# Wavelet-Based Density Estimation and Application to Process Monitoring

**A. A. Safavi, J. Chen, and J. A. Romagnoli**
ICI Process Systems Engineering Laboratory, Dept. of Chemical Engineering, The University of Sydney, NSW 2006, Australia

*An application of wavelets and multiresolution analysis to density estimation and process monitoring is presented. Wavelet-based density-estimation techniques are developed as an alternative and superior method to other common density-estimation techniques. Also shown is the effectiveness of wavelet estimators when the observations are dependent. The resulting density estimators are then used in defining a normal operating region for the process under study so that any abnormal behavior by the process can be monitored. Results of applying these techniques to a typical multivariate chemical process are also presented.*

## Introduction

The on-line monitoring and diagnosis of the process operating performance are extremely important parts of the strategies aimed at improving a process and the quality of its products in the long term. This can be achieved using the so-called multivariate statistical approaches. Several successful implementations of these approaches in chemical engineering were recently reported (see Kourti et al., 1994; Nomikos et al., 1994; Piovoso et al., 1994). The basic philosophy of multivariate statistical approaches is that the behavior of the process is characterized using data obtained when the process is operating well or in a "normal region"(in a state of control). Subsequently, future unusual events and possible abnormalities in the new data are detected by referencing the measured process behavior against the available "in-control" model. In other words, the normal region is used here as a calibration criterion of the process.

To employ such statistical approaches, an appropriate description of the normal operating conditions of the process is needed. Unfortunately, the usually available database of the process, consisting of discrete samples, can only provide a very gross impression of the normal operating conditions, and thus is not a convenient tool for such purposes. A basic tool is to find the probability density function (PDF) or simply the density function of the variables, using the available data. Then the definition of a normal region is straightforward, as shown in this article.

Correspondence concerning this article should be addressed to J. A. Romagnoli.

Several methods for estimating the density function of a set of variables are presented in the literature. Among these methods are histograms, orthogonal estimators, kernel estimators, and elliptical basis function (EBF) estimators (see Silverman, 1986; Scott, 1992; Johnston and Kramer, 1994; Chen et al., 1995). Recently, with the introduction of wavelets and multiresolution analysis (MRA) as a new and powerful mathematical medium, researchers in many branches of science and engineering have been attracted to this new tool. Not surprisingly, statisticians have also employed wavelets for many purposes, particularly density estimation. In the last few years, there has been a great deal of research and progress in the area of wavelet estimators, because they have shown an exceptional flexibility in density estimation. While almost all the usual estimators fail in dealing with dependent observations, wavelet estimators have opened new doors to the problem of density estimation for dependent observations. We introduce the method of wavelet density estimations in chemical engineering and show how the resulting estimators can be used for process monitoring.

The rest of this article is structured as follows. The following section briefly reviews the problem of density estimation and the most common density-estimation techniques. The concepts of wavelets and multiresolution analysis together with wavelet estimators are introduced in the third section. Since most readers may not be quite familiar with these concepts, this section presents some details as well as remarks on the implementation aspects. The fourth section provides a general comparison of wavelet estimators with the other pre-

viously mentioned estimators. The problem of wavelet estimation for dependent observations is briefly discussed in the fifth section, and some practical tips are given in the sixth section. The seventh section presents two case studies. The first one is a simple one-dimensional density-estimation problem that is solved using the wavelet estimators. In the second case study, a wavelet estimator is developed for a multivariable chemical engineering flow sheet. Then a typical process-monitoring investigation is performed with the aid of the developed wavelet estimator.

## Density Estimation

Probability density estimation, or simply density estimation, is a fundamental concept in statistics. The density function $f$ of any random quantity $x$ gives a natural description of the distribution of $x$. The probabilities (P) associated with $x$ can be found as follows:

$$\mathrm{P}(a < x < b) = \int_a^b f(x)\,dx \qquad \text{for all } a < b.$$

For instance, the well-known Gaussian or normal distribution has the following density function:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{1}$$

where $\mu$ and $\sigma$ are the mean and the variance of the data set, respectively.

In practice, we have a set of observed data points that are assumed to be samples from an unknown probability density function. *Density estimation* is the construction of an estimate of the density function from the observed data. Techniques to estimate the density function of a given set are classified into two categories, namely parametric approaches and nonparametric approaches. The former assumes that the data belong to one of the known families of the distributions, for example, the normal distribution. The remaining task is to estimate parameters $\mu$ and $\sigma$. Unfortunately, one cannot often assume a parametric form for the density function. Therefore, one has to estimate the density function using an unstructured approach. This is called nonparametric density estimation. Density estimation has been widely used in a variety of fields of science and engineering, and is still an active area of research (see, for example, Chen et al., 1996; Scott, 1992).

A brief description of some well-known nonparametric density estimation techniques are presented in this section. Then a wavelet-based density-estimation technique will be explained in the next section.

### *Histograms*

The simplest and oldest used density estimator is the histogram. Assume a sample of $n$ real observations $x_1, x_2, \ldots, x_n$ whose underlying density, $f(x)$, is to be estimated. Considering an origin, $x_0$, and a bin width, $h$, the bins of the histogram are considered the intervals $[x_0 + kh,\ x_0 + (k+1)h]$ for $k \in Z$. The histogram is then defined as

$$\hat{f}(x) = \frac{1}{nh}\,(\text{no. of } x_i \text{ in the same bin as } x). \tag{2}$$

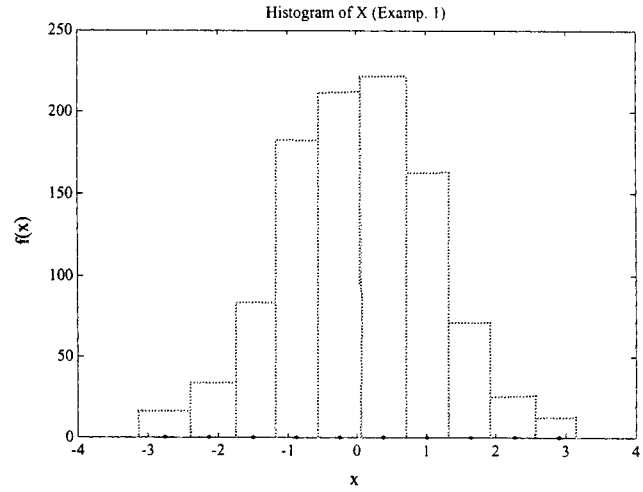


**Figure 1. Data set *X* (Example 1) taken from a normal distribution.**

Note that to construct the histogram, we have to choose both an origin and a bin width. The choice of these two parameters controls the amount of smoothing inherent in the procedure. Figure 1 gives an example of a histogram of a set of data taken from a normal (i.e., Gaussian) distribution.

Despite the simplicity of histograms, their discontinuous representation of the density function causes extreme difficulty if the derivatives of the estimate or a smooth representation of the estimate is required for some applications, as in cluster analysis and so on. Therefore, it is necessary to look for more sophisticated density-estimation methods that provide a more desirable description of the density function.

### *Kernel estimator*

The most widely studied and used density estimator is the kernel estimator. The usual kernel estimator is defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{3}$$

with the kernel function $K$ satisfying,

$$\int_{-\infty}^{\infty} K(x)\,dx = 1. \tag{4}$$

In Eq. 3, the parameter $h$ is called the window width, the bandwidth, or the smoothing parameter. The kernel function $K$ will sometimes be a symmetric probability density function (i.e., the normal density, for example). Figure 2 shows an example of constructing the density estimate from the kernel functions. Various properties of the usual kernel method as well as its extensions, such as the variable kernel method and the adaptive kernel method, which both deal with the choice

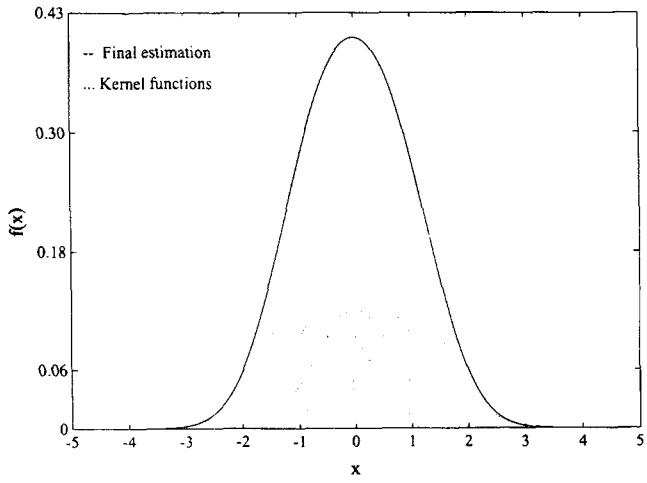


**Figure 2. Construction of the kernel estimator.**

of the window width or the smoothing parameter, are thoroughly studied in the literature, see, for example, Silverman (1986) and Chen et al. (1995).

