JIChE06
The Sixth Jordan International Chemical Engineering Conference
12-14 March 2012, Amman, Jordan

APPLICATION OF REINFORCEMENT LEARNING TO BATCH DISTILLATION

M. A. Mustafa[1] and J. A. Wilson[2]
[1]Department of Chemical Engineering,
Faculty of Engineering, University of Khartoum, Sudan
[2]Chemical and Environmental Engineering,
Faculty of Engineering, The University of Nottingham, United Kingdom

ABSTRACT
An important amount of work exists on the topic of optimal operation and control of batch distillation though it is still based on the assumption of an accurate process model being available. While this assumption is valid from a theoretical point of view, there will always remain the challenge of practical applications. Reinforcement Learning (RL) has been recognised already as a particularly suitable framework for optimizing batch process operation however no application to batch distillation has been reported. Thus, this paper presents RL as an automatic learning approach to batch distillation. The methodology is exemplified using various case studies.

INTRODUCTION
Distillation is one of the most widely used unit operations in the fine chemical, petroleum and pharmaceutical industries. It is one of the oldest methods of separation of liquid mixtures into their various components depending on differences in boiling points of liquids and relative volatility.

The rising importance of high-value-added, low-volume specialty chemicals has resulted in a renewed interest in batch processing technologies (Diewkar, 1995) and the drive for optimum operation is ever present. Batch distillation is an important and widely used separation process in batch process industry. Its main advantage over continuous operation is the ability to be used as a multi-purpose operation for separating mixtures into their pure components using a single column. Batch distillation can also handle a wide range of feed compositions with varying degrees of difficulty of separation (e.g. wide ranges of relative volatilities and product purities). Although the typical consumption of energy is more than in continuous distillation, more flexibility is provided with less capital investment (Luyben, 1992). However, besides the flexibility in the operation of batch distillation columns, a range of challenging design and operational problems occur due to its inherent unsteady state nature.

LITERATURE SURVEY
The main sequence of events in operating a batch distillation column starts with the feed charged into the reboiler. The column is then operated at total reflux until the column reaches steady state. This initial phase is known as the start-up phase. In the second phase, or production phase, light component product is collected into a product tank until its average composition drops below a certain specified value. This cut is referred to as the main cut (The 1st main cut is sometimes preceded by taking off the low boiling impurities at a high reflux ratio). After that, the first intermediate distillate fraction (off-cut or slop cut) is produced and stored in a different tank. This procedure is repeated with a second main cut and second slop cut and so on until the concentration of the heaviest component, in the reboiler of the column, reaches a specified value. At the end of the batch, the operation of the distillation column goes through a shutdown phase.

Slop cuts contain the material distilled, which does not meet specification. Considerable work in slop handling strategies has been reported in the literature ((Bonny et. al., 1994) and (Mujtaba and Macchietto, 1992)).

On the other hand, a totally different operating policy is the cyclic operation of a batch distillation column. In the case of a regular column, the cyclic operation could be characterised by repeating a three period operation (Sorensen, 1997): Filling, Total Reflux, and Dumping.

The main manipulated variable, in the process of controlling a batch distillation column, is the reflux ratio. The frequently used and conventional approach towards controlling the operation of a batch distillation column, during the production of main cuts, is either to operate at constant reflux ratio or to operate at a varying reflux ratio (constant distillate composition). During operation at constant reflux ratio, the distillate composition is allowed to vary resulting in a simpler strategy and hence it is more commonly used in industry. The second approach is conducted by maintaining a fixed overhead composition while varying the reflux ratio. The two approaches used are simple but provide sub-optimal results.

The second manipulated variable, in controlling a batch distillation column, is the boil-up rate: the quantity of liquid in the reboiler that is evaporated per unit time. In case of a batch distillation column, the boilup rate is often held at a maximum rate consistent with allowable vapour velocities and liquid capacities. In addition to the variables just mentioned, Farhat et al. (1990) used the switching time for different cuts as an extra decision variable.

Throughout the literature, the formulation of the optimal control problem in batch distillation has been categorised as either a: Maximum Distillate Problem (Converse and Gross (1963), Keith and Brunet (1971) and Diwekar et. al. (1987)); Minimum Time Problem (Coward (1967), Mayur and Jackson (1971), Egly et. al. (1979), Hansen et. al. (1986) and Mujtaba and Macchieto (1998)); Maximum Profit Problem (Kerkhof and Vissers (1978) and Logsdon et. al. (1990)).

Mujtaba and Macchietto (1997) provided an efficient framework for on-line optimization of batch distillation with chemical reaction. The technique starts by finding optimization solutions to the batch distillation with chemical reaction problem, in order to solve the maximum conversion problem. The optimization was performed for a fixed batch time and given product purity. The maximum conversion, the corresponding amount of product, optimal constant reflux ratio, and heat load profiles were plotted for different batch times. Polynomial curve fittings were then applied to the results of the optimization and were used to formulate a non-linear algebraic maximum profit problem.

Mujtaba and Hussain (1998) developed an optimization framework to tackle efficiently the optimal operation of dynamic process due to process/model mismatches. The method was applied to a batch distillation process where use is made of a neural network to predict the process/model mismatch profiles for the case study used. The Neural Network was then trained to predict the process/model mismatch, for each state variable, at the present discrete time. The mismatch then between the actual process/model (represented by error between rigorous model and simple model) and that predicted by the network was used as the error signal to train the Neural Network. The simple model was then used together with the Neural Network, to calculate the optimal reflux ratio to achieve the separation in minimum time. The results were then compared with the more rigorous model, which was used to represent the actual process in their case study. It was concluded that with the use of a simple model with mismatches, the optimal operation policy could be predicted quite accurately using the Neural Network. Although the important work by Mujtaba et. al. (1997, 1998) reduce drastically the computational time used to solve differential equations, address process/model mismatch issues and optimal operation issues in general, however exact knowledge of a mathematical process model is still assumed.

