JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

A Novel Approach to Manage Sewage Treatment Plants Based on Process Mining Technology

Jingbo Zhao[1], Bin Shao[1], Hao Jiang[2], Chao Liu[2], and Sheng Miao[1,*]

Abstract

Due to the complexity of the operation process in sewage treatment plants, there are numerous potential risks involved in the process. An appropriate business process model is necessary for effective staff management and risk detection. However, conventional modeling methods are inherently subjective in the field of sewage treatment. Designers not only have to grasp the workflow language but also need to be familiar with the whole business process. Compared to conventional data mining, process mining specializes in end-to-end processes. Consequently, process mining is better adapted to solving process problems. In this paper, a novel approach is proposed to analyze the operational risk of sewage treatment plants by using process mining technology. The ideal Petri net model and event logs are utilized for conformance checking. The results of the experiment indicate that the proposed approach can discover operational processes from event logs in the field of sewage treatment. The method can assist managers detect staff deviating from standard operating procedures. The results of the implemented process mining technology present the sewage treatment plant managers with a real analysis and understanding to make the staff's operation easier. The methodology can be extended to similar scenarios.

Keywords

Conformance Checking, Data Mining, Management, Petri Net, Process Mining

1. Introduction

There are circumstances where staff operate improperly in the operating process of the sewage treatment plant [1]. The staff's operational effectiveness is crucial for the continuous operation of the sewage treatment plant. Currently, two serious challenges are confronted by most businesses. First, the expense of business increases with the rising cost of personnel. Second, the rapid development of business leads to the complexity of the internal processes of the organization, which are becoming increasingly less efficient. Data are managed and applied by various technologies such as artificial intelligence, and machine learning [2]. This is useless if the corresponding security activities are not properly assessed. Process mining addresses the abovementioned issues of how to utilize event logs to both recognize process-relevant risks and analyze the operational behavior of staff in business management using an objective method.

Process mining is extensively applied in various domains such as healthcare process analysis, company process interpretation, industrial process management, and more [3]. It extracts and analyzes data from

the operational processes to unveil the true performance of the business. Process mining technology plays a significant role in gaining insights and improving the flow of work [4]. The German scholar Carl Adam Petri first introduced the concept of a Petri net in 1962 [5]. Petri net is used as a basic modeling language for process mining. Petri net as a mathematical model which is used to characterize the performance of concurrent systems. It is used to describe the state and logical relationships between individual components in a system. The Heuristics net is an extended model of the Petri net. The Heuristics net is designed to enhance the expressive capability of the Petri net. It is more applicable to the modeling and analysis of actual business processes. The above two modeling languages are applied in this paper.

In the field of sewage treatment, conventional modeling methods are highly subjective. Designers need to not only master the workflow language but also be familiar with the entire operational process of the business flow [6]. They have a poor understanding of business processes. The implementation of process automation is overly focused on conventional modeling approaches. As a result, important information about business processes is often overlooked by conventional approaches [7].

This paper presents a new approach to analyze and manage operational risks in sewage treatment plants using process mining technology. The operational workflow of the secondary process in a sewage treatment plant is taken as the object of research in this paper. The ideal model is constructed by this method through the expert's event log. Potential risks are identified by the conformance-checking technique in process mining. This study employs conformance checking metrics to evaluate whether staff operations align with expected behaviors. This approach provides ample value to the present situation of model construction in the sewage treatment domain. The primary contributions of this study are listed below:

- This paper innovatively utilizes process mining technology to model business operations in the field of sewage treatment.
- The process model established in this paper supplies a useful reference for the management of daily operational processes in sewage treatment plants.

The actual process model can be used to manage and improve staff operational capability. The proposed approach can provide new insights into aspects such as optimizing process efficiency in the sewage treatment domain.

2. Related Works

Sewage treatment operations take a long time to obtain feedback on the results due to the complexity of chemical reactions in sewage treatment. If the equipment is not operated properly there will be risks and losses to the sewage plant. Therefore, the process verification is extremely significant to sewage plants. Currently data mining is applied in various fields. Zhao et al. [8] used a human-computer interaction system based on artificial neural networks and disciplinary knowledge to assess the impact of buildings on the urban environment. Yera et al. [9] proposed a data mining-based approach to solving web navigation problems. Data mining is also applied to numerous studies related to the sewage treatment domain. Kusiak et al. [10] proposed a method for modeling pumping system in sewage plants using data mining. Asadi et al. [11] proposed a novel method that utilizes data mining to enhance the aeration process in sewage treatment domain. Miao et al. [12] took advantage of machine learning and data mining to aid in the management of sewage treatment at a precision chemical plant. da Silveira Barcellos and Souza [13] presented a groundbreaking method that harnesses data mining to optimize and monitor the water quality of sewage. However, there are certain limitations of conventional data mining for problems such as process optimization and detection. The models constructed by data mining are black-box models. Conclusions are presented directly, making it challenging to trace the underlying causes. Consequently,

the emerging technique of process mining is used in this paper. It presents a visualization of bottlenecks and anomalies, which are visualized through different perspectives of the process path.

Process mining is considered as a novel and cross-cutting topic of research. It can be recognized as a measure to address the gap between process management and data science [14]. The rationale for process mining is to discover process models from the event logs stored in an information system. These process models are adopted to assist people in discovering, monitoring, and improving the actual business process. The event log serves as the starting point and analysis subject of process mining [15]. As information systems evolve, many event logs are now recorded and stored in various forms [16]. The extensible markup language (XML) format is considered the most common format for documents [17]. The transmission speeds of these two conventional formats are slow, and their structures are complex. Consequently, the extensible event stream (XES) format is used as the new format for event logs. The XES format is more applicable and efficient in the process mining field.

