Neural Network Modelling and Predictive Control of a Milk Pasteurisation Plant

Khadir M. T. and Ringwood J.
Department of Electronic Engineering
National University of Ireland, Maynooth
Co. Kildare, Ireland
Fax +353-1-7086027, Tel +353-1-7046074, e-mail: tarek.khadir@may.ie

Abstract

This paper investigates the possible use of artificial neural networks (ANN), more precisely multi-layer perceptrons (MLPs), for the nonlinear modelling and predictive control of a milk pasteurisation plant. Model predictive control (MPC) schemes require the development of a predictive model. Using data gathered from an industrial milk plant, a nonlinear multi-step ahead neural network predictor model (NNM) was established. A neural predictive controller (NPC) was then designed on the same basis for the control of milk pasteurisation temperature. Simulation results are presented and conclusions are drawn.

Keywords: Milk pasteurisation, non-linear modelling, model predictive control, artificial neural networks

1 Introduction

Previous work has been done on modelling and control of milk pasteurisation using predictive control by the same authors. A first principle physical model has been developed along with a predictive functional controller (PFC) in [1]. The linear predictive controller is shown to perform better than a classical PID [1, 2] where the linear model obtained from first principles has been validated around a fixed operating point, consistent with normal operation of the pasteurisation plant. For a wider control range, for example other dairy products that require higher or lower heating requirements; a model capable of prediction over a wider range is needed and a linear model cannot achieve such performance. Moreover, analytical MPC strategies are restricted to linear models in order to compute a fixed control law [3, 4]. Therefore, for industrial processes, a linearised model is most often used as an internal model. However, physical modelling of industrial processes, due to their physical complexity, is not trivial and demands a lot of time and effort. The use of ANNs is, therefore, justified due to their ability (see Section 3) to search for a valid nonlinear model from input/output data. The established non-linear model can also be used for the implementation of a "nonlinear" neural predictive controller. The extra computation involved in NPC must be achievable within the 12 second sampling period of the pasteuriser.

2 Milk pasteurisation: process description

The pasteuriser used is a Clip 10-RM plate heat exchanger (PHE) from Alfa Laval. A PHE consists of a pack of stainless steel plates clamped in a frame. The plates are corrugated in a pattern designed to increase the flow turbulence of the medium and the product [5]. The pasteuriser is divided in five sections, S1 to S5. Section S4 and S2 are for regeneration, S1 and S3 for heating and S5 for cooling. In the Clip 10-RM the milk treatment is performed as shown in Figure 1. First the raw milk at a concentration of 4.1% enters section S4 of the PHE at a temperature of 2.0°C. It is then preheated to a temperature of 60.5°C by the outgoing pasteurised milk which as a result is reduced to a temperature of 11.5°C. Passing this section, the milk now at a temperature of 60.5°C, enters section S3 where its temperature increases to 64.5°C using hot water as a medium. The milk, before reaching the next section, is first separated from the fat then standardised and homogenised to a concentration of 3.5%. It then enters section S2, where it is preheated to a

temperature of 72°C using the already pasteurised milk as a medium. The milk is finally brought to the pasteurisation temperature in section S1 (75.0°C) using hot water at around 77.0°C as a medium. After that the homogenised pasteurised milk is held at the pasteurisation temperature for 15 s in the holding tube section before being cooled using the incoming cold milk in section S4 and section S2. Finally the pasteurised milk enters the cooling section (section S5) at a temperature of 11.5°C. The milk is chilled to a temperature of 1.0°C using propylene glycol as a medium at a temperature of -0.5°C. Note that the water for the heating sections S3 and S4 is brought to the adequate temperature in steam/water heater of type CB76 from Alfa Laval. As shown in Figure 1, milk pasteurisation temperature is a function of three inputs: steam flow injected in steam/water heater 1, steam flow injected in steam/water heater 2 and the milk input temperature, labeled as F_{v1}, F_{v2} and T_{im} respectively. The milk pasteurisation temperature is then given by a multi input single output (MISO) system, having F_{v1}, F_{v2} and T_{im} as inputs and y, the milk pasteurisation temperature, as output.

