A Fault Inference Method under Uncertainty: Case Study on Crankshafts in Fracturing Pumps

X Zhang¹, L B Zhang¹ and J Q Hu¹
¹ College of Safety and Ocean Engineering, China University of Petroleum, 102249 Beijing, China

E-mail: hujinqiu@cup.edu.cn

Abstract. Crankshaft is a pivotal mechanical unit in the power-end system of a fracturing pump and its fault inference could facilitate optimal condition-based maintenance. Fracturing pumps are equipped with advanced instrumentation systems able to acquire vibration information for crankshaft fault analysis, but there exist complex uncertain dependences between faults and symptoms as well as incomplete symptom information, further increasing the difficulty of fault inference by operators. To achieve effective fault inference in the case of uncertain or incomplete diagnosis evidences, a Bayesian network-based fault inference method for crankshafts is proposed in this article. The approach can be utilized to implement cause inference and diagnosis inference by incorporating cause nodes, fault nodes and symptom nodes into a Bayesian network (BN) model. The application of the presented approach in fault inference of crankshafts indicates its strong inference capability under uncertainty. The results from the presented BN model may offer a useful aid to repairers in their maintenance decision-making processes.

1. Introduction
In recent years, shale gas as a clean energy has attracted increasing attention over the world. To enable efficient development of shale-gas wells, large-scale fracturing pumps have been extensively utilized for shale-gas wells fracturing operations [1]. These pumps could pressurize the sand-laden fluid and then pump it into the high-pressure manifolds. Owing to the harsh working conditions such as high-pressure, large-displacement and super-power, mechanical units or components suffer from rapid degradation. In particular, the crankshaft is always subjected to rapid deterioration due to its high-speed and high-load. The enhanced equipment safety management has revealed a need for a more accurate and more intelligent fault inference method. Machinery fault inference is an inverse reasoning process from symptoms to faults. In engineering, there are multiple corresponding relations between symptoms and faults, including one-to-one, one-to-many and many-to-one correspondence [2]. Moreover, the collected symptom information is sometimes incomplete or uncertain because of the limited data acquisition conditions. Therefore, it is essential to carry out fault inference under conditions of uncertainty or incompleteness. This situation is especially suitable for the crankshafts because of the complex correspondences between faults and symptoms as well as the difficulties in collecting and analyzing crankshaft’s vibration signals.

Fault diagnosis of equipment or components has received considerable attention over the last decade [3–11]. Various artificial intelligence methods have been developed for fault diagnosis, including support vector machine [4,5,10], neural network [6–9], fuzzy logic [3], and their hybrid methods. For example, Ref. [10] have developed an intelligent fault diagnosis technique for
locomotive roller bearings by combining support vector machine and ant colony algorithm; in this technique, the ant colony algorithm was used for synchronous feature selection and parameter optimization for support vector machine. Ref. [9] have adopted artificial neural networks to isolate and identify eight different types of faults in photovoltaic systems. In this method, features were manually extracted from a number of simulations. Ref. [11] have proposed a promising tool on a basis of deep neural networks for fault feature mining and intelligent diagnosis of rotating machinery with massive data.

Although the above methods have significantly promoted the progress of fault analysis of mechanical units, they usually depend on complete signal data and thus have inadequate robustness in the presence of incomplete or uncertain diagnosis data. Hence, a novel intelligent fault inference method for crankshafts is highly required. Motivated by above requirement, this paper proposes an intelligent fault inference approach under uncertainty based on Bayesian network (BN) by taking full advantage of expert knowledge, fault causes, faults and symptoms. The proposed approach fully takes into account the complex uncertainties between causes, faults and symptoms, as well as the incomplete symptoms. Using this method, the cause inference and diagnosis inference could be implemented.

The remainder of this article is structured as follows: Section 2 gives a brief overview of BN. Section 3 describes the proposed method for developing an intelligent fault inference model for crankshafts. In Section 4, the proposed approach is applied for fault inference of a crankshaft in a fracturing pump. Finally, the conclusions are given in Section 5.