Process mining has four perspectives: control-flow, organization-flow, case-flow, and time-flow [18]. Control-flow primarily focuses on the sequence and logic of events [19]. Its results can be expressed in languages such as business process modeling notation (BPMN) and Petri net. The control perspective is a main field of research in process mining. The organization-flow pays attention to the participants, organizations, roles, etc., where the target is to analyze and furnish insights regarding the correlations of data components in the flow. The case-flow perspective concentrates on certain attributes of the cases that are performed through the processes. An alternative perspective is the time-flow, which grants relevant insights regarding the timestamp of a process. The data of industrial systems is stored in Internet of Things (IoT) devices [20,21]. There are various IoT devices that support the Modbus protocol used in the data system of sewage treatment plants [22].

Three primary techniques are used for process mining: process discovery, conformance checking and enhancement. The first technology is process discovery, which makes it possible to obtain actual process models from event logs. Process discovery algorithms include the alpha miner algorithm, inductive miner algorithm, heuristics miner algorithm, etc. [23]. Deviations and bottlenecks in business processes are detected through conformance checking [24]. It relates behavior in the event logs to trace in the existing model and compares both. It aims to identify the similarities and differences between the modeling process and the operational activities. The 3rd main technique of process mining is enhancement [25]. Its purpose is to refine or enhance current process models based on the extra knowledge presented by the event logs.
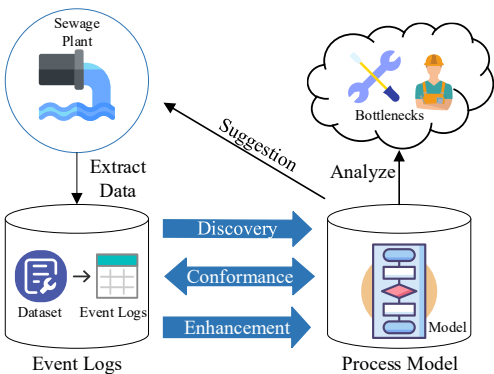
3. Methodology

Fig. 1. An approach based on process mining is proposed in sewage treatment plants.

This paper proposes a novel method to analyze operational risks in sewage treatment plants using process mining technology. The methodology is implemented in this paper consists of four main sections: data preparation, model discovery, conformance checking, and result analysis. The proposed methodology is shown in Fig. 1.

3.1 Data Preparation

Event logs are a source of data for process mining. Process mining provides a robust and established approach to obtain actual models and refine processes based on recorded data. In this paper, event logs are obtained from an industrial information system at an urban sewage treatment plant in China. The event log is a collection of events that records the operational information of a sewage treatment plant. The event logs of a sewage treatment plant must be preprocessed. This paper focuses on data collection and data cleaning of logs. The above two steps are accomplished through Python Pandas 1.4.2.

Above all, data preparation is the primary phase that initially includes extracting logs from the dataset. The managers of sewage treatment plants usually apply IoT technology to automate the treatment process. The operation information of the staff is stored in the plant's information system. Operational data from sewage treatment plants are transmitted by IoT sensors to a data transmission unit. The data are ultimately stored in the database of industrial control systems.

The second step is to clean the data. This paper considers the characteristics of the sewage treatment process. Therefore, the "staff's operation" attribute belonging to the same date is converted into the same ID. Then, any duplicate and irrelevant information in the data is required to be deleted. It is crucial to eliminate irrelevant information (e.g., outliers and noise) from the event logs to obtain better models. The activities "register", "coarse screen", and "fine screen" normally take a regular time. These activities play a minor role in our process verification. These activities are deleted in the data cleaning phase. Finally, operation names are replaced with simple activity names. The standardized event log is imported into the Process Mining for Python (PM4Py) framework. As an open-source Python library, PM4Py provides many algorithms.

Table 1. A fragment of event logs

Case_id	Activity	Timestamp	Event type	Resource
1	1HYC_HHYHLB10	2022-8-1 08:06:00	Completed	Expert
1	1SYWNB10	2022-8-1 10:23:04	Completed	Expert
2	3HYC_HHYHLB10	2022-8-2 11:03:28	Completed	Expert
1	1WNHLBf	2022-8-1 11:01:21	Completed	Expert
2	3HYC_HHYHLB10	2022-8-2 14:31:02	Completed	Expert
4	1WNHLB10	2022-8-5 10:28:15	Completed	Expert
…	…	…	…	…

Event logs are the input objects of process mining. In this paper, an expert in the field of sewage treatment is invited to manipulate the equipment for a specific period of time. The expert is well versed in each of the process aspects of a sewage treatment plant. Firstly, the event logs of the experts in sewage treatment plants are collected in this paper. Secondly, the process discovery technique is applied to convert event logs into process models. Finally, the actual operational process model is obtained. In the same way as described above, the event logs of staff members are obtained. A fragment of the expert's event log is shown in Table 1. Thus, the data are events, stored in an event log, composed of a basic set of attributes, namely the Case_id, activity, timestamp, event type, and resource attributes. The operation of the sewage treatment process is represented by the activities in the table. A timestamp is the primary

means for sequencing activities within a case as well as measuring process execution performance. Therefore, the quality of process analysis largely depends on timestamps. Table 1 describes operations performed on sewage, such as sludge absorption, sludge residue and sludge recirculation.