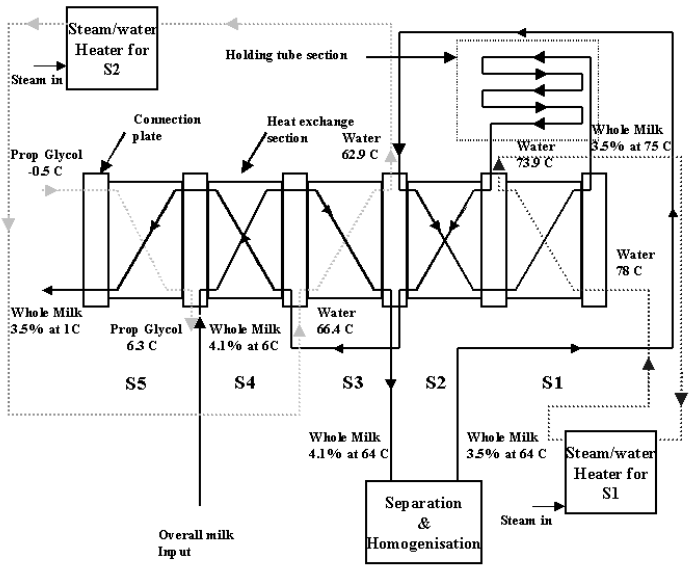


Figure 1: *General layout of the pasteuriser*

3 Modelling of the pasteurisation plant using ANN

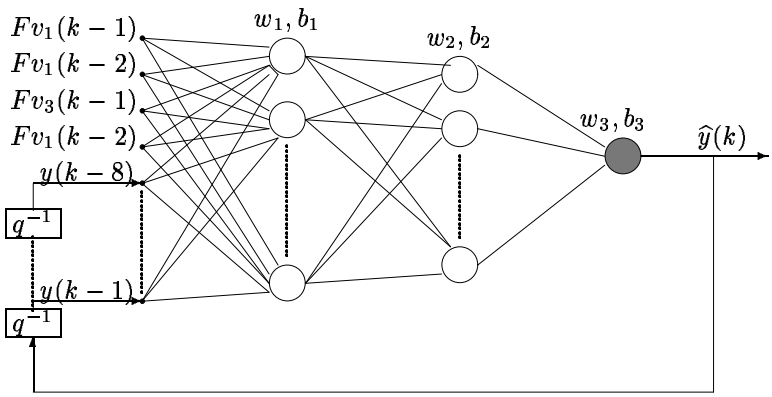
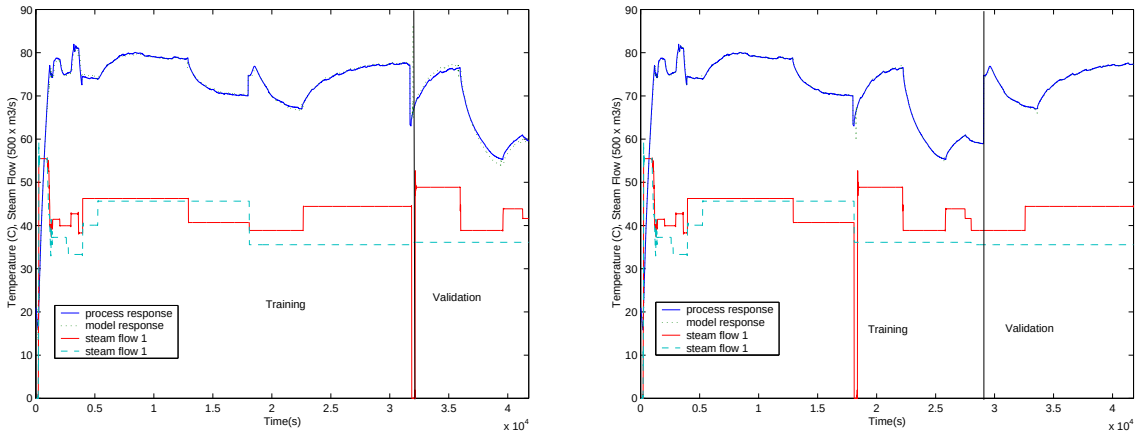


