Title	Reliability of operating window identified from process data
Author(s)	Ihalainen, H.; Ritala, R.; Saarela, Olli
Citation	23rd European Symposium on Computer Aided Process Engineering, ESCAPE 23, 9 - 12 June 2013, Lappeenranta, Finland
Date	2013
Rights	This conference paper may be downloaded for personal use only.

Andrzej Kraslawski and Ilkka Turunen (Editors) Proceedings of the 23rd European Symposium on Computer Aided Process Engineering – ESCAPE 23 June 9-12, 2013, Lappeenranta, Finland,

Reliability of operating window identified from process data

Heimo Ihalainen[a], Risto Ritala[a], Olli Saarela[b]

[a]*Tampere University of Technology, Department of Automation Science and Technology, P.O.Box 692, FIN-33101 Tampere, Finland*
[b]*VTT Technical Research Centre of Finland, P.O.Box 1000, FIN-02044 VTT, Finland*

Abstract

This paper discusses data-based operating windows as a tool for process management and development. In particular identification of the operating window and its uncertainty are analyzed. The operating window is determined by maximizing either the mutual information (static) or entropy transfer (dynamic). An industrial example shows that the entropy of the indicator variable is reduced to half by an operating window specified with only few variables, selected amongst over 3000 candidates. Test model based simulations suggest that such few-variable operating windows can be reliably identified from datasets having lengths of a few thousand observations.

Keywords: process management, operating window, mutual information, transfer entropy, uncertainty

1. Introduction

Optimization of industrial processes is a multifaceted task requiring multiple techniques. Simulation models based on first principles, deployed together with steepest gradient optimization have proven effective in improving, e.g., thermal efficiency. However, accommodation of process imperfections (such as wear and fouling) and variability in raw materials would be prohibitively laborious using that approach. For taking such factors into account the first principles approach can be supplemented by a data driven approach. Promising results have been achieved by identifying favorable operating windows from measurement data.

An operating window for a process is a logical AND rule: if all the process variables chosen to define the window are within their given ranges then the system is deemed to be operating in a satisfactory manner. Statistical Process Control has been one way of generating the operating window: the satisfactory range is defined for all process data and the range is derived from the statistical behavior of the process in the normal state. For example, if variables are within three standard deviations from their mean, the process is said to be operating in satisfactory manner, else the data is taken as indication of abnormal behavior.

In this paper we consider operating window based on process knowledge. The process knowledge is expressed as a binary indicator signal according to process experts' judgment. Then the set and ranges of key process variables are chosen, and a binary signal is formed by an AND rule of these thresholded key process signals. The set and ranges are chosen such that the AND-rule binary signal is most informative about the indicator signal. Let Y be the scalar indicator variable and let X be the real valued, typically high-dimensional process data. Let Z be the binarization of process data with some collection of ranges $\{R_i\}$. The candidate operating window signal is thus:

$$Z_n = \begin{cases} 1 & \forall i \in I : X_{n,i} \in R_i \\ 0 & else \end{cases} \qquad (1).$$

Thus the set I and the corresponding set of ranges are to be identified.

For a process operator the window is expressed as a set of upper and lower values for continuous-valued process tags (setpoint, measurements, controller output and other), e.g.

$$80 < 24\text{T}003 < 110 \quad \& \quad 20 < 24\text{F}007 < 25 \qquad (2).$$

When all inequalities are true, the system is statistically significantly more probably in a favourable state according to the binary indicator signal Y discussed above. .

Data analysis facilitates identification and remediation of many sources of process disturbances. However, an industrial process is affected by numerous disturbance sources, many of them effective only too rarely for statistical analysis. Many rarely occurring process disturbances can be too costly to address individually. Identification of favorable operating windows supports running the process in state robust against disturbances, reducing the need to address each disturbance source separately.

This paper discusses the reliability of operational windows identified from process data, an area mostly ignored despite the method being industrially applied. In Section 2, methods of identifying operational windows are discussed. Section 3 presents simulation results on the operational window reliability, and Chapter 4 describes an industrial case study. Conclusions are presented in Chapter 5.