2. Bayesian network
A Bayesian network as a directed acyclic graph (DAG) is suitable for a visual modelling of the causality with uncertainty. The analysis of a Bayesian network is based on the specification of conditional probabilities of child nodes for given states of their parent nodes, utilizing the concept of conditional probabilities table (CPT). The nodes in a BN can embrace multiple states, and the directed arrows mean the probabilistic dependencies among nodes. For more general information on BN, see Nielsen and Jensen. [12].

Bayesian networks have proved to be a powerful tool in various engineering areas; recent applications from a range of technical fields include reliability evaluation of marine floating structures [13] and kick control operation [14], risk analysis of marine transportation [15] and submarine pipelines [16], forecast of stock market index [17] and forensic assessments [18]. These studies have been motivated by the strong points of Bayesian networks that facilitate:

(1) Modelling the complex causal interdependence in a graphical way, with clear semantics and an easy-to-understand structure.

(2) Updating the probability distribution of target nodes when new evidences are available.

3. Proposed methodology
The steps for the proposed methodology is illustrated in figure 1. The steps are stated in detail as follows:

![Figure 1. Procedure for fault inference.](image_url)
3.1. Step 1: Determine nodes
The built BN model in this paper is composed of three classes of nodes: cause nodes (root nodes), fault nodes (intermediate nodes) and symptom nodes (leaf nodes). A cause is defined as an internal or external environmental factor of abnormal conditions that could lead to the loss of ability to perform a required function.

For example, poor thermal stability is the cause of crankshaft imbalance. A fault is a condition that renders an element unable to perform its required function at the expected levels of performance. A symptom may be defined as a deviation from the expected behaviour of a system or a specific phenomenon sensitive to the concerned faults. For instance, the symptoms of crankshaft misalignment may include strong axial vibration and doubling-frequency. These nodes could be determined from failure mode and effect analysis (FMEA), or fault tree analysis (FTA) of machinery or from the equipment maintenance staff.

3.2. Step 2: Establish BN structure
In fact, when various nodes are defined in Step 1, the complex uncertain relations among these nodes are also analyzed and determined. The purpose of this step is to establish the BN structure according to their complex inter-relationships. The constructed BN network contains two topological configurations of causality: topology structure between cause nodes and fault nodes as well as that between fault nodes and symptom nodes.

3.3. Step 3: Quantify BN network
The parameters in a BN consist of prior probabilities of cause nodes and conditional probabilities amongst nodes. Prior probabilities can be assigned by analyzing historical fault reports or by expert knowledge. Conditional probabilities could be estimated using parameter learning methods such as Maximum Likelihood Estimation (MLE) or expert knowledge. For more details on the methods to estimate the conditional probabilities, please refer to Reference [12].

3.4. Step 4: Fault inference
Fault inference can be further divided into diagnosis inference and cause inference. Diagnosis inference is a process of updating prior probabilities of fault nodes using subsequent available information from cause nodes or symptom nodes. Based on the observed evidence, the occurrence probabilities of each fault could be obtained using the following Bayesian theorem:

$$ P(F_i | e) = \frac{P(F_i e)}{P(e)} = \frac{P(e | F_i) P(F_i)}{\sum_{i=1}^{n} P(e | F_i) P(F_i)} $$  \hspace{1cm} (1)

in which $P(e | F_i)$ is the conditional probability of evidence $e$ given the fault $F_i$; $P(F_i)$ is the prior probability of $F_i$; $P(e) = \sum_{i=1}^{n} P(e | F_i) P(F_i)$ is the marginal probability of evidence; $P(F_i | e)$ is called the posterior probability given the evidences $e$.

Then the most probable fault can be identified. Causal inference is to estimate posterior probabilities of ancestor nodes of the observed evidences. Causal inference is applicable to computing the magnitude of a cause effect on faults, or identify the likely causes after updating the symptoms.

