Identification of Full Profile Disturbance Models for Sheet Forming Processes

Apostolos Rigopoulos and Yaman Arkun
School of Chemical Engineering, Georgia Institute of Technology, Atlanta, GA 30332

Ferhan Kayihan
Weyerhaeuser Technology Center, Tacoma, WA 98477

In this article we present a method for the on-line identification and modeling of full profile disturbance models for sheet forming processes. A particular principal components analysis technique called the Karhunen-Loève expansion is used to adaptively identify the significant features of the profile. In addition, we show how the temporal modes of the reconstructed profile can be modeled using low-order linear autoregressive (AR) processes. By simulation examples, the effect of the order of the AR model is studied, as well as the window size of the data used in the on-line application of the KL expansion, the effect of data weighting, the importance of the correct selection of the number of modes, and the frequency of updating the parameters of the AR models. Identified disturbance models can be easily incorporated into model-predictive control algorithms.

Introduction

Continuous sheet forming appears in many manufacturing processes including metal rolling, polymer film extrusion, coating, fabric dyeing, and paper making. In all these industrial applications the ultimate objective is to maintain certain properties under tight and uniform control across the sheet. For example, in paper machines, one is interested in controlling the basis weight, moisture and caliper of the paper sheet as uniformly as possible. Similarly in film processing, metal rolling and coating, thickness control is important. Finally, in textile dyeing processes, minimization of shade variation across the fabric surface is a major problem. Poor control of continuous sheet forming processes leads to unacceptable variability in product quality, significant energy and raw material losses, and production of off-spec materials. With tighter quality requirements and large quantities of "sheet" production by various process industries, the added value of any improvement in monitoring and control systems is quite apparent.

In this work we address the problem of building disturbance models which can be used for *full profile* control of continuous sheet forming processes. In the present context *full profile* refers to the evolution of the sheet properties in 2-D, that is, spatially across the sheet and temporally along the direction of the movement of the sheet. Due to our experiences with pulp and paper processes, we will use the paper machine as an example to illustrate the concepts and our approach without any loss of generality. In a paper machine, the sheet properties vary both in the direction along which the paper moves (machine direction or MD), and the direction across the width of the machine (cross direction, CD). These properties are measured either by a fixed array of optical sensors or by a scanner mounted on a carriage which moves back and forth in the cross direction (see Figure 1). In the case of an array of sensors, one has access to the full CD profile at each sampling time while only a partial zig-zag profile of the sheet is available via a scanner. In a typical machine, it is not uncommon to have hundreds of CD measurement points per scan resulting in large data sets.

To effectively control the sheet properties of interest, one needs to adequately model the disturbances that affect the system. Early attempts have assumed that the disturbance profile can be separated into three components: an MD-only, a CD-only, and a residual. Furthermore, the variance of those three terms was assumed to be independent from one another, and MD and CD components were controlled sepa-

Correspondence concerning this article should be addressed to Y. Arkun.
Current address of F. Kayihan: IETek—Integrated Engineering Technologies, 5533 Beverly Ave. NE, Tacoma, WA 98422.

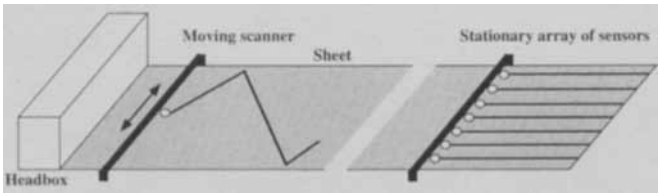


Figure 1. Property profile measurement alternatives.

rately (Wilhelm and Fjeld, 1983; Wilhelm, 1984; Taylor, 1991). Adopting the traditional assumption that CD changes occur more slowly than those of MD, Wang et al. (1993b) proposed an algorithm to estimate basis weight and moisture content variations from scanned measurements using a Kalman filter with an exponential resetting and forgetting algorithm. The same authors also implemented their estimator using fixed array sensors and noted the potential advantages (Wang et al., 1993a). The recent work of Tyler and Morari (1995) discussed the effects of the scanning rate and the noise characteristics on the CD profile estimation using a Kalman filter based on the "lifting" approach. Chen (1992), Rawlings and Chien (1996), and Bergh and MacGregor (1987) also designed Kalman filters but they did not use the "lifting" approach.

This article differs and complements the works cited above in that its main objective is to identify disturbance models directly from available full profile data, whereas most of the existing literature assumes a disturbance model and tries to estimate the full profile. For this work, we assume that the full profile is measured by fixed array sensors. This new measurement technology (such as Francis et al., 1988; Anon., 1993) is finding its way through many production facilities and offering an excellent opportunity to improve existing modeling, estimation, and control algorithms. If a scanner is used instead, then the proposed analysis would apply to the full profile after it is estimated from scanned measurements. Another very important distinction in the present work is that the emphasis is on the development of reduced-order models. In sheet forming, whether working with scanners or stationary sensors, one is confronted with large sets of correlated and noisy data which have to be efficiently managed during estimation and closed-loop control. In this work we are specifically interested in: (a) *extracting* from a large data set only the significant features of the underlying profile; (b) *identifying* reduced-order dynamic profile models to describe these features.

The method we use is the Karhunen-Loève (KL) expansion which is a principal components analysis technique. With this statistical method, spatio-temporally varying disturbance profiles are modeled by projecting the data onto a *lower* dimensional subspace spanned by orthogonal functions calculated directly from the data. The projection is performed in such a way that the dominant characteristics of the original profile are captured by the first few modes of the expansion. Thus, the method provides order reduction, feature extraction, noise filtering, and data compression. The time series of the coefficients of expansion are modeled by autoregressive (AR) models which provide P-step ahead predictions for the disturbance profile. These predictions can then be used in a model predictive control algorithm where the controller responds only to the most relevant disturbance profile at any given time.

In this work, we are not addressing the issue of how the actuators are mapped to the measured profile. Instead, we assume that we are given a model that perfectly describes the relation between the actuators and the outputs, and generate the disturbance profile by subtracting the effect of the actuators from the measured profile.

An outline of the KL expansion is provided which focuses on the on-line implementation of this method. This technique is applied to typical disturbance profiles to illustrate the effects of various parameters, such as the window size of the data onto which the KL expansion is applied on-line, the weighting of the data, the selection criterion for the number of modes, and the order of the AR models used.

Theoretical Background

KL expansion

The KL expansion is a principal components analysis technique applied to random processes. It was originally developed in a probability space to expand random functions as linear combinations of orthogonal basis functions with uncorrelated random coefficients (Karhunen, 1947; Loève, 1963). For an excellent review of the topic, see Preisendorfer (1988) and references thereof. The technique has found many applications, some of which are in pattern recognition and image analysis (Savoji and Burge, 1985; Ogawa and Oja, 1986), modeling and control of distributed parameter systems (Gay and Ray, 1988; Chen and Chang, 1992; Park, and Cho, 1996), model identification of catalytic surface reactions (Krischer et al., 1993), and low-dimensional modeling in turbulent flow (Berkooz et al., 1993; Sirovich and Everson, 1992). Here, we present only the pertinent features of the method.

Define $z_\omega(\boldsymbol{x},t)$ as the ωth realization of a random vector field where t and $\boldsymbol{x}$ denote the temporal and spatial dependencies, respectively. Next consider the deviation from the average

$$\bar{z} = \bar{z}(\boldsymbol{x},t) = z_\omega(\boldsymbol{x},t) - \langle z_\omega(\boldsymbol{x},t)\rangle_\omega \tag{1}$$

where the ensemble average $\langle\cdot\rangle$ is taken over the observed realizations. $\bar{z}(\boldsymbol{x},t)$ is the time-averaged $z_\omega(\boldsymbol{x},t)$. The KL analysis then considers the problem of finding an orthonormal basis set $\{\boldsymbol{\phi}^{(n)}(\boldsymbol{x})\}$ such that for any N the Euclidean norm of the error

$$e_N = \left\| \bar{z} - \sum_{n=1}^{N} a_n(t)\boldsymbol{\phi}^{(n)}(\boldsymbol{x}) \right\|^2 \tag{2}$$

is minimized in an average sense, i.e., $\langle e_N\rangle$ is minimum. Assuming ergodicity (that is, ensemble and temporal averages are the same), $\{\boldsymbol{\phi}^{(n)}(\boldsymbol{x})\}$ are the eigenfunctions of the integral equation

$$\int_X K(\boldsymbol{x},\boldsymbol{y})\boldsymbol{\phi}^{(n)}(\boldsymbol{y})d\boldsymbol{y} = \lambda_n\boldsymbol{\phi}^{(n)}(\boldsymbol{x}) \tag{3}$$

The kernel $\boldsymbol{K}$ is defined by the two-point correlation matrix

$$K_{ij} = \langle \bar{z}_i(x)\bar{z}_j(y)\rangle = \frac{1}{T}\int_{-T/2}^{T/2} \bar{z}_i(x,t)\bar{z}_j(y,t)dt \qquad (4)$$

where integration is done over a sufficiently long time. The temporal modes or the principal components $\{a_i(t)\}$ are calculated by projecting the data on the eigenfunctions or principal directions

$$a_n(t) = (\boldsymbol{\phi}^{(n)}, \bar{z}) = \int_X \boldsymbol{\phi}^{(n)} \cdot \bar{z}(x,t)dx \qquad (5)$$

where $(\cdot)$ denotes the inner product. One important consequence of the KL expansion is that $\{a_i(t)\}$ are uncorrelated

$$\langle a_n(t)a_m(t)\rangle = \lambda_n \delta_{nm} \qquad (6)$$

It also follows that the eigenvalues are given by $\lambda_n = \langle (\bar{z}, \boldsymbol{\phi}^{(n)})^2 \rangle$. Consequently, each eigenvalue λ_n of the KL expansion can be interpreted as a measure of the mean energy or variance explained along the direction of each eigenfunction with the total mean energy $E = \Sigma_n \lambda_n$. δ_{nm} is the Kronecker delta. Having introduced the KL expansion, we can now adapt it to one class of sheet forming processes, namely the paper machine.