2. Methods for determining the operating window

For a static system the quantitythat the operating window is to maximize is the mutual information between indicator Y and operating window Z. For a fixed set of process signals, ranges are chosen as (Au et al, 2003):

$$\{R_i\} = \arg\max_{\{R_i\}} MI(Y, Z; \{R_i\}) = \arg\max_{\{R_i\}} \sum_{k,l=0}^{1} p(Y=k, Z=l) \cdot \log\left(\frac{p(Y=k, Z=l)}{p(Y=k) \cdot p(Z=l)} \right) \qquad (3).$$

In order to determine I, this problem is to be solved for all combinations of signals. As the informativeness increases with the number of signals and as having all the signals affecting the operating window signal is rather impractical, the choice of I is a tradeoff between informativeness achieved and number of signals in the set. Even for a moderate maximal number of signals in I, this is a huge combinatorial task. Thus the maximum number of signals is usually limited to a small number, e.g. 3-10 and the combination with highest mutual information is chosen. Domain knowledge, e.g. about the correlations between process signals, can be applied to reduce further the combinations of signals considered.

Operating window (3) does not take into account the process dynamics, but can be applied successfully for sufficiently stationary processes in which changes from time step to time step are small. The simplest approach towards dynamics is to deal with delays only, thus replacing (1) by

$$Z_n = \begin{cases} 1 & \forall i \in I : X_{n-N_i,i} \in R_i \\ 0 & else \end{cases} \quad (1').$$

where signal delays N_i are additional degrees of freedom in MI optimization, and only a single N_i is chosen for each signal *i*.

Transfer entropy (Shreiber, 2000) takes dynamics into account in more detail. Let Z_i be the *i*th binarized process signal according to some ranges R_i, and let *Z* be the vector signal of Z_is. Then the transfer entropy with time window *N* is

$$TE(Y;Z,N) = S(Y_n \mid Y_{n-1},..,Y_{n-N}) - S(Y_n \mid Y_{n-1},..,Y_{n-N}; Z_{n-1},..,Z_{n-N}) \quad (4),$$

where *S(V|W)* is the conditional entropy of a random variable *V* given a variable *W*. For a set of signals *I* the operating window $\{R_i\}$ is then the one maximizing *TE(Y;Z)*. We seek the signal set I and their ranges for *Z* such that knowing the *N*-history of *Z* maximally reduces the uncertainty about *Y*. Note that transfer entropy is the amount of additional information compared to knowing the time series of indicator signal. Thus transfer entropy may be smaller than mutual information that uses the instantaneous probabilities – rather than time series – as the entropy reference value.

Quite commonly the operating windows defined through (3) are based on process properties physically and dynamically close to *Y*, whereas the root causes for *Y*, further away, are hard to discover. Then relationships between process signals can be further analyzed, e.g. with linear multivariate time series analysis, which seeks explanations to signals prominent in *Z*. This corresponds to applying the idea of transfer entropy, (4) in the linear-Gaussian system with *X* signals present in *Z* taking the role of the random variable whose transfer entropy is to be minimized on the basis of process signals not present in *Z*. A scheme of hypothesizing and excluding explanatory variables in a large set of process signals has been developed for process diagnostics (Saarela, 2002).

3. Uncertainty analysis with synthetic data

A key question when setting up the set of signals and their ranges for an operating window is the accuracy of the information and range estimates. This obviously depends on the amount and quality of data available. This section analyses an extremely simple model system to study the effect of data size and the degree and dynamics of the dependence between the process signals and the indicator.

Let us consider a set of four process variables that constitute a linear multivariate autoregressive system. The task is to find the most important process variable and its binarization such that it is most informative about the indicator variable. The system is as follows:

$$\begin{aligned} &X(n+1) = AX(n) + E(n) \\ &E(n) \sim N(0, \sigma^2 \cdot Id) \\ &P(Y(n)=1) = \frac{1}{1+\exp\left(-\gamma^{-1} \cdot (|X_k(n-d)| - X_{th})\right)} \end{aligned} \quad (5),$$

Thus the binary probability of the indicator is determined by one process variable through a sigmoid with threshold parameter X_{th} and steepness parameter γ. Two instances of this model are considered. The process signal dynamics for the cases are determined by the model matrices:

$$A^{(1)} = \begin{bmatrix} 0.8 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0 & 0.1 & 0.8 \end{bmatrix}; \quad A^{(2)} = \begin{bmatrix} 0 & 0.1 & 0 & 0.8 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.1 & 0.8 \end{bmatrix}$$