*Orthogonal Basis Estimators.* The idea of using an orthogonal basis in estimating a density function was first considered by Cencov (1962). The orthogonal basis estimators approach the density-estimation problem from quite a different point of view, though some may find some similarities between this approach and the kernel method. The idea here is to estimate $f$ by estimating the coefficients of its orthogonal projections on a basis. In fact, if $H$ is a Hilbert space and $\{\theta_k\}_{k=-\infty}^{\infty}$ is an orthonormal basis in this space, then for any $f \in H$ we have

$$f(x) = \sum_{k=-\infty}^{\infty} c_k \theta_k(x), \tag{5}$$

where $c_k$ are the orthogonal projections of the function $f(x)$ onto the basis functions, which is given by the following inner product:

$$c_k = \langle f(x), \theta_k(x) \rangle. \tag{6}$$

Here, the convergence is in the norm (Masry, 1994). If $x_1, \ldots, x_n$ are again independent and identically distributed (IID) random variables with a common density function $f \in H$, then an estimate of $f$ is given by

$$\hat{f}(x) = \sum_{k=0}^{\infty} \hat{c}_k \theta_k(x), \tag{7}$$

with the approximated coefficients $\hat{c}_k$ estimated as

$$\hat{c}_k = \mathrm{E}[\theta_k(x_i)] = \frac{1}{n} \sum_{i=0}^{n} \theta_k(x_i), \tag{8}$$

where E stands for the expectation. As $n \to \infty$, $\hat{f} \to f$ at an appropriate rate.

In general, the sum $\sum_{k=0}^{\infty} \hat{c}_k \theta_k$ will not be a proper estimate of $f$, but will converge to a sum of delta functions at the observations. This is not a desirable phenomenon. In order to obtain a useful estimate of the density $f$, it is necessary to smooth the estimation by a low-pass filtering method. The easiest way is to truncate the expansion at some point. For instance, one may choose an integer $K$ and define the estimate $\hat{f}$ as below:

$$\hat{f}(x) = \sum_{k=0}^{K} \hat{c}_k \theta_k(x). \tag{9}$$

The choice of the cutoff point $K$ determines the amount of smoothing.

## Wavelets and Density Estimation

The use of orthonormal bases to estimate the density function was introduced previously. Recently, with the introduction of wavelet orthonormal bases (Meyer, 1985; Mallat, 1989; Daubecies, 1992), increasing attention has been given to the use of these new bases for the estimation of the density function, for example, Donoho (1993) or Kerkyacharian et al. (1992). First we briefly explain the recently developed wavelet orthonormal bases and their related multiresolution analysis (MRA), after which we describe the wavelet estimators.

### *Wavelets and multiresolution analysis*

Wavelets are a new family of localized basis functions that have found many applications in quite a large area of science and engineering (Safavi, 1996; Bakshi and Stephanopoulos, 1993). These basis functions can be used to express and approximate other functions. They are functions with a combination of powerful features, such as orthonormality, locality in time and frequency domains, different degrees of smoothness, fast implementations, and in some cases compact support. Wavelets are usually introduced in the multiresolution framework developed by Mallat (1989).

Consider a function $f(x) \in L^2(R)$, where $L^2(R)$ denotes the vector space of all measurable, square integrable one-dimensional functions. Further, we assume $V_m$ is the vector space containing all possible approximations of $f(x)$ at resolution $m$. Then the ladder of spaces $V_m$, $m \in Z$ representing the successive resolution levels for $f(x)$. The properties of these spaces are as follows:

1. $\cdots \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots$ (Figure 3);
2. $\bigcap_{m \in Z} V_m = \{0\}$, $\overline{\bigcup_{m \in Z} V_m} = L^2(R)$ (Figure 3); this says that the intersection of all these spaces is {0} and their union is $L^2(R)$ (Figure 3);
3. $f(x) \in V_m \leftrightarrow f(2^m x) \in V_0$, which means all the spaces are *scaled* versions of one space;
4. $f(x) \in V_0 \to f(x-k) \in V_0$ for all $k \in Z$, which means that if $f(x) \in V_0$, then its translated version $f(x-k) \in V_0$ does as well;
5. $\exists \phi \in V_0$ so that the $\phi_{0,k}(x) = \phi(x-k)$ constitute an orthonormal basis for $V_0$.

The function $\phi$ is called a *scaling function* of the multiresolution analysis (see the following subsection), and is expressed as

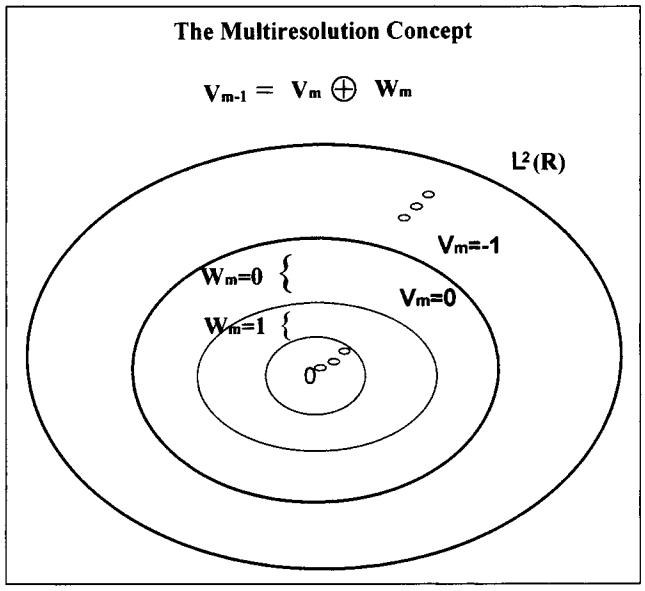


**Figure 3. Ladder of spaces $V_m$ representing a multiresolution analysis.**

$$\phi_{m,k}(x) = 2^{-m/2}\phi(2^{-m}x - k) \qquad m, k \in Z. \tag{10}$$

Let $W_m$ be the orthogonal complement of $V_m$ to $V_{m-1}$, which is $V_m \oplus W_m = V_{m-1}$. Then the orthonormal basis functions corresponding to $W_m$ (which are named wavelets denoted by $\psi_{m,k}$) can be easily obtained from $\phi_{m,k}$ (see the following) and represented as

$$\psi_{m,k}(x) = 2^{-m/2}\psi(2^{-m}x - k) \qquad m, k \in Z. \tag{11}$$

Next, $L^2(R)$ can be expressed as

$$\begin{aligned} L^2(R) &= \bigcup_{m \in Z} V_m \\ &= \cdots W_{-1} \oplus W_0 \oplus W_1 \cdots \\ &= \bigoplus_{m \in Z} W_m. \end{aligned} \tag{12}$$

Here, $W_j \perp W_m$ for $j \neq m$. We can interpret Eq. 12 as

$$f(x) = \sum_m \sum_k d_{m,k}\psi_{m,k}(x) \qquad m, k \in Z \tag{13}$$

or if we start from the approximation of the function at (for example) resolution $m = 0$, then

$$f(x) = f_0(x) + \sum_{m=-\infty}^{0} \sum_{k=-\infty}^{\infty} d_{m,k}\psi_{m,k}(x), \tag{14}$$

where

$$f_0(x) = \sum_k a_{0,k}\phi_{0,k}(x). \tag{15}$$

The simplest example of an orthonormal wavelet is the Haar function $\psi^H$, defined as

$$\psi^H(x) = \begin{cases} 1 & \text{for } 0 \le x < \dfrac{1}{2} \\ -1 & \text{for } \dfrac{1}{2} \le x < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

where its corresponding scaling function is

$$\phi^H(x) = \begin{cases} 1 & \text{for } 0 \le x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Figure 4 shows typical examples of the wavelets and their corresponding scaling functions.

*Constructing Multidimensional Wavelets.* The one-dimensional MRA developed in the previous subsection can be readily extended to higher dimensions (see Appendix A for details). Here we only show how multidimensional wavelets and scaling functions are constructed from their one-dimensional counterparts.

Consider the one-dimensional wavelets and scaling functions as below:

$$\phi_{m,k}(x) = 2^{-m/2}\phi(2^{-m}x - k) \qquad m, k \in Z \tag{18}$$

$$\psi_{m,k}(x) = 2^{-m/2}\psi(2^{-m}x - k). \tag{19}$$

The $n$-dimensional wavelets and scaling functions can then be constructed as below:

$$\begin{aligned} \Phi_{m,k}(x_1, x_2, \ldots, x_n) &= \phi_{m,k_1}(x_1)\phi_{m,k_2}(x_2)\cdots\phi_{m,k_n}(x_n) \\ \Psi^1_{m,k}(x_1, x_2, \ldots, x_n) &= \phi_{m,k_1}(x_1)\psi_{m,k_2}(x_2)\cdots\psi_{m,k_n}(x_n) \\ \Psi^2_{m,k}(x_1, x_2, \ldots, x_n) &= \psi_{m,k_1}(x_1)\phi_{m,k_2}(x_2)\cdots\psi_{m,k_n}(x_n) \\ &\vdots \\ \Psi^{2^n-1}_{m,k}(x_1, x_2, \ldots, x_n) &= \psi_{m,k_1}(x_1)\psi_{m,k_2}(x_2)\cdots\psi_{m,k_n}(x_n). \end{aligned} \tag{20}$$

It can be seen that an $n$-dimensional scaling function (i.e., $\Phi_{m,k}$) is a single $n$-dimensional function formed from a combination of one-dimensional scaling functions. The $n$-dimensional wavelet is formed in the same way, but it consists of $2^n - 1$ $n$-dimensional functions. Therefore the complexity of higher dimensional wavelets increases exponentially. The $n$-dimensional wavelets just defined are called *separable wavelets*. There is yet another approach in developing multidimensional wavelets, which leads to the so-called *nonseparable wavelets*. Detailed discussions on multidimensional wavelets are presented in Safavi (1996) and Daubechies (1992).

