Preferred citation: G.A. Dumont. System identification and adaptive control in papermaking. In **Fundamentals of Papermaking**, *Trans. of the IXth Fund. Res. Symp. Cambridge, 1989*, (C.F. Baker & V. Punton, eds), pp 1151–1181, FRC, Manchester, 2018. DOI: 10.15376/frc.1989.2.1151.

SYSTEM IDENTIFICATION AND ADAPTIVE CONTROL IN PAPERMAKING

Guy A. Dumont
Pulp and Paper Research Institute of Canada and
Department of Electrical Engineering, Pulp and Paper Centre,
University of British Columbia, Vancouver, B.C., Canada, V6T 1W5.

Abstract

The concepts and methodology of system identification and adaptive control in papermaking are presented and discussed. In particular the crucial problems of deadtime compensation and proper interpretation of scanning gauge measurements are reviewed. Two new approaches developed recently in our laboratory to overcome these problems are presented.

The key points in extending the monovariable adaptive control schemes to the multivariable case are discussed. Successful applications of self-tuning control to machine and cross-directional control of moisture and basis weight are outlined. The difficulties associated with multivariable adaptive control are illustrated by an example of colour control on a paper machine.

Finally, novel approaches to adaptive control such as genetic algorithms, neural networks, and multimodel techniques are presented. The promise of these recent advances in adaptive control is illustrated by an example of automatic compensation for species changes in kraft pulping.

INTRODUCTION

The paper machine is by far the most automated process in the pulp and paper manufacturing process. Sophisticated sensors are available to measure various properties of the sheet, ranging from simple basis weight and moisture to physical properties such as strength. Many modern identification and control methods were given their first industrial test on a paper machine (1). Unfortunately, this pioneering work has had little influence on the industrial practice of paper machine control (2). Thus, most functions on paper machines are still controlled using very simple control algorithms.

Today the amount of information available through on-line sensors is such that the design of a control system becomes a formidable task. Indeed, as more variables are measured, and thus can be controlled, it becomes more complex and difficult to model the interactions between them. Moreover, those interactions are subject to change, for example as the feedstock characteristics change. A key question is therefore: how can we control such a multivariable, nonlinear and time-varying system? Do the techniques even exist?

Today's control theory can provide partial solutions. Some promising developments may soon provide more complete answers. The characterization of unknown process relationships is the realm of system identification, while the control of unknown and time-varying system is the main goal of adaptive control research. In this paper we briefly review the concepts and methodology of system identification and adaptive control. We then discuss current and potential applications of those techniques to papermaking control problems. Finally, we discuss three novel approaches that may prove particularly appropriate to solve the multivariable nonlinear problem mentioned above.

SYSTEM IDENTIFICATION

Identification Methods

Most modern control techniques are model-based, *i.e.* they require a model of the process dynamics. Very often, modelling on the basis of first principles is not possible, or is not sufficiently reliable. The alternative is then

to perturb the process inputs, and based on the observed behaviour, to fit a simplified linear model (e.g. a transfer function) that best describes that behaviour. This is the field of process or system identification. See Ljung (3) for a comprehensive treatment of the topic.

Within the control engineering community, the most popular identification method is the least-squares method. Let a dynamic system be described by

$$y(t) = \sum_{i=1}^{n} -a_i y(t-i) + \sum_{i=1}^{n} b_i u(t-i) + w(t) \tag{1}$$

where t denotes discrete time, $u(t)$ and $y(t)$ are respectively the input and ouput of the plant and $w(t)$ is the process noise. Defining

$$\underline{\theta}^T = [\ a_1 \ \cdots \ a_n \ \ b_1 \ \cdots \ b_n]$$

$$\underline{x}^T = [\ -y(t) \ \cdots \ -y(t-n) \ \ u(t-1) \ \cdots \ u(t-n)]$$

the system above can be written as

$$y(t) = \underline{x}^T(t)\underline{\theta} + w(t) \tag{2}$$

With N observations, the data can be put in the following compact form:

$$\underline{Y} = X\underline{\theta} + \underline{W} \tag{3}$$

where

$$\underline{Y}^T = [\ y(1) \ \cdots \ y(N)]$$

$$\underline{W}^T = [\ w(1) \ \cdots \ w(N)]$$

$$X = \begin{bmatrix} \underline{x}^T(1) \\ \vdots \\ \underline{x}^T(N) \end{bmatrix}$$

If for each point we define the modelling error as

$$\epsilon(t) = y(t) - \underline{x}^T(t)\underline{\theta} \tag{4}$$

then the least-squares performance index to be minimized is:

$$J = \sum_{1}^{N} \epsilon^2(t) \tag{5}$$

or using the matrix notation above

$$J = [\underline{Y} - X\underline{\theta}]^T[\underline{Y} - X\underline{\theta}] \tag{6}$$

Differentiating with respect to $\underline{\theta}$ and equating to zero then yields $\underline{\hat{\theta}}$, the least-squares estimate of $\underline{\theta}$:

$$\underline{\hat{\theta}} = [X^T X]^{-1} X^T \underline{Y} \tag{7}$$

Unfortunately, the least-squares estimate is unbiased, *i.e.* $E(\underline{\hat{\theta}}) = \underline{\theta}$ only in two cases:

- if the noise sequence $w(t)$ is uncorrelated and has zero-mean, or
- if $w(t)$ is independent of $u(t)$ and the model is a moving average, i.e. no a term is estimated.

In practice, process disturbances are rarely totally random, but are correlated. Thus, despite its simplicity the least-squares method will in general not be used because it gives biased estimates in the presence of correlated process noise. Alternative methods such as the maximum-likelihood method must be used.

The maximum-likelihood method considers the model below where u is the input, y the output and e is zero-mean white noise with standard deviation σ:

$$y(t) = -\sum_{i=1}^{n} a_i y(t-i) + \sum_{i=1}^{n} b_i u(t-k-i) + e(t) + \sum_{i=1}^{n} c_i e(t-i) \tag{8}$$

where the parameters of a_i, b_i, c_i as well as σ are unknown. Defining

$$\underline{\theta}^T = [\ a_1 \ \cdots \ a_n \ \ b_1 \ \cdots \ b_n \ \ c_1 \ \cdots \ c_n]$$

$$\underline{x}^T = [\ -y(t) \ \cdots \ -y(t-n) \ \ u(t-k) \ \cdots \ u(t-k-n) \ \ e(t-1) \ \cdots \ e(t-n)]$$

the above model can be written as

$$y(t) = \underline{x}^T(t)\underline{\theta} + e(t) \tag{9}$$

Unfortunately, one cannot use the least-squares method on this model since the sequence $e(t)$ is unknown. In case of known parameters, the past values of $e(t)$ can be reconstructed from the sequence:

$$\epsilon(t) = y(t) + \sum_{i=1}^{n} a_i y(t-i) - \sum_{i=1}^{n} b_i u(t-k-i) - \sum_{i=1}^{n} c_i \epsilon(t-i)$$

In case of unknown parameters, $\epsilon(t)$ can be interpreted as a model error equal to $e(t)$ only when the parameters assume their true values. The maximum-likelihood method then minimizes the performance index below:

$$V = \frac{1}{2} \sum_{t=1}^{N} \epsilon^2(t)$$

Because there is no closed-form solution to this minimization problem, a numerical procedure has to be used. It can be shown that the maximum likelihood estimate is the most accurate unbiased estimate available. Hence, the maximum-likelihood method is the preferred identification technique, and is available in most software packages offering control system design and analysis capabilities.

