Application of Generalised Predictive Control to a Milk Pasteurisation Process

Khadir M.T.
Department of Electronic Engineering,
National University of Ireland, Maynooth,
Co.Kildare Ireland
email: tarek.khadir@eeng.may.ie

Ringwood, J.V.
Department of Electronic Engineering,
National University of Ireland, Maynooth,
Co.Kildare Ireland
email: john.ringwood@eeng.may.ie

ABSTRACT
In this paper, Generalised Predictive Control (GPC) is applied to a milk pasteurisation process, in order to improve pasteurisation temperature control. The controller robustness to plant/model mismatch is investigated, as the process is simulated by a nonlinear Artificial Neural Network (ANN) based model, while the embedded internal model is given by a linear First Principles (FP) model. The GPC controller is furthermore compared to an optimally tuned PID controller.

KEY WORDS
Milk pasteurisation, nonlinear modelling, generalised predictive control, artificial neural networks

1 Introduction

Model Predictive Control (MPC) grew substantially in popularity and its field of application diversified substantially since its first applications in the refining and petrochemical industry [1, 2]. It is reported [3] that MPC has been used in over 2,000 industrial applications in the chemical, pulp and paper and food processing industries, on top of the traditional refining and petrochemical sector.

The goal behind the emergence of using advanced control techniques, including MPC, is to reduce occurring variance in the Controlled Variable (CV), and then lower the control setpoint target, if possible, thus reducing cost and saving energy. MPC was found very effective to tackle such control problems due to valuable prediction given by the embedded model as well as integrated constraint handling.

The milk pasteurisation problem may be considered as a classic example for highlighting MPC benefits. It is often the case in pasteurisation processes to aim for a high setpoint in order to avoid any violation of the 72.0°C pasteurisation temperature in case of disturbances, sometimes heating as high as 76.0°C. This may alter the milk constituents, especially if the temperature variance is large. The target for an MPC controller is, then to decrease the variance obtained by a classical controller (*i.e.*, PID) and shift the setpoint target as shown in Figure 1. Note that pasteurisation involves holding the milk at the pasteurisation temperature for at least 15s, which introduces a pure time delay to the system, as well as the existence of input constraints on the Manipulated Variable (MV), generally the position of a steam valve. Pure time delay and the existence of input constraints furthermore encourage the use of MPC [4].

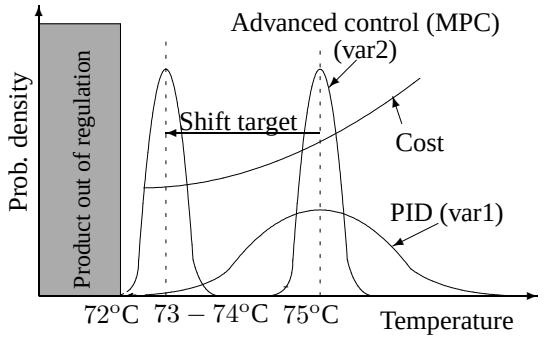


Figure 1. *Goal of an MPC approach: Squeeze the variance shift the target*

2 Physical description of the pasteurisation plant

The pasteuriser studied is based on a Clip 10-RM plate heat exchanger (PHE) from Alfa Laval. A PHE description can be found in [5]. The pasteuriser is divided into 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 3. The raw milk enters section S4 of the PHE at a temperature of 2.0°C. It is then routed to the remaining sections S3, S2 and finally S1 where it reaches pasteurisation temperature (75.0°C). The milk is heated using hot water in S3 and S1 and by the already pasteurised milk in S4 and S2. In the latter sections the already pasteurised milk is also cooled as a result. The milk is finally chilled to a temperature of 1.0°C using propylene glycol as a medium, at a temperature of