*Implementing Wavelet Algorithms.* Having gained some understanding of the wavelets and their properties, a basic question is how to find or build wavelets. Unfortunately, despite the recent existence of a large body of literature on

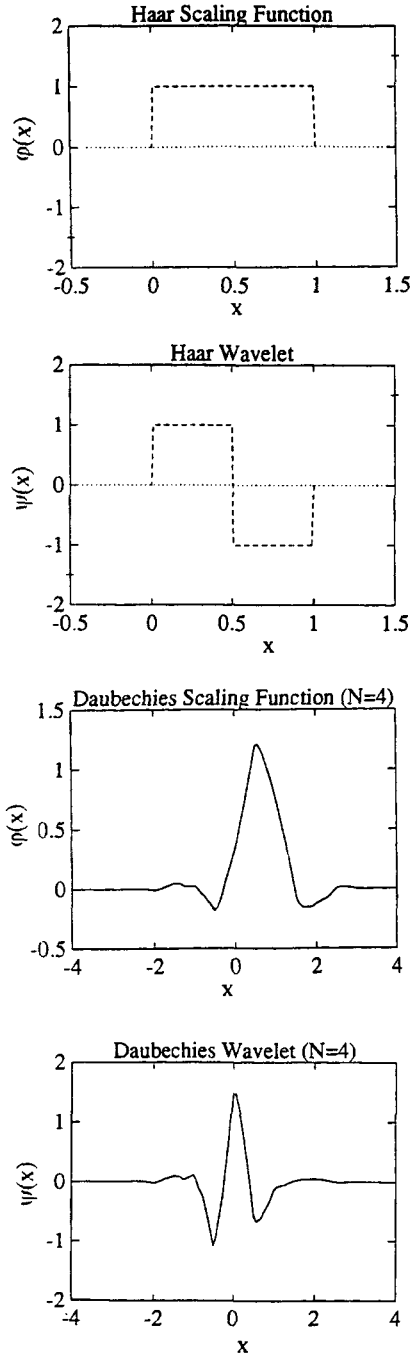


**Figure 4. Typical examples of wavelets and their corresponding scaling functions.**

wavelets, not many of them clearly address this issue. For a detailed discussion on this matter, refer to Safavi (1996) and Strang (1989). We present a brief discussion here.

Assume that $\phi(x) = \phi_{0,0}(x)$ is a basis function of the approximation space $V_0$ (see Figure 3). Further, the approximation space $V_0$ is embedded in $V_{-1}$. That is to say $\phi(x) \in V_0 \subset V_{-1}$. Thus, one can write

$$\phi(x) = \sum_k c_k \phi(2x - k), \qquad k \in Z. \tag{21}$$

In general, $c_k$ depend on the choice of the function $\phi$. For instance, if we choose $\phi$ to be the indicator function of the interval [0, 1]—we have already called it the Haar scaling function $\phi^H$—then we can write

$$\phi^H(x) = \phi^H(2x) + \phi^H(2x - 1). \tag{22}$$

It is common to look for a solution normalized by $\int \phi(x)\, dx = 1$. One should note that several conditions can be imposed on $c_k$ to obtain $\phi$'s with certain desirable properties. With the scaling function $\phi(x)$ expressed as in Eq. 21, the wavelet $\psi(x)$, as a basis function perpendicular to the scaling function, is simply derived as

$$\psi(x) = \sum_k (-1)^k c_{1-k} \phi(2x - k). \tag{23}$$

For instance, if one implies Eq. 23 on the $\phi^H$ defined in Eq. 22 with $c_0 = 1$ and $c_1 = 1$, then the resulting wavelet is

$$\psi^H(x) = \phi^H(2x) - \phi^H(2x - 1),$$

which is the Haar wavelet, as described earlier in this section.

The coefficients $c_k$ can completely describe a wavelet and a scaling function. These coefficients are available in the literature (e.g., Daubechies, 1992; Cohen et al., 1993). Therefore, in most cases, one does not need to create a new wavelet or try to find such coefficients. On the other hand, in applications such as signal decomposition, the whole analysis can be done using only these coefficients and Mallat's fast computing algorithm (Mallat, 1989). However, one may also need to construct the wavelet (or the scaling function) itself for some studies, as we employ them in this article. Three methods for constructing wavelets and scaling functions from coefficients $c_k$ are proposed by Strang (1989). One of these methods is a convenient approach based on the concept of *eigenvectors* and is explained in Appendix B.

*Wavelets and the Problem of Finite Length.* The MRA described here is actually based on an infinite set of data (i.e., it is defined on the whole real line). But in practice, one has to work on a finite set of data (i.e., only a finite interval of the real line). Therefore, we need to deal with the boundary problems. Cohen et al. (1993) present three different approaches to overcome this problem. The first two methods are based on the extension of the original function. For instance, if the original function $F(x)$ is assumed to be defined on the interval [0, 1], then one may assume $F(x) = 0$ for the outside of interval [0, 1]. A second option is that one can assume that the function $F$ is periodic by extending the function with periodic replicas beyond each endpoint. These methods are simple to implement but have some disadvantages, as is explained in Safavi (1996) and Cohen et al. (1993).

A more natural approach is to consider wavelets that are defined on [0, 1] (in fact, their corresponding MRA is also defined on [0, 1]). Such wavelets are presented in the literature, for example, Cohen et al. (1993). In general, however, these wavelets are not just simple dilations and translations of a single mother wavelet. For each mother wavelet, they consist of "left-edge," "right-edge," and "interior" wavelets (the "interior" wavelets are the usual wavelets). Figure 5

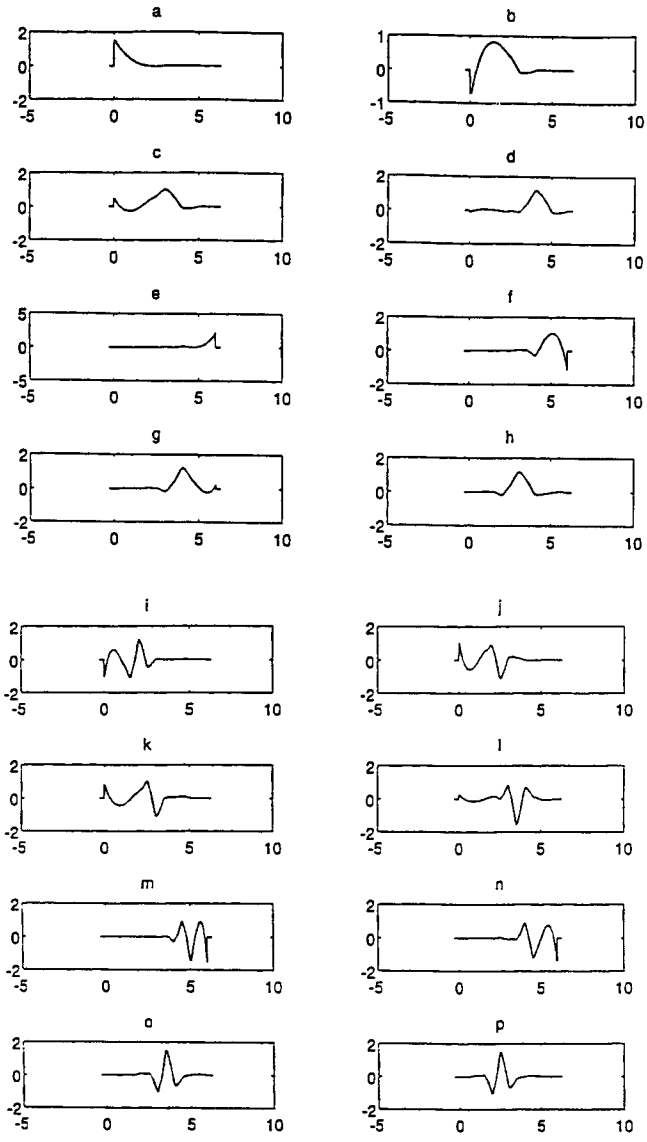


**Figure 5. Wavelets on the intervals: "edge" scaling functions and wavelets corresponding to Daubechies $N = 4$ wavelet.**

(a)–(d) $\phi_i^{left}$, $i = 0, 1, 2, 3$; (e)–(h) $\phi_i^{right}$, $i = 0, 1, 2, 3$;(i)–(l) $\psi_i^{left}$, $i = 0, 1, 2, 3$; (m)–(p) $\psi_i^{right}$, $i = 0, 1, 2, 3$; for more details, see Cohen et al. (1993) or Safavi (1996).

shows examples of the left-edge and right-edge wavelets and scaling functions for the Daubechies ($N = 4$) wavelet and scaling function shown in Figure 4. The number of left-edge and right-edge wavelets required (as well as the number of required scaling functions) depends on the support of the mother wavelet. It should be noted that there is also a restriction imposed on using these wavelets. The usual assumption for these wavelets is that no interior wavelet should touch both ends, that is, it should not touch both $x = 0$ and $x = 1$. Therefore, to use these wavelets, one must not work at the coarsest resolutions where their wavelets touch both ends.

Although this approach may increase the complexity of the wavelet analysis to some extent, it is a more natural way to work with data on the interval [0, 1]. It is worth noting that the Haar wavelet is already a wavelet defined on the interval.

### *Wavelet-based density estimators*

The localized nature of wavelets together with their orthonormality and multiresolution framework make these basis functions very attractive for many applications as well as density estimation. In the following, we explain the use of wavelets for density estimation.