3.2 Process Mining

Process mining deems the discovery of models as a fundamental step in building graphical structural models based on real events. There are several discovery modeling techniques that can be used to support process discovery, such as alpha miner, fuzzy miner, heuristics miner, and inductive miner. The heuristics miner algorithm is selected to discover the process model through comparative analysis. This paper uses three state-of-the-art algorithms to obtain process models and compares the results. This paper obtains process models using three common algorithms. Obviously, alpha miner and inductive miner struggle with generating accurate models in the sewage treatment domain. This paper chooses to use the heuristics miner algorithm to discover the actual process model. It serves as an extension to the alpha miner algorithm, which addresses many of the shortcomings of the alpha miner algorithm. Process models are constructed by heuristics miner algorithm utilizing the dependencies among activities in event logs. The basic steps of this algorithm include (i) generating dependency tables by the frequency of activities; (ii) generating dependency graphs using heuristic rules; (iii) creating a causal matrix based on a dependency graph; (iv) using the alpha miner algorithm to convert causal matrices into Heuristics nets. Heuristics nets and Petri nets can be freely converted between them by PM4Py.

Petri net is a language for process modeling that supports the concurrent syntactic modeling. A Petri net N is composed of a tuple (P, T, F). The net needs to fulfill the following conditions at the same time. The basic conditions are shown in Eqs. (1)–(3):

$$P \cap T = \emptyset \tag{1}$$

$$P \cup T \neq \emptyset \tag{2}$$

$$F \subseteq (P \times T) \cup (T \times P), \tag{3}$$

P and T represent a set of places and transitions, respectively, and F represents the connection between them. F is represented by the flow relationship. P is shown as a circle, T is shown as a box, and F is shown as an oriented arrow. A token is a place-based dynamic object that can be moved from one place to another. The condition for a transition to be enabled is that all previous places of this transition must contain a token. The graphical description of the Petri net is shown in Fig. 2.
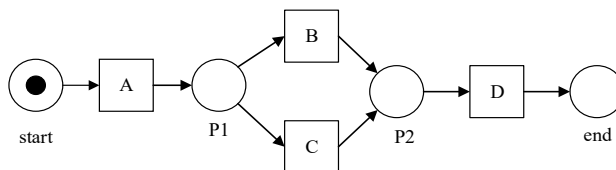

Fig. 2. Example of a Petri net.

Heuristics net is a model in process mining technology used to describe and analyze behaviors and flows in business processes. It is an extension of the alpha miner algorithm but considers the frequency of sequencing relations. It is designed to enhance the expressive power of Petri nets and make them more suitable for modeling and analyzing in real business processes. Accordingly, this approach is suitable for recognizing the primary behaviors recorded in the event logs, excluding duplicate events and certain non-free-choice structures. It adds the following features:

- Weight, which is used to indicate the importance, frequency, or time-consuming nature of changes and arcs, etc.
- Temporal, which denotes the sequence of implementation of activities in a Heuristics net.

3.3 Process Discovery and Conformance Checking

Event logs are taken as input object for creating models by process discovery techniques. Process discovery models can be obtained through several process mining algorithms. To obtain a better process model, parameters can be set by some process discovery algorithms. Unusual behavior can be ignored by parameters. The event logs are compared with process models for the purpose of conformance checking. The deviations between them can be used to diagnose the information. Operational errors and deviations are the main cause of process inefficiencies and risks. Therefore, it is important to identify them. It is crucial for the relevant staff to improve the process to eliminate these deviations. Event logs can be replayed on the model, which is considered essential for analysis. There are two replay techniques for conformance checking. The token-based replay and alignment-based replay are widely used in specific fields. The replay technique used in this paper is the token-based replay technique.

The event logs of the expert in the sewage treatment plant are first used as input for process discovery. The PM4Py library is then invoked by the Anaconda environment. The operational process model of the sewage treatment plant is discovered by the heuristics miner. The important aspect of this section stresses the conformance checking between the expert model and the staff's event logs. In this paper, the token-based replay feature of the PM4Py library is used to discover deviations between the event logs and the model in the actual process. The token-based replay technique calculates the fitness based on the quantitative relationship of all the tokens. The degree of the event log that can be replayed on the model is represented by fitness. The range of fitness values is from "0" to "1." The fitness is "1," which means that all traces in the event log can be replayed on the process model. Conversely, if the fitness is "0," then it indicates that the log is completely divorced from the actual model.

$$\text{fitness}\,(L, N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times P_{N,\sigma}}\right). \qquad (4)$$

The fitness between the event log and the Petri net is represented by Equation (4). For each phase, there are four counters shown: $P_{N,\sigma}$ (produced tokens), $c_{N,\sigma}$ (consumed tokens), $m_{N,\sigma}$ (missing tokens), and $r_{N,\sigma}$ (remaining tokens). By calculating the degree of fitness, the manager of a sewage treatment plant can check whether there is a significant deviation between the staff's operation and the expert's operation. If the behavior recorded in the event logs of staffs at a sewage treatment plant deviates significantly from the operational process of the expert, and the staffs can be evaluated or trained by the plant manager. This is advantageous for reducing the cost of a sewage treatment plant, as well as effectively preventing accidents caused by irregularities in the operation of the staff. Dangerous elements in the sewage treatment process can be identified by the conformance-checking technology of process mining. The technology is used to provide a data analysis basis for the risk assessment of sewage treatment processes. The parts that affect process performance can be easily identified by the technology. The process efficiency of sewage treatment plant maintenance equipment can be improved by conformance checking.

As a technique for checking conformance in process mining, the main idea of token-based replay is to compute fitness metrics by recording and replaying tokens. Researchers in process mining typically aim to understand and analyze the processes' behavior, e.g., to be able to identify potential problems, detect anomalous behavior, assess performance, etc. Token-based replay techniques are commonly used in areas such as bottleneck detection, performance analysis, and process optimization.