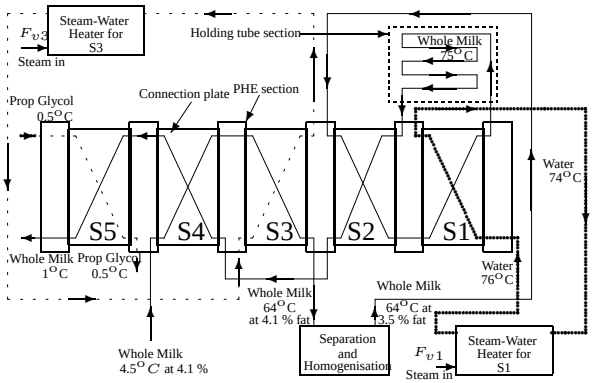


Figure 2. *General layout of the pasteuriser*

-0.5°C.

Note that the water for the heating sections S3 and S4 is brought to the adequate temperature in steam/water heaters of type CB76 from Alfa Laval. If only the pasteurisation section is considered (Section S2, S1 and the steam/water heater 1, Figure 3), milk pasteurisation temperature is a function of two inputs: steam flow injected in steam/water heater 1, and the milk input temperature coming out of section S3, defined as F_{v1} and T_{op3} respectively. The milk pasteurisation temperature is then given by a Multi Input Single Output (MISO) system, having F_{v1} and T_{op3} as inputs and T_{op1}, the milk pasteurisation temperature, as output. In the remaining of this paper, only the pasteurisation section model is considered as it is the one where milk pasteurisation takes place.

3 Models for the pasteurisation process

MPC control strategies require the use of an internal model for prediction, in the case of GPC the internal model must be linear and of a Controlled Auto-Regressive Integrated Moving Average (CARIMA) form [6]. However, for simulation purposes, a model is also needed as a process model. This model does not have to be linear, in fact, it is preferred that it exhibits the true process behavior (including nonlinearities), this will further test the controller robustness to plant/model mismatches during simulation. Thus, two models are used in this paper:

- An FP linear model used as the GPC internal model, and
- an ANN based model used as a process model.

3.1 Internal model

The modelling approach used to develop a relatively simple linear FP model, yet with an appreciable level of detail, is based on the energy balance equations of the heat transfer process. It considers every section of the PHE as a single plate, where the fluids, product and medium, pass at either side (for heat transfer through a wall. The models for each section are then concatenated in order to give the full PHE model. A similar approach is used to model the steam/water heater, considering that the total steam energy is transferred to the water to be heated [7]. The pasteurisation model (considering sections S1, S2 and steam/water heater 1) is given in its final transfer function form in equation (1).

$$T_{op1} = \frac{c_0 + b_1 \mathrm{s} + c_2 \mathrm{s}^2}{a_0 + a_1 \mathrm{s} + a_2 \mathrm{s}^2 + a_3 \mathrm{s}^3 + a_4 \mathrm{s}^4} T_{op3} + \frac{b_0 + b_1 \mathrm{s}}{a_0 + a_1 \mathrm{s} + a_2 \mathrm{s}^2 + a_3 \mathrm{s}^3 + a_4 \mathrm{s}^4} e^{-15\mathrm{s}} F_{v1} \qquad (1)$$

The expressions for the parameters a_i, b_i and c_i were derived from first principles modelling of the pasteurisation plant (also given in [7]), and their numerical values are given in Table 1.

	Parameter subscript, i		
Parameters	0	1	2
a_i	0.0850	7.4286 10^2	5.1639 10^5
b_i	28.4400	5.1192 10^3	-
c_i	0.0850	64.8000	1.1664 10^4
Parameters	3	4	
a_i	1.1664 10^8	8.3980 10^9	
b_i	-	-	
c_i	-	-	

Table 1. Continuous pasteurisation model parameters.

