中国地质大学
CHINA UNIVERSITY OF GEOSCIENCES
武汉 · WUHAN

Frequent Event Pattern Extraction of Drilling Time Series Using Change Point Detection and Event Sequence Generation

Yupeng Li, Wenkai Hu, Weihua Cao, R. Bhushan Gopaluni, Liang Cao, Chao Gan, Min Wu

School of Automation, China University of Geosciences, Wuhan 430074, China
Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.

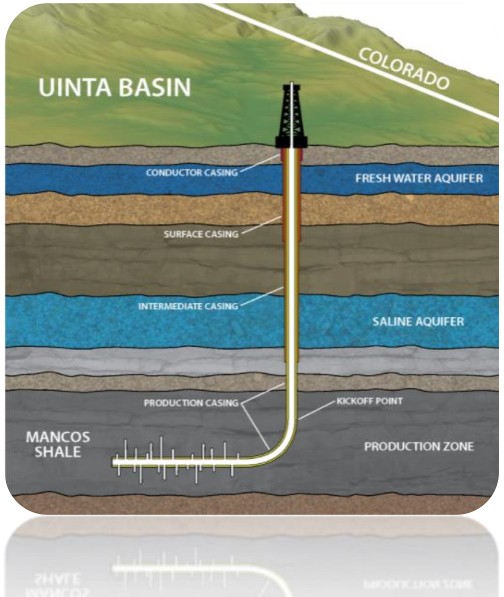
1. Introduction

➢ Significance of deep exploration

- Way to obtain underground resources
- Important for mineral exploration or geothermal exploitation
- In China, over 40% of oil and 60% of natural gas resources are buried deep

➢ High safety risk in deep drilling process

- Risk of downhole faults increases with the drilling depth
  - ✓ High temperature, high pressure, high steep structure stratum
  - ✓ Alternations of hard and soft strata
- Challenges in deep drilling processes
  - ✓ Incomplete data makes it difficult to accurately obtain the downhole condition
  - ✓ High porosity, swelling and other factors increase the risk of downhole incidents



1. Introduction

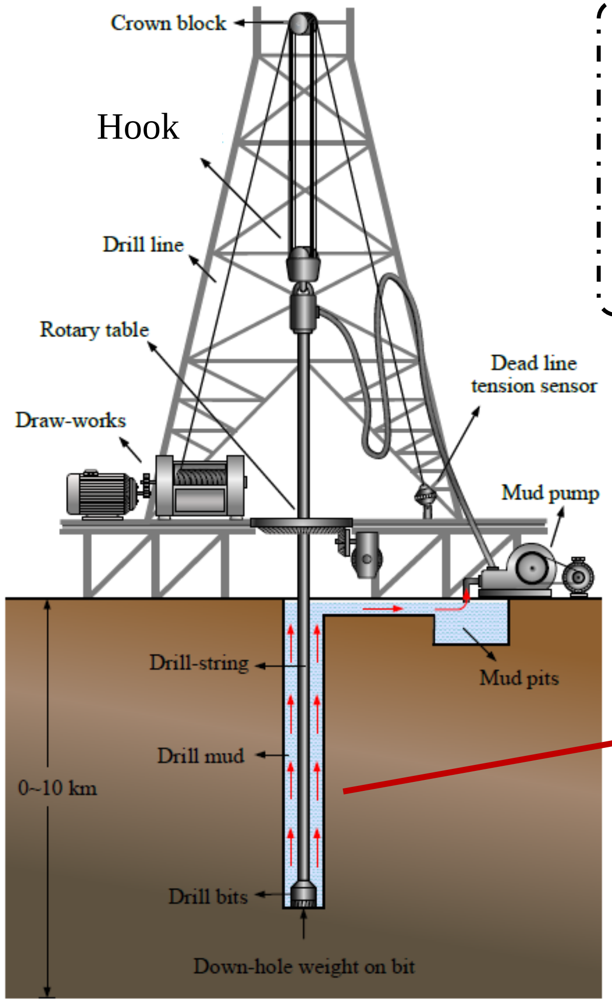
➢ Drilling accidents led to disastrous results