Figure 2: *NNM topology and input signals used for training*

It has been shown by Cybenko [6], that back-propagation neural networks, with one hidden layer, can approximate any non-linear function and generate complex decision regions for input-output mapping. For ease of training and overall reduction for neuron counts, a multi-layer network with an input layer, an output layer and two hidden layers is used. The inputs to the NNM, for training, are chosen to be a sequence of data from, F_{v1}, F_{v2} and a set of delayed values of the output signal y, see

Figure 2. The neurons in the two hidden layers are *tan sigmoid* neurons, where the output layer neuron is a *linear* neuron. The prediction will be given by the NNM obtained after appropriate training on the form of a Nonlinear Moving Average with Exogeneous Variable (NARMAX) model, given by:

$$\widehat{y}(k) = NNM[\widehat{y}(k-1), \widehat{y}(k-2) \cdots \widehat{y}(k-8), F_{v1}(k-1), F_{v1}(k-1), F_{v2}(k-1), F_{v2}(k-1)] \quad (1)$$



(a) Data set 1 (b) Data set 1

Figure 3: *Training and validation of the ANN model (NNM)*

The choice of the inputs has been heavily dictated by the *a-priori* information gathered from the first principle physical model used in [1]. Where the output pasteurisation temperature can be modelled by an eighth order linear system, this justifies the use of eight delayed signals of $\widehat{y}(k)$ in equation (1). The input milk temperature T_{im} is not used in the NNM as the milk is kept at a relatively constant temperature of 2^oC, and its use, in the training process, will only introduce a random disturbance to be modelled. The choice of a good network topology is not a straightforward task. There are no hard rules or theorems to find an optimal topology for a given set of input/output data. However, an appropriate topology may be found by performing network pruning or network growing. Starting with a sufficiently big topology, the NNM is pruned by eliminating the links containing insignificant weights using a *weight elimination method*, for example the Optimal Brain Damage (OBD) method developed by Le Cun [7]. Alternatively, starting with a small NNM (for example a 1-2-1 topology). The network is grown until reaching a size which gives a good prediction model. In this paper, we chose a topology large enough to permit good modelling and possible network pruning. The modelling approach chosen in this paper is, in the main, similar to the one described by Nørgaard *et al* in [8]. Folowing such experimentation, an 8-12-1 MLP is chosen, the network was trained for a total of 20,000 epochs using a set of data that consists of data subsets obtained during a series of test protocols, where F_{v1} and F_{v2} were varied around the operating region. Four subsets of data were used, three for training and a separate subset was used for validation in order to make sure of the validity of NNM. To avoid overtraining i.e., deterioration of the model as it tries to fit the training set, a sum squared error (SSE) on the validation set is plotted along the epoch number and the NNM parameters are chosen when the SSE is minimum. Overtraining and its effects are explained in detail in [9]. A cross validation method, also called *cross model selection*, is used for validation. This method consist of using all the data subsets in turn for validation, obtaining a number of $NNMs$. The

