A MATHEMATICAL PROCEDURE FOR ESTIMATING CROSS AND MACHINE DIRECTIONS MOISTURE PROFILES IN PAPERMAKING PROCESSES

Refaat A. El-Attar
Department of Engineering Mathematics and Physics, Faculty of Engineering,
Alexandria University, Alexandria, Egypt.

RACT

mathematical procedure has been developed for estimating cross and machine direction (CD and MD) oisture profiles. A model is first presented that describes mathematically the process under consideration. The esence of noise is taken into account. The procedure consists of a least squares parameter identifier for timating CD profile deviations and a Kalman filter for estimating MD profiles. Simulation results of the ocedure are given and are followed by the results of its application to industrial data. The developed ocedure can be extended to other industrial models.

RODUCTION

per making processes, the moisture content of as it is being made, is measured using a single scanning sensor located after the paper machine The sensor in mounted and guided to move in achine direction (CD) perpendicular to the nt of the paper in the machine direction (MD). or traverses from front to back and from back to tinuously, taking measurements at fixed intervals Due to the MD movement of the paper and the ement of the sensor, the resulting measurement the paper form a zig-zag pattern as shown in). The measured profile is a composite of CD profiles. In [1], the author postulated and ntally verified the measured profile to be a function of CD and MD profiles. This ip forms the model for the present work.

ective of the measurement of moisture on paper rol the final sellable product [2]. The primary n to control the MD profile is the Dryer steam Increasing the steam pressure in the dryer a paper machine will result in decreasing the content. The CD profile is controlled either by eating boxes at intervals across the machine's l/or by remoisturizing showers. Therefore, in control the moisture, we must first estimate the MD profiles from the composite measured paper, we develop a mathematical procedure to deduce the CD and MD profiles from the zig-zag measurements. In section 2, we give a description of the moisture model based on the work of Lindeborg [1]. The proposed mathematical procedure is presented in section 3. Based on this procedure, we give, in section 4, the computational requirements in terms of computer operations knowing that the calculation is carried out in real time and that computer time is at a premium. The simulation results are illustrated in section 5; and the application of the proposed procedure to an industrial process is given in section 6.

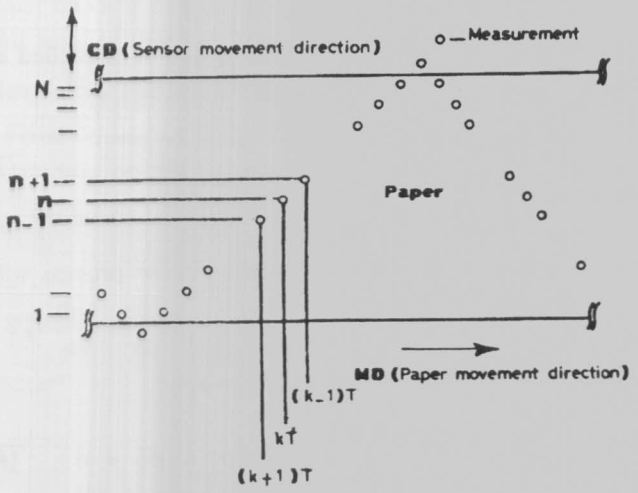

Figure 1. Zig-zag measurements due to CD and MD movements of sensor and paper.

2. MATHEMATICAL MODEL

The moisture variations can be described as [1]:

$$P_k^n = M + p^n + (1+Bp^n)u_k + O(h) \quad (1)$$

where:

- P_k^n is the (percentage) moisture content at CD position n, $(1 \le n \le N)$ at time kT (Figure (1)); N is the total number of CD points at which measurements are taken; T is the sampling period between measurements and k is an integer.
- M is the known reference level (percentage) moisture content.
- p^n is the (percentage) profile deviation from the reference level in CD at position n. By definition: $\sum_{n=1}^{N} p^n = 0$.
- u_k is the (percentage) MD variation at time kT.
- B is a constant (a function of M and other constants), and
- O(h) represents higher order terms.

The machine direction variations can be described as:

$$u_k = \bar{u} + \xi_k \quad (2)$$

where:

- $\bar{u}$ is the mean (percentage) moisture content in the MD.
- $\{\xi\}$ is a zero mean stochastic process.

