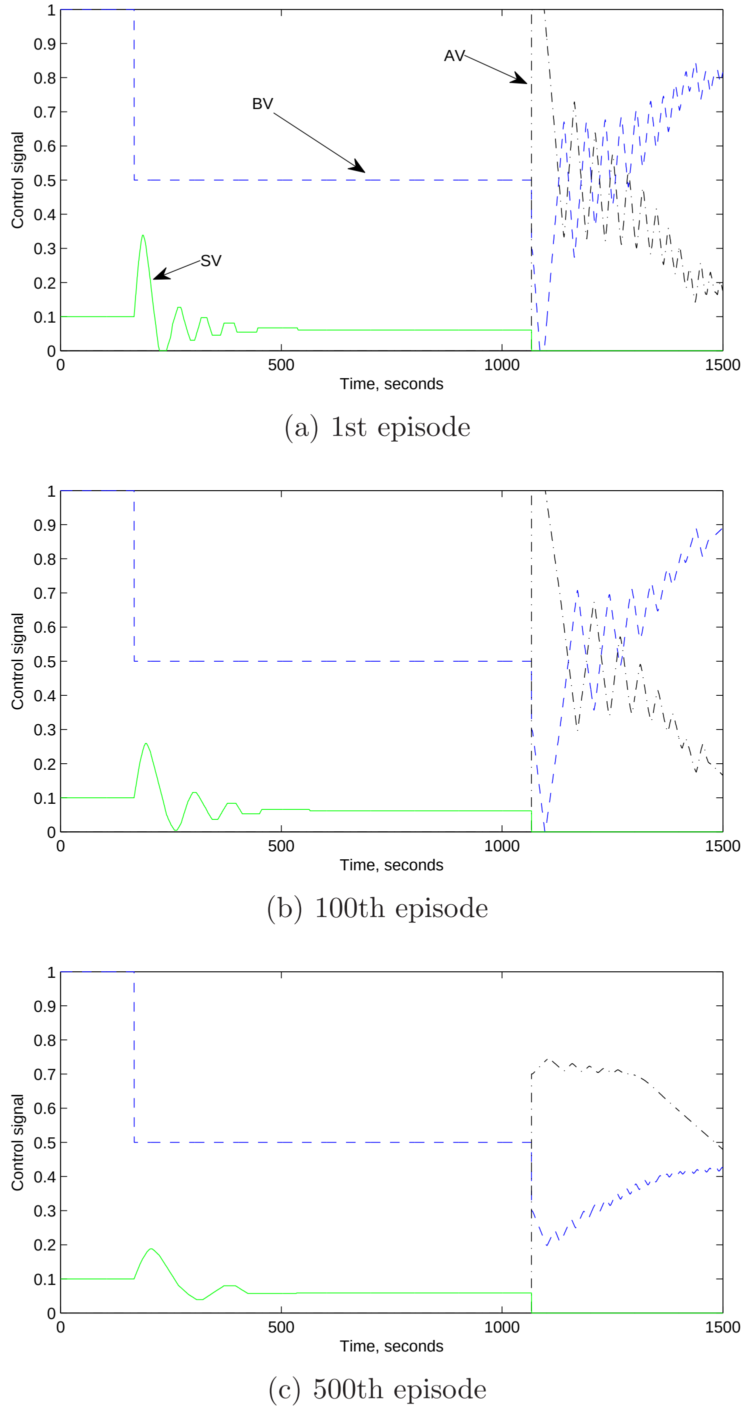


(a) Evolution of the temperature at episodes 1, 100 and 500 (Heating from 200s to 1050s; Cooling from 1050s)



(b) Detail of evolution of pressure during Cooling at episodes 1 (left), 100 (center) and 500 (right)

Fig. 7. Evolution of temperature and pressure during learning on the virtual plant



(a) 1st episode

(b) 100th episode

(c) 500th episode

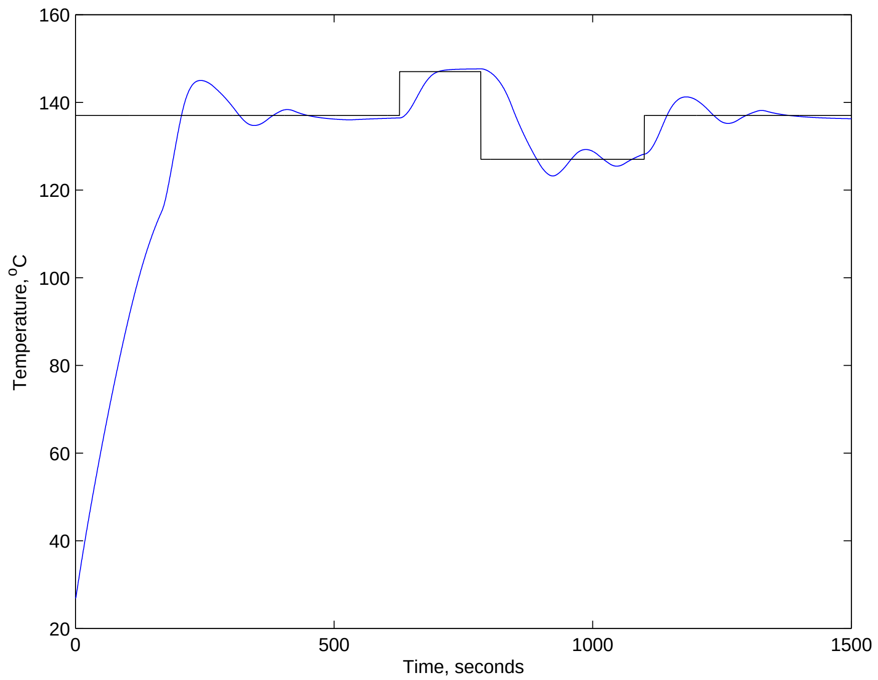
Fig. 8. Control signals during learning on the virtual plant (Heating from 200s to 1050s; Cooling from 1050s)

