Comparison of Linear Distributed-Parameter Filters to Lumped Approximants

Optimal distributed-parameter filters are commonly implemented using approximating lumped Kalman filtering theory. The effect of such an approximation is investigated. A theoretical development shows that there is a loss in the spatial noise correlation for the lumped approximants. Two numerical examples of engineering significance illustrate that one result of this loss is slower filter convergence for the lumped approximants relative to the full distributed-parameter filters.

D. J. COOPER
W. F. RAMIREZ
and D. E. CLOUGH

Department of Chemical Engineering
University of Colorado
Boulder, CO 80309

SCOPE

Simplifying approximations are often made when designing a distributed-parameter filter for real-time implementation. Early lumping is one simplifying approximation that is commonly employed. Lumping (spatial discretization) is a process that must occur at some point during filter development so that the equations can be programmed for digital computation. Lumping can occur at several points in the filter development. The earlier that lumping occurs, the more simplifying the effect. At one extreme, early lumping involves discretizing the state model partial differential equations and approximating them as a system of ordinary differential equations. Application of optimal filtering theory to the system of ordinary differential equations results in a lumped Kalman filter that approximates the distributed-parameter filter. This is in contrast to late lumping, at the other extreme, where the full distributed-parameter filter partial differential equations are derived before they are lumped for numerical computation.

Advantages of early lumping include being able to use simpler theoretical and numerical methods for filter design and implementation. Also, the early lumped filtering algorithm requires less computation for filter equation solution, so less computer resources are used when on-line. The early lumping approximation does have negative consequences, however.

In this work, we present a side-by-side comparison of early lumping vs. late lumping as applied to the least-squares filtering of linear distributed-parameter systems. This investigation includes a general theoretical formalization that is applicable to such lumping methods as finite differencing, eigenfunction expansion, orthogonal collocation, and Galerkin's method. The results of the theoretical investigation are illustrated by means of two numerical examples. The first example considers the optimal estimation of dynamic temperature distributions in a heated bar. The second example considers the optimal estimation of dynamic particle-size distributions (PSD) in a fluidized bed.

CONCLUSIONS AND SIGNIFICANCE

The early lumping approximation results in a loss in the spatial noise correlation that is associated with distributed-parameter systems. This spatial noise correlation is preserved in the late-lumped distributed filter.

In a spatially correlated system, an occurrence at one location will directly affect neighboring locations. The early lumping approximation causes a loss of spatial correlation that is reflected in the theoretical development in that the spatially discrete covariance matrices, $Q(t)$, $R(t)$, and P_o, can be partitioned into diagonal forms. Each submatrix along the diagonal of these partitioned matrices contains covariance information that is particular to a specific spatial location. It is the off-diagonal elements that provide the direct mathematical communication between neighboring locations in the spatial domain. The lack of these elements in the early-lumped approximation filter means that a poor state model prediction, a bad measurement, or a poor initial estimate at one location will not be directly accounted for in the optimal estimate computations of neighboring locations. The late-lumped distributed filter retains the covariance matrices as full matrices, providing the mathematical communication necessary for spatial correlation.

The loss in spatial correlation from early lumping is manifested in the numerical examples by a slower convergence rate relative to the late-lumped distributed filter when starting from a poor *a priori* initial estimate. The loss is also reflected in that the early-lumped filter does not track system dynamics as well as the late-lumped filter when burdened with a poor dynamic model. The early-lumped approximation does result in a filter form that requires less computation to solve, however, and this can be an important consideration in the real-time implementation of linear distributed-parameter filters.

D. J. Cooper is now with the Chemical Engineering Dept. at the University of Connecticut, Storrs, CT 06268.

INTRODUCTION

Real-time implementation of optimal distributed-parameter (DP) filters can be extremely challenging. One reason is that considerable expertise is required to apply optimal filtering theory to the complex state models that often arise in the mathematical description of DP systems. Sophisticated numerical techniques are then required to obtain rapid and accurate solutions to the resulting filter algorithms. Another reason is a mechanical consideration: the filtering algorithm computational burden cannot exceed the computer speed or memory size. Computer system limitations often require a careful and clever filter design. A final reason is a practical consideration: not all applications require that the full complexity of the distributed system be included in the filter for it to perform satisfactorily. For some or all of these reasons, simplifying approximations are made by many investigators when attempting a DP filter implementation.

Lumping (spatial discretization) is a process that must occur at some point during the filter derivation so that the filter equations can be programmed on a digital computer. Lumping can occur at several points in the filter development. The earlier it occurs, the more simplifying the effect. Early lumping , at one extreme, is a commonly employed simplifying approximation. Early lumping involves discretizing the state model partial differential equations and approximating them as a system of ordinary differential equations. Application of optimal estimation theory to the system of ordinary differential equations results in a lumped Kalman filter that serves as an approximation to the linear DP filter. This is in contrast to late lumping, at the other extreme, where the full DP filter partial differential equations are derived before they are lumped for numerical computation.

One advantage of early lumping is that the application of lumped Kalman filtering theory is relatively straightforward. Another advantage is that a system of lumped filter ordinary differential equations can be solved more readily in real time. A disadvantage of early lumping, however, is a loss in the spatial noise correlation associated with distributed systems. This loss can negatively affect filter performance.

In this work, we present a side-by-side comparison of early lumping with late lumping when applied to the least-squares filtering of linear DP systems. This investigation includes a theoretical development of the two filters that is general to such lumping methods as finite differencing, eigenfunction expansion, orthogonal collocation, and Galerkin's method. The results of the theoretical investigation are illustrated by two numerical examples. The first example considers the optimal estimation of dynamic temperature distributions in a heated bar. The second example considers the optimal estimation of dynamic particle-size distributions (PSD) in a fluidized bed.

Background

A number of papers have discussed lumping as applied to DP filtering theory and have summarized some of the more popular lumping methods (Van den Bosch, 1978; Ray, 1981; Seidman, 1982). These and other investigators (e.g., Koda and Seinfeld, 1978; Lausterer et al., 1978) are in general agreement that by delaying the lumping process for as long as possible, the distributed nature of the system is best preserved. There has been little published, however, that formalizes the difference between early vs. late lumping as applied to the filtering of linear DP systems, or that investigates the character and extent of such approximations on filter performance.

The lumping methods that are found most often in the literature for the implementation of linear DP filters include finite differencing, eigenfunction expansion, orthogonal collocation, and Galerkin's method. Finite differencing has been used for many years, and is probably the most straightforward of the above methods to apply. It is not a very efficient technique to use, however, and this can be important if computer resources are limited. Sage (1968) presents a good general development of finite differencing as a tool for both early lumping and late lumping as applied to linear DP filtering theory. There are many examples found in the literature of both early- and late-lumped filter implementation using finite differencing, including the work of Collins and Khatri (1969), Desalu et al. (1974), Koda and Seinfeld (1978), Koda (1979), Kitamura and Taniguchi (1981), Ramirez (1982), and Cooper and Clough (1985).

