A Methodology for modeling Wastewater Treatment Process based on Uncertainty Theory

Dan Li 1,a, Tao Yuan 2,b, and Xiao Feng Liang 3,c,*

1,2 School of Environmental Science and Engineering, Shanghai Jiaotong University, China
3 State Key Laboratory of Ocean Engineering, Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration(CISSE), Shanghai Jiaotong University, China

Keywords: Wastewater treatment system; Uncertainty theory ;Bayesian network; Modified sequencing batch reactor.

Abstract. Wastewater treatment is a complicated dynamic process affected by microbial, chemical and physical factors. These variables are always uncertain. Due to the complex biological reaction mechanisms, the highly time-varying and multivariable aspects, the traditional analysis method in the description and simulation of complex reaction process and mechanism of wastewater treatment encountered challenges. However, we can use uncertainty theory to mine the rules behind the data and find the relationship between them. The Bayesian network is a powerful knowledge representation tool that deals explicitly with uncertainty. This paper employed the Bayesian network to make active exploration on the modeling of wastewater treatment system. An example is given to illustrate how to build a BN based sewage treatment system model.

Introduction

Wastewater treatment is a complex process, the effectiveness of which is affected by microbial, physical, chemical and many other factors [1]. These factors are often nonlinear and uncertainty. These reasons make the real-time prediction and diagnosis analysis in wastewater treatment system is still a big problem. The traditional analysis method in the description and simulation of complex reaction process and mechanism of wastewater treatment encountered challenges. Moreover, the traditional methods are powerless for the nonlinearity and uncertainty factors in wastewater treatment.

Bayesian networks (BNs) are an alternative technique to conventional modelling for investigating multi-factor problems [2, 3]. BNs characterise uncertainties in knowledge and use probability theories to manage these uncertainties by explicitly representing the conditional probability dependencies between variables [4, 5]. Given their solid theoretical foundation, flexible inference capability and convenient decision support mechanism, BNs have a great potential in the modeling of WWTSs [1]. Scholars have attempted to diagnose WWTSs using BN tools [6, 7]. In the present work, we proposes a method for modelling a WWTS based on MSBR via BNs. When the BN based model is established, the inference mechanism of BN can be used to predict, diagnose and optimize the wastewater treatment system.

Characteristics of the wastewater treatment system based MSBR

In this work, a six-pool MSBR was employed. The process flows of the MSBR are shown in Fig. 1.
The MSBR comprises two functional areas, the anaerobic-anoxic-oxic and SBR functional areas. Sewage initially enters the anoxic zone and then enters the anaerobic zone before eventually entering the aerobic aeration zone. During this process, organic matter is decomposed by phosphate-accumulating organisms and denitrifying bacteria in the anaerobic and anoxic zones, thereby significantly decreasing the COD concentration. Simultaneously, nitrogen and phosphorus are partially removed. After the anaerobic-anoxic-oxic reaction process, the sewage alternatively enters two SBR zones through an automatic control device. Within the SBRs, alternative anaerobic–aerobic treatments further enhance the effects of nitrogen and phosphorus removal[8].

**BN model building methods.**

The methods for building a BN model could be divided into three types: Expert experience method, Parameter learning method and Structure learning method[2, 3, 9]. Building the BN model using the first method is unrealistic because of the complexity of the WWTS. We usually choose the second or third method to build BN models based on the richness of the underlying data. In this study, the second method mentioned above was used to construct the BN for two reasons. First, the structure of a BN reveals the qualitative relationships between variables [10], and previous studies [8] could be used to develop the initial BN structure for MSBR. Second, the BN parameters reveal the quantitative relationships between variables. In addition, quantitative relationships could not be reliably obtained through expert opinion because only a few studies have quantitatively investigated MSBR systems. Consequently, data mining was employed to derive the quantitative relationships used in the BN.