4. Case study
In this section, the proposed BN method is applied to fault inference of crankshafts. Crankshafts play a significant role in ensuring the smooth operation of the pumps. In the construction site, multiple fault modes such as imbalance, misalignment and bending of crankshaft could result in performance degradation of pumps and even undesirable events. These faults may have some identical fault symptoms; moreover, one type of fault with varying degrees may have different symptoms, indicating a complex uncertainty relations between the faults and the symptoms. Therefore, fault inference for crankshafts has become a challenging but also intellectually fascinating task in engineering.
4.1. Step 1: Determine nodes
In this case, five main fault modes of a crankshaft are selected as fault nodes, as shown in Table 1. Through field investigation and literature research, ten cause nodes and twelve symptom nodes are identified and listed in Table 1. The symptom nodes can be roughly categorized as four classes: vibration direction nodes (S1 and S2), axis orbit nodes (S3, S4 and S5), characteristic frequency nodes (S6, S7, S8 and S9) as well as secondary frequency nodes (S10, S11 and S12).

| No. | Node | Description               | No. | Node | Description             |
|-----|------|---------------------------|-----|------|--------------------------|
| 1   | C1   | Initial bending           | 15  | F5   | Crankshaft cracks        |
| 2   | C2   | Misalignment of bearing block | 16  | S1   | Abnormal axial vibration |
| 3   | C3   | Uneven material           | 17  | S2   | Abnormal radial vibration|
| 4   | C4   | Poor thermal stability    | 18  | S3   | Axis orbit with single ellipse |
| 5   | C5   | Polluted lubricant        | 19  | S4   | Axis orbit with double ellipse |
| 6   | C6   | Unreasonable axle bush   | 20  | S5   | Irregular axis orbit     |
| 7   | C7   | Excessive lubricant temp | 21  | S6   | 0.5X characteristic frequency |
| 8   | C8   | Excessive load            | 22  | S7   | 1X characteristic frequency |
| 9   | C9   | Frequent starting         | 23  | S8   | 2X characteristic frequency |
| 10  | C10  | Fatigue running           | 24  | S9   | 3X characteristic frequency |
| 11  | F1   | Crankshaft Misalignment   | 25  | S10  | 1X secondary frequency   |
| 12  | F2   | Crankshaft bending        | 26  | S11  | 2X secondary frequency   |
| 13  | F3   | Crankshaft imbalance      | 27  | S12  | Higher harmonics         |
| 14  | F4   | Oil-film whirl            |      |      |                          |

4.2. Step 2: Establish BN structure
By analysing the uncertain causal relations among the identified nodes, the BN network is built and provided in Figure 2. As observed in Figure 2, there exist complex dependencies among nodes. For instance, uneven material may lead to three faults at the same time and abnormal axial vibration may be caused by five faults.

4.3. Step 3: Quantify BN network
Quantifying BN structure is a challenging issue because of the complicated causality. Limited fault and maintenance records could be accessed due to their confidentiality; thus, the prior probabilities of cause nodes are assigned by means of expert knowledge and given in Table 2. The conditional probabilities between cause nodes and fault nodes are also determined using expert judgment. For basic sensing information relevant to symptoms such as axis orbit, advanced instrumentation system in pumps provide an access to them; so sensing information from February 2018 to June 2018 is
collected and utilized to learn conditional probabilities between fault nodes and symptom nodes. Appendix provides a part of dataset used in this case study. Taking four nodes (F1, F4, F5 and S4) as an example, the CPT among them is given in Table 3.

