**Novel Panoramic Indicators for Process Operation Stability Assessment through Clustering and Frequency Analysis**

Tetsuya Wada, Yoshiyuki Yamashita

Department of Applied Physics and Chemical Engineering

Tokyo University of Agriculture and Technology

2-24-16 Naka-cho Koganei, Tokyo 184-8588, Japan

Tetsuya.wada@chem-vision.com

**Abstract**

The paper highlights the significance of ensuring stable operation and profitability in a process plant. It introduces new panoramic and frequency analysis-based indicators for evaluating process stability. The study also addresses the challenge of quantitatively assessing operational instability from clustering analysis results. By quantifying instability, priorities for operational stabilization can be set within engineering constraints. The paper further examines how fluctuations in individual variables affect overall process stability by applying wavelet transform to isolate and analyze high-frequency band noise. This approach, validated in several chemical process plants, provides a quantitative tool for assessing operational instability across various process plants.

**Keywords**: Operational Stability, Fault Detection, Machine Learning, Clustering, Wavelet Transformation

1. **Introduction**

Unstable operation can lead to significant fluctuations in operating variables, making it impossible to control product quality near target values. Anomaly detection has been studied in three areas: knowledge-based, model-based, and data-driven studies. Data-driven anomaly detection has been extensively studied and does not require theoretical formulas or experts. Most past research has focused on understanding sudden anomalies, but there are cases where anomalies occur all the time and are the norm. These are called continuous anomalies.

An abnormality in plant operation refers to a condition where the dynamics occurring in plant operation exhibit unusual behavior [1][2]. In general, fault detection in plant operation means detecting abnormalities that occur suddenly in a continuous operating state. These sudden abnormalities can be caused by sticking at control valves, sudden mechanical abnormalities at rotating machines, or sudden blockages in distillation or reaction columns. When detecting this sudden change in behavior, the operating behavior before the sudden change is considered stable, not abnormal. However, abnormalities in operation are not limited to those that occur suddenly. Some equipment continues to operate with abnormalities in operating behavior without any sudden change. This may be caused by insufficient tuning of control system parameters, smaller cavitation and surging at rotating machines, entrainment such as jet-flooding and weeping in distillation columns, as well as distribution abnormalities in reaction columns.

In actual unit operation, more units operate with continuous anomalies than with sudden fluctuations. Therefore, when detecting anomalies related to the operation of a process plant, it is more important to detect anomalies in behavior that occur on a steady basis in operation than to detect anomalies with sudden changes. This regularly occurring operational abnormality is called unstable operation. However, while technology has been developed to detect suddenly occurring abnormalities, little focus has been placed on developing technology to analyze unstable operation. There are no clear criteria for quantitatively indicating stability/unstable operation. Process plant operators have a sensory perception of operational instability, but after a period of time, unstable conditions become ordinary.

In this study, we aimed to quantify the stability/instability of process operation using clustering analysis techniques from time series operation data of multiple variables. Additionally, we attempted to quantify the stability/instability of individual variables in the frequency domain using wavelet transform.

1. **Clustering Analysis Applied to Observe Abnormalities**

The Distributed Control System (DCS) has a limit to the information it can display, and operational stability/instability recognition is generally managed by alarm management with upper and lower thresholds, which are set with empirical consideration. Therefore, instability occurring within the upper and lower thresholds cannot be checked. An indicator that can quantitatively assess operational stability/instability and is independent from the upper and lower thresholds and from the experience or skills of the operator would be very useful for establishing safe and efficient operation through further improvement. In addition, if continuous instability is occurring due to damage of the hardware, etc., evaluation by means of indicators will contribute significantly to identifying the location of such instability.

Clustering analysis, classified as unsupervised learning, has been used to identify abnormal conditions due to sudden changes in operation from historical big data [3]. Attempts have also been made to develop offline monitoring models to identify it [4]. However, while it is used as an anomaly detection technique to find sudden anomalies, it has not been applied to check continuous operational stability/instability.

Clustering analysis is a technique that assigns cluster numbers to data from a group of variables related at different times so that the distance from the center of each cluster is minimized. Fluctuations that occur suddenly are easy to detect because the cluster numbers assigned change significantly. However, it has not been applied to instabilities that occur continuously because there is no significant change in the cluster number.