**Conceptual BN of MSBR**

**Selection of BN node.** Theoretically, all factors affecting MSBR should be selected as variables in the BN. However, an overly complex model would encounter many problems, such as difficulty in collecting experimental data. Thus, the selection of variables for describing MSBR requires a compromise between model objectivity and operability. On the one hand, the selected variables must reflect the characteristics of the core processes in MSBR. On the other hand, those variables must be measurable and controllable by a treatment operator.

Our BN variables for MSBR were selected based on the reasons mentioned above. As shown in Table 1, the variables were categorised into four groups, namely, variables related to the (1) influent loads, (2) control parameters, (3) effluent concentration and (4) environment.

**Network structure development.** The network was developed on knowledge derived from the literature (Yang, 2001) and elicited from experts (Prof. Hai Zhen Yang, the inventor of MSBR; Prof. Guo Wei Gu) in MSBR technology. Qualitative relationships among the selected variables were established upon consultation with the aforementioned MSBR experts. Firstly, the conceptual BN models for nitrogen, phosphorus and COD removal were established. Afterwards, the integrated BN...
model of the MSBR system was finally obtained by incorporating the three models.

**Conceptual BN model for nitrogen removal.** The total nitrogen concentration of influent \((TN_{in})\) directly affects the total nitrogen concentration of the effluent \((TN_{out})\). Thus, the \(TN_{out}\) and \(TN_{in}\) are connected by a direct arc, which indicates the qualitative influence relationship between these two variables. Biological nitrogen removal involves ammonification, nitrification and denitrification. Ammonification can be carried out under aerobic or anoxic conditions and has a rapid reaction rate, which is not the key control factor in wastewater nitrogen removal. The major factors affecting nitrification include temperature, dissolved oxygen \((DO)\) level, pH and sludge age, and the major factors affecting denitrification are temperature, pH and \(DO\) level. Therefore, the organic concentration \((COD_{in})\), temperature \((WT)\), \(DO\), pH, sludge age \((SRT)\) and \(TN_{out}\) are connected by a direct arc. The major control parameters of the MSBR system are hydraulic retention time \((HRT)\), cycle period of SBR \((C_C)\), reflux ratio of mixture \((R)\), cycle period of SBR \((C_{tb})\), anoxic mixing time \((T_{an})\), aerobic aeration time \((T_{aa})\) and static settling time \((T_{se})\). These factors and \(TN_{out}\) are also connected by a direct arc. The removal of nitrogen, phosphorus and organic matter is affected by pH, whereas the influent loads of \(COD_{in}, TP_{in}\) and \(TN_{in}\) can affect pH. Therefore, \(COD_{in}, TP_{in}, TN_{in}\) and pH are connected by a direct arc. As shown in Figure 2, the preliminary BN model for nitrogen removal was developed on the basis of the above analysis.

During the experiment, municipal wastewater was generally neutral; therefore, its pH can be considered constant. To play the system performance, the \(DO\) of each pool was controlled within a certain range without significant changes. Therefore, \(DO\) can also be considered constant. \(Q\) was approximately 250 m\(^3\)/day, and \(R\) was fixed. Thus, both of them can also be considered constants. \(HRT\) is affected by \(Q\) and \(R\), but the actual \(HRT\) is difficult to observe and measure. Therefore, \(HRT\) was considered an intermediate variable.

On the basis of the above analysis, the preliminary BN model in Figure 2 should be modified. In model modification, the constants are removed from the conceptual model. Intermediate variables are represented by dotted lines to indicate the qualitative relationship between them; however, these variables are not involved in the parameter learning and inference calculation. The amended BN model for nitrogen removal is shown in Figure 3.