One of the first applications of Artificial Intelligence as the central part of batch distillation automation was by Cressy et. al. (1993). They made use of neural networks in order to learn the control profiles of a batch distillation with a binary mixture: methanol and water. Two Neural Networks were used in the methodology: Neural Emulator (used to approximate the input/output function defined by the forward dynamics of the column) and a Neural Controller. The trained Neural Network achieved an error of less than 3% over a narrow range of conditions. Over a wider range, the results were not uniformly good. Furthermore, the amount of training data of 4080 training patterns would justify such a good fit to the observed data. The immediate concern is the issue of acquiring such an amount of data in practice.

Stenz and Kuhn (1995) managed to integrate operator's knowledge, using fuzzy technology, into the automation of the batch distillation process. They concluded that fuzzy logic is not a superior method, but is rather an addition to the toolbox of the automation engineer, which is potentially useful. Although fuzzy logic presents the operator’s know how as a sequence of acting steps, it still does not aim at giving the optimum solution.

Wilson and Martinez (1997) proposed a novel approach towards batch process automation involving simultaneous reaction and distillation. The methodology proposed combined fuzzy modelling and RL. The RL part of the methodology meant that the controller implemented is geared towards incrementally achieving goals, using rewards obtained as guideline. However, a large amount of data (1000 randomly chosen batches) is still required for learning, which is well beyond the small number of initial batch runs that would be practically available in industry.

Further important work to determine efficient time profiles still depends upon having an accurate process model ((Barolo and Cengio, 2001), (Kim, 1999), (Lopes and Song, 2010) and (Pommier et. al., 2008)). In practice such models are never available partly because conditions and parameters vary from one batch to another. Furthermore, the classical open loop time profile can not react to measurements during the progress of a batch. The industry is faced with composition analyzers which are again often not available and seldom instantaneous (Luyben, 1992). Despite all those problems human operators have managed so far to incrementally drive those processes to near optimal operation. Thus it is the aim of this work to provide a software tool to mimic the operator’s interactive learning approach.

METHODOLOGY

If an analysis of our learning during childhood is made, we find that (for example) we learn to walk without the help of an explicit teacher. Also learning how to talk or even how to behave in society when we are growing up. We tend to learn according to trial and error interaction with our environment and then go on reinforcing those actions we took and resulted in better situations. Following this natural process provides us with wealth of knowledge and information about cause and effect, the results of different actions and hence what to do to achieve certain goals.

Reinforcement Learning (RL) algorithms could be seen as a way of providing a computational approach focused on goal-directed learning and decision making from interaction. Following the book on the subject by Sutton and Barto (1998), one could define RL as simply being the mapping of situations to actions so as to maximize a numerical reward. An important point to add is that during learning, the algorithm is not told which actions to take but must explore and exploit to discover actions that yield the most reward by trying those actions. The RL algorithm tends to learn an optimum control policy by gathering data from a series of batch runs. The particular suitability of RL as a framework for optimizing batch process operation has been recognized already (Martinez et. al., 1998a,b,c)). The proposed hybrid predictive model (which form part of the RL algorithm) delivered adequate performance in previous applications to batch reactors, however there are no applications to batch distillation.

Batch distillation problems fit nicely with a typical Reinforcement Learning problem, characterized by setting of explicit goals, breaking of problem into decision steps, interaction with environment, sense of uncertainty, sense of cause and effect. The main elements of RL comprise of an agent (e.g. operator, software) and an environment (Sutton and Barto (1998)). The agent is simply the controller, which interacts with the environment by selecting certain actions. The environment then responds to those actions and presents new situations to the agent. The agent's decisions are based on signals from the environment, called the environment's state. Fig. 1 shows the main framework of RL.

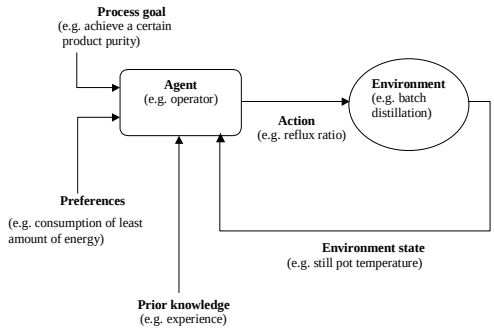

Fig.1: Main framework of Reinforcement Learning

The developed RL approach is composed of a combination of integrated techniques such as Neural Networks (Carling, 1992), Dynamic Programming (Bellman, 1957) and Wire Fitting (Baird and Klopf, 1993). Furthermore, predictive models are used to mimic the forward dynamics of the process. The 'value function' which represents the objective function in the optimization problem defined as follows:

$$Q(s_t,a_t)=\begin{cases}\leftarrow +PI, \text{ if } a_t \text{ is a final action and the goal has been achieved,}\\ \leftarrow -1., \text{ if } a_t \text{ is a final action and the goal has not been achieved,}\\ \leftarrow \max\limits_{a_{t+1}\in\Omega} Q(s_{t+1},a_{t+1}), \text{ otherwise.}\end{cases} \quad (1)$$

where PI is the Performance Index (a function of the final conditions at time T). Penalty of -1 is nominal value and it may be appropriate to use other values in particular problems.

The RL algorithm then aims to optimize the operation of the process through the following control law:

$$a^* = \arg\left(\max_{a\in\Omega} Q(s,a)\right) \quad (2)$$

where Ω represents the set of feasible control actions.