Table 1
Heating Control Parameters

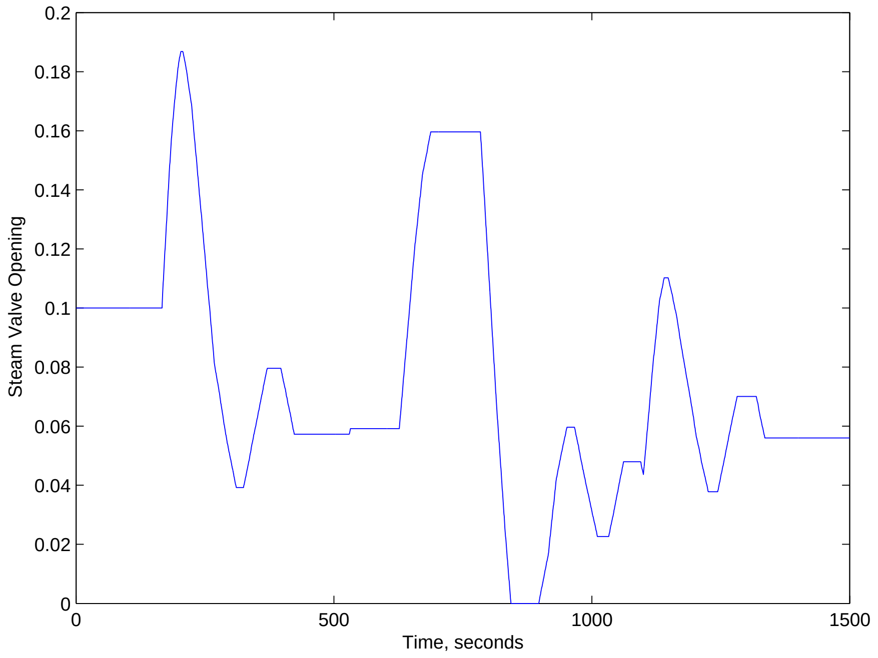
parameters	value	units
learning rate, α	0.1	-
forgetting factor, γ	0.98	-
number of states, $2h+1$	21	-
span of goal state, d	2	^{o}C
limited error exploration, h	29	^{o}C
overlapping degree, β	5	-
wait action, a_w	801	-
controller gain, k	1×10^{-5}	-
upper limit, $\overline{\Delta u}$	0.008	kg/s
lower limit, $\underline{\Delta u}$	0.0001	kg/s

Table 2
Cooling Control Parameters

parameters	value	units
learning rate, α	0.1	-
forgetting factor, γ	0.98	-
number of state, $2h+1$	21	-
span of goal state, d	100	Pa
limited error exploration, h	1×10^4	Pa
overlapping degree, β	10	-
wait action, a_w	601	-
controller gain, k	1×10^{-5}	-
upper limit, $\overline{\Delta u}$	0.006	kg/s
lower limit, $\underline{\Delta u}$	0.0001	kg/s