Consider a paper machine with N CD positions. In the case of the array sensors we have access to N CD measurements taken in M time intervals or MD positions ($N < M$). These measurements are stored in a matrix $\boldsymbol{Z}$ of size $N \times M$ where each column vector contains the full CD profile corresponding to a particular MD position. In the case of the moving scanner the data matrix $\boldsymbol{Z}$ is constructed similarly by replacing the missing CD measurements with their estimates which can be obtained from an estimator.

The first step in the KL procedure is to mean-center the data by subtracting the temporal average from each CD lane or row of $\boldsymbol{Z}$ to form a time-averaged matrix $\bar{\boldsymbol{Z}}$ with entries

$$\bar{Z}(n,k) = Z(n,k) - \frac{1}{M}\sum_{i=1}^{M} Z(n,i)$$
$$n = 1, \ldots, N; \quad k = 1, \ldots, M \qquad (7)$$

Then, the elements of the two-point spatial correlation matrix $\boldsymbol{K}$ are computed from

$$K(i,j) = \frac{1}{M}\sum_{k=1}^{M} \bar{Z}(i,k)\bar{Z}(j,k) \qquad i,j = 1, \ldots, N \qquad (8)$$

Let now $\{\boldsymbol{\phi}^{(n)}\}$ be the set of eigenvectors of $\boldsymbol{K}$ (also called the spatial modes). n is the lane index. Then, the finite KL expansion of the data set over the finite sample space can be written as a linear combination of the eigenvectors of $\boldsymbol{K}$, with the coefficients of the expansion being uncorrelated random variables on that space, that is

$$\bar{z}(k) = \sum_{n=1}^{N} a_n(k)\boldsymbol{\phi}^{(n)} \qquad (9)$$

where the column vector $\bar{z}(k)$ is the time-averaged full CD profile at time k. The temporal modes $\{a_n\}$ can be computed as

$$a_n(k) = \bar{z}(k)'\boldsymbol{\phi}^{(n)} \qquad (10)$$

This procedure essentially diagonalizes the correlation matrix of the original data since

$$\boldsymbol{\Phi}'\boldsymbol{K}\boldsymbol{\Phi} = \boldsymbol{\Lambda} = E\{\boldsymbol{A}\boldsymbol{A}'\} \qquad (11)$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of $\boldsymbol{K}$; $\boldsymbol{A} = [\boldsymbol{a}_1 \cdots \boldsymbol{a}_n]'$ and $\boldsymbol{\Phi} = [\boldsymbol{\phi}^{(1)} \cdots \boldsymbol{\phi}^{(n)}]$. $\boldsymbol{A}$ is the matrix of temporal vectors; $\boldsymbol{a}_n$ is the nth temporal vector. Equation 11 also shows that the temporal modes $\{\boldsymbol{a}_i\}$ are uncorrelated as indicated earlier.

The KL expansion has some additional properties that deserve special attention. First of all, instead of using *a priori* chosen basis functions, it uses the data set itself to determine a natural set of coordinates to analyze it. For this reason, the eigenvectors are often referred to as *empirical orthogonal functions*. Among the class of all linear decompositions, the KL method is optimal in the sense that the KL expansion $\Sigma_{n=1}^{L} a_n(k)\boldsymbol{\phi}^{(n)}$ on a subspace of lower dimension $L < N$ retains the most average energy possible (Berkooz et al., 1993). In other words the % variance $100 \times \Sigma_{n=1}^{L}\lambda_n / \Sigma_{n=1}^{N}\lambda_n$ is maximized, where each eigenvalue represents the mean energy of its mode. In practical applications given a desired level of % variance (such as 90%), one can compute the number of significant modes necessary to meet the specified accuracy. The KL expansion including only the first L modes will extract the main features (to the given accuracy) from the data and projects to lower dimension "filters" the random noise along with other insignificant modes. In the literature such applications of the KL method are all performed off-line to a given batch of data. In our work we implement it on-line in an adaptive fashion as described next.

On-line implementation of the KL analysis

For on-line implementation, the KL procedure needs to be automated and repetitively applied in order to take advantage of the new measurements becoming available at each sampling time. Also, old data should be discarded as their importance in predicting subsequent disturbance profiles is minimal, and may prohibit the KL method from quickly identifying a change in the profile. Therefore, we use a trailing window estimator where at each sampling time k, the KL expansion is applied to a data matrix $\boldsymbol{Z}$ of size $N \times W$ where W denotes the number of past CD profiles contained in the moving window. From one sampling time to another, the data matrix $\boldsymbol{Z}$ is updated by incorporating the most recent data vector and dropping the oldest one, and the KL expansion is repetitively applied. The window size W is treated as a tuning parameter. Also, exponential weighting (spatially uniform) can be applied to each column vector of $\boldsymbol{Z}$ to forget the influence of older data so that the KL expansion can quickly adapt to new profile patterns (see Figure 2).

A critical point in the above adaptive application of the KL expansion is the selection of the subspace order L, alternatively defined as the number of modes retained in the expan-

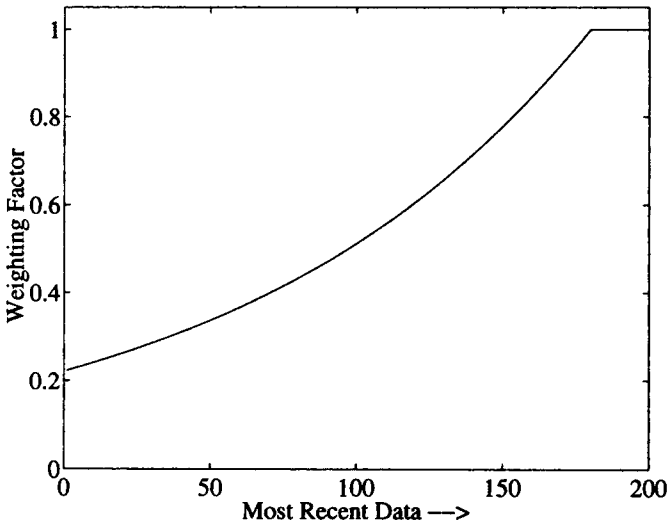


Figure 2. More emphasis can be placed on the most recent data by multiplying the data through a weighting function.

sion. A small value could fail to capture all the significant patterns in the particular window of data leading to poor disturbance predictions. On the other hand, selecting a large number of modes will capture the significant profile features, but data filtering would be reduced since some of the included modes would correspond to noise or insignificant patterns, which we do not want to model. A criterion is needed for the selection of the number of modes; below, we give two criteria.

% Variance Criterion. Here the number of significant modes is selected to satisfy a desired value of % variance, as shown in Figure 3. It is simple to implement on-line and works very well when the data do not contain significant additive white noise. However, as the signal-to-noise ratio decreases, it becomes increasingly more difficult to specify the cut-off value for the desired percent variance. Too small a value filters the noise but misses some significant modes, while too large a value would start modeling the noise.

Scaled Eigenvalues Criterion. For additive white noise, the % variance varies almost linearly with the number of modes, since each noise mode contributes almost equally to the total variance. To separate these modes from the rest, the % variance is differenced and plotted against the number of modes. Differencing amplifies the contribution of the modes which correspond to the noise-free signal, thus making the cut-off point less ambiguous. When normalized by the % variance contribution of the most significant mode, this plot is equivalent to the distribution of the scaled eigenvalues λ_i/λ_{max}. The distribution curve starts with a large slope (due to the most significant modes) and reaches an almost straight line with a small slope (due to insignificant modes and additive noise), as shown in Figure 3. This test has also appeared in the factor analysis literature (Cattell, 1966) as the "scree" test. Here we implement this criterion on-line by discarding the modes which are aligned on a straight line in the least-squares sense by checking the coefficient of determination r^2 of the line against a preset desired value.

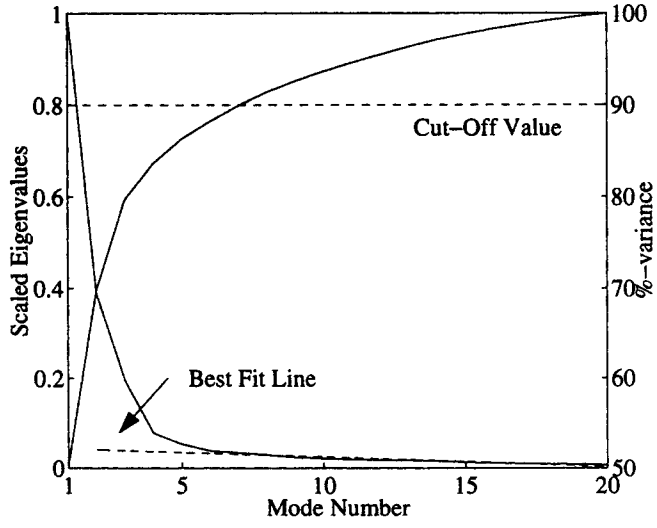


Figure 3. Typical plot of scaled eigenvalues and % variance.