The RL approach could be seen (i.e. with reference to Wire Fitting approximations) as a means of learning to identify the optimal wire, or wires for the different states. This is achieved through learning the weights and biases in the Neural Network. The change in weights (Δ weights) is calculated as follows:

$$\Delta \text{ weights} = -\alpha\left(\frac{\delta \text{ Mean squared Bellman Error}}{\delta \text{ weights}}\right) \quad (3)$$

where α is referred to as the learning rate.
At the end, the RL algorithm converges to the actual optimal value function when Eq. 4 is true .

$$Q^*(s_t,a_t^*)=\begin{cases}\leftarrow PI^*, \text{ if } a_t \text{ is a final action}\\ \leftarrow \max\limits_{a_{t+1}\in\Omega} Q^*(s_{t+1},a_{t+1}), \text{otherwise}\end{cases} \quad (4)$$

During incremental learning of the optimal value function, differences occur which define the error: Bellman error. The mean squared Bellman error (Bellman, 1957), E_B, is then used in the approach to drive the learning process to the true optimal value function (Eq. 5 defines E_B for a given state-action pair (s_t,a_t)).

$$E_B=\begin{cases}\leftarrow \frac{1}{2}E\left[\left\{PI^*-Q^*(s_t,a_t^*)\right\}^2\right], \text{ if } a_t \text{ is a final action.}\\ \leftarrow \frac{1}{2}E\left[\left\{\max\limits_{a_{t+1}\in\Omega} Q^*(s_{t+1},a_{t+1})-Q^*(s_t,a_t^*)\right\}^2\right], \text{otherwise.}\end{cases} \quad (5)$$

A detailed description of the RL algorithm and Matlab code is provided by Mustafa (2001).

CASE STUDY

The RL approach developed involves only the most generic form of a priori knowledge in relation to the physical properties or distillation characteristics of the feed mixture. Starting with data from a small set of experimental batch runs the approach 'learns', through batch-to-batch incremental improvement, how to operate the plant in a near optimal fashion using a crude generic model of the behaviour of batch distillation systems.

The RL technique is applied to a batch distillation case study which involves a 10-tray batch distillation column with a binary mixture having a relative volatility of 2.5. Simulations of the batch distillation column were conducted using Smoker's equation for a binary mixture. The still is initially charged with a feed of 1 kmol containing 0.7 mole fraction of the more volatile component. The specification for the product purity was set at 0.98 mole fraction.

The strategy for operating and simulating the batch distillation column was then as follows:

1. Three periods of operation each at a fixed reflux ratio (i.e. three decision steps as shown in Fig. 2).
2. Still temperature measured and used to decide on change to reflux ratio when still pot contents lie at 1.0, 0.68 and 0.48 kmol (those values were selected following an analysis of optimal operation of case study).
3. Each batch is terminated when still pot contents falls to 0.35 kmol.
4. Constant vapour boilup rate of 0.2 kmol/h.
5. The target for the RL algorithm is then set to achieve the goal of obtaining a product purity of 0.98 mole fraction. In addition, the preference is given to meeting the goal in the minimum amount of time so as to achieve the maximum profit. The Performance Index (PI) is defined as follows:

$$PI = D.\ P_r - V\ .\ BxTime\ .C_s \qquad (6)$$

where D is the amount of product distilled (kmol), P_r is the sales value of product (£/kmol), V is the vapour boilup rate (kmol/h), BxTime is the time for completion of batch and C_S is the heating cost £/kmol.

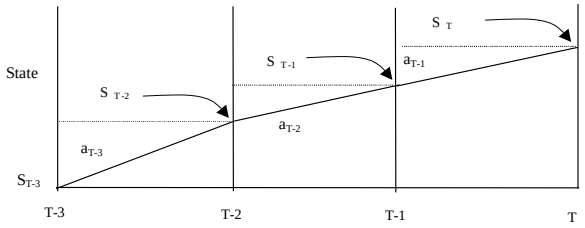

Fig. 2: Three-decision step case study

The process starts at state S_{T-3}, corresponding to the initial state and terminates at state S_T (at time interval T). During different time intervals (T-3, T-2 and T-1), samples of the state of the process are taken, and accordingly 3 actions are chosen (a_{T-3}, a_{T-2}, and a_{T-1}). States are the bubble point temperature except for the final state where it represents the product purity. As for the actions, they are the reflux ratios.

Three additional case studies, with the same feed and product specifications as in base case, were added with feed mixtures having different relative volatilities and with different numbers of column trays (Table 1).

Table 1: Description of various case studies

Case Study	No of trays	Relative volatility
Base	10	2.5
1	10	3
2	12	2
3	17	1.5

A comparison between the different case studies is possible through the measure defined by Kerkhof and Vissers (1978), σ_{diff}, which indicates the degree of difficulty of separation:

$$\sigma_{diff} = \frac{x_{D,preset} - x_F}{x_F\left[1 - x_{D,preset}\right]\left[\rho^{N+1} - 1\right]} 100\% \qquad (7)$$

where $x_{D,preset}$ is the pre-set product purity (mole fraction), x_F is the feed purity (mole fraction), ρ is the relative volatility and N is the number of theoretical plates in the column. They further categories the results into the following:

Easy separation	σ_{diff} <1%
Moderate separation	1%< σ_{diff} <10%
Difficult separation	σ_{diff} >10%
Very difficult separation	σ_{diff} >15%

Hence, according to the above criteria, base case (σ_{diff} = 0.08%), Case study 1 (σ_{diff} = 0.01%), Case study 2 (σ_{diff} = 0.24%) and Case Study 3 (σ_{diff} = 1.35%) represent easy to moderate degrees of difficulty of separation.