- The Deepwater Horizon oil spill
  - ✓ 11 people killed, 17 people injured
  - ✓ 4 million barrels of oil flowed from the damaged well
  - ✓ Tremendous damage to the ecology, economy, and society
- Blowout in Kai County, Chongqing, China
  - ✓ Over 200 deaths in the hydrogen sulfide blowout
  - ✓ Economic losses of more than 64 million yuan

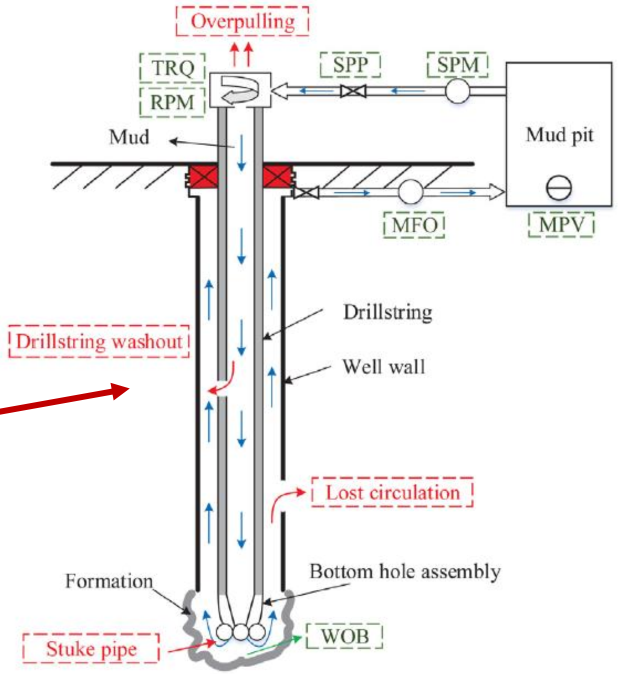
Process monitoring and fault detection are significant to ensure drilling safety and reduce maintenance costs

1. Introduction

➢ A typical geological drilling process



- **Key variables to monitor:**
  - ✓ Axial direction: Weight on bit (WOB), Hook load (HKL)
  - ✓ Torsional direction: Torque (TRQ), rotary per minute (RPM)
  - ✓ Circulation system: mud flow in/out (MFI, MFO), mud pit volume (MPV), SPP



- **Common faults**
  - ✓ Lost circulation: Mud enters the formation
  - ✓ Washout: Drillstring cracks
  - ✓ Stuck pipe
  - ✓ Bit bounce

Broken segment of drillstring

1. Introduction

CHINA UNIVERSITY OF GEOSCIENCES
WUHAN

Mechanism model-based methods

- Hydraulic or kinematic model described by differential equations

- Single-phase hydraulic model
  - A downhole fault detection method was proposed based on adaptive observers and single-phase hydraulic model [Willersrud et al., *CEP*, 2015. Willersrud et al., *IFAC Proceedings*, 2013.]

- Multiphase flow model
  - A fault classification method was developed based on a pressure-temperature coupling model of drilling fluid system, where the UKF was used to estimate the pressure loss and loss rate[Jiang et al., *JPSE*, 2020.]

- Differential equation model of drill string
  - Considering that the motion characteristics of drillstring were related to downhole faults, differential equation models describing the stick-slip fault were established [Kamel et al., *JSB*, 2014.]

Establishing high precision mechanism models is rather tricky in geological drilling processes due to the limited measurable variables

1. Introduction

Data-driven methods

- Learning the underlying structure of the drilling process from historical data without the first principle model

- Downhole fault classification and diagnosis
  - ✓ Transform the fault diagnosis into a binary or multi-classification problem, supervised learning approaches were adopted to diagnose downhole faults based on SVM and NN [Zhang et al., *JPC*, 2021. Li et al., *PESP*, 2020]

- Distance-based fault detection
  - ✓ Analyze drilling time series based on distance metrics, such as the morphological distance, local distance, and dynamic time warping [Zhao et al., *JNASE*, 2019,. Zhang et al., *JPC*, 2021.]

