

Fault Detection and Diagnosis in a Food Pasteurization Process with Hidden Markov Models

Figen (Kosebalaban) Tokatli¹ and Ali Cinar^{2*}

¹ *Izmir Institute of Technology, Food Engineering Department, Izmir, 35430 Turkey*

² *Illinois Institute of Technology, Chemical and Environmental Engineering Department, 10 W.33rd Street, Chicago, IL 60616*

Fast and accurate detection and diagnosis of process faults has a strong impact on safe and profitable operation of food processing systems and safety of products. In an earlier study, multivariate statistical process monitoring charts based on Hotelling's T^2 and squared prediction error (SPE) statistics and contribution plots, and parity space techniques were shown to be useful tools for monitoring the system and detecting and diagnosing the abnormalities of a pilot-scale high-temperature-short-time (HTST) pasteurization plant (Kosebalaban and Cinar, 2001). It was shown that these techniques could alarm the malfunctioning of sensors at the same time or earlier than univariate statistical process monitoring charts. In this study, the same experimental data are used to test fault detection and diagnosis by using Hidden Markov Models (HMM) based techniques. Case studies show that some of the HMM state variables detected certain types of faults while others discriminated between magnitudes of faults. This provides the opportunity to develop diagnosis schemes by HMMs for different faults that can occur in the HTST pasteurization process.

The paper is organized as follows. The next section summarizes the use HMMs for process monitoring and fault diagnosis. Then, HMMs are briefly introduced in the section titled Hidden Markov Models. The pilot-scale high-temperature-short-time (HTST) pasteurization plant where experiments were performed, the data acquisition system (DAS), and the implementation of the faults are described in the section titled HTST Dairy Pasteurization and Types of Faults. The use of HMM techniques for fault detection and diagnosis in HTST pasteurization data is presented and results are discussed in the "Results" section. Conclusions are presented in the last section.

Process Monitoring and Fault Diagnosis with HMMs

Detailed reviews and theoretical background of HMMs can be found in the literature (Rabiner, 1989). An HMM can be defined as a statistical model. In the communication literature, it is the probabilistic functions of Markov chains. HMMs have found widespread application in speech recognition and computational biology. HMMs were first implemented for speech recognition studies in the late 1960s and early 1970s. In bioinformatics, HMMs are used to statistically describe protein family's sequences. Rabiner (1989) provides a detailed introduction to HMMs along with applications in speech recognition. Cohen (1998) presents discrete density HMM (DD-HMM) and continuous density HMM (CD-HMM) and use of HMM in biomedical signal processing.

The use of HMMs for monitoring dynamic systems is discussed by several researchers. Smyth (1994) describes the use of HMMs in

Hidden Markov Models (HMM) are used to detect abnormal operation of dynamic processes and diagnose sensor and actuator faults. The method is illustrated by monitoring the operation of a pasteurization plant and diagnosing causes of abnormal operation. Process data collected under the influence of faults of different magnitude and duration in sensors and actuators are used to illustrate the use of HMM in the detection and diagnosis of process faults. Case studies with experimental data from a high-temperature-short-time pasteurization system showed that HMM can diagnose the faults with certain characteristics such as fault duration and magnitude.

On a employé des modèles cachés de Markov (HMM) afin de détecter le fonctionnement anormal des procédés dynamiques et de diagnostiquer les défaillances des appareils de détection et de mise en marche. La méthode est illustrée par le suivi du fonctionnement d'une installation de pasteurisation et en diagnostiquant les causes d'un fonctionnement anormal. On a eu recours à des données de procédés recueillies lors de défaillances d'importance plus ou moins grave et de durée différente des appareils de détection et de mise en marche pour illustrer l'utilisation des HMM dans la détection et le diagnostic des défaillances de procédés. Des études de cas utilisant des données expérimentales pour un système de pasteurisation rapide et à température élevée montrent que les HMM peuvent diagnostiquer les défaillances avec certaines caractéristiques telles que la durée et l'importance de la défaillance.

Keywords: hidden Markov models, fault diagnosis, food processing

monitoring dynamic systems with a case study of antenna fault monitoring. Kwon and Kim (1999) used HMMs for accident identification of a nuclear power plant by using a real-time test simulator including 22 variables. In another study (Ma and Shi, 2000), a stochastic approach based on HMM to distinguish a certain process behavior from a fault case is developed for transformer protection. Wong et al. (1998) present a trend detection and classification study for a single variable of a continuously stirred tank reactor (CSTR) by using HMM. In their recent work (Wong et al.,

* Author to whom correspondence may be addressed. E-mail address: cinar@iit.edu

2001), two methods of multiple process faults classification are presented for a simulation of a first order, non-isothermal CSTR with six variables. In both studies, simulated data are first denoised by a wavelet-based smoothing algorithm. Then, the process data are fuzzified by using triangular symbols before they were introduced to an HMM algorithm.

HTST pasteurization is a multivariable continuous process with energy recycle. As a high-speed integrated process, it yields a challenging fault diagnosis problem. HMMs provide a powerful framework for recognizing patterns in data and diagnosing process faults in the HTST pasteurization process. The procedure used determines first whether the mean values of HMM state variables for each process variable have changed significantly from their in-control values. Then, by tracking these HMM states, the trends of change in the HMM states are used to identify the faults that caused the variation in behavior. Two different approaches have been tested for developing the HMMs. In one approach, the whole data set that contains all the faults that need to be diagnosed is used for developing the HMM. Similar faults are represented by the same HMM states. The second approach uses a moving horizon (sliding time window) observation sequence. An HMM with a specific number of states (K) and specific observation vector of length T (at each sampling time t) on a time window length L is used. As new data (observation vector) are collected and included in the training set, the oldest observation vector is eliminated from the set. Each time new data are collected, a new HMM with the same K and T is formed. The advantage of this approach is its ability to detect and diagnose small magnitude faults. In the Results section, the diagnostic power of this moving horizon approach is illustrated by detecting small temperature increases that were not detected by the first HMM approach or the parity space technique and the state space identification integrated with multivariate statistical process monitoring technique (Kosebalaban and Cinar, 2001).

There are no strong guidelines for choosing the number of states (Wong et al., 1998, 2001). A small number of HMM states K allows faster training time and analysis. However, this usually results in poor estimation of data. A large number of HMM states improves the estimation but slows down the training time. The performance of HMM in representing data can be measured by the residuals of the model or the correlation coefficients of observed and estimated values of process variables. The residuals have normal distribution ($\sim N(0, \sigma^2)$) if there is no systematic information left in them. The normality plots (QQ plot) of residuals of important process variables can provide useful information about model quality.

