Establishing Independence of Continuous Process Product Samples

Reed L. Christiansen
Eastman Chemical Company, Kingsport, TN 37662

New product qualification frequently requires that a manufacturer provide customers with several independent samples for evaluation. For continuous manufacturing processes, equipment residence time distributions (RTDs) introduce serial dependence among the samples. Determining RTD by pulse or step modulation of a tracer or process variable can produce off-spec material and add costs by limiting the opportunity to conduct other product or process development experiments. A novel method of determining a continuous process' RTD does not risk production of off-spec material. Through analysis of serial component analytical data, the process' washout curve can be inferred. The theoretical basis for the method is presented and experimental requirements enumerated. Applicability to well-mixed, laminar flow and convective-dispersive processes is demonstrated. Experimental validation in a specialty polymer pilot plant resulted in an error of about 5% in mean residence time, when compared to estimates from pulse tests and calibrated level measurements.

Introduction

New product qualification frequently requires that a manufacturer provide customers several statistically independent samples for evaluation. Sample independence allows customers to interpret fitness-for-use tests using basic statistical tools. While demonstration of sample independence is straightforward for batch or semi-batch processes, it is more difficult for continuous processes, where the presence of equipment residence time distributions (RTDs) introduces dependence among serial samples.

Existing options for demonstration of continuous process sample independence are to perform intrusive RTD testing or to separate the samples by shutdown/clean equipment/startup cycles. For a typical new product, at least one of the raw materials is likely to be costly and in short supply. Both of these options diverts the scarce raw material into production of off-specification product, whether due to tracer contamination, operation at nonstandard conditions, or transient operation during process shutdown and startup. Any raw material that is expended for these purposes cannot be used for valuable product and process development experiments.

On a pilot and semiworks scale, early in a process' life cycle, catalyst and additive mix-to-mix concentration variation and feed flow rate variation are likely to be larger than seen for more mature processes at the manufacturing scale. However, there may be safety, economic, or technical constraints precluding frequent measurements of mix concentrations. Similarly, for economic reasons, inlet flow rate regulation and measurement at a small scale may also be inaccurate and imprecise. While mature literature exists on identification of statistical transfer functions from observed inputs and outputs (Box et al., 1994), this becomes much more difficult when input measurements are not readily available.

An empirical property of continuous processes is that the autocorrelation coefficients of concentrations of nonreactive components in product samples decay with increasing lag and become statistically insignificant for very long lag times. This property is assumed implicitly in the development of the extent of attenuation of the variance of a species' concentration during blending in Danckwerts (1953). It is also invoked explicitly in the development of stochastic models for continuous flow systems in Václavek (1967). The corresponding assumption in viscous mixing analysis, where the spacial correlation of species' concentrations is assumed to become negligible beyond a characteristic length scale, is discussed in Ottino and Chella (1983). It is shown in the sequel that this long lag time decay of the autocorrelation function can be linked to the physical washout of the analyzed component.

The present work describes a means of determining the time interval between independent samples from a continuous process, making use of only sequential analyses of nonreactive components in the product. No overt manipulation of process inputs is attempted, nor are input or disturbance variables assumed to be measured. The theoretical basis for this method is developed and its limitations are enumerated. Mathematical examples demonstrating the applicability of the technique to well-mixed, laminar flow, and dispersive-convective systems are then provided. Lastly, an experimental example showing good correspondence between the present method and results of pulse testing on a pilot-scale polymerization process is presented.

Discrete RTD Theory

We assume that the output concentration of the system sampled at a fixed time interval Δt can be represented by a linear difference equation in time with constant coefficients

$$Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + \sum_{j=0}^{q} b_j U_{n-d-j-1} \tag{1}$$

where Y_n is the measured output at time t_n, U_n is the unmeasured input at time t_n, and d is the number of intervals of dead-time due to transportation lag in the system. Without loss of generality, we will factor out the d units of dead-time and refer to

$$X_n \equiv U_{n-d} \tag{2}$$

so that

$$Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + \sum_{j=0}^{q} b_j X_{n-j-1} \tag{3}$$

The continuous-time analog of Eq. 3 is discussed in the review (Nauman, 1981). It is capable of representing single-input systems with recycle streams and flow bypassing. Cascades of perfectly-backmixed reactors with stagewise varying residence times have been similarly represented (Fan et al., 1969). The ability of Eq. 3 to represent perfectly-mixed, laminar flow, and dispersive-convective systems is demonstrated as examples below.