- Distribution-based approaches
  - ✓ The generalized Gaussian distribution was used to describe the drilling data, enabling a detection decision by comparing the KL divergence from the distribution of normal data to that of online data[Li et al., *CEP*, 2021.]

Some normal drilling operations may also lead to false alarms, and how to eliminate these alarms is rarely considered

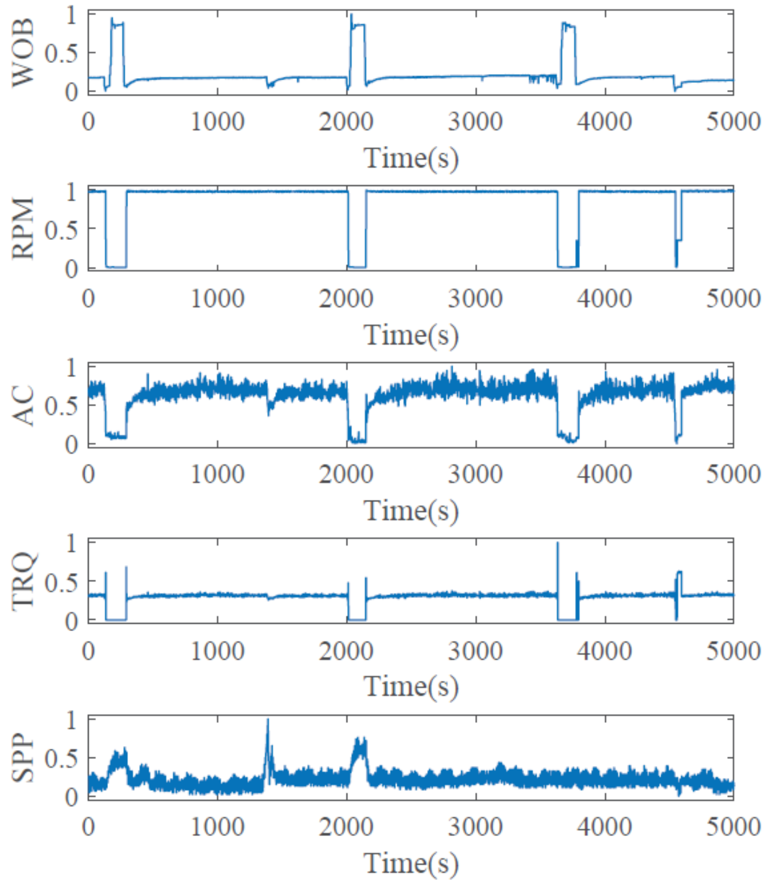
1. Introduction

➢ Limitations

- The drilling operation is not always in a stationary phase due to the changes in formation and shifts between operating conditions

- Non-stationary phases are prone to cause false alarms and deteriorate alarm system performance

- Existing methods assume that the templates of all drilling modes are known and well-prepared



1. Introduction

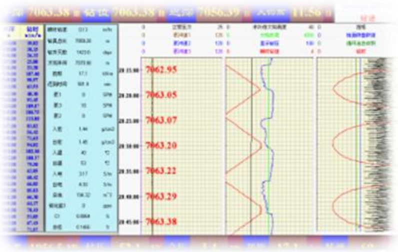
CHINA UNIVERSITY OF GEOSCIENCES

Motivations

- Non-stationary phases are prone to cause false alarms and deteriorate alarm system performance
- Lack of abnormal data and presence of multiple operating states

Main contributions

- A non-stationary phase detection method is proposed to extract drilling frequent event patterns
- An event sequence generation method is proposed to express drilling frequent event patterns with a group of symbols



The Proposed Method

3. The Proposed Method

Dominant feature extraction based on t-SNE

- Drilling process is a typical multivariate process （WOB, TRQ, RPM, AC, and SPP）
- A dominant feature needs to be extracted
- Transform the high-dimensional data into a low-dimensional space using similarity probabilities

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|x_i - x_k\|/2\sigma_i^2)},$$