Dashed line is the best fit line passing through the last 14 CD positions, thus picking six significant modes.

Modeling of temporal modes

We have shown how the KL expansion can be used adaptively to identify the significant patterns in the measurements. However, for control purposes, we need a dynamic model to describe the profile. In the subspace defined by the KL expansion, this implies that we need to model the spatial $\{\phi^{(n)}\}$ and temporal $\{a_n(k)\}$ modes and use their predicted values to reconstruct the profile P steps into the future. The temporal modes are modeled using rth order linear autoregressive models AR(r)

$$a_n(k) = \sum_{i=1}^{r} \theta(n,i) a_n(k-i) + e_n(k) \qquad (12)$$

e_n is the residual of the AR fitting of the nth temporal vector. $\theta(n,i)$ is the autoregressive coefficient i ($i = 1, \ldots r$), for the nth temporal vector. P-step full profile predictions are then computed from

$$\hat{z}(k+P|k) = \sum_{n=1}^{L} \hat{a}_n(k+P|k)\boldsymbol{\phi}^{(n)} \qquad (13)$$

In the above expression we assume that the spatial eigenvectors $\phi^{(n)}$ remain constant for the entire prediction horizon, but they are updated at every sampling time k. $\hat{a}_n$ is the estimate of the nth temporal vector.

In general, the selection of the order of the AR processes requires several tests based on criteria such as the Akaike Information Criterion. However, such tests require considerable computation time, if they are to be performed at every sampling time. Therefore, in the simulations we have fixed the order to one ($r = 1$) and two ($r = 2$) and evaluated the overall results, accordingly. Next, we present some examples to demonstrate the application of the method for batch and on-line analysis.

Batch Analysis

In this section we apply the KL expansion to three types of stochastic disturbance profiles. To easily convey the main ideas, the analysis is first done in a batch manner deferring the on-line features to a later example: data are generated and stored in matrix $\mathbf{Z}$; they are subsequently time-averaged and KL expansion is applied; finally, the reconstructed profile is generated based on the number of modes selected by our criterion.

Example 1: Uncorrelated white noise profile

The entire disturbance profile is composed of uncorrelated two-dimensional white noise with zero mean and identity covariance matrix. Theoretically, since the data are completely uncorrelated, the KL decomposition should produce N equal eigenvalues. However, in practice, the randomness of the data is as good as the algorithm used to generate them. Increasing the number of samples taken makes the data generated approach better a "true" 2-D white noise profile. Figure 4 shows a plot of the % variance and the scaled eigenvalues of a white noise profile where 50,000 full-scan samples were generated. One can see that the eigenvalues are very close to each other, whereas the % variance plot is linear. The linearity of the % variance plot is characteristic of a profile where the data are completely uncorrelated. All directions are equally likely; therefore, all eigenvalues are almost equal.

For such a profile, our selection criterion indicates that zero modes should be used and that the profile should be described by its mean.

Example 2: Autoregressive profile driven by spatially correlated white noise

Frequently in the literature (such as Bergh and MacGregor, 1987), the disturbance profile is modeled as an autoregressive process driven by white noise that is spatially correlated. For illustration purposes, we assumed a first-order autoregressive model of the form

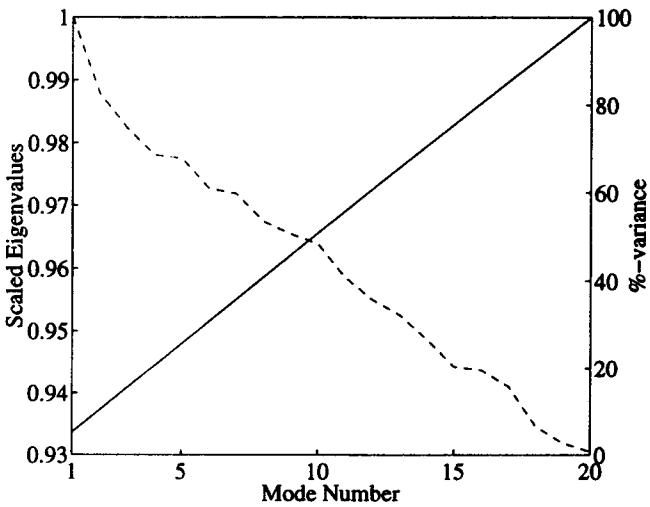


Figure 4. Scaled eigenvalues (--) and % variance plots (——) for an uncorrelated white noise profile.

$$\mathbf{y}(k) = 0.9\,\mathbf{I}_N\mathbf{y}(k-1) + \mathbf{w}(k)$$
$$\mathbf{z}(k) = \mathbf{y}(k) + \mathbf{v}(k) \tag{14}$$

$\mathbf{y}(k)$ is the measurement noise-free CD profile vector at time k where $\mathbf{w}(k)$ is an N-dimensional white noise vector with zero mean and a Toeplitz symmetric covariance

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ & 1 & \rho & \cdots & \rho^{N-2} \\ & & \ddots & & \vdots \\ & & & \ddots & \rho \\ \text{sym} & & & & 1 \end{bmatrix} \sigma_w^2 \tag{15}$$

and $\mathbf{z}(k)$ is the measured profile at time k corrupted by measurement noise $\mathbf{v}(k)$ which was assumed to be uncorrelated white noise with signal-to-noise ratio equal to 5. σ_w^2 is the variance of each lane.

ρ, the degree of correlation between adjacent lanes, can assume values between 0 and 1, implying pure randomness and complete correlation in the CD direction, respectively. We can argue intuitively that when ρ is very small, the KL expansion will need a lot more modes to describe the profile than when ρ is large. In the limit, when ρ is zero, all N modes will be needed, whereas when $\rho = 1$, only one mode will be sufficient.

Figure 5a illustrates the % variance and the scaled eigenvalues resulting from a KL expansion of the data matrix for three ρ values. Notice the similarity in the shape of both curves for $\rho = 0$ with the curves of the white noise profile analyzed above. Another indication of the correlation between adjacent lanes can be seen in Figure 5b where the first spatial mode is plotted as a function of ρ. One can see that the spatial excitation of the mode is reduced as ρ increases.

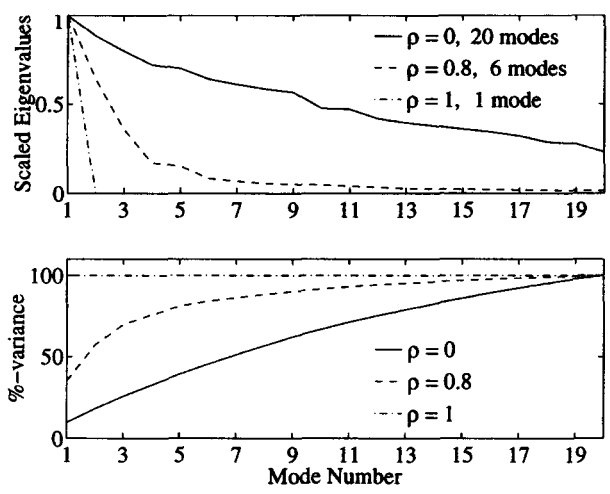
For $\rho = 0.8$, our selection criterion indicated that 6 modes are necessary to capture the significant profile patterns. For this case, the measurement noise-free profile to be approximated is shown in Figure 6a, whereas its reconstructed profile is shown in Figure 6b. Comparison shows that projection of data to a lower dimensional subspace filters the high frequency modes of the profile. The difference between the original (measured) profile and the reconstructed one is shown in Figure 7a, whose 2-D cross-correlation shown in Figure 7b suggests that the reconstruction error has white noise characteristics.

Example 3: Autoregressive profile driven by white noise filtered through a Toeplitz disturbance gain matrix

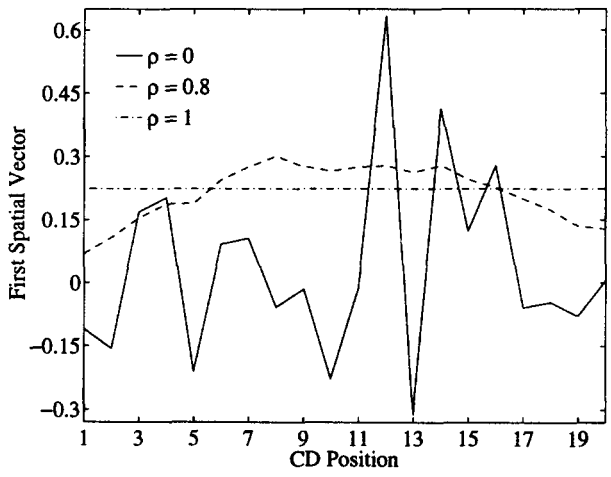
An alternative representation of the disturbance profile is to assume that spatially uncorrelated white noise is integrated through a Toeplitz disturbance gain matrix (Rawlings and Chien, 1996)

$$\mathbf{y}(k) = 0.9\,\mathbf{I}_N\mathbf{y}(k-1) + \mathbf{G}\mathbf{w}(k)$$
$$\mathbf{z}(k) = \mathbf{y}(k) + \mathbf{v}(k) \tag{16}$$

where $\mathbf{G}$ is a Toeplitz symmetric matrix; $\mathbf{w}$ has zero mean and covariance $\mathbf{Q}_w = q\mathbf{I}_N$; ($\mathbf{I}_N$ is an identity matrix of size N)