For the Model 1 $k=1$ and $d=1$, for the Model 2 $k=2$ and $d=2$. In both cases $\sigma^2=0.1$. In the first case the operating window should be set by thresholding X_1, whereas in the second by thresholding X_4. However, in the Model 2, the shortest delay dynamics is between Y and X_1 as both strongly determined by X_4 with lag 2 and 1 dynamics, respectively. The two models were studied with sigmoid parameters: $X_{th}=1$, $\gamma=0.02$; $X_{th}=0.6$, $\gamma=0.02$; and $X_{th}=1$, $\gamma=0.1$ with synthetic data of lengths M = 500, 1000, 2000, and 4000. The maximum of mutual information/transfer entropy, resp. position of the maximum as the threshold estimate, with respect to cause variable threshold was determined based on 10 samples in each case, and the estimates and their standard deviations were determined from the data. All the information/entropy variables are given in nats. A reference in the information/entropy scale is provided by assuming stepwise relationship between the binary probability of Y and the affecting process signal. Then for $X_{th}=1$ the total entropy of Y is 0.39 nats and for $X_{th}=0.6$, 0.66 nats.

Table I summarizes the maximal mutual information/transfer entropy and the thresholds providing this maximum for the Model 1 and the three sigmoid parameters and for Model 2 with $X_{th}=0.1$, $\gamma=0.02$. The data length M=4000. The maximum MI/TE values are to be compared with those variables not affecting the indicator Y; these values are in all cases of the order of 0.01. Table I shows that with Model 1 and M=4000, the correct signal is chosen for operating window and its optimal threshold estimate is stable for all TE, whereas the MI threshold values are somewhat smaller due to the delay from X_1 to Y. With Model 2 TE($Y;Z,N$) for N=2 and larger, the thresholds are stable.

Table I. Maximal information and optimal thresholds

	Model 1; thresholding X_1						Model 2; X_1		Model 2; X_4	
	$X_{th}=1$, $\gamma=0.02$ (S(Y)=0.39)		$X_{th}=1$, $\gamma=0.1$ (S(Y)=0.39)		$X_{th}=0.6$,$\gamma=0.02$ (S(Y)=0.66)		$X_{th}=1$, $\gamma=0.02$ (S(Y)=0.39)			
	max	thr_{opt}	Max	thr_{opt}	Max	thr_{opt}	Max	thr_{opt}	max	thr_{opt}
MI	0.039	0.94	0.043	0.92	0.063	0.77	0.030	0.97	0.029	0.93
TE($Y;Z$,1)	0.087	1.01	0.061	0.97	0.152	0.62	0.020	0.93	0.024	0.92
TE($Y;Z$,2)	0.097	1.00	0.065	0.92	0.187	0.60	0.023	0.80	0.083	1.00
TE(Y;Z,3)	0.099	1.00	0.070	0.94	0.194	0.60	0.030	0.78	0.097	0.99

Table II shows the threshold uncertainties (as standard deviations amongst 10 samples) for Model 1, sigmoid parameters $X_{th}=1$, $\gamma=0.02$; and $X_{th}=1$, $\gamma=0.1$. Results in Table II show the importance of dynamics for the accuracy of the thresholding. Although in model 1 there is only lag 1 in dynamics and the change in X is rather low, MI accuracies are on the same order of as for TE only for large data sets M = 4000. For the case $\gamma=0.1$, in which the degree X determines the indicator variable is lower, larger data sets are needed for TE($Y;Z$,1), whereas for higher order TE, even the modest amount of data M=500 appears sufficient for obtaining an accurate estimate for the optimal threshold.

Table II. Uncertainty of the optimal threshold, standard deviation between 10 samples, as a function of data length. Model 1, X_{th}=1, γ=0.02 (results denoted method/1), 0.1 (method/2).

M	MI/1	TE1/1	TE2/1	TE3/1	MI/2	TE1/2	TE2/2	TE3/2
500	0.33	0.013	0.022	0.017	0.30	0.121	0.056	0.082
1000	0.30	0.017	0.020	0.020	0.30	0.061	0.063	0.071
2000	0.147	0.016	0.015	0.019	0.115	0.051	0.042	0.032
4000	0.089	0.015	0.020	0.013	0.080	0.054	0.050	0.060

4. Industrial case study

In an industrial case study data from a paper mill was analyzed. The goal of this study was to find favorable operating windows with respect to web breaks. A multitude of disturbances can cause web breaks. Thus an operating window, robust against many disturbances is desirable. A performance indicator was formed based on the number of web breaks in a 24 hour time window, eliminating transient situations. Figure 1 depicts the original web break signal and the performance indicator for a one week period.