Let $\{x_i\}_{i=0}^{n-1}$ be a stationary process and $f$ the density function of $x$. Then, invoking wavelets and multiresolution analysis, the approximation of $f$ at resolution $m$ is the orthogonal projection of $f$ on $V_m$, which is $f_m$,

$$f_m(x) = \sum_{k=-\infty}^{\infty} a_{m,k}\,\phi_{m,k}(x), \tag{24}$$

or, in general, $f(x)$ is expressed as (see previous arguments)

$$f(x) = \sum_k a_{m,k}\,\phi_{m,k}(x) + \sum_{l=-\infty}^{m} \sum_k d_{l,k}\,\psi_{l,k}(x), \qquad k \in Z, \tag{25}$$

with

$$a_{m,k} = \langle \phi_{m,k}, f \rangle,$$
$$d_{l,k} = \langle \psi_{l,k}, f \rangle. \tag{26}$$

Since $f$ is not available, the unbiased estimation of each coefficient in Eq. 26 can be computed as (see the subsection on orthogonal basis estimators)

$$\hat{a}_{m,k} = \frac{1}{n} \sum_{i=0}^{n-1} \phi_{m,k}(X_i),$$
$$\hat{d}_{l,k} = \frac{1}{n} \sum_{i=0}^{n-1} \psi_{l,k}(X_i). \tag{27}$$

Now, a simple approach to estimate $f$ is to ignore the wavelet terms in Eq. 25, focusing attention on the scaling functions at a particular resolution $m$ only, which yields,

$$\hat{f}(x) = \sum_k \hat{a}_{m,k}\,\phi_{m,k}(x). \tag{28}$$

This is almost identical to the kernel estimator. The performance of this estimator can often be enhanced by incorporating the wavelet terms from Eq. 25. It is, however, important not to include all the wavelet terms of Eq. 25. Otherwise, the resulting estimation may not be reasonably smooth and useful. In order to achieve an estimation with the desirable degree of smoothness, several methods are proposed in the literature (Kerkyacharian and Picard, 1992; Hall and Patil, 1993; Pinheiro and Vidakovic, 1995). The simplest is to express the estimation as

$$\hat{f}(x) = \sum_k \hat{a}_{m,k}\,\phi_{m,k}(x) + \sum_{l=-K}^{m} \sum_k \hat{d}_{l,k}\,\psi_{l,k}(x), \tag{29}$$

where $K$ is the cutoff point based on the following measure:

$$\sup |\hat{f}(x+h)-\hat{f}(x)| < \alpha, \qquad \text{for all } h>0. \tag{30}$$

Here, $\alpha$ and $h$ are some constants. Two other techniques that are widely used in wavelet estimator literature are called *soft-thresholding* and *hard-thresholding*. The final estimation is generally expressed as

$$\hat{f}(x)=\sum_k \hat{a}_{m,k}\phi_{m,k}(x)+\sum_{l=-q}^{m}\sum_k \hat{d}_{l,k}w(\hat{d}_{l,k}/\delta)\psi_{l,k}(x), \tag{31}$$

where $q$ and $\delta$ are adjustable constants, and $w$ is a threshold function satisfying

$$w(u)=\begin{cases} 0 & \text{for } 0<u<c_1 \\ \in[0,1] & \text{for } c_1\le u\le c_2 \\ 1 & \text{for } u>c_2 \end{cases}$$

for constants $0<c_1<c_2<\infty$. For hard-thresholding, $w$ is

$$w(u)=\begin{cases} 0 & \text{for } 0<u<1 \\ 1 & \text{for } u>1, \end{cases},$$

and for soft-thresholding, we have

$$w(u)=\begin{cases} 0 & \text{for } 0<u<c_1 \\ (u-c_1)/c_2 & \text{for } c_1\le u\le c_2 \\ 1 & \text{for } u>c_2. \end{cases}$$

In this article we follow the measure, Eq. 30.

## Wavelet Estimators vs. Other Estimators

The nature of wavelets and multiresultion analysis makes the wavelet estimators superior to the other estimators in general. We present some comparisons here.

The greatest feature of orthogonal estimators is their ease of computation. However, since the usual orthogonal bases (Fourier basis, for example) consist of global basis functions, their corresponding estimators may not be easily adapted to local behavior of the underlying density function. Wavelet estimators not only provide all the beauty of typical orthogonal estimators but they also allow local learning and manipulation of the estimated density function. Further, with the availability of several families of orthogonal wavelets, a high degree of flexibility in terms of convergence rate and smoothness will be available with wavelet estimators.

The convergence rates of kernel estimators, or indeed their very construction in terms of the kernel choice and the bandwidth, are often dictated by the assumption of the smoothness of the density function to be estimated. This assumption denies the existence of a wide variety of practical problems where the function to be estimated may not be smooth in the classic sense. The functions may, for example, contain high-frequency oscillation, or discontinuities and invoke techniques that can be easily adapted to differing local levels of smoothness. Wavelet methods, however, enjoy exceptional potential for adaptive smoothing. They permit two different levels of smoothing, one globally, in terms of the frequency of the scaling functions, and the other locally, via the scale of the wavelet functions. Due to the possible mutual orthogonality of the wavelets and scaling functions, these two levels can be mutually independent. The global level, analogous to bandwidth choice for the kernel estimator, may be chosen so as to provide an amount of smoothing that is appropriate in an "average" sense for lower-frequency parts of the desired function. On the other hand, the local level provides a tuning method for a wide variety of adjustments and corrections in different places. Viewing the method in this way, one could say that wavelet methods may be seen not so much as an alternative to the kernel approach, but as a way of enhancing that technique. In fact, the "basic" wavelet estimator may be seen as a special case of the kernel estimators.

Another interesting feature of the wavelet density estimators is that the resulting approximations are provided at several resolutions. This allows the user to choose the degree of accuracy vs. the simplicity (or the speed) of estimations. Besides having such multiresolution representation of the estimation, one can also optimize the number of employed wavelets in a final representation. This is a direct result of the well-known phenomenon in the wavelet representation of data: most of the total energy is concentrated in a few wavelet coefficients. Therefore, ignoring those coefficients that make a small contribution to the total representation can lead to a simpler representation. The change in the approximation error can then be worked out as follows (see Safavi, 1996; Mallat, 1989):

$$\text{error}^2_{\text{new}}-\text{error}^2_{\text{old}}=\sum_{m,k}\hat{d}^2_{m,k}, \qquad m,k\in S_-, \tag{32}$$

where $\{\hat{d}_{m,k}\}_{m,k\in S_-}$ is the set of wavelet coefficients ignored in the representation. More complete discussions of the features of wavelet estimators are presented in Pinheiro and Vidakovic (1995), Hall and Patil (1996), and Nason (1996).

## Density Estimation for Dependent Observations

In the previous discussion, we assumed that for all the estimators there is no dependency between the data. That is to say, the observations are independent and identically distributed (IID). Though such observations may sometimes be produced, this isn't always the case. In practice, the existence of some degree of dependency between the observed data is quite possible. It is therefore important to know how flexible the density estimation technique is when one deals with dependent observations. Since the dependence of data can be of several types, here we more precisely focus on the *stationary* observations that are more common in practice. This means while the data are correlated, this correlation depends on the lag between the samples, and the mean $\mu=E(x_i)$ (Beran, 1994).

While other estimators fail to deal with dependent observations, wavelet estimators challenge dependent observations in several ways. *Their exceptional flexibility in generating smooth*

*estimations together with their capability in local manipulation of the data allows a better treatment of dependent observations. Several studies (Tewfik and Kim, 1992; Wang, 1995) show that the auto and cross-correlation functions of wavelet coefficients decay at a much faster rate than do the original data.* Therefore, wavelet estimators could in general decorrelate the data to some extent. Nevertheless, several data-selection and cross-validation methods for wavelet estimators are introduced significantly in the literature (Nason, 1996; Wang, 1995). A relatively simple method of this kind is also presented in this article. Beside the preceding argument, Masry (1994) and Hall et al. (1994) detail formulations of the $L^2$ approximation error of the estimated density function. They then show that for wavelet estimators such measures could remain valid for dependent random variables, provided the tuning parameters associated with the estimators are chosen judiciously and allowing for some simple assumptions on using multiresolution analysis and the underlying density function $f$. Therefore, the advantages of wavelet estimators over other estimators can be seen even in the case of dependent observations.

### *Data-selection method for dependent observation*

A method based on data selection plus wavelet estimation is proposed here. This method is most appropriate when there is only short-range dependence among the data. For long-range dependence or long-memory data, the reader is referred to Wang (1995).

Consider a set of observations $\{x_i\}_{i=1}^{i=n}$ with a common density function $f$. To reduce dependence among the data, a subset of the original set is generated by selecting all the odd-indexed data. Next, wavelet estimation is performed for the original set, while using one of the wavelet threshold (i.e., the termination criterion) methods described in the earlier sections. The estimation will be

$$\hat{f}^1(x) = \sum_k \hat{a}^1_{m0,k}\,\phi_{m0,k}(x) + \sum_m \sum_k \hat{d}^1_{m,k}\,\psi_{m,k}(x), \quad (33)$$

where $m0$ is the starting resolution level. For the subset, the wavelet estimation proceeds until it is using the same number of basis functions as in the previous estimation. This led to

$$\hat{f}^2(x) = \sum_k \hat{a}^2_{m0,k}\,\phi_{m0,k}(x) + \sum_m \sum_k \hat{d}^2_{m,k}\,\psi_{m,k}(x).$$

The procedure can continue with further subsets. Here the final estimation is obtained by averaging the two previous ones, saying

$$\bar{f}(x) = \sum_k \bar{a}_{m0,k}\,\phi_{m0,k}(x) + \sum_m \sum_k \bar{d}_{m,k}\,\psi_{m,k}(x), \quad (34)$$

where each of the preceding coefficients is obtained as

$$\bar{a}_{m0,k} = \frac{1}{2}\sum_i \hat{a}^i_{m0,k}, \qquad \bar{d}_{m,k} = \frac{1}{2}\sum_j \hat{d}^j_{m,k}. \quad (35)$$

The effectiveness of this approach is demonstrated in the following section.

