# Maximal information-based nonparametric exploration of condition monitoring data

Yang Hu[1], Thomas Palmé[2], and Olga Fink[1*]

[1]*Zurich University of Applied Sciences, Rosenstr. 3, Winterthur, 8401, Switzerland,*

*yang.hu@zhaw.ch*
*olga.fink@zhaw.ch*

[2]*General Electric (GE) Switzerland, Brown Boveri Str. 7, Baden, 5401, Switzerland*

*thomas.palme@ge.com*

## ABSTRACT

The system condition of valuable assets such as power plants is often monitored with thousands of sensors. A full evaluation of all sensors is normally not done. Most of the important failures are captured by established algorithms that use a selection of parameters and compare this to defined limits or references.

Due to the availability of massive amounts of data and many different feature extraction techniques, the application of feature learning within fault detection and subsequent prognostics have been increasing. They provide powerful results. However, in many cases, they are not able to isolate the signal or set of signals that caused a change in the system condition.

Therefore, approaches are required to isolate the signals with a change in their behavior after a fault is detected and to provide this information to diagnostics and maintenance engineers to further evaluate the system state.

In this paper, we propose the application of Maximal Information-based Nonparametric Exploration (MINE) statistics for fault isolation and detection in condition monitoring data.

The MINE statistics provide normalized scores for the strength of the relationship, the departure from monotonicity, the closeness to being a function and the complexity. These characteristics make the MINE statistics a good tool for monitoring the pair-wise relationships in the condition monitoring signals and detect changes in the relationship over time.

The application of MINE statistics in the context of condition monitoring is demonstrated on an artificial case study. The focus of the case study is particularly on two of the MINE indicators: the Maximal information coefficient (MIC) and the Maximum Asymmetry Score (MAS).

MINE statistics prove to be particularly useful when the change of system condition is reflected in the relationship between two signals, which is usually difficult to be captured by other metrics.

## 1. INTRODUCTION

Monitoring the condition of complex systems is usually achieved by analyzing the signals collected by a number of sensors located in different parts of the system. The obtained condition monitoring data is typically high dimensional and single signals or extracted features are characterized by different degrees of correlation to other signals. For many impending faults, the first indication is changes in the relationship between the signals and can be evaluated with correlation analysis. Correlation analysis is, therefore, often applied to assess the relationship between the signals and also to detect a change in the relationship (Dai & Gao, 2013).

Many of the applied correlation methods show several limitations and drawbacks. They are either limited to linear relationships, e.g. the Pearson correlation coefficient, or monotonic functions, e.g. the Spearman coefficient, or sensitive to outliers and noise in the data. Information theoretical approaches overcome some of the limitations of the previously mentioned approaches. They are not limited to linear relationships and are not sensitive to outliers (Ando & Suzuki, 2006; Wu & Wang, 2013). However, they are either difficult to interpret or highly dependent on the approach used to approximate the underlying distributions.

Mutual information has been applied in fault detection in several applications (Jiang & Yan, 2014; Kappaganthu & Nataraj, 2011; Verron, Tiplica, & Kobi, 2008). Mutual information has been either applied for feature selection such as in (Jiang & Yan, 2014) and (Kappaganthu & Nataraj, 2011) or plant-wide process monitoring as in (Verron et al., 2008). Recently, a new approach has been introduced to overcome some of the existing limitations: maximal information coefficient (MIC) and the pertinent maximal information-based nonparametric exploration that covers additional coefficients (Reshef et al., 2011).

The approach is not sensitive to noise or outliers. It can be directly interpreted and compared since it is normalized to values in the interval [0,1]. Additionally, it is not limited to monotonic functions but is able to detect different types of functional relationships. While MIC captures the strength of the relationship, a further interesting parameter as part of the maximal information-based nonparametric exploration is the Maximum Asymmetry Score (MAS). MAS captures the departure from monotonicity and is able to detect non-linear relationship in the data.

Particularly, these two metrics (MIC and MAS) are useful for applications of condition monitoring data. They can be applied to extract the relationships between different features and also detect the changes in the relationship that cannot be detected by commonly applied metrics.