(a)

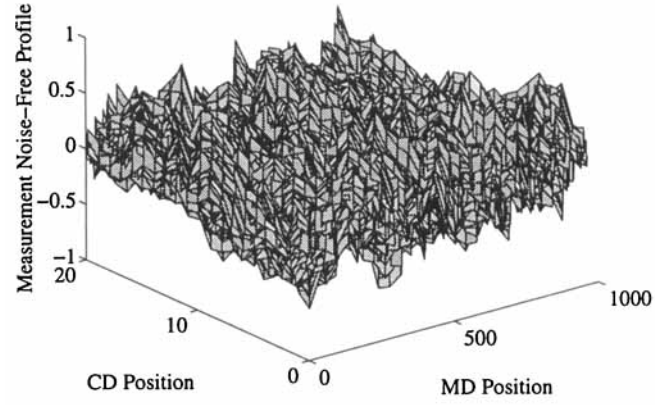


(b)

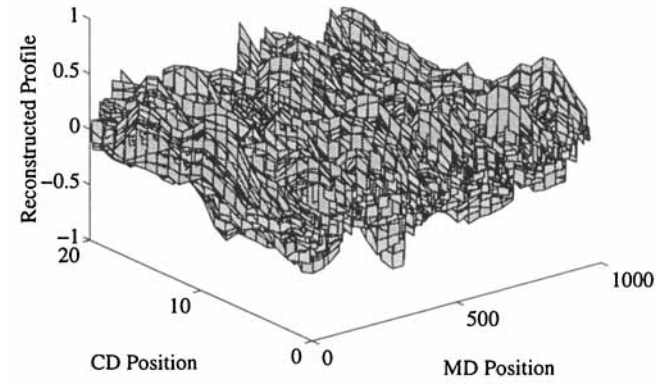
Figure 5. (a) Scaled eigenvalues and % variance for an AR profile with covariance given by Eq. 15 for different ρ values; (b) corresponding first KL spatial vectors.

and the other quantities are as described in the previous example. This disturbance is different from the previous one, given by Eqs. 14 and 15, due to the fact that both Σ and G are constrained to be Toeplitz symmetric. Given Σ, in general, one cannot find a corresponding Toeplitz symmetric G. Under limiting conditions, however, the two representations coincide: a full G matrix with all its elements equal corresponds to a Σ with $\rho = 1$; a diagonal G matrix corresponds to a Σ matrix with $\rho = 0$. The more nonzero bands G has, the more CD lanes will be correlated.

For G = Toeplitz[1 .8 .6 .4 .2 0 $\cdots$ 0] of size 20×20 with diagonal elements equal to 1, first upper and lower diagonal entries are equal to .8, and so on, and a representative measurement noise-free profile is illustrated in Figure 8a, whereas its reconstructed one using seven modes selected by our criterion is shown in Figure 8b. A comparison of the original profiles in Figures 6a and 8a indicate that matrix G acts as a "filtering" mechanism to the input noise.



(a)



(b)

Figure 6. (a) Original measurement noise-free profile with $\rho = 0.8$; (b) reconstructed profile using six modes.

The scaled eigenvalues and the % variance of the disturbance profile generated with different G matrices is shown in Figure 9a. Again, the conclusions are similar to the previous case: The more correlated the CD lanes are, the less modes are needed to effectively describe the disturbance profile.

Finally, for the profile shown in Figure 8a, the first and eighth temporal modes are shown in Figure 9b. The eighth temporal mode which is the first mode not included in the KL expansion is a lot noisier than the first one, an indication that it attempts to model higher frequency noise contribution or less significant profile patterns.

Adaptive Analysis

As discussed in the previous section, due to the nonstationary nature of disturbances, the KL procedure needs to be repetitively applied at every sampling time to take advantage of the new measurements. Furthermore, since we are primarily interested in making P-step ahead predictions of the disturbance profile, it becomes obvious that old data need to be discarded; otherwise, the effect of emerging disturbances will not be quickly identified leading to possibly poor profile predictions.

In this section we will illustrate through a simulation example the adaptive application of the method and discuss the

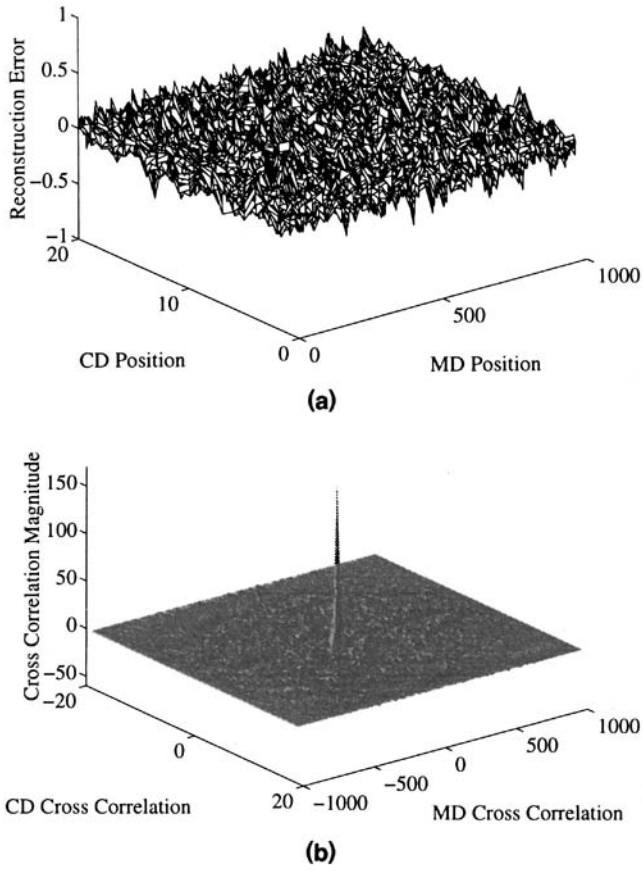


Figure 7. (a) Residual between original and reconstructed profile; (b) residual cross correlation.

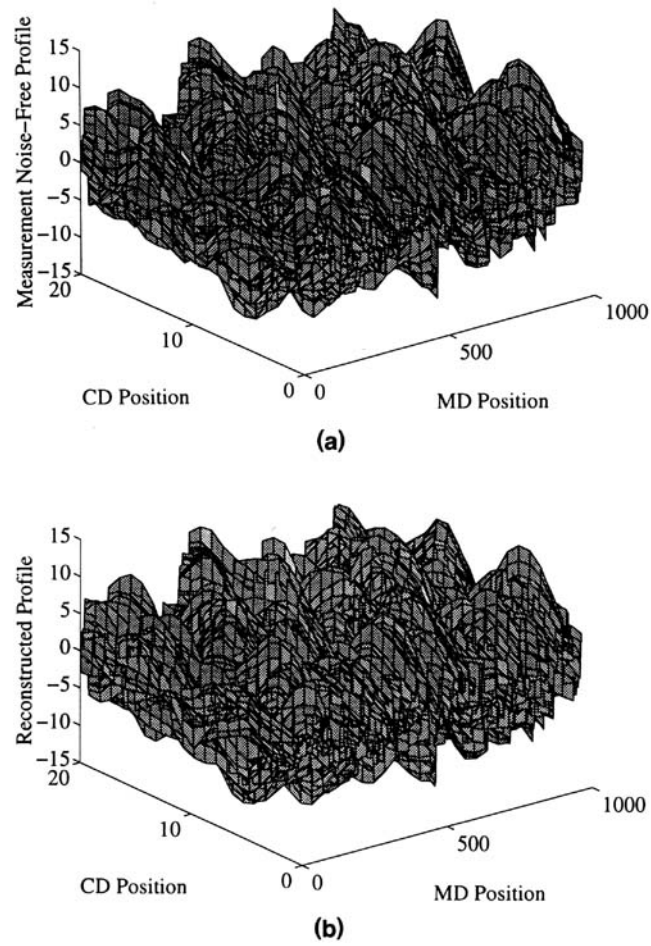


Figure 8. (a) Original measurement noise-free profile for a 9-diagonal G matrix; (b) reconstructed profile using seven modes.

effect of the window size selection, and the order of the AR model used for the temporal modes. Also, we will show how exponential weighting of the data matrix prior to the application of the KL expansion leads to better disturbance predictions during transition regions.