Hidden Markov Models

HMM is a model in which a sequence of observations is modeled by a probabilistic relation between observations and a sequence of hidden, unobservable states and a Markovian transition structure linking the hidden states. HMMs are the learning probabilistic models of time series.

A HMM can be represented by the compact notation $\lambda = (\pi, \mathbf{A}, \mathbf{B})$. All members of a HMM are defined by:

1. A set of K states, $S = S_1, S_2, \dots, S_p, \dots, S_K$
2. Initial probability distribution, the probability of being in state S_i at time $t = 1$ is π_i . The initial probability vector is defined by $\pi = [\pi_1, \pi_2, \dots, \pi_K]$.
3. A probability distribution to describe the transitions between states:

$$a_{ij} = P(S_j \text{ at time } t | S_i \text{ at time } t-1) \quad (1)$$

This transition matrix \mathbf{A} of dimensions $K \times K$ is defined by:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \dots & a_{Kk} \end{bmatrix} \quad (2)$$

where $a_{ij} \geq 0$ and $\sum_{j=1}^K a_{ij} = 1$

4. Assuming that discrete signals (observations) in an observation sequence

$O = (O_1, O_2, \dots, O_p, \dots, O_T)$ belong to a set of M distinct symbols of

$V = (V_1, V_2, \dots, V_k, \dots, V_M)$, $b_j(k)$ is the probability that the k^{th} symbol will be generated by the j^{th} state:

$$b_j(k) = P(O_t = V_k | S_t = j) \quad (3)$$

where $j = 1, 2, \dots, K$ and $k = 1, 2, \dots, M$. Therefore, $\mathbf{B} = [b_j(k)]$ is $K \times M$ observation probability matrix.

There are two important stages in hidden Markov modeling. The first step is the training stage where parameters of HMM are estimated. Training is a learning process. Proper estimates of π , \mathbf{A} and \mathbf{B} are learned in an algorithm. The most popular algorithms are forward-backward (or Baum-Welch) and Viterbi algorithms. Model parameters π , \mathbf{A} and \mathbf{B} are estimated in each iteration to find $\lambda = (\hat{\pi}, \hat{\mathbf{A}}, \hat{\mathbf{B}})$. This is for improving the probability of observation sequence O being observed from the model ($P(O|\lambda)$). The iteration continues as long as $P(O|\hat{\lambda}) > P(O|\lambda)$. The final values of the estimates are called maximum likelihood estimates of the HMM. The second step is the evaluation of new data. It focuses on computing the probability that the observed sequence was produced by the model, or how well a model matches a given observation sequence. Discrete HMMs use discrete probability distributions to model the observation probabilities $b_j(k)$. Continuous density HMMs use continuous probability functions to describe the statistics within HMM states.

Represent the sequence of continuous observation vector by \mathbf{y}_t where $t = 1, 2, \dots, T$. The state transition probabilities, $P(S_t | S_{t-1})$ are specified by a $K \times K$ transition matrix \mathbf{A} . For a continuous observation vector, $P(\mathbf{y}_t | S_t)$ is the observation probability and can be modeled in many different forms. Gaussian type is the most popular form (Ghahramani and Jordan, 1997; Cohen, 1998). For a continuous observation vector \mathbf{y}_t , the probability density function of continuous observations in terms of Gaussian function is:

$$P(\mathbf{y}_t | S_t) = \frac{1}{S^{1/2} (2\pi)^{p/2}} \exp\left\{-\frac{1}{2} (\mathbf{y}_t - \boldsymbol{\mu})^T S^{-1} (\mathbf{y}_t - \boldsymbol{\mu})\right\} \quad (4)$$

where p is the number of variables, \mathbf{S} is $p \times p$ covariance matrix, and $\boldsymbol{\mu}$ is the mean vector of HMM states S of length p . The mean vector and covariance matrix are estimated by the forward-backward (Baum-Welch) algorithm in the training step.

One problem in building the HMM is to find a state sequence which best explains the real observations. For this, a new variable γ is defined in terms of the forward (α) and

backward (β) variables of Baum-Welch algorithm (Rabiner, 1989).

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^K \alpha_t(i)\beta_t(i)} \quad (5)$$

where $\sum_{i=1}^K \gamma_t(i) = 1$. $\gamma_t(i)$ is the probability of being in state j at time t and γ matrix has the size of $T \times K$. By using γ , the HMM state trajectories can be monitored over time sequence T .

HTST Dairy Pasteurization and Types of Faults

Pasteurization is a heat treatment process of foods to secure destruction of pathogenic bacteria without markedly affecting the physical and chemical properties of the end product. In HTST pasteurization of milk, the standard time-temperature combination is 72°C (161°F) with a holding time of 15 sec is used. Then, the pasteurized product is cooled and sent to storage. The HTST pasteurization plant that provided the data for evaluation of HMM based monitoring and diagnosis is located at the National Center for Food Safety and Technology (NCFST) of Illinois Institute of Technology. A process flow and instrument diagram is given in Figure 1. The pilot pasteurization plant consists of a plate heat exchanger, a centrifugal pump, a flow diversion valve (FDV) (A-P-C 3A standard 08-17 type, model 1-1/2"), a boiler and a homogenizer. The heat exchanger is a multipass, plate heat exchanger by the APV Company. The homogenizer (APV) is used as a timing pump and has a capacity of 15-50 gallon/L. There are two regulatory valves. One is steam injection valve to the boiler and the other is hot water flow valve in preheater section.

The incoming raw milk passing through the regenerator section goes first to the preheater section where it exchanges heat with hot water for controlling raw product temperature entering the homogenizer. After the homogenizer, raw milk flows to the main heat exchanger and follows the same procedure as in the generic pasteurization plant.

The primary source of heat is hot water. The hot water is heated by direct steam injection in the hot water heater. Three PID controllers are used to control product temperature. The first control loop regulates the raw product temperature leaving the preheater. The second loop controls product temperature entering the holding tube. The last loop controls the temperature of the pasteurized product leaving the cooler. The raw product temperature at the exit of the preheater is controlled by manipulating the flow of hot water through the preheat heater exchanger. The product temperature at the holding tube inlet is controlled by manipulating the steam flow rate into the hot water heat exchanger. The cooler product temperature is controlled by manipulating the flow rate of cold water through the cooler heat exchanger. The flow diversion valve is controlled by pasteurized milk temperature at the holding tube exit. The measured variables are hot water, holding tube inlet, holding tube outlet, and preheater exit temperatures and the steam valve and preheater valve signals.

Data Acquisition and Control System

The HTST system was equipped with a data acquisition system (DAS) wired to sensors for monitoring the process variables.

