

in turns is connected to the filter module via a valve. The filter module contains two sections. The upper section is a single-in, single-out system with the feed at the inlet, and the outlet contains the retentate. The inlet is placed to the side of the filter module while the outlet is above the filter module. This will ensure that the section will be filled with fluid for ultrasonic scan. The lower section is only an outlet which is the permeate side. The membrane is placed between the upper and the lower section. With the diameter of 90 mm, the MF Millipore membrane of $0.5\mu\text{m}$ is used in the experiment. The permeate is collected at a connected tank which is kept at a pressure less than 1 atm by the use of a vacuum pump. The oil is at the permeate side while the retentate side consists of impure oil that needs to be recycled to the feed tank. There is a small water tank above the filter module which facilitates the movement of the transducer when scanning and the transducer are linked to a computer. The feed is $1\text{g}/100\text{ml}$ oil and phosphate and hexane are added at a ratio of 1:3 to lower the energy consumption during filtration [7]. The feed tank temperature is kept at 30°C and the feed flow rate at $11\text{ cm}/\text{s}$. In situ monitoring by ultrasonic wave is carried out for the duration of the filtration.

III. ULTRASONIC DATA

A. Data Structure

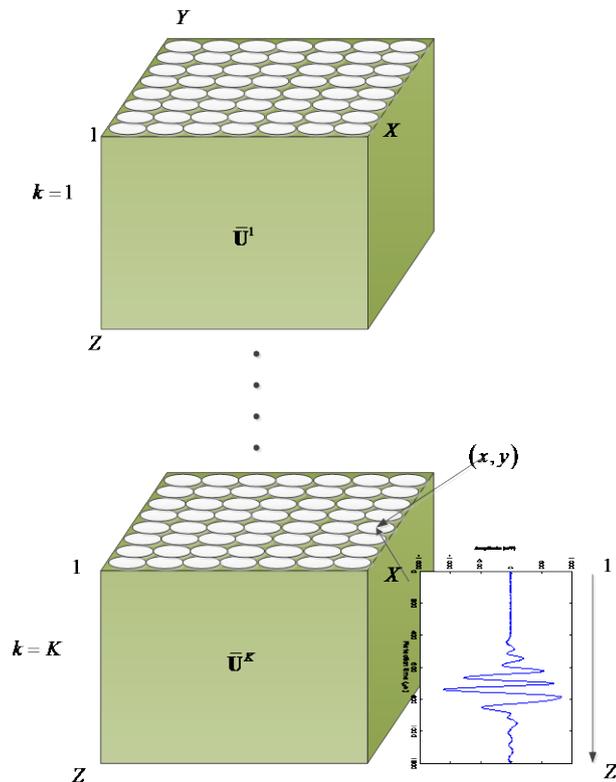


Fig. 2. Data structure of UDTR

During the cross flow filtration, at time k , ultrasonic scan is carried out on each section of the membrane of $200\mu\text{m} \times 200\mu\text{m}$. The time taken for one scan is about 10s. Compared with the total time of the operation of the filtration, it is considered short enough for the error due to time difference in scanning, so the time difference is insignificant.

Under such assumptions, the data in each area of scan of the cross-flow filtration can be represented as shown in Fig. 2. At each sampling time k , x and y represent the positions of the scan with z representing the reflected signal at position (x, y) . The graph on the right side of Fig. 2 shows the signal obtained at time point K at a certain point with the vertical representing the reflection time and the horizontal, the amplitude (mV). When the transducer is fixed at a certain position, the echo signal set collected at the fixed point (x, y) upon the observed zone can be expressed as

$$\mathbf{u} = [u(1), \dots, u(Z)]^T \quad (1)$$

Thus, the data at k sampling is $\bar{\mathbf{U}}^k (Z \times X \times Y)$ and the complete data $\bar{\mathbf{U}} = \{\bar{\mathbf{U}}^1, \dots, \bar{\mathbf{U}}^k, \dots, \bar{\mathbf{U}}^K\}$, forming a 4-dimensional observation data whose size is $Z \times X \times Y \times K$.