The general structure of the predictive models for the various stages is provided by Eq. 8 -10. The predictive models are as follows:

$$s_T = f\ (s_{T-1}, a_{T-1}) \qquad (8)$$

for the last decision stage at T-1

$$s_{T-1} = f\ (s_{T-2}, a_{T-2}) \qquad (9)$$

for the intermediate decision stage at T-2

where s_t (state at time t) denotes the bubble point temperature of the mixture in the still pot (representing the composition of the mixture), with the exception of the last decision stage T-1 where it represents the final product purity (mole fraction), and a_t (action taken at time t) denotes the reflux ratio.

For the initial stage there is again a slight difference in the predictive model, since all batches were assumed to start from the same initial point. This would mean that the predictive model would have no dependency on the initial state, and hence the state at T-2 (still pot temperature at T-2) becomes only a function of the action at T-3 (reflux ratio at time T-3).

$$s_{T-2} = f\ (a_{T-3}) \tag{10}$$

RESULTS AND DISCUSSION

Starting with an initial training data set of six batch runs, the RL algorithm with an embedded hybrid predictive model was applied using Matlab. Results demonstrating the RL based optimization are presented involving a stepped reflux ratio policy to meet a minimum overhead purity specification where the performance criteria used are product 'give away' and the number of 'off spec' batches produced during an initial production campaign of 50 batches. Give-away is a common term in industry and is used when dealing with problems where a hard constraint has to be met and could not be violated. For example the goal in case study is to meet a product purity of 0.98 mole fraction. If the batch distillation is controlled in practice along that value of product purity, the controller is bound to produce off-spec batch runs some of the time. Hence in industry, they are willing to give away a slightly more pure product on average, so as to reduce the risk of loosing money through production of off-spec batches. Hence, the term give-away in this context refers to the amount of average product purity that one could give-away above the fixed product specification. Concerning the analysis in the following sections, the product specification is set throughout at 0.98 mole fraction. Give-away values of 0.005, 0.01, 0.015 and 0.02 are used to reflect how all batches produced to a product purity of 0.975, 0.97, 0.965 and 0.96 mole fraction respectively are accepted as being on-spec.

Unfortunately, an average of 75% of off-spec additional batches (i.e. which did not meet the goal) were produced. Whilst linear predictive models delivered adequate performance in previous applications of this approach to batch reactors, in batch distillation the strong non-linearity present, especially with high purity products, leads to difficulties (slow convergence) even with binary feeds. It was not clear at this stage where the problem actually occurred. Thus, the embedded predictive models in the RL algorithm were replaced with the actual mathematical model of the process (i.e. no process/model mismatch) and the RL algorithm was rerun. This way any problems in the RL algorithm itself could be detected and further addressed. The RL algorithm converged directly to the optimal solution starting from an initial training data set of only two batch runs. The convergence of the RL algorithm here proves the effectiveness of the methodology as an optimization method, but also reveals the central aspect about applications of RL: The importance of the Predictive Models. The objective of subsequent work was thus to identify suitable general form of predictive model.

Following the unsuccessful implementation of the hybrid predictive model in RL applications to base case, different predictive model forms were used so as to observe the performance of the RL algorithm. Predictive models in the form of a linear function, a second order polynomial and a Neural Network (using one node in the hidden layer) were used in place of the generalised hybrid predictive model proposed by Martinez (1998a). Table 2 shows a description of the different predictive model forms used.

Table 2: General description of various predictive model forms used in base case

Predictive model form	*Description*
Linear model	$s_{t+1} = m\ s_t + n\ a_t + p$
2nd order polynomial model	$s_{t+1} = m\ s_t + n\ a_t + j\ (s_t)^2 + k\ (a_t)^2 + l\ (s_t\ a_t) + p$
Neural Network	Number of inputs: 2 Number of nodes in hidden layer: 1 Number of output nodes: 1 Number of free parameters: 5 Activation function: Tansigmoidal function
Hybrid predictive model	Weighted predictions of a combination of linear models and extrapolated values using "slopes". The "slopes" represent the sensitivity of a one-step ahead prediction of states (using finite difference) towards slight changes in initial input conditions

Starting from the same initial training data set, the RL algorithm was rerun using the different predictive models. For each different predictive model, six batch runs of training data were followed by 21 batches used for control testing. This was repeated five times for each different predictive model used, starting from different initial weights in the Neural Network. The results of using the different predictive models are shown in Fig 3, where the worst performance (the most off-spec batches produced) is found using a Neural Network predictive model. This was probably due to the small training data set of six initial batch runs, since the Neural Network had enough free parameters (five weights and biases) to fit the six data points observed. On the other hand, the second order predictive model performed always better than the linear predictive model. Although there is a slight improvement with the 2nd order polynomial model,

these results are well beyond what could be considered acceptable in terms of off-spec batches produced. The use of higher order polynomial model forms was not pursued in this study, since larger amounts of training data would be required to fit the additional model parameters (current initial training data set consists of six batch runs only).