More efficient lumping methods, such as the weighted residual methods, are becoming increasingly popular. Lumping via a weighted residual method involves expanding the solution of the partial differential equation as a set of known basis functions, $\Phi_i(y)$, i.e.,

$$x(y,t) \approx \sum_{i=1}^{N} a_i(t)\Phi_i(y) \tag{1}$$

The problem is then to determine the coefficients $a_i(t)$ that yield the best approximation to the solution based on some goodness-of-fit criterion. Also required is to determine the number of terms, N, that must be retained in the series approximation to obtain a satisfactory solution. It is important to note that increasing the number of terms, N, generally increases solution accuracy and always increases solution computation load.

There are three methods of weighted residuals that appear often in the literature for the implementation of linear DP filters. These methods are eigenfunction expansion, orthogonal collocation, and Galerkin's method. The difference between these methods lies in the choice of the basis function used for solution approximation. Finlayson (1972) and Villadsen and Michelsen (1978) have written textbooks on weighted residual methods for differential equation solution. Ray (1981) has included a section in his textbook that presents a good overview of weighted residual methods as applied to DP filtering theory. There are many examples in the literature of both early- and late-lumped implementation of linear DP filters using a weighted residual method. Some examples of filter implementation using eigenfunction expansion include the works of Ajinkya (1975), Ajinkya et al. (1975), Lausterer et al. (1978), Lausterer and Ray (1979), and Omatu and Seinfeld (1981). Some examples of filter implementation using orthogonal collocation include the works of Seinfeld and Chen (1971), Tanner (1972), Viskanta et al. (1975), Van den Bosch (1978), Clement et al. (1980), and Sorensen et al. (1980). Some examples of filter implementation using Galerkin's method include the works of Prabhu and McCausland (1972), Polis et al. (1973), Orner et al. (1975), Ito et al. (1978), and Yoshimura and Campo (1982).

In this work, we formalize the differences between early lumping and late lumping. The theoretical development will not be general to all lumping methods; rather, it will be specific to approximations methods that reduce the state equation to the form:

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t) \tag{2}$$

Where $\mathbf{x}(t)$ is a vector of spatially discrete states $x(y,t)$, and $\mathbf{A}(t)$ is a lumped spatial-derivative approximation operator. Finite differencing, eigenfunction expansion, orthogonal collocation, and Galerkin's method all fit this requirement.

We begin the investigation by outlining the derivation of a general optimal linear DP filter and the lumping of the resulting filter partial differential equations (late lumping). Next, the general distributed state model is lumped, and the derivation of a lumped Kalman filter is outlined (early lumping). The matrix

forms of the final filtering algorithms for both cases are then compared. The side-by-side comparison of the theoretical results gives the reader an intuitive appreciation for the differences between the early-lumped approximation filter and the late-lumped distributed filter. We conclude with numerical examples that illustrate the difference in filter performance for the two cases, along with a discussion of the work.

THEORETICAL DEVELOPMENT

Late Lumping

Let Ω be a bounded open set in Euclidean N-space with any point in Ω denoted as $y = (y_1, \ldots, y_N)$. Let t denote time on the fixed interval $[t_0,t_f]$. We consider the general linear distributed parameter (DP) system described by:

$$\frac{\partial x(y,t)}{\partial t} = \mathcal{A}_y x(y,t) + u(y,t) \tag{3a}$$

$$z(y,t) = M(y,t)x(y,t) + v(y,t) \tag{3b}$$

$$x(y,t_0) = x_0(y) + u(y,t_0) \tag{3c}$$

Where the n-dimensional system state is defined as:

$$x(y,t) = \begin{bmatrix} x_1(y,t) \\ \cdot \\ \cdot \\ \cdot \\ x_n(y,t) \end{bmatrix} \tag{4}$$

and $\mathcal{A}_y$ is a well-posed linear spatial-differential operator. $z(y,t)$ is the measured output of the system state and $M(y,t)$ is a linear distributed modeling operator that describes the relationship between the system state and the measured output. $x(y,t_0)$is the initial state of the system and is unknown, and $x_0(y)$ is an *a priori* estimate of the initial state.

The variables $u(y,t)$ and $v(y,t)$ are Gaussian-distributed, zero-mean noise vectors representing the uncertainties in the state model and the measured output model, respectively. Both are assumed to be white in time and to have covariance matrices of the form

$$E[u(y,t)u^T(y',\tau)] = Q(y,y',t)\delta(t-\tau) \tag{5a}$$

$$E[v(y,t)v^T(y',\tau)] = R(y,y',t)\delta(t-\tau) \tag{5b}$$

and with a cross covariance of zero (uncorrelated). The *a priori* initial state estimate $x_0(y)$ has a covariance matrix of the form:

$$E[x_0(y)x_0^T(y')] = P_0(y,y') \tag{6}$$

The optimal filtered estimate is obtained by minimizing the quadratic least-squares error criterion:

$$\begin{aligned} J = \frac{1}{2}\int_\Omega\int_\Omega [x(y,t_0)-x_0(y)]^T P_0^{-1}(y,y')\,[x(y',t_0)-x_0(y')]dy'dy \\ + \frac{1}{2}\int_{t_0}^{t_f}\int_\Omega\int_\Omega [z(y,t)-M(y,t)x(y,t)]^T \\ R^{-1}(y,y',t)[z(y',t)-M(y',t)x(y',t)]dy'dydt \\ + \frac{1}{2}\int_{t_0}^{t_f}\int_\Omega\int_\Omega [u^T(y,t)Q^{-1}(y,y',t)u(y',t)]dy'dydt \end{aligned} \tag{7}$$

The problem, to minimize J subject to the conditions expressed in Eq. 3a–c, is treated via the calculus of variations by introducing a Lagrange multiplier function and considering the augmented functional:

$$J_{aug} = J + \int_{t_0}^{t_f}\int_\Omega \lambda(y,t)\left[\frac{\partial x(y,t)}{\partial t} - \mathcal{A}_y x(y,t)\right]dydt \tag{8}$$

As detailed in the literature (Sage, 1968; Meditch, 1971; Tzafestas, 1978), if a solution of the form:

$$x(y,t) = \int_\Omega P(y,y',t)\lambda(y',t)dy' + \chi(y,t) \tag{9}$$

is assumed for the relevant Euler-Lagrange equations and associated transversality conditions, the following pair of simultaneous linear DP filter equations results:

$$\frac{\partial \chi(y,t)}{\partial t} = \mathcal{A}_y\chi(y,t) + \int_\Omega\int_\Omega P(y,y',t)R^{-1}(y',y'',t)[z(y'',t) - M(y'',t)\chi(y'',t)]dy''dy' \tag{10a}$$

$$\begin{aligned} \frac{\partial P(y,y',t)}{\partial t} = \mathcal{A}_y P(y,y',t) + P(y,y',t)\mathcal{A}_{y'}^T + Q(y,y',t) \\ - \int_\Omega\int_\Omega P(y,y''',t)M^T(y''',t)R^{-1}(y''',y'',t) \\ M(y'',t)P(y'',y',t)dy'''dy'' \end{aligned} \tag{10b}$$

where $\chi(y,t)$ is the optimal state estimate and $P(y,y',t) = P(y',y,t)$.