The real-time data acquisition computer system consists of a HP 75000 System with a HP 75000 card gage and a Compaq Presario 7132 personal computer. The measurement card on HP 75000 includes 5.5 digit multimeter, two 16 channel thermocouple relay multiplexers, a quad 8-bit digital input/output card and a 4-channel D/A converter. The software used for data acquisition, display, storage and process control is HPVEE 3.12. The control signals, which are signals to pneumatic steam valve in main heating loop, are generated by PID controllers designed in HPVEE software and send to the HTST system. The digital 4-20 mA signal is converted to 20.7–103 kPa (3–15 psig) pneumatic signal by a model 1000, I/P transducer (Bellofram Co. Precision Controls Division, Burlington, MA).

The variables used in process modeling and fault diagnosis are four temperature measurements (°F) and two PID controller outputs (mA). Hot water temperature, preheater outlet temperature of raw product, holding tube inlet temperature of pasteurized product and holding tube outlet temperature of pasteurized product are the output variables of the process (Variables 1–4, respectively). The input (manipulated) variables of the process are the PID controller output to the steam valve (variable 5) that regulates the holding tube inlet temperature of product and the PID controller output to preheater hot water valve (variable 6) that regulates the preheater outlet temperature of raw product.

Implementation of Faults

Sensor and actuator faults were implemented to the HTST pasteurization system. Experiments were conducted with different fault magnitudes and durations.

Sensor Faults

Sensor failures were introduced to the system through the software by using the process control computer. To generate a sensor failure for a specific time period, a real number is added to the actual sensor reading, which is transmitted to the computer from the process. Instead of the actual reading of the sensor, the modified sensor reading is sent to the PID controllers. The controllers compute the manipulated variable settings based on the failed sensor reading. Therefore, the system receives a false action through controllers even though there was no need for any adjustments in the flow rates of steam and hot water, and the fault implemented on the sensors propagates through the system. The magnitude of faults introduced to the sensors changes between variations of -0.83°C and 0.83°C (-1.5°F and 1.5°F). The duration of faults changes between 2 s and 30 s. The type of disturbance introduced is bias change.

Actuator Faults

The faults in the actuators can be caused by the controllers or any mechanical failure in the actuator itself. To implement actuator faults in the system, the controllers are turned off for a specific time period. A constant signal is sent to the actuators.

Faults can be considered as malfunctions in the controllers that send signals to the actuators. During the fault implementation, the controlled variables show the signs of abnormality in the system. However, the controller and actuator cannot respond to the needs of the process until the implementation is over.

Results

In this study, it was observed that similar faults were represented by a particular HMM state variable. The methodology used

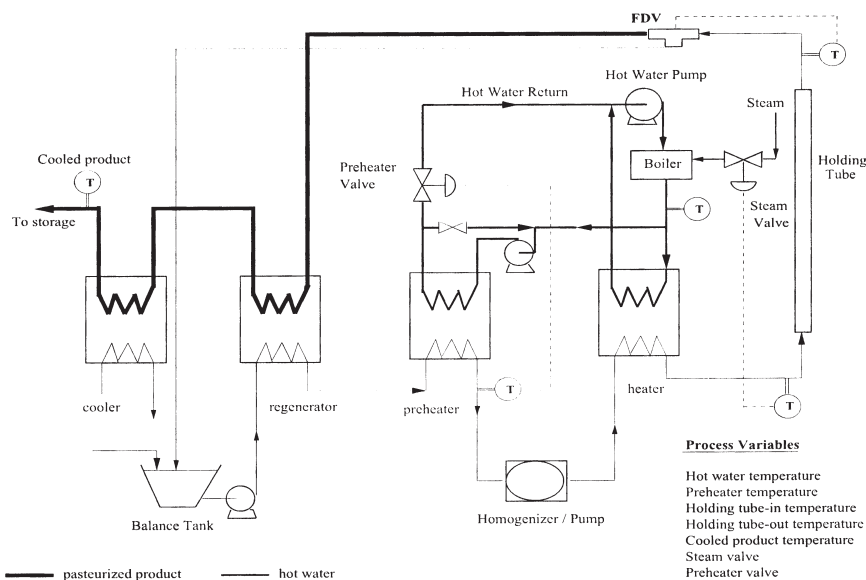


Figure 1. Schematic diagram of the HTST pasteurization plant.

here is first to determine whether the mean values of HMM state variables for each process variable change significantly compared to in-control HMM states. Then, by tracking these HMM states, the increase in probabilities of these states is investigated when corresponding faults occur in the system.

Some HMM state variables alarmed certain type of faults while other HMM state variables alarmed only faults with different characteristics. After construction of a HMM for the given data sequence, the mean values of K state variables (μ matrix) are checked first for an abnormal increase or decrease in state variables. Then, the probability values which these particular states take during a specific time period are examined by using the γ matrix to determine the time and duration of faults.

Both HMM based on a single model development for fault diagnosis (Method 1) and HMM based on sliding data windows (Method 2) are tested with HTST pasteurization data. The performance of HMM in representing HTST data is assessed by the residuals of the model and the correlation coefficients of observed and estimated values of process variables. The normality plots (QQ plot) belonging to residuals of some important process variables (holding tube inlet temperature and steam valve setting, variables 3 and 5, respectively) were checked after the HMM of an observation sequence is obtained. For the single HMM developed using all the data with several faults (Method 1), residuals have autocorrelation most of the time, at least up to lag 6 and also at lags 13-16. The normality plots are affected by some extreme values of faulty signals since the model may not perform well to estimate the measurements at times of fault implementation. Otherwise, the middle portion of the QQ plots indicates Normal distribution. The autocorrelation problem is not specific to method 1. The same problem would occur in method 2, if the short data segment contains too many faults. In case of too many outliers in the observation sequence, it is very likely that distribution of residuals will not have Normal distribution.

The number of HMM states (K) is an indicator of model performance. Low K values are not considered for HTST pasteurization data because they yield HMMs with poor predictions. High K values improve the predictions as expected but

Table 1. Effect of the number of HMM states K on SPE and r values of variables 3 and 5.

K	SPE	r	r
		(Holding tube-inlet estimated and actual values)	(Steam valve signals estimated and actual values)
30	112.8748	0.9667	0.9374
50	60.9471	0.9890	0.9825
70	33.6863	0.9971	0.9969
90	28.5711	0.9972	0.9939

increase the computational load. Table 1 shows the squared prediction errors (SPE) of a data set by HMM with different K values and the r values for holding tube inlet temperature and steam valve signals. A second disadvantage of large K values is over-fitting. Some states become too specific for certain process behavior. For example, each temperature increase in holding tube inlet sensor can be represented by different states even though there are other faults with the same magnitude causing similar reactions in the system.