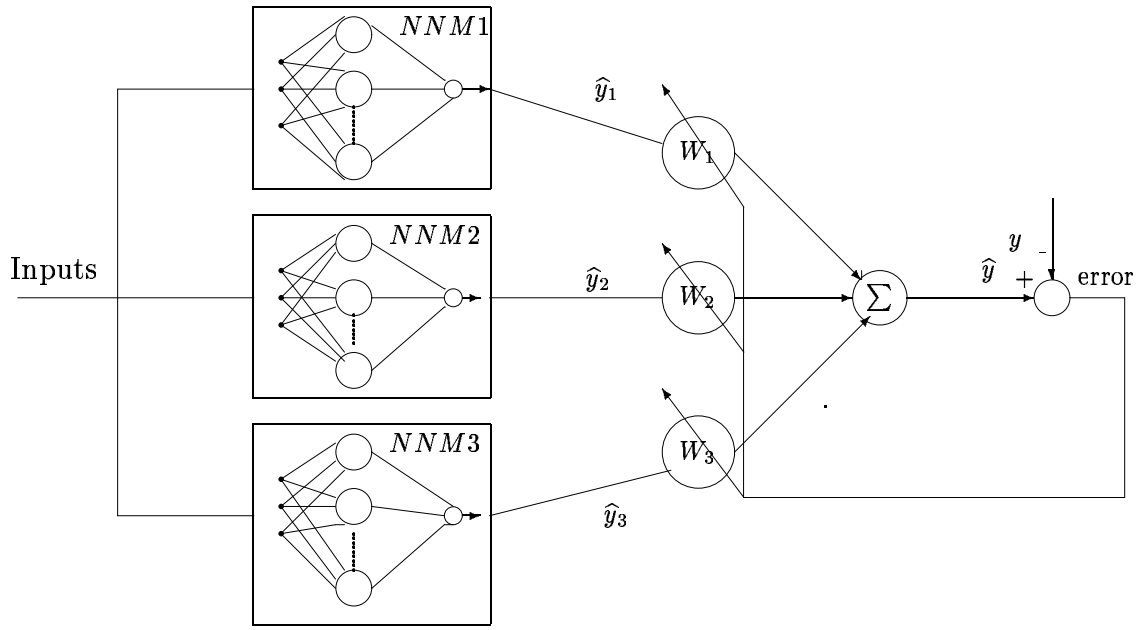


Figure 5: *Overall NNM model: Linear combiner*

method proved to be useful when the number of sample points is constrained. Moreover, having several validation estimates covering the entire training data set gives a better confidence degree to the estimates. The best results, obtained after a number of training runs, for two validation subsets are shown in Figure 3. The model and process response to a change in F_{v1} and F_{v2}, given in m^3/s (shown scaled on the graph by a factor of 500), can be seen for the training and validation data sets. On the other hand, the validation SSE versus epoch number, for the first set, is given in Figure 4. We can see that the SSE after 5,500 epoch start to increase (overtraining occurring), the model parameters are then chosen at that moment in the training process. A number of simulations were run for each validation subset, as we are not guaranteed to reach a global, or at least a decent local minimum in one training session. The SSE for the second and third set have a similar shape, therefore their corresponding graphs, have been omitted. Using the three validation subset we obtained three ANN models, $NNM1$, $NNM2$ and $NNM3$, the corresponding minimum mean square errors MSEs for validation are given in Table 1 along with the corresponding epoch number and the training MSE values. The $NNMs$ where trained using the Matlab Artificial Neural Network Toolbox [10], where the training algorithm used is the well known back propagation algorithm described in detail in [11]. The definitive NNM is then given by a

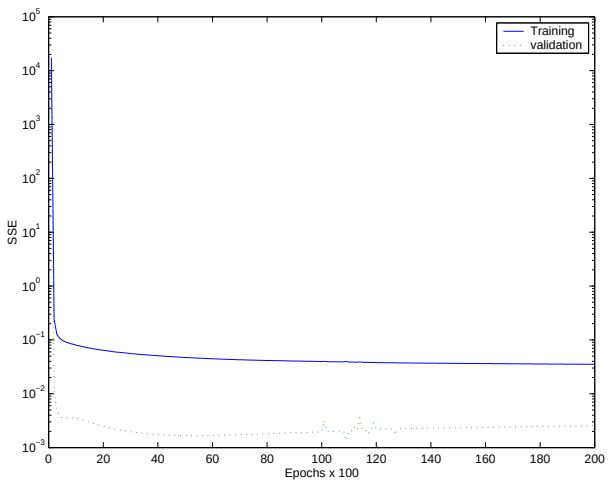


Figure 4: *Validation SSE progression versus training epoch number*

linear combination of $NNM1$, $NNM2$ and $NNM3$, as shown in Figure 5. The linear combiner is trained over the complete data set in order to obtain the weight parameter values; using a simple least square method. The weights W_1, W_2 and W_3 are found to be 0.0689, 0.7703 and 0.1617 respectively.