The linear DP filter equations are lumped by discretizing the spatially continuous variables at N spatial locations, and approximating the spatial-differential state operator at each $n = 1,\ldots,N$ as:

$$\mathcal{A}_y\chi(y_n,t) = \sum_{i=1}^{N} A(y_n,y_i)\chi(y_i,t) \tag{11}$$

If N process measurements are made at the N spatial locations, the linear DP filter is described at each n by the ordinary differential equations:

$$\dot{\chi}(y_n,t) = \sum_{i=1}^{N} A(y_n,y_i)\chi(y_i,t) + \sum_{i=1}^{N}\sum_{j=1}^{N} P(y_n,y_i,t)R^{-1}(y_i,y_j,t)[z(y_j,t) - M(y_j,t)\chi(y_j,t)] \tag{12a}$$

$$\begin{aligned} \dot{P}(y_n,y_k,t) = \sum_{i=1}^{N} A(y_n,y_i)P(y_i,y_k,t) + \\ \sum_{i=1}^{N} A(y_k,y_i)P(y_n,y_i,t) + Q(y_n,y_k,t) \\ - \sum_{i=1}^{N}\sum_{j=1}^{N} P(y_n,y_i,t)M(y_i,t)R^{-1}(y_i,y_j,t)M(y_j,t)P(y_j,y_k,t) \end{aligned} \tag{12b}$$

with the initial conditions:

$$\chi(y_n,t_0) = x_0(y_n) \tag{12c}$$

$$P(y_n,y_k,t_0) = P_0(y_n,y_k) \tag{12d}$$

where $n,k = 1,\ldots,N$.

These equations can be expressed in matrix notation by defining the discretized optimal estimate vector:

$$\chi(t) = \begin{bmatrix} \chi(y_1,t) \\ \cdot \\ \cdot \\ \cdot \\ \chi(y_N,t) \end{bmatrix} \tag{13}$$

Similarly appropriate definitions for the other discretized variables result in the late-lumped linear DP filter matrix representation:

$$\dot{\chi}(t) = A\chi(t) + P(t)R^{-1}(t)[z(t) - M(t)\chi(t)] \quad (14a)$$

$$\dot{P}(t) = AP(t) + P(t)A^T + Q(t) - P(t)M^T(t)R^{-1}(t)M(t)P(t) \quad (14b)$$

with,

$$\chi(t_0) = x_0 \quad (14c)$$

$$P(t_0) = P_0 \quad (14d)$$

Early Lumping

When the system state is lumped using the approximation defined in Eq. 11, the linear DP system described by Eqs. 3a–c is approximated at each n with the ordinary differential equations:

$$\dot{x}(y_n,t) = \sum_{i=1}^{N} A(y_n,y_i)x(y_i,t) + u(y_n,t) \quad (15a)$$

$$z(y_n,t) = M(y_n,t)x(y_n,t) + v(y_n,t) \quad (15b)$$

$$x(y_n,t_0) = x_0(y_n) + u(y_n,t_0) \quad (15c)$$

where $n = 1,\ldots,N$. The lumped least-squares error criterion for this early lumping is:

$$J = \sum_{i=1}^{N}\frac{1}{2}[x(y_i,t_0) - x_0(y_i)]^T P_0^{-1}(y_i,y_i)[x(y_i,t_0) - x_0(y_i)] + \sum_{i=1}^{N}\frac{1}{2}\int_{t_0}^{t_f}[z(y_i,t) - M(y_i,t)x(y_i,t)]^T R^{-1}(y_i,y_i,t)[z(y_i,t) - M(y_i,t)x(y_i,t)]dt + \sum_{i=1}^{N}\frac{1}{2}\int_{t_0}^{t_f}[u^T(y_i,t)Q^{-1}(y_i,y_i,t)u(y_i,t)]dt \quad (16)$$

Where $P_0^{-1}(y_i,y_i)$, $R^{-1}(y_i,y_i,t)$, and $Q^{-1}(y_i,y_i,t)$ are nxn matrices that contain covariance information for the n independent variables at spatial location y_i. The discrete covariance matrix, $\mathbf{P}_0^{-1}$, can be partitioned into the NxN diagonal matrix form:

$$\mathbf{P}_{0,\text{diag}}^{-1} = \begin{bmatrix} P_0^{-1}(y_1,y_1) & 0 & \cdots & 0 \\ 0 & P_0^{-1}(y_2,y_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_0^{-1}(y_N,y_N) \end{bmatrix} \quad (17)$$

With similar definitions for the other discrete covariance matrices, Eq. 16 can be expressed in matrix notation as:

$$J = \frac{1}{2}[x(t_0) - x_0]^T\mathbf{P}_{0,\text{diag}}^{-1}[x(t_0) - x_0] + \frac{1}{2}\int_{t_0}^{t_f}[z(t) - M(t)x(t)]^T\mathbf{R}_{\text{diag}}^{-1}(t)[z(t) - M(t)x(t)]dt + \frac{1}{2}\int_{t_0}^{t_f}[u^T(t)\mathbf{Q}_{\text{diag}}^{-1}(t)u(t)]dt \quad (18)$$

Minimization of J via the calculus of variations yields the classical lumped Kalman filter, expressed in matrix form as:

$$\dot{\chi}(t) = A\chi(t) + P(t)R_{\text{diag}}^{-1}(t)[z(t) - M(t)\chi(t)] \quad (19a)$$

$$\dot{P}(t) = AP(t) + P(t)A^T + Q_{\text{diag}}(t) - P(t)M^T(t)R_{\text{diag}}^{-1}(t)M(t)P(t) \quad (19b)$$

with

$$\chi(t_0) = x_0 \quad (19c)$$

$$P(t_0) = P_{0,\text{diag}} \quad (19d)$$

Comparison of Matrix Forms

A comparison of Eqs. 14 and 19, the matrix forms for the late-lumped distributed filter and the early-lumped approximation filter, respectively, allows for an appreciation of the information lost by early lumping. The filter algorithms have identical form, yet the early-lumped approximation filter has spatially discrete model and measurement noise covariance matrices that are specified as diagonal; i.e., $\mathbf{Q}(t)$ and $\mathbf{R}(t)$ are reduced to $\mathbf{Q}_{\text{diag}}(t)$ and $\mathbf{R}_{\text{diag}}(t)$ respectively. Also, the initial condition on the spatially discrete state error covariance matrix is specified as diagonal, i.e., $\mathbf{P}_0$ is reduced to $\mathbf{P}_{0,\text{diag}}$.