## Practical Tips

In developing wavelet estimators, the following remarks may be found helpful.

- Without loss of generality, any function defined on the real line can be scaled on the interval [0, 1]. This is very common in mathematical formulation. When making any wavelet-related analysis, and particularly when using wavelets adapted to intervals, it is most often easier to work on the interval [0, 1]. Therefore, we will work on this interval throughout this article. Of course, scaling all the results to the whole real line is a simple matter.
- Many standard density-estimation procedures may yield a density estimator with negative values. This could also happen for wavelet estimation. Although such a phenomenon is usually confined to the tails and sparsely represented regions of the density, reporting an estimator that takes negative values might be disturbing to the user. Pinheiro and Vidakovic (1995) suggest two methods to overcome such results. However, neither of those methods was used in this article.
- When one deals with a discrete analysis (i.e., a discrete calculation or discrete representation of a function), the accuracy of the analysis or representation could depend on the frequency of the sampling used. This is also the case when doing wavelet analysis. Thus, the caveat here is that a sufficient number of samples should be available to perform an accurate wavelet estimation. This sufficient number of samples could vary for the different wavelets used.
- As the dimensionality of a problem increases, the computational complexity of multidimensional wavelets increases greatly. However, this does not mean that the problem cannot be solved. Under such circumstances, ones may start the analysis from a relatively high resolution (i.e., $m \gg 0$), where the number of scaling functions is large enough to solve the problem. This results in a dramatic reduction in the complexity of the solution.

## Case Studies

Two case studies are presented in this section. In the first, a one-dimensional wavelet estimator is developed for a set of random samples of a Gaussian distribution. This is just to show the capability of wavelet estimators in approximating a well-known and standard distribution function. The second one presents a more detailed study of wavelet estimators for a typical multivariable chemical process. Then the use of the resulting density function for process monitoring is investigated. All the implementations were performed within the MATLAB environment.

## Example 1

A set $X$ of 1,000 random numbers belonging to a Gaussian distribution (i.e., here, mean 0.030 and variance 0.998) is generated using MATLAB. Figure 1 shows the histogram of $X$. The task is to approximate the density function of this data set. Three cases are studied for this purpose. First, using the existing set $X$, the density function is approximated. For demonstrative reasons, a Haar wavelet is initially employed in the estimation. Figure 6 shows the resulting approximations at various resolutions. As is observed from the figure, the Haar wavelet results are similar to the histogram, but in a

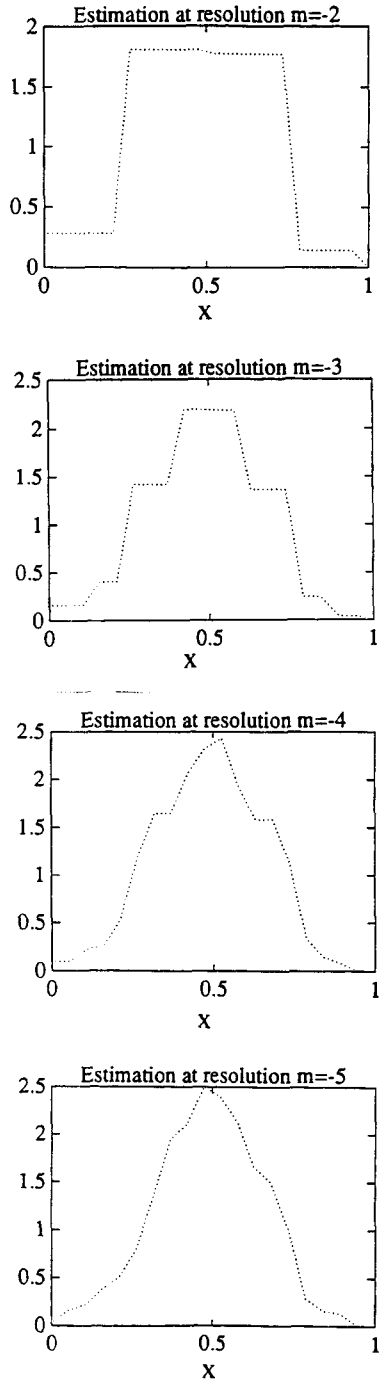


**Figure 6. Multiresolution estimation of the density function of the case study variables using Haar wavelets.**

multiresolution framework (i.e., like histograms with different bin widths). This is not a very useful representation and thus is not considered again in our studies. Employing a Daubechies ($N = 4$) wavelet adapted to the intervals (Figures 4 and 5) results in a reasonably smooth approximation, as shown in Figure 7. In the second case, a 5% random white noise, $\epsilon$, is added to the data set, which yields

$$X^{\text{static}} = X + \epsilon, \qquad X = \{x_i\}_{i=1}^{1,000}.$$

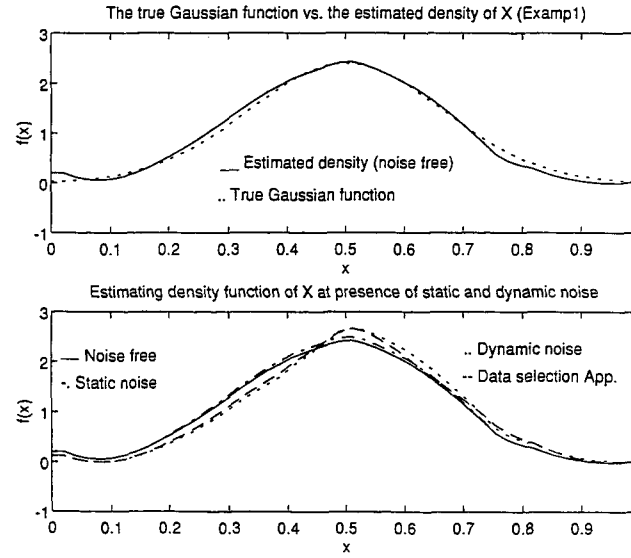


**Figure 7. Density estimations corresponding to Example 1.**

The new set $X^{\text{static}}$ is thus used to approximate the density function (Figure 7). In the third case, we assume that the noise has some dynamics, which is

$$\epsilon^{\text{dynamic}}(i) = 0.5\epsilon^{\text{dynamic}}(i-1) + 0.2\epsilon^{\text{dynamic}}(i-2) + e(i), \qquad i = 1, 2, \ldots, 1{,}000,$$

where $\{e\}$ is a set of white noise. The new dynamic set is then obtained as

$$X^{\text{dynamic}} = X + E^{\text{dynamic}}, \qquad E = \{\epsilon^{\text{dynamic}}(i)\}_{i=1}^{1,000}.$$

Now, density estimation is performed using the new dynamic set. The results of these estimations are given in Figure 7. This figure shows the capability of wavelet estimators for dealing with problems, as mentioned earlier.

### *Example 2*

*Process Description.* The case study process consists of two continuous stirred-tank reactors (CSTRs) in series with an intermediate mixer for the introduction of the second feed (Figure 8). A single, irreversible, exothermic, first-order reaction ($A \rightarrow B$) takes place in each reactor (Appendix C). A detailed description of this flow sheet can be found in Safavi (1996) and Hennin (1991). The independent variables in this system are input compositions ($C_F^1$ and $C^2$), two input flow rates ($Q_F^1$, $Q^2$), and input temperatures ($T_F^1$ and $T^2$). In our study, the compositions and temperatures of both feeds—$C_F^1 = C^2$ and $T_F^1 = T^2$—are kept similar. The outputs of this system are reactors' compositions ($C_1$ and $C_2$), reactors' temperatures ($T_1$ and $T_2$), and mixer's composition and temperature ($C_m$ and $T_m$).

Since the process conditions are going to be monitored in this study, all or some of the outputs of the process can be analyzed. Since some of the outputs may be correlated, a multivariate statistical technique (e.g., principal component analysis (PCA); partial least squares (PLS); method, etc.)