**Table 2. Prior probability of cause nodes.**

| Node | State | Prior probability | Node | State | Prior probability |
|------|-------|-------------------|------|-------|-------------------|
| C_1  | Yes   | 3%                | C_6  | Yes   | 7%                |
|      | No    | 97%               |      | No    | 93%               |
| C_2  | Yes   | 5%                | C_7  | Yes   | 15%               |
|      | No    | 95%               |      | No    | 85%               |
| C_3  | Yes   | 8%                | C_8  | Yes   | 10%               |
|      | No    | 92%               |      | No    | 90%               |
| C_4  | Yes   | 7%                | C_9  | Yes   | 18%               |
|      | No    | 93%               |      | No    | 82%               |
| C_5  | Yes   | 13%               | C_10 | Yes   | 15%               |
|      | No    | 87%               |      | No    | 85%               |

**Table 3. Conditional probabilities among four nodes.**

| Fault | S4 (%) |
|-------|--------|
| F1    |        |
| Yes   | 100    |
| Yes   | 100    |
| Yes   | 100    |
| No    | 100    |
| No    | 97.27  |
| Yes   | 94.46  |
| No    | 94.35  |
| No    | 96.89  |
| No    | 4.62   |

4.4. Step 4: Fault inference

Before conducting a case study, the concepts of hard evidences and soft evidences must be defined. Hard evidences refer to certain symptoms whose probabilities of occurrence are 100%, while soft evidences are those whose probabilities are between 0 and 100%. Next, a case with simple evidences and two cases with complex evidences will be provided. Simple evidences is a combination of abnormal symptoms in which the class number of hard evidences is no more than three, while complex evidences refers to the combination of abnormal symptoms in which the class number of hard evidences is more than three. Worth noting that hard evidences in the complex symptoms may come from the cause layer or the symptom layer.

4.4.1. Case 1: fault inference under simple evidences. In the first case, only two hard evidences appear: abnormal radial vibration (S2 is yes) and obvious 1X characteristic frequency in the frequency spectrum (S7 is yes). The probability of single-ring ellipse (S3) is set to be 30%. Obviously, it can be seen qualitatively from figure 2 that crankshaft bending (F2) or crankshaft imbalance (F3) are likely to occur. The above symptoms are input into the BN model to update prior probabilities. The obtained posterior probability of each fault node are listed in Table 4. The posterior probability of F3 (42.87%) is the largest, followed by F2 (1.38%), indicating the most likely fault is crankshaft imbalance. In the cause layer, the probability of polluted lubricant oil C5 (16.68%) is maximum, showing that C5 is the cause of F3. In order to reduce the
damage to fracturing pump caused by impurities in lubricating oil, it is necessary to replace the oil filter or to replace the polluted lubricant in time

Table 4. Information of fault inference.

| No. | Evidence | Complexity | Description | Fault inference | Diagnosis inference | Cause inference |
|-----|----------|------------|-------------|-----------------|---------------------|-----------------|
|     |          | Simple     | S2 is yes   | P(F1|e) = 3.54%   | F3 is yes.          | C5 is yes.      |
|     |          | (symptom   | S7 is yes   | P(F2|e) = 11.38%  |                     |                 |
|     |          | layer)     | P(S3) = 30% | P(F3|e) = 42.87%  |                     |                 |
|     |          |            |             | P(F4|e) = 1.86%  |                     |                 |
|     |          |            |             | P(F5|e) = 0.27%  |                     |                 |
| Case 2 | Complex | Complex     | S2 is yes   | P(F1|e) = 50.04%  | F1 and F4 are yes. | C7 and C3 are yes. |
|        |         | (symptom   | S4 is yes   | P(F2|e) = 3.89%   |                     |                 |
|        |         | layer)     | S8 is yes   | P(F3|e) = 6.74%   |                     |                 |
|        |         |            | S10 is yes  | P(F4|e) = 45.50%  |                     |                 |
|        |         |            |             | P(S1) = 20%    |                     |                 |
| Case 3 | Complex | Complex     | C10 is yes  | P(F1|e) = 52.05%  | F1 and F5 are yes. | C3 is yes. |
|        |         | (symptom     | S1 is yes   | P(F2|e) = 16.47%  |                     |                 |
|        |         | and cause   | S2 is yes   | P(F3|e) = 4.52%   |                     |                 |
|        |         | layer)     | S5 is yes   | P(F4|e) = 3.27%   |                     |                 |
|        |         |            | S8, S10 are yes | P(F5|e) = 46.78% |                     |                 |