Real process data may show strong autocorrelation, crosscorrelation and include noise. HMMs may require high number of states to truly represent the system. In this study 50 states were chosen by considering the advantages and disadvantages selecting a high K value.

Case Study with Single HMM Developed for the Whole Data Set

A number of sensor and actuator faults were implemented in the system to test the performance of HMM technique in the classification of faults. Various fault magnitudes and durations were used on controlled and manipulated variables of the process. Two cases of actuator faults in steam valve and a case of sensor faults in holding tube inlet sensor are discussed for illustrating the method.

Sensor faults were introduced in the holding tube inlet

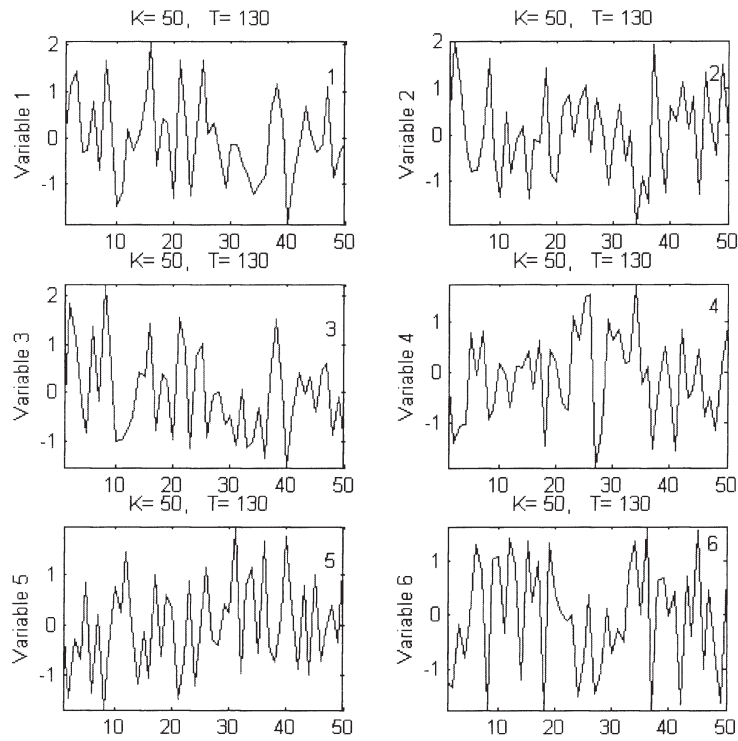


Figure 2. Mean values of HMM states for 6 process variables of in-control data sequence.

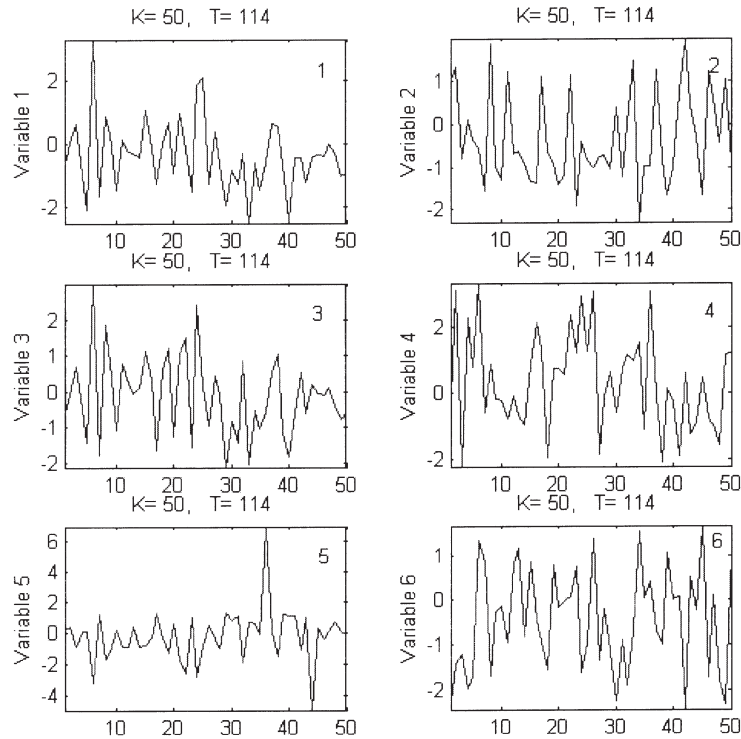
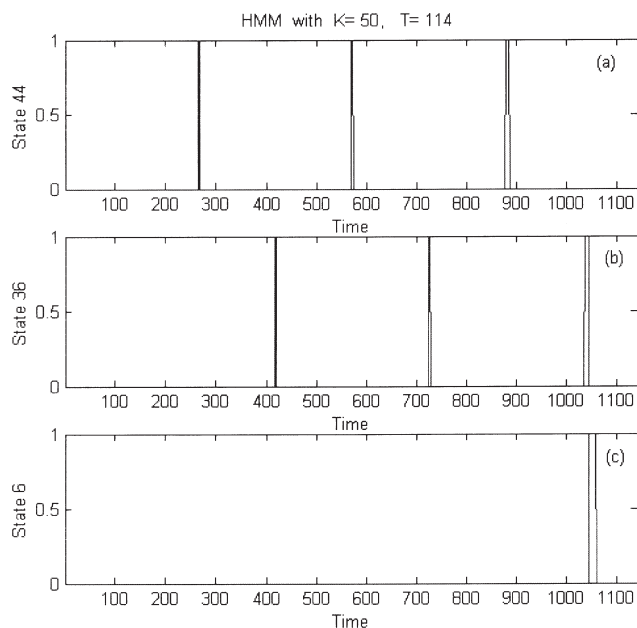
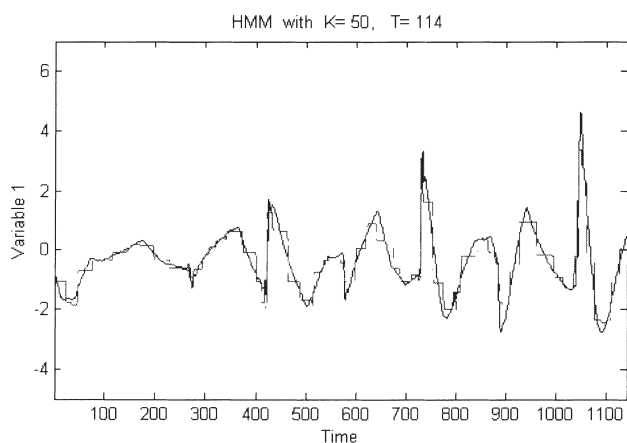