3.2 Process model

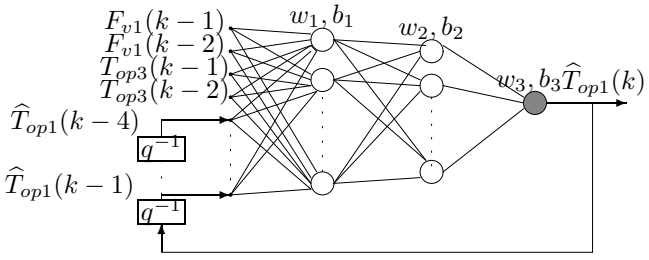


Figure 3. *ANN topology and input signals used for training*

The ANN model used in this paper has been established and validated in [7, 8]. For ease of training and overall reduction in neuron count, a multi-layer network with input, output, and two hidden layers is used. The inputs to the

Models	MAE (°C)
Linear (FP)	0.6349
Nonlinear (ANN)	0.2187

Table 2. Linear and nonlinear model accuracy.

ANN model are step delayed versions of F_{v1} and T_{op3}, and four delayed values of the estimated output $\widehat{T}_{op1}$, see Figure 3. The prediction will be given by the function f_{ANN} obtained after appropriate training, of the form:

$$\widehat{T}_{op1}(k) = f_{ANN}[\widehat{T}_{op1}(k-1), \widehat{T}_{op1}(k-2) \cdots \widehat{T}_{op1}(k-4)$$

$$, F_{v1}(k-1), F_{v1}(k-2), T_{op3}(k-1), T_{op3}(k-2)] \quad (2)$$

The choice of the inputs has been heavily dictated by the *a-priori* information gathered from the first principle physical model (Section 3.1). Since the output pasteurisation temperature can be modelled by a fourth order linear system, this justifies the use of four delayed values of $\widehat{T}_{op1}(k)$ in equation (2). A topology large enough to permit good modelling and possible network pruning [9] is chosen, i.e., 8-10-1. The network was trained for 20,000 epochs using a set of data containing subsets obtained during a series of test protocols on an industrial Clip 10-RM pasteuriser at a sampling rate of 12 s, where F_{v1} and T_{op3} were varied around the operating region of interest. Four subsets of data were used for training where a separate subset was used for validation in order to obtain an appropriate model. To avoid overtraining i.e., deterioration of the model as it tries to fit the training set (see [7, 10]), a Sum Squared Error on the validation set ($\mathrm{SSE_v}$) is plotted and the model parameters are chosen when $\mathrm{SSE_v}$ is minimum i.e., early stopping. A cross validation method [10] is used, where one data sets is used for validation at each time, while the rest of the data is used for training. This method has proven to be useful when the number of data points is constrained. Moreover, this will give a better degree of confidence to the estimates. The definitive ANN model is then given by a linear combination of the models obtained with each validation set in the cross validation process.
Both ANN and FP model responses are shown in Figure 4, with their relative accuracy detailed in Table 2.

4 Generalised predictive control (GPC)

In GPC, the internal model used is a CARIMA model, given in equation (3).

$$A(q^{-1})y(k) = B(q^{-1})u(k-d) + C(q^{-1})\frac{\xi(k)}{\Delta(q^{-1})} \quad (3)$$

where, A, B and C are polynomials in the backward shift operator q^{-1}:

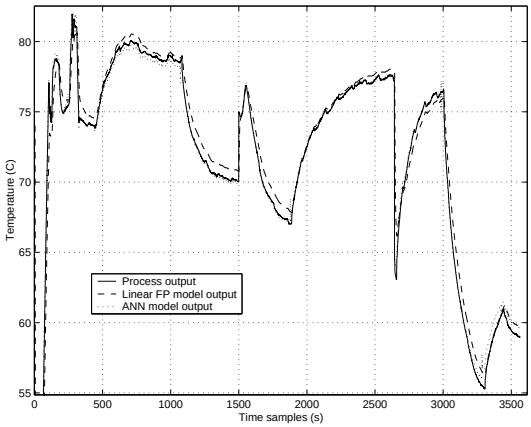


Figure 4. Linear FP model and the ANN model responses

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{nb} q^{-nb}$$
$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}$$
$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}$$

$\xi(k)$ is an uncorrelated random sequence (disturbance signal) and $\Delta(q^{-1})$ is the delta operator $(1 - q^{-1})$.

