The following publication Man, Y., Hu, Y., & Ren, J. (2019). Forecasting COD load in municipal sewage based on ARMA and VAR algorithms. Resources, Conservation and Recycling, 144, 56–64 is available at https://doi.org/10.1016/j.resconrec.2019.01.030.

COD load forecasting model of municipal sewage for wastewater treatment plants based on ARMA and VAR algorithms

Yi Man[1,2], Yusha Hu[1], Jigeng Li[1], Mengna Hong[1], Jingzheng Ren[2,] *

1. State Key Laboratory of Pulp and Paper Engineering, South China University of Technology, Guangzhou, 510640, China

2. Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China

* Corresponding author:

Email: jzhren@polyu.edu.hk (Jingzheng Ren)

Abstract

Due to different sources and the water using habits, the influent COD of municipal sewage fluctuate sharply over time. To ensure the treatment quality of sewage, the wastewater treatment plants (WWTP) often over-aerate the air and over-add the chemicals. This results in a waste of energy consumption and increases the operation cost for WWTP. With the rapid expansion of industrialization and urbanization, municipal sewage has increased by years. Energy conservation and sustainable water management for municipal WWTP are becoming an urgent issue that needs to be solved. This paper proposes a COD load forecasting model for municipal WWTP using hybrid artificial intelligence algorithms. The auto-regressive moving average (ARMA) algorithm is used for sewage inflow forecasting, and a vector auto-regression (VAR) algorithm is used for COD forecasting. The real-time data from a municipal WWTP is used for model verification. Besides the proposed ARMA+VAR model, the BPNN, LSSVM, GA-BPNN based COD load forecasting models are also studied as the contrasting cases. The verification results reveal that the ARMA+VAR model is superior to the other forecasting models for future application in the wastewater treatment plants. The accuracy of the proposed model is as high as 99%.

Keywords: municipal sewage; wastewater treatment plants; COD load; forecasting model; sustainable water management

1. Introduction

With the rapid expansion of industrialization and urbanization, the quantity of municipal wastewater effluent has been growing at a rate of 5% per year over the past decade (Yang, et al., 2017). The energy consumption for municipal sewage treatment is constantly rising. How to reduce energy consumption is an issue that must be solved in municipal wastewater treatment plants (WWTP) (López-Morales & Rodríguez-Tapia, 2019).

The activated sludge process that contains secondary bio-treatment process has been applied in most of municipal WWTP in China (Man et al., 2017). The energy consumption is mainly concentrated in the influent pump station for improving sewage and the aeration system for the secondary bio-treatment process. The energy consumption of these two operation units accounts for about 70% of total energy consumption (Man et al., 2018). The power consumption of the aeration system generally accounts for 40% to 50% of the whole plant (Li et al., 2017). It is the largest power consumption operation unit in the municipal WWTP.

Chemical oxygen demand (COD) is one of the most commonly measured items in water quality monitoring and analysis. It directly reflects the extent of contamination of the water which is polluted by reducing substances (Wang at al., 2018). COD is one of the most important indicators to demonstrate whether the effluent fits the discharge standard after treatment. Municipal sewage mainly comes from the urban human living area, precipitation, and some industrial wastewater. Unlike the industrial wastewater, the influent COD load of municipal sewage plants has changed greatly due to the difference in climate change and living habits of residents. To ensure that the treated effluent can meet the discharge standard, the COD content of the discharged effluent should be monitored in WWTP. The aeration rate and chemicals dosage should be controlled according to the COD content of the discharged effluent (Babu & Reddy, 2014). However, the COD detection needs a quite long time and it is an off-line operation process, which will cause problems such as time lag and inaccurate feedback during the process control. Meanwhile, due to the wide range of municipal sewage sources and the large fluctuation of influent mass flow, a large design margin is often reserved for aeration process in WWTP. In the treatment process, the air flow is often over-aerated and chemicals are over-added in order to ensure treatment quality of sewage when the sewage inlet mass flow or the COD content fluctuates sharply. However, in spite of well effluent quality control, this operation not only sacrifices a large amount of unnecessary energy input, but also causes problems such as secondary contamination of chemicals (Sen et al., 2016). Moreover, the excessive dissolved oxygen will cause the destruction of the flocculating agent and result in poor settling of suspended solids, thus reducing the quality of effluent. If the aeration rate and the chemicals dosage in the treatment process can be accurately controlled by establishing a "feed-forward and feedback" control system, the energy consumption and cost of the treatment process can be both reduced on the premise of ensuring the effluent quality. However, the influent COD load of the sewage must be forecasted for establishing such "feed-forward and feedback" control system.