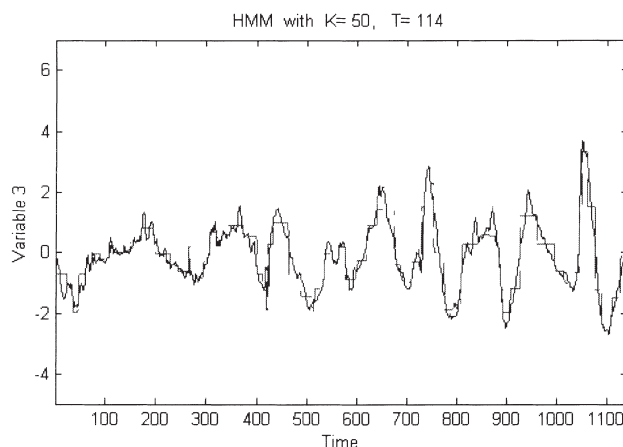
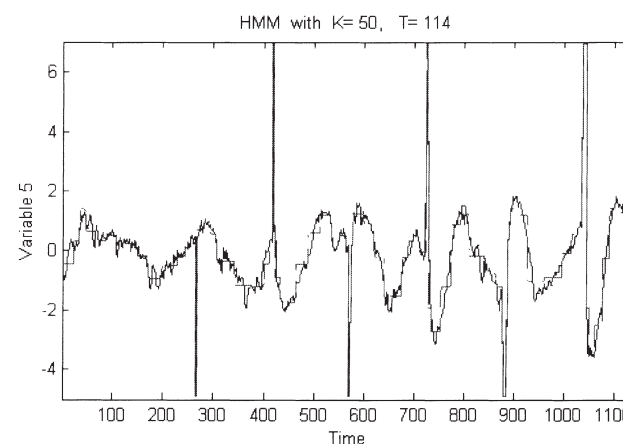
Figure 3. Mean values of HMM states for 6 process variables of data sequence with steam valve actuator fault (Case I).

Table 2. Steam valve faults (Case I): times and magnitudes of faults.

Fault	Fault Time (s)	Steam Valve Signal (mA)	Duration (s)
1	266	6.0	2
2	418	12.0	2
3	570	6.0	4
4	724	12.0	4
5	878	6.0	8
6	1036	12.0	8

**Figure 4.** Probability values of states 44, 36 and 6 of HMM of data sequence: steam valve actuator fault (Case I)**Figure 5.** Hot water temperature measurements (solid line) and estimated values (dashed line): steam valve actuator fault (Case I).

temperature sensor. Since this is the controlled variable of the process, any deviation in its measurements causes the controllers of the system to respond to that change and modify

**Figure 6.** Holding tube inlet temperature measurements (solid line) and estimated values (dashed line): steam valve actuator fault (Case I).**Figure 7.** Steam valve signals (solid line) and estimated values (dashed line): steam valve actuator fault (Case I).**Table 3.** Steam valve fault (Case II): times and magnitudes of faults.

Fault	Fault Time (s)	Steam Valve Signal (mA)	Duration (s)
1	102	6.0	2
2	304	12.0	6
3	510	12.0	2
4	712	6.0	6

process operation. The actuator faults were introduced in steam controller valve. Since this is the manipulated variable, any fault in controllers or valves causes all process variables to behave abnormally depending on the magnitude and duration of the fault.

Any fault in holding tube inlet sensor causes a prompt response by the steam valve that supplies the heating medium into the system. The product temperature at the exit of the

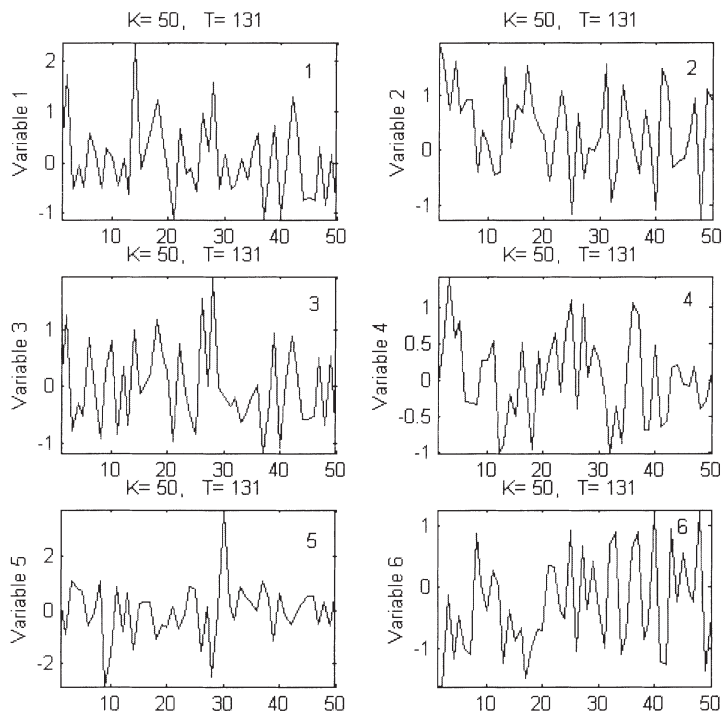


Figure 7. Mean values of HMM states for 6 process variables of data sequence: steam valve fault (Case II).

heater is an important variable in HTST food pasteurization. The hot product is then introduced to the holding tube. The temperature at the exit of the holding tube is the critical control point and it is very strongly correlated with the holding tube inlet temperature. Therefore, any kind of failure related to both holding tube inlet sensor and steam valve may cause undesirable consequences as far as consumer health is concerned.

Differences in the mean values of HMM state variables for two different operations are depicted in Figures 2 and 3. Figure 2 displays the mean values of HMM states of process variables for data collected under normal operating conditions. For example, mean values of states for steam valve signals (variable 5) change between -2 and $+2$ approximately. Mean values of HMM states of process variables for data sequences with steam valve actuator faults are displayed in Figure 3 (Case I). Mean values of states for steam valve signals (variable 5) vary between -5 and $+7$ approximately. Therefore, HMM state variables can be monitored to determine their behavior with respect to their mean values for particular process variables under various operating conditions.