B. Wavelet Based Ultrasonic Data

To enrich the frequency resolution from ultrasonic transient signals, a discrete wavelet packet transform (DWPT) is applied here [8]. The basic idea of DWPT is to decompose a time series as a weighted sum of shifted and scaled versions of the wavelets that are suited for capturing the local behavior of non-stationary series, such as sharp changes with different characteristics of frequency at the same time intervals. To do this, the family of discrete wavelets with different scales and time parameters is given by $\psi_{m,n}(z) = a_0^{-m/2} \psi(a_0^{-m}z - nb_0)$, where m, n are integer. The wavelet coefficients are obtained by computing the correlation between the scaled and time shifted version of the wavelets and the analyzed part of the series from the ultrasonic signals. The coefficients in the linear combinations are computed by a factored or recursive algorithm. As a result, expansions in wavelet packet base have low computational complexity. The vector of coefficients at scale j is represented by

$$\begin{aligned} \mathbf{w}_{j,2n} &= \mathbf{H}_0 \mathbf{w}_{j-1,n} \\ \mathbf{w}_{j,2n+1} &= \mathbf{H}_1 \mathbf{w}_{j-1,n} \end{aligned} \quad (2)$$

where \mathbf{H}_0 and \mathbf{H}_1 are the orthonormal wavelet transform matrix for a low pass filter and a high pass filter respectively. They are gotten from a sequence of linear filtering operations. $\mathbf{w}_{0,0} = \mathbf{u}$. \mathbf{u} is the collections of the measured signals at equal space points. $\mathbf{w}_{j,2n}$ and $\mathbf{w}_{j,2n+1}$ are the projections on the high-pass and the low-pass components respectively at scale j . For notational simplicity, we drop the index $\mathbf{x}\mathbf{y}$ of the signal set $(\mathbf{u}_{\mathbf{x}\mathbf{y}})$ here.

To reduce the dimensionality of the feature vectors and provide good class separation, the energies of the wavelet coefficients at scale J are used

$$s_n = \|\mathbf{w}_{J,n}\|, \quad n = 0, 1, \dots, 2^J - 1 \quad (3)$$

which is obtained by calculating the root mean square value of the wavelet coefficients [9].

Thus, each echo signal at the fixed location (x, y) consists of all the wavelet energy features at different frequencies,

$$\mathbf{s} = [s_0 \quad s_1 \quad \cdots \quad s_{2^j-1}]^T \quad (4)$$

These energies are employed as elements of the feature vector. Eq.(4) is calculated only for the fixed location (x, y) in the xy -plane. For the whole plane, the average of $s_n^{x,y}$ among all the locations is applied,

$$\bar{s}_n = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y s_n^{x,y} \quad (5)$$

where X and Y are the number of measured locations in the xy -plane. Thus, the average value for the whole space signal at time k can be written as $\bar{\mathbf{s}}^k = [\bar{s}_0^k \quad \bar{s}_1^k \quad \cdots \quad \bar{s}_{2^j-1}^k]^T$, which can be regarded as a feature pattern that contains both spatial and frequency domain information. For notational simplicity, in the following discussion, the hat at the top of the notations $\bar{\mathbf{s}}$ and \bar{s} is neglected.

IV. DYNAMIC TRANSIENT FEATURE EXTRACTION

A. Classifications

The data are obtained when the cross-flow filtration starts till the blockage occurs, so Eq.(5) yields $\mathbf{S} = [(\mathbf{s}^1)^T, \cdots, (\mathbf{s}^k)^T, \cdots, (\mathbf{s}^K)^T]$ and $\mathbf{s}^k = [s_1^k, \cdots, s_n^k, \cdots, s_N^k]$ for K sampling time points on N band. If I batch operations are carried out, the data containing the normal as well as abnormal operations are $\bar{\mathbf{S}}(I \times N \times K)$. The normal and the abnormal operation data represent the accepted and the unaccepted filtration quality respectively. The normalization of $\bar{\mathbf{S}}$ ensures the data is of the same scale value for different bandwidth. The bandwidth attribute is a numerical value. It implies a large decision tree in the binary tree of C4.5. To reduce the computation load, the observed data is classified qualitatively for different regions using the Gaussian smoothing [10].