The Petri net in Fig. 2 is used in this paper to explain the token-based replay technique. An example trace is replayed by this Petri net: $\sigma = <A, B, D>$. Initially, the token-based replay technique requires four counters. Consumed tokens and missing tokens are not included in this log, so $c = m = 0$. Therefore, only p and m need to be calculated. The initial values of p and m are zero. Above all, the library is given a token by the system. P counter is increased: $p = 1$. Then A is stimulated, and the token is passed to B. A token is generated and subsequently consumed within the system. Counters are increased: $p = 2$ and $c = 1$. Then B is triggered, $p = 3$ and $c = 2$. Finally, D is triggered, $p = 4$ and $c = 3$. A token in the termination library is consumed by the system. The eventual result is $p = c = 4$, $m = r = 0$. The result is then substituted into Eq. (4). The result of token-based replay is fitness = 1.

3.4 Modeling

A model for process mining is typically represented as a Petri net that represents actual business processes. The heuristics miner automatically obtains process models by event logs from an information system in a sewage treatment plant. The following settings are made in this section using the PM4Py tool for process mining: (1) the heuristics miner is used to build models, (2) the dependency-threshold of the heuristics miner is set to 0.5. The remaining parameter settings are set by default, (3) the token-based replay method is applied for conformance checking, (4) the logs of the four staff are cleaned using the preprocessing method of Section 3.1 to obtain four standard event logs, and (5) the event logs of each of the four staff members are used to check conformance with the existing model. The quality dimensions of the process mining are used to identify staff members who are operating incorrectly. The result of the analysis is intimately related to whether a sewage treatment plant wishes to improve the actual implementation of processes.

The PM4Py framework does not support the visualization function of conformance checking for the time being. In this paper, the conformance results are visualized with the developable platform in the process mining (ProM) framework tool. The visualization plug-in used is the multi-perspective process explorer (MPE) plug-in in the ProM framework. Deviations can be found by analyzing the replay results on the model. The Flow diagram of the presented modeling strategy is shown in Fig. 3.
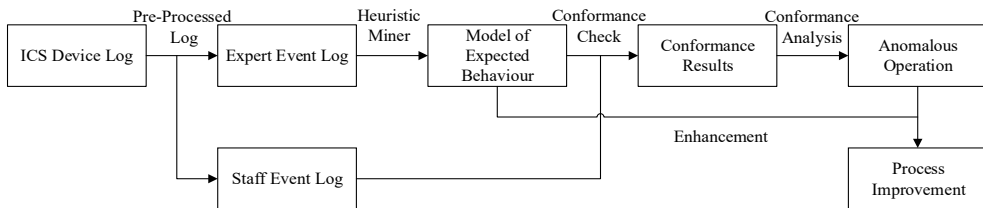

Fig. 3. Flowchart of the proposed modeling approach.

4. Experimental Analysis of Conformance Checking

In this section, this paper applies the method developed in the previous sections to an actual sewage treatment plant to verify the validity and reliability of the methodology. The modeling and analysis of the entire experiment is based on the PM4Py framework for Python. The experimental environment is set up as follows: Operating system, Ubuntu 20.04 LTS; CPU, 2 × Intel Xeon Silver 4210R; GPU, NVIDIA RTX 3090 (24 GB); memory capacity, 256 GB; Python version 3.7; PM4Py version 2.4.0; ProM version 6.9.

4.1 Dataset

This paper collected the dataset from the automation system from the sewage treatment plant. The dataset is generated from January 2022 to October 2022. The preprocessing method presented in Section 3.1 is used in this paper and the original data are processed in this paper using Python scripts. The format of the event log is then converted by this paper. The formal event log for this period is obtained in this method. Four event logs from the same time are presented in this paper. The four staff members are represented by A, B, C, and D, respectively. This paper ensures conformance by extracting event logs for the staff using the Python script. The expert event log has 343 cases, 2,094 events, and 10 event classes. In the treatment of sewage, each device has different functions. The device names corresponding to each activity name are shown in Table 2.

Table 2. The name of device that corresponds to the activity

Activity	Device name
1HYC_HHYHLB10	mixed liquid recirculation pump #1
2HYC_HHYHLB10	mixed liquid recirculation pump #2
3HYC_HHYHLB10	mixed liquid recirculation pump #3
1SYWNB10	residual sludge pump #1
2SYWNB10	residual sludge pump # 2
1WNHLBf	Frequency of sludge recirculation pump #1
1WNHLB10	sludge recirculation pump #1
2WNHLB10	sludge recirculation pump # 2
3WNHLB10	sludge recirculation pump #3
2HD_XNJ10	mud suction machine # 2

4.2 Model of Expected Behavior

The modeling approach presented in Section 3.4 is used in this section. The expert's event log is used as input object for process discovery. Three different process discovery algorithms are used to model expert event logs in this paper. These algorithms are supported in the PM4Py framework. The purpose is to select the process discovery algorithm that is most suitable for the operational process. The result of the alpha miner is shown in Fig. 4. The result shows that it is not possible to express the complete process. The approach is overly simplistic, failing to capture the inherent complexity of the process logic. The Petri net obtained by the inductive miner algorithm is shown in Fig. 5. To guarantee the accessibility of the Petri net, the noise threshold of the inductive miner is limited to 0.2. Low-frequency behaviors are filtered by the inductive miner. However, these behaviors are important for understanding and improving processes. Clearly, the Petri net shown in Fig. 5 is difficult to interpret. Moreover, there are many black boxes in the Petri net, which are invisible transitions. The Petri net has more arcs and invisible transitions. There are invisible transitions in the process model, which are not present in the event log. The invisible transition is generated from the process discovery algorithm. It is used in the Petri net only for the purpose of routing