We can appreciate intuitively that when considering a distributed system, a poor state model prediction, a bad measurement, or a poor initial estimate at one location will affect the optimal state estimate of neighboring locations. The early-lumped approximation does not directly account for this phenomenon because it lacks the off-diagonal terms of the covariance matrices that provide the mathematical communication between neighboring spatial locations. This loss in communication results in a loss in the spatial noise correlation associated with distributed systems. A loss in spatial correlation means that the early-lumped approximation filter, working with less information, cannot perform as well as the late-lumped distributed filter in many cases. This is borne out in the examples that follow.

This theoretical result has implications for current work. Many investigators have applied late lumping to the full DP filter partial differential equations with the intent of preserving the distributed nature of the system for as long as possible. Upon implementation, however, any who then define the covariance matrices $\mathbf{Q}(t)$, $\mathbf{R}(t)$, and $\mathbf{P}_0$ with diagonal forms have specified numerical conditions that reduce their late-lumped distributed filter to the form of an early-lumped approximation filter.

NUMERICAL EXAMPLES

Temperature Distribution in a Bar

Lumped via Orthogonal Collocation. Consider an insulated bar with an initial temperature distribution $\phi(y,\tau_0) = 1.0$ and with ends that are maintained such that the boundary conditions are $\phi(0,\tau) = \phi(1,\tau) = 0.0$. With suitable assumptions, the dimensionless linear distributed system can be described as:

$$\frac{\partial\phi(y,\tau)}{\partial\tau} = \frac{\partial^2\phi(y,\tau)}{\partial y^2} + u(y,\tau) \quad (20a)$$

$$z(y,\tau) = M(y,\tau)\phi(y,\tau) + v(y,\tau) \quad (20b)$$

$$\phi(y,\tau_0) = \phi_0(y) + u(y,\tau_0) \quad (20c)$$

Orthogonal collocation is used as the lumping method for both early and late lumping in this example. The second-spatial derivative is approximated as:

$$\frac{\partial^2\phi(y_n,\tau)}{\partial y^2} = \sum_{i=1}^{N} B(y_n,y_i)\phi(y_i,\tau) \quad (21)$$

where $\mathbf{B}$, the matrix representation for $B(y_n,y_i)$, is the second-derivative orthogonal collocation operator using the shifted Legen-

dre polynomial series as the basis function (Finlayson, 1972; Villadsen and Michelsen, 1978). The first seven terms of the series accurately describe the solution to the partial differential equations. Therefore, there are only nine (seven internal collocation points plus two end points) discrete computation nodes needed to describe the solution for this example, i.e., $N = 9$ and $\mathbf{B}$ is a 9×9 symmetric matrix.

The discretized measurement vector, $\mathbf{z}(\tau)$, contains an individual temperature measurement for each of the nine spatially discrete collocation points. For simplicity, we assume that the measurements are made in the same units as $\phi(\tau)$ at each time step. Thus, the discretized modeling matrix, $\mathbf{M}$, is an identity matrix. We simulate the measurements by adding random noise to a model of the true system state.

It is most physically realistic to assume that random noise in a measurement signal at one location does not cause noise in neighboring measurement signals; that is, to assume that there is no spatial correlation among the discrete measurements. Therefore, the discretized covariance matrix, $\mathbf{R}(\tau)$, is diagonal in the late-lumped filter as well as the early-lumped filter for this example. As derived in the Theoretical Development section, however, we do consider that there is spatial correlation associated with the state model and the *a priori* state estimate and assume it can be described as an exponential spatial decay. The above assumptions lead to late-lumped covariance matrices for this example of the form:

$$Q(y_n, y_k, \tau) = d_Q e^{-|y_n - y_k|} \tag{22a}$$

$$R_{\text{diag}}(y_n, y_n, \tau) = d_R \tag{22b}$$

$$P_{0,\text{diag}}(y_n, y_n)\ e^{-|y_n - y_k|} \tag{22c}$$

and early-lumped matrices of the form:

$$Q_{\text{diag}}(y_n, y_n, \tau) = d_Q \tag{23a}$$

$$R_{\text{diag}}(y_n, y_n, \tau) = d_R \tag{23b}$$

$$P_{0,\text{diag}}(y_n, y_n) = d_P \tag{23c}$$

where $n,k = 1, \ldots, 9$ and all off-diagonal elements of the diagonal matrices are zero. Because temperature is the only independent variable in this example, d_Q, d_R, and d_P are scalars, not partitioned submatrices. In theory, these variables represent the variance from the true value due to the Gaussian-distributed, zero-mean noise associated with the state model, the measured output, and the *a priori* state estimate, respectively. In practice, these variables are assigned values based on the investigators' confidence in the relative accuracy of these terms.

Numerous off-line studies were made to compare the performance of the early-lumped approximation filter to the late-lumped distributed filter. The studies involved comparing convergence rate and tracking ability for different values for the variables d_Q, d_R, and d_P. For many of the cases, the two filters yielded optimal estimates that were within a few percent of each other. This was especially true for cases where d_Q was assigned very small values relative to d_R, indicating an accurate state model relative to the measurements. In no case did the early-lumped approximation filter perform better than the late-lumped filter.

One study in which the late-lumped filter performed significantly better than the early-lumped filter relates to the rate of convergence of the optimal estimate to the true state when starting from a poor *a priori* estimate ϕ_0. As stated previously, the true state of the bar is initially at a temperature $\phi(y,0) = 1.0$. The fact that the endpoints are maintained at $\phi(0,\tau) = \phi(1,\tau) = 0.0$ causes the temperature distribution in the bar to fall in a parabolic shape toward $\phi(y,\infty) = 0.0$. In this particular study, however, we assigned a value of $\phi_0 = 0.50$ as the *a priori* state estimates for both filters. To reflect this very poor initial estimate, we assigned the initial state error covariance matrix the relatively large value of $d_P = 0.25$. Both the state model and noisy measurements were considered to be equally accurate and assigned values of $d_Q = d_R = 0.01$ reflect this.