MSE/Data subsets	1 st	2 nd	3rd
Training	$5.7825\ 10^{-5}$	$2.0862\ 10^{-5}$	$2.2391\ 10^{-5}$
Validation	$1.0457\ 10^{-4}$	$1.0400\ 10^{-6}$	$4.3521\ 10^{-5}$
Epoch stop	2100	5500	2800

Table 1: Overall SSE values.

4 Neural predictive controller design

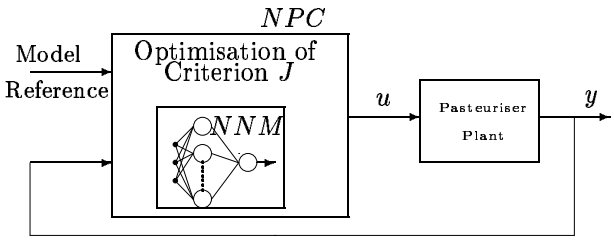


Figure 6: *Structure of the neural predictive controller NPC*

Predictive control is becoming a valuable control strategy for higher control requirements i.e, tighter, faster regulation or tracking in the industrial world. MPC has been used in over 2,000 industrial applications in the refining, petrochemical, chemical, pulp and paper and food processing industries [12]. Some examples of industrial predictive controllers include, PFC from ADERSA, Dynamic Matrix Control (DMC) from DMC Corp and Robust Model Predictive Control Technology (RMPCT) from Honeywell. Most of these algorithms rely on linear or linearised internal models. Using predictive control, a process is regulated by specifying the desired plant output at a particular instance or instances in the future and then calculating the controller action which minimises the predicted error either in the form of an equation, analytical solution for linear internal models or using an optimiser, in the case of a nonlinear model. In this section, a neural predictive controller NPC is designed as shown in Figure 6. The model NNM obtained in Section 3 is used to produce prediction data to the optimisation algorithm. The control variable u is obtained by minimising a criterion function, J given in equation (2), where $N1$ is the time delay (if any), N the prediction horizon and $N2$ the control horizon.

$$J_{=} \sum_{i=N1}^{N1+N} [y_r(k+i) - \widehat{y}(k+i|k)]^2 + \sum_{i=0}^{i=N2} \lambda[u(k+i)]^2 \qquad (2)$$

At each instant k, the predicted output $\widehat{y}(k+i|k)$ is compared to a reference $y_r(k+i)$ describing the optimal trajectory to reach the target, in the regulation case a constant C. The control variable u is represented physically by F_{v1}, where F_{v3} is used to act on the intermediate temperature at the output of section S3 (see Figure 1) and can be considered in the control of the pasteurisation temperature as a disturbance. An optimiser has to be used in order to obtain the value of u that minimises J since an analytical solution is not possible. The NPC is based on a *gradient* optimiser from the NAG toolbox used with Matlab [13] and developed by Gill and Murray in [14]. The optimal u is found by applying different control variable values to the NNM, Figure 6 until finding the value that minimises J. Other, and maybe more efficient optimisers can be