To derive the corresponding discrete RTD, we recall that the RTD of a linear system is simply its impulse response function, irrespective of the actual input sequence experienced by the system (Nauman, 1981). Accordingly, assume that a tracer pulse of concentration A is introduced at the inlet of the system described by Eq. 3 at time $-d$ so that

$$X_n = \delta_{n0} A \tag{4}$$

where δ_{n0} is the Kronecker delta

$$\delta_{ij} \equiv \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{5}$$

The normalized RTD

$$C_n \equiv Y_n / \sum_{i=0}^{\infty} Y_i$$

is

$$C_n = \begin{cases} 0, & n \leq 0 \\ \sum_{i=1}^{p} a_i C_{n-i} + \left(1 - \sum_{j=1}^{p} a_j\right) b_{n-1} \Big/ \sum_{k=0}^{q} b_k, & 0 < n \leq q+1 \\ \sum_{i=1}^{p} a_i C_{n-i}, & n > q+1 \end{cases} \tag{6}$$

where we have made use of the overall material balance

$$\sum_{i=0}^{\infty} Y_i = A \sum_{n=0}^{q} b_n \Big/ \left(1 - \sum_{j=1}^{p} a_j\right) \tag{7}$$

The cumulative RTD

$$F_n \equiv \sum_{i=0}^{n} C_i$$

is given by

$$F_n = \begin{cases} 0, & n \leq 0 \\ \sum_{i=1}^{p} a_i F_{n-i} + \left(1 - \sum_{j=1}^{p} a_j\right) \sum_{k=0}^{n} b_{k-1} \Big/ \sum_{\ell=0}^{q} b_\ell, & 0 < n \leq q \\ \sum_{i=1}^{p} a_i F_{n-i} + \left(1 - \sum_{j=1}^{p} a_j\right), & n > q \end{cases} \tag{8}$$

Of direct interest is the fraction of tracer remaining at time-step n, which is given by the washout function $W_n \equiv 1 - F_n$

$$W_n = \begin{cases} 1, & n \leq 0 \\ \sum_{i=1}^{p} a_i W_{n-i} + \left(1 - \sum_{j=1}^{p} a_j\right) \sum_{k=n+1}^{q} b_k \Big/ \sum_{\ell=0}^{q} b_\ell, & 0 < n \leq q \\ \sum_{i=1}^{p} a_i W_{n-1}, & n > q \end{cases} \tag{9}$$

W_n represents the fraction of the fluid which entered n time steps ago and is still remaining in the reactor. It is also the fraction of fluid exiting the reactor that was present n time steps in the past. Evidently, $F_n \equiv 1 - W_n$ is therefore the probability that samples taken n time steps apart are independent.

ARMA Model of Time Series

In the case where we have no knowledge of the inlet concentrations to the system at each time-step, it is still possible to represent the outlet concentration time series in terms of an empirical autoregressive moving average (ARMA) model

(Box et al., 1994)

$$y_n = \sum_{i=1}^{p} \alpha_i y_{n-i} + \sum_{j=0}^{q} \beta_j \epsilon_{n-j} \tag{10}$$

where $y_n \equiv Y_n - \overline{Y}$, $\overline{Y}$ is an average value of Y_n, and ϵ_n is the residual error (or "residual") of the model at time-step n. The residual ϵ_n is equal to the model prediction error at time step n. Such a model is said to be of order (p, q) and is referred to as an ARMA (p, q) model. The terms in the first summation are the autoregressive terms; those in the second are the moving average terms.

Many efficient numerical algorithms exist for determining the coefficients α_i and β_j (cf. Therrien, 1992). Most commercially available desktop statistical software will compute these coefficients quite rapidly. Conceptually, these algorithms are based on analysis of the sample estimates of the autocorrelations of y_n at lag k

$$r_k = \sum_{i=0}^{N-k} y_i y_{i+k} \Big/ \sum_{i=0}^{N} y_i y_i,\ k \ll N \tag{11}$$

where N is the number of samples included in the analysis. For the case of the least-squares solution, the autoregressive coefficients are determined from the autocorrelations at lags $q-p+1$ to $q+p$ from the solution of the linear equations

$$r_{q+k} = \sum_{i=1}^{p} \alpha_i r_{q+k-i}, \quad k = 1, \ldots, p \tag{12}$$

The moving average parameters are then found by solving a set of nonlinear equations involving the derived moving average process w_n, defined as

$$\begin{aligned} w_n &\equiv y_n - \sum_{i=1}^{p} \alpha_i y_{n-i} \\ &= \sum_{j=0}^{q} \beta_j \epsilon_{n-j} \end{aligned} \tag{13}$$

If r'_k are the sample estimates of autocorrelations of the derived w_n, defined analogously to Eq. 11, the parameters β_j satisfy the equations

$$r'_k = \frac{\beta_k + \beta_1 \beta_{k+1} + \ldots + \beta_{q-k} \beta_q}{1 + \beta_1^2 + \ldots + \beta_q^2}, \quad k = 1, \ldots, q \tag{14}$$

and $\beta_0 = 1$. Both (Therrien, 1992; Box et al., 1994) describe algorithms for solving these nonlinear equations.

Heuristics for determination of the model order (p, q), described in Box et al. (1994), generally rely on the structure of the autocorrelations, as well as that of the related partial autocorrelation functions. In general, the lower-lag autocorrelations will have contributions from both the autoregressive and moving average portions of the model, while the larger-lag autocorrelations will depend only on the autoregressive coefficients, as described above.

Several numerical criteria for model order selection have been derived, which balance increases in the number of model parameters vs. improvements in accuracy resulting from those increases. Schwarz's BIC criterion provides suitably parsimonious estimates of the model order

$$\min_{p,q} \mathrm{BIC}(p, q) = \ln\left(\sigma_a^2\right) + (p + q + 1) \frac{\ln(N)}{N} \tag{15}$$

where σ_a^2 is the variance of the ARMA (p, q) model residuals.