For simplicity, $C(q^{-1})$ is chosen to be 1 to give the model:

$$A(q^{-1})y(k) = B(q^{-1})u(k-d) + C(q^{-1})\frac{\xi(k)}{\Delta(q^{-1})} \quad (4)$$

For the prediction interval j:

$$y(k+j) = E_j B(q^{-1})\Delta u(k+j-d) + F_j y(t) + E_j \xi(k+j) \quad (5)$$

As E_j is degree $j-1$, the noise component are all in the future so that the optimal predictor, given measured output data up to time k and any given $u(k+i)$ for $i > 1$, is clearly:

$$\widehat{y}(k+j|k) = G_j \Delta u(k+j-1) + F_j y(t) + E_j y(k) \quad (6)$$

where $G_j = E_j B$.

For the derivation of a j-step ahead predictor of $y(k+j)$ based on (6), consider the identity:

$$1 = E_j(q^{-1})A(q^{-1})\Delta(q^{-1}) + q^{-j}F_j(q^{-1}) \quad (7)$$

where E_j and F_j are polynomials uniquely defined given $A(q^{-1})$ and the prediction interval j.

Denoting $\widetilde{A} = A\Delta = 1 + \widetilde{a}_1 q^{-1} + \cdots + \widetilde{a}_{na} q^{-na+1}$, from equation (7),

$$B(q^{-1}) = E_j(q^{-1})B(q^{-1})\widetilde{A}(q^{-1}) + B(q^{-1})q^{-j}F_j(q^{-1})$$

$$E_j B(q^{-1}) = B_j(q^{-1})[1 - q^{-j}F_j(q^{-1})]/\widetilde{A}(q^{-1})$$

$$G_j = B_j(q^{-1})[1 - q^{-j}F_j(q^{-1})]/\widetilde{A}(q^{-1}) \qquad (8)$$

Using recursing diophantine equation (8) so that polynomials E_{j+1} and F_{j+1} can be obtained given the values of E_j and F_j [6].

For GPC, a whole set of predictions are considered for which j runs from a minimum up to a larger value: these are termed the minimum and maximum prediction horizon. For $j < k$ the prediction process $\widehat{y}(k+j|k)$ depends entirely on available data, but for $j \geq k$ assumptions need to be made about future control action.

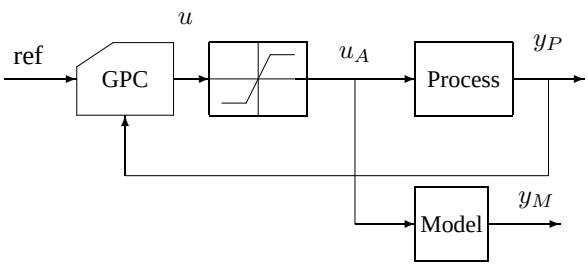


Figure 5. *Handling of input constraints*

4.1 Control law and choice of control parameters

Assuming a future set-point or reference sequence $[w(k+j); j = 1, 2, \cdots N_2]$ is available, in most cases $w(k+j)$ is a constant value w_0, *i.e.*, the regulation case. The objective of the predictive control law is to drive future plant outputs $y(k+j)$ close to w_0.

GPC normally specifies a cost function of the form:

$$J = \sum_{j=N_1}^{N_2}[w_0 - y(k+j)]^2 + \sum_{j=1}^{N_2}\lambda(j)[\Delta u(k+j-1)]^2 \quad (9)$$

where: N_1 is the minimum costing horizon, N_2 is the maximum costing horizon for error, and control signal λ is a control weighting sequence (often chosen as $\lambda(j) = \lambda = 0.5$).