In this paper, the application of these two metrics is demonstrated on an artificial case study in which different types of relationships between signals are introduced. The metrics and their performance are compared to the Pearson correlation coefficient.

The proposed approach proves to be a very useful tool for condition monitoring applications. It can be applied not only to detect relationships between two signals or features, but also to monitor the evolution of a feature over time and compare it to the previously observed patterns. The proposed approach can be particularly useful for fault detection of the complex industrial equipment, in which faults are characterized by changes in highly non-linear relationships of the high-dimensional condition monitoring data.

## 2. THEORETICAL BACKGROUND ON MAXIMAL INFORMATION-BASED NONPARAMETRIC EXPLORATION

Generally, mutual information of two random variables is an information theoretical measure and quantifies the mutual dependence of these two variables (Kraskov, Stögbauer, & Grassberger, 2004). It is based on the concept of entropy of a random variable. Mutual information quantifies the amount of information gained about one variable given that the other variable is known. Generally, if two variables are independent mutual information will be zero and if the dependence between two variables is large, mutual information will be large.

Different approaches have been introduced to calculate the mutual information between the variables, such as by the histogram method (Fraser & Swinney, 1986) or by the kernel density estimators (Moon, Rajagopalan, & Lall, 1995). One of the limitations is that different approaches provide different results and are therefore difficult to interpret. Additionally, mutual information is not normalized and may be difficult to compare between different types of relationships.

The Maximal information criterion is based on the assumption that for any relationship, there exists a grid that can capture the relationship between the two variables. Therefore, the MIC algorithm examines different x-y-grid combinations and determines the grid with the largest possible mutual information that can be achieved by any x-y-grid. To find the grid with the largest mutual information, all grids up to a maximal grid resolution are evaluated, dependent on the size of the sample, computing for every pair of integers (x,y) the largest possible mutual information achievable by any x-by-y grid applied to the data (Reshef et al., 2011). This is contrary to the other mutual information computations which only consider the mutual information in the given orthogonal x-y-grid and do not vary different number of rows and column combinations.

Given the data $D$, $I^*(D,x,y)$ is the maximum mutual information achieved by any grid with x columns and y rows on the data $D$:

$$I^*(D,x,y) = \max I(D|_G) \tag{1}$$

The characteristic matrix $M$ of the set of data $D$ is given by the maximum mutual information achieved by any grid with x columns and y rows and normalized by the maximal possible mutual information $\log\min\{x,y\}$ :

$$M(D)_{x,y} = \frac{I^*(D,x,y)}{\log\min\{x,y\}} \tag{2}$$

Thereby, the mutual information is normalized to be in the scale between 0 and 1 which enables a comparison between different types of relationships. The maximal information criterion is then defined as the maximum value in the characteristics matrix M:

$$MIC(D) = \max_{xy<B(n)} \{M(D)_{x,y}\} \tag{3}$$

Where the bin size is determined by $\omega(1) < \mathrm{B(n)} \le \mathrm{O(n^{1-\varepsilon})}$ for some $o < \varepsilon < 1.$

Maximum asymmetry score measures the deviation from monotonicity and is given by:

$$MAS(\mathrm{D}) = \max_{xy<B} \left| M(D)_{x,y} - M(D)_{y,x} \right| \tag{4}$$

For more details of the calculation of MICA and MAS, interested reader can refer to (Reshef et al., 2011).

## 3. CASE STUDY IN FAULT DETECTION

### 3.1. Applied datasets

To test the application of Maximal Information-based Nonparametric Exploration (MINE) statistics in condition monitoring applications, we apply them on datasets where we know the actual relationship between the signals, which will be referred as the "ground truth". This provides us an opportunity to interpret the results correctly. We generate three different types of faults to evaluate the performance of the algorithm in different types of changes in the relationships between the signals.

Each of the datasets comprises 2000 patterns. Whereby, the first 1000 patterns are in the normal state and the following 1000 patterns are in a faulty state. The three different fault types are the following:

1) A fault characterized by an abrupt change in the relationship between two signals
2) A fault characterized by a gradual drift in a single signal
3) A fault in which one independent signal is not affected by the fault, but all the dependent signals are affected by a gradual drift.

The equations for generating the normal state signals are shown in Table 1.