The data for different bandwidth $\mathbf{S}_n (I \times K)$ is

$$\mathbf{S}_n = [s_{1,n} \quad \cdots \quad s_{i,n} \quad \cdots \quad s_{I,n}]^T \quad (6)$$

where $s_{i,n}$ represents the data i of band n ,

$$\mathbf{s}_{i,n} = [s_{i,n}^1, \cdots, s_{i,n}^k, \cdots, s_{i,n}^K]^T \quad (7)$$

Since $\mathbf{S}_n (I \times K)$ is of high dimension, PCA is used to

decompose \mathbf{S}_n

$$\mathbf{S}_n = \mathbf{t}_n^1 (\mathbf{p}_n^1)^T + \mathbf{t}_n^2 (\mathbf{p}_n^2)^T + \cdots \quad (8)$$

$\mathbf{p}_n^1, \mathbf{p}_n^2, \cdots$ are the load vectors and $\mathbf{t}_n^1, \mathbf{t}_n^2, \cdots$ are the scores. The first 2 components (\mathbf{t}_n^1 and \mathbf{t}_n^2) are used for analysis because the first two components explain more than 90% of the total variance. Using Haar wavelet as the template and the aforementioned experiment, three-level wavelet packet decomposition and feature extraction is carried out on the obtained data and normalization is done on one bandwidth data of \mathbf{S}_1 . The scores of the first 2 loadings are illustrated in Fig. 3.

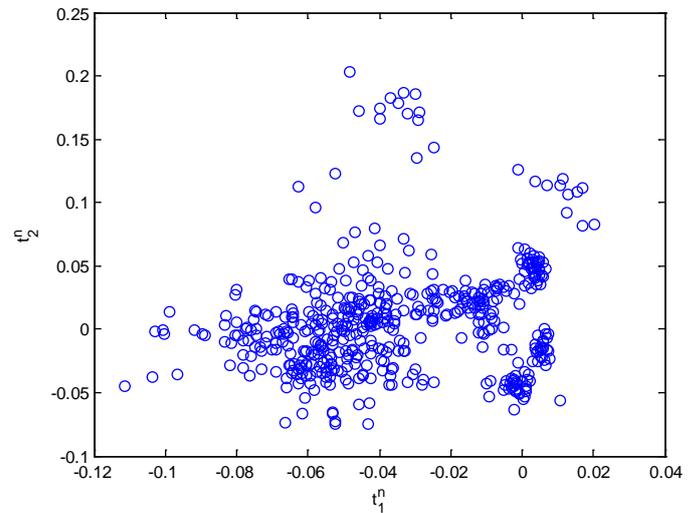


Fig. 3. First 2 scores of \mathbf{S}_1 , $x : t_1^1$, $y : t_1^2$

To identify the group within band n , Gaussian smoothing is used to construct the surface distribution. The effectiveness and feasibility of the proposed auto-clustering method are presented with one frequency which contains all the experimental data in all the operation conditions. Fig. 4(a) shows the exact number of clusters is not clear in the original intensity energy histogram. It is the Gaussian smoothing result at initial $\sigma = 7.51$. In the figure, the solid circle points denote the center of each cluster. In Fig. 4(b), after the first iteration, there are 40 peaks which are reduced subsequently. The spread parameters of smoothing are iteratively updated in Fig. 4(c), in which the input signal is smoothed. Eventually only 8 peaks are obtained, which means that there are 8 clusters at the current band. According to this auto-clustering algorithm, the cluster center $c_{n,c}$ can be computed. The collection of the energy data $s_{n,i}$, $n = 1, 2, \cdots, 2^j - 1$, $i = 1, 2, \cdots, I$ is partitioned into these clusters. Due to the length limitation, the detailed procedures are not explained here, but they will be published in another paper.

The resulting 8 clusters for bandwidth $n = 1$ are A_1, B_1, \cdots, H_1 as shown in Fig. 5. Fig. 6 shows 8 clusters for bandwidth $n = 2$. Thus, the original data for each band can be represented by an array as shown in Fig. 7. At band n , the

numerical value of the original data in $S_n (I \times K)$ is changed to vector $(I \times 1)$ of cluster symbols. This is done for all N band and the data is transformed from $\bar{S}(I \times N \times K)$ to $S(I \times N)$.

