

Application of artificial neural networks in modeling and optimization of batch crystallization processes

Petia Georgieva and Sebastião Feyo de Azevedo¹

Resumo - Este artigo é focado em assuntos de modelação do comportamento dinâmico e optimização baseada em modelo de processos industriais de cristalização em lotes. A estratégia das redes neuronais artificiais foi implementada como a ferramenta computacional. O objectivo é conduzir o processo ao estado óptimo de maximização do lucro e minimização dos custos. Os resultados de simulação demonstram que os objectivos conflituosos, que dizem respeito ao tempo final do processo, são simultaneamente atingíveis na presença de restrições fortes impostas ao processo.

Abstract — This paper is focused on issues of dynamic process modeling and model-based optimization of batch and fed-batch industrial crystallization processes applying the concept of artificial neural networks as computational tools. The objective is to drive the process to its optimal state of profit maximization and cost minimization. The simulation results demonstrate that the very tight and conflicting end-point objectives are simultaneously feasible in the presence of hard process constraints.

I. INTRODUCTION

The phenomenon of crystallisation occurs in a large group of pharmaceutical, biotechnological, food and chemical processes. These kind of industrial productions are usually performed in a batch or fed-batch mode which is related with the formulation of a control problem in terms of economic or performance objective at the end of the process (*end-point property control*). The crystallisation quality is evaluated by the particle size distribution (PSD) at the end of the process which is quantified by two parameters - the average (in mass) particle size (MA) and the coefficient of particle variation (CV). The main challenge of the batch production is the large batch to batch variation of the final PSD. This lack of process repeatability is caused mainly by improper control policy and results in final product recycling and loss increase.

Due to the highly competitive nature of the today's crystallization industry, model-based optimization becomes

increasingly accepted as one of the approaches that can overcome the problem of repeatability and can drive the process to its optimal state of profit maximization and cost minimization [1], [2]. However, the crystallisation occurs through the complex mechanisms of particle nucleation, subsequent particle growth and agglomeration or aggregation, phenomena that are physically not well understood therefore their reliable modelling is still a challenging task [3]. For example many of the reported crystallizer models neglect the agglomeration effect but it leads in general to biased estimation of CV and MA [4].

Development of a reliable model facilitates effectively all subsequent steps in process optimization, control and operation monitoring. There are two main modelling paradigms - analytical (based on the first principles rules) which has been the traditional way of process modelling since many years and data-driven (based on the process data) which became nowadays practically meaningful due to the rapid growth of computational resources. One of the most successful data-driven modelling techniques are the artificial neural networks (ANNs). Their ability to approximate complex non-linear relationships without prior knowledge of the model structure makes them a very attractive alternative to the classical modelling techniques [5]- [7].

The purpose of this paper is twofold. On one hand we discuss and evaluate the benefits of applying ANNs at two stages of the process control –dynamic behaviour modelling and model-based optimization. On the other hand it is an attempt to determine a systematic procedure for modelling and optimisation of batch crystallization class of processes.

II. CRYSTALLIZATION PROCESS MODELLING

A. Analytical prior knowledge approach (white box model)

The traditional way of process modelling for many years has been by mathematical equations. Since the analytical models capture physical behaviour they have the potential to extrapolate beyond the regions for which the model was constructed. The general first principles model describing a batch crystallization process consists of three parts [4].

Mass balance:

P. Georgieva is with the Department of Electronics and Telecommunications, University of Aveiro, 3810-193 Aveiro, Portugal (corresponding author, phone: +351-234-370531; fax: +351-234-370-545; e-mail: petia@det.ua.pt).

S. Feyo de Azevedo is with the Department of Chemical Engineering, Faculty of Engineering, University Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal (e-mail: sfeyo@fe.up.pt).

The mass of all participating solid and dissolved substances are included in a set of conservation mass balance equations

$$\dot{M} = f(M(t), F(t), P1), \quad 0 \leq t \leq t_f, \quad M(0) = M_0 \quad (1)$$

where $M(t) \in R^q$ and $F(t) \in R^m$ are the mass and the flow rate vectors, with q and m dimensions respectively, and t_f is the final batch time. $P1$ is the vector of physical parameters as density, viscosity, purity, ect.

Energy balance:

The general energy (E) balance model is

$$\dot{E} = f(E(t), M(t), F(t), P2), \quad 0 \leq t \leq t_f, \quad E(0) = E_0 \quad (2)$$

where $P2$ incorporates the enthalpy terms and specific heat capacities derived as functions of physical and thermodynamic properties.