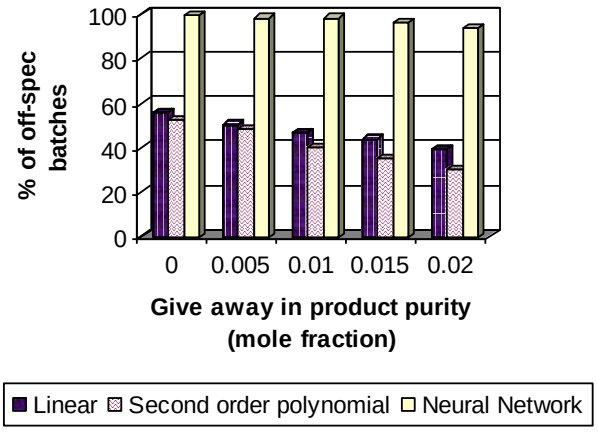

Fig.3: % of off-spec batches versus giveaway in product purity for linear, second order polynomial and Neural Network predictive models

Development of Predictive Models

The focus was shifted towards trying, through analysis of simulated batch runs, to identify a fingerprint that is particular to batch distillation. Hence, through observing the behaviour (i.e. behaviour of the system separate from RL) of a distillation column for a series of different scenarios, regressive model forms could be chosen to capture relationships between variables of interest. Starting from different initial states (still pot temperatures), and applying a range of actions (reflux ratio's), the resulting states (still pot temperatures or product purity for last but one stage) were calculated using base case and plotted as shown in Fig. 4 and 5.

Fig. 4 and 5 show the trends followed in the last but one stage. Fig. 4 shows a set of curves that represent the final product purity as a function of the reflux ratio. Each different curve in Fig. 4 specifies a fixed still pot temperature at T-1. It is clear how the goal (final product purity of 0.98 mole fraction) is only achieved if starting from higher temperatures (higher curves in Fig. 4) in the still pot at T-1 otherwise the goal is never achieved. Fig. 5 shows the trends followed for the final product purity as a function of different still pot temperatures at T-1 at constant reflux ratio. It is clear from Fig. 5 that the relationship is linear and that the lines are approximately parallel to each other. Similar trends to those shown in Fig. 4 are observed for stage T-2 and T-3 with lines curving initially and then gradually reaching asymptotic values. This is also true for Fig. 5 where linear relationships for previous stages prevail.

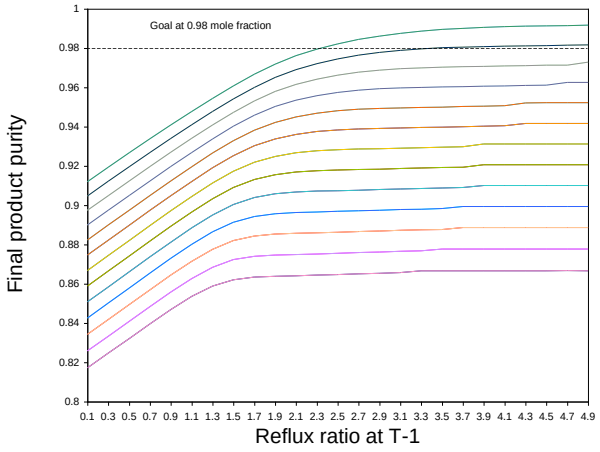

Fig. 4: Final product purity as a function of reflux ratio at T-1 (lines of constant still pot temperature at T-1 with temperatures increasing from bottom to top) for the distillation column in base case

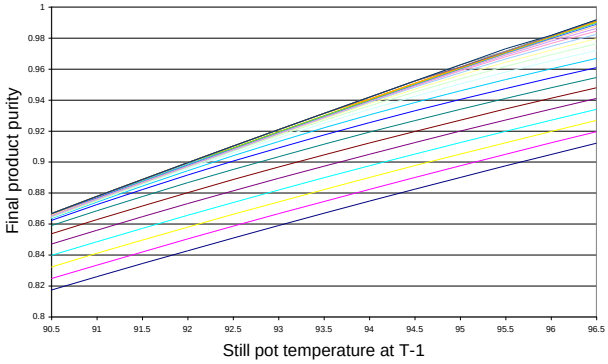

Fig 5: Final product purity as a function of still pot temperature at T-1 (lines of constant reflux ratio with values increasing from bottom to top) for the distillation column in base case

Following the analysis of the behaviour of the batch distillation system (i.e. separate from RL), a predictive model could be constructed as follows:

1. For fixed reflux ratio a linear model could approximate the one step ahead prediction of temperatures (or purity for last decision stage) at t+1 as a function of temperature at t as follows (assuming constant reflux ratio):
2. $$s_{t+1} = m\, s_t + p \qquad (11)$$
   where m and p are free parameters.
3. Concerning the relationship of state at t+1 as a function of reflux ratio, for fixed temperature at t, the following model form is proposed to produce a curve which would gradually reach an asymptotic value for increasing values in reflux ratio (assuming constant state at t):
   $$s_{t+1} = n\left(1 - e^{\frac{-a_t}{\beta}}\right) + p \qquad (12)$$

where n, p and β are free parameters, state s refers to still pot temperature and action a refers to the value of the reflux ratio.

Eq. 11 and 12 are then combined to form the general predictive model forms for the different stages as shown in Table 3. It could be noticed that Model M is a simpler version of Model P and that Model Q is a simpler version of Model R. The only difference is the addition of the parameter β which allows more flexibility for the curvature of the relationships.

Table 3: Proposed predictive model forms for different stages (Models M & P depend on reflux ratio only, since it is assumed that the system always starts from the same initial still pot temperature)

Decision Stage	*Name of model*	*Equation of model*
Initial stage T-3 to T-2 (Predictive model PM3)	M	$s_{t+1} = n\,(1-e^{-a_t}) + p$
	P	$s_{t+1} = n\,(1-e^{-a_t/\beta}) + p$
T-2 to T-1 and T-1 to T (PM2 & PM1)	Q	$s_{t+1} = m\,s_t + n\,(1-e^{-a_t}) + p$
	R	$s_{t+1} = m\,s_t + n\,(1-e^{-a_t/\beta}) + p$

where s is state (temperature or final product purity composition) and a is action (reflux ratio).