Original dataset X = {x_1, x_2, ..., x_M}

Low-dimensional dataset Y = {y_1, y_2, ..., y_m}

- Minimize the difference between pj|I and qj|i

$$E = \sum_i \mathrm{KL}(P_i\|Q_i) = \sum_i\sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}}.$$

$$\frac{\partial E}{\partial y_i} = 4\sum_j (p_{ij}-q_{ij})(y_i - y_j)\left(1+\|y_i - y_j\|^2\right)^{-1}.$$

the distribution P_i given all datapoint x_i to the distribution Q_i given all datapoint y_i

Low-dimensional signal Y is obtained and prepared for extracting the non-stationary phase

3. The Proposed Method

CHINA UNIVERSITY OF GEOSCIENCES

Non-stationary phase detection using change point detection

- As non-stationary phases caused by set-point adjustments, mode switching, formation changes, usually result in dynamic changes in the time series
- Change point detection method → extract non-stationary phases in historical data
- The PE divergence PE($P \| P_r$) between $P(y)$ and $P_r(y)$ is given as

$$\mathrm{PE}(P\|P_r) = \frac{1}{2}\int P_r(y)\left(\frac{P(y)}{P_r(y)} - 1\right)^2 dy,$$

$P(y)$ - distribution of Y, $P_r(y)$ - distribution of a stable reference signal.

- Estimate the density ratio P(A)/P(B) based on Least-Squares Importance Fitting algorithm

$$\begin{aligned}\mathrm{PE}_\alpha(P\|P_r) &= \mathrm{PE}(P\|g_\alpha)\\ &= \frac{1}{2}\int P_r(y)\left(\frac{P(y)}{g_\alpha(y)} - 1\right)^2 g_\alpha(y)dy,\end{aligned}$$

$$g_\alpha(y) = \alpha P(y) + (1-\alpha)P_r(y) \qquad \alpha \in (0, 1).$$

3. The Proposed Method

中国地质大学 CHINA UNIVERSITY OF GEOSCIENCES 武汉 · WUHAN

Non-stationary phase detection using change point detection

- α-relative density ratio is given by $r_\alpha(y) = \frac{P(y)}{g_\alpha(y)}$,

- r (y) is estimated based on the sum of kernel models as

$$\hat{r}_\alpha(y) = h(y) = \sum_{i=1}^{n} \phi_i \mathcal{K}(y, y_i), \qquad \mathcal{K}(y, y_i) = \exp\left(-\frac{\|y - y_i\|^2}{2\sigma^2}\right),$$

$$\hat{\mathrm{PE}}_\alpha(P\|P_r) = \frac{1}{2m}\sum_{j=1}^{m}\left(2\hat{h}(y_j) - \alpha\hat{h}(y_j)^2\right) - \sum_{i=1}^{m}\frac{(1-\alpha)}{2\tilde{n}}\hat{h}(y_i^r)^2 - \frac{1}{2}.$$

- Determine the threshold by historical data using cumulative distribution as

$$P(s \le \mathrm{PE}_{th}^s) = \int_0^{\mathrm{PE}_{th}^s} \Gamma(s)ds = \beta.$$

- The change point is extracted using the following hypothesis testing

$$\begin{cases} \mathrm{PE}_\alpha^s \le \mathrm{PE}_{th}^s : y(k) \text{ is not a change point}, \\ \mathrm{PE}_\alpha^s > \mathrm{PE}_{th}^s : y(k) \text{ is a change point}. \end{cases}$$