Recursive Identification

There are many situations when it is preferable to perform the identification on-line, such as in adaptive control. In this case, the previous identification methods need to be implemented in a recursive fashion, i.e. the parameter estimate at time t should be computed as a function of the estimate at time $t-1$ and of the incoming information at time t. The field of recursive identification has been the subject of intensive research in the last two decades. For an in-depth look at this field, the reader is referred to the book by Ljung and Söderström (4) on the topic.

The recursive least-squares (RLS) algorithm is given by the equations below:

$$\hat{\underline{\theta}}(t+1) = \hat{\underline{\theta}}(t) + \frac{P(t)\underline{x}(t+1)}{1 + \underline{x}^T(t+1)P(t)\underline{x}(t+1)}[y(t+1) - \underline{x}^T(t+1)\hat{\underline{\theta}}(t)] \tag{10}$$

$$P(t+1) = P(t) - \frac{P(t)\underline{x}(t+1)\underline{x}^T(t+1)P(t)}{1+\underline{x}^T(t+1)P(t)\underline{x}(t+1)} \tag{11}$$

The recursive least-squares algorithm is the exact mathematical equivalent of the batch least-squares algorithm , and thus has the same properties. The matrix P is proportional to the covariance matrix of the estimate, and is thus called the covariance matrix. The algorithm has to be started with initial values for $\underline{\hat{\theta}}(0)$ and $P(0)$. Generally, $P(0)$ is set to αI where I is the identity matrix and α is a large positive number. The larger α is, the less confidence is put in the initial estimate $\underline{\hat{\theta}}(0)$.

Whereas it is possible to develop an exact recursive version of the least-squares algorithm, it is not possible to do so for the maximum-likelihood method. Instead, one has to develop approximations. The approximation that has been shown to be best, both from theory and from simulation studies, is the approximate maximum-likelihood method (AML) where $e(t)$ is approximated by the residual $\eta(t)$ defined as

$$\eta(t) = y(t) - \hat{y}(t|t) = y(t) - \underline{x}^T(t)\underline{\hat{\theta}}(t)$$

Defining the $\underline{\theta}$ vector and the $\underline{x}$ vector as in:

$$\underline{\theta}^T = [\ a_1 \ \cdots \ a_n \ \ b_1 \ \cdots \ b_n \ \ c_1 \ \cdots \ c_n]$$

$$\underline{x}^T = [\ -y(t) \ \cdots \ -y(t-n) \ \ u(t-k) \ \cdots \ u(t-k-n) \ \ \eta(t-1) \ \cdots \ \eta(t-n)]$$

then results in a scheme that is described by the same equations as the recursive least-squares above. The AML estimate can be shown to converge to the true parameter values under fairly weak conditions.

ADAPTIVE CONTROL

Monovariable Adaptive Control

Adaptive control is a complex and still somewhat immature field, despite more than thirty years of research which became particularly intense after the seminal 1973 paper by Åström and Wittenmark (5). For a comprehensive view of the current state of the art in the field, readers are referred to the book by Åström and Wittenmark (6).

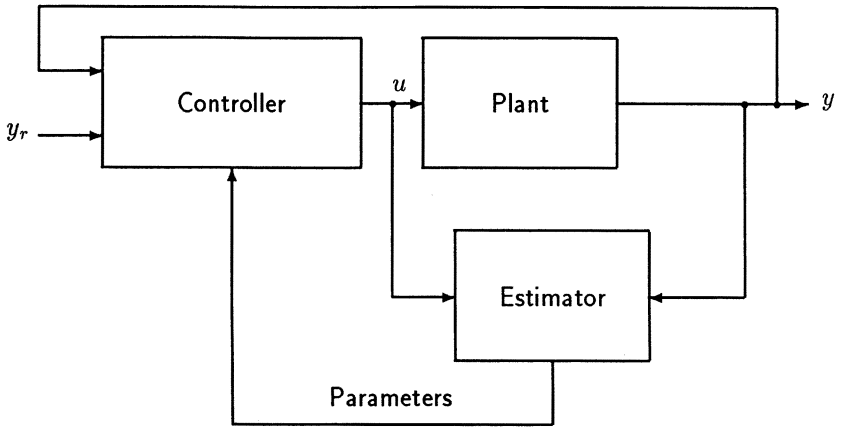

Fig 1– Block diagram of an adaptive controller

When a control engineer wants to develop a control scheme for a particular process, one of the first tasks to be accomplished is the identification of the process dynamics. The second major task is then to design the controller that will achieve the desired control objective. Adaptive control can be thought of as an automation of that design procedure, *i.e.* on-line identification and control design. A typical block diagram of such an adaptive controller is shown on Figure 1. A major difference is that the entire procedure is performed online, in real-time and without human supervision. Because of this lack of human supervision, extra care has to be taken when implementing such a scheme. One of the current research trends is to develop an emulation of that human supervisor using an expert system overlooking both the implementation and the operation of the adaptive controller. Such a system would greatly improve the reliability of current adaptive control schemes. However, a major drawback with current adaptive control methods is that they require prior knowledge of the structure of the plant dynamics, *i.e* of the order and dead time of the process transfer function. Wrong assumptions may lead to instability, obviously an undesirable feature.

One way to design a robust adaptive control requiring minimal a priori

information and capable of handling time-delay plants (common in process control) is to abandon the usual transfer function models and instead develop an unstructured adaptive control scheme using an orthonormal series representation. This is the approach recently developed in our laboratory (7), (8). The set of Laguerre functions is particularly appealing because it is simple to represent and is similar to transient signals. It also closely resembles Padé approximants. The continuous Laguerre functions, can be represented by the simple and convenient ladder network shown in Figure 2, where s is the Laplace transform operator. The output of the plant $y(t)$ is described by,

$$y(t) = \sum_{i=1}^{N} c_i l_i(t) + w(t) = \underline{c}_0^T \underline{l}(t) + w(t) \tag{12}$$

where $\underline{c}_0^T = [c_1 \quad c_2 \ldots c_N]$, $\underline{l}^T(t) = [\ l_1(t) \quad l_2(t) \quad \ldots \quad l_N(t)\ ]^T$, where the l_i's are the outputs from each block in Figure 1, and $w(t)$ is the process noise. Some of advantages in using the above series representation are that, (a) because of its resemblance to the Padé approximants, time-delays can be very well represented as part of the plant dynamics and, (b) the model order N does not theoretically affect the coefficients c_i. It can be shown that under fairly weak conditions, $\hat{\underline{c}}$, the least-squares estimate of $\underline{c}$ will be unbiased.

An adaptive control scheme based on the above formulation uses the recursive least-squares (RLS) method to identify the parameter vector $\underline{c}$.

$$\hat{\underline{c}}(t) = \hat{\underline{c}}(t-1) + \frac{P(t-1)\underline{l}(t)}{1 + \underline{l}^T(t)P(t-1)\underline{l}(t)}[y(t) - \hat{\underline{c}}^T(t-1)\underline{l}(t)] \tag{13}$$

$$P(t) = P(t-1) - \frac{P(t-1)\underline{l}(t)\underline{l}^T(t)P(t-1)}{1 + \underline{l}^T(t)P(t-1)\underline{l}(t)} \tag{14}$$

A predictive control law can then easily be developed. Theorems proving the global convergence and stability of this scheme are presented in Zervos and Dumont (8).