Table 1. Equations for normal state signal generation

| | |
|---|---|
| $S(i,1) \sim N\left(0, 0.5^2\right)$ <br> $i = 1, 2, \ldots, 2000$ | $S(i,2) = aS(i,1) + b$ |
| $S(i,3) = c * \cos\left(dS(i,1)\right)$ | $S(i,4) = eS^3(i,1) + fS^2(i,1)$ <br> $+ gS(i,1)$ |
| $S(i,5) = l * \exp\left(hS(i,1)\right)$ | a=2, f=-1.006, b=5, g=0.36, c=2, h=0.2, d=10, l=0.06, e=0.6 |

In Table 1, $S(i,j)$ represents the $i$-th pattern of signal $j$. Signal 1 represents the independent signal based on Gaussian distribution. Signals 2 to 4 are generated based on signal 1. They have a linear, a cosine, a polynomial and exponential relationship to signal 1.

The purpose of generating such signals is to simulate different types of relationships, which provide us an opportunity to analyze the sensitivity of MINE statistics on them. In addition, the dependent relationships between these five signals are designed based on a real application profile. Typically, for many industrial applications, only few working parameters are independent, and most of monitored signals are related to the independent parameters (the relationships are depend on the sensor measurement principle) The signals in the normal state will be used as a basis for injecting the different fault types on them to simulate different changes in relationships between dependent and independent parameters.

For the first fault type (dataset 1), which represents the fault characterized by an abrupt change in the relationship between two signals, an abrupt change of the relationship between signal 1 and 3 is injected in the following way:

1. Generate $S_{temp}(i,j)$ based on the equations for $S(i,j)$ in Table 1;
2. Inject the abrupt change of relationship between signal 1 and 3 from the 1001-th pattern:

$$S^{F1}(i,j) = \begin{cases} S_{temp}(i,j), \ 1 \le i \le 1000, \ j = 1,2,\ldots,5 \\ c * \cos\left(d' S_{temp}(i,1)\right) \ 1001 \le i \le 2000, \ j = 3 \\ S_{temp}(i,j), \ 1001 \le i \le 2000, \ j = 1,2,4,5 \end{cases}$$

3. Normalize $S^{F1}(i,j)$ to be in the interval $[0,1]$.

The injected fault is characterized by changing the parameter $d$ to $d' = 50$. In Figure 1, signal 3 in normal state and in the faulty state is presented.

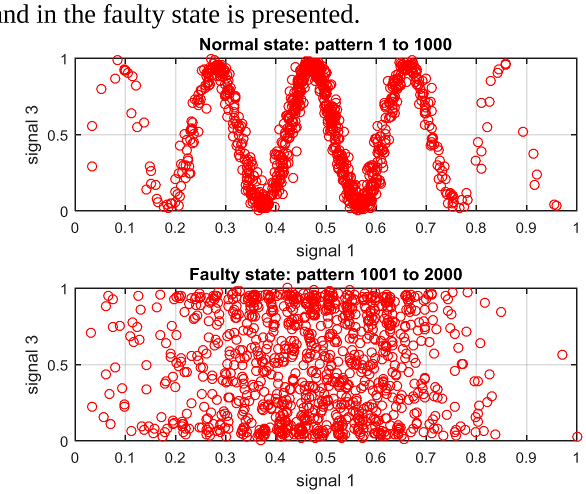


Figure 1. Relationship of signal one before (top) and after the fault injection (bottom)

The second fault type (dataset 2) represents a gradual drift that is injected in signal 3. The following steps are followed to generate the signals:

1. Generate $S_{temp}(i,j)$ based on the equations for $S(i,j)$ in Table 1;
2. Inject the gradual drift in signal 3, from the 1001-th pattern:

$$S^{F2}(i,j) = \begin{cases} S_{temp}(i,j), \ 1 \le i \le 1000, \ j = 1,2,\ldots,5 \\ S_{temp}(i,j) \ + C_1 * Q * i \ , \ 1001 \le i \le 2000, \ j = 3 \\ S_{temp}(i,j), \ 1001 \le i \le 2000, \ j = 1,2,4,5 \end{cases}$$

$$with \qquad Q = \max_{p=1}^{50}\left(S_{temp}(p,j)\right)$$

3. Normalize $S^{F2}(i,j)$ to be in the interval $[0,1]$.

The parameter determining the gradual drift is set to $C_1 = 0.005$. Figure 2 shows the fault in signal 3, starting from pattern 1001. The other signals are not affected by the fault.