3. The Proposed Method

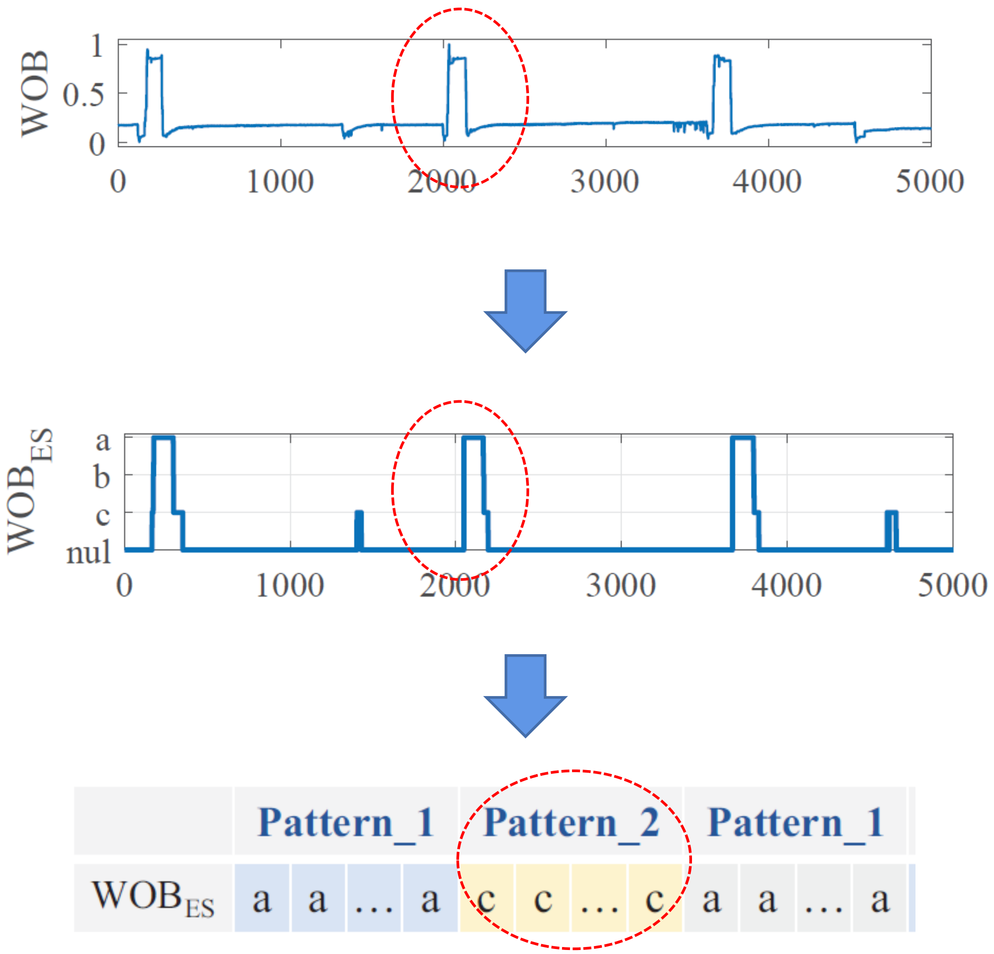
Event sequence generation

- Drilling signals are transformed into event sequences to express non-stationary patterns
- Symbolic aggregate approximation is used to convert the MTS $X(k)$ into the event sequence matrix $E(k)$ as

$$\mathcal{E} = \{E_i(k) | \forall i \leq k, E_i(k) = f_i(\mathcal{X}_i(k), \eta_i)\},$$

- Every element $x_i(j)$ of $x_i(k)$ is mapped to one of the discrete intervals with a certain event symbol as

$$e_i(j) = \begin{cases} a \;\; \eta_i^u < x_i(j) \\ b \;\; \eta_i^l \leq x_i(j) \leq \eta_i^u \\ c \;\; x_i(j) < \eta_i^l \end{cases}$$