The choice of the parameter p in the Laguerre functions is not crucial. However, it influences the accuracy of the approximation of the plant dynamics as a truncated series. For a given plant, there exists an optimal p

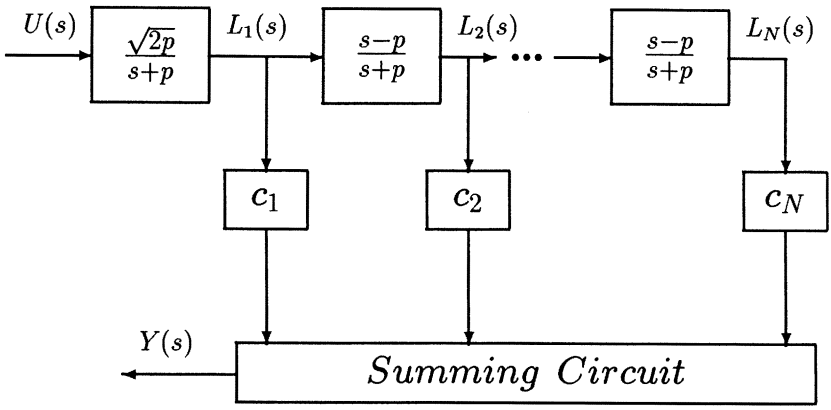

Fig 2– Laguerre ladder network

that minimizes the number of filters required to achieve a given accuracy. The chain of all-pass filters in the Laguerre network provides good representation of a time delay τ, in particular when $p = 2N/\tau$. The actual plant order has little bearing on the number of filters N. The horizon of the predictive control law is automatically adjusted on-line to assure closed-loop stability.

Multivariable Adaptive Control

In theory, the generalization of monovariable adaptive control schemes to the multivariable case is rather straightforward. The main difficulty appears for the case of multiple delays, in which case the design of the controller is significantly more complex. Recent advances in multivariable predictive control appear to provide a satisfactory solution. The major difficulty in multivariable adaptive control arises from the significantly increased number of parameters to be estimated on-line. This increases the computational load, but more importantly, slows convergence down to the point of making the algorithms impractical in many situations. Thus, in multivariable adaptive control, it is crucial to utilize as much prior knowledge as possible to reduce the number of parameters to be estimated on-line. Ideally, one would like to estimate physical parameters of the system

directly, but due to the nonlinear nature of most systems, this is not a trivial task. We shall discuss later some novel techniques that may overcome this problem.

APPLICATIONS TO PAPERMAKING CONTROL

Estimation of MD and CD Variations

While the gauges traverse the web, the sheet moves in the machine direction at very high speed. For instance, on a 4m wide newsprint sheet moving at 900 m/min, during a 20s scanning time, 300 m of paper has passed under the gauge. Thus, the path of the gauge on the sheet is at a very shallow angle (0.76 deg) relative to the machine direction. Therefore, the raw profile measurement contains a vary significant machine direction component. The problem is then: is it possible to retrieve the true profile from this raw signal?

On a machine with a steady profile and nearly white machine direction variations, the retrieval of the profile is a rather trivial task and the standard method which follows works well. Let $x^n(t)$ denote the raw measurement obtained at the measurement box n during the scan s. The standard method is simply an exponential filter

$$y^n(s) = (1 - a)y^n(s - 1) + ax^n(s) \tag{15}$$

where $y^n(s)$ is the estimated profile for the measurement box n and the scan s, and $0 < a < 1$ is the exponential filter pole. The exponential filter pole a is generally set such that $a \leq 0.3$, which means that after a step change in the profile it takes about 8 scans to get within 5% of the new profile. The simple filter described above is crude and rather slow to converge to the true profile. Moreover, it is not optimal. Because of the inherent process nonlinearity, for moisture profiles it will give biased estimates. Efforts to overcome this problem have resulted in a significant amount of patented work, all of which belongs to the vendors of paper machine control system.

More recently in our laboratory, Natarajan et al. (9) proposed a scheme consisting of a least-squares parameter identifier for estimating CD profile

deviations and a Kalman filter for estimating MD profiles. The Lindeborg (10) model for moisture deviations can be written as:

$$y_n^k = p^n + (1 + Bp^n)u_k + v_k \tag{16}$$

$$u_k = \bar{u} + \xi_k \tag{17}$$

where

- y_k^n is the measured profile deviation from the reference level at CD position n and time instant k
- p^n is the percentage deviation from the reference level at the CD position n.
- B is a constant
- u_k is the percentage MD variation at time k
- v_k is sensor noise assumed to be Gaussian white noise
- $\bar{u}$ is the mean moisture content in the MD
- ξ is a zero mean stochastic process

This model can be expressed in a state-space form given by:

$$x_{k+1} = Ax_k + W_k \tag{18}$$

$$y_k^n = p^n + C^n x_k + v_k \tag{19}$$

where:

$$x_k = \begin{bmatrix} \bar{u} \\ \xi_k \end{bmatrix}; \qquad A = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix};$$

$$W_k = \begin{bmatrix} 0 \\ w_k \end{bmatrix};$$

$$C^n = \begin{bmatrix} (1 + Bp^n) & (1 + Bp^n) \end{bmatrix} \tag{20}$$

If p^n and B are known, by using Equations (18) and (19) the estimation of $\bar{u}$ and ξ_k can be approached as a Kalman filtering problem. Conversely, if $\bar{u}$ and ξ_k are known, then the estimation of p^n and B can be attempted as a least squares parameter identification problem. The proposed algorithm is a bootstrap algorithm combining these two ideas. Using the present estimates of p^n and B in a Kalman filter, we predict $\bar{u}$ and ξ_{k+1} (at next

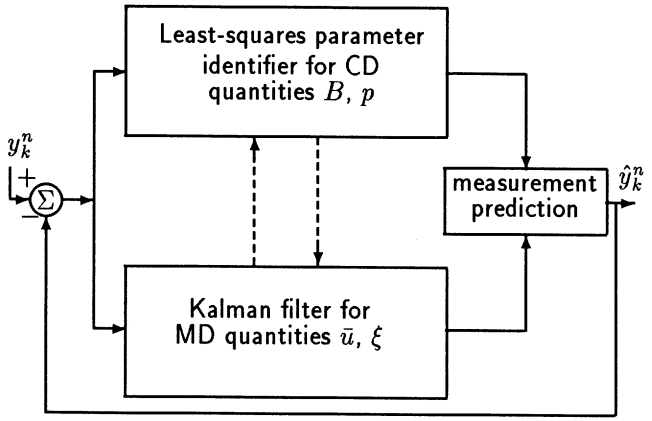

Fig 3– Structure of the profile estimation algorithm

instant), then using this prediction and the measurement $y_{k+1}^{n\pm1}$, we update $p^{n\pm1}$, B and so on. The overall structure of the algorithm is shown in Figure 3. The least squares identifier and the Kalman filter are well described in the literature, as for example in (11). Thus, for every measurement box, the least-squares algorithm provides an estimate of the CD profile deviation and the Kalman filter provides an estimate of the MD variation. Simulation and industrial test results show the algorithm to be more accurate and faster to converge to the true profile than other techniques. It can be tuned to provide optimal (in the least-squares sense) estimates. Any shortening of the paper machine, for example by successful use of impulse drying, will shorten dead time, making MD control at the current rate of once a scan unattractive. Because it gives several estimates of the MD variations per scan, this filter will allow MD control at a much increased rate. Obviously, the use of full CD non-scanning gauges would render such a filtering algorithm redundant.