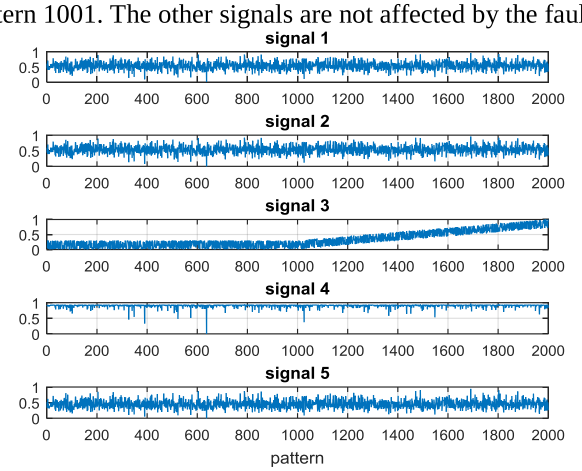


Figure 2. Dataset 2 with signal 3 being affected by gradual drift

For dataset 3 that represents a fault affecting all dependent signals, the procedures are similar to dataset 2. In this case, the gradual drift is injected into signals 2 to 5, while signal 1 remains unaffected by the fault. The following procedure is followed to generate the dataset.

1. Generate $S_{temp}(i,j)$ based on the equations for $S(i,j)$ in Table 1;
2. Inject the gradual drift in signal 2,3,4,5, from the 1001-th pattern:

$$S^{F3}(i,j) = \begin{cases} S_{temp}(i,j), \; 1 \le i \le 1000, \; j = 1,2,\ldots,5 \\ S_{temp}(i,j) \; + C_1 * Q * i \;, \; 1001 \le i \le 2000, \; j = 2,3,4,5 \\ S_{temp}(i,j), \; 1001 \le i \le 2000, \; j = 1 \end{cases}$$

3. Normalize $S^{F3}(i,j)$ into interval $[0,1]$.

$Q$ and $\boldsymbol{C_1}$ are defined in the same way as for dataset 2.

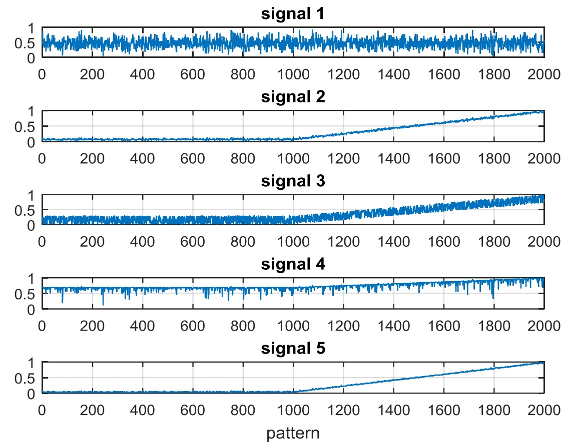


Figure 3. Dataset 3 with signal 2,3,4,5 being affected by gradual drift

A Gaussian measurement noise is imposed on all of the signals of these three system states. The standard deviation of the noise is set as ten percent of the standard deviation of the original signal, as shown in equations (5).

$$\begin{aligned} S_f^{F1} &= S^{F1} + N\left(0, 0.1\sigma_{S^{F1}}\right) \\ S_f^{F2} &= S^{F2} + N\left(0, 0.1\sigma_{S^{F2}}\right) \\ S_f^{F3} &= S^{F3} + N\left(0, 0.1\sigma_{S^{F3}}\right) \end{aligned} \tag{5}$$

where $S^{F1}$, $S^{F2}$, $S^{F3}$ are the original signals of the three datasets, and $S_f^{F1}$, $S_f^{F2}$, $S_f^{F3}$ are the sets of final signals to which the MINE statistics are applied in the following sections.