The analysis of data indicates that ξ can be modelled as a first order process described by:

$$\xi_{k+1} = a\xi_k + w_k \quad (3)$$

where:

- a is a known constant, and
- {w} is a zero mean Gaussian white noise process with known variance q.

Equation (1) may be rearranged to read:

$$y_k^n = p^n + (1+Bp^n)u_k + v_k \quad (4)$$

where:

- y_k^n is the measured profile deviation from the reference level M, at CD position n and time instant kT, and
- {v} comprises of sensor noise and neglecting higher order terms and is assumed to be a Gaussian white noise process with known variance R.

The implication of the zig-zag measurement path in equation (4) is that when k increases by one, n increases or decreases by one (accordingly as sensor movement is front to back or back to front). This relationship must be kept in mind through the rest of the paper.

The problem now is to develop a recursive algorithm to estimate p^n $(1 \le n \le N)$, B, $\bar{u}$, ξ_k (at each instant k). To make this estimation, we get at each instant k (and corresponding n) the measurement, i.e. the left hand side of equation (4). It is assumed that we know, or have good estimates for the parameters a, q and R. Also, on the assumption that the CD quantities vary slowly, if at all, in comparison with the MD quantities, the total number of parameters and states to be estimated are N + 1 (CD quantities) and 2 (MD quantities) respectively.

Equations (2), (3) and (4) can be amalgamated into a state space form given by:

$$x_{k+1} = Ax_k + W_k \quad (5)$$

$$y_k^n = p^n + C^n x_k + v_k \quad (6)$$

where

$$x_k = \begin{bmatrix} \bar{u} \\ \xi_k \end{bmatrix}; \; A = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}; \; W_k = \begin{bmatrix} 0 \\ w_k \end{bmatrix}; \; C^n = \begin{bmatrix} (1+Bp^n) & (1+Bp^n) \end{bmatrix} \quad (7)$$

3. MATHEMATICAL PROCEDURE

Using equations (5) and (6), and assuming that p^n and B were known, the estimation of $\bar{u}$ and ξ_k can be approached as a Kalman filtering problem. Conversely, if $\bar{u}$ and ξ_k were known, then the estimation of p^n and B can be strived as a least squares parameter identification problem. The proposed procedure 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 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 procedure is shown in Figure (2). The least squares identifier and the Kalman filter are well described in the literature, as for example in [3].

To increase the robustness of the procedure, a few heuristic modifications are needed. These are introduced as follows:

(i) With each CD position n, we associate two quantities p^n (as before) and B^n (in place of a single B). At the end of each scan of the sensor, we average the N values of B^n estimated during the scan and replace the B^n by this average $\bar{B}$. With this artificially introduced set B^n, the CD parameters to be estimated are collected into a 2N×1 vector θ given by:

$$\theta = [\, p^1 \;\; B^1 \;\; p^2 \;\; B^2 \;\; \ldots \;\; p^N \;\; B^N \,]^T \qquad (8)$$

where T denotes transposition.

(ii) At the end of each scan we compute the average of the N values of p^n estimated during the scan. This average is subtracted from each p^n (to ensure $\sum_{n=1}^{N} p^n = 0$) and finally added to the currently estimated $\bar{u}$.

These two modifications arose from simulation and practical experience. Intuitively, (ii) above can be seen by the fact that if the right hand side of equation (4) is averaged over a scan, with $\sum_{n=1}^{N} p^n = 0$, we have $\bar{u}$ as the approximate result.

From [3], the least squares identifier for θ in equation (8) is given by equations (9), (10) and (11), below.

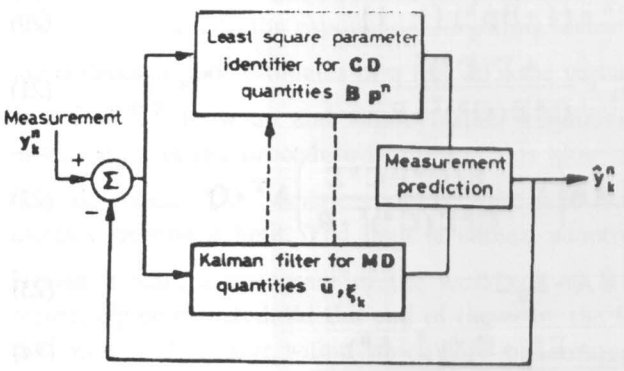

Figure 2. Overall structure of the procedure.