![Fig. 2. Preliminary Bayesian network model for nitrogen removal](image)

---

**Table 1. Categories of MSBR variables**

| Variable | Type | Name                          | Symbol |
|----------|------|-------------------------------|--------|
| Related to influent loads | | Influent flow | \(Q\) |
| Related to influent loads | | Influent COD concentration | \(COD_{in}\) |
| Related to influent loads | | Influent total phosphorus concentration | \(TP_{in}\) |
| Related to influent loads | | Influent total nitrogen concentration | \(TN_{in}\) |
| Related to control parameters | | Cycle period of SBR | \(C_C\) |
| Related to control parameters | | Hydraulic retention time | \(HRT\) |
| Related to control parameters | | Sediment retention time | \(SRT\) |
| Related to control parameters | | Reflux ratio of mixed liquor | \(R\) |
| Related to control parameters | | Sludge return ratio | \(R\) |
| Related to control parameters | | Anoxic mixing time | \(T_{an}\) |
| Related to control parameters | | Aerobic aeration time | \(T_{aa}\) |
| Related to control parameters | | Static settling time | \(T_{se}\) |
| Related to effluent concentration | | Effluent COD concentration | \(COD_{out}\) |
| Related to effluent concentration | | Effluent total phosphorus concentration | \(TP_{out}\) |
| Related to effluent concentration | | Effluent total nitrogen concentration | \(TN_{out}\) |
| Environmental variables | | pH | \(pH\) |
| Environmental variables | | Dissolved oxygen level | \(DO\) |
| Environmental variables | | Water temperature | \(T\) |
Conceptual BN model for phosphorus removal. For phosphorus removal, the related factors include sludge age, DO level, organic matter and temperature. Same as the nitrogen removal BN, we built a conceptual BN model for phosphorus removal, given in Figure 4.

The model shown in Figure 4 was also modified to obtain a BN model for phosphorus removal, as shown in Figure 5:

Conceptual BN model for COD removal. The factors influencing COD removal include COD loading, sludge age and DO level. Similarly, a preliminary BN model for phosphorus removal was developed, as shown in Figure 6.

Similarly, the model shown in Figure 6 was modified to obtain a BN model for COD removal, as shown in Figure 7:

The Integrated BN model for MSBR. Figures 3, 5 and 7 show three models and describe the MSBR system properties from different aspects. One or more BN models can be used to analyse the MSBR system. However, the function of a MSBR system is not to perform nitrogen removal, phosphorus removal, or COD removal alone. Thus, using a decomposition model for analysis is not comprehensive and objective. For example, the monitoring value of $TP_{out}$ was abnormal, and we employed the decomposition model of phosphorus removal for diagnosis. The diagnosis result indicated that the long $T_{as}$ caused the abnormal $TP_{out}$. Therefore, we correspondingly adjusted the $T_{as}$,
after which $TP_{\text{out}}$ normalised; however, this adjustment led to $TN_{\text{out}}$ becoming abnormal. Therefore, an integrated model is needed to carry out diagnostic analysis.

Combining the models in Figures 3, 5 and 7 produced an integrated BN model for removing nitrogen, phosphorus and COD, as shown in Figure 8.

**Model parameterisation**. The choice of number of states is an important step in developing a BN because it often affects the model sensitivity. Choosing the number of states requires a compromise between model simplicity and accuracy [11]. Pradhan suggested that two or three states are often sufficient for a BN to analyse a system[12]. The state definitions of MSBR variables are shown in Table 2.

![Fig. 8. Integrated Bayesian network of MSBR](image)

### Table 2. State definition of MSBR variables

| Variable | Symbol | Unit | State definition |
|----------|--------|------|------------------|
| Related to influent loads | | | |
| Influent COD concentration | $COD_{\text{in}}$ | mg/L | Low ≤ 200, Mid 200–400, High ≥ 400 |
| Influent TP concentration | $TP_{\text{in}}$ | mg/L | Low ≤ 3, Mid 3–6, High ≥ 6 |
| Influent TN concentration | $TN_{\text{in}}$ | mg/L | Low ≤ 35, Mid 35–70, High ≥ 70 |
| Water temperature | $T$ | °C | Low ≤ 10, Mid 10–20, High ≥ 20 |
| Related to control parameters | | | |
| Cycle period of SBR | $C_p$ | h | Short ≤ 5, Mid —, Long ≥ 5 |
| Anoxic mixing time | $T_{\text{an}}$ | min | Low ≤ 35, Mid 35–45, High ≥ 45 |
| Aerobic aeration time | $T_{\text{ao}}$ | min | Low ≤ 60, Mid 60–85, High ≥ 85 |
| Reflux ratio of mixed liquor | $R$ | | Low ≤ 1.5, Mid —, High ≥ 1.5 |
| Related to effluent concentration | | | |
| Effluent COD concentration | $COD_{\text{out}}$ | mg/L | Normal ≤ 100, Abnormal —, ≥ 100 |
| Effluent TP concentration | $TP_{\text{out}}$ | mg/L | Normal ≤ 1, Abnormal —, ≥ 1 |
| Effluent TN concentration | $TN_{\text{out}}$ | mg/L | Normal ≤ 35, Abnormal —, ≥ 35 |