## 3.2. Fault detection with maximal information-based statistics

In this section, the maximal information-based statistics are applied to detect the three different fault types described in the previous section. The performance of the statistics is compared to the Pearson Correlation Coefficient (PCC) that is a commonly applied correlation coefficient to quantify the relationship between two variables.

The comparisons between PCC and the maximal information-based statistics are twofold. In the first step, the normal states in all of the three datasets are compared to the faulty states. In the second step, it is assumed that the condition monitoring data arrives in batches of 200 measurements each and the relationship of the signals per batch is analyzed to compare the detection ability of the two approaches.

Two different statistics of the maximal information-based statistics are selected: MIC capturing the strength of the relationship, and the MAS capturing the departure from monotonicity. Generally, MAS is always smaller than MIC. Figures 4 to 6 present the heatmaps of MIC, MAS and PCC for the first dataset. Since MIC and MAS are scaled to be in the interval [0,1], they can be easily compared. PCC can generally take values in the interval [-1,1]. However, for the datasets applied, only values in the interval [0,1] were observed. Therefore, the same scale for all the values is used to compare the relationship between the signals. For dataset 2 and 3, a similar behavior of the three indicators is observed. However, due to space limitation in this paper, the plots are not displayed.

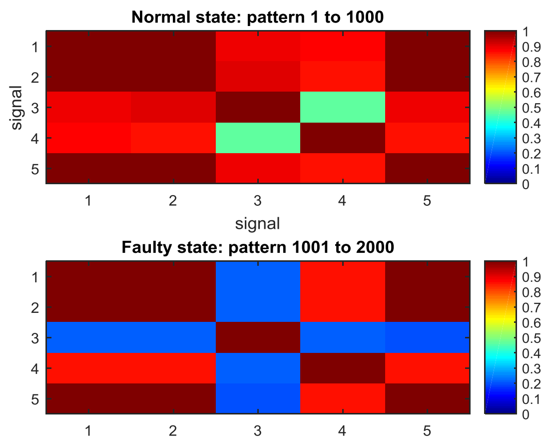


Figure 4. MIC for healthy and faulty signals Dataset 1

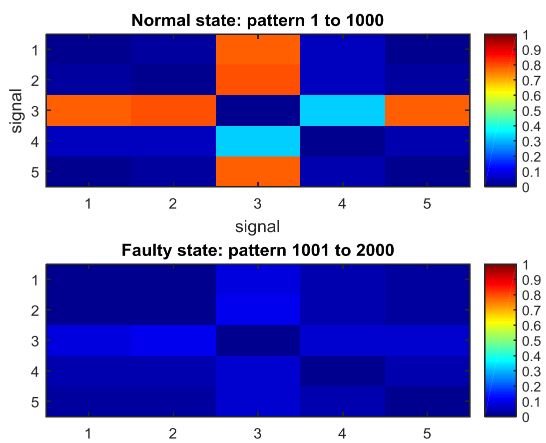


Figure 5. MAS for healthy and faulty signals Dataset 1

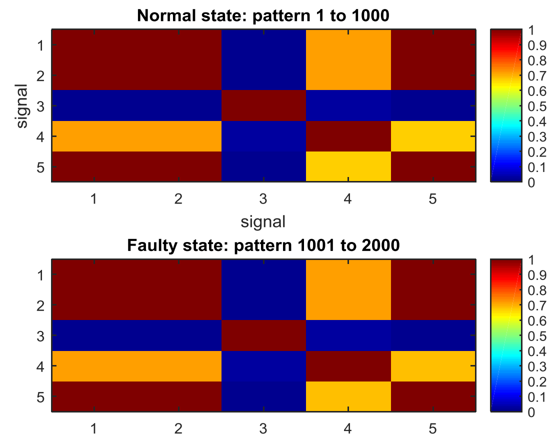


Figure 6. PCC for healthy and faulty signals Dataset 1

To analyze the ability of MIC and MAS to timely detect the changes in the relationship between two signals, a more detailed analysis is performed. The three indicators are computed for every 200 patterns. Additionally, the sum of the indicators of each signal is displayed in Figures 7 to 9. If the relationships of the signal to all other signals are strong, the maximum value of 5 will be achieved (including the relationship of the signal to itself).