After comparative analysis, this paper presents the final execution results using the heuristics miner. The Heuristics net aims to enhance the expressive ability of the Petri net. The result of the Heuristics net is shown in Fig. 6. The result of the model mined by the heuristics miner begins with a green circle and crosses the arrows to reach the end orange circle, whereas activities are represented by each box (Darker/ thicker blue color = more times). In the Heuristics net, sequential relationships between activities are represented by arcs. The execution frequency is represented by the value on the connecting arcs. The

total frequency of activity implementation is visualized by the numbers in the transitions. The result of the Heuristics net is shown in Fig. 6. The operating procedures for external and internal re-circulation in biochemical treatment are relatively independent. The operation of the residual sludge pump is most frequently carried out in the entire process.
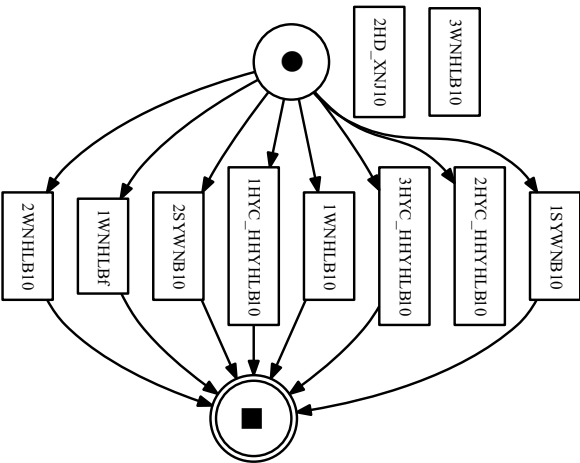

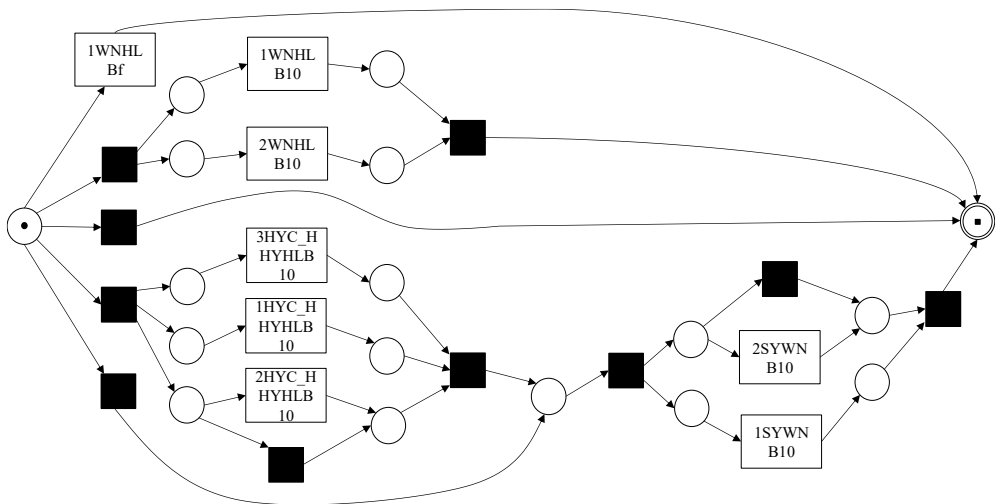**Fig. 4.** Petri net from the alpha miner.

Fig. 5. A Petri net model from the inductive miner.

The 1SYWNB10 is regulated 942 times. The 2SYWNB10 is regulated 440 times. Due to the significance of the sludge liquid level, the operation of the residual sludge pump is relatively frequent. In addition, there are two extremely low-frequency activities in the process. The 3WNHLB10 and 2HD_XNJ10 are executed 8 times and 1 time, respectively. Through the analysis conducted in this article, it can be concluded that this Heuristics net is consistent with real-world sewage treatment processes. The secondary treatment phase of a sewage treatment plant is the core stage of the entire process. Microbial flora is used throughout the process to remove organic pollutants, nitrogen, and phosphorus from the sewage.
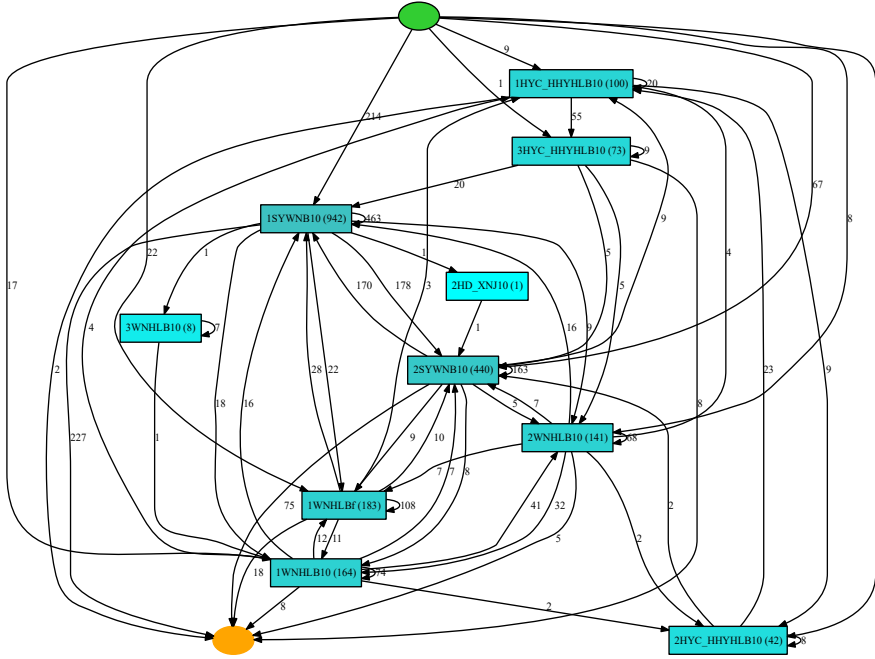
Fig. 6. A heuristics net model from heuristics miner, describing the process of sewage treatment.