Figure 1 illustrates the behavior of the early-lumped and late-lumped optimal estimates as they rise to converge with the true state, which is itself experiencing dynamics as it falls toward zero. This figure illustrates the system for dimensionless times $\tau = 0.01$ and $\tau = 0.03$, respectively. Although the optimal estimates and the true state are shown as spatially continuous in these figures, note that these are lumped computations made at only nine spatial locations. The nine noisy measurements used for these computations, located at the nine collocation points, are depicted on the figures as o's. Orthogonal collocation interpolation, as described in Finlayson (1972), was used to compute the continuous true state from the discrete values.

As illustrated in these figures, the convergence rate of the late-lumped distributed filter is superior to that of the early-lumped approximation filter as they converge on the true state. Figure 2 shows the history of the noisy measurement and the optimal estimates at the mid-point of the bar, $y = 0.5$. This figure further illustrates that the late-lumped optimal estimate does a superior

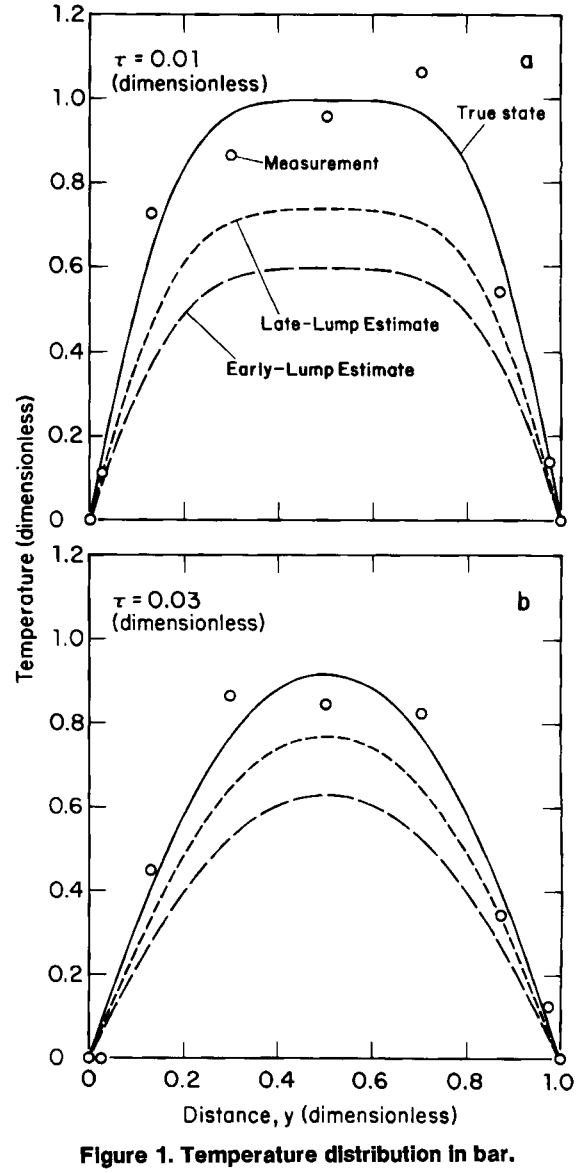

Figure 1. Temperature distribution in bar.

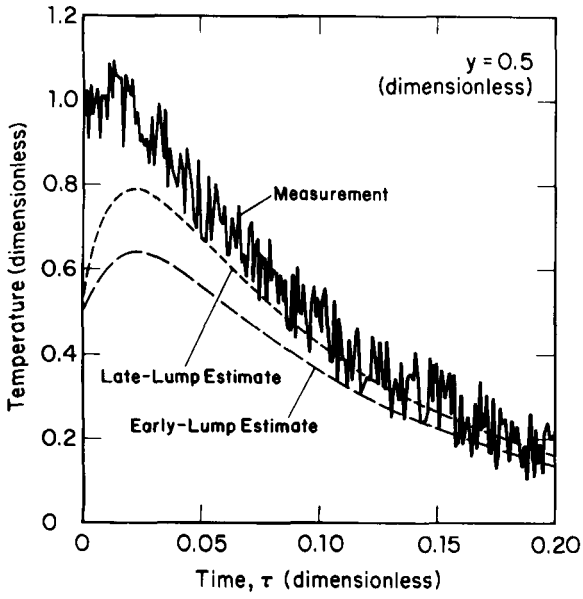

Figure 2. Temperature history of bar at point $y = 0.5$.

job in converging from the poor initial condition toward the noisy measurements, which are zero-mean distributed around the true state.

PSD in a Fluidized Bed

Lumped via Finite Difference. PSD (particle-size distribution) is a parameter of fundamental importance to the performance and operation of fluidized beds. In this example, we consider the optimal estimation of PSD as it changes with time in a fluidized bed. In a solids processing fluidized bed (Levenspiel et al., 1968; Weimer and Clough, 1980), a diagram of which is shown in Figure 3, fresh particles of known PSD, $p_o(r)$, are fed at a constant rate to a fluidized bed with a PSD, $p_b(r,\tau)$. They are discharged either by an overflow pipe or by entrainment via the fluidization gas. Considering the case where no reaction is occurring, and under the assumption of constant bed density and volume, the following general relation can be derived for spherical particles of radius r for $0 < r \leq R_{max}$:

$$w\frac{\partial p_b(r,t)}{\partial t} = F_0 p_0(r) - F_1(t)p_1(r,t) + w\frac{\partial}{\partial r}\left\{ p_b(r,t)\left(\frac{\partial r}{\partial t}\right)\right\} \quad (24)$$

with the requirement that at all times, the probability density function $p_b(r,t)$ be normalized such that:

$$\int_0^{R_{max}} p_b(r,t)dr = 1.0 \quad (25)$$

The form of the dynamic model can be recast by assuming an attrition mechanism in which tiny fragments are worn or abraded off the particles, are immediately carried overhead (elutriated), and are not considered as part of the solids population, i.e.:

$$\left(\frac{\partial r}{\partial t}\right) = kr \quad (26)$$

With a further assumption that the fragments abraded off the spherical particles are the only particles elutriated, and that they immediately leave the system and are not considered as part of the

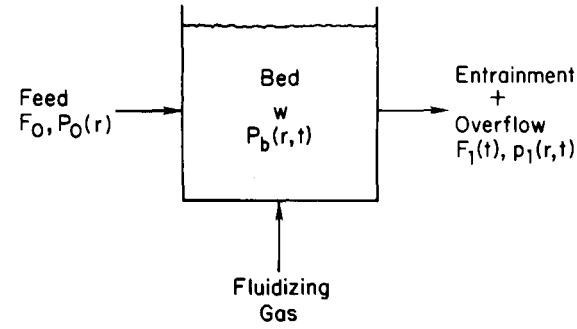

Figure 3. Fluidized bed, schematic diagram.