Correspondence Between ARMA Model and Discrete RTD

This section describes conditions under which a system's discrete RTD may be inferred from an empirical ARMA model. The presence of feedback control complicates this task, with perfect feedback control completely masking information about the system under control (Box and MacGregor, 1974; MacGregor and Fogal, 1995). However, recent work in model predictive control has led to the development of algorithms which penalize control actions which would prevent model parameter identification, generally allowing the controlled variable freedom to wander in some range of the set point, cf. Genceli and Nikolaou (1996), Vuthandam and Nikolaou (1997), Eker and Nikolaou (1998), and Kothare et al. (1998). The treatment in the sequel is sufficiently general to apply to systems under any form of feedback control, so that conditions allowing identification of the RTD may be expressed in a general manner.

We consider a system with discrete RTD given by Eq. 3 whose outlet concentration Y_n is periodically measured

$$Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + \sum_{j=0}^{q} b_j \left(U_{n-d-j-1} + \nu'_{n-d-j-1}\right) + \xi'_n \tag{16}$$

Here, ν'_n represents any time-varying disturbances which may be affecting the inlet concentration. Similarly, ξ'_n represents the combination of any measurement errors and disturbances, also possibly time-varying, which may affect the outlet concentration.

Without loss of generality, we assume that the outlet concentration is being maintained at some possibly time-varying set point $Y_{sp,n}$ by a feedback controller, whose control law may also be time-varying

$$e_n \equiv Y_{sp,n} - Y_n \tag{17}$$

$$U_n = U_{n-1} + \sum_{\ell=0}^{r} c'_{\ell,n} e_{n-\ell-d_c} \tag{18}$$

Here, e_n is the deviation from set point, $c'_{\ell,n}$ are the coefficients for a feedback controller in "velocity" form (Ogunnaike and Ray, 1994), and d_c represents the minimum deadtime introduced by the controller. Note that Eq. 18 could describe an actual automatic feedback controller or some sporadic control actions taken by an operator. In the case where the controller time delay increases from its minimum value d_c, some of the leading coefficients of $c'_{\ell,n}$ would be zero. In the case where no feedback control actions are taken, all of the $c'_{\ell,n}$ would vanish.

By applying the z-Transform to Eqs. 16–18, substituting the expressions for the transforms of U_n and e_n into that for the transform of Y_n, and inverting, we obtain

$$Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + \sum_{j=0}^{q} \sum_{\ell=0}^{r} \sum_{m=0}^{\infty} b_j c'_{\ell,n-d-d_c-j-\ell-m-1} \times \left(Y_{\mathrm{sp},n-d-d_c-j-\ell-m-1} - Y_{n-d-d_c-j-\ell-m-1} \right) + \sum_{j=0}^{q} b_j \nu'_{n-d-j-1} + \xi'_n \quad (19)$$

It is assumed that set point changes are made only infrequently during the production campaign, on a time-scale greater than the time constant associated with RTD washout. This mode of operation is typical when several grades of product would be made and/or a designed experiment carried out; the amount of time spent at each set of conditions would be maximized to minimize the relative yield losses occurring due to transition off-class. Under such an assumption, the effects of set point changes can be minimized by subjecting the data to a high pass filter, cf. Ljung (1999). Use of such a filter would also provide the concomitant benefits of at least partially whitening the disturbance time series and eliminating long-time cross-correlations between the time series of the output and disturbance variables. A suitable filter would exhibit zero gain in the low frequency limit and unity gain in the high frequency limit. It would also be linear, such as the filter

$$y_n \equiv Y_n - \overline{Y}_n$$
$$\overline{Y}_n = \alpha \overline{Y}_{n-1} + (1-\alpha) Y_{n-1} \quad (20)$$

with $0 < \alpha < 1$. Application of this filter to Eq. 19 results in

$$y_n = \sum_{i=1}^{p} a_i y_{n-1} - \sum_{j=0}^{q} \sum_{\ell=0}^{r} \sum_{m=0}^{\infty} b_j (c_\ell y)_{n-d-d_c-j-\ell-m-1} + \sum_{j=0}^{q} b_j \nu_{n-d-j-1} + \xi_n \quad (21)$$

with

$$(c_\ell y)_n \equiv c'_{\ell,n} Y_n - \overline{c'_\ell Y_n}$$
$$\overline{c'_\ell Y_n} = \alpha \overline{c'_\ell Y}_{n-1} + (1-\alpha) c'_{\ell,n-1} Y_{n-1} \quad (22)$$
$$\nu_n \equiv \nu'_n - \overline{\nu}_n$$
$$\overline{\nu}_n = \alpha \overline{\nu}_{n-1} + (1-\alpha) \nu'_{n-1} \quad (23)$$
$$\xi_n \equiv \xi'_n - \overline{\xi}_n$$
$$\overline{\xi}_n = \alpha \overline{\xi}_{n-1} + (1-\alpha) \xi'_{n-1} \quad (24)$$

Note that we have made use of the linearity property of the filter of Eq. 20 to obtain these results.