Starting again from the same initial training data set of six batches, the RL algorithm was rerun five times for each combination of the proposed predictive models at the different stages (Table 3). Each run was terminated when 21 batches were added to the initial training data set. The percentage of off-spec batches was then calculated out of a total of 105 batch runs using the same previous basis. Fig. 6 4 shows the results of using different combinations of proposed predictive models at various stages of the batch run.

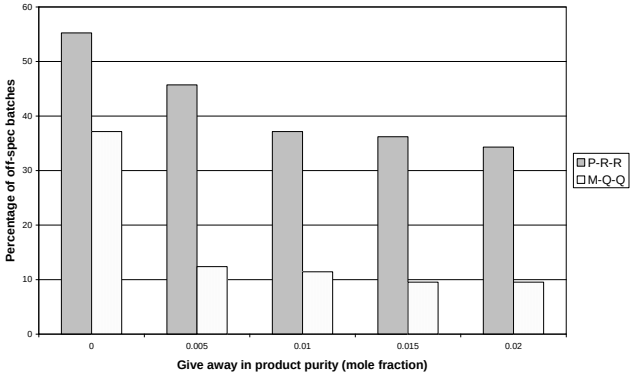

Fig. 6: Percentage of off-spec batches produced for different give-aways in product purity while using different predictive model combinations (at different stages) in RL applications to base case

A close look at the results presented in Fig. 6 reveals the following about RL applications to base case using the proposed predictive models:

1. The results for those combinations of models are quite impressive if taken into account that the algorithm has learned the Value Function without knowledge of VLE data and with only six batch runs initially.
2. Since a certain amount of give-away in product purity is certainly needed to reduce risk of producing off-spec batches, the criterion for the best model is the smallest amount in give-away in product purity. Thus, the use of models M-Q-Q provides better performance than models P-R-R although the later is only a more general version. This is mainly due to the addition of an extra parameter (β in model P & R) which leads to the requirement of more training data.

Wider Applications of Proposed Predictive Model

Predictive model M-Q-Q was used (following the best performance criteria in Fig. 6) in RL application to the newly proposed case studies. Starting from different sets of initial training data for each case study, the algorithm was rerun five times each time until 21 batch runs were added to the initial training data set.

Fig. 7 shows that the performance of the proposed predictive model (M-Q-Q) is very satisfactory for Case Study 1. On the other hand, a very high percentage of 73.33% of off-spec batches was produced in Case Study 2 to achieve on-spec product (i.e. give-away = zero). As for Case Study 3, all batches produced were off-spec. Hence more analysis was required to evaluate and compare the behaviour of all case studies.

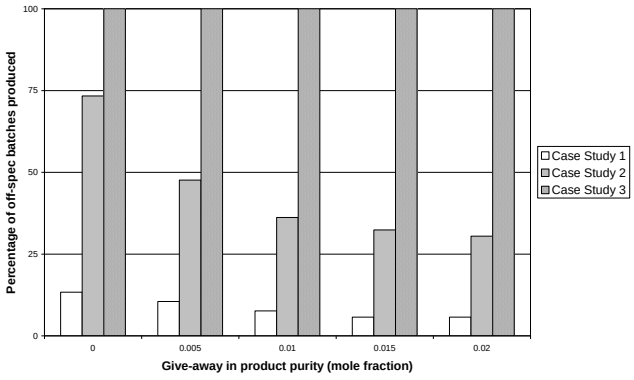

Fig. 7: Percentage of off-spec batches produced with various product give-aways using the M-Q-Q predictive model in RL applications to Case Studies 1, 2 and 3

The detailed behaviour of the batch distillation systems for the three new case studies (Case Studies 1, 2 and 3) was examined to see if the same trends followed in base case was maintained. It became apparent that relationships observed were still within

the mapping capabilities of Model P & R (due to the flexibility offered by the addition of the extra free parameter β). Furthermore, this assumption was investigated through fitting of Models P & R to observed data points from the different batch distillation systems. Fig. 8 shows how Model P provides a good fit for the final product purity as a function of reflux ratio at T-1 for all case studies. The dotted lines (next to the solid lines in Fig. 8) show the prediction of the proposed predictive models (Model P).

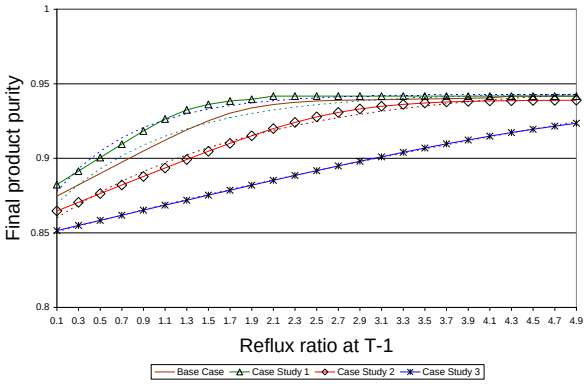

Fig. 8: Use of Model P for different case studies at stage T-1 to T to fit the relationship between final product purity as a function of reflux ratio at T-1 (dotted lines represent predictions of Model P)

The mean squared error achieved during the fitting of the regressive models together with the value of coefficients in Model P are shown in Table 4.