Example 4: A disturbance profile composed of sinusoids

The profile is composed of three sections with different frequency characteristics and varying CD correlation. Such variations may be attributed to inconsistencies in pulp flow rate and chemistry coming out of the headbox. We have assumed that the disturbance can be modeled as a linear combination of three patterns

$$Y_1(n,k) = \sin\left(\frac{4\pi k}{K} + \frac{10\pi}{2(N-1)}\right),\ k = 1, \ldots, 1{,}000$$

$$Y_2(n,k) = \begin{cases} 0.5\sin\left(\dfrac{20\pi k}{K} + \theta_2(n)\right), & k = 1, \ldots, 500 \\ 0.5\sin\left(\dfrac{(20 - 0.05\lfloor (k-499)/2 \rfloor)\pi}{K} + \theta_2(n)\right), & \\ \qquad k = 501, \ldots, 1{,}000,\ \theta_2(n) = \dfrac{(n-1)\pi}{2(N-1)} & \end{cases}$$

$$Y_3(n,k) = \sin\left(\frac{4\pi n}{N} + \frac{4\pi(1{,}000 - k + 1)n}{KN}\right),$$

$$k = 1, \ldots, 1{,}000 \quad (17)$$

where $K = 1{,}000$, $N = 20$, and $\lfloor x \rfloor$ rounds x to the nearest integer that is smaller than x. The distinguishing features of the above patterns are the following:

- Y_1 describes a profile with MD variations only. (Y is the measurement noise-free profile.)
- Y_2 uses a spatially varying phase shift θ_2 to allow for nonuniform velocity profiles among the CD lanes. Furthermore, its temporal evolution is much faster than for Y_1 and it allows for more complicated MD variations. In general, when the arguments of the sinusoids are functions of both the MD and CD locations, the generated profile has its peaks and troughs change lanes as a function of time producing a "wandering" effect. When these expressions are linear, the lane shifting is also a linear function of position.
- Y_3 allows for nonlinear "wandering." We used a simple nonlinear dependency of the MD and CD positions in the argument of the sinusoid.

The sections of the disturbance profile are as follows:

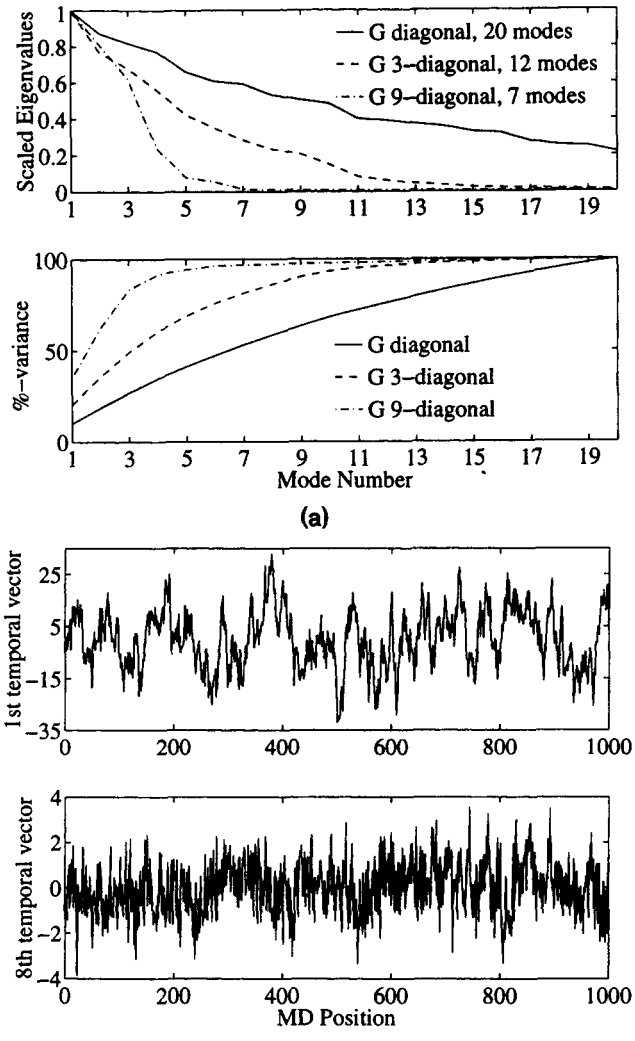


Figure 9. (a) Scaled eigenvalues and % variance for an AR profile given by Eq. 16 for different G matrices; (b) first and eighth temporal modes for a 9-diagonal G matrix.

- *Section 1:* MD Positions 1 to 500. Only MD disturbances are assumed, hence, only Y_1 is included.
- *Transition Region 1:* MD Positions 501 to 600. In addition to the MD-only disturbances, CD variations in the form of Y_2 are also beginning to appear in a gradual way.
- *Section 2:* MD Positions 601 to 900. Y_1 and Y_2 are both present.
- *Transition Region 2:* MD Positions 901 to 1,000. The MD-only disturbances Y_1 gradually disappear leaving Y_2 as the only disturbance.
- *Transition Region 3:* MD Positions 1,001 to 1,100. Nonlinear wandering effects are introduced through Y_3.
- *Section 3:* MD Positions 1,101 to 1,500. Y_2 and Y_3 are both present.

Figure 10 shows a 2-D contour plot of the entire profile with the dashed lines indicating the beginning and end of the transition regions. Notice how the uniform CD profile in section 1 is replaced by CD variations where the peaks are linearly evolving in section 2, and how they are drifting among adjacent lanes in section 3.

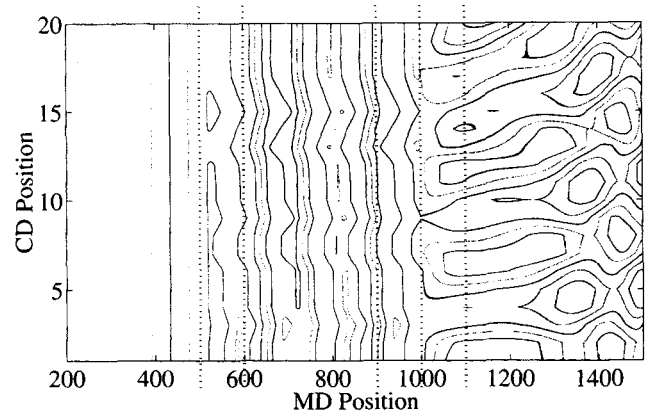


Figure 10. 2-D contour plot of the measurement noise-free profile used in subsequent simulations.

Dashed lines outline the beginning and end of the transition regions.

Effect of the window size W

To examine the effect of the window size W on the number of modes selected, we simulated the above profile for two window sizes, $W = 200$, and $W = 100$. The results are presented in Figures 11a and b, respectively. The main conclusions that can be drawn from these plots are:

- On average, the number of modes selected is higher for $W = 200$. This is because a larger profile contains more information which requires more modes to describe it.
- There is less mode switching for the larger window size, because the effect of the inclusion of the new data vector is smaller than when the window size is smaller.

The above remarks suggest that the window size should be viewed as a tuning parameter. It should be stressed that apart

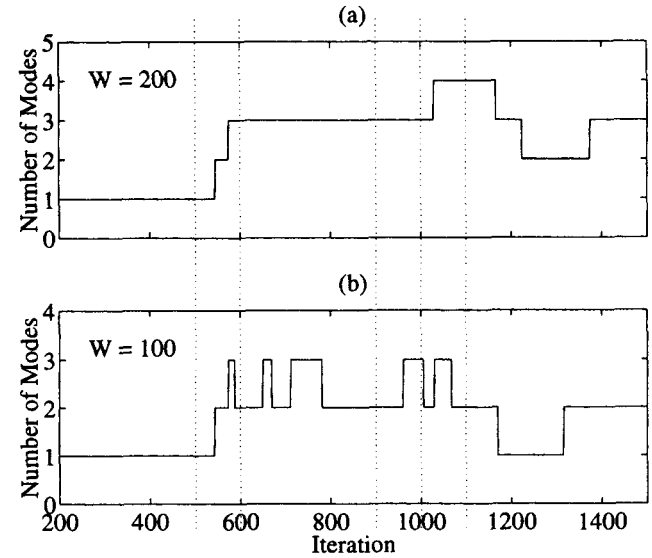


Figure 11. Number of modes selected using the criterion of scaled eigenvalues for window sizes $W = 200$, and $W = 100$.

Dotted lines define the boundaries of the transition regions.

from correctly identifying the disturbance profile, our second main objective is to use the KL procedure for profile predictions. Consequently, the building of a dynamic model for a fast changing disturbance profile should rely on recent data as the inclusion of older data would hinder the tracking of the emerging disturbances. On the other hand, a very small window size would make the parameters of the dynamic model more sensitive causing "aggressive" profile predictions which is also undesirable.

Effect of the order of AR model and frequency of parameter updating

To assess the effect of the autoregressive modeling on the profile predictions we performed simulations for three cases: (a) when the temporal modes are modeled using an AR(1) process whose parameter is updated at every sampling time; (b) same as (a) but an AR(2) model is used; (c) an AR(2) model is used again, but its parameters are updated only when the number of modes changes. Of course, as it was mentioned earlier, a more elaborate approach would identify the most appropriate model order by using an Information Criterion, but this is a very time-consuming operation and it was omitted in the on-line implementation. In all the simulations the window size was set to 200 and the data were equally weighted.