We now calculate the autocorrelations of the time series represented by Eq. 21, at lags $q-p+1$ to $q+p$ to determine sufficient conditions for equality of the autoregressive coefficients α_i of Eq. 10 and the a_i of the discrete RTD of Eq. 3. Assuming that the filtered data time series y_n is not correlated with the filtered inlet disturbance process time series ν_n for lags $k > d+q+1$, and that the time series y_n is also not correlated with the filtered outlet disturbance process time series ξ_n for all lags, the autocorrelations of y_n for lags $q-p+1$ to $q+p$ are given by

$$r_k = \sum_{i=1}^{p} a_i r_{k+i} + \sum_{j=0}^{q} \sum_{\ell=0}^{r} \sum_{m=0}^{\infty} b_j \sum_{n}^{N-k} y_{n+k} (c_\ell y)_{n-d-d_c-j-\ell-m-1} \Big/ \sum_{n=1}^{N} y_n^2 \quad (25)$$

with the lower limit of the summation over n being $n = d + d_c + j + \ell + m + 2$. Comparison of Eqs. 12 and 25 corroborates the literature cited above that the presence of control actions causes the autoregressive parameters α_i to be biased estimates of the discrete RTD a_i. The magnitude of this bias could be reduced by making the size of control actions as small as possible, while taking control action as infrequently as possible. Increasing the minimum controller time delay d_c would be another strategy to decrease bias, since an increase in d_c causes higher-order lags of the cross-correlations to appear in the bias term. Note that an excessive increase in controller time delay would likely lead to an unacceptable increase in required control action magnitude, however, which could be counterproductive.

Assuming that actions are taken to minimize the bias in estimating the a_i of the discrete RTD, it remains to analyze the derived residual process that would be identified from the system represented by Eq. 21

$$w_n = \sum_{j=0}^{q} b_j \nu_{n-d-j-1} + \xi_n \quad (26)$$

By comparing this equation to Eq. 13, we postulate that the ARMA residual error time series ϵ_n is a linear combination of the filtered disturbance process time series ν_{n-d-1} and ξ_n

$$\epsilon_n = \phi_1 \nu_{n-d-1} + \phi_2 \xi_n \quad (27)$$

where ϕ_1 and ϕ_2 are constants to be determined. Assuming that the filtered disturbance process time series ν_n and ξ_n are uncorrelated, we can equate the autocorrelations of w_n obtained from Eqs. 13 and 26 to obtain

$$0 = \sum_{i=0}^{q} \sum_{j=0}^{Q} \left(b_i b_j - \phi_1^2 \beta_i \beta_j \right) r_{\nu\nu(k+i-j)} \quad (28)$$

$$0 = r_{\xi\xi}(k) - 2\phi_2 \sum_{j=0}^{q} \beta_j r_{\xi\xi}(k+j) + \left(\phi_2 \sum_{j=0}^{q} \beta_j \right)^2 r_{\xi\xi}(k) \quad (29)$$

where $r_{\nu\nu}(k)$ and $r_{\xi\xi}(k)$ represent the k-th autocorrelations of ν_n and ξ_n. Under the additional assumption that the filtered disturbance time series ξ_n is white noise, we conclude that solutions for ϕ_1 and ϕ_2 can be found satisfying

$$b_j = \phi_1 \beta_j, \quad j = 0, \ldots, q \tag{30}$$

$$\phi_2 = \left(\sum_{j=0}^{q} \beta_j \right)^{-1} \tag{31}$$

Equation 30 describes relationships between each coefficient b_j, which throughout this development multiplies the inlet concentration and its disturbances, and the corresponding coefficient β_j, which multiplies residuals of the outlet concentration. For these relationships to hold, we physically require that the yield of the tracer analyzed at the outlet of the system, with respect to the levels introduced at the system inlet, remain constant over the course of the time-series.

Note that Eqs. 6, 8, and 9, which describe the properties of the normalized discrete RTD, refer only to ratios of the b_j, and not to their absolute values. Therefore, if the constant yield requirement of Eq. 30 is satisfied, knowledge of the identified β_j is sufficient to completely specify the normalized discrete RTD.

To summarize, proper identification of the a_i parameters of the discrete RTD requires that feedback control actions be minimal and infrequent. Subsequent identification of the b_j parameters requires use of a tracer whose input-output yield remains constant over the course of the time-series. Taken together, these criteria suggest that catalysts, additives, or tie species that have the reputation with plant operations of having low-fuss ‘‘set and forget’’ open-loop feed systems would be ideal tracer candidates.

Concerning Sampling Interval Δt

The fidelity of the representation of a continuous RTD by its discrete counterpart will depend on the choice of sampling interval Δt. If our intent was to identify a discrete model for process control purposes, selection of Δt could be a critical decision. That is not the case for the present problem.

Let us assume that customers are being sent railcars, and analytical samples are being taken at the start of filling each. If subsequent analysis indicates that there is no correlation of serial samples, one may conclude that a customer could safety be sampled every other railcar, with reasonable assurance of sample independence. Although the data collection frequency was not rapid enough to develop a detailed discrete RTD model, it was sufficient to answer the problem at hand.