$$\Psi_k^n = \left| \frac{d\hat{y}_k^n}{d\theta} \right|_{\hat{\theta}_k} \qquad (9)$$

$$V_k = V_{k-1} - \frac{V_{k-1}\Psi_k^n(\Psi_k^n)^T V_{k-1}}{1+(\Psi_k^n)^T V_{k-1}\Psi_k^n} \qquad (10)$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + V_k \Psi_k^n (y_k^n - \hat{y}_k^n) \qquad (11)$$

where $\hat{}$ denotes estimate of the quantity under it, and V_k is the 2N×2N covariance matrix at instant k, with value V_o (likewise $\hat{\theta}_1$ is an initial value).

In reading equations (9), (10) and (11), the connection between n and k should not be invoked for those quantities for which n does not appear as an explicit superscript. In all such variables all 2N (or 2N×2N) CD quantities are simultaneously updated from the scalar measurement y_k^n and its estimate.

In an industrial environment, the identifier is to be implemented in a low-cost portable computer. Typically, in this environment, N is 50 and the sampling period T is one second. In such an environment, since the sampling period is not large enough to finish all the computations in a low-cost computer, the least squares identifier expressed in equations (9), (10) and (11) cannot be used directly. Therefore we would like to reduce its size. Ideally we would like to update θ_k and V_k locally, i.e. only the two elements and the 2×2 sub-block corresponding to the CD position n associated with k. Such a local update can be achieved if we place restrictions on V_o and Ψ_k^n. The restriction on V_o is that it should be block diagonal with each block a 2×2 positive semidefinite matrix. Physically this implies that we assume that adjacent profiles are not correlated. The restriction on Ψ_k^n will be developed next.

From equations (9) and (6), with the understanding that everything on the right hand side of (12) is evaluated at $\hat{\theta}_k$, we have:

$$\Psi_k^n + \frac{dp^n}{d\theta} + \frac{dC^n}{d\theta}\hat{x}_k + \left\{C^n \frac{d\hat{x}_k}{d\theta}\right\}^T = \Gamma^n + \Omega^n \hat{x}_k + \left\{C^n \frac{d\hat{x}_k}{d\theta}\right\}^T \qquad (12)$$

where:

Γ^n is a 2N×1 vector with all elements zero except the element 2n-1 which is 1.

Ω^n is a 2N×2 matrix with all elements zero except those of the sub-block formed by the rows 2n-1

and 2n. The non-zero sub-block is given by the matrix Φ in equation (13).

$$\Phi = \begin{bmatrix} \hat{B}^n & \hat{B}^n \\ \hat{p}^n & \hat{p}^n \end{bmatrix} \tag{13}$$

From equation (12), it is apparent that the first two terms have the structure that we seek. However, the third term does not satisfy our requirements as the 2N×2 matrix can in general have non-zero elements throughout. This is best seen by expressing the recursive equation for $d\hat{x}_k/d\theta$ for the conditions of our problem [3]. In view of this we neglect the third term in equation (12). The dropping of this term can be justified only by simulation.

Since we will now be dealing only with the local versions of equations (9), (10) and (11), we introduce some additional notation. P^n denotes the sub-block of V_k corresponding to the covariance of $\left[\hat{p}^n \quad \hat{B}^n\right]^T$ while the corresponding sub-block of Ψ_k^n is denoted ψ_k^n and is given by equation (14).

$$\psi_k^n = \left[1 + \hat{B}^n(\hat{\bar{u}}_k + \hat{\xi}_k) \quad \hat{p}^n(\hat{\bar{u}}_k + \hat{\xi}_k)\right]^T \tag{14}$$

To provide the least squares identifier with the ability to track slowly varying p^n, B^n we use an exponential forgetting factor α, $0 < \alpha \leq 1$.

The Kalman filter design, for estimating x_k, is standard. In equation (5), the covariance of W_k is given by Q in equation (15) with $q_1 = 0$.

$$Q = \begin{bmatrix} q_1 & 0 \\ 0 & q \end{bmatrix} \tag{15}$$

However, with $q_1 = 0$, the gain for $\hat{\bar{u}}$ will asymptotically go to zero. To avoid this, a small value of q_1 is used. Σ and K will be used below to denote the Kalman filter's covariance and gain respectively.