MD Control of Dry Weight and Moisture

Machine direction control considers the control of the scan averages of basis-weight and moisture. Control of dry weight manipulates the thick

stock flow setpoint to obtain the desired fiber flow to the headbox. Usually, the thick stock consistency is controlled and feedforward to the thick stock flow setpoint is provided as well. Moisture is generally controlled from the steam pressure in the last dryer section. As seen earlier, the dynamics of the basis-weight and moisture control loops is heavily dominated by dead time. Also, on a newsprint machine, the setpoints for moisture and in particular basis-weight are rarely changed. Thus, for these loops one should really solve the regulation problem, which requires characterization of the process disturbance as a stochastic process.

However, in practice most systems installed in mills are based on deterministic control design and often tuned for the servo problem, *i.e.* setpoint tracking. The Dahlin algorithm is used extensively on paper machines. It is a simple and straightforward way to perform dead-time compensation. Depending on the process disturbance characteristics, it may in some instances be tuned to near minimum-variance performance, even though it does not use a model of the process noise. To solve the desired control problem for the basis-weight and moisture control loops requires minimum-variance control. Indeed, these crucial qualities of the final product must meet stringent specifications, hence it is important to obtain good control. In addition, any reduction in the moisture variations allows an increase in the moisture content, thus saving steam or permitting a production increase.

The first application of minimum-variance control to a paper machine was reported in 1967 by Åström (12). In practice, true minimum variance is rarely desirable because of the generally excessive control energy it requires. This is particularly true on a paper machine where actuators are valves that should not be subjected to excessive wear.

More recently, predictive control has attracted much interest in the process control community. It can be used as a way of reducing the control variance. Another problem with minimum-variance control is that it is impossible to tune manually. It relies on model of the plant, which because of furnish or grade changes, or other variations of the equipment or the environment, may need to be updated from time to time.

One way to automate this task is through adaptive control. Self-tuning control for the paper machine was proposed very early. Indeed two of the first industrial tests of self-tuning controllers were performed on a paper machine (13), (14). In both cases MD control of basis weight and moisture was considered. For moisture control, even the gains of the feedforward loop from the couch vacuum for decoupling control were adapted, using a controller of the form:

$$\Delta u(t) = b_1\Delta u(t-1)+b_2\Delta u(t-2)+a_0y(t)+a_1y(t-1)+c_0\Delta v(t)+c_1\Delta v(t-1) \tag{21}$$

where Δu is the incremental control signal, y is the error in moisture content, and Δv is the incremental couch vacuum.

Attempts by a vendor to develop a system based on the above work did not lead to a commercial product for reasons not disclosed (15). In Canada, a pulp and paper company attempted to apply the above methodology to a linerboard machine but failed to develop a scheme reliable enough for continous operation. The intended scheme used feedforward from several variables, resulting in the estimation of a large number of parameters and identifiability problems. This emphasizes the fact that adaptive control is not yet an *off-the-shelf* technology. Nevertheless, another pulp and paper company in Canada claims to have successfully developed and operated a self-tuning regulator for MD control of a fine paper machine for several years (16).

Lastly, one may look at the entire paper machine as a multivariable system. Although it has long been known that the headbox is a multivariable system, and should be controlled as such, it is still controlled by two monovariable loops. The headbox has been used in many simulation studies as a benchmark for testing both monovariable and multivariable adaptive control schemes.

Non-adaptive multivariable control has been applied to the entire paper machine. The paper machine multivariable control system *MUVAR* was developed and patented by the *Centre Technique de l'Industrie des Papiers, Cartons et Celluloses* in France (17). This is a 4×4 system considering the multivariable control of the headbox stock concentration and

level, the jet velocity and the basis weight by action on the air valve, the thin stock and thick stock valves and the slice opening, taking interactions into account. With a jet velocity sensor, the basis weight at the slice can also be controlled. This allows control of perturbations in a bandwidth above that controllable by the reel basis-weight control loop. This system is claimed to bring significant improvements over existing schemes, such as a faster response, better decoupling, and fast, efficient speed and grade changes. However, despite impressive results, this system failed to be a commercial success, probably because of its complexity.

CD Control of Dry Weight and Moisture

The most impressive progress in paper machine control over the past 10 years or so has been in the area of CD control, thanks to the development of innovative actuators and control algorithms. When 50 to 100 such actuators are distributed across the web, close interactions are bound to exist. A typical response of basis-weight to a change in a single slice actuator spreads to other locations. Fortunately, in general the only dynamics present, apart from dead-time, come from the actuators, and in most cases can be neglected. There are basically two approaches for modelling this kind of responses. The first uses an interaction matrix, the second a non-causal spatial impulse response (18).

The first approach considers the spread across the web in reponse to an actuator is finite, identical for all actuators, and symmetric. The response can therefore be represented by a small number of coefficients. Then, given an initial deviation from target profile Δy_0, where each element corresponds to a particular measurement box, and Δu is a vector containing the actuators moves, the resulting deviation from target profile Δy is given by

$$\Delta y = K\Delta u + \Delta y_0 \tag{22}$$

where K is a band diagonal matrix such as:

$$\begin{bmatrix} a & b & c & 0 & \cdots & 0 & 0 & 0 & 0 \\ b & a & b & c & 0 & \cdots & 0 & 0 & 0 \\ c & b & a & b & c & 0 & \cdots & 0 & 0 \\ 0 & c & b & a & b & c & 0 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & & & & \vdots \\ 0 & 0 & \cdots & 0 & c & b & a & b & c \\ 0 & 0 & 0 & \cdots & 0 & c & b & a & b \\ 0 & 0 & 0 & 0 & \cdots & 0 & c & b & a \end{bmatrix}$$

If the interaction matrix K is square, and if zero-deviation from the target profile is desired, then from equation (22) one can obtain

$$\Delta y = K\Delta u + \Delta y_0 = 0 \tag{23}$$

or

$$\Delta u = -K^{-1}\Delta y_0 \tag{24}$$

This equation is the equivalent to a deadbeat controller. Generally, to allow more conservative adjustments, the following is used

$$\Delta u = -\alpha K^{-1}\Delta y_0 \tag{25}$$

with $0 < \alpha < 1$. As seen on Figures 4 and 5, even for a simple K matrix with a fairly narrow interaction band, the K^{-1} matrix is a full and rather complex matrix. In practice, as the number of measurement boxes is generally larger than the number of actuators, the matrix K is non-square and thus non-invertible. Moreover, large control actions are also undesirable, and thus a better problem is to find Δu that minimizes J where

$$J = \Delta y^T Q \Delta y + \Delta u^T P \Delta u \tag{26}$$

and P and Q are symmetric, positive definite matrices. Differentiating with respect to Δu and equating to zero gives the control law:

$$\Delta u = -(K^T Q K + P)^{-1} K^T Q \Delta y_0 \tag{27}$$

Note that, when $P = 0$ and $Q = I$, one obtains

$$\Delta u = -(K^T K)^{-1} K^T \Delta y_0$$

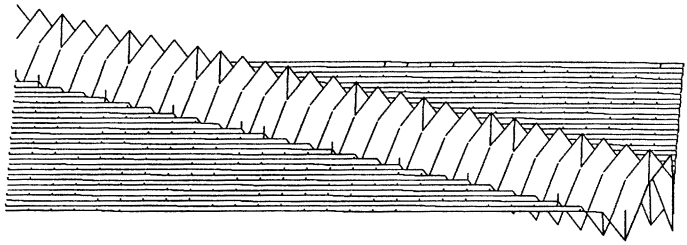

Fig 4– 3-D plot of the band diagonal interaction matrix

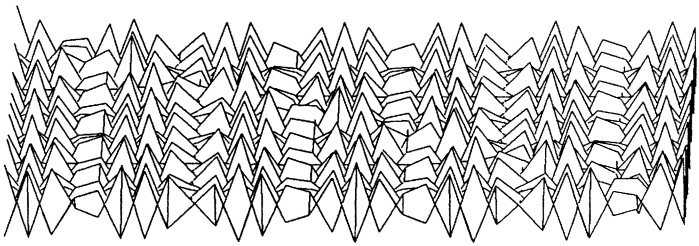

Fig 5– 3-D plot of the inverse of the band diagonal interaction matrix

which is the least-squares solution of equation (23) when K is non-square.

Obviously, to implement this control law, one needs to know the matrix K. One approach is to use a least-squares estimator to update it on-line, *i.e* having a self-tuning scheme able to track changes in the profile response, see (18). This is one of the most successful commercial applications of adaptive control. The fact that it is static and only three parameters are estimated may have a lot to do with that success.

The other approach to modelling is to view the spread across the web as a spatial impulse response, which differs from the familiar temporal impulse response by the fact that on the negative x-axis it is non-zero, and thus is non-causal. Assuming symmetry of the response, such a system can be approximated by:

$$\sum_{i=-n}^{n} a_i \Delta y_i = \sum_{j=-m}^{m} b_i \Delta u_j \tag{28}$$

where Δy_i is the profile deviation at CD location i, and Δu_j is the movement of actuator j. One can then design a linear quadratic controller that minimizes the following performance index:

$$J = \sum_{i=0}^{n} [(\Delta y(ih))^2 + \rho \left. \frac{d^2 \Delta u(x)}{dx^2} \right|_{x=ih}] \tag{29}$$

where h is the actuator spacing. The second term in this performance index, the second derivative of the slice lip deflection curve is proportional to the bending moment, and is there to protect the slice lip from permanent damage. For computation, this term is approximated by finite differences, see (18).

Control of Colour

Colourimetry studies indicate that the human eye is a trivariant system. As a result colours are defined by three numbers or tristimulus values. Currently, the ISO standard frame of reference is the $L\ a\ b$ system. Scanning colour gauges based on three-dimensional measurements have been developed and used to provide an on-line measurement of the colour on

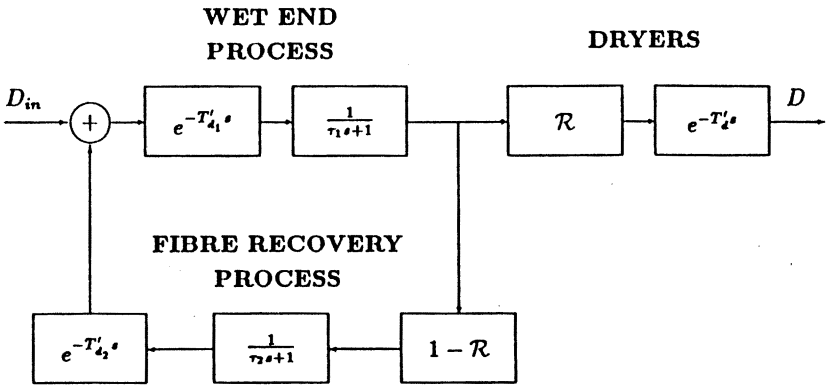

Fig 6– Dyeing process representation

paper machines.

The paper dyeing process can be modelled as a mass tranfer process. It is a reasonable to assume that the three dyes have the same transport dynamics. Indeed, the general understanding is that the dyes fix very rapidly to the fibers, and consequently that dye retention is the same as fiber retention. If such is the case, then the dyeing process can be represented by the block diagram of Figure 6 where R is the fiber retention. The colour at the reel then relates to the dye content at the reel through a complicated nonlinear function based on the Kubelka-Munk theory. The objective of the control is to maintain the paper colour at a desired setpoint despite the addition of recycled paper (broke) to the pulp. Good response to setpoint change is also desirable. Interest in such continuous colour monitoring and control is growing due to the demand for higher and more uniform paper quality.

The function relating dye concentration and colour on the $L\ a\ b$ scale is a non-linear one. For control it has to be linearized around the operating point. In so doing, the colour deviations relate to the dye deviations

through the colour matrix Λ:

$$\Delta \underline{C} = \Lambda \Delta \underline{D} \tag{30}$$

where $\Delta \underline{C}$ and $\Delta \underline{D}$ represent the deviations of the colour and dye vectors from the operating points. Note that the desired dye flows are not known unless Λ is known.

An estimate of Λ is required by any controller. However, Λ is a function of the setpoint and the dye characteristics, and it is affected by process variables such as the pH of the pulp. Thus, one might consider a multivariable adpaptive control scheme for this process where identification of the process is in large part the identification of an improved estimate of Λ. However, it is not possible in an input-output model to separate Λ from the other parameters. Thus even with a simple first-order model for dye transport, the number of parameters to be estimated on-line is rather large. Simulations show that such an adaptive controller fails to outperform and sometimes performs worse than a simple multivariable Dahlin controller with fixed gains (19). The adaptive controller simply cannot cope with the large number of parameters to be estimated and therefore displays poor tracking ability. This is a common problem with multivariable adaptive controllers. Ideally, in this case, to outperform the fixed controller one would like to estimate the colour matrix Λ directly.

Total Control of the Paper Machine

The current trend for paper machine control systems is toward on-line control of the physical properties of paper (20). Novel sensors to measure sheet strength in three directions (MD, CD and ZD) are becoming available. However, before controlling these properties on-line, we may have to develop new sensors and actuators. Moreover, we still lack clear understanding of how papermaking variables interact to define the sheet properties. The relationships involved are likely to vary in an unforeseeable fashion, and thus will need to be estimated on-line using information from various sensors. As emphasized in (20), there are many variables in the papermaking process, and thus the identification of those relationships will not be a trivial task. Therefore, to control all those properties on-line

will require the solution of a high dimensional, nonlinear time-varying control problem. From our earlier discussion of the colour control problem, it is obvious that new estimation and adaptive control techniques will be required. We present some promising approaches in the next section.

NOVEL APPROACHES

Multimodel Adaptive Control

The principle of multimodel adaptive control is illustrated in Figure 7. The unknown plant is assumed to lie within a finite set of models. The plant is characterized by a weighted sum of all those models. The weights vary from zero to one, their sum is one, and they reflect the probability of a particular model to adequately represent the plant. The adapter adjusts the weights on-line to minimize a performance index, generally the squared modelling error summed over a finite window. The main challenge is to develop a simple search algorithm that guarantees convergence to the global minimum, as the performance index is likely to be a multimodal function.

A multimodel adaptive controller presents various advantages. The bank of models might correspond to a model parameterized as a function of an unknown physical parameter assuming different probable values. This is particularly appealing in the case of a multivariable system, as it will likely result in fewer parameters to be estimated. Simulations show such multimodel systems have the ability to track very fast changes in dynamics. Another advantage of this approach is its inherent robustness. Indeed, even with a fixed weight distribution, such a control system will be rather insensitive to process dynamics changes.