In another instance, let us assume that boxes are being filled, with analytical samples again being taken at the start of filling each. We may find that the time required to fill, for example, eight boxes, is the time required for RTD washout. In this instance, if analytical samples were not required from each box to complete certificates of analysis for quality assurance purposes, the analytical sampling interval could certainly be increased without loss of information. An analysis of optimal sampling interval selection in the presence of noisy data may be found in Ljung (1999).

In short, as long as samples are taken at least as frequently as product containers are being filled, the choice of Δt is not critical.

Example: Applicability to a Continuous Stirred Tank Reactor

In this example, we show that it is possible to represent the sampled input-output relationship of a perfectly-mixed continuous stirred tank reactor in terms of the discrete difference Eq. 1.

We consider a constant-density well-mixed tank, with inlet concentration $U(t)$, outlet concentration $Y(t)$, and mean residence time τ. The outlet concentration is described by

$$\frac{dY}{dt} = \frac{1}{\tau}[U(t) - Y(t)] \tag{32}$$

Integrating this equation between times $t = (n-1)\Delta t$ and $t = n\Delta t$, while assuming that $U(t) \approx U((n-1)\Delta t)$ during this time interval, yields

$$Y(n\Delta t) = \exp(-\Delta t/\tau) Y((n-1)\Delta t) + (1 - \exp(-\Delta t/\tau)) U((n-1)\Delta t) \tag{33}$$

This is of the form of Eq. 1, with $p = 1$, $q = 0$, $d = 0$, and

$$a_1 \equiv \exp(-\Delta t/\tau)$$
$$b_0 \equiv 1 - \exp(-\Delta t/\tau) \tag{34}$$

Example: Applicability to a Newtonian Fluid in Circular Pipe

In this example, we show that it is possible to represent the sampled input-output relationship of a laminar flow system in terms of the discrete difference Eq. 1. We consider the case of a pipe of length L with circular cross-section of radius R bearing a Newtonian fluid with parabolic velocity profile $\upsilon(r)$

$$\upsilon(r) = \frac{2L}{\tau}\left(1 - \frac{r^2}{R^2}\right) \tag{35}$$

where r is the distance from the centerline and τ is the mean residence-time. At the pipe inlet, a tracer of time-varying concentration $c_0(t)$ is introduced. The cup mixing concentration $c_L(t)$ at the pipe's outlet is given by

$$c_L(t) = \frac{2\tau}{LR^2}\int_0^R r\upsilon(r) c_0[t - L/\upsilon(r)] dr \tag{36}$$

Let us sample the inlet and outlet concentrations with sampling period Δt. Between samples, we assume that the inlet concentration remains constant, so that

$$c_0(t) \approx c_0[(n-1)\Delta t], \quad (n-1)\Delta t \le t < n\Delta t \tag{37}$$

The sampled outlet concentration is then given by

$$c_L(n\Delta t) = \frac{2\tau}{LR^2} \sum_{j=j_{\min}}^{n} c_0[(n-j)\Delta t] \int_{r_a(j)}^{r_b(j)} rv(r)\,dr \quad (38)$$

where $j_{\min}$ is the smallest j for which Eq. 35 can be inverted to solve for a real-valued radius as a function of velocity

$$\begin{aligned} r(v) &= v^{-1}(L/j\Delta t) \\ &= R\sqrt{1-\tau/2j\Delta t} \end{aligned} \quad (39)$$

and $r_a(j)$ and $r_b(j)$ of Eq. 38 are given by

$$r_a(j) \equiv R\sqrt{1-\tau/2j\Delta t} \quad (40)$$

$$r_b(j) \equiv R\sqrt{1-\tau/2(j+1)\Delta t} \quad (41)$$

Upon inserting the velocity profile (Eq. 35) into Eq. 38, integrating and simplifying, a closed-form solution for c_L is found in terms of c_0

$$c_L(n\Delta t) = \sum_{j=j_{\min}}^{n} b'_j c_0[(n-j)\Delta t] \quad (42)$$

with

$$b'_j \equiv \frac{\tau^2}{4\Delta t^2}\left[\frac{1}{j^2} - \frac{1}{(j+1)^2}\right] \quad (43)$$

A truncated version of the results (Eq. 42) would therefore be in the form of the originally assumed discrete difference Eq. 1, with $Y_n \equiv c_L(n\Delta t)$, $U_n \equiv c_0\ (n\Delta t)$, $b_i \equiv b'_{j+1}$, and $d \equiv j_{\min} - 1$.