Table 2 gives the magnitude, duration, and time of occurrence of six different faults implemented in steam valve for the first case study. The length of data sequence T used for HMM development is 114. The number of states of the HMM is 50. Figure 3 shows the mean values of HMM states for the data sequence collected under faults implemented in steam valve as given in Table 2. State 36 of steam valve signals (variable 5) in Figure 3 has the highest mean value among the HMM states and state 44 has the lowest mean value. Figure 4 shows the probability values that these particular states took. It is clearly seen that state 44 alarms faults in which the steam valve signals are low, namely 6 mA. On the other hand, state 36 alarms the faults in which steam valve signals are higher,

12 mA. HMM state 6 has large values among the states of both hot water temperature sensor (variable 1) and holding tube inlet temperature sensor (variable 3) (Figure 3). This state gets a low mean value for steam valve signals (variable 5 in Figure 3). As seen in Figure 4, state 6 gets high probability value after fault 6 (magnitude of 12 mA with duration of 8 in Table 2), which causes an increase in both temperature sensors. This is a typical behavior when a fault occurs in the steam valve of the HTST pasteurization system. Since the steam valve opens up to 12 mA and introduces large amounts of steam into the heater, the hot water temperature and consequently the product temperature in holding tube inlet increase as seen in the scaled measurements of the variables (Figures 5, 6 and 7).

Similar results are observed in the second case of steam valve faults (Case II). Table 3 gives the magnitude, duration and time of four different faults implemented in steam valve. The length of data sequence, T , during HMM development is 131. The number of states is 50. First, the mean values of HMM states are checked in Figure 8. State 30 of variable 5 (steam valve signals) gets the highest mean value among the HMM states. State 9, on the other hand, gets the lowest mean value. Figure 9 shows the probability values that these particular states took. State 9 alarms faults when steam valve signals are low (6 mA) and state 30 alarms the faults when steam valve signals are high (12 mA) (Table 3). State 14 of this HMM is also worth of examining. In Figure 8, state 14 of hot water temperature measurements (variable 1) gets the highest mean value. State 14 gets the probability value of 1 during the time intervals just after the second fault in steam valve at time 304 sec with the magnitude of 12 mA and duration 6 (Figure 9). Clearly, this fault causes an extreme increase in the temperature of hot water and this is shown in the scaled measurements

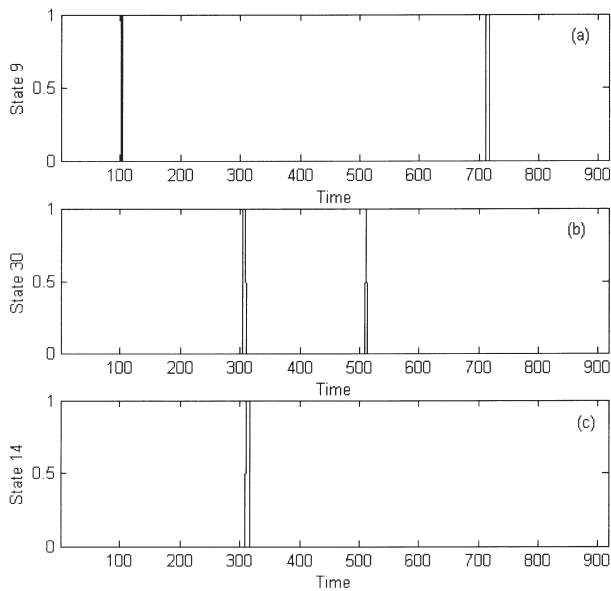


Figure 9. Probability values of states 9, 30 and 14 of HMM of data sequence: steam valve fault (Case II).

of the variables (Figures 10 and 11).

Faults in holding tube inlet temperature sensor (variable 3) are investigated in the third case (Case III). The time of occurrence, magnitude, and duration of six sensor faults are given in Table 4. The magnitudes of faults implemented to the holding tube inlet temperature sensor, are $+0.39^{\circ}\text{C}$ and $+0.83^{\circ}\text{C}$ ($+0.7^{\circ}\text{F}$ and $+1.5^{\circ}\text{F}$). The fault duration is between 2 and 8. The length of data sequence, T , during HMM development is taken as 131. The number of states is 50. Figure 12 shows the mean values of HMM states for six process variables. State 43 of steam valve signals (variable 5) has the lowest value among all HMM states and the highest value for the holding tube inlet temperature sensor (Variable 3). Whenever a fault occurs in the holding tube inlet sensor, the steam valve responds promptly. Consequently, the same HMM state variable represents the changes in both variables. Since the faults in the sensor are positive magnitude faults, they cause a reduction in signal magnitudes to the steam valve. Figure 13 shows the probability values of state 43 during process operation. State 43 alarms the faults with magnitudes of $+0.8^{\circ}\text{C}$ ($+1.5^{\circ}\text{F}$), which cause stronger responses than the faults numbered 1, 3, and 5 with fault magnitudes of $+0.4^{\circ}\text{C}$ ($+0.7^{\circ}\text{F}$) in Table 4. State 35 of steam valve (variable 5), which has the largest value, indicates all faults except the first one (Figure 13 (b)). The probabilities become 1 during the time intervals after the faults occur. This state contributes to situations where the steam valve opens and injects additional steam just after the closing action because of increasing temperature in the holding tube inlet sensor. The second largest mean value belongs to HMM state 13 among the holding tube inlet temperature states. The final plot in Figure 13 shows the probability values of state 13 during the operation. This state alarms faults 2, 3, 4, 5 and 6 (Table 4). The first fault in the holding tube inlet is hard to detect since it does not cause serious deviation in any of the process variables (Figures 14 and 15).

In the case of holding tube inlet sensor faults, the same HMM states contribute to changes in the holding tube inlet

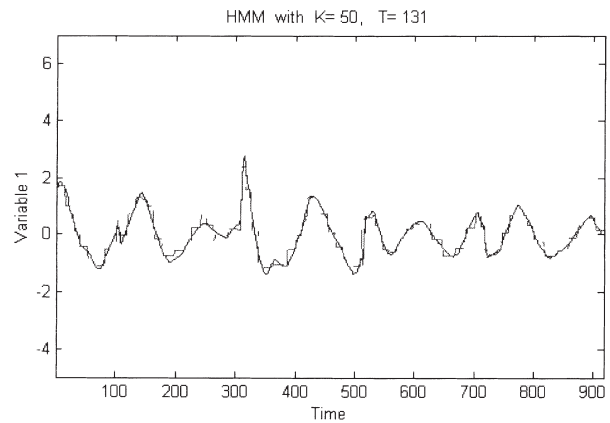


Figure 10. Hot water temperature measurements (solid line) and estimated values (dashed line): steam valve fault (Case II).

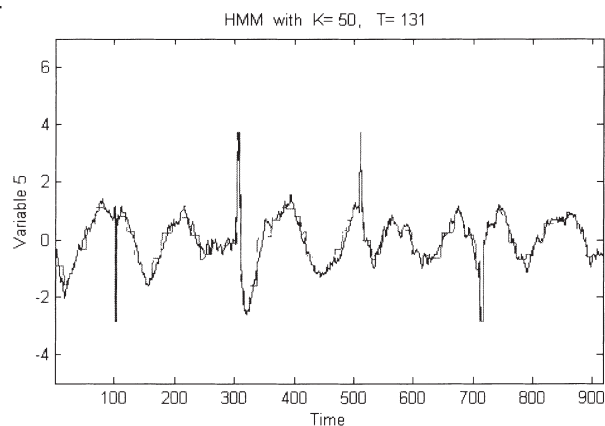


Figure 11. Steam valve signals (solid line) and estimated values (dashed line): steam valve fault (Case II).