Population balance:

Mathematical representation of the crystallization rate can be achieved through basic mass transfer considerations [8] or by writing a population balance represented by its moment equations [9]. Employing a population balance is generally preferred since it allows to take into account initial experimental distributions and, most significantly, to consider complex mechanisms such as those of size dispersion and/or particle agglomeration/aggregation. Hence

$$\dot{\eta}_i = f(\eta_i(t), \tilde{B}_0, G, \beta'), \quad 0 \leq t \leq t_f, \quad i = 0, 1, 2, \dots, \eta_i(0) = \eta_{i0} \quad (3)$$

where η_i is the j -th moment of the mass-size particle distribution function, $\tilde{B}_0$, G and β' are the kinetic variables nucleation rate, linear growth rate and the agglomeration kernel, respectively. The PSD measures (MA and CV) are derived from (3) as follows

$$MA = \eta_1 / \eta_0 \quad (4.1)$$

$$CV = (\eta_0 \eta_2 / \eta_1^2 - 1)^{1/2} \quad (4.2)$$

It is difficult to formulate physically based analytical models for the kinetic variables (see Fig. 1). Here, the empirical correlations have a long tradition and there exist in the literature a large number of empirical equations for them [4], [10], [11]. The decision which of them provides the best approximation of the crystallisation process in hand is very difficult.

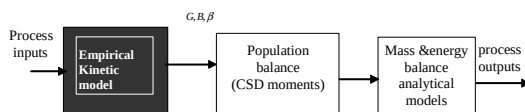


Fig 1. Analytical model

B. ANN (black box model)

Data based modelling techniques are methods that are able to extract process knowledge from measured data. The ANNs are the most celebrating data driven approach. A neural network consists of a set of elementary computing units termed *neurons*, which are combined in a parallel-serial mode. The neuron consists of a set of input signals, multiplied by respective weights and processed through an activation function to compute the neuron output. The parallel combination of a number of such units supplied with same inputs is called layer. The most typical ANN structure consists of three layers – input, output and hidden layers. In case the information (the signals) move only in one direction (from left to right) the network is termed *feedforward* neural network (FFNN). This is in contrast to the *recurrent* neural networks (RNN) having feedback connections from the output layer to the input or hidden layers. More details on the ANN concept can be found elsewhere [5].

The topology of the network is defined by the number of network inputs and outputs, the number of layers and neurons per layer and the associated number of weights and biases. While the particular choice of network inputs and outputs is physically motivated and therefore comparatively straightforward task, the choice of layer numbers and neurons per layer is less clear. It depends on the available process data, the desired model accuracy and the assumed model complexity. There is no recipe for choosing the right ANN topology. In most of the applications the ANN contains usually one or rarely two hidden layers. In general, the chosen topology is a trial and error compromise between model accuracy and complexity.

ANN training:

The ANN parameters (weights and biases) are identified by an adaptation algorithm known as network training. Among various ANN training algorithms the Backpropagation (BP) algorithm is the most widely implemented for modeling purposes. Standard BP is a gradient descent algorithm in which the network parameters are moved in the steepest descent direction i.e. along the negative of the gradient of the performance index. This is the direction in which the performance index is decreasing most rapidly. The term *backpropagation* refers to the manner in which the gradients are computed for nonlinear multilayer networks. It starts computing the gradient of the performance index with respect to the network parameters in the last (the output) layer and continuing subsequently with the previous layers. One (k) iteration of this algorithm can be written as

$$x_{k+1} = x_k - \alpha_k g_k, \quad g_k = \frac{\partial P_k}{\partial x_k} \quad (5)$$

where x_k is a vector of current network parameters, g_k is the current gradient of the performance index (P_k) and α_k is the learning rate. The steepest descent algorithm suffers of convergence problems. Though the performance index decreases most rapidly along the negative of the gradient, this does not guarantee the fastest convergence. The

conjugate gradient algorithm [12] (with many variations) and the Newton's method [13] are alternatives which overcome the problem of the optimization speed. The basic step of Newton's method is

$$x_{k+1} = x_k - H_k^{-1} g_k, \quad H_k = \frac{(\partial P_k)^2}{\partial x_k \partial x_k} \quad (6)$$

where H_k is the Hessian matrix (second derivatives) of the performance index at the current values of the ANN parameters. Newton's method often converges faster than steepest descent and conjugate gradient methods. Unfortunately, it is complex and expensive to compute the Hessian matrix for ANN. In case the performance index has the form of a sum of squares of the errors (typically in training of ANNs)

$$P_k = e_k^2, \quad e_k = y_k^{NN} - t_k$$

where y_k^{NN} , t_k and e_k are the network output, target and error vectors respectively, then there are solutions (known also as quasi-Newton methods) to approach second-order training speed without having to compute the Hessian matrix. We apply here the Levenberg-Marquardt algorithm where, at each iteration, instead of direct calculation of second derivatives, the Hessian matrix is approximated as

$$H_k = J_k^T J_k, \quad (7)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the network parameters

$$J_k = \frac{\partial e_k}{\partial x_k}, \quad (8)$$

Then the gradient is computed as

$$g_k = J_k e_k \quad (9)$$

and the Levenberg-Marquardt algorithm updates the parameters in the following way:

$$x_{k+1} = x_k [J_k^T J_k + \mu I]^{-1} J_k^T e_k \quad (10)$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift towards Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm.