4.3 Conformance Checking

The presented method is assessed on an authentic industrial dataset from the sewage treatment plant. The level of staff in the sewage treatment plant is assessed using quality indexes for process mining. Precision represents the behavioral part of the process model and can also be found in the event log. The precision value ranges from "0" to "1." The precision is "1," which means that all traces generated by the process model are included in the event log. If the value is close to "1," the model is more accurate, and conversely, if the value is close to "0," the model is underfitted. The precision index is as follows: $en_L(e)$ is the number of activities that can be observed in the event log; $en_M(e)$ is the number of tasks that are enabled in the model. The precision is calculated as shown in Eq. (5). The event log is a set of ε. Activity in the event log is represented by e. The event log is represented by L and the process model is represented by N.

$$\text{precision}(L, N) = \frac{1}{|\varepsilon|} + \sum_{e \in \varepsilon} \frac{|en_L(e)|}{|en_M(e)|}. \tag{5}$$

The F-measure index is used in this study, which is the reconciled average of fitness and precision. It can evaluate fitness and precision comprehensively. The operations of the staff are assessed by this index in this study. The formula is as follows:

$$F(L, N) = 2 \times \frac{\text{fitness}(L, N) \times \text{precision}(L, N)}{\text{fitness}(L, N) + \text{precision}(L, N)}. \tag{6}$$

The quality metrics of event logs are calculated by this paper through the PM4Py framework. The procedure is as follows: the Heuristics net shown in Fig. 6 is directly converted to Petri net using the PM4Py framework. Then the Petri net is taken as an input model for conformance checking. Four

individual event logs are available for checking conformance to the process model. Then the quality metrics function for conformance checking is called in this paper. The results of fitness, precision, and F measure obtained in this section using Eqs. (4)–(6) are shown in Table 3. The value of F-measure ranges from "0" to "1." Where the measurement of F is closer to "1," the better the quality of the sample. The F-measure for A is 0.912. The behavior is basically in accordance with the expected operating process. The high F-measure values of logs B and D indicate that almost all recorded activity sequences can be explained by this model. However, the F-measure for log C is lower than normal, demonstrating that most events cannot be replayed by the model. Therefore, the operation of the staff needs to be detected in this paper.

The Heuristics net shown in Fig. 6 can be directly transformed into a Petri net by PM4Py. In this paper, the results are visualized using the MPE plug-in. The event log and the Petri net are used as inputs for conformance checking. The result of the conformance analysis of log C in the process model is shown in Fig. 7. The frequency of execution of the activity is indicated by the shade of the color of the change. The frequency of activity execution is indicated by the number on the trace. The average time for activity execution is marked by number of times on the trace. As shown in Fig. 7, most of the model's traces are replayed on the event log. Residual sludge pumps are operated for a higher average time, totaling 3.4 hours. At the same time, the results reveal a potential problem of abnormal operation. At the same time, a potential anomalous operation problem is revealed by the visualization graphic. We find that the frequency regulation of the 1WNHLBf is the underlying cause of the reduced accuracy. This operation is not performed by Employee C from start to finish. The frequency of sludge return pumps is preset by experts. The frequency of this device is not regulated by the staff. However, the frequency of the sludge return pump needs to be regulated. If only the switching state is changed without adjusting the frequency, the cost of sludge treatment will increase.

Table 3. Results of fitness, precision, and F-measure

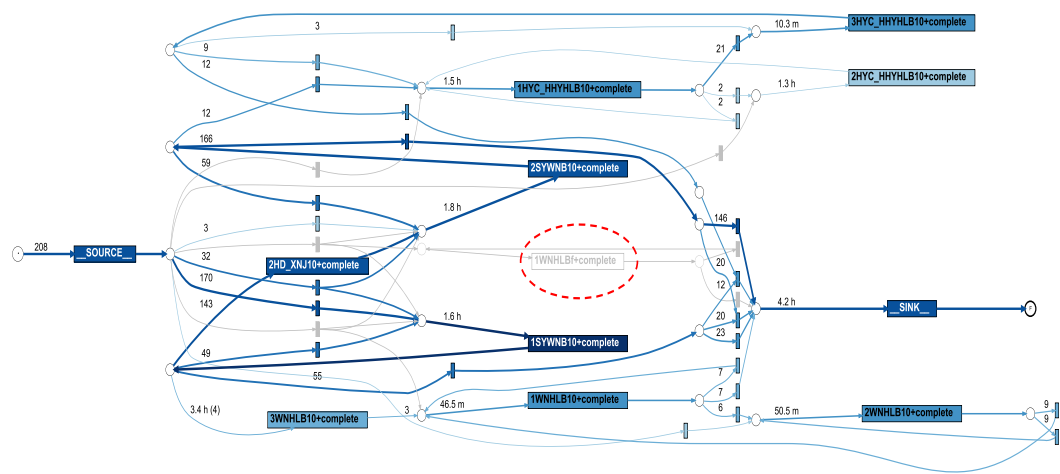|	**A**	**B**	**C**	**D**
Fitness	0.986	0.985	0.847	0.983
Precision	0.848	0.839	0.532	0.701
F-measure	0.912	0.906	0.654	0.818

Fig. 7. The result of the conformance checking.