Table 4. Holding tube inlet sensor fault (Case III): times and magnitudes of faults.

Fault	Fault Time (s)	Steam Valve Signal (mA)	Duration (s)
1	63	+0.7	2
2	215	+1.5	2
3	367	+0.7	4
4	521	+1.5	4
5	675	+0.7	8
6	833	+1.5	8

temperature measurements and steam valve actuator behavior. This situation is not observed in HMM of data sequences with the faults in steam valve (Case I and II).

Case Study with HMM Developed Using Sliding Window Data

The most important achievement of this application is the identification of faults with small magnitude. For example, in case III, the first fault is the 0.4°C (0.7°F) temperature increase in the

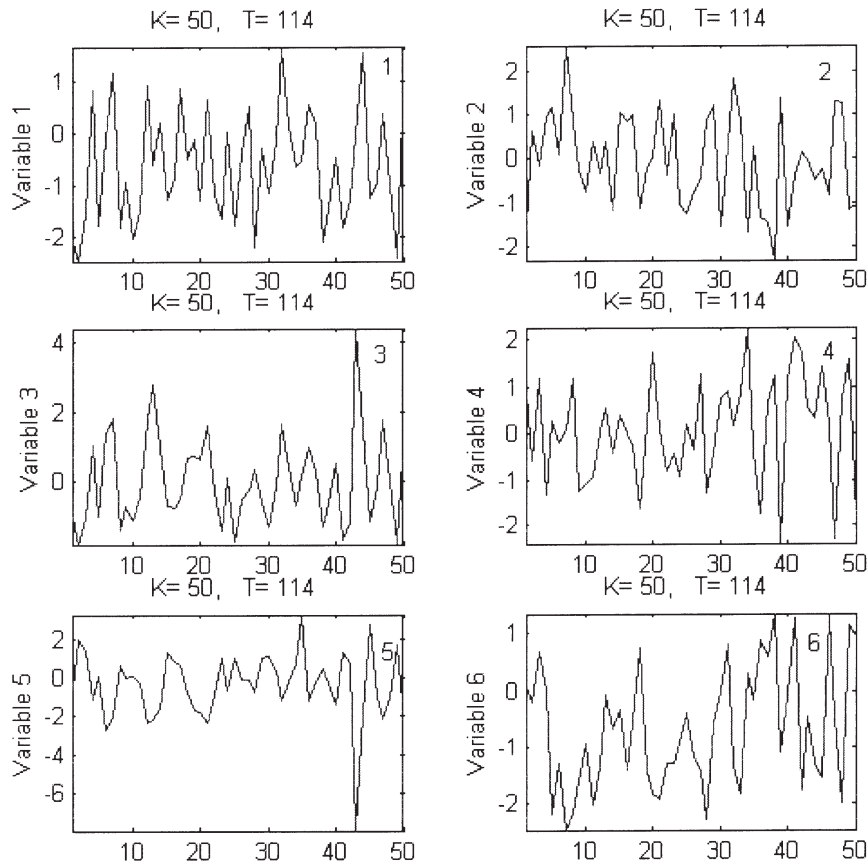


Figure 12. Mean values of HMM states for 6 process variables of data sequence: holding tube inlet sensor faults (Case III).

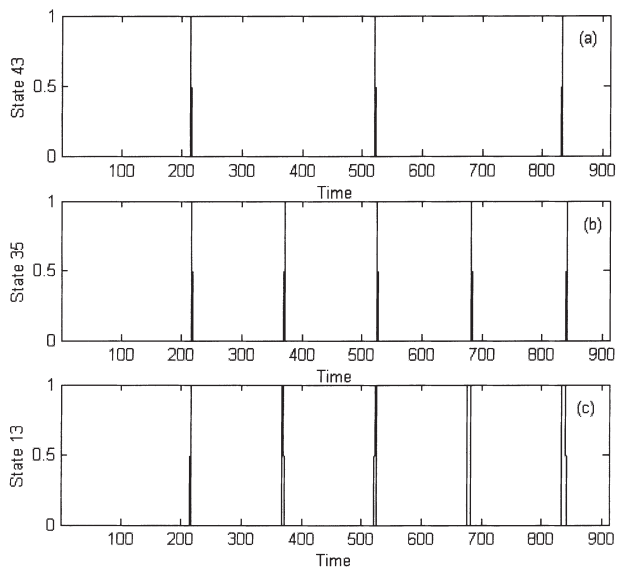


Figure 13. Probability values of states 43,35 and 13 of HMM of data sequence: holding tube inlet sensor faults (Case III).

holding tube-inlet measurement. These types of small magnitude faults are difficult to identify since they may not cause any large

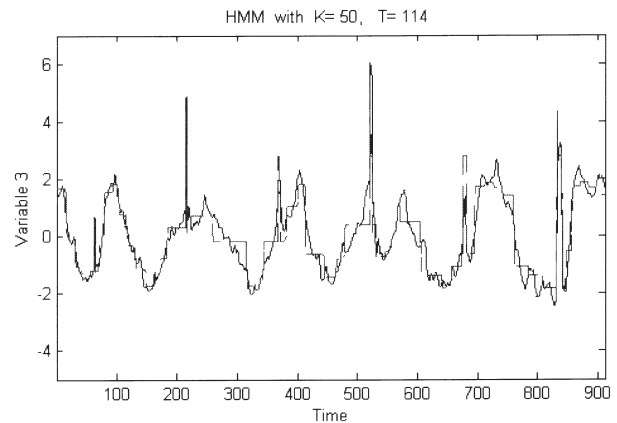


Figure 14. Holding tube inlet temperature measurements (solid line) and estimated values (dashed line): holding tube inlet sensor fault (Case III).

change in the system at that particular time. In the earlier study by Kosebalaban and Cinar (2001), neither parity space nor state space identification integrated with multivariate charting techniques could detect this small magnitude fault. However, HMM of the same process data diagnosed this failure in the temperature sensor by applying the sliding window technique described above online. The diagnosis power of HMM for such

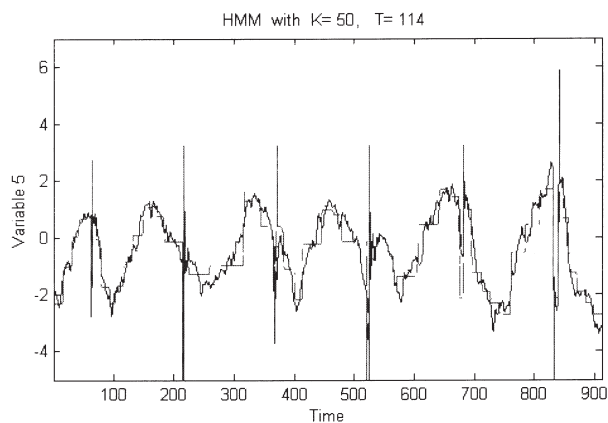


Figure 15. Steam valve signals (solid line) and estimated values (dashed line): holding tube inlet sensor fault (Case III).