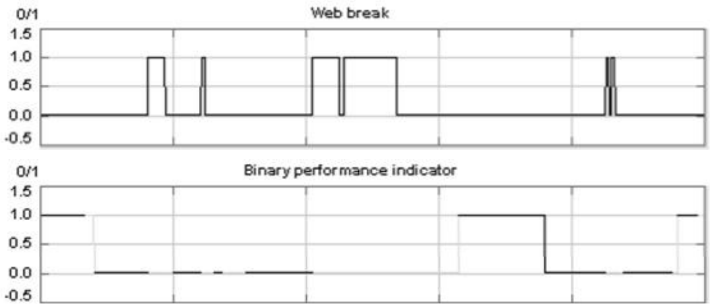

Figure 1. Example of break signal and the derived binary performance indicator for one week of process time. Transient situations and the durations of the web breaks have been excluded from the analysis.

The data included 1573 measurement and set point time series from the process. As the goal was to find operational windows that could be maintained for an extended period of time, a time resolution of one hour was deemed adequate. Both hourly averages and hourly standard deviations were included, making the number of variables in the search space 3146. The time span of the data set was one year, M=4640 after removal of transient situations, web breaks, and shutdowns. Results from Table II indicate that operating window thresholds can be identified for univariate models to an accuracy of the order of 0.1 times the standard deviation of the variable in the data set. An evolutionary algorithm was used to search for favorable operating windows judged by MI within this search space. Operating windows defined by 1-6 variables were searched. With the evolutionary algorithm the search could be carried out in roughly $O(n \log n)$ time complexity with respect to the number of variables in the search space, facilitating the analysis to include a large number of process variables.

The time resolution of the analysis, selected for identification of favorable long-term operating windows, did not facilitate evaluation of causality. This was utilized as a partial validation of the analysis technique: the known consequences of web breaks (e.g., surface level in the broke tower) were included in the search. These known causes were systematically included as variables in the identified operational windows, as analysis was carried out for different paper grades and for different time intervals. This supports the reliability of the analysis results.

In the final analysis known consequences of the web breaks were excluded. The two process variables discovered most significant were hourly averages of a line pressure in the 1st press section and the output of the controller for a steam group pressure, both known a priori to affect the drying profile of the paper web, and consequently its strength at different parts of the process. Table III summarizes the operating hours inside (all the inequalities true) and outside (at least one inequality false) the operating window identified by these two variables. This operational window has been able to capture nearly all favorable values of the binary performance indicator, while including only 20% of the unfavorable values. For 12 hours of process operation the truth values of the inequalities could not be determined, due to missing measurement data.

Table III. Operating hours inside/outside of the identified operational window for a paper grade.

	Process in window	Process out window	Process unknown
Indicator = 1	219	4	3
Indicator = 0	127	527	9

The entropy of the indicator is 0.57 nats and the mutual information of the indicator and process window is 0.28 nats. Based on Section 3, the maximal mutual information is high enough to exclude accidental contributing signals from the operating window.

5. Conclusions

Operating windows supplement process diagnostics. As a data analysis task the identification of such favorable operating windows has desirable characteristics: identification is relatively insensitive to outlier values and the analysis results are easily communicated to, and evaluated by process personnel.

According to the results in Section 3, if the operating window consists of only few signals, the thresholds can be identified with adequate accuracy from data sets consisting of a few thousand observations. Taking dynamics into account by maximizing transfer entropy instead of mutual information increases thresholding accuracy and provides hints of root causes. If the operating window is to be identified amongst several thousand candidates, one should be cautious in analyzing information contributions and identifying thresholds. The results should always be analyzed also from domain knowledge perspective. When operating windows are identified with evolutionary algorithms, the method provides several window candidates amongst which the final choice can be made based on process knowledge.

References

Wai-Ho Au, K. C. C. Khan, X. Yao, A Novel Evolutionary Data Mining Algorihtm With Appliation to Churn Prediction, IEEE Trans. Evolutionary Computation, 7, (6):532-545 (2003).

O. Saarela, 2002. Multivariate Autoregressive Analysis in Locating the Origin of Fluctuation in Continuous Industrial Processes. Thesis (PhD), Tampere University of Technology, 137 p.

T. Schreiber, Measuring Information Transfer, Phys. Rev. Lett., 85, 2:461-4 (2000).