Example: Applicability to Dispersive-Convective System

In this example, we show that it is possible to represent the sampled input-output relationship of a dispersive-convective system in terms of the discrete difference Eq. 1. We consider a dimensionless partial differential equation describing tracer concentration as a function of distance and time, subject to Danckwerts boundary conditions:

$$\frac{\partial Y}{\partial t} = Pe^{-1}\frac{\partial^2 Y}{\partial x^2} - \frac{\partial Y}{\partial x}$$

$$\text{ICs:} \quad \frac{\partial^2 Y}{\partial x^2}(x,0) = \frac{\partial Y}{\partial x}(x,0) = Y(x,0) = 0$$

$$\text{BCs:} \quad Pe^{-1}\frac{\partial Y}{\partial x}(0,t) = Y(0,t) - U(t)$$

$$\frac{\partial Y}{\partial x}(1,t) = 0 \quad (44)$$

where $U(t)$ is the concentration at the system inlet $x = 0$ and Pe is the Peclet Number, representing the importance of convection relative to axial dispersion. The outlet concentration $Y(1,t)$ is obtained by applying the Laplace transform to Eq. 44 and solving by means of the corresponding Green's function, as explained in Stakgold (1979) and Gay (1989). The solution in the Laplace domain is

$$Y(1,s) = g(s)U(s) \quad (45)$$

where

$$g(s) = \frac{(r_1 - r_2)Pe}{r_1^2 \exp(-r_2) - r_2^2 \exp(-r_1)} \quad (46)$$

and $r_1(s)$ and $r_2(s)$ are the quadratic roots

$$r_1(s), r_2(s) = \frac{Pe}{2} \pm \sqrt{\left(\frac{Pe}{2}\right)^2 + Pe \cdot s} \quad (47)$$

Note that Eqs. 46–47 are equivalent to those tabulated in Wen and Fan (1975) for these boundary conditions.

Given a dimensionless sampling interval Δt, which has been normalized by the dispersionless residence time L/v, these Laplace domain results may be represented in the z-domain using the identity $s \equiv \ln z/\Delta t$

$$Y(1,z) = g(z)U(z) \quad (48)$$

The transfer function $g(z)$ may be expanded as a power series in z^{-1}

$$g(z) = \sum_{j=1}^{q+1} b'_j z^{-j} \quad (49)$$

Note that with knowledge of the coefficients b'_j, Eq. 48 may be inverted to a time-domain difference equation of the desired form of Eq. 1, with $b_j \equiv b'_{j+1}$. These real-valued coefficients may be determined through the inversion of the Laplace domain transfer function (Eq. 46)

$$b'_j = \frac{1}{2\pi i}\int_{v-i\infty}^{v+i\infty} g(s)\exp(st)\,|_{t=j\Delta t}\,ds \quad (50)$$

where $i \equiv \sqrt{-1}$. Here, v is an arbitrary real number greater than the real parts of all of the singularities of $g(s)$. This inversion may be implemented numerically through the use of the algorithm of Honig and Hirdes (1984), which also provides an estimate of the total numerical (discretization plus truncation) error. The efficacy of this algorithm has been explored by Seidel-Morgenstern (1991).

Table 1 displays the first 10 coefficients b'_j resulting from simulating the case of $Pe = 5$ and $\Delta t = 0.5$, along with their

Table 1. Expansion Coefficients b'_j from Eq. 50 and Estimated Total Numerical Error for $Pe = 5$, $\Delta t = 0.5$

j	b'_j	Error	j	b'_j	Error
1	8.99960505×10^{-1}	2.3×10^{-13}	6	1.68637442×10^{-2}	3.0×10^{-14}
2	6.99559779×10^{-1}	3.9×10^{-13}	7	6.38542492×10^{-3}	3.0×10^{-14}
3	2.99994829×10^{-1}	6.9×10^{-13}	8	2.41713934×10^{-3}	6.7×10^{-15}
4	1.16755680×10^{-1}	3.6×10^{-13}	9	9.14921582×10^{-4}	3.4×10^{-15}
5	4.44836814×10^{-2}	2.0×10^{-14}	10	3.46305141×10^{-4}	2.2×10^{-15}

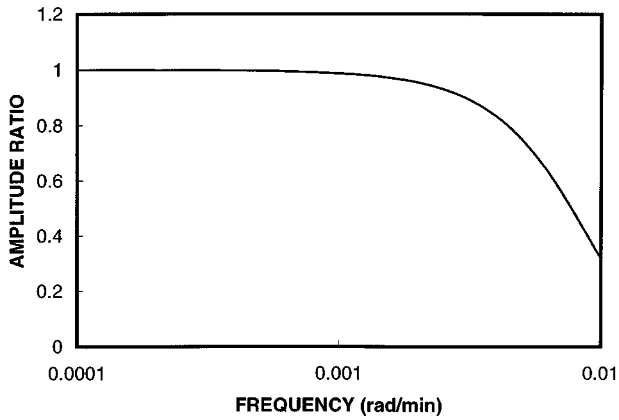

Figure 1. Amplitude ratio vs. frequency from pilot-plant pulse test.

estimated numerical errors. From the rapid rate at which the magnitudes of the coefficients diminish with an increasing index, it may be concluded that this dispersive-convective system can be represented by a few leading terms in the form of Eq. 1.

Example: Application to Polymer Pilot Plant

Pulse testing was conducted on a continuous specialty polymer pilot plant, using standard techniques, cf. Clements and Schnelle (1963) and Ogunnaike and Ray (1994). Figure 1 depicts the amplitude ratio vs. frequency diagram from the analysis. The high-frequency asymptote indicates that the system is approximately first-order. The corner frequency of 0.0046 rad/min corresponds to a time constant of approximately 220 min. Figure 2, summarizing the frequency content of the input pulse vs. frequency, suggests that the quality of the results deteriorates as frequency increases beyond about 0.01 rad/min.