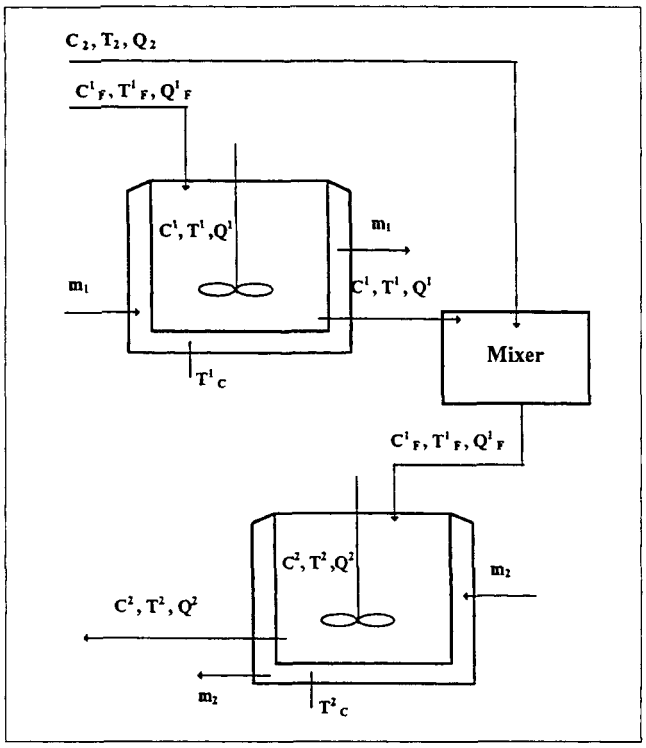


**Figure 8. Double CSTR's flow sheet.**

could be used to reduce the dimensionality of the system (Chen et al., 1996; Kourti and MacGregor, 1995; Nomikos and MacGreger, 1994). Chen et al. (1996) also proposes a robust PCA for statistical process monitoring. These methods can reduce the computational load while providing a reasonably acceptable representation of the system. In practice, two or three principal components are often sufficient to explain most of the predictable variations in a process with a large number of variables (Kourti and MacGregor, 1995). Here, the dimensionality of the system is reduced to two using the PLS technique, and finally we attain two (latent) variables $t_1$ and $t_2$, which are employed to analyze the process. It is worth noting, however, that we could have proceeded with the estimation of the density function without using the PCA techniques, by working with a larger number of process variables. In such a case, the process-monitoring results of the following section could not have been visualized, though they could still have been determined.

*Density Estimation.* For our analysis, the system inputs (i.e., $C_F^1$, $T_F^1$, $Q_F^1$, $Q^2$) are pseudorandomly changed within the 5% range of their operating point at normal conditions. Each step change is held longer than the process time constant until the process almost reaches its steady state value for that particular change. One thousand pseudorandom changes are applied to the inputs. Figure 9 shows the typical moves of some of the process outputs corresponding to the input variations.

First, 1,000 steady-state data points are collected around the process operating point. This set is considered to be *static* and *noise-free* observations. Using the wavelet estimation techniques, with the Daubechies ($N = 4$) wavelet (see Figures 4 and 5) adapted to the interval [0, 1], an approximation of the density function is obtained, as shown in Figure 10.

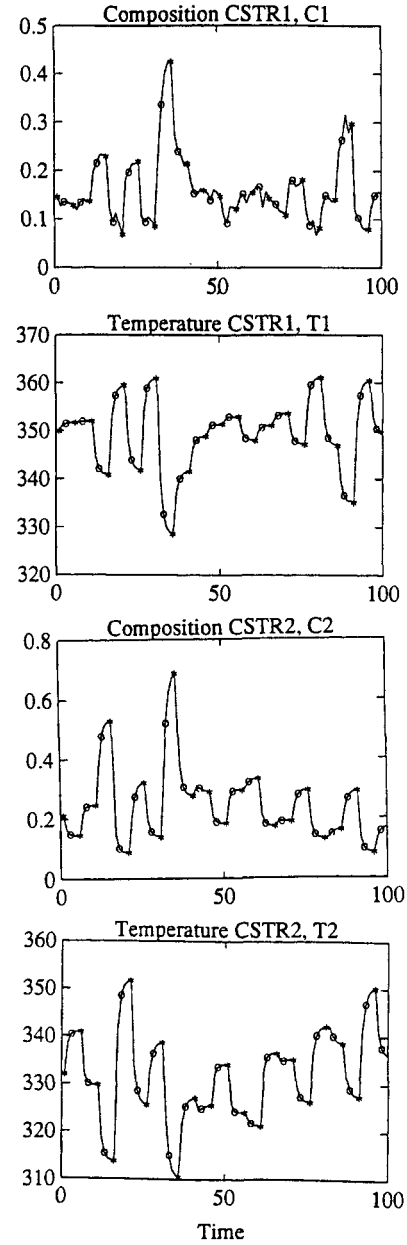


**Figure 9. Typical moves of process outputs in response to the variations of the inputs.**

*'s show the steady-state data and o's show the dynamic data selected for estimating the density function.

In practical situations, the collection of such noise-free and independent (or steady-state) observations may not always be possible. To create a more practical example, sample the process outputs between the steady-state points to generate a new data set from the simulation results. These points are shown in Figure 9. Since the process's dynamic behavior appears at these points, the collected observations should be considered dependent (i.e., or the data may be considered independent data plus dependent noise). The new data set is called the dynamic set. To approximate general practical observations, the original steady-state data set is mixed with the new dynamic set, and the resulting observations are used for

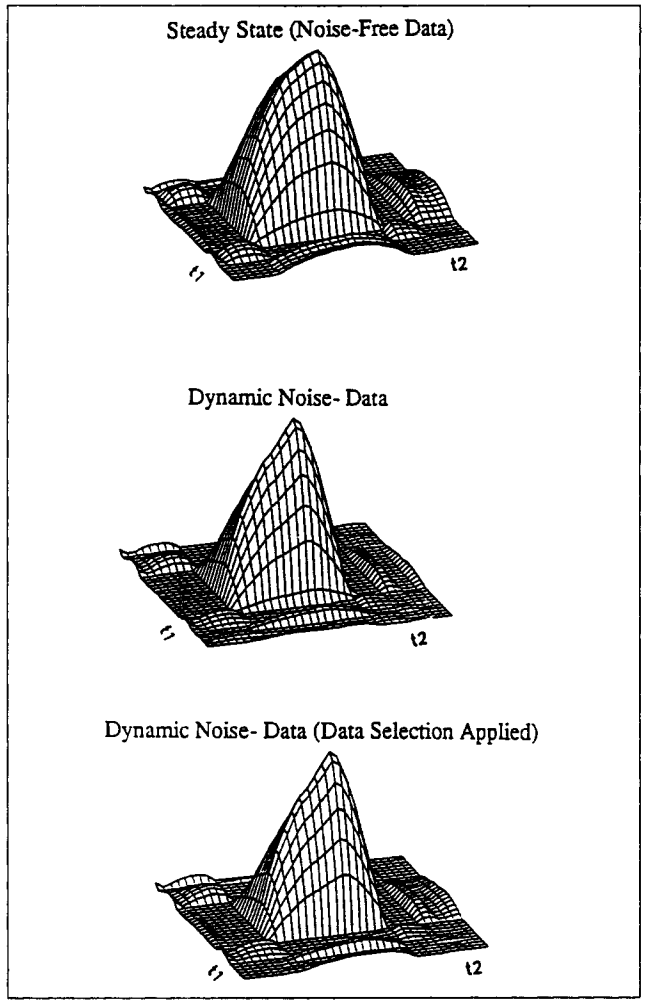


**Figure 10. Estimations of the density function of the case study process for the original noise-free observations and dependent observations, using Daubechies wavelet $N = 4$.**

density estimation. This new wavelet density estimation is also shown in Figure 10. As can be seen in the figure, this new density estimation is reasonably close to the original one.

The data-selection method described in the earlier section may also be used for a closer approximation to the original density function. In fact, we employ the data-selection method to reduce the dependence among the data before applying the wavelet methods. In this study, three data sets are considered, the mixed data set, every second sample of the mixed data set, and every third sample of the data set. The corresponding wavelet methods were then applied, resulting in the density function shown in Figure 10, which is seen to be a closer approximation to the original density function.

*Process Monitoring.* Since we obtained a smooth estimation of the density function (Figure 10) we can find a normal operating process region for a particular confidence level. Here we consider a 90% confidence level. If the number of variables (i.e., here latent variables) are two, then the normal region can be visualized via a contour in a two-dimensional plane. This is demonstrated in the following.

For our case of two latent variables, one must first decide a confidence level that will define a corresponding contour of the density function, assuming $f_n(z) = f_{constant} = f^*$ at this contour. This should represent our normal region. To attain process monitoring, we then need to calculate $\hat{f}(z^{new})$ for each newly available process data $z^{new}$ (or data pair). If $\hat{f}(z^{new}) \geq f^*$, it indicates that the new projected data are within the normal region, or, in fact, that the system is in its normal operating condition. Otherwise, the process conditions have been changed either as a result of a fault or a disturbance, and further actions should thus be taken. Figure 11 shows the normal region in our case, plus the moves of the process data toward the outside of normal region that was caused by a 20% increase (i.e., disturbance) in the normal value of the first feed flow rate $Q_F^1$. The direction of this data movement depends on the type and effect of the disturbance.

The preceding process-monitoring results can also be used as inputs to some process fault-detection algorithms to find more about the cause of such disturbances.

In this case study the final number of variables used to define the normal region was two, and sufficient detail for process monitoring was presented earlier. If only one variable is used, then the problem is even simpler and proceeds as before. If the number of variables (i.e., here latent variables) is more than two, the region can still be determined as explained below. In such a case, there are two options. The first option is to proceed with the earlier approach and determine the normal region; however, the region cannot be visualized as in the one- or two-dimensional case. For process monitoring, one again needs to calculate $\hat{f}(z^{new})$ for each newly available process data $z^{new}$ (or data pair) to see if $\hat{f}(z^{new}) \geq f^*$, which means that the new projected data are within the normal region. But the moves of process data toward or away from the normal region cannot be visualized. The second option is to split the process measurements into several groups and apply PLS to each group, leading to one or two latent variables for each group (see MacGregor and Jaeckle (1994) for details). The process monitoring as described in this case study can then proceed.