*A dash (—) indicates that the value is unspecified for this variable. Only short and long states of this variable were specified.

Our research group carried out numerous studies on a small pilot of MSBR to obtain the control range of the operating parameters. To understand further the characteristics of the MSBR system, we conducted a pilot study based on laboratory experiments. In the pilot study, the sewage was taken from Shanghai Confluence Pretreatment Wastewater Plant Project from a grass town in Pudong, Shanghai; half of inoculated sludge was obtained from the excess sludge of secondary sedimentation tanks in the Eastern Shanghai WWTP, and the other half of inoculated sludge was obtained from the dewatering sludge of primary settling tanks in the Shanghai Cao Yang WWTP. The influent flow rate was approximately $250 \text{ m}^3/\text{day}$, and pH was controlled between 6.5 and 7.5.

Several working conditions were tested according to the operation mode, load condition and environmental condition. We have collected a large number of data for each condition, and more than 2000 packets of data on MSBR monitoring variables were collected. In consideration of the limited space, only a part of the test data is shown in Table 3:
Table 3 Partial test data

| No. | WT (°C) | R | C_m (h) | T_m (min) | T_a (min) | \(\text{COD}_m\) | \(\text{TN}_m\) | \(\text{TP}_m\) | \(\text{COD}_{out}\) | \(\text{TN}_{out}\) | \(\text{TP}_{out}\) |
|-----|---------|---|---------|----------|----------|----------------|-------------|-------------|----------------|-------------|-------------|
| 1   | 28      | 2 | 4       | 50       | 40       | 158.5         | 20.8        | 2.16        | 46.1          | 8.3         | 0.60        |
| 2   | 24      | 2 | 6       | 50       | 90       | 129.0         | 20.6        | 2.29        | 40.9          | 4.4         | 0.70        |
| 3   | 18      | 1 | 6       | 50       | 90       | 140.0         | 30.5        | 3.02        | 47.5          | 20.7        | 0.95        |
| 4   | 14      | 1 | 6       | 40       | 85       | 156.0         | 29.1        | 2.31        | 51.1          | 22.8        | 0.71        |
| 5   | 14      | 2 | 6       | 40       | 85       | 230.0         | 25.0        | 4.15        | 57.8          | 18.0        | 1.63        |
| 6   | 28      | 1 | 4       | 30       | 60       | 496.0         | 72.7        | 8.80        | 124.0         | 60.6        | 4.19        |
| 7   | 24      | 2 | 4       | 30       | 60       | 359.9         | 51.3        | 7.54        | 76.6          | 34.1        | 3.42        |
| 8   | 10      | 1 | 7       | 50       | 80       | 157.3         | 38.6        | 2.07        | 68.4          | 27.9        | 0.48        |
| 9   | 8       | 1 | 8       | 50       | 90       | 214.5         | 29.9        | 3.06        | 82.2          | 31.4        | 1.27        |
| 10  | 12      | 1 | 8       | 55       | 90       | 178.8         | 99.3        | 1.89        | 67.6          | 46.7        | 0.78        |

The network parameters of the nodes were computed based on training data from the pilot test. Figure 9 shows that \(\text{COD}_m\), \(\text{TP}_m\), \(\text{TN}_m\), \(C_m\), \(T_m\) and \(T\) are the root nodes. Hence, their prior probabilities are also the marginal probabilities. The marginal and joint probabilities are shown in Table 4. The conditional probabilities are not listed because of space constraints.