An obvious advantage of the ANN modelling is its universal character in approximating different physical phenomena with similar computational structure. It saves time and efforts for identifying parameters, in contrast to the case when an analytical model is designed. Therefore ANNs are nowadays known as powerful computing structures for data processing and information storage. However, they have some remarkable disadvantages. The ANN approach suffers of the lack of transparent structure and physical understanding of the network parameters. The resulting black-box (input-output) model in general does not provide the transparency desired to enhance the process understanding. It relies only on the recorded data and does not exploit any other source of knowledge available for the process in hand.

C. Knowledge-based hybrid modeling (grey box model)

Knowledge-based hybrid modelling (KBHM) is a quite efficient alternative of the two modelling techniques discussed above [14]. The idea of KBHM is to complement the analytical model with the data-driven approach. In the design of such models it is possible to combine theoretical and experimental knowledge as well as process information from different sources: theoretical knowledge from physical and mass conservation laws; experimental data from laboratory plant experiments; experimental data from real plant experiments; data from regular process operation; knowledge and experience from qualified process operators. The clear advantages of KBHM compared with the data-based modelling are first with respect to more physical transparency of the model parameters and secondly less training data is required [15].

Our solution for a KBHM of crystallization processes combines a partial analytical model reflecting the mass, energy and population balances (1-3) with an ANN for modelling the crystal growth, nucleation rate and the agglomeration kernel (see Fig. 2). The ANN parameters were tuned applying the Levenberg-Marquardt optimisation procedure described above.

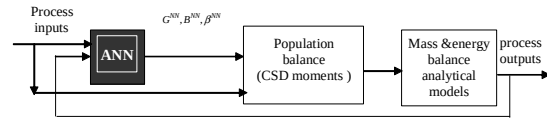


Fig. 2 KBHM

III. ANN PARAMETERISED END-POINT OPTIMIZATION

End-point process optimization, i.e. achievement of the best outcome of a finite end-time process through appropriate manipulation of its input variables is the main control concern of batch applications. For example, in a crystallization process the final time values of the quality parameters MA , and CV are common properties of interest

for optimization. The optimization problem can be mathematically formulated as:

$$\max_{u_{\min} \leq u(t) \leq u_{\max}} J_f = \varphi(x(t_f), P), \quad (11)$$

subject to:

$$\dot{x} = f(x(t), u(t), P), \quad 0 \leq t \leq t_f, \quad x(0) = x_0 \quad (12.1)$$

$$y(t) = h(x(t), P) \quad (12.2)$$

$$g_j(x) = 0, \quad j = 1, 2, \dots, p \quad (13.1)$$

$$v_j(x) < 0, \quad j = 1, 2, \dots, l \quad (13.2)$$

where (11) is the performance index, (12) is the process model, function f is the state-space description, function h is the relationship between the output and the state and P is the parameter vector. $x(t) \in R^n$, $u(t) \in R^m$ and $y(t) \in R^p$ are the state, the manipulated input, also known as the control decision variable, and the control output vectors, respectively. The manipulated inputs, the state and the control outputs are subject to the following constraints, $x(t) \in X$, $u(t) \in Z$, $y(t) \in Y$ in which X , Z and Y are convex and closed subsets of R^n , R^m and R^p . g_j and v_j are the equality and inequality constraints with p and l dimensions respectively.

We apply here the practical approach of reformulating the optimization problem (11) and the process constraints (13) in an unified structure through the use of a penalty function in the performance index. It is a simple and intuitive way to include the constraints in the optimization procedure as an extra term in the performance function

$$\max_{u_{\min} \leq u(t) \leq u_{\max}} J = \alpha_1 J_f + \alpha_2 \sum_{j=1}^p (g_j)^2 - \alpha_3 \sum_{j=1}^l [\max(0, v_j)]^2 \quad (14)$$

where g_j and v_j are the constraints defined in (13) and α_i , $i = 1, 2, 3$ are weighting factors accounting for the relative importance of each component of (14).

Several methods can tackle the optimisation problem (14). The derivative based deterministic methods are rather sensitive to the model initial conditions and do not guarantee a global optimum but they are faster than the stochastic methods and are usually more appropriate to on-line optimisation. Since the aim was to find *off-line* the optimal profiles of the manipulated inputs that lead to maximisation of the process performance index at the end of the process, the stochastic approach appears to be more efficient solution. In particular, the evolutionary programming was considered as the method with less sensitivity to the scaling of the

multidimensional performance index and good convergence approaching the optimum [16], [17].

To relax the numerical procedure the optimization variables are usually parameterised as piece wise constant, linear or polynomial functions. However, it means less freedom in determining their final values and leads to suboptimal profiles. To deal with this problem we express each optimization variable as a general nonlinear time function [18]

$$F_i(t) = f_i(a_1, a_2, \dots, a_k, t), \quad i = \{s, f\}, \quad (15)$$

where t is the time and $a_1, a_2, \dots, a_k$ are parameters to determine. Then employ an ANN with a single layer and radial-basis functions (RBF) as the activation units to approximate (15)

$$f_i(a_1, a_2, \dots, a_k, t) = \sum_{j=1}^k w_j \Phi_j, \quad i = \{s, f\} \quad (16)$$

where w_j are the ANN weights and Φ_j are Gaussian functions, determining bell shaped relationships:

$$\Phi_j(t) = \exp\left(-\frac{(t-c_j)^2}{\rho_j^2}\right). \quad (17)$$

Note that c_j is the equidistantly divided time grid for RBF and ρ_j define the shape of RBF.