Some research achievements have been made on the forecasting of municipal sewage quality based on different mathematics or mechanism models, such as regression model (Park & Engel, 2015; Suchetana et al., 2019), grey forecasting model (Chen et al., 2010), neural network (Vrečko et al., 2011; Gebler et al., 2018), and auto-regressive moving average (ARMA) (Yuan et al., 2016; Barak & Sadegh, 2016). Due to the simple mathematical structure, the mechanism model has the advantages of fast convergence speed and high forecasting accuracy for stable data sequence. However, the accuracy will largely decrease when the raw data fluctuates sharply because such models usually pay much attention to data fitting for the search of data sequence rule. The heuristic algorithms such as neural network, particle swarm optimization (PSO), etc. with strong adaptability and learning ability are usually used for dealing with nonlinear and uncertain problems. However, they are easy to appear the shortcoming such as long learning time and local optimization, which results in non-convergence and reduces the industrial application scope (Son & Kim, 2017; Ye et al., 2018). The time series based forecasting method is a kind of intelligent algorithms based on the essential law of data reflected by time series. Compared with other intelligent algorithms, the most outstanding advantage of the time series based algorithms is that they can rapidly capture the trends of the data sequence. The rapid calculation process opens up possibilities for its industrial applications. In recent years, although there are many applications in of forecasting models based on time series algorithms (Deng & Wang, 2017; Deng et al., 2015), it is still in the initial stage for the application for sewage treatment.

In order to increase the control accuracy of the aeration process, this paper proposes a COD load forecasting model for municipal sewage based on ARMA and vector auto-regression (VAR) algorithms. The industrial real-time data is used for modeling and model verification. The proposed COD load forecasting model will provide a scientific basis for precise control of the aeration rate, which will reduce energy consumption and operation cost.

2. Materials and methodology

In the WWTP, the COD load is usually used as an indicator of aeration and chemical dosage. The COD load is the product of the sewage mass inflow and the absolute value of influent COD. Since the amount of sewage mass inflow and COD are two independent variables, they can be modeled separately and thereby obtaining the forecasting model of influent COD load.

Affected by residents living habits and precipitation, the mass flow of municipal sewage influent presents the characteristics of strong timeliness and seasonality. Therefore, the ARMA algorithm is used to model municipal sewage inflow in this paper. The influent COD is related to many internal correlation factors or variables. It is necessary to analyze the influence of the variables on the influent COD in time series. Therefore, the VAR algorithm is used to forecast the COD of municipal sewage inflow in this paper.

This research consists of 4 steps, as shown in Fig. 1. (1) Data collection: The data in this paper come from the real-time data from a municipal WWTP. (2) Data pre-processing: The real-time data usually have problems such as data missing and error, it is necessary to filter the error data and fill up the missing data. The data preprocessing will help to improve the accuracy of the model. (3) Modeling: The sewage influent mass flow forecasting model is established based on ARMA algorithm, and the influent COD forecasting model is established based on VAR algorithm. (4) Forecasting and verification: The influent COD load is forecasted and the industrial real-time data are used to verify the accuracy of the forecasting model.