Our research on several real datasets has found that clustering analysis techniques are effective for the qualitative analysis of continuous instabilities. In stable operations, cluster numbers change infrequently at consecutive times, while in unstable operations, cluster numbers change frequently. Therefore, by graphing the clustering analysis results and judging them visually, a certain degree of stability/instability analysis is possible. However, it may be difficult to make qualitative judgments in the case of multiple datasets with different sampling periods. It is also impossible to quantitatively assess the results of the analysis.

An example of quantitative stability evaluation with clustering analysis is in the field of biometrics, where multiple predictions were made using the Bootstrap Resampling method and the available Gold Standard. The Jaccard Coefficient was used to carry out the evaluation [5]. However, this method requires the Gold Standard to evaluate stability and also evaluates by similarity using silhouette analysis [6]. In other words, it is an evaluation method that assumes stability if the number of data contained in a certain cluster is high and is not suitable for stability analysis of time-series data.

Therefore, we have developed a new evaluation method called Process Operation Stability Indicator (POSI) as an operational stability indicator that can be calculated from clustering analysis results. The advantages of this POSI for stability/instability assessment are:

1. Stability/instability is expressed quantitatively from clustering analysis results,
2. It is independent from the number of variables in the data set and the sampling period,
3. Multiple sections can be assessed using the same criteria, and
4. It can be calculated continuously or at regular intervals to Trend management of stability/ instability is possible.

The -means method is a method to find the cluster that minimize in Equation (1).

where is the mean value in cluster .

In addition, we developed our own Process Operation Stability Indicator (POSI), shown below, as an indicator for evaluating stability/instability.

 *,* where *,*

if the observation lies in cluster , is the number of elements in *V*.

To put it differently, this index is calculated by counting the number of times there is no change in adjacent cluster numbers and not counting when there is a change, based on the results of the cluster analysis calculation, for all time axes in the sampling window. The number of counts obtained from this process is then expressed as a percentage of the total number of time axes. If the value of POSI is large, the behavior of the process operation is relatively stable because the number of adjacent clusters rarely changes.

1. **Clustering Examples**
	1. *Stability/instability evaluation of a column with positive degrees of freedom*

One of the key factors for stable operation of a distillation column is to have zero operational degrees of freedom. This degree of freedom () is the number of disturbances

and constraints and products () minus the number of independent control systems (), which must be .

 

Fig. 1. Zero Degree of Freedom Fig. 2. One Degree of Freedom

For example, if the flow rate and temperature of the feedstock are the disturbances, the levels at the top receiver and bottom of the column are the constraints, and the outflow at the top of the column is the product, then . So, the distillation column needs five independent control systems. In general, the main controls will be control of the feed flow rate, control of the reboiler heat rate, and control of the top of the column pressure. Also, the top product flow and bottom withdraw are used to meet the constraints of the top receiver and bottom level of the column. The reflux flow rate is dependent on the feedstock flow rate and the reboiler flow rate and cannot be used to calculate the degrees of freedom.

Figures 1 and 2 show the results of the clustering analysis and POSI calculations for a distillation column operating under similar conditions. Figure 1 shows the results for a distillation column with top pressure control (zero degrees of freedom), and Fig. 2 shows the results for a distillation column without top pressure control (one degree of freedom). When the degree of freedom is zero, the operation of the distillation column is normally very stable, but when the degree of freedom is set to one or positive value, the operation of the distillation column becomes very unstable.

* 1. *Stability/instability analysis of continuous BTX columns*

Figure 3 shows three consecutive BTX distillation columns (Benzene, Xylene, and Xylene columns) that separate each component from a mixture of Benzene, Toluene, Xylene, and AC9+. The column top pressures are 0.002 MPaG, 0.005 MPaG, and 0.016 MPaG, respectively. The top pressure of each column is maintained by injecting hydrogen into the top section. The same amount of hydrogen is injected into each of the three columns. Figures 4, 5 and 6 display the results of the Clustering analysis and POSI calculations for these three distillation columns. From these POSI results, the operation of the Benzene column is stable, while the Toluene and Xylene columns are unstable.