GPC assumes that the future control actions, after an interval $N_c < N_2$, are null, an idea originally used in DMC (Cutler and Ramaker, 1980).

$$\Delta u(k+j-1) = 0 \qquad j > N_c \qquad (10)$$

The value of N_c is called the "control horizon". In cost function terms, this is equivalent to placing effectively infinite weight on control changes after some future time.

The computational advantages of having $N_c < N_2$, the solution vector $\widetilde{u}$ is then of dimension N_c and the prediction equations reduce to:

$$\widehat{y} = G_1\widetilde{u} + f \qquad (11)$$

Minimising the cost function J, with respect to u with no constraints on future controls, results in the projected control increment vector:

$$\widetilde{u} = (G_1^T G_1 + \lambda I)^{-1} G_1^T (w - \mathrm{f}) \qquad (12)$$

where:

$$\widetilde{u} = [\Delta u(k), \Delta u(k+1), \cdots, \Delta u(k+N_c-1)]^T$$

$$\mathrm{f} = [f(k+1), f(k+2), \cdots, f(k+N_2)]^T$$

where f is obtained by solving the Diophantine equation (8) [6].

The first j terms in $G_j(q^{-1})$ are the parameters of the step response and therefore, $g_{ij} = g_j$ for $j = 0, 1, 2, \cdots, < i$ independent of the particular G polynomial.

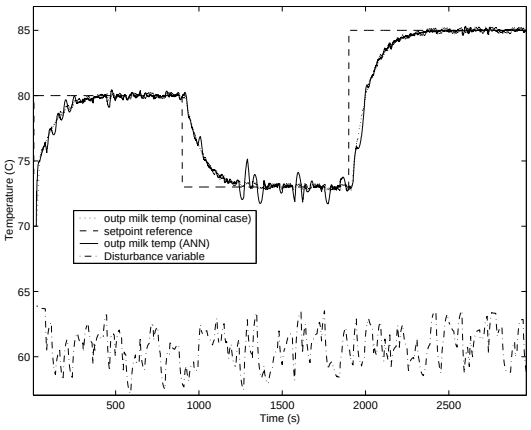
where

$$G = \begin{bmatrix} g_0 & 0 & \cdots & 0 \\ g_1 & g_0 & \cdots & 0 \\ . & .. & \cdots & . \\ : & . & & g_0 \\ : & : & & : \\ g_{N_2-1} & g_{N_2-2} & \cdots & g_{N_2-N_c} \end{bmatrix}$$

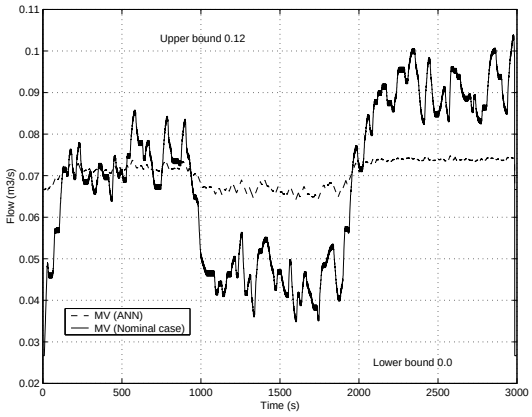
The matrix G involved in the calculation of equation (12), is of the much reduced dimension $N_2 \times N_c$. In particular if $N_c = 1$ (usually chosen for a simple plant), this reduces to a scalar computation. The dimension of G can be further reduced if the process has a known pure time delay, d. Then, the start of the prediction horizon N_2 is set to start at $N_1 = d$. N_2, in theory, should be equal to the process rise time, however in practice this may lead to very large values (especially for slow processes). Smaller values can be selected for N_2 as long as they are larger than the order of polynomial $B(q^{-1})$ [6].