IV. CASE-STUDY – BATCH EVAPORATIVE SUGAR CRYSTALLIZATION PROCESS

A. Process Operation

Crystallisation occurs through the mechanisms of nucleation, growth and agglomeration. The process is characterised by strongly non-linear and non-stationary dynamics and can be divided into several sequential phases.

Charging. During the first phase the pan is partially filled with a juice containing dissolved sucrose (termed liquor).

Concentration. The next phase is the concentration. The liquor is concentrated by evaporation, under vacuum, until the supersaturation reaches a predefined value. At this stage seed crystals are introduced into the pan to induce the production of crystals. This is the beginning of the third (crystallisation) phase.

Crystallisation (main phase). In this phase as evaporation takes place further liquor or water is added to the pan in order to guarantee crystal growth at a controlled supersaturation level and to increase total contents of sugar in the pan. In most cases, due to economical reasons, the liquor is replaced by other juice of lower purity (termed syrup).

Tightening. The fourth phase consists of tightening which is principally controlled by the evaporation capacity. The

pan is filled with a suspension of sugar crystals in heavy syrup, which is dropped into a storage mixer. At the end of the batch, the final massecuite undergoes centrifugation, where final refined sugar is separated from the (mother) liquor.

The unit contains 15 sensors for the following properties and operating variables: *i) inside the pan* - massecuite temperatures at three locations; brix of solution; level; massecuite consistency; stirrer current; vacuum pressure and temperature. *ii) feed conditions* - temperature, brix and flow rate of feed liquor and feed syrup. *iii) steam conditions* - temperature, pressure and flow rate of steam.

Brix is the concentration of total dissolved solids (sucrose plus impurities) in the solution. Supersaturation is not a measured variable but can be determined from the available measurements. The feed flow rates of sugar liquor/syrup and the steam supply are considered as process inputs. The final crystal contents and the end-point PSD characterise the product quality. More details about the process can be found elsewhere [4], [11].

B. Knowledge-based hybrid model (analytical+ ANN)

Based on the available process measurements a detailed first-principles model was developed and identified following the general structure (1-4). Due to the lack of on-line measurements of the PSD, the states related with the population balance (3) are not estimable and it makes the model not suitable for optimization or control purposes. To overcome this problem a KBHM was obtained according to the strategy presented in section II. Short description of the model follows below.

Mass balance:

The mass of water (M_w), impurities (M_i), dissolved sucrose (M_s) and crystals (M_c) are included in the following set of conservation mass balance equations

$$\frac{dM_w}{dt} = F_f \rho_f (1 - B_f) - J_{vap} \quad (18)$$

$$\frac{dM_i}{dt} = F_f \rho_f B_f (1 - Pur_f) \quad (19)$$

$$\frac{dM_s}{dt} = F_f \rho_f B_f Pur_f - J_{cris} \quad (20)$$

$$\frac{dM_c}{dt} = J_{cris} \quad (21)$$

where Pur_f , B_f and ρ_f are the purity (mass fraction of sucrose in the dissolved solids), brix and the density of the incoming feed. J_{cris} is the crystallisation rate and F_f is the feed flowrate considered as the process input.

Energy balance

$$\frac{dT_m}{dt} = aJ_{cris} + bF_f + cJ_{vap}(F_s) + d \quad (22)$$

where a , b , c , d are the components of parameter vector $P1$ and J_{vap} is the evaporation rate which is a function of the steam supply rate (F_s), considered as the second process input.

Population balance (in volume coordinates):

The kinetics mechanisms of nucleation, crystal growth and particle agglomeration are defined by the population balance. The population balance is expressed by the leading moments of PSD in volume coordinates ($\tilde{\mu}_i$) since agglomeration must obey mass conservation law,

$$\frac{d\tilde{\mu}_0}{dt} = \tilde{B}_0 - \frac{1}{2} \beta' \tilde{\mu}_0^2 \quad (23)$$

$$\frac{d\tilde{\mu}_1}{dt} = G_v \tilde{\mu}_0 \quad (24)$$

$$\frac{d\tilde{\mu}_2}{dt} = 2G_v + \beta' \tilde{\mu}_1^2 \quad (25)$$

$$\frac{d\tilde{\mu}_3}{dt} = 3G_v \tilde{\mu}_2 + 3\beta' \tilde{\mu}_1 \tilde{\mu}_2 \quad (26)$$

The crystallisation rate is determined as

$$J_{cris} = \rho_c \frac{d\tilde{\mu}_1}{dt}$$

The nucleation rate ($\tilde{B}_0$), the growth rate (G_v) and the agglomeration kernel (β') are replaced by a feed-forward ANN with 4 inputs, 3 outputs and one hidden layer with 9 sigmoid activation functions. The temperature of massecuite (T_m), the supersaturation (S), the purity of the solution (Pur_{sol}) and the volume fraction of crystals (v_c) are considered as the networks inputs because they all affect directly the kinetic parameters.