particle population (Levenspiel et al., 1968), Eq. 26 leads to the mass balance:

$$F_1 = F_0 - 3kw \quad (27)$$

We also specify that the fluidization velocity is high enough (approximately three times minimum fluidization or greater) such that:

$$p_1(r,t) \approx p_b(r,t) \quad (28)$$

Substitution of Eqs. 26, 27, and 28 into Eq. 24 leads to the general form of the dimensionless linear fluidized bed dynamic model:

$$\frac{\partial p_b(r,\tau)}{\partial \tau} = \alpha(r)\frac{\partial p_b(r,\tau)}{\partial r} + \beta(r,\tau)p_b(r,\tau) + \gamma(r) \quad (29)$$

where

$$\alpha(r) = \frac{wkr}{F_0} \quad (30)$$

$$\beta(r,\tau) = \left[\frac{4kw}{F_0}\right] - 1 \quad (31)$$

$$\gamma(r) = p_0(r) \quad (32)$$

and

$$\tau = \frac{tF_0}{w} \quad (33)$$

Finite differencing is used as the lumping method for both early and late lumping in this example. The fluidized bed PSD, $p_b(r,\tau)$, is approximated as a vector of 50 discrete states, each at a specific radius, i.e.

$$\mathbf{p_b}(\tau) = \begin{bmatrix} p_b(r_1,\tau) \\ p_b(r_2,\tau) \\ \cdot \\ \cdot \\ \cdot \\ p_b(r_{50},\tau) \end{bmatrix} \quad (34)$$

and the spatial derivative is approximated via classical finite difference as:

$$\frac{\partial p_b(r,\tau)}{\partial r} = \mathbf{D}\mathbf{p_b}(\tau) \quad (35)$$

where **D**, a 50×50 banded matrix, is a central-difference, first-spatial derivative approximation operator.

We consider that a continuous PSD measurement is such that the discretized measurement vector, $\mathbf{z}(\tau)$, contains a value for

each of the 50 states (radius sizes) at every time step. We again consider that the measured values are in the same units as the states; therefore, M is a 50 × 50 identity matrix.

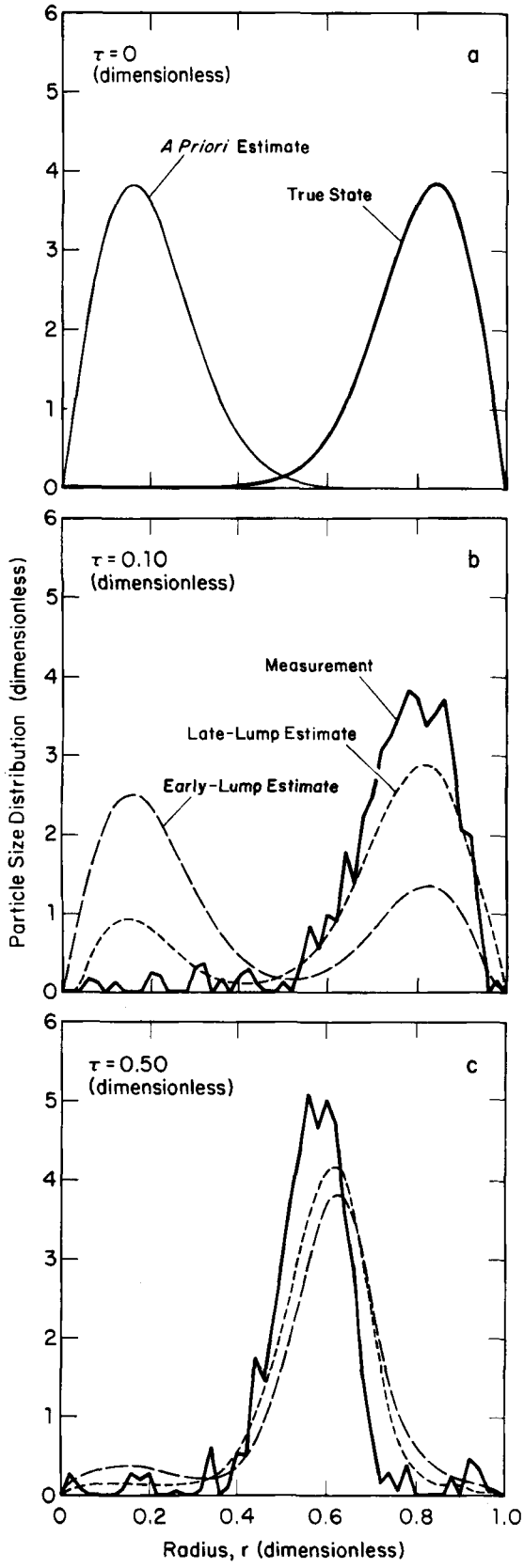

Figure 4. Particle size distribution in fluidized bed.

Figure 4a shows the true initial state of the system and the *a priori* estimate of it used for the optimal estimate initial conditions. The *a priori* estimate, shown as a mirror image of the true system state, is a deliberately poor initial estimate so that relative convergence rates can be compared. The parameter values and relations found in Table 1 are used in the computations for this example.

In the example, we test relative filter convergence rate; we also test relative ability to track system dynamics. We do this by using a bad dynamic model in the filter derivations. The bad dynamic model is created in a very straightforward manner: the attrition rate causes the system dynamics in these examples because the feed PSD is the same as the initial bed PSD (see Table 1). Attrition (particle breakage) causes all particles to become smaller. Because we start initially with all large particles, the attrition causes the mode of the true distribution to travel from right to left across the domain of particle radius as viewed in Figure 4. With a zero attrition rate, the bed PSD would not change.

A bad model can be created by using a different attrition rate in the filtering algorithm than the one used in the model generating the noisy measurements. For the case presented, we used a value of $k = 0.25$ for the attrition rate constant in the model used in deriving the filtering algorithms, and used a value of $k = 0.75$ in the model used to generate the noisy measurements. This difference in attrition rate has a dramatic effect, with the filtering algorithm model substantially underestimating the true system dynamics.

We increase the rate of spatial decay for the spatial correlation in this example by defining the covariance weighting matrices as:

$$Q(y_n,y_k,\tau) = d_Q e^{-10.0|y_n-y_k|} \tag{36a}$$

$$R_{\text{diag}}(y_n,y_n,\tau) = d_R \tag{36b}$$

$$p_0(y_n,y_k) = d_P e^{-10.0|y_n-y_k|} \tag{36c}$$

Note that again there is only one independent variable so that the variables d_Q, d_R, and d_P are scalars, and that we do not consider spatial correlation in the measurements. To account for the very poor dynamic model and the very poor initial estimate, we assigned the values $d_Q = d_P = 0.10$. The measurements were considered to be relatively good, so the value $d_R = 0.01$ was assigned to reflect this.