4.2 Constraint handling

In the original GPC algorithm [6], it is not specified on how to handle constraints. For the input constraints on F_{v1} considered in this paper, GPC uses a sub-optimal approach. A simple solution will be to feed the model not with the MV calculated by the GPC algorithm, but with its constrained value. The model output y_M is then, calculated with the new applied MV, Figure 5. This technique is also used in Predictive Functional control (PFC) [11].



(a) Output product temperature T_{op1}



(b) Manipulated variable F_{v1}

Figure 6. GPC response facing input disturbances when full model used as internal model

5 Optimised benchmark PID

The PID controller transfer function is usually given by (13):

$$C(s) = K_p + \frac{K_i}{\mathrm{s}} + K_d \mathrm{s} \tag{13}$$

where K_p, K_i and K_d are respectively the proportional, integral and derivative gain. A digital version of the classical PID is used when the plant is operated by any digital/computer based controller, and can be given by the following set of equations, assuming backward difference approximations to derivative and integral terms:

$$e_r(k) = w_0(k) - y(k) \tag{14}$$

$$s(k) = s(k-1) + e_r(k) \tag{15}$$

$$u(k) = K_p \left(e_r(k) + \frac{T_s}{T_i} s(k) + \frac{T_d}{T_s}(e_r(k) - e_r(k-1)) \right) \tag{16}$$

where: $T_i = \frac{K_p}{K_i}$, $T_d = \frac{K_d}{K_p}$ and T_s is the sampling period.

From equation (16) we can see that $u(k)$ is a function of the PID parameters K_p, K_i and K_d. The quadratic criterion J, equation (9), can in turn, be written in terms of K_p, K_i and K_d. An optimisation method based on a quasi-Newtonian algorithm is used for the function minimisation in order to calculate the PID parameters that give the optimal $u(k)$ [12]. This assumes a consistent control objective between the GPC and PID controllers.

6 Simulation results

The following GPC design parameters were chosen:

- A sampling period, T_s, of 12s was found economical and still satisfies the usual Shannon sampling theorem [13] as well as sampling requirements for industrial processes given by equation (17) [11].

$$T_s = \frac{T_r}{N}, \qquad \text{with} \qquad 30 < N < 50 \tag{17}$$

  where T_r is the process rise time to reach steady state, and N a constant.
- An exponential reference trajectory,
- a costing horizon, N_1, chosen equal to $d = 1$, (12s, approximation of the system time delay of 15s),
- a prediction horizon $N_2 = 7$ (84 s), superior to the order of $B(q^{-1})$,
- a control horizon N_c of 1 (12 s), and
- λ is taken equal to 0.5, equation (9).

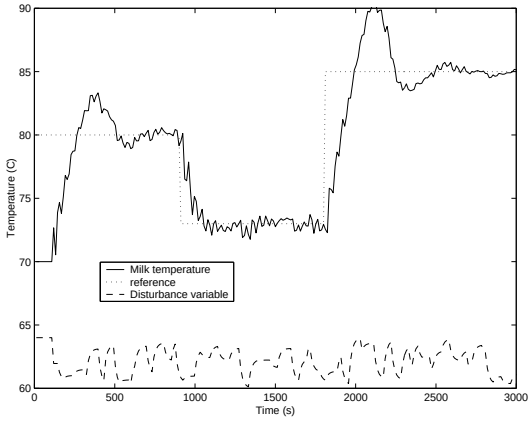
The pasteurisation sub-model expressed in equation (1), is reformulated in a CARIMA form (as in equation (3)).

The Diophantine equation, (7), is then solved in order to find the coefficients $E_j(q^{-1})$ and $F_j(q^{-1})$. The control law is already given in equation (12).