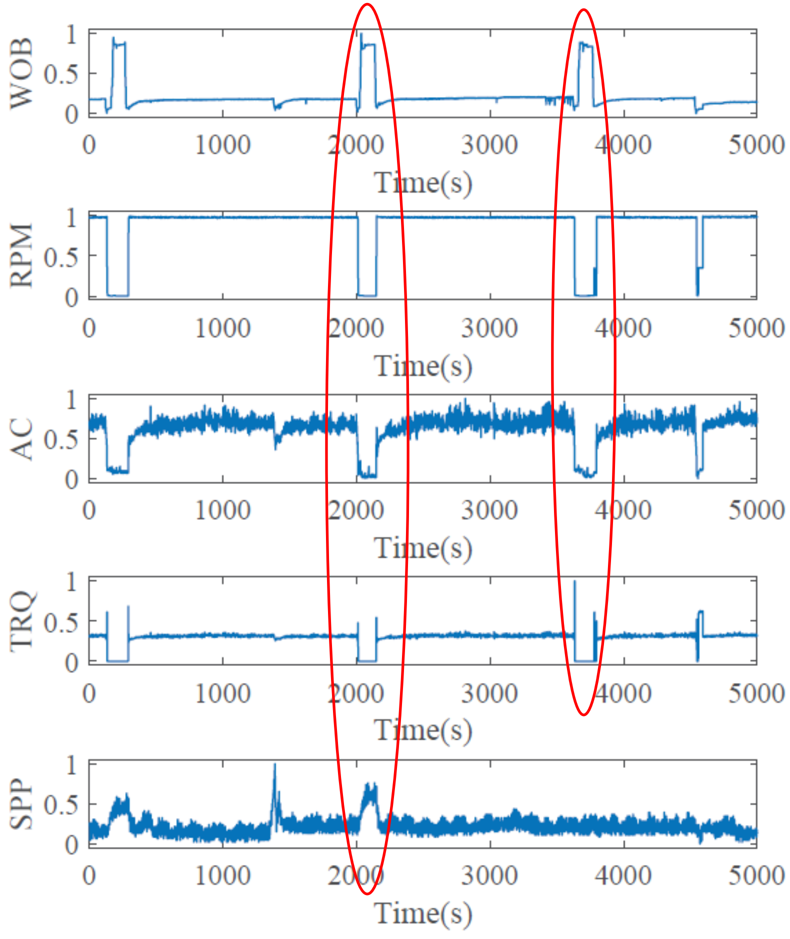
中国地质大学
CHINA UNIVERSITY OF GEOSCIENCES
武汉 · WUHAN

Industrial case study

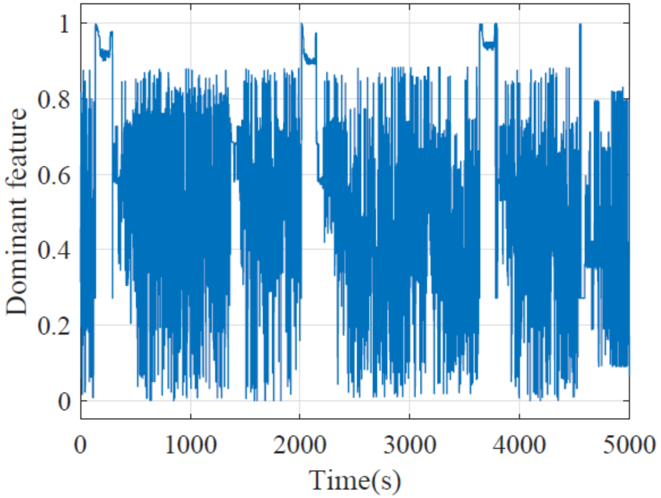
4. Industrial case study

➢ Change point detection result

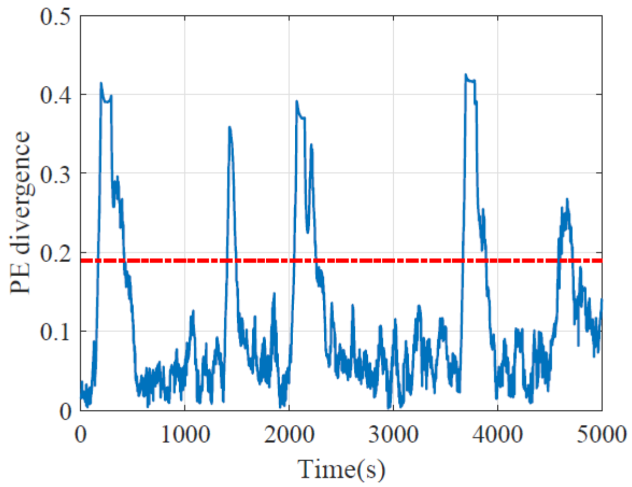
- The industrial data from a real drilling project located in Shandong province, China, is provided