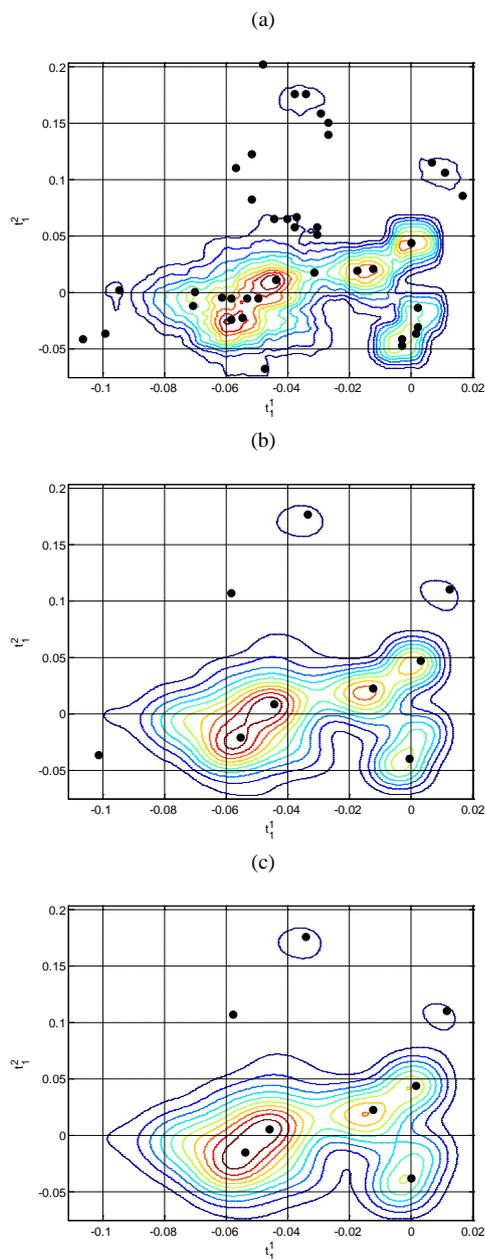


Fig. 4. Gaussian smoothing extracts the proper number of clusters from (a) at $\sigma = 7.51$ to (c) at $\sigma = 7.27$.

A. Rule Extraction Using Decision Tree

After the frequency energies are characterized and the different operation states are identified, it is necessary to find out how to generate knowledge which correlates the frequency energy scales with operational states. To do this, the data structure for the above clustering of all the study classes is shown in Fig. 7. In Fig. 1, given a number of samples, each row is described by a set of attributes.

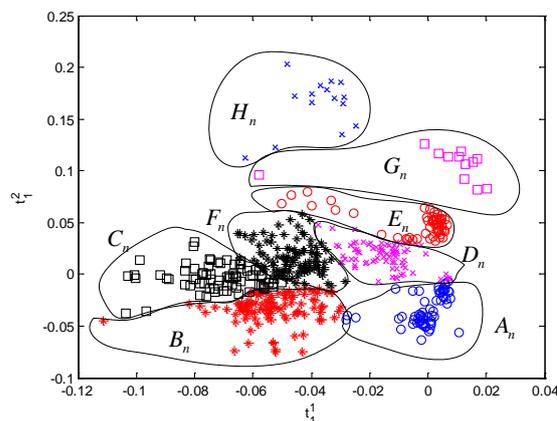


Fig. 5. Clusters in bandwidth $n = 1$

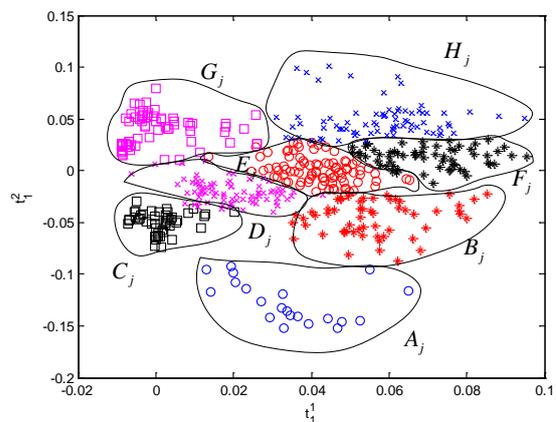


Fig. 6. Clusters in bandwidth $n = 2$

(a)

Sample	n=1				n=2				Class ID
	1	2	K	1	K	
s_1	0.84	0.23	0.62	0.23	1.34	1
s_2	0.41	1.38	-0.22	1.12	-0.95	1
\vdots									\vdots
s_{j-1}	-0.24	0.72	-1.87	0.53	1.22	c
s_j	1.00	4.26	1.69	-0.41	-1.49	c
\vdots									\vdots
s_{j-1}	1.30	1.97	2.42	-1.78	-2.19	C
s_j	1.86	2.10	1.87	-1.01	-1.80	C