used, as a large number of optimisation routines are available in the literature. As an example, the interior-point method based optimiser is proven to be efficient for MPC design [15]. However, the optimiser used in this paper is found to give satisfactory results, as the sampling time is $12s$, which makes, in this case, the convergence speed of the optimisation algorithm not a priority. During the simulation, the process and the internal models NNM are subject to a disturbance F_{v3} modelled as shown in Figure 7. As can be seen, the value of F_{v3} varies in a pseudo random manner around a given value. In real life, this value would be around 0.08 m^3/s, a steam flow necessary to bring the intermediate temperature, at the output of section S3 in Figure 1, to 64°C. The large change of F_{v3} at time $3,500s$ is an extreme condition and is improbable, but has been modelled to test the capabilities of NPC under severe disturbance changes. Figure 8(a), shows the results obtained using NPC for target temperatures of 74°C and 80°C. Milk pasteurisation is performed at around 74/75°C, however this temperature might be changed in the case of reconstructed milk or pasteurisation of other dairy product. For this reason NPC behaviour must be examined for higher temperature set points. Figure 8(b) shows the computed control variable, it can be seen that NPC takes into account the upper and lower bounds of F_{v1}, namely 0.1 m^3/s and 0 m^3/s respectively. The prediction horizon has been chosen to be 25 samples while a single step control horizon is chosen since, in practice such control action is sufficient.

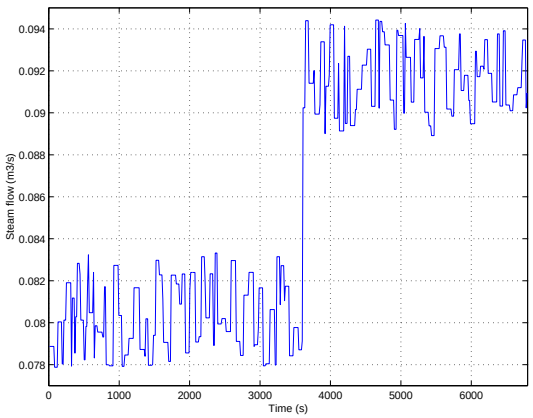
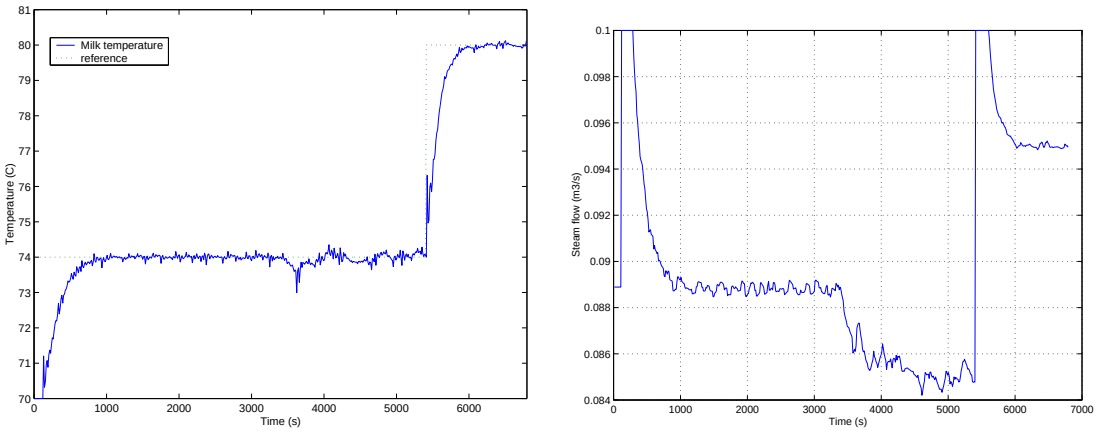


Figure 7: *Input disturbance*



(a) Temperature control

(b) control variable

Figure 8: *NPC behaviour to disturbance and set point changes*

5 Conclusions

Artificial neural networks are a formidable tool for function approximation and classification problems, but yet very easy to use, thus results can be easily achieved. That feature makes ANN