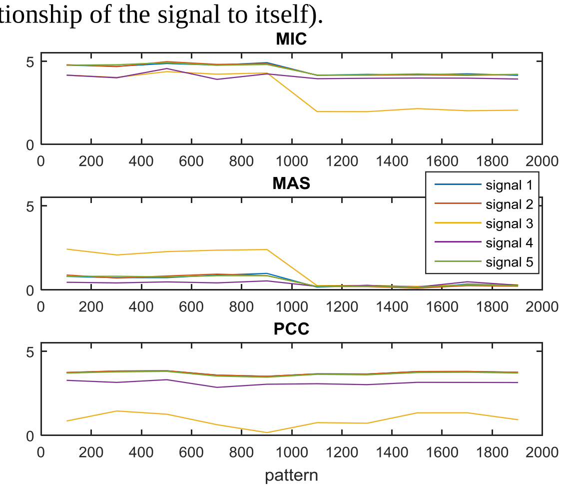


Figure 7. Sum of MIC, MAS and PCC of all signals (Dataset 1)

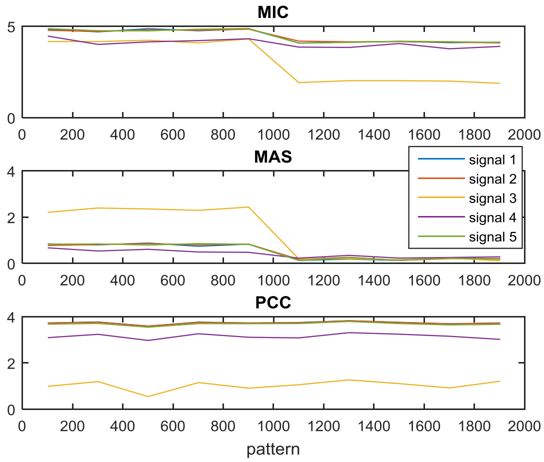


Figure 8. Sum of MIC, MAS and PCC of all signals (Dataset 2)

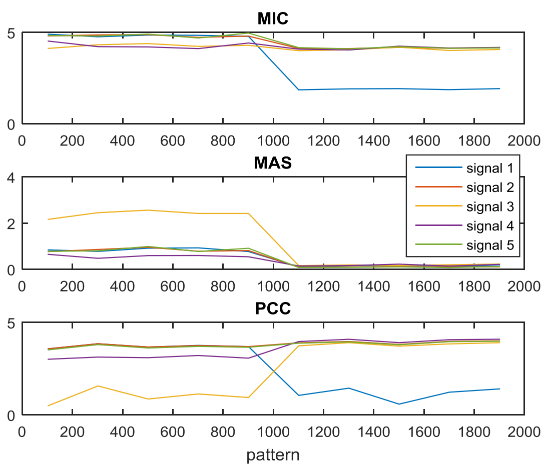


Figure 9. Sum of MIC, MAS and PCC of all signals (Dataset 3)

## 4. RESULT INTERPRETATION AND DISCUSSION

MIC and MAS show a good ability to detect non-linear relationship in the data and also to capture the change in the relationship. These characteristics make them suitable for the application of fault isolation, particularly in the scenario of unsupervised deep feature learning.

For the dataset 1, PCC is not able to detect the relationship between signal 3 and other signals in the normal state. It is therefore also not able to detect the change in the relationship in the faulty state. On the other hand, both MIC and MAS show a strong difference of the relationship among signal 3 and other 4 signals in normal and faulty states.

In dataset 2, the performance of MIC, MAS is similar to that on dataset 1. The PCC also shows a small distinguishable difference of the relationship between signal 3 and the other signals. The main reason that the this change in the relationships is detected by PCC but the one in dataset 1 not is that a linear relationship plays a dominant role in the injected fault type in signal 3. Linear relationships are easier for PCC to detect.

In dataset 3, the relationships between the four dependent signals are not changed. However, the relationships of the dependent signals to the independent signal 1 change. This change in the relationship can be clearly interpreted based on the MIC values in Figure 9. The strength of the relationship of signal 1 to the other four signals decreases to approximately 0.2 in the faulty state from being in the interval of [0.85,1] in the normal state. For the sum over all the signals in the faulty state, 1.8 is achieved. In the dependent signals, only the relationship to signal 1 is lost, the strength of the relationship to the other signals remains at the same level. Therefore, the sum of the MIC values of the other signals is around 4.0.