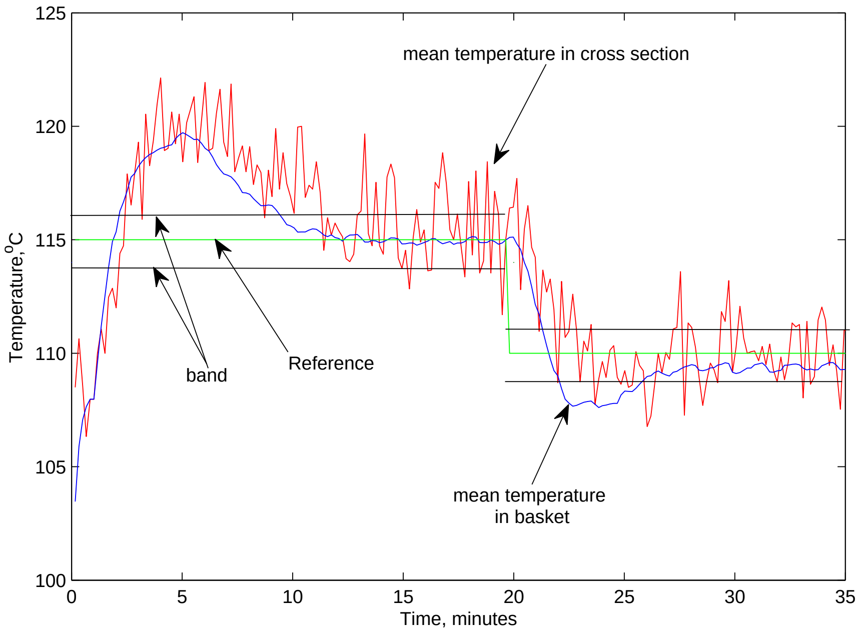


(a) Temperature

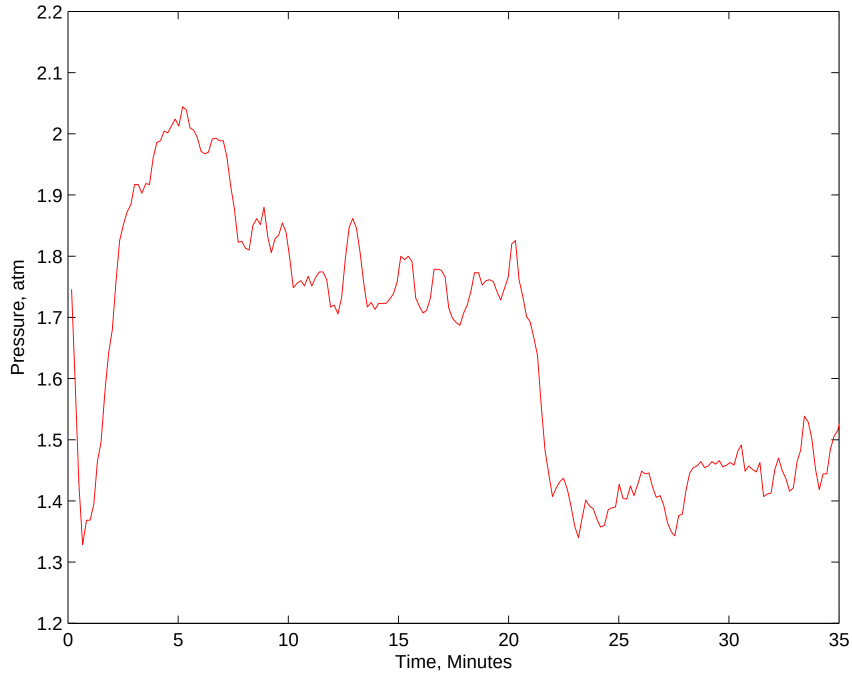


(b) Control signal

Fig. 9. Temperature responses under changes in the temperature setpoint during Heating



(a) Temperature



(b) Pressure

Fig. 10. Temperature and pressure measured in the laboratory plant during Heating, using the proposed control strategy

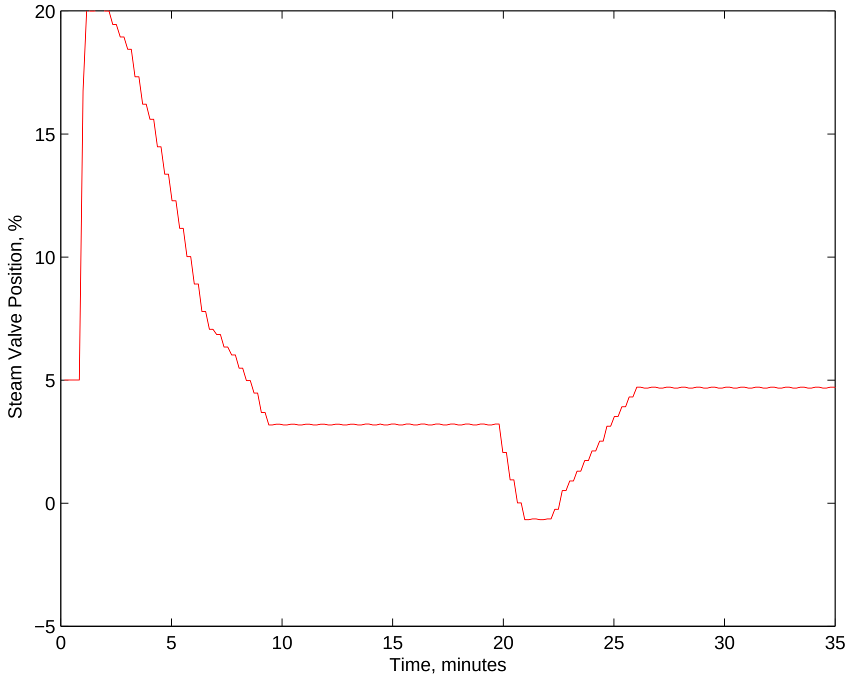


Fig. 11. Steam valve signal calculated by the controller for the experiment in Fig. 10

Table 3
Control Performances

index	parameters
Time-to-target	2 minutes
Settling time	7 minutes
Maximum overshoot	3 ^{o}C