(b)

Sample	Attribute		Class ID
	n=1	n=2	
s_1	A_1	D_2	1
s_2	E_1	G_2	1
\vdots	\vdots	\vdots	\vdots
s_{j-1}	D_1	C_2	c
s_j	C_1	F_2	c
\vdots	\vdots	\vdots	\vdots
s_{j-1}	B_1	F_2	C
s_j	F_1	H_2	C

Fig. 7. I batches of S training data with N attributes and C classes. The data structure (a) with numerical values and (b) with the attributes of the membrane filtration for knowledge clustering. The rows represent observations from UTRD; the columns represent frequency scales.

I batches of C class array is used to construct the rules. C4.5 is chosen because it allows different number of branches at each node. The algorithm uses the greatest information gain for partition and consists of two phases, the growth and the pruning of the tree respectively. A classification scheme is designed for grouping the quantitative numbers (shown in Fig. 7(a)) into a number of classes (shown in Fig. 7(b)) so that instances within a class are similar in some respect, but they are distinct from other classes. The frequencies around the operation of a certain membrane filtration process may cover different frequency regions. DT can be built from given attribute sets. The tree can be converted into a set of IF-THEN rules. In contrast to the data-driven models (such as neural networks), one can understand the set of rules while numerical weights of neural networks cannot be easily deciphered.

B. Real-time Fault Monitoring

In this research an real-time monitoring method is proposed. It can promptly detect the error in the operation. The difference here lies with the incomplete data set at a particular time in the operation. At time k , the data collected from the start of the operation is

$$\begin{bmatrix} \mathbf{s}_{new}^1 & \cdots & \mathbf{s}_{new}^k \end{bmatrix} = \begin{bmatrix} s_{new,1}^1 & \cdots & s_{new,1}^k \\ \vdots & & \vdots \\ s_{new,n}^1 & \cdots & s_{new,n}^k \\ \vdots & & \vdots \\ s_{new,N}^1 & \cdots & s_{new,N}^k \end{bmatrix} \quad (9)$$

To complete the data set, the average values from the past operation, $[m_{c,n}^1, \dots, m_{c,n}^k, \dots, m_{c,n}^K]$, is inserted for the future duration.

$$\mathbf{S}_{new+c} = \begin{bmatrix} s_{new,1}^1 & \cdots & s_{new,1}^k & m_{c,1}^{k+1} & \cdots & m_{c,1}^K \\ \vdots & & \vdots & \vdots & & \vdots \\ s_{new,n}^1 & \cdots & s_{new,n}^k & m_{c,n}^{k+1} & \cdots & m_{c,n}^K \\ \vdots & & \vdots & \vdots & & \vdots \\ s_{new,N}^1 & \cdots & s_{new,N}^k & m_{c,N}^{k+1} & \cdots & m_{c,N}^K \end{bmatrix} \quad (10)$$

With this data, the trained decision tree is used to classify the data \mathbf{S}_{new+c} . Thus, the status of the current operation can be induced. In the event that none fall into the class, then new error has occurred and training has to be carried out again. This is repeated for the duration of the operation.

V. EXAMPLES

The proposed method is applied to the cross-flow filtration system to verify its performance. In this system, the feed rate is the most prominent affection of the formation of cake. Too fast a feed flow means the particle will not deposit as readily on the membrane, so the cake growth would be slowed down and the flow has the cleansing effect on the membrane, but the filtrate is more resistant to passing through the membrane. On the other hand, too slow a flow rate means that filtration

performance will be lowered. Concentration of the feed also affects the filtration. If the concentration is too high, cake growth is accelerated and longer filtration time is needed. On the contrary, if the concentration is too low, cake growth is slowed down. The status of the operation is listed in Table 1.