extensively used in modelling of industrial processes unfortunately, not always at their best. ANN are often used as a black box modelling strategy, which they are. However, an insight knowledge of the process can help the designer to find quickly an adequate topology for the neural model. Indeed, having an idea on the order of the system, existence of possible time delays and correlated inputs can influence the choice of the number of inputs used for the network, dictating roughly, the complete ANN topology, this in turn makes a possible network pruning operation more efficient. Nevertheless, ANNs should be used in predictive control only when an internal linear model does not give satisfaction and a development of a valid nonlinear first principle model tend to be too expensive and time consuming; in this case, it is important for the designer to have some physical knowledge of the process. Once a model is established, the designer needs to make sure of its validity and to be aware of any overtraining that might occur. In this paper, the neural model NNM is based on a physical first principle model of the plant developed earlier [1]. A *cross model selection* method is used for validation (picking for each validation subset) the neural model that gives a minimum SSE. Finally, the three neural networks obtained are combined linearly to give the final NNM. This approach increases the chances to reach a global (or at least a decent local) minimum and gives more confidence to the overall model validity.

Predictive control is a powerful control design strategy, for its ability to handle time delays, non minimum phase and a wide range of constrains, which suits industrial applications. However MPC relies heavily on the quality of the internal model. Therefore, if a valid ANN model, describing a complex non linear industrial process can be achieved, then MPC can ensure excellent control. However, computation time (and thus the choice of the optimiser algorithm) becomes an important issue if we are dealing with a fast process. In this case, the pasteurisation plant being a slow process, gives the designer enough time and freedom for the choice of the prediction horizon, internal model complexity and the optimisation algorithm. Indeed, the sampling time of $12s$ is a comfortable margin for an online optimisation even using a relatively slow optimisation algorithm.

References

[1] Khadir, M.T., Richalet, J., Ringwood, J. and O'Connor, B. *Modelling and Predictive Control of Milk Pasteurisation in a Plate Heat Exchanger.* Proc. Foodsim, Nante, France, June 2000.

[2] Khadir, M.T., Ringwood, J. and O'Connor, B. *Comparison of Model Predictive and PID Controllers for Dairy Systems.* Proc. ISSC, Dublin, Rep of Ireland, June 2000.

[3] Clarke, D.W., Mohtadi, C. and Tuffs, P.S. *Generalised Predictive Control-part I. The Basic Algorithm.* Automatica, Vol 23, No2, 137-148, 1987.

[4] Richalet, J. *Pratique de la Commande Predictive.* Traite des Nouvelles Technologies, Serie Automatique Hermes, 1993.

[5] Alpha Laval *Dairy Processing Handbook.* Tetra Pak, 1995.

[6] Cybenko, G. *Aproximation by superposition of a sigmoidal function.* Math. Cont. Sig. syst. vol 2, 303-314 1989.

[7] Le Cun, Y., Denker, J.S. and Solla, S.A. *Optimal brain damage.* Addvanced in Neural information, vol 2, pp126-142 1989.

[8] Nørgaard, M., Poulsen, N.K. and Hansen, L.K. *Neural networks for modelling and control of dynamic systems.* Springer-Verlag, London, UK 2000.

[9] Sjöberg, J. *Non-linear system identification with neural networks.* Ph.D. Thesis No 831, Division of Automatic Control, Department of Electrical Engineering, Linkping University, Sweden, 1995.

[10] Artificial Neural Network Toolbox for use with Matlab. Users guide. The Mathworks Inc. 1994.

[11] Heicht-Nielsen, R. *Neurocomputing.* Addison-Wesley, Reading, Massachusetts 1989.

[12] Quin, S.J. and Badgwell, T.A. *An overview of industrial model predictive control technology.* Chemical Process Control. AIChe Symposium Series. Kantor., J., Garcia, C. and Carnahan, B. Eds. New York. pp 232-256 1997.

[13] NAG Foundation Toolbox for use with Matlab, Numerical Algorithms group Ltd, 1995.

[14] Gill, P.E and Murray, W. *Minimisation subject to bounds on the variables.* Report NAC 72, National Physical Laboratory 1976.

[15] Rao, C.V., Wright, S.J. and Rawlings, J.B. *Application of Interior-Point Mehods to Model Predictive Control.* Journal of Optimisation Theory and Applications. Vol. 99, No 3, pp 723-757 1998.