In any identification algorithm bounds have to be applied whenever excessively high or low parameter estimates are obtained. The bounds are selected from apriori knowledge of the range of values for the parameters being estimated. Thus whenever the estimate of a variable at any stage of the procedure (equations (16) through (24)) is higher or lower than the appropriate bound, the estimate is replaced by the transgressed bound.

Based on the ideas presented in this section, we present the proposed procedure in the order in which it is to be executed, in equations (16) through (24). On the first pass through these equations, all variables not defined earlier are to be provided as initial estimates. When the same variable appears on both sides of an equation, the current value of the variable is to be used in evaluating the entire right hand side and the results assigned to the same variable as its new value. Any equations following, when referring to this variable, will use the new value. At this point attention is drawn again to the relationship between n and k imposed by paper and sensor movement. The averaging operations on $\hat{p}^n$, $\hat{B}^n$ at the end of each scan are not shown in the equations below.

$$z_k = \hat{\bar{u}} + \hat{\xi}_k \tag{16}$$

$$\psi_k^n = \left[1 + \hat{B}^n z_k \quad \hat{p}^n z_k\right] \tag{17}$$

$$P^n = \frac{1}{\alpha}\left\{P^n - \frac{P^n \psi_k^n (\psi_k^n)^T P^n}{\alpha + (\psi_k^n)^T P^n \psi_k^n}\right\} \tag{18}$$

$$\left[\hat{p}^n \quad \hat{B}^n\right]^T = \left[\hat{p}^n \quad \hat{B}^n\right]^T + P^n \psi_k^n (y_k^n - \hat{y}_k^n) \tag{19}$$

$$C^n = (1 + \bar{B}\hat{p}^n)\,[1 \quad 1] \tag{20}$$

$$K_k = \frac{A\Sigma(C^n)^T}{C^n\Sigma(C^n)^T + R} \tag{21}$$

$$\Sigma = A\left(\Sigma - \frac{\Sigma(C^n)^T C^n \Sigma}{C^n\Sigma(C^n)^T + R}\right)A^T + Q \tag{22}$$

$$F = A - K_k C^n \tag{23}$$

$$\hat{x}_{k+1} = F\hat{x}_k + K_k(y_k^n - \hat{p}^n) \tag{24}$$

4. COMPUTATIONAL REQUIREMENTS

Although the procedure appears to require considerable computation, it should be noted that no more than 2×2 matrix is involved in each equation. The most complex set of calculations is at the edge of the sheet (at the end of a scan). In this case, before processing the next

measurement, 60 additions, 60 multiplications, and 20 comparisons are needed. The operations count are rounded to the nearest higher multiple of 10. End of scan averaging is assumed to be done recursively. The comparison operations arise out of the bounds imposed on the estimated values. In the operations count, the symmetry of the covariances in exploited. Partial products used repeatedly are assumed to be computed once and stored. The storage space for these partial products is taken into account in the estimate of the memory requirements given later. Assuming that each operation takes 40μs, the worst case computation time is 6ms. Assuming memory access times of the same order, we require, in the worst case, 20ms for computation. This is an order of 50 less than the sampling period used currently in the industry.

The memory space required by data for N CD positions, exploiting the symmetry in the covariances, is, given, rounded up, by (70+4N) times the word length for the number representation. With a word length of 4 bytes and for N=50, we require 1.5Kb of memory for data storage. Assuming the program size is similar, the total memory requirement is about 3Kb. Thus, in terms of speed and memory, the proposed procedure can be installed on a low-cost portable computer.

5. SIMULATION RESULTS

The proposed procedure of section 3 was first tested by simulation. In long duration runs of the procedure, it was observed that due to the exponential forgetting factor and the occasional poor estimates of a few $\hat{B}^n$, the variances of a few $\hat{B}^n$ blow up and vitiate further estimates. A modification of the procedure to prevent this blow up is now described. The variances of $\hat{B}^n$ are allowed to increase beyond a limit. The limit is chosen adaptively. During a scan, a histogram of the variances of $\hat{B}^n$ is recursively constructed. At the end of the scan, the limit is chosen as that value within which 85% of the area of the histogram is enclosed. When the variance of $\hat{B}^n$ exceeds the limit, two actions are taken: The variance of $\hat{B}^n$ is set to the limit and the cross-covariance of $\hat{p}^n$ and $\hat{B}^n$ is set to zero. The latter action is taken to ensure that the covariance of $[\hat{p}^n \ \ \hat{B}^n]^T$ (a 2×2 matrix) does not turn negative definite.