There are various situations in papermaking for which such an approach could prove useful. For instance, in the previously discussed colour control problem, the main unknowns are the dye absorption and scattering coefficients. However, we know that these lie in a given range, and we could therefore easily build a bank of models corresponding to a finite number of values of absorption and scattering coefficients K and S. The adapter would then automatically adjust the weights to choose the most appropriate values of K and S at any given time. Such a scheme could handle

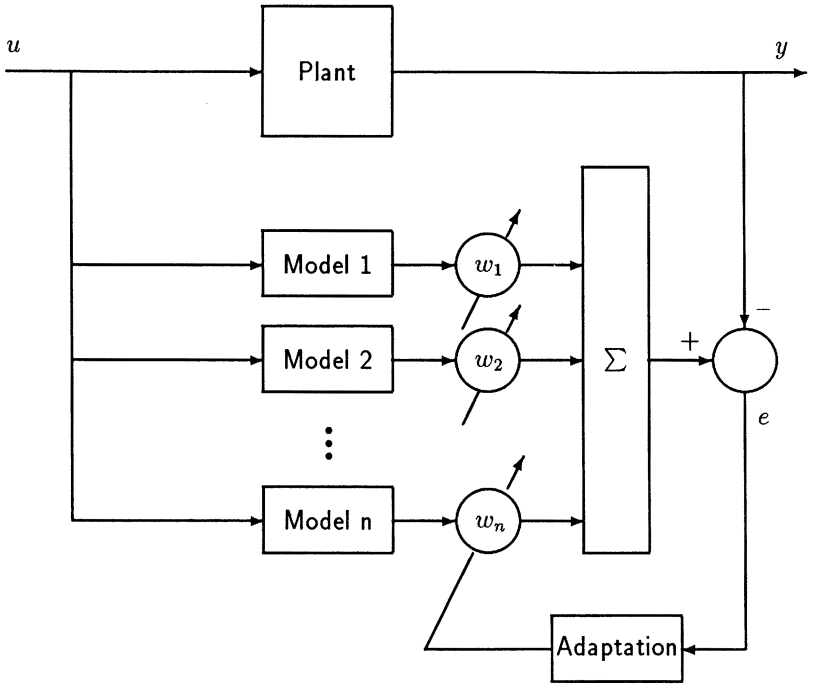

Fig 7– Multimodel adaptive control

grade and dye changes.

Another situation is when the raw material is known to consist of an unknown mix of a finite number of known species (21). If adequate models can be developed for individual species, then a bank of models can be built and the weights adjusted on-line. This not only provides automatic species compensation to the controller but, perhaps more importantly, automatic wood species identification to the operator. We illustrate this through a simple example involving chemical cooking of an unknown mixture of two wood species.

Consider the chemical pulping of two different wood species where we know the Hatton's equation (22) relating the final Kappa number to the

H-factor and the initial effective alkali EA for each species:

$$\kappa_i = \alpha_i - \beta_i(\log H)(EA)^{n_i} \qquad \text{for } i = 1,2 \tag{31}$$

Assume that the pulping takes place in a hypothetical reactor that can be described by first-order dynamics with pole a plus dead time k. Now, assume the feed stream is composed of a mix of two species and let w_i be the weight fraction of species i. The Kappa number out of the reactor is then described by:

$$\kappa_t = a\kappa_{t-1} + \sum_{i=1}^{2} w_i b_i u_{t-k} + d\sum_{i=1}^{2} v_{i,t-k} \tag{32}$$

where

$$b_i = -(1-a)[\beta_i(EA)^{n_i}\log H]_{t-k}$$

$$d = 1 - a$$

$$v = \alpha_i$$

$$u = \log H$$

From this model, we can easily design a κ-number multispecies controller based on a Dahlin controller and feedforward:

$$u_t = \frac{1-aq^{-1}}{\sum_{i=1}^{2} b_i w_i}\frac{1-p}{1-pq^{-1}-(1-p)q^{-k}}(\kappa^* - \kappa_t) - \frac{d}{\sum_{i=1}^{2} b_i w_i}\sum_{i=1}^{2} w_i v_{i,t} \tag{33}$$

where κ^* is the setpoint and p is the desired closed-loop pole. It is easily shown that the multispecies controller can be expressed as a linear combination of the single species controllers using the same control error signal, i.e:

$$u_t = \sum_{i=1}^{2} \lambda_i u_{i,t} \tag{34}$$

where

$$\lambda_i = \frac{b_i w_i}{\sum_{i=1}^{2} b_i w_i} \tag{35}$$

The multispecies controller above assumes knowledge of the feed stream composition, i.e. of the w_i. If the feed stream composition is unknown, then we must find the right values for λ_i or w_i. In this example, we choose

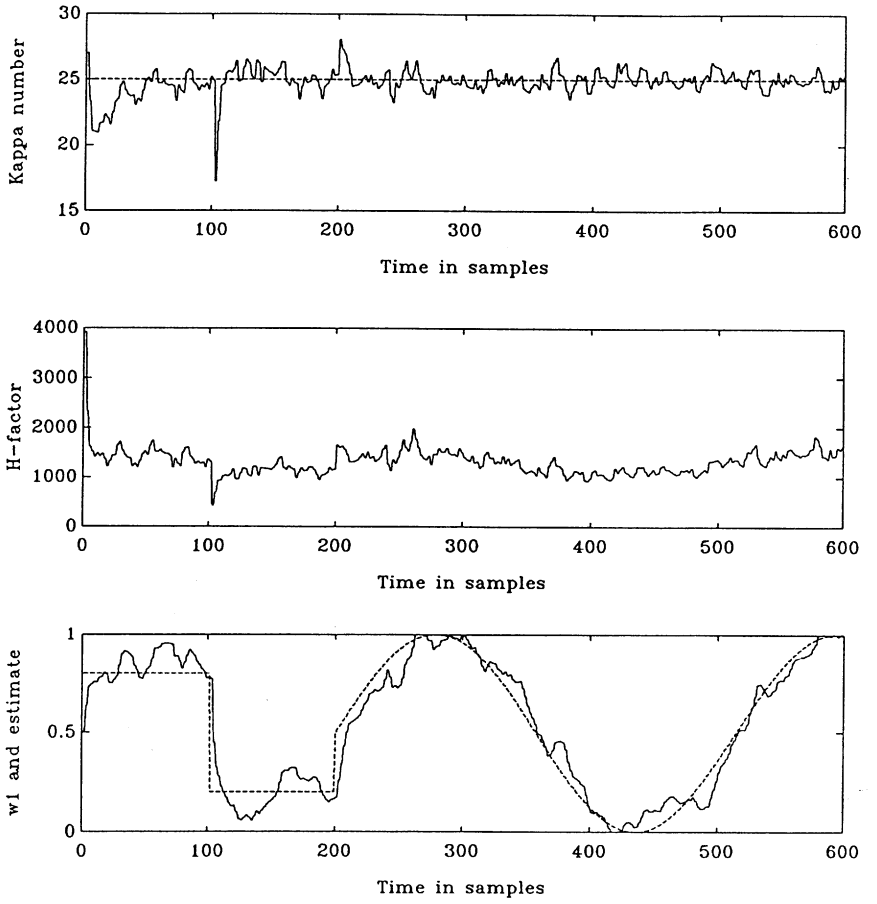

Fig 8– Adaptive species compensation in kraft pulping of a two-species mix

the second option, i.e. direct identification of the w_i. For this, we choose the simple projection algorithm:

$$\hat{w}_{1,t} = \hat{w}_{1,t-1} + \frac{c\kappa_{i,t-1}}{1 + \phi_{t-1}^T \phi_{t-1}} [\kappa_t - \phi_{t-1} \hat{\theta}_{t-1}] \tag{36}$$

where c is a constant and

$$\phi_t = [\kappa_{1,t} \kappa_{2,t}]^T$$

$$\hat{\theta}_t = [\hat{w}_{1,t} \hat{w}_{2,t}]^T$$

$$\hat{w}_{2,t} = 1 - \hat{w}_{1,t}$$

Figure 8 shows the behaviour of such a scheme when controlling the chemical pulping of a variable mixture of hemlock and cedar. Noise has been superimposed to make this simulation a bit more realistic. It is seen that the Kappa number is well controlled despite wide variations in the species mix, and that the composition of the feed is accurately estimated. Although this example concerns chemical pulping, it is straightforward to apply it to a species-dependent papermaking control problem, such as refining.

Genetic Algorithms

Genetic algorithms (GA) are search algorithms based on concepts of natural selection and genetics, see Goldberg (23). Using the Darwinian principle of survival of the fittest together with some simple genetic operators, genetic algorithms perform a randomized but directed search. Contrary to gradient-based algorithms that only go from one point to another in the search space, a GA uses a population of points in the search space and is therefore less likely to miss the global optimum for a local one. Thus, it is particularly well suited to multimodal functions. Also, because it does not require gradient information, it can easily handle discontinuous functions.

A GA works with a population of binary strings just as nature works with chromosomes. The binary strings are made from a coding of the parameters which the algorithm should find or identify. Each parameter corresponds to a fixed length binary substring of j bits $[0, \ldots, 2^j - 1]$. The

value of the substring is mapped to an interval of the real numbers $[l, u]$ so the precision of the coding is $(u - l)/(2^j - 1)$. With n parameters, the final string consists of n concatenated substrings. The initial population is generated randomly and the population size is kept constant throughout the process. The algorithm only requires payoff information (fitness) for each string.

A genetic algorithm in its simplest form consists of 3 steps: reproduction, crossover and mutation. For reproduction, strings are chosen according to their normalized fitness. The strings with above average fitness will have more offsprings than those with below average fitness. This directs the search towards the best individuals. Next, new individuals have to be generated by crossover, the main search operator. This operator combines substrings from individuals selected for mating to generate a population of better fit individuals. To apply this operator two strings are mated at random and a point on the interval $1 \leq k \leq j - 1$ is chosen randomly. Two new strings are then created by changing all characters between position 1 and k inclusively. This can best be explained by example. Suppose there are two strings *00000000* and *11111111* and assume a randomly chosen crossover point of 3. Then the new strings will be *11100000* and *00011111*. These two operators give genetic algorithms much of their power. The mutation operator, which simply flips the state of a bit, can be viewed as a background(secondary) operator to insure against loss of information in some bit positions and as a way of getting the algorithm out of local optimum.

Although the transition rules are probabilistic, this approach is not a random search. By use of operators taken from population genetics the algorithm efficiently explores part of the search space where the probability of finding improved performance is high. Although genetic algorithms have been shown to behave well on multimodal functions, there is no known necessary and sufficient condition under which a function is genetically optimizable. Genetic algorithms are inherently parallel. Indeed, all strings or individuals in a population evolve simultaneously without central coordination. To realize their full potential, they must be implemented on parallel computer architectures. Genetic algorithms allow direct identification of physical parameters, difficult with current methods due to severe nonlin-

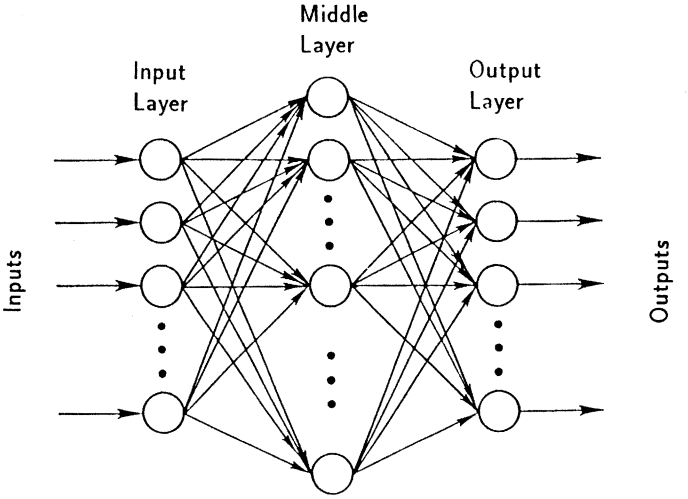

Fig 9– A three-layer neural network

earities, and thus have great potential in adaptive control, as shown in our laboratory see Kristinsson and Dumont (24) and Dumont and Kristinsson (25).

Neural Networks

Neural networks are non-algorithmic computer architectures based on our current understanding of how the brain encodes and processes information (26). They consist of densely interconnected networks of simple processing elements called neurons. Figure 9 shows a three-layer network. Artificial neurons are crude emulations of their biological counterparts. The output of an artificial neuron is computed as the output of a nonlinear function operating on a weighted sum of all inputs. The inputs x_i correspond to stimulation levels, the weights w_i to synaptic strengths, the non-linearity f(generally a sigmoid function) emulates the activation threshold:

$$y = f(\sum_{i=1}^{n} w_i x_i) \qquad \text{with } f(s) = \frac{1}{1 - e^{-s}} \tag{37}$$

Neural networks can learn by training. A successful application is in recognition of handwritten characters. Neural networks have proved good at extracting patterns, and thus their biggest potential is in pattern recognition. An interesting capability of multilayer neural networks is highlighted in a theorem due to Kolmogorov (27) which implies that any continuous mapping from $\boldsymbol{R}^n$ to $\boldsymbol{R}^m$ can be implemented as a three-layer neural network with n input neurons, m output neurons, and $2n + 1$ middle-layer neurons. This property could be used to exactly map the profile response on paper machines, without assumptions of the type described previously. We are currently investigating this approach in our laboratory, see Dumont (28).

CONCLUSIONS

We have explored the extent to which techniques of system identification and adaptive control have been applied to papermaking. Despite some successes, those techniques are not widely used, primarily because their use is not straightforward. In the case of monovariable systems, the techniques are fairly well developed and work well. With some clever engineering, they can be made easy to use. In the case of multivariable systems, the current schemes suffer from identifiability and convergence problems when dealing with high dimension systems, as demonstrated by the colour control problem. Novel approaches such as multimodel adaptive control, genetic algorithms and neural networks may provide innovative solutions to these nonlinear, high dimensional problems likely to be encountered as we move towards total control of the paper machine.

REFERENCES

1. Bolam, F. (Editor), *Papermaking Systems and their Control*, The BP & BMA (1970).

2. Dumont, G.A.,"Application of Advanced Control Methods in the Pulp and Paper Industry", *Automatica*, **2**(2), 143-153 (1986).

3. Ljung, L., *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, New-Jersey (1987).

4. Ljung, L. and Söderström, T., *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, Massachusetts (1983).

5. Åström, K.J. and Wittenmark, B., "On Self-Tuning Regulators", *Automatica*, **9**(2), 185-199 (1973).