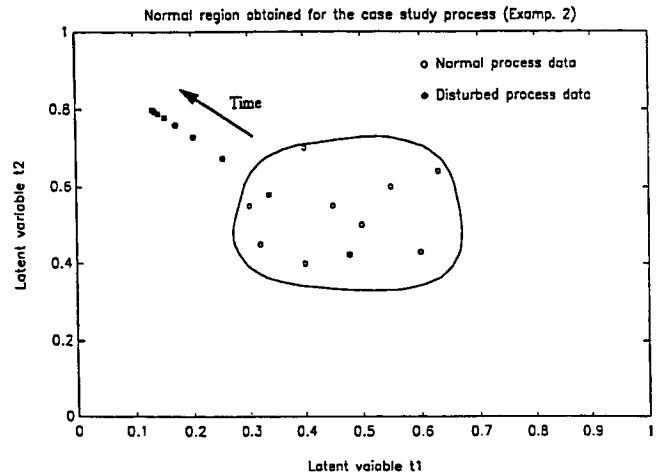


**Figure 11. Normal region for the case study process in Example 2.**

As is observed, the disturbed-process data move away from the normal region as time evolves.

## Concluding Remarks

In this article, a density-estimation technique based on wavelets and multiresolution analysis was developed. The technique was also used to estimate the density of a typical multidimensional chemical process, in the presence of dependent observations. The resulting density function was then used to define a normal operating region for the process so that any future abnormal changes in the process could be easily detected.

In the current study, observations with short-range dependence were considered, while a more detailed study of wavelet methods for observations with long-range dependence is under study and will be reported in the near future. In addition, a more rigorous formulation of wavelet estimators for dependent data could be a useful subject for further research.

## Notation

$Q_2$ = flow rate of feed stream 2, $m^3/s$
$C^1$ = concentration out of the first CSTR, $mol/m^3$
$T^1$ = temperature out of the first CSTR, K
$Q^1$ = flow rate out of the first CSTR, $m^3/s$
$COOL^1$ = output temperature cooling 1, K
$COOL^2$ = output temperature cooling 2, K
$m_1$ = cooling water flow for the first CSTR, $m^3/s$
$m_2$ = cooling water flow for the second CSTR, $m^3/s$
$C_p$ = heat capacity of the mixture
$DH$ = heat of the reaction lumped parameter
$\Delta H_R$ = heat of the reaction
$K_0$ = rate of the reaction coefficients
$V$ = volume of the tank
$\rho$ = density of the mixture
$UA$ = heat-transfer coefficient

### *Variables and constraints*

$K_0 = 2.7e0.8$
$DH1 = 5.0$
$C_F^1 = C_2 = 20$ $(mol/m^3)$
$COOL^2 = 275.0$
$T^1 < 350$
$COOL^2 < 20$
$C^2 < 0.3$
$E/R = 6{,}000$
$DH2 = 5.0$
$T_F^1 = T_2 = 300$ K
$m_1 = 0.35$ $(m^3/s)$
$T^2 < 350$
$Q_F^1 > .05$
$UA = 0.35$
$C_p = 1.0$
$COOL^1 = 300.0$
$m_2 = 0.8$ $(m^3/s)$
$COOL^1 < 30$
$Q_2 > 0.5$

## Literature Cited


Bakshi, B. R., and G. Stephanopoulos, "Wave-Net: A Multi-resolution, Hierarchical Neural Network with Localized Learning," *AIChE J.*, **39**(1), 57 (1993).

Beran, J., *Statistics for Long Memory Processes*, Chapman & Hall, London (1994).

Chen, J., J. A. Bandoni, and J. A. Romagnoli, "A Kernel Approach to Normal Region in Multivariate Statistical Process Monitoring for Gross Errors Detection and Process Fault Detection," AIChE Meeting, Miami, FL, p. 12 (1995).

Chen, J., J. A. Bandoni, and J. A. Romagnoli, "Robust Statistical Process Monitoring," *Comput. Chem. Eng.*, **20**, (Suppl. A), 497 (1996).

Cencov, N. N., "Evaluation of an Unknown Distribution Density from Observations," *Sov. Math.*, **3**, 1559 (1962).

Cohen, A., I. Daubechies, and P. Vial, "Wavelets on the Interval and Fast Wavelet Transform," *Appl. Comput. Harmonic Anal.*, **1**, 54 (1993).

Daubechies, I., *Ten Lectures on Wavelets*, SIAM, Philadelphia (1992).

Donoho, D. L., "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," *Proc. Symp. Applied Math.*, Preprint, p. 173 (1993).

Hall, P., and P. Patil, "On Wavelet Methods for Estimating Smooth Function," Res. Rep. No. CMA-SR12-93, CMA, Australian National University (1993).

Hall, P., and P. Patil, "On the Choice of Smoothing Parameter, Threshold, and Truncation in Nonparametric Regression by Non-Linear Wavelet Methods," *J. R. Stat. Soc. B*, **58**(2), 361 (1996).

Hall, P., P. Patil, and Y. K. Truong, "On Wavelet Methods for Curve Estimation in Time Series," Res. Rep. No. SRR-034-94, CMA, Australian National University, (1994).

Hennin, S. R., "Structural Decisions in On-Line Optimisation," First Quarterly Rep., Imperial College, London (1991).

Johnson, L. P. M., and M. A. Kramer, "Probability Density Function Using Elliptical Basis Functions," *AIChE J.*, **40**(10), 1639 (1994).

Kerkyacharian, G., and D. Picard, "Density Estimation in Bescov Spaces," *Stat. Probab. Lett.*, **13**, 15 (1992).

Kourti, T., and J. F. MacGregor, "Multivariate SPC Methods for Monitoring and Diagnosing of Process Performance," *Proc. PSE 94*, p. 739 (1994).

Kourti, T., and J. F. MacGregor, "Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods," *Chemometrics Intelligent Lab. Syst.*, **28**, 3 (1995).

MacGregor, J. F., and C. Jaeckle, "Process Monitoring and Diagnosis by Multiblock PLS Methods," *AIChE J.*, (1994).

Mallat, S. G., "A Theory for Multi-resolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Anal. Mach. Int.*, **PAMI-11**(7), 674 (1989).

Masry, E., "Probability Density Estimation from Dependent Observations Using Wavelets Orthonormal Bases," *Stat. Probab. Lett.*, **21**, 181 (1994).

Meyer, Y., "Principe d'Incertitude, Bases Hilbertiennes et Algebres d'Operateurs," *Semin. Bourbaki*, **662** (1985–1986).

Nason, G. P., "Wavelet Shrinkage Using Cross-validation," *J. R. Stat. Soc. B.*, **58**(2), 463 (1996).

Nomikos, P., and J. F. MacGreger, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**(8), 1994.

Pinheiro, A., and B. Vidakovic, "Estimating the Square Root of Density via Compactly Supported Wavelets," Duke Univ., Durham, NC, Preprints (1995).

Piovoso, M. J., and K. A. Kosanovich, "Application of Multivariate Statistical Methods to Process Monitoring and Controller Design," *Int. J. Control*, **59**(3), 743 (1994).

Safavi, A. A., "Wavelet Neural Networks and Multiresolution Analysis with Applications to Process Systems Engineering," PhD Thesis, Dept. of Chemical Engineering, The University of Sydney, Sydney, Australia (1996).

Scott, D. W., *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley, New York (1992).

Silverman, B. W., *Density Estimation Statistics and Data Analysis*, Chapman & Hall, London (1986).

Strang, G., "Wavelets and Dilation Equations: A Brief Introduction," *SIAM Rev.*, **31**(4), 614 (1989).

Tewfik, A. H., and M. Kim, "Correlation Structure of the Discrete Wavelet Coefficients of the Fractional Brownian Motion," *IEEE Trans. Inform. Theory*, **IT-38**, 904 (1992).

Wang, Y., "Function Estimation via Wavelet Shrinkage for Long-Memory Data," Preprint, (1995).


## Appendix A: Derivation of *N*-Dimensional MRA

Consider the 1-D wavelets and scaling functions

$$\phi_{m,k}(x) = 2^{-m/2}\phi(2^{-m}x - k) \qquad m, k \in Z \qquad \text{(A1)}$$

$$\psi_{m,k}(x) = 2^{-m/2}\psi(2^{-m}x - k) \qquad \text{(A2)}$$

and the (one-dimensional) multiresolution representation of a function $F(x) \in L^2(R)$,

$$F(x) = \sum_{k=-\infty}^{+\infty} a_{0,k}\phi_{0,k}(x) + \sum_{m=0}^{+\infty}\sum_{k=-\infty}^{+\infty} d_{m,k}\psi_{m,k}(x). \quad \text{(A3)}$$

Recall the approximation spaces $V_m$ and the detail spaces $W_m$ in a one-dimensional (orthogonal) multiresolution scheme, and their relation as

$$V_{m-1} = V_m \oplus W_m \qquad m \in Z, \quad \text{(A4)}$$

where the symbol $\oplus$ denotes the "orthogonal sum." In the two-dimensional case, one can define a set of vector spaces $\bar{V}_m$ for $L^2(R^2)$ as

$$\bar{V}_m = V_m \otimes V_m \qquad m \in Z, \quad \text{(A5)}$$

where $\otimes$ denotes the "tensor product," and the $\bar{V}_m$ satisfy a multiresolution ladder in $L^2(R^2)$,

$$\cdots \subset \bar{V}_2 \subset \bar{V}_1 \subset \bar{V}_0 \subset \bar{V}_{-1} \subset \cdots$$

$$\bigcap_{m \in Z} \bar{V}_m = \{0\}, \qquad \overline{\bigcup_{m \in Z} \bar{V}_m} = L^2(R^2).$$