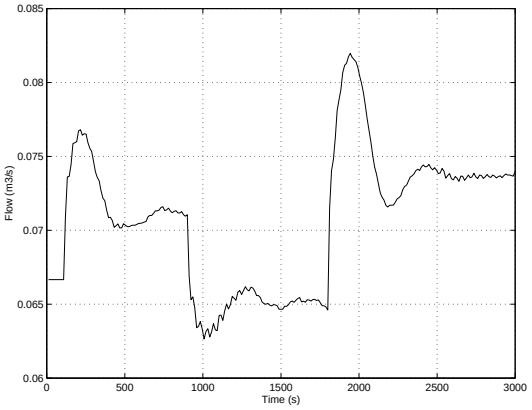
The optimal PID control performance is shown in Figure 7.

In the case of disturbance on the input variable, T_{op3}, it can be seen from Figure 6(a) and 7(a), that the GPC response is better than the PID one. The variance in the output temperature T_{op1} is minimised to around 0.5°C when ANN_P is used as a process model, and to around 0.2°C in nominal case (when the FP model is used as a process model).

Table 3 gives the control performances of the GPC and the optimal PID controllers, in terms of Mean Absolute Error (MAE), Maximum Variance in Steady State (MVSS) and Maximum overshoot. Not that, for GPC, the regulation at



(a) Controlled variable T_{op1}



(b) Manipulated variable F_{v1}

Figure 7. Optimal PID performance for set-point changes and disturbance rejection

Control strategy	PID	GPC Internal model	
		FP	ANN
MAE(^{o}C)	1.5	1.1	0.8
Max. MVSS (^{o}C)	2.4	0.3	1.2
Max. Overshoot (^{o}C)	7.2	0.0	0.0

Table 3. GPC and PID control performances.

a set-point of 73.0^oC is not as good as the results for other set-points, as shown in Figure 6(a). It is strongly suspected that this is due to poor behavior of the ANN model in the region of 72-73^{o}C, as it was not appropriately trained at these temperatures due to the lack of experimental data.

7 Conclusion

For the pasteurisation process presented, a GPC control approach gives better results than a classical PID controller, even tuned optimally with respect to the same quality criterion. The faster response with minimised variance and no overshoot given by the GPC can considerably reduce the control variance, making possible a shift of the set-point to lower temperatures saving energy and avoiding milk overheat.

References

[1] J. Richalet, A. Raul, J.L. Testud and J. Papon, Algorithmic control of industrial processes. *Proceedings of the 4$^{\text{th}}$ IFAC Symposium on Identification and System Parameters Estimation*. 1976, 1119-1167.

[2] C.R. Cutler and R.L. Rameker, Dynamic matrix control: A computer control algorithm, *AIChE 86$^{\text{th}}$ National Meeting*, April 1982.

[3] S.J. Quin and T.A. Badgwell, An overview of industrial model predictive control technology, *Chemical Process Control. AIChe Symposium* Series. Kantor, J., Garcia, C. and Carnahan, B. Eds. New York. 1999, 232-256.

[4] E.F. Camacho and C. Bordons, *Model predictive control.* London: Springer-Verlag, 2000.

[5] G. Bylund, *Dairy Processing Handbook.* Tetra Pak, 1995.

[6] D.W. Clarke, C. Mohtadi and P.S. Tuffs, *Generalised predictive control-part I. The basic algorithm*. Automatica, Vol 23(2),1987, 137-148.

[7] M.T. Khadir, *Modelling and predictive control of a milk pasteurisation plant*. Ph.D. Thesis, Dept. of Electronic Engineering, NUI Maynooth, Ireland, 2002.

[8] M.T. Khadir and J. Ringwood, Linear and nonlinear model predictive control design for a pasteurisation plant. *Control and Intelligent Systems*, 31(1), 2003,1-8.

[9] Y. Le Cun, J.S. Denker and S.A. Solla, Optimal brain damage, *Advances in Neural Information, 2*, 1989, 126-142.

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

[11] J. Richalet, La commande predictive, *Techniques de l'Ingenieur Traite Mesure et Control*, R7 423, 1998, 1-17.

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

[13] J.R. Higgins, *Sampling theory in Fourier and signal analysis: foundations*, New York: Oxford University Press, 1996.