Table 1. Types of operation

Fault Type	Fault description	Symbol
Normal	Normal operation condition	c_1
Fault 1	High inlet flow rate	c_2
Fault 2	Low input flow rate	c_3
Fault 3	High inlet concentration	c_4
Fault 4	Low inlet concentration	c_5

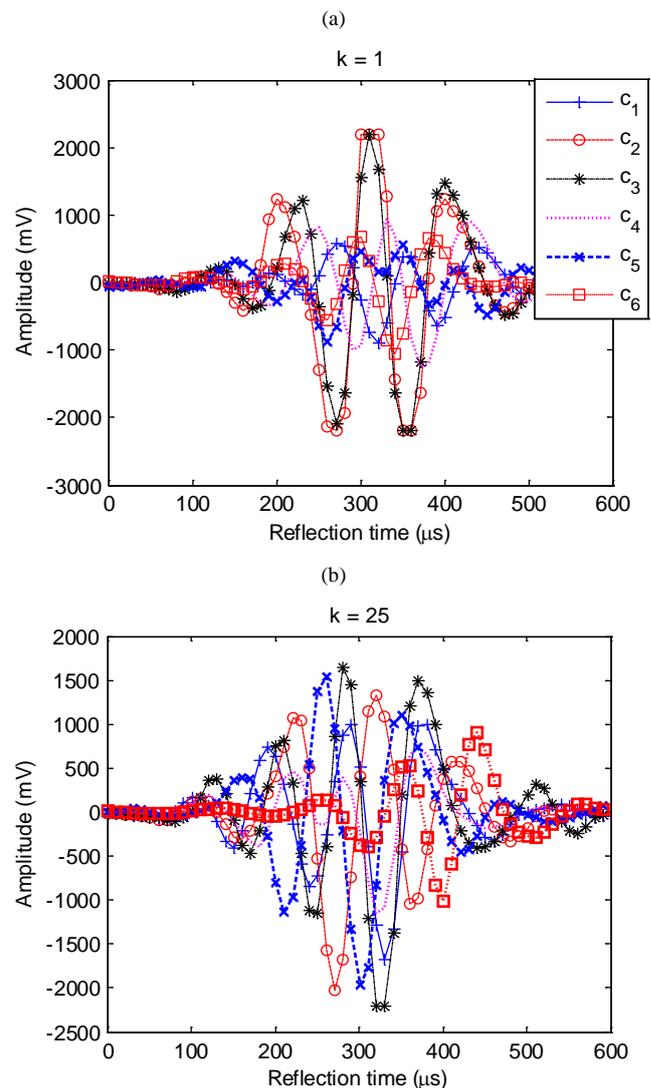


Fig. 8. Six different operations at 4 time points, (a) $k=1$ and (b) $k=25$

Fig. 8 shows the RF signals for 2 different time points in 6 different operations. It is difficult to deduce the error from simply looking at these signals. In this experiment, the sample size ($X \times Y$) is (5×5) with 1,200 observation data. The detailed training procedures are not further discussed due to the space limitation of the paper, and only results in the testing condition are shown here.

With complete training, new cross-flow operation is carried out and data collected. As for real-time diagnosis, the currently unavailable data of the data set is filled with average value as outlined in the previous section. The data are partitioned based on the shortest distance between the training data and the classified centers. Fig. 9 shows monitoring of c_1 type data. It shows the type classification against the sampling time. At $k=1$, c_1 , c_2 , c_4 and c_5 are the possible types. From $k=2 \sim 4$, c_1 and c_4 are the possible types. c_2 and c_5 types have been eliminated because the additional data do not conform to the prediction. At $k=5 \sim 25$, the system is deduced to be c_1 , which is in accordance with the actual operation. The monitoring system, therefore, can be achieved successful classification after 5 samplings. Fig. 10 shows the experimental result of C_4 type. The method takes 8 samplings to replace most types with only type C_1 and C_4 as the possible types. After $k=12$, the method has been able to deduce the correct mode of the operation. From the experimental study, it is found that the proposed method is good at monitoring the cross-flow filtration. The insufficient data at the beginning implies that the classification can only be reliable after certain time and generally the time required is about 1/3 of the operation time. This prompt detection of any error can be beneficial to the operation.