Time series plots of drilling process signals



Time series plot of the extracted dominant feature based on t-SNE



Change point detection result for the one-dimensional feature based on PE divergence

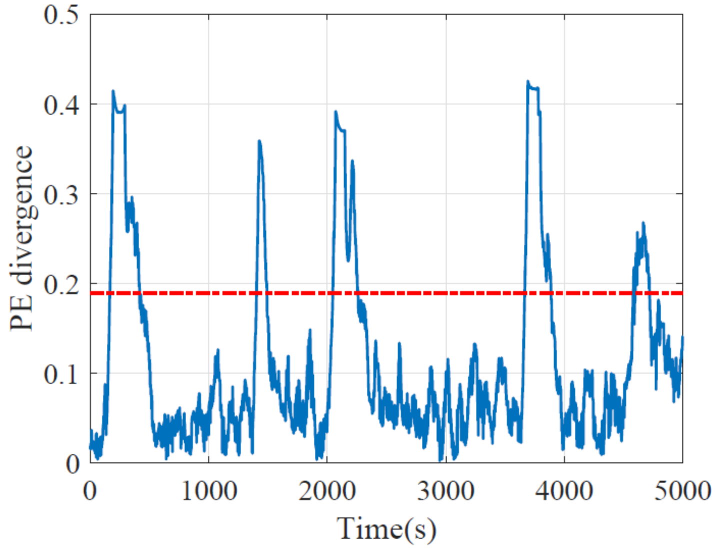
- Original signals are normalized to 0 to 1
- Multiple signals were reduced to a dominant feature
- PE divergence is calculated to detect change points

4. Industrial case study

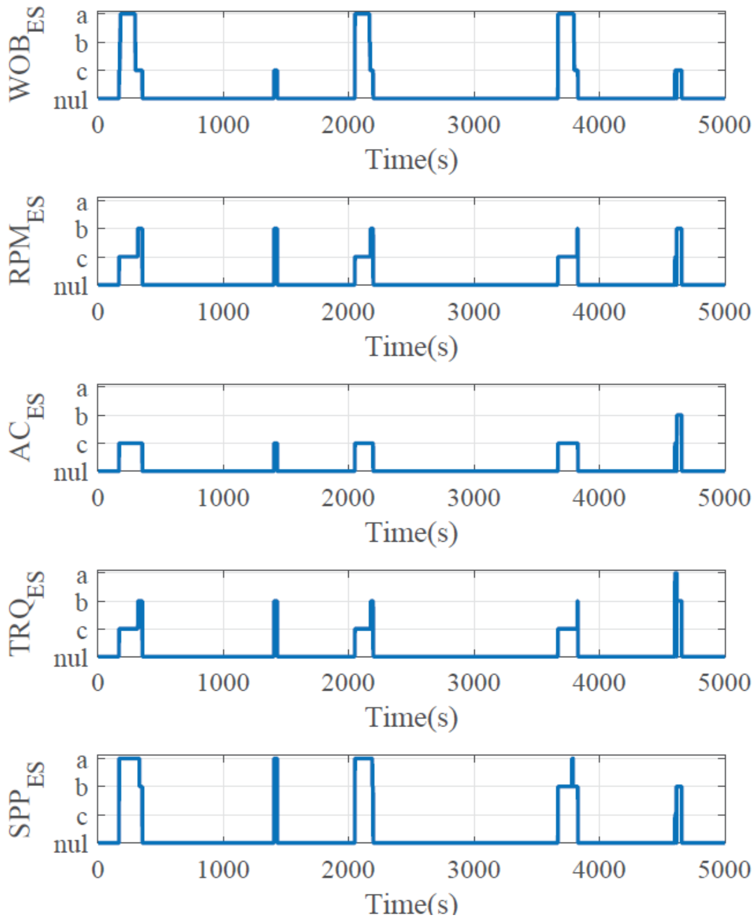
CHINA UNIVERSITY OF GEOSCIENCES

➢ Non-stationary phases extraction

- Assessing whether the PE divergence violates the red dashed line representing the threshold
- Expressed by a group of Event Sequences
- The symbol 'a' denotes high, 'b' indicates normal, 'c' represents low, and 'nul' correspond to stationary phases