Table 4: Values of mean sum of squared errors and coefficients (n, β and p) for fitted Model P in Fig. 8 to Case Study 1 to 3

Case Study	Mean sum of squared error at end of model fitting	Values of coefficients		
		n	β	p
Base case	0.002385	0.081	1.07	0.86
1	0.002218	0.075	0.75	0.87
2	0.002605	0.095	1.89	0.86
3	0.000455	0.21	10.97	0.85

A value of β of 1.07 explains why the simpler Q model converged in previous application to base case. Since the model was roughly not a function of β. Furthermore, with a value of the coefficient β in Case Study 1 equal to 0.75, which is very near to the value of 1 for base case, the execution of the RL algorithm was repeated for the next more difficult Case Study 2 (relative volatility=2 and number of trays=12) with the calculated value of coefficient β in model P-R-R equal to 1.89. The results produced are shown in Fig. 9 and reveal how the performance of the RL algorithm dramatically improves with 14.29 % off-spec batches produced in comparison to 47.26 % produced when models M-Q-Q is used for a give-away of 0.005 in product purity. This proves that model P-R-R could potentially be used as a truly general predictive model for RL applications to batch distillation if knowledge of parameter β is available.

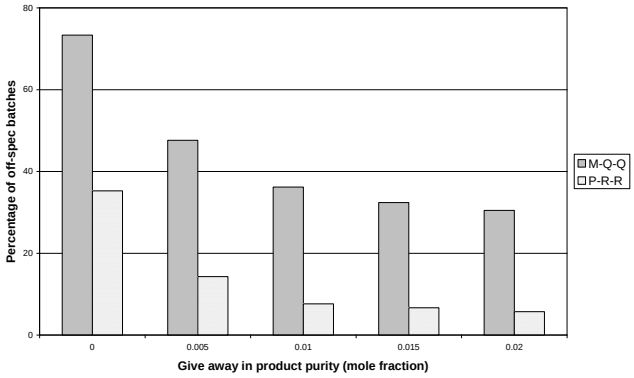

Fig. 9: Improvement in RL applications to case study 2 with the use of the newly fitted model P-R-R (coefficient β =1.89) in comparison to the previous performance of the best proposed predictive model M-Q-Q

Finally, using the P-R-R predictive model the RL algorithm was applied. Fig. 10 shows the PI performance of the system slowly improving although occasionally the product purity goal is not met. On the other hand, it is clear that the product purity has converged around the goal of 0.98 mole fraction.

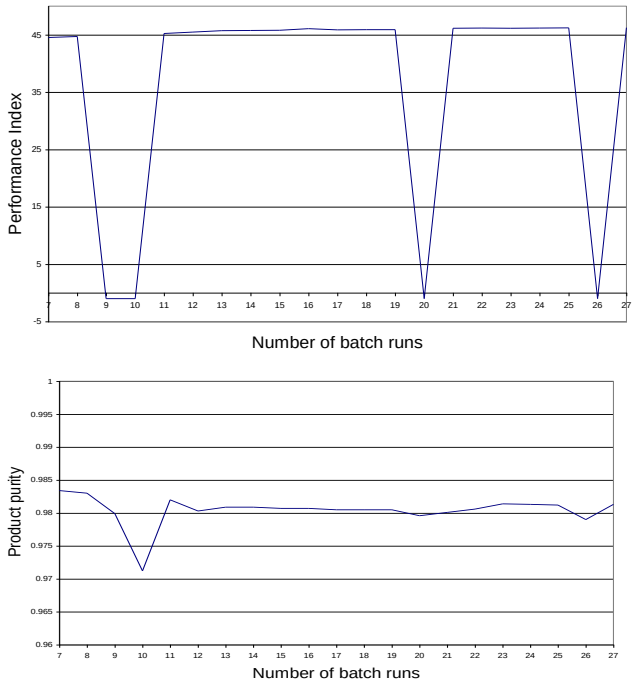

Fig. 10: Performance Index & product purity versus number of batch runs starting from the 1st additional batch run to the initial training data set (7th batch run)

CONCLUSION

The RL application has shown huge potential and a step towards full automation of batch distillation. The results obtained are quite impressive if taken into account that the algorithm has learned the Value Function without knowledge of VLE data and with only six batch runs initially. Although the amount of off-specification batches is well above acceptable level, however work conducted reveals that the predictive model is crucial to the RL approach. Furthermore, following the analysis of data from different case studies, a new predictive model has being put forward. It was shown how predictive model P-R-R has being able to capture the different trends for the different case studies. The results produced are encouraging, although the determination of coefficient β (model P-R-R) is still open to further research. Thus the proposed predictive model is still one factor short of achieving a truly general predictive model for efficient RL applications to batch distillation processes.

Acknowledgement

The work described in this paper was carried out while at the University of Nottingham.

NOMENCLATURE

a	Control action
BxTime	Time for completion of batch (h)
Cs	Heating cost (£/kmol)
D	Amount of product distilled (kmol)
E	Squared error
P	Sales value (£/kmol)
PI	Performance Index
PM1	Predictive model for stage T-1 to T
PM2	Predictive model for stage T-2 to T-1
PM3	Predictive model for stage T-3 to T-2
Q (s,a)	Value Function for state action pair
RL	Reinforcement Learning
m, n, p	Free parameters
s	Process state
T	Final stage
T-1	Last decision stage
T-2	Intermediate decision stage
T-3	Initial decision stage
V	Vapour boilup rate (kmol/h)
X	Product purity

Greek Letters

α	Learning rate
β	Free parameter
σ	Measure (Kerkhof and Vissers, 1978)
ρ	Relative volatility
Ω	Set of feasible control actions