The sum of square prediction errors normalized by the number of data points (NSSPE) for $P = 1$ and $P = 10$ are shown in Table 1 for each data region separately. The results indicate that for all regions an AR(2) model yields better results than using an AR(1) model. We have also done simulations with an AR(3) model, but the results were very similar to the ones obtained using an AR(2) model and hence are not shown here. The fact that an AR(2) model is deemed appropriate is, of course, related to the profile under consideration. If an "easier" profile were to be modeled, perhaps an AR(1) model would be as effective as an AR(2) model. More interesting, however, are the results from case (c) where the parameters of the AR(2) model are updated only during mode changes. For the profile at hand, this occurred only seven times as seen in Figure 11a, therefore saving significant computation time. The results indicate that the additional NSSPE generated by not constantly updating the parameters of the AR model has increased on the average for the entire simulation by 12% for $P = 1$, and 23% for $P = 10$. The reason we do not see a much higher error may be attributed to the fact that the AR parameters change drastically only during transitions 1 and 3 (Figure 12). We also considered the case

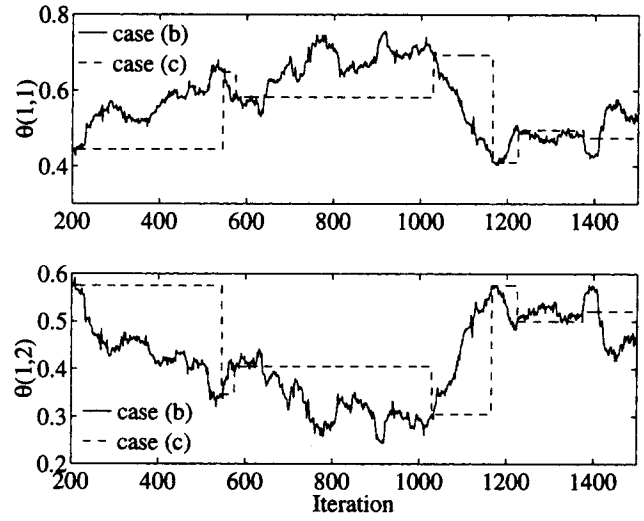


Figure 12. AR coefficients for the first temporal mode, $\theta(1,1)$ $\theta(1,2)$, as functions of the iteration number for cases (b) and (c) discussed in the text.

where the parameters were not updated at all. We used the AR parameters evaluated at $k = 1{,}050$ to simulate the remaining of the profile. For $P = 1$, the NSSPE for region 3 jumped to 0.0310, a 265% increase over case (b), justifying the adaptation.

Effect of data weighting

As discussed earlier, in order to improve profile predictions spatially uniform weighting may be applied to the data window prior to the application of the KL procedure. In this way the more recent full profile data vectors will be more influential in the estimation of the AR parameters. There are, of course, several different shapes the weighting function can take. Such a weighting can be easily accomplished by applying a matrix operator $\boldsymbol{F}$ on the data matrix $\boldsymbol{Z}$. Here, we assumed $\boldsymbol{F}$ to be diagonal, with its diagonal elements taking on values between 0 and 1. We let the most recent N data vectors have weights equal to 1 and the remaining $W - N$ data vectors be exponentially weighted (see Figure 2).

We simulated our profile with window $W = 200$ using an AR(2) model. The number of modes selected by our criterion is illustrated in Figure 13. Table 2 shows the NSSPEs for $P = 1$, and 10 for the different regions of interest. One can

Table 1. NSSPE for AR(1), AR(2), and AR(2) with Infrequent Parameter Updating (Only When Number of Modes Change) for $P = 1$, and $P = 10$

	Region 1	Transition 1	Region 2	Transition 2	Transition 3	Region 3
			$P = 1$			
AR(1)	0.0026	0.0109	0.0056	0.0061	0.0117	0.0106
AR(2)	0.0010	0.0094	0.0034	0.0040	0.0101	0.0085
AR(2)+infreq. upd	0.0011	0.0096	0.0044	0.0046	0.0102	0.0096
			$P = 10$			
AR(1)	0.0715	0.1215	0.1359	0.1227	0.0607	0.0711
AR(2)	0.0058	0.0364	0.0661	0.0641	0.0384	0.0432
AR(2)+infreq. upd	0.0071	0.0474	0.0860	0.0710	0.0432	0.0566

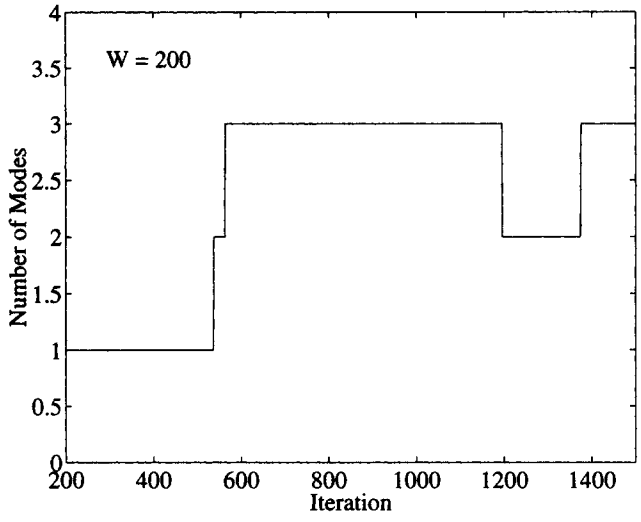


Figure 13. Number of modes selected using the criterion of scaled eigenvalues for exponentially weighted data.

see the significant improvement in the NSSPE values of the weighted case over the equally weighted case during transition regions, while the results are comparable for regions 1, 2 and 3.

Effect of the number of modes on P-step ahead predictions

Figures 14a and 14b show one profile section. Using the KL expansion adaptively with $W = 200$, equal data weighting, and modeling of the temporal modes using an AR(2) model, the profile's one- and 10-step ahead predictions are shown in Figures 14c and 14d, respectively. It is clear that the identification (or modeling) procedure has been able to effectively filter the measurement noise and capture the profile patterns for both $P = 1$ and $P = 10$. Most importantly, the number of modes selected using our criterion during this successive application of the algorithm was two or three only (see also Figure 11a), achieving significant data compression. On the other hand, if instead of using our modes selection algorithm, we fixed the number of modes throughout the entire simulation period to one or two modes only, in order to obtain even higher data compression, we would fail to take into account some significant variations as shown in Figure 15b, where the number of modes is fixed to one. Figure 15a exhibits the same result by plotting the error between the measurement noise-free profile and one-step ahead predicted profiles, where the number of modes is fixed to one or two.

The retained temporal modes for the window at $k = 760$ are illustrated in Figure 16a. We notice that as we move from the first mode to the third, the noise content increases as expected. We also notice that the peaks and troughs are not occurring at the same MD locations. This out-of-phase behavior should be compared with the original noise-free profile (Figure 14b) whose individual lanes exhibit exactly the same phenomenon (Figure 16b), and is clearly missed when the profile is modeled using one mode only (Figure 15b).

To further assess the importance of the number of modes that should be retained in the KL expansion, we performed simulations with the number of modes fixed throughout the entire period, and evaluated the NSSPE for comparison purposes. We simulated the profile with 1, 2, 3, 4, 5, and 10 modes for $P = 1$ and $P = 10$. The results are tabulated in Table 3 for each region separately. The following conclusions can be drawn.

- For all regions, when the number of modes selected is smaller than what our selection criterion indicated, the NSSPE deteriorates *significantly.*
- Depending on the region, the inclusion of more modes than what our selection criterion indicates may or may not produce a better NSSPE. In particular, for $P = 10$, even when 10 modes are included, the NSSPEs are almost the same as what we get by applying the selection criterion. For $P = 1$, the NSSPEs for regions 1 and 2 indicate that there is no advantage of increasing the number of modes. For the transition regions, however, it appears that incorporating more modes than what our criterion suggests yields a better NSSPE. This can be explained by the fact that our criterion cannot immediately identify the beginning of the transition region, thus the increase in NSSPE. Comparison of these results with the exponentially weighted case shown in Table 2, however, indicates that for $P = 1$ the results are very similar, whereas for $P = 10$, the weighted case is overall superior to the best constant mode case.
- Above a certain number of modes (which depends on the region) the inclusion of more modes produces no advantage at all as expected, since the additional modes have a very small magnitude and correspond to noise.

Overall, the results of this table indicate that using our selection criterion, we are able to effectively describe the disturbance profile, *while achieving the "best" possible compression ratio,* in the sense that the inclusion of more modes does not necessarily reduce the NSSPE appreciably.

Example 5: Disturbance profile composed of sinusoids and high frequency components

Here, we will combine the behaviors of the previous two examples. Specifically, we will add to the profile composed of sinusoids an autoregressive profile driven by white noise fil-

Table 2. NSSPE for Equally and Exponentially Weighted Data for $P = 1$, and $P = 10$ Using AR(2) Model and $W = 200$

	Region 1	Transition 1	Region 2	Transition 2	Transition 3	Region 3
			$P = 1$			
Equally weighted data	0.0010	0.0094	0.0034	0.0040	0.0101	0.0085
Exponentially weighted data	0.0011	0.0059	0.0035	0.0038	0.0068	0.0061
			$P = 10$			
Equally weighted data	0.0058	0.0364	0.0661	0.0641	0.0384	0.0432
Exponentially weighted data	0.0058	0.0336	0.0680	0.0558	0.0274	0.0301