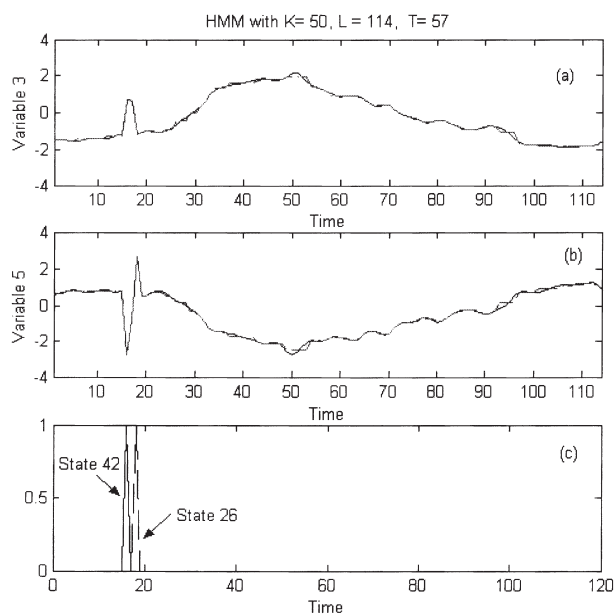


Figure 16. Online fault detection with HMM. (a) holding tube inlet temperature measurements (solid line) and estimated values (dashed line). (b) steam valve signals (solid line) and estimated values (dashed line). (c) Probability trajectories of State 42 (solid line) and State 26 (dashed line)

small changes in the process is a significant advantage.

The predictions of process variables (variable 3 and 5) and probability trajectories of HMM states 42 and 26 for case III are used to illustrate fault detection and diagnosis. As shown in Figure 16 (a) and (b), the predictions of process variables are significantly improved with the sliding window technique. The correlation coefficients r for estimated versus actual values of holding tube-inlet temperatures (variable 3) and steam valve (variable 5) are 0.9981 and 0.9969, respectively. The small magnitude change in the temperature is determined in the probability trajectories of state 42 (decreasing steam valve signals) and state 26 (increasing steam valve signals after the fault in holding tube-inlet) as shown in Figure 16 (c). This is because the less of the observed data is fed

into HMM and the small magnitude fault is not masked by other faults.

HMM as fault detection and diagnosis tool is successful for HTST data because it identifies the faulty status of the system both in method 1 and method 2. Further interpretation of diagnosis results can be achieved by post processing of the information, for example, in a knowledge-based system.

Conclusions

In this study, the HMM technique is used with experimental data to detect and diagnose the faults of a pilot HTST pasteurization system. The methodology used determines first whether the mean values of HMM state variables for each process variable change significantly and take extreme values with respect to in-control HMM states. Based on the results of μ matrix, the corresponding HMM states are examined in terms of their probabilities. The columns of γ matrix (dimension $n \times K$) corresponding to those particular HMM states are plotted to find out at the time of occurrence of maximum probabilities. It was observed that similar faults were represented by a particular HMM state variable depending on their magnitude and duration.

In this food processing example, sensor and actuator faults, which may cause health risks in the end product, must be detected and diagnosed correctly for product safety. HMM provides an aid to classify similar faults in the trajectories of state variables. This can lead to developing a powerful fault diagnosis scheme for HTST pasteurization systems.

Acknowledgements

Financial assistance provided by USDA (NRI Grant 97-35503-4887), the scholarship given by the Turkish Government to Figen Kosebalaban through the faculty member development program, the comments and suggestions provided by Dr. J. Schlessor of FDA and Mr. Bruce Williams are gratefully acknowledged.

Nomenclature

A	state transition matrix
B	observation probability matrix
S	covariance matrix
K	number of HMM states
L	data window length
M	number of symbols in V
O	observation sequence
S_i	i^{th} HMM state
T	length of an observation sequence
V	set of M symbols
$b_f(k)$	elements of B matrix
n	length of data sets
p	number of process variables
r	correlation coefficient

Greek Symbols

α	the forward variable of Baum-Welch algorithm
β	the backward variable of Baum-Welch algorithm
$\gamma_i(t)$	probability of being in state S_i at time t
λ	HMM model
α_{ij}	elements of A matrix
π	initial probability matrix
μ	mean matrix

References

- Cohen, A., "Hidden Markov Models in Biomedical Signal Processing", in "Proc. 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society", Hong-Kong, China

-
- (1998), **20**(3), pp. 1145–1150.
- Ghahramani, Z. and M. I. Jordan, "Factorial Hidden Markov Models", *Machine Learning* **29**, 245–273 (1997).
- Kosebalaban, F. and A. Cinar, "Integration of Multivariate SPM and FDD by Parity Space Technique for a Food Pasteurization Process", *Comp. Chem. Engin.* **25**, 473–491 (2001).
- Kwon, K. C. and J. H. Kim, "Accident Identification in Nuclear Power Plants Using Hidden Markov Models", *Engineering Applications of Artificial Intelligence* **12**, 491–501 (1999).
- Ma, X. and J. Shi, "A New Method for Discrimination Between Fault and Magnetizing Inrush Current Using HMM", *Electric Power Systems Research* **56**, 43–49 (2000).
- Rabiner, L. R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", in "Proc. IEEE" (1989), **77**, pp. 257–286.
- Smyth, P., "Hidden Markov Models for Fault Detection in Dynamic Systems", *Pattern Recognition* **27**, 149–164 (1994).
- Wong, J. C., K. McDonald and A. Palazoglu, "Classification of Abnormal Plant Operation Using Multiple Process Variable Trends", *J. Proc. Cont.* **11**, 409–418 (2001).
- Wong, J. C., K. McDonald and A. Palazoglu, "Classification of Process Trends Based on Fuzzified Symbolic Representation and Hidden Markov Models", *J. Proc. Cont.* **8**, 395–408 (1998).
-

Manuscript received January 14, 2003; revised manuscript received June 16, 2004; accepted for publication November 4, 2004.