Change point detection result for the one-dimensional feature based on PE divergence



Time series plots of the generated event sequences for drilling signals

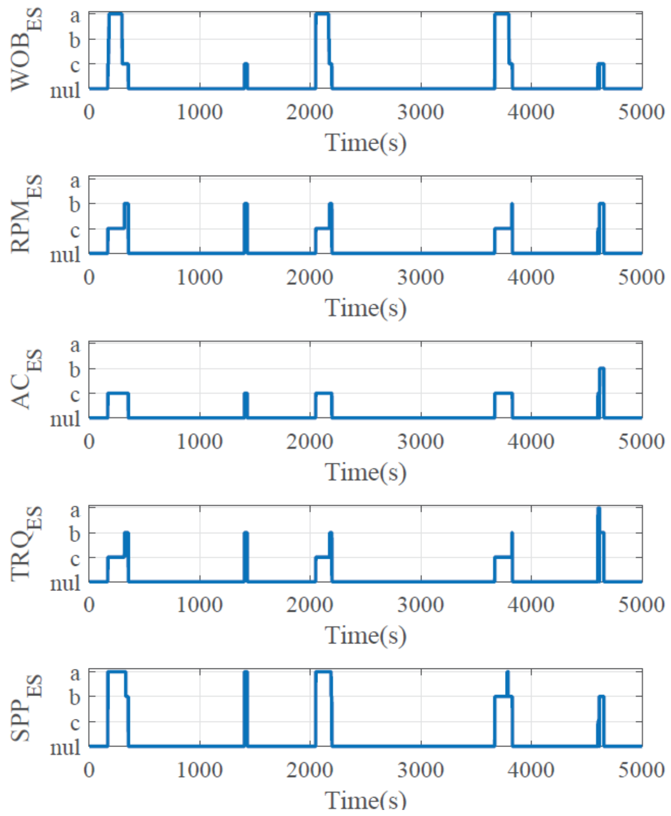
4. Industrial case study

➢ Event Sequence Generation

- Frequent event patterns correspond to non-stationary intervals include [168, 198], [1380, 1410], [2050, 2080]
- The first and third non-stationary phases share identical sequences, and thus are grouped into the same pattern

	Pattern_1				Pattern_2				Pattern_1				Pattern_3				Pattern_4			
WOB_{ES}	a	a	…	a	c	c	…	c	a	a	…	a	a	a	…	a	c	c	…	c
RPM_{ES}	c	c	…	c	b	b	…	b	c	c	…	c	c	c	…	c	c	c	…	b
AC_{ES}	c	c	…	c	c	c	…	c	c	c	…	c	c	c	…	c	c	c	…	b
TRQ_{ES}	c	c	…	c	b	b	…	b	c	c	…	c	c	c	…	c	a	a	…	b
SPP_{ES}	a	a	…	a	a	a	…	a	a	a	…	a	b	b	…	b	c	c	…	b

Part of frequent event patterns extracted from historical data



Time series plots of the generated event sequences for drilling signals

Frequent event patterns corresponding to false alarms can be obtained, which provides a path to improve the safety monitoring performance by reducing false alarms

Conclusion

5. Conclusion

中国地质大学
CHINA UNIVERSITY OF GEOSCIENCES
武汉 • WUHAN

Summary

- An original framework is proposed for extracting frequent event patterns
- The non-stationary phase is identified from the dominant feature detection
- Frequent patterns associated with the non-stationary phase are determined and transformed into event sequences

Future work

- Discover the fault-sensitive feature based on the extracted frequent patterns
- Design a similarity analysis index for frequent patterns to design process monitoring systems

Thank you!
Questions?

Dr. Yupeng Li

School of Automation, China University of Geosciences, Wuhan 430074, China

Email: yupengli@cug.edu.cn