Hybrid ANN training – sensitivity approach:

The training of an ANN requires that the network weights are determined in such a way that the error between the network output and the corresponding target output becomes minimal. In the hybrid system, however, the target outputs are not available since the kinetic parameters are not measured. Therefore, a new training procedure was developed. Our solution was to build a hybrid ANN training structure where the network outputs go through some fixed (known) part of the analytical model and to compare this hybrid model output with the available data (Fig. 3).

The error for updating the network weights is a function of the observed error and the gradient of the hybrid model output with respect to the ANN output. The mass of crystals is considered as most appropriate to serve as a target output in the hybrid ANN training.

According to equations (21), (23), the mass balance of crystals can be rewritten as

$$\frac{dM_c^{hyb}}{dt} = f_2(\tilde{\mu}_0^{hyb}, M_c^{hyb}, G^{NN}) \quad (27)$$

(27) is incorporated in the hybrid training structure but in order to integrate it $\tilde{\mu}_0$ is required. Therefore its balance equation is also involved in the network training stage,

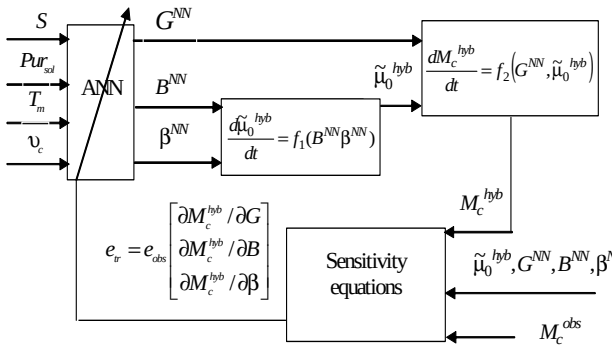


Fig.3: Hybrid ANN training procedure

$$\frac{d\tilde{\mu}_0^{hyb}}{dt} = B^{NN} - \frac{1}{2} \beta^{NN} \left(\tilde{\mu}_0^{hyb} \right)^2. \quad (28)$$

Superscripts *hyb* and *NN* are used to point out variables obtained during the hybrid network training. The network outputs give estimates of the growth rate, nucleation and agglomeration kinetic parameters. These estimates are propagated through (27-28). The error signal for updating the network parameters is

$$e_{tr} = e_{obs} [\lambda_G \quad \lambda_B \quad \lambda_\beta]^T \quad (29)$$

It is obtained by multiplying the observed error

$$e_{obs} = M_c^{obs} - M_c^{hyb} \quad (30)$$

with the gradient of the hybrid model output with respect to the network outputs

$$\lambda_G = \frac{\partial M_c^{hyb}}{\partial G} \quad (31)$$

$$\lambda_B = \frac{\partial M_c^{hyb}}{\partial B} \quad (32)$$

$$\lambda_\beta = \frac{\partial M_c^{hyb}}{\partial \beta}. \quad (33)$$

The gradients (31-33) can be computed through integration of the sensitivity equations

$$\frac{d\lambda_G}{dt} = \frac{\partial f_2}{\partial M_c^{hyb}} \lambda_B + \frac{\partial f_2}{\partial G}, \quad \lambda_G(0) = 0 \quad (34)$$

$$\frac{d\lambda_B}{dt} = \frac{\partial f_2}{\partial M_c^{hyb}} \lambda_B + \frac{\partial f_2}{\partial \tilde{\mu}_0^{hyb}} \left(\frac{\partial \tilde{\mu}_0^{hyb}}{\partial B} \right), \quad \lambda_B(0) = 0 \quad (35)$$

$$\frac{d\lambda_\beta}{dt} = \frac{\partial f_2}{\partial M_c^{hyb}} \lambda_\beta + \frac{\partial f_2}{\partial \tilde{\mu}_0^{hyb}} \frac{\partial \tilde{\mu}_0^{hyb}}{\partial \beta}, \quad \lambda_\beta(0) = 0, \quad (36)$$

Note, that while λ_G can be straightforwardly obtained, λ_B and λ_β depend on the gradients of $\tilde{\mu}_0$ with respect to $\tilde{B}_0$ and β , respectively. In order to determine them the same strategy is applied leading to integration of the following sensitivity equations with zero initial conditions

$$\frac{d\chi_B}{dt} = \frac{\partial f_1}{\partial \tilde{\mu}_0} \chi_B + \frac{\partial f_1}{\partial B}, \quad \chi_B(0) = 0 \quad (37)$$

$$\frac{d\chi_\beta}{dt} = \frac{\partial f_1}{\partial \tilde{\mu}_0} \chi_\beta + \frac{\partial f_1}{\partial \beta}, \quad \chi_\beta(0) = 0, \quad (38)$$

where $f_1 = B^{NN} - \frac{1}{2} \beta^{NN} \left(\tilde{\mu}_0^{hyb} \right)^2$, $\chi_B = \frac{\partial \tilde{\mu}_0}{\partial B}$, and $\chi_\beta = \frac{\partial \tilde{\mu}_0}{\partial \beta}$.