6. Åström, K.J. and Wittenmark, B., *Adaptive Control*, Addison-Wesley, Reading, Massachusetts (1989).

7. Dumont, G.A. and Zervos, C. ,"Adaptive Control Based on Orthonormal Series Representation", *2nd IFAC Workshop on Adaptive Systems in Signal Processing and Control*, Lund, Sweden, 371-376 (1986).

8. Zervos, C. and Dumont, G.A.,"Deterministic Adaptive Control Based on Laguerre Series Representation", *International Journal of Control*, **48**(6), 2333-2359 (1988).

9. Natarajan, K., Dumont, G.A. and Davies, M.S.,"An Algorithm for Estimating Cross and Machine Direction Moisture Profiles on Paper Machines", *1988 IFAC Symp. on Identification and Syst. Param. Estimation*, Beijing, China 1503-1508 (1988).

10. Lindeborg, C.,"A Process Model of Moisture Variations", *Pulp and Paper Canada*, **87**(4), T142-147 (1986).

11. Goodwin, G.C. and Sin, K.S., *Adaptive Filtering, Prediction and Control*, Prentice Hall Inc., Englewood Cliffs, New-Jersey (1984).

12. Åström, K.J.,"Computer Control of a Paper Machine — An Application of Linear Stochastic Control Theory" *IBM J. Res. Dev.*, **11**, 389-405 (1967).

13. Borisson, V. and Wittenmark, B.,"An Industrial Application of a Self-Tuning Regulator", *4th IFAC Symp. on Digital Comp. Appl. to Proc. Control*, Zurich (1974).

14. Cegrell, T. and Hedqvist, T.,"Successful Adaptive Control of Paper Machines", *Automatica*, **11**(1), 53-60 (1975).

15. Kelly, W.H.,"Self-Tuning Control Strategies: Modern Solutions to Papermaking Control problems", *Can. Industr. Comp. Soc. Conf.*, Hamilton, (1982).

16. Gilbert, A.F. and Venhola, M.V.,"Application of Self-Tuning regulators at Great Lakes Forest Products",*Can. Industr. Comp. Soc. Conf.*, Ottawa, 48.1-48.5, (1986).

17. CTP,"MUVAR–Un Nouveau Système de Commande et de Régulation Multivariable", *Centre Technique de l'Industrie des Papiers, Cartons et Cellulose*, Grenoble, France (1982).

18. Chen, S.C., Snyder, R.M. and Wilhelm, R.G.,"Adaptive Profile Control for Sheetmaking processes", *Proc. of the 6th Int. IFAC/IMEKO Conference on Instrum. and Autom. in the Paper, Rubber, Plastics and Polymerization Industries*, Akron, OH, 77-83 (1986).

19. Bond, T. and Dumont, G.A.,"On Adaptive Control of Colour for Paper Machines", *1988 American Control Conference*, Atlanta, 1965-1970 (1988).

20. Baum, G.A.,"On the Verge of Total Control of the Paper Machine", *PTI*, **29**(8) (1988).

21. Dumont, G.A., "Wood Species Compensation and Identification via Multimodel Adaptive Control", *in preparation* (1989).

22. Hatton, J.V.,"Development of Yield Prediction Equations in Kraft Pulping", *Tappi*, **56**(7), 97-100 (1973).

23. Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Massachusets (1989).

24. Kristinnson, K. and Dumont, G.A.,"Genetic Algorithms in System Identification", Third IEEE International Symposium on Intellligent Control, Arlington, Virginia (1988).

25. Dumont, G.A. and Kristinnson, K., "On Adaptive Control and Genetic Algorithms", *Journal of Intelligent and Robotic Systems*, To be submitted for review.

26. Rumelhart, D.E. and McLelland, J.L. (Editors), *Parallel Distributed Processing — Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, Massachusetts (1986).

27. Kolmogorov, A.N.," On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of one Variable and Addition", *Amer. Math. Soc. Translation* , 2, 55-59 (1957).

28. Dumont, G.A., "Paper Machine Profile Control via a Neural Network", *in preparation* (1989).

Transcription of Discussion

SYSTEM IDENTIFICATION AND ADAPTIVE CONTROL IN PAPERMAKING

G. A. Dumont

ERRATUM
Equation (37) on page 1177 of Vol.2 should read:

$$f(s) = \frac{1}{1 + e^{-s}}$$

P.T. Herdman Wiggins Teape R & D Ltd

There is considerable interest in the use of expert control techniques although these are still in the early stages. Could you tell us how this relates to the techniques you have just described?

G.A. Dumont

What is known currently as "expert control" is a combination of expert systems technology with adaptive control. The essential purpose is to overcome the problem of having to have a control expert available 24 hours a day. The expert systems are designed to do what a process control expert would do; he would look at what the control system is doing; try to tune it; change the structure and so on.

M.I. MacLaurin

This has just reminded me about decoupled machine direction control. I first saw machine direction basis weight control on a paper machine when I joined Wiggins Teape in 1957. It was on Stoneywood number 8 machine with a hard valve system that took a beta gauge signal and applied it to a machine speed, without any decoupling so it did not really work too well.

Having said that you will appreciate that I did not really understand the significance of that Neural Network theorem. I

wonder if you could explain; how does it relate to the CD response?

G.A. Dumont

If you look at the response between 50 or 60 actuator movements on the slice and the 2-300 measurement values on a profile you have a continuous function. The problem is how do you characterise this response? This is a multi dimensional problem of many variables. What people have done is to use the simplistic model of the band-diagonal matrix which is the only thing that can be managed currently. The advantage of the neural network approach is that you would not need to make any assumptions at all such as that all actuators have a similar response and there is no edge effect. You could directly identify the full matrix. There may be some problems associated with the speed of convergence and that is being worked on currently.

Dr. G.A. Baum James River

Thank you for a very interesting paper. Please could you give an example of genetic algorithms?

G.A.Dumont

We have used it to find the friction coefficients for the drive in a robotic arm. That is something which is quite difficult to do using standard techniques because the system has a discontinuity. Genetic algorithms do not worry if there are discontinuities. One of the major obstacles in applying these techniques is that they are extremely slow to run on sequential computers, they are essentially parallel algorithms that require parallel computers. These are coming up slowly so within 5 or 10 years we will be able to use these techniques more widely.

A.R. Guesalaga UMIST (England)

The model described in Equation 16 on page 1161 in the proceedings for MD and CD variations present a high non linearity for the variables being identified. In figure 3 on page 1162 you then present a structure to tackle the estimation problem which has some similarity to the so called Extended Kalman Filters.

From my experience using Extended Kalman Filters, this method shows poor robustness even for very soft nonlinearities in the models being identified, and for some initial values of the estimates and covariance matrix,it can lead to wrong estimates.

In your paper you do not refer much to the robustness of your method, so I would like to know if your approach overcame the convergence problem, so it can be applied with confidence to a production process?

G.A. Dumont

The scheme we use is not an Extended Kalman Filter (EKF) but a bootstrap parameter estimator. An EKF simultaneously estimates parameters and states, and is thus rather complex. Because a bootstrap estimator provides separate estimation of parameters and states, it is computationally simpler. Because of the nonlinear nature of such a system, there is no general proof of convergence. However, various simulation and application studies have shown the superiority of bootstrap algorithms.