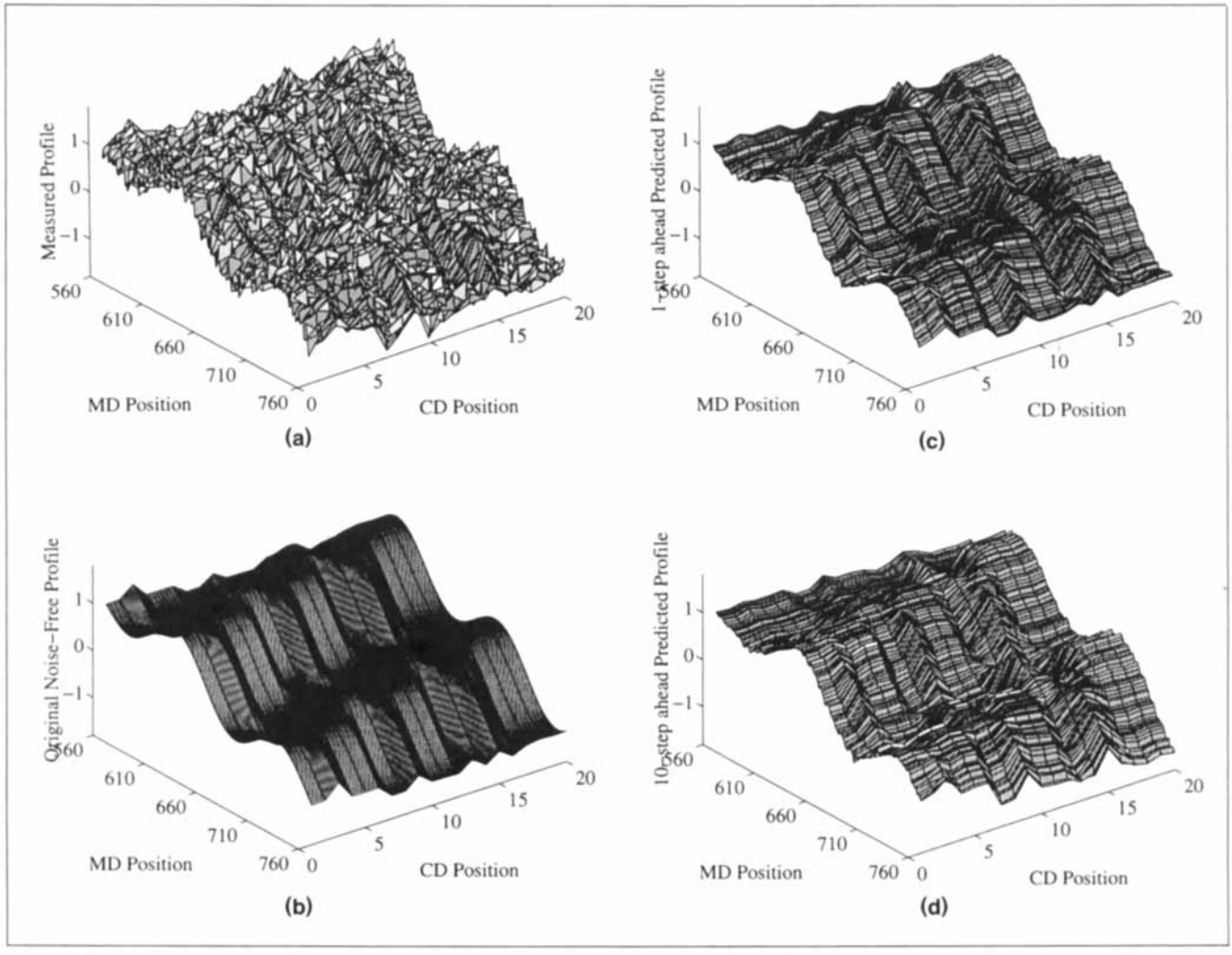


Figure 14. (a) Representative measured profile, (b) corresponding measurement noise-free profile (c), (d) 1 and 10-step ahead predicted profiles for $W = 200$, using an AR(2) model for the temporal modes.

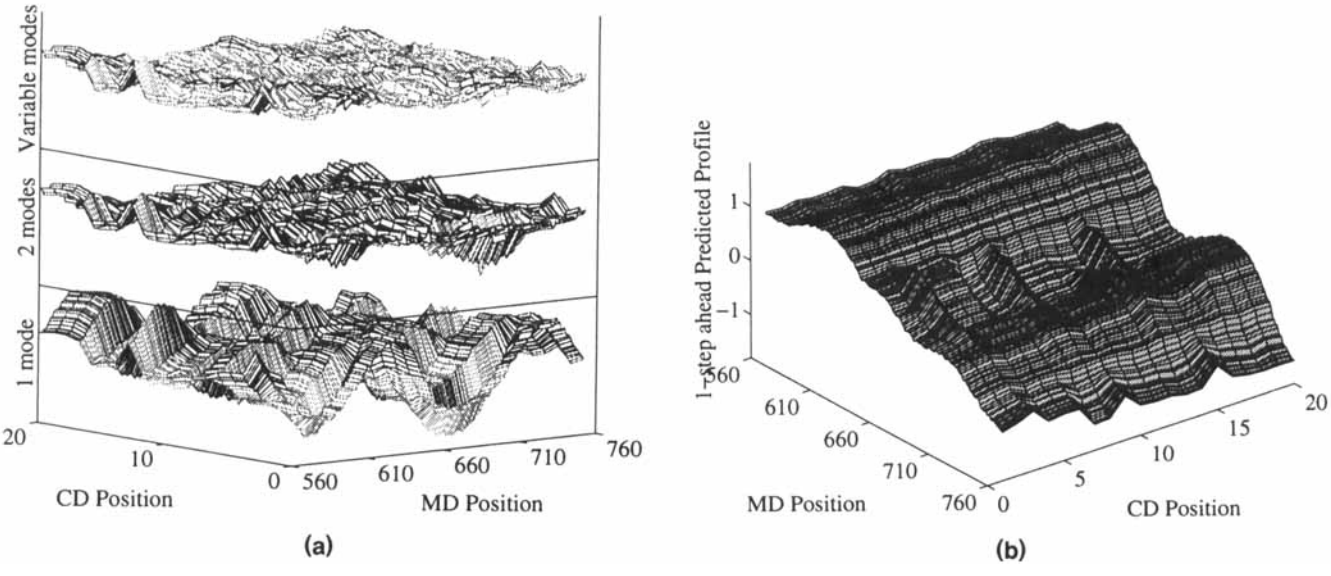
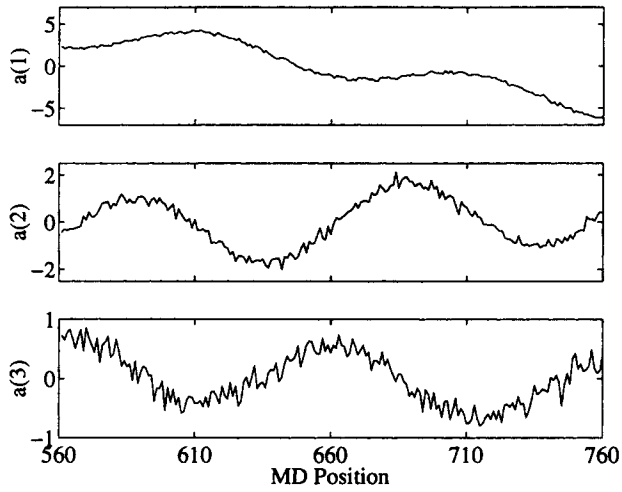
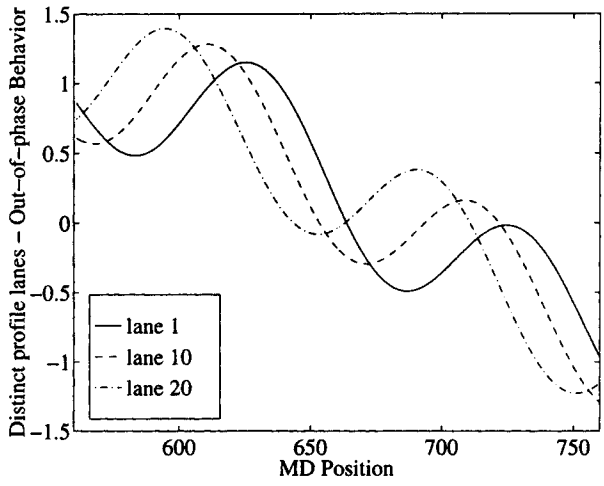


Figure 15. (a) One-step ahead prediction errors; (b) one-step ahead predicted profile fixing the number of modes to 1.



(a)



(b)

Figure 16. (a) Retained temporal modes at $k = 760$; (b) three lanes from the original noise-free profile.

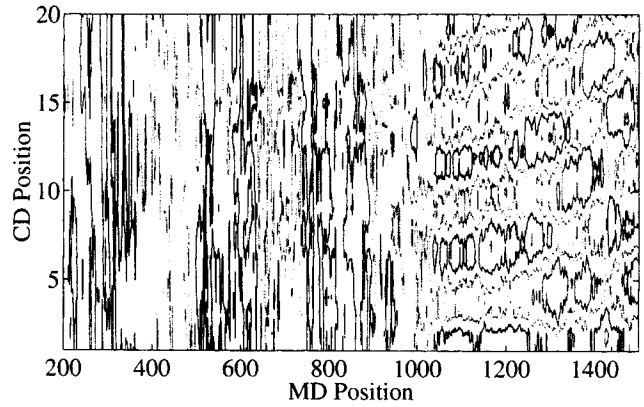


Figure 17. 2-D contour plot of the measurement noise-free profile used in the fifth example.

tered through a Toeplitz disturbance gain matrix. The addition of the autoregressive profile allows for higher frequency input variations that enter through the headbox and cannot be modeled as sine waves to be included in the overall disturbance model. The profile is expressed as

$$y(k) = y_{\sin} + f y_{\mathrm{ar}}$$
$$z(k) = y(k) + v(k) \tag{18}$$

where $y_{\sin}$ is given by Eq. 17; y_{ar} is given by Eq. 16, with G being the 9-diagonal matrix used before; f is a scalar regulating the relative magnitude of y_{ar} vs. $y_{\sin}$. Here, we chose f to be such that the mean standard deviation of $y_{\sin}$ is twice that of y_{ar}. $z(k)$ is the measured profile at time k corrupted by measurement noise $v(k)$, assumed white, with signal-to-noise ratio equal to 5.