### Conclusion

Wastewater treatment is a complicated dynamic process affected by microbial, chemical and physical factors. Due to the complex biological reaction mechanisms, the highly time-varying and multivariable aspects, modeling of wastewater treatment system are still difficult in practice. In this research, we use uncertainty theory to mine the rules behind the data and find the relationship between them. This paper employed the Bayesian network to make active exploration on the modeling of wastewater treatment system. An example is given to illustrate how to build a BN based sewage treatment system model. The established model can provide a strong support for the prediction, diagnosis and optimization analysis of wastewater treatment system.

Table 4 Marginal and joint probabilities

| variables | Marginal and joint probabilities |
|-----------|---------------------------------|
|           | High   | Mid    | Low    |
| \(\text{COD}_m\)   | 7.89   | 65.79  | 26.32  |
| \(\text{TN}_m\)   | 39.47  | 28.95  | 31.58  |
| \(\text{WT}\)     | 39.47  | 39.49  | 21.04  |
| \(\text{T}_m\)    | 65.79  | 20.09  | 14.12  |
| \(\text{C}_m\)    | 42.11  | 31.58  | 26.31  |
| \(\text{R}\)      | 68.42  | 31.58  | 50.00  |
| \(\text{COD}_{out}\) | 26.73  | 73.27  |       |
| \(\text{TN}_{out}\)| 28.71  | 71.29  |       |
| \(\text{TP}_{out}\) | 29.70  | 70.30  |       |
References

[1] D. Li, H.Z. Yang, X.F. Liang, Prediction analysis of a wastewater treatment system using a Bayesian network, Environmental Modelling & Software, 40 (2013) 140-150.
[2] J. Pearl, Bayesian Networks, in: Handbook of Brain Theory and Neural Networks, 1995, pp. 149-153.
[3] F.V. Jensen, Introduction to bayesian networks, Springer-Verlag New York, Inc., 1996.
[4] D. Nash, D. Waters, A. Buldu, Y. Wu, Y. Lin, W. Yang, Y. Song, J. Shu, W. Qin, M. Hannah, Using a conceptual Bayesian network to investigate environmental management of vegetable production in the Lake Taihu region of China, Environmental Modelling & Software, 46 (2013) 170-181.
[5] D. Nash, M. Hannah, Using Monte-Carlo simulations and Bayesian Networks to quantify and demonstrate the impact of fertiliser best management practices, Environmental Modelling & Software, 26 (2011) 1079-1088.
[6] H.G. Chong, W.J. Walley, Rule-based versus probabilistic approaches to the diagnosis of faults in wastewater treatment processes, Artificial Intelligence in Engineering, 10 (1996) 265-273.
[7] B.S.G.E. Sahely, D.M. Bagley, Diagnosing upsets in anaerobic wastewater treatment using bayesian belief networks, Journal of Environmental Engineering, 127 (2001) 302-310.
[8] H.Z. Yang, Mechanism and technology process experiment to MSBR process for nitrogen and phosphorus removals of municipal wastewater in, Tongji university, Shanghai, 2001.
[9] N.L. Zhang, H.P. Guo, Introduction to Bayesian Networks, Science Press, Beijing, 2006.
[10] F.V. Jensen, T.D. Nielsen, Bayesian networks and decision graphs, Springer Verlag, 2007.
[11] P. Myllymaki, T. Silander, H. Tirri, P. Uronen, B-course: A web service for Bayesian data analysis, in, Institute of Electrical and Electronics Engineers Inc., Dallas, TX, United states, 2001, pp. 247-256.
[12] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, K. Huang, The sensitivity of belief networks to imprecise probabilities: An experimental investigation, Artificial Intelligence, 85 (1996) 363-397.