Subscripts

B	Bellman
diff	Difficulty
D	Product
F	Feed
r	Product
T	Time
T	Final time step

Superscripts

*	Optimum
N	Number of theoretical plates in column

REFERENCES

1. Baird, L.C. and A.H. Klopf. 1993. Reinforcement Learning with High-dimensional Continuous Actions, Technical Report WL-TR-93-1147, Wright Laboratory, Wright Patterson Air Force Base.
2. Barolo, M. and P.D. Cengio. 2001. Closed-loop optimal operation of batch distillation columns. Computers and Chemical Engineering 25: 561-569.
3. Bellman, R. 1957. Dynamic Programming, Princeton University, Press, Princeton, New Jersey.
4. Bonny, L., S. Domentech, P. Floquet and L. Pibouleau. 1994. Recycling of slop cuts in multicomponent batch distillation. Computers and Chemical Engineering 18:S75-S79.
5. Carling, A. 1992. Introducing Neural Networks, SIGMA Press, UK.
6. Converse, A.O. and G.D. Gross. 1963. Optimal distillate-rate policy in batch distillation. Industrial Engineering and Chemistry Fundamentals 2:217-221.
7. Coward, I. 1967. The time optimal problem in binary batch distillation. Chemical Engineering Science 22:503-516.
8. Cressy, D.C., I.T. Nabney and A.M. Simper. 1993. Neural control of a batch distillation. Neural Computing and Applications 1:115 – 123.
9. Diwekar, U.M., R.K. Malik and K.P. Madhavan. 1987. Optimal reflux rate policy determinations for multicomponent batch distillation columns. Computers and Chemical Engineering 11:629-637.
10. Diwekar, U.M. 1995. Batch distillation: Simulation, optimal design and control. Carneige Mellon University, Pittsburg, Pennsylvania.
11. Egly, H., N. Ruby and B. Seid. 1979. Optimum design and operation of batch rectification accompanied by chemical reaction. Computers and Chemical Engineering 3:169-174.
12. Farhat, S., M. Czernicki, L. Pibouleau and S. Domenech. 1990. Optimization of multiple-fraction batch distillation by nonlinear programming. AIChE Journal 36:1349-1360.
13. Hansen, T. T. and S.B. Jorgensen. 1986. Optimal control of binary batch distillation in tray or packed columns. Chemical Engineering Journal 33:151-155.

14. Keith, F. M. and Brunet. 1971. Optimal operation of a batch packed distillation column. Canadian Journal of Chemical Engineering 49:291-294.
15. Kerhof L. H. and H.J.M. Vissers. 1978. On the profit of optimum control in batch distillation. Chemical Engineering Science 33:961-970.
16. Kim, Y.H. 1999. Optimal design and operation of a multi-product batch distillation column using dynamic model. Chemical Engineering and Processing 38: 61-72.
17. Logsdon, J.S., U.M. Diwekar and L.T. Biegler. 1990. On the simultaneous optimal design and operation of batch distillation columns. Chemical Engineering Research and Design 68:434-444.
18. Lopes, M.M. and T.W. Song. 2010. Batch distillation: Better at constant or variable reflux? Chemical Engineering and Processing: Process Intensification 49:1298-1304.
19. Luyben, W.L. 1992. Practical Distillation Control. Van Nostrand Reinhold, New York, USA.
20. Martinez, E.C., R.A. Pulley and J.A. Wilson. 1998a. Learning to control the performance of batch processes. Chemical Engineering Research and Design 76:711-722.
21. Martinez, E.C. and J.A. Wilson. 1998b. A hybrid neural network first principles approach to batch unit optimization. Computer and Chemical Engineering 22:S893-S896.
22. Martinez, E.C, J.A. Wilson and M.A. Mustafa. 1998c. An incremental learning approach to batch unit optimization. The 1998 IChemE Research Event, Newcastle, UK.
23. MATLAB: Version 4, The Math Works Inc, Natick, Massachusetts.
24. Mayur, D.N. and R. Jackson. 1971. Time optimal problems in batch distillation for multicomponent mixtures columns with hold-up. Chemical Engineering Journal 2:150-163.
25. Mujtaba, I.M. and S. Macchietto. 1992. An optimal recycle policy for multicomponent batch distillation. Computers and Chemical Engineering 16:S273-S280.
26. Mujtaba, I.M. and S. Macchietto. 1997. Efficient optimization of batch distillation with chemical reaction using polynomial curve fitting. Industrial and Engineering Chemistry Research 36:2287-2295.
27. Mujtaba, I.M. and M.A. Hussain. 1998. Optimal operation of dynamic processes under process-model mismatches: Application to batch distillation. Computers and Chemical. Engineering 22:S621-S624.
28. Mustafa, M.A. 2001. Reinforcement Learning in Batch Chemical Processing, The University of Nottingham, School of Chemical, Environmental and Mining Engineering (SChEME), Nottingham, United Kingdom.
29. Pommier, S., S. Massebeuf, B. Kotai, P. Lang, P. Baudouin, P. Floquet and V. Gerbaud. 2008. Heterogeneous batch distillation processes: Real system optimization. Chemical Engineering and Processing: Process Intensification 47:408-419.
30. Sorensen, E. 1997. Alternative ways of operating a batch distillation column. Institution of Chemical Engineers Symposium Series 142:643-652.
31. Stenz, R. and U. Kuhn. 1995. Automation of a batch distillation column using fuzzy and conventional control. IEEE Transactions on Control Systems Technology 3:171-176.
32. Sutton, R.S. and A.G. Barto. 1998. Reinforcement Learning: An Introduction, The MIT Press, Cambridge, Massachusetts, London, England.
33. Wilson, J.A. and E.C. Martinez. 1997. Neuro-fuzzy modeling and control of a batch process involving simultaneous reaction and distillation. Computers and Chemical Engineering 21:S1233-S12.