C. KBHM-based optimization procedure

Process performance index: The crystallizer performance index J_f (related to the final time objectives) has several components. The first objective is to achieve crystals with a desired final size, which is quantified by the quality variable *MA*. It is practically more relevant instead of defining a fixed end-setpoint for *MA* to choose a tight zone around the desired value. For the process considered these are $MA_{min} = 0.55$ and $MA_{max} = 0.60$. Therefore

$$MA_{min} \leq MA(t_f) \leq MA_{max} \quad (39)$$

The second objective is reducing the quality variable CV as much as possible but in practice CV less than a predefined maximum value is enough for good performance. Then

$$CV(t_f) \leq CV_{\max}, \text{ with } CV_{\max} = 30\% \quad (40)$$

To guarantee sufficient efficiency of the production, the crystals should occupy a certain minimum volume of the pan. This objective is quantified by the crystal content (wc)

$$wc_{\min} \leq wc(t_f), \text{ with } wc_{\min} = 50\% \quad (41)$$

The main process constraints are related to the supersaturation and the volume of the pan during the batch. In case the supersaturation is below a minimum value, the crystals start dissolving and if the supersaturation is above a maximum value, undesired secondary nucleation takes place. Therefore

$$S_{\min} \leq S(t) \leq S_{\max} \quad (42)$$

where for the process in hand reasonable limiting values are $S_{\min} = 1.02$, $S_{\max} = 1.25$. The total volume constrain is determined by the physical dimension of the pan. Hence

$$V(t) \leq V_{\max}, \text{ with } V_{\max} = 35 \quad (43)$$

The decision variables of the optimisation (the process inputs) are the feeding rate (F_f) and the steam supply rate (F_s). The optimal switching time between liquor and syrup supply t_{syr} was also considered as an optimisation parameter. Due to technological limits of the process equipment, F_f and F_s are limited which is considered as a hard constrain of the optimisation procedure. The input constraints are summarised in Table 1.

TABLE 1. INPUT MIN AND MAX VALUES

$F_{s\min}$	1.1 [kg/s]
$F_{s\max}$	2.1 [kg/s]
$F_{f\min}$	0.0 [m ³ /s]
$F_{f\max}$	0.015 [m ³ /s]

The final process time is fixed to 90min. Taking into account (39-43), the general multi-objective optimization performance index in (14) has the following particular structure:

$$\begin{aligned} J = & -\alpha_1 [\max(0, AM_{\min} - AM(t_f))]^2 - \\ & \alpha_2 [\max(0, AM(t_f) - AM_{\max})]^2 - \\ & \alpha_3 [\max(0, CV(t_f) - CV_{\max})]^2 - \\ & \alpha_4 [\max(0, wc_{\min} - wc(t_f))]^2 - \\ & \alpha_5 [\max(0, S_{\min} - S(t))]^2 - \\ & \alpha_6 [\max(0, S(t) - S_{\max})]^2 - \\ & \alpha_7 [\max(0, V(t) - V_{\max})]^2. \end{aligned} \quad (44)$$

and the optimisation problem (11) can be stated as follows:

$$\max_{\{F_{\min} \leq F_f(t) \leq F_{\max}; F_{s\min} \leq F_s(t) \leq F_{s\max}; t_{\text{syr}}\}} J \quad (45)$$

subject to: equations (18-28)

The iterative optimisation procedure can be summarized in the following steps:

- i) Assignment of initial values of the process states, initial values of all parameters (w_j) subject to optimization (16) and the switching time between liquor and syrup, t_{syr} .
- ii) Computation of the manipulated inputs (15) by solving (16) and (17) for each variable.
- iii) The tentative inputs are propagated through the KBHM (see Fig. 2) and the values of the components of the performance index are obtained (CV, MA, V, S, wc).
- iv) The overall performance index (44) is computed. In case, an improvement with the previous iteration is registered, the evolutionary programming technique generates a new set of w_j and t_{syr} . The procedure repeats starting from the step 2.
- v) If no improvements of the performance index (44) is achieved within a predefined iteration number or the relative iteration-to-iteration change is insignificant the procedure is stopped and the final optimal profiles are generated.

V. SIMULATION RESULTS

The simulation results are summarised in Figs. 4-7. The optimal profiles of the steam supply rate and the feeding rate of liquor/syrup with the respective switching time between liquor and syrup are depicted in Fig. 4. Based on these optimal input profiles (determined by the optimization procedure), the optimal trajectories for the process outputs are estimated by the model. The optimal profiles of the

supersaturation (Fig.5a), the brix (Fig.5b) and the massecuite temperature (Fig.5c) can be used then as setpoints in a feedback control framework.