A typical simulation result is depicted in Figure (3).

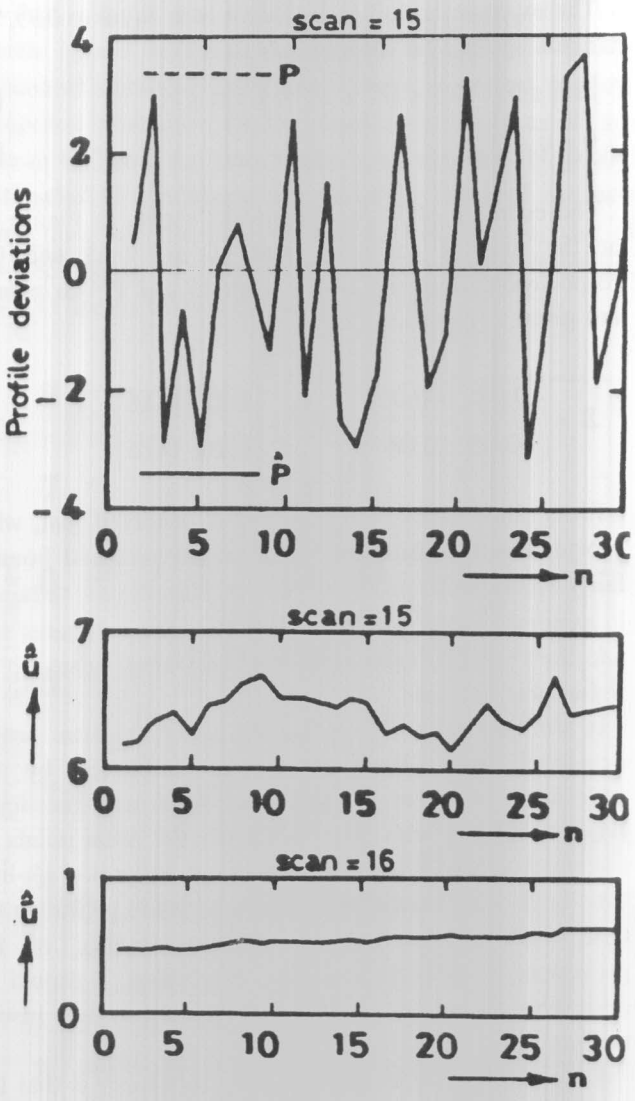

Figure 3. Simulation results.

In this simulation, the process is modelled by equations (2), (3) and (4). The process data assumed in N=30, a=0.9753, q=0.015, R=0.0025, $\bar{u}$=0.5 and B=0.5. Profile deviations are generated as uniform random deviates within ±3.5. The initial parameters for the procedure are chosen as:

$$\hat{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.31 \end{bmatrix} \tag{25}$$

$$\begin{bmatrix} \hat{p}^n \\ \hat{B}^n \end{bmatrix} = \begin{bmatrix} 0 \\ 0.10 \end{bmatrix}; \quad P^n = \begin{bmatrix} 100 & 0 \\ 0 & 4 \end{bmatrix}; \quad 1 \le n \le N \tag{26}$$

The exponential forgetting factor α is equal to 0.95. The bounds applied on the estimates are

$$|\hat{p}^n| \le 5; \quad 0.1 \le \hat{B}^n \le 1; \quad 1 \le n \le N \tag{27}$$

$$0 \le \hat{\bar{u}} \le 2; \tag{28}$$

The estimate of B, for the results shown in Figure (3), at the end of scans 15 and 16 are 0.43 and 0.45 respectively. Σ at the end of scan 15 and P^{15} in scan 15 are given in equation (29).

$$\Sigma = \begin{bmatrix} 0.05 & -0.05 \\ -0.05 & 0.06 \end{bmatrix}; \quad P^{15} = \begin{bmatrix} 0.22 & 0.16 \\ 0.16 & 0.16 \end{bmatrix} \tag{29}$$

From simulations of this type it is observed that, when provided with reasonable initial parameters and bounds, fairly accurate CD profile deviation estimates are obtained, in general, within ten scans but that the estimates of $\bar{u}$ and B, typically, take from ten to twenty scans to be reasonably accurate.