The 2-D contour plot of the disturbance profile is shown in Figure 17. Comparison of this profile with that in Figure 10 indicates that the addition of the autoregressive term made the profile a lot more irregular than before in both the MD and CD directions. There is now no distinction between regions 1 and 2, whereas one can still distinguish the nonlinear

Table 3. Comparison of NSSPEs for Cases where Number of Modes is Kept Constant to 1, 2, 3, 4, 5, and 10, and When Using the Mode Selection Criterion with AR(2), for $P = 1$, and $P = 10$

	Region 1	Transition 1	Region 2	Transition 2	Transition 3	Region 3
			$P = 1$			
Mode selection	0.0010	0.0094	0.0034	0.0040	0.0101	0.0085
1 mode	0.0010	0.0379	0.0714	0.0694	0.0895	0.1255
2 modes	0.0010	0.0076	0.0117	0.0158	0.0841	0.0377
3 modes	0.0010	0.0051	0.0034	0.0040	0.0138	0.0080
4 modes	0.0011	0.0050	0.0034	0.0040	0.0076	0.0059
5 modes	0.0011	0.0050	0.0035	0.0040	0.0065	0.0059
10 modes	0.0011	0.0050	0.0036	0.0041	0.0066	0.0059
			$P = 10$			
Mode selection criterion	0.0058	0.0364	0.0661	0.0641	0.0384	0.0432
1 mode	0.0058	0.0548	0.1305	0.1187	0.1324	0.1707
2 modes	0.0058	0.0361	0.0735	0.0738	0.1246	0.0710
3 modes	0.0058	0.0355	0.0661	0.0641	0.0418	0.0434
4 modes	0.0058	0.0355	0.0661	0.0641	0.0376	0.0427
5 modes	0.0058	0.0355	0.0661	0.0641	0.0376	0.0427
10 modes	0.0058	0.0355	0.0661	0.0641	0.0376	0.0427

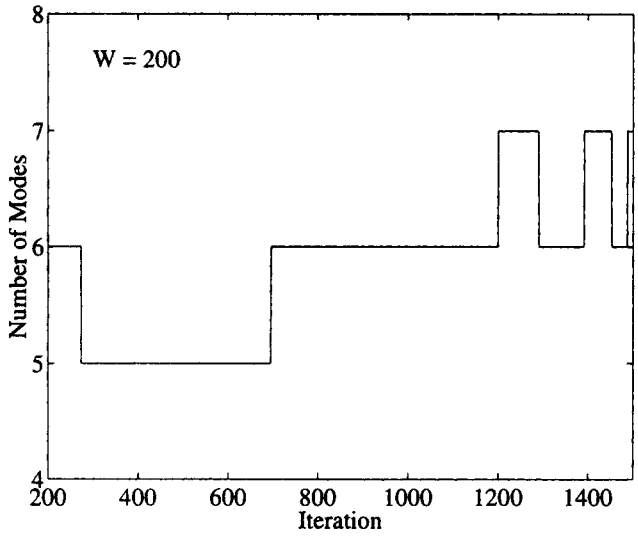


Figure 18. Number of modes selected using the criterion of scaled eigenvalues for equally weighted data for the profile illustrated in Figure 17.

wandering effect in section 3. The number of modes selected using our criterion is shown in Figure 18 for $W = 200$ using equally weighted data. If the disturbance profile were composed solely of y_{ar}, six modes would have been needed throughout. Now that it is composed of two terms, the number of modes selected is *not* the sum of the modes required by each profile separately. This occurs because of two reasons: (a) profiles y_{sin} and y_{ar} are not orthogonal to each other; (b) a disturbance pattern, considered important enough to be included in the expansion for any of the two profiles separately, may be treated as negligible in the overall profile.

Conclusions

We presented a procedure for the on-line identification of full profile disturbance models for sheet forming processes. We showed how the KL expansion can be applied on-line to reveal the significant patterns of the disturbance profile, while providing noise filtering and data compression. We illustrated how to compute P-step ahead disturbance predictions by modeling the temporal modes through low-order linear autoregressive processes. The effect of the window size used in the expansion, the order of the AR model as well as the frequency of updating its parameters, the effect of data weighting, and the significance of the number of modes retained in the expansion were illustrated through simulation examples. The advantage of this method over others presented elsewhere is that it makes no assumption about the structure of the disturbance, but rather it builds a disturbance model based on actual measurements. Under this framework, it is straightforward to transform these models into a state-space form, which facilitates the implementation of a control strategy such as model predictive control.

Acknowledgments

The authors gratefully acknowledge the financial support of the Weyerhaeuser Co.

Literature Cited

Anon., "Gauging System Gives Small Converter a Big Edge," *Converting Magazine*, 44 (1993).

Bergh, L. G., and J. F. MacGregor, "Spatial Control of Sheet and Film Forming Processes," *Can. J. Chem. Eng.*, **65**, 148 (1987).

Berkooz, B., P. Holmes, and J. L. Lumley, "The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows," *Ann. Rev. Fluid Mech.*, **25**, 539 (1993).

Cattell, R. B., "The Scree Test for the Number of Factors," *Multivariate Behavioral Res.*, **1**, 245 (1966).

Chen, C., and H. C. Chang, "Accelerated Disturbance Damping of an Unknown Distributed System by Nonlinear Feedback," *AIChE J.*, **38**, 9, 1461 (1992).

Chen, S.-C., "Full-Width Sheet Property Estimation from Scanning Measurements," *Preprints, Control Systems '92*, p. 123 (1992).

Francis, K., C. Kleinsmith, and M. Stenbak, "The Stationary Bone-Dry Weight Sensor," in *Eng. Conf. Tappi Proc.*, p. 561 (1988).

Gay, D. H., and W. H. Ray, "Application of Singular Value Methods for Identification and Model Based Control of Distributed Parameter Systems," *IFAC Symp. on Model-Based Control*, Atlanta (1988).

Karhunen, K., "Über lineare methoden in der wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fenn., Ser. A, I. Math. Phys.*, **37**, 79 (1947).

Krischer, K., R. Rico-Martinez, I. G. Kevrekidis, H. H. Rotermund, G. Ertl, and J. L. Hudson, "Model Identification of a Spatiotemporally Varying Catalytic Reaction," *AIChE J.*, **39**, 1, 89 (1993).

Loève, M. M., *Probability Theory*, 3rd ed., University Ser. in Higher Mathematics," Van Nostrand, Princeton, NJ (1963).

Ogawa, H., and E. Oja, "Projection Filter, Wiener Filter, and Karhunen-Loève Subspaces in Digital Image Restoration," *J. Math. Anal. Appl.*, **114**, 37 (1986).

Park, H. M., and D. H. Cho, "The Use of the Karhunen-Loève Decomposition for the Modeling of Distributed Parameter Systems," *Chem. Eng. Sci.*, **51**, 1, 81 (19960.

Preisendorfer, R. W., *Principal Component Analysis in Meteorology and Oceanography*, Vol. 17 of *Development in Atmospheric Science*, Elsevier Science Publishers (1988).

Rawlings, J. B., and I. Chien, "Gage Control of Sheet and Film Forming Processes," *AIChE J.*, **42**, 3, 753 (1996).

Savoji, M. H., and R. E. Burge, "On Different Methods Based on the Karhunen-Loève Expansion and Used in Image Analysis," *Comp. Vision, Image and Signal Proc.*, **29**, 2, 259 (1985).

Sirovich, L., and R. Everson, "Management and Analysis of Large Scientific Datasets," *Int. J. Supercomputer Appl.*, **6**, 1, 50 (1992).

Taylor, B. F., "Optimum Separation of Machine-Direction and Cross-Direction Product Variations. The Scanning Measurement Challenge," *Tappi J.*, 87 (1991).

Tyler, M. L., and M. Morari, "Estimation of Cross Directional Properties: Scanning versus Stationary Sensors," *AIChE J.*, **41**, 846 (1995).

Wang, X. G., G. A. Dumont, and M. S. Davies, "Basis Weight Estimation from Array Measurements," *Proc. IFAC World Congress*, Vol. 6, Sydney, Australia, p. 55 (1993a).

Wang, X. G., G. A. Dumont, and M. S. Davies, "Estimation in Paper Machine Control," *IEEE Control Systems Magazine*, **13**, 34 (1993b).

Wilhelm, R. G., "On the Controllability of Cross-Direction Variations of Sheet Properties," *Eng. Conf. Proc. of Techn. Assoc. of Pulp and Paper Ind.*, p. 621 (1984).

Wilhelm, R. G., and M. Fjeld, "Control Algorithms for Cross Directional Control: The State of the Art," *Proc. of Int. IFAC/IMECO Conf. on Instrument. in Paper, Rubber, Plastics and Polymer. Ind.*, Antwerp, Belgium, p. 163 (1983).

Manuscript received June 21, 1996, and revision received Oct. 9, 1996.