Fig. 6 shows the main process quality variables along the batch. Though the PSD objectives are related only with the final values, the smooth behaviour of MA (Fig.6a) and CV (Fig.6b) contribute to a higher process internal performance. Note that the MA final value (0.6mm) is within the margins defined by (39) and the CV= 28.2% is also less than the upper limit defined by (40). Moreover it is much less than the average values of CV (37%-39%) obtained in the real plant production [11]. The third objective, quantified by (41), is also satisfied. The crystal content (Fig.7a) at the process end was 57% of the total volume (Fig. 7b). The main process constraints related to the supersaturation and the total volume (42-43) remain within the predefined limits. $V(t_f)=29\text{m}^3$ corresponds to approximately 90% utilisation of the working volume which is considered as a reasonable compromise between productivity and safeness.

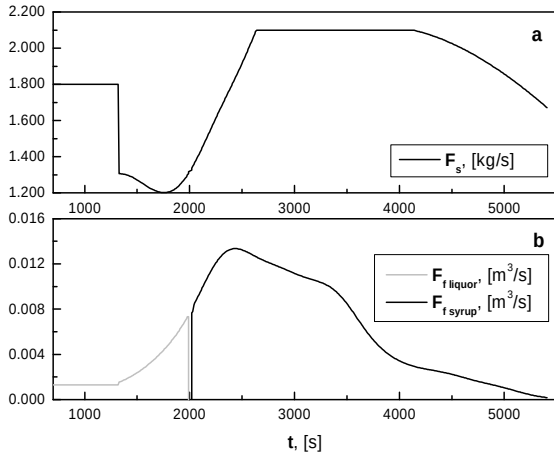


Fig.4. Optimised steam supply (a) and feeding rate (b) profiles, and the optimal switching time.

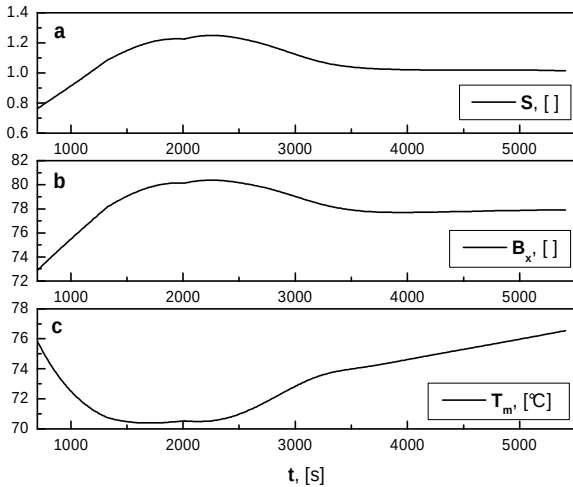


Fig. 5. Supersaturation (a), brix (b) and temperature (c) profiles resulting from the optimised control strategy.

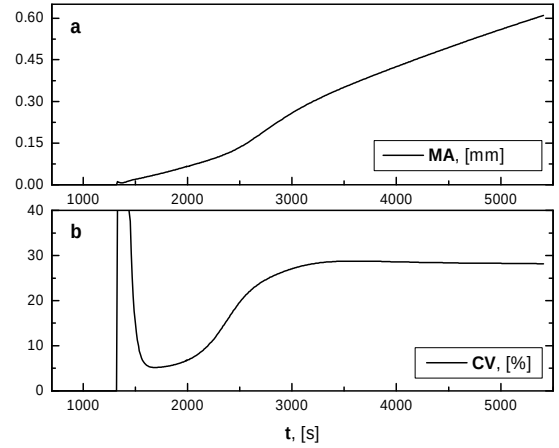


Fig. 6. Average (in mass) crystal size MA (a) and coefficient of variation CV (b) resulting from the optimised control strategy.

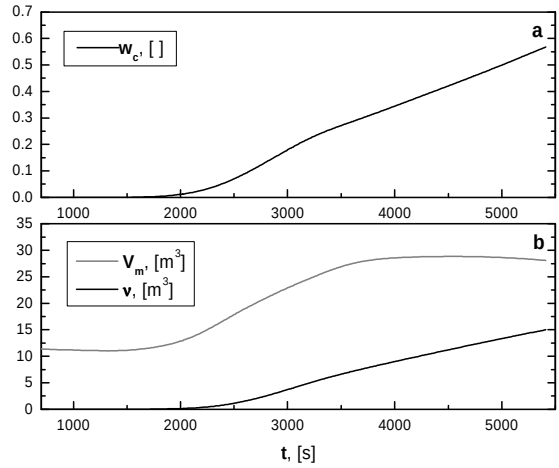


Fig. 7. Crystal content (a), the total (massecuite) volume and the crystal volume (b) resulting from the optimised control strategy.

VI. CONCLUSIONS

The application of ANN at two stages of batch crystallization process automation, namely modelling and optimisation, is presented.