4.4.2. Case 2: fault inference under complex evidences. In the second case, there are four hard evidences and one soft evidence from the symptom layer: abnormal radial vibration (S2 is yes), axis orbit with double ellipse (S4 is yes), and obvious 2X characteristic frequency (S8 is yes) along with 1X secondary frequency (S10 is yes). Because the axial vibration is slightly worse than that in normal condition, the occurrence probability of S1 is set to be 20%. The inference results are listed in table 4. As shown in table 4, the crankshaft misalignment F1 and the oil-film whirl F4 have the similar posterior probabilities (50.04% and 45.50% respectively), indicating that there is a composite fault of F1 and F4. In the cause layer, excessive lubricating oil temperature C7 and uneven material C3 are the most likely causes of the above faults.

4.4.3. Case 3: fault inference under complex evidences. In the third case, the observed evidences come from cause layer and symptom layer. The hard evidences include fatigue running of crankshaft (C10 is yes), irregular axis orbit (S5 is yes), abnormal radial vibration and axial vibration (S1 is yes and S2 is yes), and obvious 2X characteristic frequency (S8 is yes) along with 1X secondary frequency (S10 is yes). The structure of BN denotes that fatigue state of crankshaft C10 would give rise to cracks on its body. However, the inference results show that the crankshaft misalignment F1 and crankshaft cracks F5 are likely to exist at the same time because of their similar posterior probabilities (52.05% and 46.78%). Moreover, the cause inference indicates F1 is caused by uneven material of crankshaft C3.

5. Conclusion
Fault inference of crankshafts in fracturing pumps is a challenging but also intellectually appealing task in engineering. Although fracturing pumps are installed with data acquisition systems that could acquire sensing information for fault analysis, there exist complex uncertain relations between faults and symptoms as well as incomplete symptom information, further increasing the difficulty of fault inference by human.
To perform fault inference under conditions of uncertainty and incompleteness, this paper proposes an intelligent fault inference method for crankshafts using a Bayesian network. The presented method integrates cause nodes, fault nodes and symptom nodes into a BN model. Using
forward-backward reasoning, fault inference and cause inference can be achieved. The application of the presented approach in fault inference of crankshafts indicates its strong inference ability under uncertain and incomplete information. The obtained results can provide scientific information for equipment maintenance support decision. Further research is required to develop real-time diagnosis system based on the proposed method and to broaden its application range.

Acknowledgements
This work was supported by the National Natural Science Foundation of China [Grant numbers: 51574263], and Science Foundation of China University of Petroleum-Beijing [Grant numbers: 2462015YQ0403, Grant numbers: C201602]. The BN model in this paper was created using BNT toolbox in MATLAB developed by Kevin Murphy. The authors would like to thank the engineers in the field for their valuable inputs to this article.

Appendix
Table A1: Partial dataset used in this study

| No. | F1 | F2 | F3 | F4 | F5 | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1   | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No |
| 2   | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No |
| 3   | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No |
| …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  |
| 105 | Yes| No | No | No | No | Yes| No | No | Yes| No | No | Yes| No | Yes| No | No |
| 106 | Yes| No | No | No | No | Yes| No | No | Yes| No | No | Yes| No | Yes| No | No |
| 107 | Yes| No | No | No | No | Yes| No | No | Yes| No | No | Yes| No | Yes| No | No |
| …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  |
| 976 | No | No | Yes| No | No | No | Yes| No | No | Yes| No | No | No | No | No | No |
| 977 | No | No | Yes| No | No | No | Yes| No | No | Yes| No | No | No | No | No | No |
| 978 | No | No | Yes| No | No | No | Yes| No | No | Yes| No | No | No | No | No | No |
| …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  |
| 4836| No | Yes| No | No | No | No | Yes| No | No | Yes| No | No | No | Yes| No | Yes|
| 4837| No | Yes| No | No | No | No | Yes| No | No | Yes| No | No | No | No | No | Yes|
| 4838| No | Yes| No | No | No | Yes| Yes| Yes| No | Yes| No | No | No | No | Yes| Yes|
| …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  |
| 9580| No | No | No | Yes| No | No | Yes| No | Yes| No | Yes| No | No | Yes| No | No |
| 9581| No | No | Yes| Yes| No | No | Yes| No | Yes| No | Yes| No | No | Yes| No | No |
| 9582| No | No | Yes| Yes| No | No | Yes| No | Yes| No | Yes| No | No | Yes| No | No |
| …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  | …  |