Fig.3. BTX Separation Columns

  

Fig.4. Benzene Column Fig.5. Toluene Column Fig.6. Xylene Column

1. **Frequency Analysis with Wavelet Transformation**

“Wavelet transformation” is often used to detect anomalies in individual variables [7][8]. In particular, it is still the subject of much research in the detection of anomalies in rotating equipment, including in combination with modern techniques such as machine learning. However, the main focus of most studies is to use wavelet transformation as a filter to find the frequency band containing fault features.

In process operation, it is not only rotating equipment that exhibits fluctuating features in specific frequency bands. Similar characteristics appear in unstable control systems and the variables affected by these control systems. When these characteristics appear in the high frequency band, they are often due to operation beyond the capacity of the equipment or inadequate adjustment of control parameters.

This indicator is calculated using the following formula. This indicator is also named Frequency Analysis Stability Indicator (FASI). The Daubechies wavelet was applied as the orthogonal wavelet in this study, but no significant differences were found in the calculation results using other orthogonal wavelets.

where is the weighted factor and is the standard deviation in decomposed noise at **n**th decomposition.

1. **Example**

Table-1 shows the results of FASI calculations for a group of variables related to a certain distillation column: a high FASI value which means that there is a possibility of periodic fluctuations. If the variable is a control system and the FASI value is high, it is necessary to tune the parameters first. Also, if the FASI value of the Indicator is high, it is likely to be affected by the control system.

Thus, the periodic instability of the control system can be identified from the FASI calculation, so it is possible to identify which parameters of which variables should be tuned. In addition, as the instability can be quantified, it is easy to prioritise which variables should be considered first.

**6．Conclusion**

In process operation, abnormality detection is commonly used to manage the trend of operating data for each variable with alarms of upper and lower limit values. However, this method cannot detect operational abnormalities that can be detected with operational

Tabel 1 FASI Calculation Example



data of multiple variables or that occur in a specific frequency band. Furthermore, no technology has been developed to practically quantify operational stability, including abnormality detection by trend.

POSI, described in the previous section, can identify and quantify abnormalities that occur routinely in plant operation from a bird's-eye view of the operating data of several relevant variables. FASI can also identify and quantify cyclical anomalies in individual variables. By using these two indicators, it is possible to provide operators and engineers with quantitative information on operational stability/instability, which was not possible with the conventional alarm management of upper and lower limits. If the causes of operational instability are identified and countermeasures taken before alarms occur, it is possible to nip abnormalities in the bud before alarms occur, and achieving incident-free operation is not just a dream.

**References**

[1] D. M. Himmelblau, Fault detection and diagnosis in chemical and petrochemical processes, Elsevier Scientific Pub. Co., Amsterdam, 1978.

[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S. N. Kavuri, “A review of process fault detection and diagnosis Part I: Quantitative model-based methods,” *Computers and Chemical Engineering*, Vol 27, pp. 293-311, 2003.

[3] I. A. Udugama, C. L. Gargalo, Y. Yamashita, M. A. Taube, A. Palazoglu, B. R. Young, K. V. Gernaey, M. Kulahci, C. Bayer, “The Role of Big Data in Industrial (Bio)chemical Process Operations,” *Industrial & Engineering Chemistry Research,* Vol. 59, no. 34, pp.15283-15297, 2020.

[4] M. C. Thomas, W. Zhu, J. A. Romagnoli, “Data mining and clustering in chemical process databases for monitoring and knowledge discovery,” *Journal of Process Control*, Vol. 67,
pp. 160-175, 2018.

[5] J. A. Bernabé-Díaz, M. Franco, J.-M. Vivo, M. Quesada-Martínezc, J. T. Fernández-Breis, “An automated process for supporting decisions in clustering-based data analysis,” *Computer Methods and Programs in Biomedicine*, Vol. 219, 106765, 2022.

[6] P. J. Rousseuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53-65, 1987.

[7] R. Yan, R.X. Gao, X. Chen, “Wavelets for fault diagnosis of rotary machines:
A review with applications,” *Signal Processing*, Vol. 96, pp. 1–15, 2014

[8] T. Guo, T. Zhang, E. Lim, M. Lopez-Benitez, F. Ma, L. Yu, “A review of wavelet analysis and its applications: challenges and opportunities,” *IEEE Access*, Vol.10, pp.58869-59803, 2022