Under the assumption that the 220 min first-order time constant from the pulse testing was related to the residence time of the largest vessel in the line, the vessel's level indicators were calibrated by filling the vessel with a known quantity of solvent. The vessel was then drained in stages, weighing the amount of material drained over time and noting the level indications. Based on these data, and accounting for the density differences between polymer and the calibration solvent, it was estimated that the residence time of the vessel under normal operating conditions was approximately 223 min, agreeing very closely with the results of the pulse testing.

For a time period immediately preceding the pulse tests, process development experiments had been run on the pilot line with polymer product samples taken at 4 h intervals. These samples were submitted for chemical analysis of catalyst residues. A total of 206 data points resulted, not including 20 interspersed missing values. These data are displayed in Figure 3. For the purposes of propriety, the data in this figure have been mean centered and normalized by their sample standard deviation.

Since missing values generally signal a deviation from normal operations, either in the pilot plant or the analytical lab, a conservative approach was taken in estimating autocorrelation coefficients for subsequent analysis. Autocorrelation coefficients were calculated for contiguous sequences of data which had been centered on the mean value for the entire run, according to Eq. 11. A sequence-length-weighted average of each coefficient was then calculated, using estimates from only those sequences that were at least five times as long as the lag whose autocorrelation coefficient was being estimated. This is in keeping with the caveat of Box et al. (1994) that Eq. 11 provides a suitable estimate of the autocorrelation coefficients for sample sizes greater than four times the lag being estimated.

ARMA models were then fit to the autocorrelation estimates; the lowest BIC metric of Eq. 15 was attained by the third-order autoregressive model. However, the apparent 29 h dominant time constant of this model was found to be caused by catalyst target changes during the campaign. To mask the effect of these changes, the high-pass filter of Eq. 20 was applied. A filter parameter $\alpha = 0.75$ was chosen, matching the best-fit first-order autoregressive coefficient of the unfiltered process. The initial condition for the filter was determined by applying the filter to the reversed time-series,

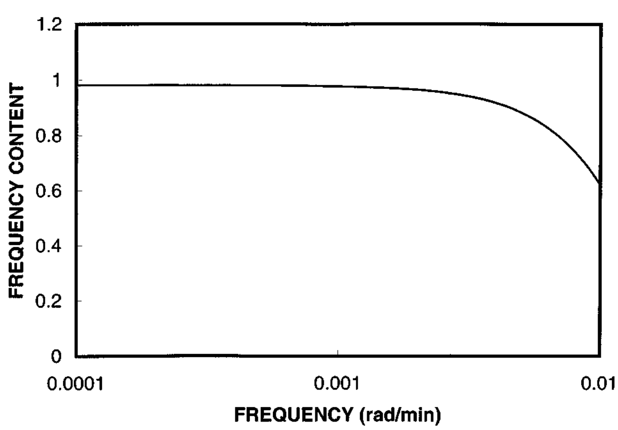

Figure 2. Normalized frequency content vs. frequency from pilot-plant pulse test.

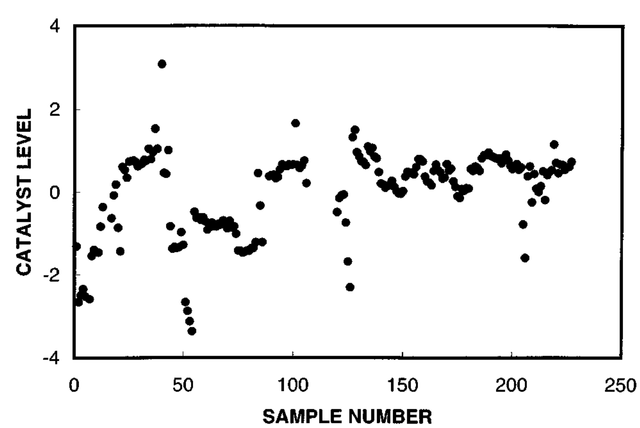

Figure 3. Time series of pilot-plant catalyst analytical data, mean centered and normalized by sample standard deviation.

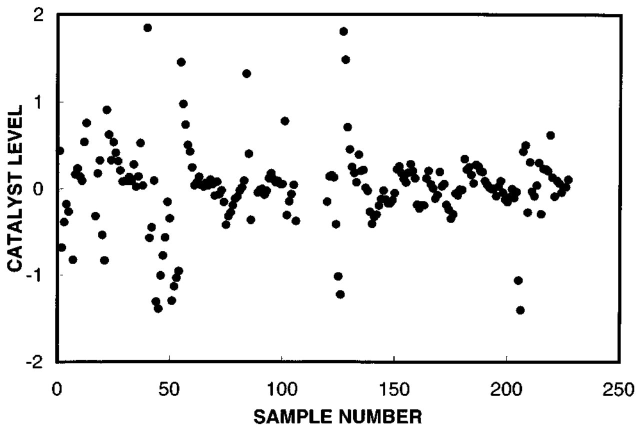

Figure 4. Result of subjecting data of Figure 3 to high-pass filtering ($\alpha = 0.75$).

that is, by forecasting past $\overline{Y}_k$ from future data points (Therrien, 1992). The filtered time series is depicted in Figure 4, again normalized by the original data's standard deviation, for purposes of propriety.