References
[1] Zhang X, Zhang L and Hu J 2018 Real-time diagnosis and alarm of down-hole incidents in the shale-gas well fracturing process Process Saf. Environ. Prot. 116 243–53
[2] Hu J, Zhang L, Ma L and Liang W 2010 An integrated method for safety pre-warning of complex system Saf. Sci. 48 580–97
[3] Simani S, Farsoni S and Castaldi P 2015 Wind turbine simulator fault diagnosis via fuzzy modelling and identification techniques Sustain. Energy, Grids Networks 1 45–52
[4] Ayodeji A and Liu Y kuo 2018 Support vector ensemble for incipient fault diagnosis in nuclear plant components Nucl. Eng. Technol. 1–8
[5] Yang D, Liu Y, Li S, Li X and Ma L 2015 Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm Mech. Mach. Theory 90 219–29
[6] Huang S R, Huang K H, Chao K H and Chiang W T 2016 Fault analysis and diagnosis system for induction motors Comput. Electr. Eng. 54 195–209
[7] Janssens O, Slavkovikj V, Vervisch B, Stockman K, Locuffier M, Verstockt S, Van de Walle R and Van Hoecke S 2016 Convolutional Neural Network Based Fault Detection for Rotating Machinery J. Sound Vib. 377 331–45

[8] Merabet H, Bahi T and Halem N 2015 Condition Monitoring and Fault Detection in Wind Turbine Based on DFIG by the Fuzzy Logic Energy Procedia 74 518–28

[9] Chine W, Mellit A, Lughi V, Malek A, Sulligoi G and Massi Pavan A 2016 A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks Renew. Energy 90 501–12

[10] Zhang X L, Chen W, Wang B J and Chen X F 2015 Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization Neurocomputing 167 260–79

[11] Jia F, Lei Y, Lin J, Zhou X and Lu N 2016 Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data Mech. Syst. Signal Process. 72–73 303–15

[12] Nielsen T D and Jensen F V 2009 Bayesian Network and Decision Graph

[13] Abaie M M, Abbassi R, Garaniya V, Chai S and Khan F 2018 Reliability assessment of marine floating structures using Bayesian network Appl. Ocean Res. 76 51–60

[14] Sule I, Khan F, Butt S and Yang M 2018 Kick control reliability analysis of managed pressure drilling operation J. Loss Prev. Process Ind. 52 7–20

[15] Baksh A A, Abbassi R, Garaniya V and Khan F 2018 Marine transportation risk assessment using Bayesian Network: Application to Arctic waters Ocean Eng. 159 422–36

[16] Li X, Chen G and Zhu H 2016 Quantitative risk analysis on leakage failure of submarine oil and gas pipelines using Bayesian network Process Saf. Environ. Prot. 103 163–73

[17] Malagrino L S, Roman N T and Monteiro A M 2018 Forecasting stock market index daily direction: A Bayesian Network approach Expert Syst. Appl. 105 11–22

[18] Holický M, Marková J and Sýkora M 2013 Forensic assessment of a bridge downfall using Bayesian networks Eng. Fail. Anal. 30 1–9