For dataset 3, PCC shows a change in the relationships. However, the results cannot be interpreted.

The results obtained on the three different types of faults demonstrate that the indicators based on the maximal information criterion, particularly MIS and MAS, are not only able to indicate the change in the relationship between two signals but also enable an interpretation of the obtained detection results and enable to distinguish different fault types based on the combination of the relationships between the signals.

## 5. CONCLUSIONS

Fault detection and prognostics based on unsupervised deep feature learning approaches requires fast and robust fault isolation approaches.

The two maximal information-based nonparametric exploration indicators: maximal information criterion and the maximal asymmetry score provide a good approach to detect changes in the relationship between dependent and independent signals. The main advantage of applying MIC and MAS is to extract non-linear relationships and the change of these relationships when the state condition changes.

In the described case study, we have applied the MIC and MAS to extract the pairwise relationships between the different condition monitoring signals. For the 3 simulated datasets, the proposed approach was shown to be useful.

The application is particularly suitable for condition monitoring applications where the monitoring data is analyzed batch-wise. This is for example the case, when the data is uploaded once a day from the plant and is then analyzed by the diagnostics engineers. It provides a good tool for accurate detections. In addition, the maximal information-based nonparametric exploration of condition monitoring data can also be applied on online monitored data in a moving window.

An interesting characteristic of the two applied maximal information-based features is that they are not only considering the behavior of one signal, but the relationship between two signals. In this case study, the MIC and MAS were applied as the only features for detection the faults in the condition monitoring data and detecting the changes in the relationships. However, they can also be applied in combination with other features to complement the extracted information. This is subject for further research.

A limitation of the maximal information-based criteria is that they are only considering pairwise-relationship between two signals. They are not able to integrate relationships between more than two signals which is often required for condition monitoring tasks of complex systems. Additionally, MIC can also be sensitive the size of the signal windows that used to extract the relationships and compare the distributions.

### ACKNOWLEDGEMENT

This research was funded by the Swiss Commission for Technology and Innovation under grant number 17833.2 PFES-ES.

### REFERENCES

Ando, S., & Suzuki, E. (2006). An information theoretic approach to detection of minority subsets in database. In *Proceedings - IEEE International Conference on Data Mining, ICDM* (pp. 11–20). http://doi.org/10.1109/ICDM.2006.19

Dai, X., & Gao, Z. (2013). From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, *9*(4), 2226–2238. http://doi.org/10.1109/TII.2013.2243743

Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, *33*(2), 1134–1140. http://doi.org/10.1103/PhysRevA.33.1134

Jiang, Q., & Yan, X. (2014). Plant-wide process monitoring based on mutual information-multiblock principal component analysis. *ISA Transactions*, *53*(5), 1516–27. http://doi.org/10.1016/j.isatra.2014.05.031

Kappaganthu, K., & Nataraj, C. (2011). Feature Selection for Fault Detection in Rolling Element Bearings Using Mutual Information. *Journal of Vibration and Acoustics*, *133*(6), 061001. http://doi.org/10.1115/1.4003400

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *69*(6 Pt 2), 066138. http://doi.org/10.1103/PhysRevE.69.066138

Moon, Y.-I., Rajagopalan, B., & Lall, U. (1995). Estimation of mutual information using kernel density estimators. *Physical Review E*, *52*(3), 2318–2321. http://doi.org/10.1103/PhysRevE.52.2318

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., … Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science (New York, N.Y.)*, *334*(6062), 1518–24. http://doi.org/10.1126/science.1205438

Verron, S., Tiplica, T., & Kobi, A. (2008). Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control*, *18*(5), 479–490. http://doi.org/10.1016/j.jprocont.2007.08.003

Wu, S., & Wang, S. (2013). Information-theoretic outlier detection for large-scale categorical data. *IEEE Transactions on Knowledge and Data Engineering*, *25*(3), 589–602. http://doi.org/10.1109/TKDE.2011.261