In simulations, we have knowledge of the true profile deviations and other parameters estimated by the procedure. However for industrial data such knowledge requires off-line tests done at exactly the same points on the sheet at which the parameters are estimated on-line. This kind of coordinated off-line tests while providing the best method of checking the performance of the procedure, is difficult and time consuming. It would be desirable to have some other independent on-line testing method.

A standard result in Kalman filtering theory is that the innovations sequence (prediction error) of the Kalman filter which matches the model of the measurement, is white [4]. If all the information about the parameters and states are being successfully extracted by the procedure, then it is reasonable to expect that the innovations sequence could be used as a performance measure.

6. APPLICATION TO INDUSTRIAL DATA

Raw data was collected on-line for a period of approximately 3.5 hours at a paper mill (Stone-Consolidated Inc., Laurentide Division, Quebec, Canada). The data was collected with the MD moisture profile automatic control system turned off and included a number of large MD transients. The number of CD points N is 58 and the average sampling period between points is 1.03s. The data comprised of 175 scans and included 9 breaks in continuity. All but 3 of the breaks were of 1 scan duration. The 3 larger breaks had a duration of 5, 16 and 6 scans.

The analysis of the data was started by first collecting the contiguous scans into "strings", numbered from 1 to 10. Out of the 10 strings, 3 strings in which the transients were relatively small (string 1 (14 scans), string 3 (28 scans) and string 7 (15 scans)) were used for fixing the parameters of the procedure. M in equation (1) was chosen as 5.50, a rounded up value of the average (5.385) of all data in string 1. The MD variance was estimated by using Dahlin's analysis of variance method [5] on strings 1 and 7. On both strings, the MD variance estimate was approximately 0.4. The value of R which results are reported here, is 0.01 and corresponds to a sensor range error of 0.6% ($6\times\sqrt{R}$). The value of R has been varied from 0.09 (sensor range error = 1.8%) to 0.0001 (sensor range error = 0.06%). The qualitative conclusions drawn from runs with different values of R are the same as that reported here for the case of values of a (correspondingly $q=0.4\times\{1-a^2\}$) varied from 0.5 to 0.99 in steps of 0.05. For each value of a the auto-correlation of the innovations sequence was plotted. It was judged (with a search grid size of 0.05) that at a=0.85 the auto-correlation best satisfies the whiteness criterion. For this value of a, q=0.11. The exponential forgetting actor used in all runs is 0.95. The initial states, parameters and their covariance were set as per equations (30) and (31).

$$\hat{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.30 & 0 \\ 0 & 0.401 \end{bmatrix} \tag{30}$$

$$\begin{bmatrix} \hat{p}^n \\ \hat{B}^n \end{bmatrix} = \begin{bmatrix} 0 \\ 0.50 \end{bmatrix}; \quad P^n = \begin{bmatrix} 100 & 0 \\ 0 & 4 \end{bmatrix}; \quad 1 \le n \le N \tag{31}$$

The bounds applied on the parameter estimates are given in equation (32). No bounds were applied on the state estimates $\hat{\bar{u}}$ and ξ.

$$|\hat{p}^n| \le 6; \quad 0.1 \le \hat{B}^n \le 1; \quad 1 \le n \le N \tag{32}$$

The procedure was run on all 175 scans continuously. The starting scan direction after each break, is found from that of the last scan before the break and the number of missing scans during the break. The initial estimates of the

parameters, states and covariances at the starting scan after each break are that estimated at the scan before the break.

The profile estimation part of Lindeborg's algorithm [1] and Dahlin's algorithm [5] were also run on this data. In Figure (4) we show the profiles estimated by all three algorithms at scan 25 and 150. In this figure, M = 5.50 has been added to the profile deviation estimates obtained from Lindeborg's and the new procedure. The Dahlin and Lindeborg algorithms provide very similar shapes for the profile estimates. They differ (approximately) by the MD mean which in the present case causes the Dahlin profile estimates to jump up or down relative to the Lindeborg profile estimates due to the presence of large MD transients. In Table 1, under the column a = 0.85, are shown The mean of the squared differences between profile deviation estimates obtained from Lindeborg's and the new procedure.