The basis functions for $\bar{V}_m$ are defined as

$$\begin{aligned}\Phi_{m,k}(x_1, x_2) &= \phi_{m,k_1}(x_1)\phi_{m,k_2}(x_2) \\ &= 2^{-m}\Phi(2^{-m}x_1 - k_1, 2^{-m}x_2 - k_2) \\ &\qquad m, k_1, k_2 \in Z. \quad \text{(A6)}\end{aligned}$$

In the two-dimensional case, Eq. A4 becomes

$$\begin{aligned}\bar{V}_{m-1} &= V_{m-1} \otimes V_{m-1} \\ &= (V_m \oplus W_m) \otimes (V_m \oplus W_m) \\ &= (V_m \otimes V_m) \\ &\quad \oplus [(V_m \otimes W_m) \oplus (W_m \otimes V_m) \oplus (W_m \otimes W_m)] \quad \text{(A7)} \\ &= \bar{V}_m \oplus \bar{W}_m, \quad \text{(A8)}\end{aligned}$$

and thus we have,

$$L^2(R^2) = \bigoplus_{m \in Z} \bar{W}_m.$$

Now, it should be noted that the detail space $\bar{W}_m$ (regarding Eq. A8) consists of three components, with three corresponding basis functions as shown below:

$$\begin{aligned}\Psi^1_{m,k}(x_1, x_2) &= \phi_{m,k_1}(x_1)\psi_{m,k_2}(x_2) \\ \Psi^2_{m,k}(x_1, x_2) &= \psi_{m,k_1}(x_1)\phi_{m,k_2}(x_2) \\ \Psi^3_{m,k}(x_1, x_2) &= \psi_{m,k_1}(x_1)\psi_{m,k_2}(x_2), \quad \text{(A9)}\end{aligned}$$

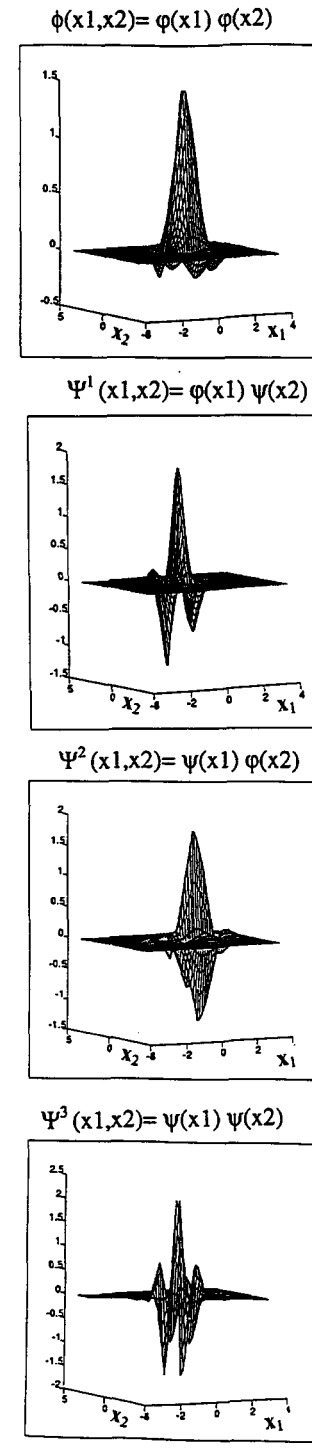


**Figure A1. Two-dimensional wavelets and scaling function corresponding to Daubechies ($N = 4$) wavelet.**

where each of $\Psi^1_{m,k}$, $\Psi^2_{m,k}$, and $\Psi^3_{m,k}$ denote the detail in the "horizontal," "vertical," and "diagonal" orientations, respectively (Daubechies, 1992). Therefore, compared to the one-dimensional case, the situation is somewhat more complicated with three two-dimensional wavelets. Figure A1 depicts the two-dimensional Daubechies ($N = 4$) scaling function and wavelets constructed from their one-dimensional counterpart using Eqs. A6 and A9.

## Appendix B: Wavelet Construction

The approach of constructing wavelets and scaling functions from the coefficients $c_k$ is adopted from Strang (1989). Here we use this method to construct the Daubechies ($N=2$) scaling function defined on (0 3) (i.e., which is also defined on [−1 2]). A similar method is applied to the other scaling functions or wavelets.

The Daubechies scaling function just mentioned has the following coefficients (Daubechies, 1992; Strang, 1989):

$$c_0 = \frac{1}{4}(1+\sqrt{3}),$$

$$c_1 = \frac{1}{4}(3+\sqrt{3}),$$

$$c_2 = \frac{1}{4}(3-\sqrt{3}),$$

$$c_3 = \frac{1}{4}(1-\sqrt{3}).$$

Assume that $\phi$ is known at the integers $x=j$'s. The following recursion,

$$\phi(x) + \sum_k c_k \phi(2x-k), \qquad \text{(A10)}$$

provides values of $\phi$ at the half-integers. It then gives values of $\phi$ at the quarter-integers, and recursively at all dyadic points $x=k/2^j$ (see Figure A2). This is programmed very fast. For instance, for the just mentioned scaling function, we have $\phi(x)=0$ for $x$ not belonging to [0 3]. Now we write Eq. A10 at $x=1$ and $x=2$ (with $c_k$ defined as before),

$$\phi(1) = c_1\phi(1) + c_0\phi(2)$$

$$\phi(2) = c_3\phi(1) + c_2\phi(2).$$

This is $\phi = L\phi$, where $\phi = [\phi(1)\ \phi(2)]^T$ and matrix entries $L_{ij} = c_{2i-j}$. The eigenvalues of this matrix are $\lambda=1$ and $\lambda=1/2$. The eigenvector corresponding to $\lambda=1$ has the components as follows:

$$\phi(1) = \frac{1}{2}(1+\sqrt{3}),$$

$$\phi(2) = \frac{1}{2}(1-\sqrt{3}).$$

For the description of the eigenvector corresponding to $\lambda=1/2$, see Strang (1989). With the recursion (Eq. A10) and using the value of $\phi$ at $x=1$ and $x=2$, one can obtain everything. It is also worth emphasizing that the corresponding equation for the wavelet can be written as

$$\psi(x) = \sum_k \bar{c}_k \phi(2x-k),$$

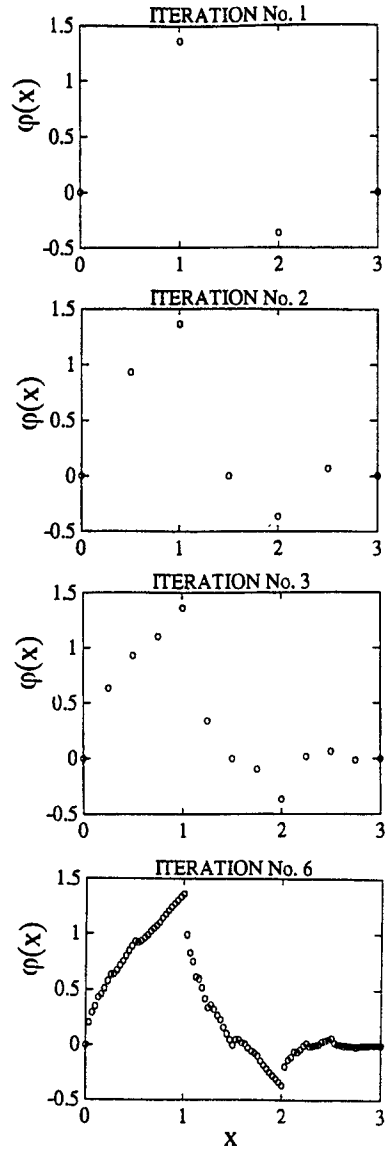


**Figure A2. Recursive calculation of the values of the Daubechies ($N=2$) scaling (on [0, 3]) at dyadic points $x=k/2^{-m}$ (see text for details).**

where

$$\bar{c}_k = (-1)^k c_{1-k}.$$

Now, using the $c_k$ for the scaling function and the values of the scaling function at dyadic points ($k/2^{-m}$), one can find the values of the wavelet at dyadic points.

## Appendix C: Double CSTR's Flowsheet

The system in Figure 8 is described here in more detail. Some simplifying assumptions in the model are:

1. The densities of the mixtures are constant at all temperatures and concentrations.

2. The heat capacities are also constant at all temperatures and concentrations.

The reaction is modeled in terms of the raw material ($A$) concentration, $C$:

$$\text{Rate} = -K_0 \cdot C \cdot V \cdot e^{-E/RT}, \tag{A11}$$

which yields a rough balance of

$$V\frac{d(C)}{dt} = \text{Rate} + Q_F(C_F - C). \tag{A12}$$

The reactors are cooled by coolants passing through jackets surrounding each reactor,

$$\text{COOL} = UA(T - T_c). \tag{A13}$$

The energy balance for the reaction is then

$$V\frac{d(T)}{dt} = DH \cdot \text{Rate} + Q_F(T_F - T) - \text{COOL},$$

where

$$DH = -\frac{\Delta H_R}{\rho \cdot C_p}.$$

The reactors are considered to be fixed volume, with a volume of

$$V = V_{ss} = 5.0,$$

which gives the following overall balance at steady state,

$$0 = \frac{d(V)}{dt} = Q_F - Q.$$

*Manuscript received Sept. 28, 1995, and revision received Oct. 28, 1996.*