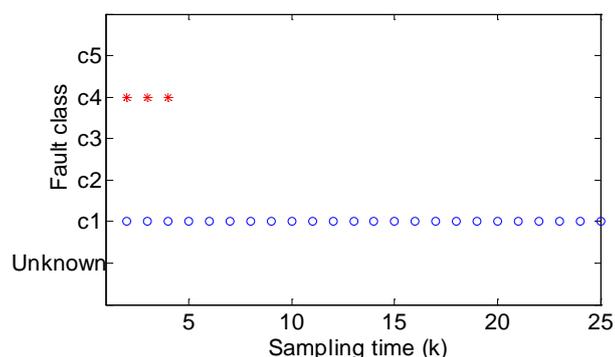


Fig. 9. Real-time monitoring of C_1 type

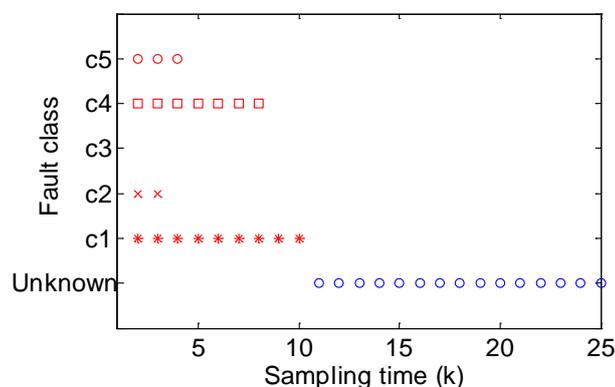


Fig. 10. Real-time monitoring of C_4 type

VI. CONCLUSION

Past research on monitoring of membrane filtration focused on cake formation. There was little diagnosis system developed to detect system anomaly. UTDR technique is capable of membrane fouling noninvasively under realistic

operating conditions. The ultrasonic technique can measure the changes on the membrane surface in a flat-sheet geometry. In this research the data mining technique which is the combination WT and DT from the UTDR signals of the fouling membranes is evaluated. The method is applied to a filtration system with observation carried out on a sample of size 5×5 to obtain the information for the operation. With Haar wavelet as a basis wavelet, packet transform is carried out on the RF signal and feature extracted for different operation modes. PCA and Gaussian smoothing applied to the data lower the classification load using C4.5. It is believed that the problem addressed here has not been fully studied before although it is important in membrane fouling diagnosis. Experimental results show that the proposed method can detect the fault early without waiting till the end point of the whole filtration stage.

REFERENCES

- [1] S. Yao, M. Costello, A. G. Fane, J. M. Pope, "Non-invasive observation of flow profiles and polarisation layers in hollow fiber membrane filtration modules using NMR micro-imaging," *J. Membr. Sci.*, 99 pp.207-216, 1995.
- [2] J. Altmann, S. Ripperger, "Particle deposition and layer formation at the crossflow microfiltration," *J. Membr. Sci.*, 124 pp.119-128, 1997.
- [3] S. Chang, A. G. Fane, S. Vigneswaran, "Experimental assessment of filtration of biomass with transverse and axial fibers," *Chem. Eng. J.*, 87 pp.121-127, 2002.
- [4] D. Hughes, U. K. Tirlapur, R. Field, Z. Cui, "In situ 3D characterization of membrane by yeast suspensions using two-photon femtosecond near infrared non-linear optical imaging," *J. Membr. Sci.*, 280 pp.124-133, 2006.
- [5] A.P. Mairal, A.R. Greenberg, W.B. Krantz, "Investigation of membrane fouling and cleaning using ultrasonic time-domain reflectometry," *Desalin.*, 130 pp.45-60, 2000
- [6] J. Li, D.K. Hallbauer, R.D. Sanderson, "Direct monitoring of membrane fouling and cleaning during ultrafiltration using a non-invasive ultrasonic technique," *J. Membr. Sci.*, 215 pp.33-52, 2003
- [7] J. M.L.N. de Moura, L.A.G. Gonçalves, J.C.C. Petrus, L. A. Viotto, "Degumming of vegetable oil by microporous membrane," *J. Food Eng.*, 70 pp.473-478, 2005.
- [8] Z. Chen, Y. Shi, B. Jiao and H. Zhao, "Ultrasonic nondestructive evaluation of spot welds for zinc-coated high strength steel sheet based on wavelet packet analysis," *J. Mater. Process. Technol.* 209 pp.2329-2337, 2009.
- [9] T. Chang, C.C.J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans Image Process.* 2 pp.429, 1993.
- [10] J. Chen, J. Liu, "Using mixture PCA networks to extract fuzzy rules from data," *Ind. Eng. Chem. Res.*, 38 pp.1478-1488, 1999.