Table 1. Mean squared differences between profile deviation estimates obtained from Lindeborg's algorithm and the new procedure.

Scan	a = 0.85	a = 0.50	Scan	a = 0.85	a = 0.50
20	0.1953	0.0915	100	0.0800	0.0404
25	0.1833	0.0738	105	0.0451	0.0375
30	0.1600	0.0594	110	0.0489	0.0381
35	0.1486	0.0590	115	0.1128	0.0709
40	0.1078	0.0591	120	0.0894	0.0513
45	0.1133	0.0609	125	0.0652	0.0367
50	0.1051	0.0467	130	0.0754	0.0368
55	0.1158	0.0478	135	0.0965	0.0494
60	0.1030	0.0431	140	0.1191	0.0775
65	0.1014	0.0453	145	0.0765	0.0441
70	0.0837	0.0557	150	0.0584	0.0341
75	0.0804	0.0409	155	0.0502	0.0305
80	0.0476	0.0254	160	0.0511	0.0289
85	0.0366	0.0200	165	0.0549	0.0314
90	0.0520	0.0278	170	0.0336	0.0217
95	0.0627	0.0368	175	0.0237	0.0159

From this table and Figure (4), it is apparent that there are fairly appreciable differences between Lindeborg and the new profile estimates. We illustrate, in Figure (5), the initial portion of the auto-correlations (8192 points) of the innovations sequence obtained from the proposed procedure for the cases of a = 0.5 and a = 0.85. From this figure, it is seen that the auto-correlation for the case a = 0.85 is much more white than that of case a = 0.5. The mean of the squared differences between the profile estimates obtained from Lindeborg's algorithm and the proposed procedure for the cases a = 0.85 and a = 0.5 are shown in Table 1. From Table 1, we see that the profile estimates of Lindeborg are closer to those of the new procedure for the case of a = 0.5 than for the case of a = 0.85. However, from the auto-correlation functions, a = 0.85 seems to give better estimates.

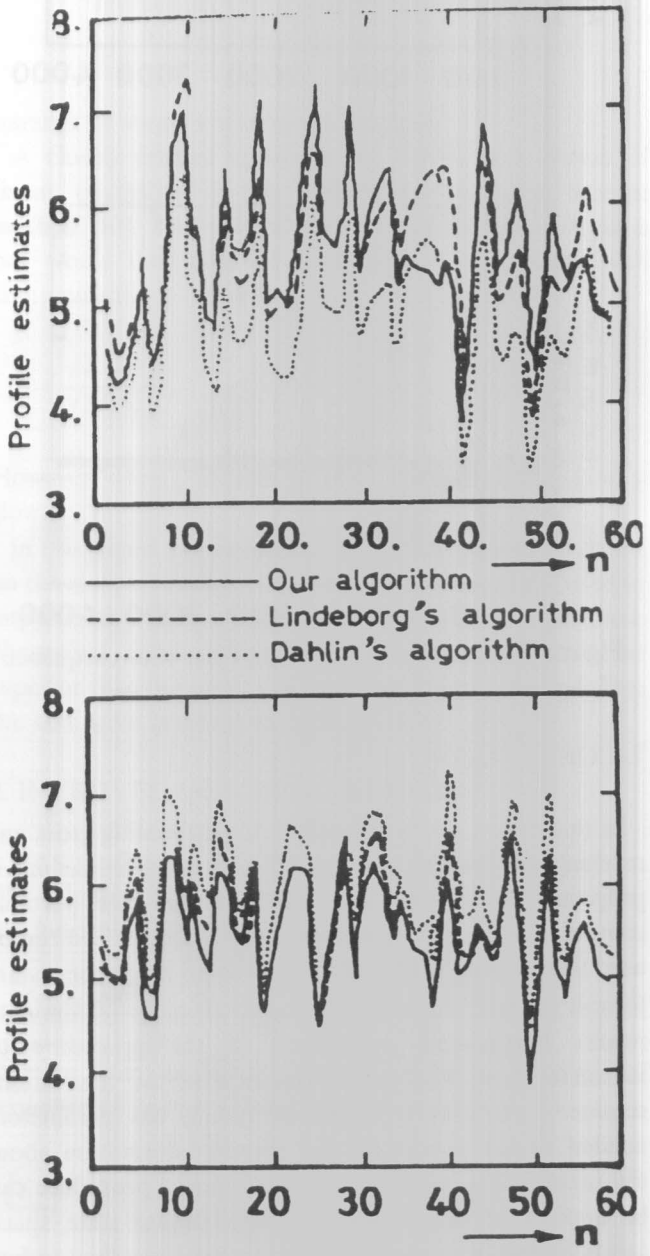

Figure 4. Profiles estimated by different procedures.