4.4 Discussion

Because business process modeling is extensively used in process risk analysis and has excellent results, the method is process mining in this study. The expected process model is proposed in this paper. In the end, to make the research more convincing, four event logs are used in this paper to check the conformance with the process model, and the results are shown in Table 3 and Fig. 7. The results show that the expected process model is proposed in this study. The results are reflected in the staff's performance metrics. There are popular process discovery algorithms used in this paper for comparative analysis. The heuristics miner algorithm is selected for model discovery through comparative analysis. Inaccurate behavior in the process of the staff is explained by conformance checking in this study. The F-measure values are 0.912, 0.906, 0.654, and 0.818, respectively. The method can assist managers in identifying staff who are operating incorrectly.

Currently, research on data mining is mainly focused on water quality prediction, aeration control and other areas. However, it cannot model the process of event logs. Process modeling is very important for sewage treatment plants. Process mining provides the ability to analyze the operations from a process perspective. From the abovementioned discussion, it is observed that compared to conventional data mining, the process mining can be utilized to enhance the management and risk analysis of sewage treatment plants.

There is no clear set of standardized processes for all aspects of sewage treatment. In this study, a realistic operational process model of a sewage treatment plant is obtained through experimentation. This study provides new insights into the field of sewage treatment and management. The process model obtained in this paper plays an essential role in the intelligent management and operation of sewage treatment plants. This study provides new perspectives for the sewage treatment domain.

5. Conclusion

There are many operational risks associated with sewage treatment plants. To protect the safety and cost savings of sewage treatment plants, it is essential to monitor the sewage operation process reliably and effectively. This study presents a novel method to analyze the operational risk of sewage treatment plants using process mining technology. A case study of a municipal sewage treatment plant is presented in detail in this study. The ideology of process mining is used to model the sewage operation process. The quality metrics of the process mining are used to evaluate the operational level of the staff. The evaluation is based on fitness, precision, and F-measure. The results obtained from this case study are analyzed in this paper. Bottlenecks in operations are accurately identified by conformance checking. This paper overcomes the limitations of conventional modeling. Process mining is applied in sewage operation process flow to discover the actual operation process. Process models provide experts and employees with suggestions for improving existing operational processes. In addition, sewage treatment plants can improve the proficiency of staff through better management and training. Sewage treatment plant accidents can be avoided, and the cost of sewage treatment can be reduced.

Given the work in this paper, the paper outlines several fields in which future work might be carried out. This paper obtains a model of the expected operation of sewage secondary treatment. In the future work, this paper will extend the model to include the overall operational process. The overall process model will help managers evaluate the level of operation of all staff in sewage treatment. In future work, multiple aspects of the attributes will be added, including indicators for equipment and organizational

perspectives, among others. With the addition of this information, fault warning and quick start-up of the equipment will be realized.

Conflict of Interest

Funding

Acknowledgments

References

[1] I. Tuser and A. Oulehlova, "Risk assessment and sustainability of wastewater treatment plant operation," *Sustainability*, vol. 13, no. 9, article no. 5120, 2021. https://doi.org/10.3390/su13095120

[2] E. Shin, H. Yu, S. Bae, and H. B. Chang, "The impact of enterprise security performance on business performance in industrial convergence environment," *Human-centric Computing and Information Sciences*, vol. 12, article no. 33, 2022. https://doi.org/10.22967/HCIS.2022.12.033

[3] C. dos Santos Garcia, A. Meincheim, E. R. F. Junior, M. R. Dallagassa, D. M. V. Sato, D. R. Carvalho, E. V. P. Santos, and E. E. Scalabrin, "Process mining techniques and applications: a systematic mapping study," *Expert Systems with Applications*, vol. 133, pp. 260-295, 2019. https://doi.org/10.1016/j.eswa.2019.05.003

[4] J. Park, J. Y. Jung, and W. Jung, "The use of a process mining technique to characterize the work process of main control room crews: a feasibility study," *Reliability Engineering & System Safety*, vol. 154, pp. 31-41, 2016. https://doi.org/10.1016/j.ress.2016.05.004

[5] T. Murata, "Petri nets: properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541-580, 1989. https://doi.org/10.1109/5.24143

[6] G. Liu and K. Barkaoui, "A survey of siphons in Petri nets," *Information Sciences*, vol. 363, pp. 198-220, 2016. https://doi.org/10.1016/j.ins.2015.08.037

[7] W. M. Van der Aalst, M. Weske, and D. Grunbauer, "Case handling: a new paradigm for business process support," *Data & Knowledge Engineering*, vol. 53, no. 2, pp. 129-162, 2005. https://doi.org/10.1016/j.datak.2004.07.003

[8] D. Zhao, Y. Liu, G. Zeng, X. Wang, S. Miao, and W. Gao, "A knowledge-based human-computer interaction system for the building design evaluation using artificial neural network," *Human-Centric Computing and Information Sciences*, vol. 13, article no. 2, 2023. https://doi.org/10.22967/HCIS.2023.13.002

[9] A. Yera, I. Perona, O. Arbelaitz, J. Muguerza, J. E. Perez, and X. Valencia, X. (2021). Automatic Web

Navigation Problem Detection Based on Client-Side Interaction Data. *Human-Centric Computing and Information Sciences*, vol. 11, article no. 17, 2021.