Upon fitting ARMA models to the filtered time-series, the lowest BIC metric was attained by the first-order autoregressive model. The time constant of this model was 230 min., which is in excellent agreement with the estimates of 220 min and 223 min from the pulse test and the calibrated level indicators, respectively. The washout curve of this process can be estimated using Eq. 9, with $p = 1$, $q = 0$, and $a_1 = 0.3845$. 95% of an instantaneous pulse of tracer would be expected to wash out over 12.5 h. Accordingly, it may be said with 95% confidence that two samples separated by at least 12.5 h would be statistically independent.

Conclusions

A novel method of determining a continuous process' residence time distribution (RTD) has been developed which does not risk production of off-spec material. Through analysis of serial component analytical data, the process' washout curve can be inferred. Theoretical justification for the method highlights the need to choose an analyzed component whose yield is invariant over the experimental campaign. Limitations with respect to control of the analyzed component were also outlined. Applicability to well-mixed, laminar flow, and convective-dispersive systems was shown. Experimental comparison of pilot-plant pulse experiments and calibrated level measurements to the current method suggests that a ca. 5% error in mean residence time determination is possible. This error would be expected to increase for systems where feedback control has introduced bias into the parameter estimates or where the analyzed component's yield is not invariant over the course of the time-series.

Acknowledgments

The author appreciates the assistance of Dr. Valerie Adams, Mr. William Murdoch, and Mr. Kenneth Roberts in executing the pilot-plant pulse test and tracer chemical analyses, as well as that of Dr. Joseph Knight for residence time estimates from the pilot-plant calibrated vessel level indicator.

Literature Cited

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed., Prentice Hall, Englewood Cliffs, NJ (1994).

Box, G. E. P., and J. F. MacGregor, "The Analysis of Closed-Loop Dynamic-Stochastic Systems," *Technometrics*, **16**, 391 (1974).

Clements, Jr., W. C., and K. B. Schnelle, Jr., "Pulse Testing for Dynamic Analysis," *I&EC Proc. Des. Dev.*, **2**, 94 (1963).

Danckwerts, P. V., "Continuous Flow Systems: Distribution of Residence Times," *Chem. Eng. Sci.*, **2**, 1 (1953).

Eker, S. A., and M. Nikolaou, "Model Predictive Control and Identification: A New Approach to Closed-Loop Identification and Adaptive Control," AIChE Meeting, Miami (1998).

Fan, L. T., M. S. K. Chen, Y. K. Ahn, and C. Y. Wen, "Mixing Models with Varying Stage Size," *Can. J. Chem. Eng.*, **47**, 141 (1969).

Gay, D. H., "Identification and Control of Distributed Parameter Systems by Means of the Singular Value Decomposition," PhD Thesis, Univ. of Wisconsin, Madison (1989).

Genceli, H., and M. Nikolaou, "New Approach to Constrained Predictive Control with Simultaneous Model Identification," *AIChE J.*, **42**, 2857 (1996).

Honig, G., and U. Hirdes, "A Method for the Numerical Inversion of Laplace Transforms," *J. Comput. Appl. Math.*, **10**, 113 (1984).

Kothare, M. V., R. Shinnar, I. Rinard, and M. Morari, "Controlling Complex Chemical Processes using Partial Control Techniques," AIChE Meeting, Miami (1998).

Ljung, L., *System Identification: Theory for the User*, 2nd ed., Prentice Hall, Upper Saddle River, NJ (1999).

MacGregor, J. F., and D. T. Fogal, "Closed-Loop Identification: the Role of the Noise Model and Prefilters," *J. Proc. Control*, **5**, 163 (1995).

Nauman, E. B., "Residence Time Distributions and Micromixing," *Chem. Eng. Comm.*, **8**, 53 (1981).

Ogunnaike, B. A., and W. H. Ray, *Process Dynamics, Modeling, and Control*, Oxford University Press, New York (1994).

Ottino, J. M., and R. Chella, "Laminar Mixing of Polymeric Liquids: A Brief Review and Recent Theoretical Developments," *Polym. Eng. Sci.*, **23**, 357 (1983).

Seidel-Morgenstern, A., "Analysis of Boundary Conditions in the Axial Dispersion Model by Application of Numerical Laplace Inversion," *Chem. Eng. Sci.*, **46**, 2567 (1991).

Stakgold, I., *Green's Functions and Boundary Value Problems*, Wiley-Interscience, New York (1979).

Therrien, C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall, Englewood Cliffs, NJ (1992).

Václavek, V., "Stochastic Linear Operator as a Characteristic of a Continuous Flow System," *Chem. Eng. Sci.*, **22**, 1209 (1967).

Vuthandam, P., and M. Nikolaou, "Constrained MPCI: A Weak Persistent Excitation Approach," *AIChE J.*, **43**, 2279 (1997).

Wen, C. Y., and L. T. Fan, *Models for Flow Systems and Chemical Reactors*, Marcel Dekker, New York (1975).

Manuscript received Jan. 25, 1999, and revision received Jan. 27, 2000.