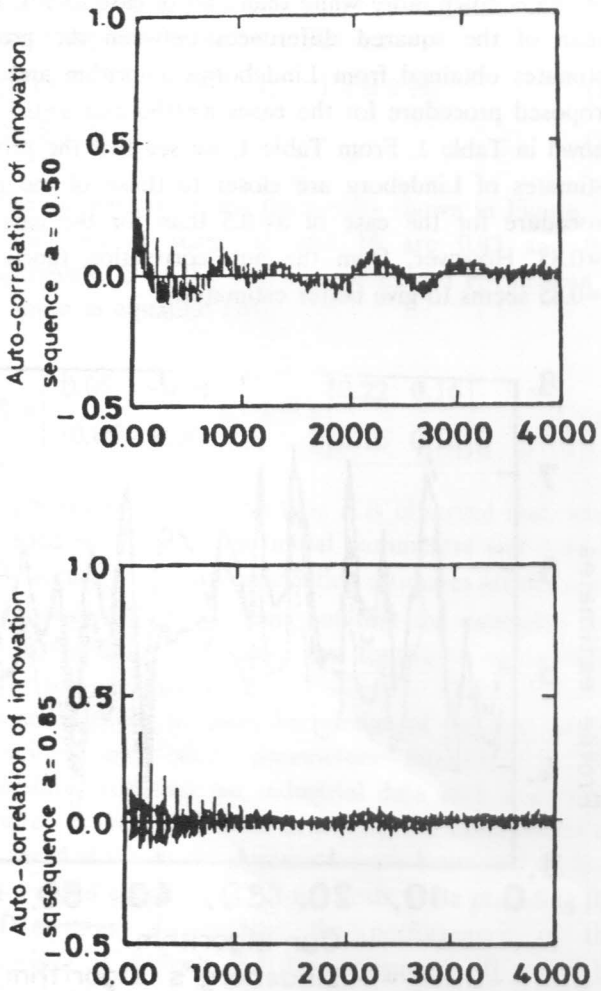

Figure 5. Auto-correlation of the innovations sequence of the industrial data.

7. CONCLUSION

In this paper, a new procedure for estimating cross and machine direction moisture profiles an a paper machine is proposed. It can be installed in a low-cost portable computer. It provides pointwise MD estimates and can be used to speed up MD control, as it is no longer necessary to wait for end of scan averaging to estimate $\bar{u}$. Simulate results encouraged application of the procedure to industrial data. The auto-correlation of the innovations sequence provides a means of tuning the parameters needed by the procedure and reveals information about the process. The structure of the proposed procedure can be applied to other problems with similar state space forms by the development of an appropriate (low order) ψ.

Acknowledgement

The author would like to thank Dr. J. Bland of Proconex Consultants Inc., Montreal, Quebec, Canada, for valuable discussions. This work was carried out while the author was on a study leave from Alexandria University in 1989.

REFERENCES

[1] C. Lindeborg, "A process model of moisture variations", *Pulp and Paper Canada*, 87:4, T142-147, 1986.

[2] R.A. El-Attar, "Multivariable control for paper making processes", *The Bulletin of the Faculty of Engineering*, Alexandria University, vol. XXIV, pp. 107-118, 1985.

[3] G.C. Goodwin and K.S. Sin, *Adaptive filtering, prediction and control*, Prentice Hall Inc., 1984.

[4] B.D.O. Anderson and J.B. Moore, *Optimal filtering*, Prentice Hall Inc., 1979.

[5] E.B. Dahlin, "Computational methods of a dedicated computer system for measurement and control on paper machines", *TAPPI 24th Engineering Conference*, San Francisco, California, Sept 14-19, 1969.