[10] A. Kusiak, Y. Zeng, and Z. Zhang, "Modeling and analysis of pumps in a wastewater treatment plant: a data-mining approach," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 7, pp. 1643-1651, 2013. https://doi.org/10.1016/j.engappai.2013.04.001

[11] A. Asadi, A. Verma, K. Yang, and B. Mejabi, "Wastewater treatment aeration process optimization: a data mining approach," *Journal of Environmental Management*, vol. 203, pp. 630-639, 2017. https://doi.org/10.1016/j.jenvman.2016.07.047

[12] S. Miao, C. Zhou, S. A. AlQahtani, M. Alrashoud, A. Ghoneim, and Z. Lv, "Applying machine learning in intelligent sewage treatment: a case study of chemical plant in sustainable cities," *Sustainable Cities and Society*, vol. 72, article no. 103009, 2021. https://doi.org/10.1016/j.scs.2021.103009

[13] D. da Silveira Barcellos and F. T. de Souza, "Optimization of water quality monitoring programs by data mining," *Water Research*, vol. 221, article no. 118805, 2022. https://doi.org/10.1016/j.watres.2022.118805

[14] J. F. Rodriguez-Quintero, A. Sanchez-Diaz, L. Iriarte-Navarro, A. Mate, M. Marco-Such, and J. Trujillo, "Fraud audit based on visual analysis: a process mining approach," *Applied Sciences*, vol. 11, no. 11, article no. 4751, 2021. https://doi.org/10.3390/app11114751

[15] H. M. Marin-Castro and E. Tello-Leal, "Event log preprocessing for process mining: a review," *Applied Sciences*, vol. 11, no. 22, article no. 10556, 2021. https://doi.org/10.3390/app112210556

[16] A. Augusto, J. Mendling, M. Vidgof, and B. Wurm, "The connection between process complexity of event sequences and models discovered by process mining," *Information Sciences*, vol. 598, pp. 196-215, 2022. https://doi.org/10.1016/j.ins.2022.03.072

[17] M. Fischer, A. Hofmann, F. Imgrund, C. Janiesch, and A. Winkelmann, "On the composition of the long tail of business processes: implications from a process mining study," *Information Systems*, vol. 97, article no. 101689, 2021. https://doi.org/10.1016/j.is.2020.101689

[18] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, et al., "Process mining for healthcare: characteristics and challenges," *Journal of Biomedical Informatics*, vol. 127, article no. 103994, 2022. https://doi.org/10.1016/j.jbi.2022.103994

[19] R. Sarno, F. Sinaga, and K. R. Sungkono, "Anomaly detection in business processes using process mining and fuzzy association rule learning," *Journal of Big Data*, vol. 7, article no. 5, 2020. https://doi.org/10.1186/s40537-019-0277-1

[20] J. Jin, "Electrical automatic control system based on the Internet of Things," *Journal of Information Processing Systems*, vol. 18, no. 6, pp. 784-793, 2022. https://doi.org/10.3745/JIPS.04.0258

[21] B. Chen and D. Liu, "Remote fault diagnosis method of wind power generation equipment based on Internet of Things," *Journal of Information Processing Systems*, vol. 18, no. 6, pp. 822-829, 2022.

[22] J. Lim, "A resource management scheme based on live migrations for mobility support in edge-based fog computing environments," *KIPS Transactions on Software and Data Engineering*, vol. 11, no. 4, pp. 163-168, 2022. https://doi.org/10.3745/KTSDE.2022.11.4.163

[23] M. R. Dallagassa, C. dos Santos Garcia, E. E. Scalabrin, S. O. Ioshii, and D. R. Carvalho, "Opportunities and challenges for applying process mining in healthcare: a systematic mapping study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 165-182, 2022. https://doi.org/10.1007/s12652-021-02894-7

[24] J. Lismont, A. S. Janssens, I. Odnoletkova, S. vanden Broucke, F. Caron, and J. Vanthienen, "A guide for the application of analytics on healthcare processes: a dynamic view on patient pathways," *Computers in Biology and Medicine*, vol. 77, pp. 125-134, 2016. https://doi.org/10.1016/j.compbiomed.2016.08.007

[25] W. M. Van Der Aalst, H. A. Reijers, A. J. Weijters, B. F. van Dongen, A. A. De Medeiros, M. Song, and H. M. Verbeek, "Business process mining: an industrial application," *Information Systems*, vol. 32, no. 5, pp. 713-732, 2007. https://doi.org/10.1016/j.is.2006.05.003

Jingbo Zhao https://orcid.org/0000-0001-5493-1992

He received Ph.D. degree in control theory and control engineering from the Harbin Engineering University, in 2007. He is currently a professor in the Department of Information and Control Engineering, Qingdao University of Technology, Qingdao, China. His current research interests include Network control, robot engineering, image processing.

Bin Shao https://orcid.org/0000-0002-9789-923X

He received B.S. degree in School of Information and Electronic Engineering from Shandong Technology and Business University in 2021. Since September 2021, he is with the School of Information and Control Engineering from Qingdao University of Technology as a M.S. candidate. His current research interests include process mining.

Hao Jiang https://orcid.org/0000-0001-6174-7595

He received B.S. degree in School of Environmental and Municipal Engineering from Qingdao University of Technology in 2021. Since September 2021, he is with the School of Environmental and Municipal Engineering from Qingdao University of Technology as a M.S. candidate. His current research interests include water pollution control and environmental system analysis.

Chao Liu https://orcid.org/0009-0001-7842-9064

She is an associate professor in the School of Information and Control Engineering, Qingdao University of Technology. She received Ph.D. degree in School of Environmental and Municipal Engineering from Qingdao University of Technology in 2018. Her current research interests include water environment system analysis and water pollution control.

Sheng Miao https://orcid.org/0000-0001-6176-3624

He is an associate professor in the School of Information and Control Engineering, Qingdao University of Technology. He received his Ph.D. degree from Towson university, Maryland, USA in 2017. His research interests include machine learning, smart healthcare, and intelligence systems. He has published multiple high quality research papers in journals and conferences in recent years.