Figure 4 illustrates the behavior of the early-lumped and late-lumped optimal estimates as they converge toward the noisy measurements for dimensionless times $\tau = 0.10$ and $\tau = 0.50$, respectively. The noisy measurements shown in these figures are zero-mean distributed around the true state (not shown), which is itself undergoing dynamics as the attrition of particles cause the distribution mode to shift. The values assigned to the covariance weighting variables, d_Q, d_R, and d_P, are very illustrative for the comparison of filter performance. Figure 4 illustrates that the late-lumped distributed filter is converging faster and tracking state dynamics better. Figure 5, illustrating the history of the system for particle radius r = 0.5, shows that the late-lumped filter performs better in estimating the true system state over all time for the conditions used.

TABLE 1. VALUES AND RELATIONS USED IN EXAMPLE 2

Parameter	Value/Relation	Units
Feed rate	$F_0 = 1.0$	Mass/time
Feed PSD	$p_0(r) = 40.0re^{-20.0r^2}$	—
Initial bed PSD	$p_b(r,\tau_0) = 40.0re^{-20.0r^2}$	—
Max. radius	$R_{\max} = 1.0$	—
Bed weight	$w = 5.0$	Mass

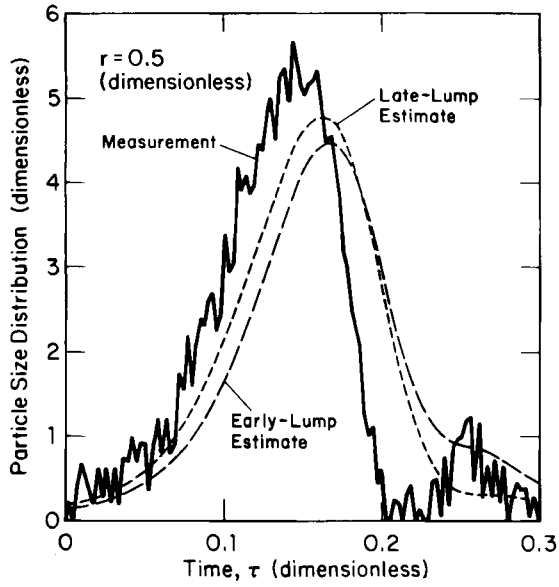

Figure 5. History of fluidized bed for particle radius $r = 0.5$.

CONCLUDING REMARKS

Early lumping results in spatially discrete covariance matrices that can be partitioned into diagonal form. The early-lumped filter approximation cannot directly account for the spatial noise correlation of distributed systems because it is the off-diagonal terms that provide the direct mathematical communication between the discrete locations. The examples illustrate that the late-lumped distributed filter converges faster and tracks system dynamics better than the early-lumped approximation filter.

As stated previously, one very important advantage of using an early-lumped approximation is computation load. A filter is of no use for real-time application if the resulting algorithm cannot be solved on the process computer in real time. To illustrate the burden of computation for the late-lumped distributed algorithm, an algorithm that contains full covariance weighting matrices, consider the following: two 50 × 50 diagonal matrices can be multiplied together in just 50 multiplication operations, a 50 × 50 diagonal matrix multiplied by a full 50 × 50 matrix requires 2,500 multiplications, and two full 50 × 50 matrices multiplied together require 125,000 multiplications. When we consider that a filtering algorithm must be repeatedly solved over time, the computer burden considerations become significant.

In general, a good programmer can expect to implement the early-lumped approximation filter at a minimum of half the computation load required for the late-lumped distributed filter. This was the computation reduction we achieved for the heated-bar example. For the fluidized bed particle-size distribution example, the early-lumped approximation filter was implemented at about one-tenth the computation load of the late-lumped distributed filter. This tremendous savings in computation load was achieved because the use of a central divided difference approximation for the first spatial derivative coincidentally caused the error covariance matrix, P, to remain diagonal throughout the early-lumped approximation simulation. Recognition of this when programming the simulation led to the tremendous computational savings.

NOTATION

a_i = weighting coefficient of basis function in Eq 1
$\mathcal{A}$ = well-posed linear spatial-differential operator
A = discrete approximation of $\mathcal{A}$
$\mathbf{A}$ = matrix notation for A
B = second spatial derivative orthogonal collocation operator
$\mathbf{B}$ = matrix notation for B
$\mathbf{D}$ = central-difference first-derivative matrix, defined in Eq. 35
d_P = numeric value of diagonal terms in $\mathbf{P}(t)$
d_Q = numeric value of diagonal terms in $\mathbf{Q}(t)$
d_R = numeric value of diagonal terms in $\mathbf{R}(t)$
F_0 = feed rate in second example, mass/time
F_1 = discharge rate in second example, mass/time, defined in Eq. 27
J = quadratic least-squares error criterion, defined in Eq. 7
J_{aug} = augmented form of J, defined in Eq. 8
k = attrition rate constant, 1/time, defined in Eq. 26
M = linear distributed modeling operator
$\mathbf{M}$ = matrix notation for spatially discretized M
N = number of discrete spatial locations and measurements
p_0 = particle size distribution of feed in second example
p_1 = particle size distribution of discharge in second example
p_b = particle size distribution of fluidized bed in second example
$\mathbf{p}_b$ = matrix notation for spatially discretized p_b
P = state error covariance matrix
$\mathbf{P}$ = matrix notation for spatially discretized P
P_0 = covariance matrix of x_0, defined in Eq. 6
$\mathbf{P}_0$ = matrix notation for spatially discretized P_0
$\dot{P}$ = first time-derivative of P
$\dot{\mathbf{P}}$ = matrix notation for spatially discretized $\dot{P}$
Q = covariance matrix of u, defined in Eq. 5a
$\mathbf{Q}$ = matrix notation for spatially discretized Q
r = any radius in the interval $[0,R_{max}]$, dimensionless
R = covariance matrix for v, defined in Eq. 5b
$\mathbf{R}$ = matrix notation for spatially discretized R
R_{max} = maximum dimensionless radius
t = any time in the interval $[t_0,t_f]$
t_0 = initial time
t_f = final time
u = noise vector representing uncertainty in the state model
$\mathbf{u}$ = matrix notation for spatially discretized u
v = noise vector representing uncertainity in the measurement model
w = weight of fluidized bed (mass)
x = the n-dimensional system state
x_0 = *a priori* estimate of system state at initial time
$\mathbf{x}_0$ = matrix notation for spatially discretized x_0
y,y' = any spatial location in Ω
z = measured output, defined in Eq. 3b
$\mathbf{z}$ = matrix notation for spatially discretized z

Greek Letters

α = dynamic model coefficient, defined in Eq. 30
β = dynamic model coefficient, defined in Eq. 31
δ = Dirac delta function
γ = dynamic model coefficient, defined in Eq. 32
λ = Lagrange multiplier function
$\dot{\lambda}$ = first time-derivative of λ
Ω = bounded open set in Euclidean n-space
ϕ = dimensionless temperature used in first example
Φ = basis functions in solution approximation, defined in Eq. 1
τ = dimensionless time used in both examples

χ = optimal state estimate
$\dot{\chi}$ = first time-derivative of χ
$\underline{\chi}$ = matrix notation for spatially discretized χ

LITERATURE CITED

Ajinkya, M. B., "The Optimization and State Estimation of Distributed Parameter Systems," Ph.D. Thesis, State Univ. of New York at Buffalo, 116, (1975).