At the modelling stage, a knowledge based hybrid model (KBHM) of the process was designed that possesses the advantages of both analytical and pure data based process models. The KBHM offers a reasonable compromise between the extensive efforts to get a fully parameterised structure, as are the analytical models and the poor generalisation of the complete data-based modelling approaches.

Optimisation strategy based on the KBHM model was proposed and the simulations show that the very tight and conflicting end-point objectives are simultaneously feasible in the presence of hard process constraints. Moreover it led to a significant improvement in the CV measure of the industrial sugar crystallisation process as compared to statistically averaged value of CV achieved by the industrial

data.

For successful implementation of the optimised control strategy in practice, accurate tracking of the optimised profiles is required. However, the presence of inevitable disturbances occurring in the process variables like brix, purity and temperature of feeding solution, vacuum pressure or steam temperature can make the manipulated inputs not optimal any more. A closed loop control is usually the most effective solution where an on-line input correction is performed based on the current measurements. These issues are not treated in this paper but work on them is now in progress [19].

ACKNOWLEDGMENT

This work was pursued in the framework of EC RTN Project BatchPro HPRN-CT-2000-0039. It was further financed by the Portuguese Foundation for Science and Technology within the activity of the Research Unit IEETA-Aveiro, which is gratefully acknowledged.

REFERENCES

- [1] J. B. Rawlings, S. M. Miller, and W. R. Witkowski, "Model identification and control of solution crystallization processes: a review", *Ind. Eng. Chem. Res.*, 32(7):1275-1296, July 1993.
- [2] Nagy Z. K. and R. D. Braatz, "Robust nonlinear model predictive control of batch processes", *AIChE J.*, 49:1776-1786, 2003.
- [3] M. Fujiwara, Z. K. Nagy, J. W. Chew, and R. D. Braatz, "First-principles and direct design approaches for the control of pharmaceutical crystallization", *Journal of Process Control*, 15:493-504, 2005 (invited).
- [4] S. Feyer de Azevedo, J. Chorão, M. J. Gonçalves and L. Bento, "On-line Monitoring of White Sugar Crystallization through Software Sensors", *Int. Sugar JNL.*, 96:18-26, 1994.
- [5] S. Haykin, *Neural Networks: A comprehensive foundation*, Prentice Hall, UK, 1994.
- [6] N.T. Russell and H.H.C. Bakker, "Modular modelling of an evaporator for long-range prediction" *Artificial Intelligence in Engineering*, 11, 347-355, 1997.
- [7] L.F.M. Zorzetto, R. Maciel Filho and M. R. Wolf-Maciel, "Process modelling development through artificial neural networks and hybrid models", *Computers and Chemical Engineering*, 24, 1355-1360, 2000.
- [8] P. Dittl, L. Beranek and Z. Rieger, "Simulation of a stirred sugar boiling pan", *Zuckerind*, 115, 667-676, 1990.
- [9] A.D. Randolph and M.A. Larson, *Theory of Particulate Processes – Analyses and Techniques of Continuous Crystallization*, Academic Press, 1988.
- [10] P. Lauret, H. Boyer, and J.C. Gatina, "Hybrid modelling of a sugar boiling process", *Control Engineering Practice*, 8, 299-310, 2000.
- [11] P. Georgieva, M. J. Meireles and S. Feyer de Azevedo, "Knowledge based hybrid model of a fed-batch sugar crystallization when accounting for nucleation, growth and agglomeration phenomena", *Chem Eng Sci*, 58:3699-3711, 2003.
- [12] M. T. Hagan, H. B. Demuth and M. H. Beale, *Neural Network Design*, Boston, MA: PWS Publishing, 1996.
- [13] M. T. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no.6, pp. 989–993, 1994.
- [14] D.C. Psychogios and L. H. Ungar, "A hybrid neural network - first principles approach to process modeling", *AIChE J.*, 38(10), 1499-1511, 1992.
- [15] J. Graefe, P. Bogaerts, J. Castillo, M. Cherlet, J. Werenne, P. Marenbach and R. Hanus, "A new training method for hybrid models of bioprocesses", *Bioprocess Engineering*, 21, 423-429, 1999.
- [16] D.J. Fogel, *Evolutionary Computation: toward a new philosophy of machine intelligence*. New York, IEEE Press, 1995.
- [17] Q.T. Pham, "Dynamic Optimization of Chemical Engineering Processes by an Evolutionary Method", *Computers and Chemical Engineering*, 22:1089-1097, 1998.
- [18] V. Galvanauskas, R. Simutis, N. Volk, and A. Lübbert, "Model Based Design of a Biochemical Cultivation Process", *Bioprocess and Biosystems Engineering*, 18(3):227-234, 1998.
- [19] P. Georgieva and S. Feyer de Azevedo, "Application of feed forward neural networks in modeling and closed loop control of a fed-batch crystallization process", accepted for publication at the International Conference on Neural Networks, Vienna, Austria, 29-31 March 2006.