Ajinkya, M. B., et al., "The Experimental Implementation of a Distributed Parameter Filter," *Automatica*, **11**, 571 (1975).

Clement, K., S. B. Jorgensen, and J. P. Sorensen, "Fixed Bed Reactor Kalman Filtering and Optimal Control," *Chem. Eng. Sci.*, **35**, 1,231 (1980).

Collins, P. L., and H. C. Khatri, "Identification of Distributed-Parameter Systems Using Finite Differences," *J. Basic Eng.*, **91**, 239 (1969).

Cooper, D. J., and D. E. Clough, "Real-Time Estimation of Dynamic Particle-Size Distributions in a Fluidized Bed: Theoretical Foundation," *AIChE J.*, **31**, 1,202 (July 1985).

Desalu, A. A., L. A. Gould, and F. C. Schweepe, "Dynamic Estimation of Air Pollution," *IEEE Trans. Auto. Control*, **AC-19**, 904 (1974).

Finlayson, B. A., *The Method of Weighted Residuals and Variational Principles*, Academic Press, New York, 96 (1972).

Ito, A., H. Kanoh, and M. Masubuchi, "MWR Approximation and Modal Control of Parallel- and Counter-Flow Heat Exchangers," *Control of Distributed Parameter Systems*, S. P. Banks and A. J. Pritchard, Eds., Pergamon Press, Elmsford, NY, 423 (1978).

Kitamura, S., and Y. Taniguchi, "Parameter Estimation of One- and Two-Dimensional Heat Conduction Processes Using a Minicomputer," *Int. J. Control*, **34**, 955 (1981).

Koda, M., "Finite Difference Implementation of Distributed Parameter Filters," *Automatica*, **15**, 687 (1979).

Koda, M., and J. H. Seinfeld, "Estimation of Urban Air Pollution," *Automatica*, **14**, 583 (1978).

Lausterer, G. K., and W. H. Ray, "Distributed Parameter State Estimation and Optimal Feedback Control—An Experimental Study in Two Space Dimensions," *IEEE Trans. Auto. Control*. **AC-24**, 179 (1979).

Lausterer, G. K., W. H. Ray, and H. R. Martens, "Real-Time Distributed-Parameter State Estimation Applied to a Two-Dimensional Heated Ingot," *Automatica*, **14**, 335 (1978).

Levenspiel, O., D. Kunii, and T. Fitzgerald, "The Processing of Solids of Changing Size in Bubbling Fluidized Beds," *Powder Technol.*, **2**, 87 (1968/69).

Meditch, J. S., "Least-Squares Filtering and Smoothing for Linear Distributed-Parameter Systems," *Automatica*, **7**, 315 (1971).

Omatu, S., and J. H. Seinfeld, "Filtering and Smoothing for Linear Discrete-Time Distributed Parameter Systems Based on Wiener-Hopf Theory with Application to Estimation of Air Pollution," *IEEE Trans. Systems Man, Cyber.*, **SMC-11**, 785 (1981).

Orner, P. A., P. F. Salamon, Jr., and W. Yu, "Least-Squares Simulation of Distributed Systems," *IEEE Trans. Auto. Control*, **AC-20**, 75 (1975).

Polis, M. P., R. E. Goodson, and M. J. Wozny, "On Parameter Identification for Distributed Systems Using Galerkin's Criterion," *Automatica*, **9**, 53 (1973).

Prabhu, S. S., and I. McCausland, "Optimal Control of Linear Diffusion Processes with Quadratic Error Criteria," *Automatica*, **8**, 299 (1972).

Ramirez, W. F., "Identification for Distributed Parameter Chemical Engineering Systems—Fixed-Bed Tubular Reactors," *Distributed Parameter Control Systems*, S. G. Tzafestas, Ed., Pergamon Press, Elmsford, NY, 391 (1982).

Ray, W. H., *Advanced Process Control*, McGraw-Hill, New York, 201 (1981).

Sage, A. P., *Optimum Systems Control*, Prentice-Hall, Englewood Cliffs, NJ, 151 (1968).

Seidman, T. I., "Approximation Methods for Distributed Parameter Systems," *Distributed Parameter Systems Control*, S. G. Tzafestas, Ed., Pergamon Press, Elmsford, NY, 273 (1982).

Seinfeld, J. H., and W. H. Chen, "Estimation of Parameters in Partial Differential Equations from Noisy Experimental Data," *Chem. Eng. Sci.*, **26**, 753 (1971).

Sorensen, J. P., S. B. Jorgensen, and K. Clement, "Fixed-Bed Reactor Kalman Filtering and Optimal Control," *Chem. Eng. Sci.* 35, 1,223 (1980).

Tanner, R., "Estimating Kinetic Rate Constants Using Orthogonal Polynomials and Picard's Iteration Method," *Ind. Eng. Chem. Fund.*, **11**, 1 (1972).

Tzafestas, S. G., "Distributed Parameter State Estimation," *Distributed-Parameter Systems—Identification, Estimation and Control*, W. H. Ray and D. G. Lainiotis, Eds., Dekker, New York, 135 (1978).

Van den Bosch, B. A. J., "Identification of Parameters in Distributed Chemical Reactors," *Distributed-Parameter Systems—Identification, Estimation and Control*, W. H. Ray and D. G. Lainiotis, Eds., Dekker, New York, 555 (1978).

Villadsen, J., and M. L. Michelsen, *Solution of Differential Equation Models by Polynomial Approximation*, Prentice-Hall, Englewood Cliffs, NJ, 143 (1978).

Viskanta, R., et al., "Thermal Remote Sensing of Temperature Distribution in Glass," *Automatica*, **11**, 409 (1975).

Weimer, A. W., and D. E. Clough, "Dynamics of Particle Size/Conversion Distributions in Fluidized Beds: Application to Char Gasification," *Powder Technol.*, **26**, 11 (1980).

Yoshimura, T., and A. Campo, "Identification of a Heat Transfer System Using the Galerkin Method," *Int. J. Systems Sci.*, 13, 247 (1982).

Manuscript received June 15, 1984, and revision received Feb. 12, 1985.