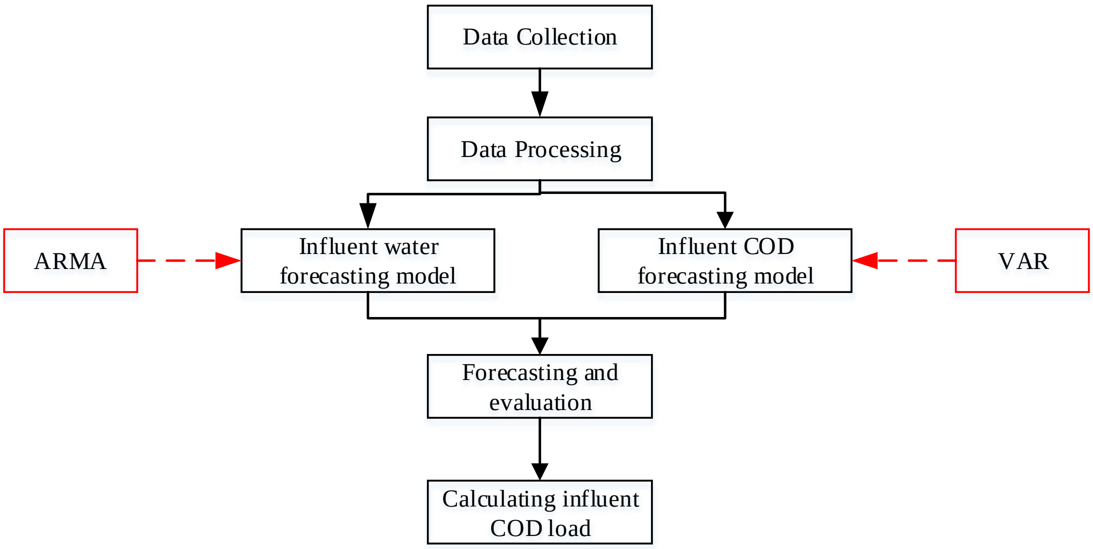


Figure 1. Roadmap of the research

2.1. Data preparation

The original data used in this paper are collected from a municipal WWTP in Qingyuan, Guangdong Province. The annual treatment capacity of this WWTP is 5 million tons. The temperature of sewage varies from around 8 °C to 30 °C. This research is carried out based on the A2O wastewater treatment technology. The collected real-time data is obtained from the historical database of the WWTP. The sampling frequency of sewage inflow is every 1 hour and influent COD is every 1 minute.

Since the object of this paper is to obtain the phase forecasting model of sewage inflow and influent COD based on time series analysis, the relevant factors that affect the influent water inflow and influent COD are analyzed and selected in this section.

Unlike the industrial WWTP, the sewage inflow of municipal WWTP is mainly related to human water using habits and natural precipitation. The former enables the water inflow to present a strong cyclical change, and the latter results in an abrupt change of water inflow in the time series. Therefore, it is necessary to introduce precipitation data during the modeling process to forecast the sewage inflow of municipal WWTP.

The influent COD is affected by the pH, the concentration of ammonia and nitride (NH_3-N), and the influent sewage temperature (T). The time sequence $\{Z_{t1}\}$ of the influent COD (mg•L-1), the time sequence $\{Z_{t2}\}$ of influent pH, the time sequence $\{Z_{t3}\}$ of NH_3-N ($mg{\cdot}L^{-1}$), and the time sequence $\{Z_{t4}\}$ of influent sewage temperature (T) are selected as the model input variable.

2.2. ARMA algorithm based forecasting model

ARMA algorithm is an effective method to forecast time series based data sequence, which can be explained by the time-delay term and random error term of variable μ. ARMA algorithm can find a suitable forecasting model on the premise of the given data pattern. The algorithm of the auto-regressive moving-average model (p, q) is as shown in Eq. (1) (Wang et al, 2018):

$$\mu_t = c + \varphi_1 \times \mu_{t-1} + \cdots + \varphi_p \times \mu_{t-p} + \varepsilon_t + \theta_1 \times \varepsilon_{t-1} + \cdots + \theta_q \times \varepsilon_{t-q}, \quad t = 1, 2, \ldots, T \quad (1)$$

Where, c is a constant, $\varphi_1, \varphi_2, \ldots, \varphi_p$ are the autoregressive model coefficient, P is autoregressive model order; ε_t is white noise series that the mean value is 0 with the variance δ^2, μ is a constant parameter, $\theta_1, \theta_2, \cdots, \theta_q$ are coefficients of the q-order moving average model.

The ARMA based sewage inflow forecasting model has two main procedures: Firstly, based on the preprocessed data, the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) are calculated to identify the model and the estimate the parameters; secondly, the preliminary model and the estimated model parameters shall be certified. The Akaike Information Criterion (AIC) method is used to certify and determine the appropriate order of the model. The flow diagram of the ARMA modeling flowchart is shown in Figure 2.

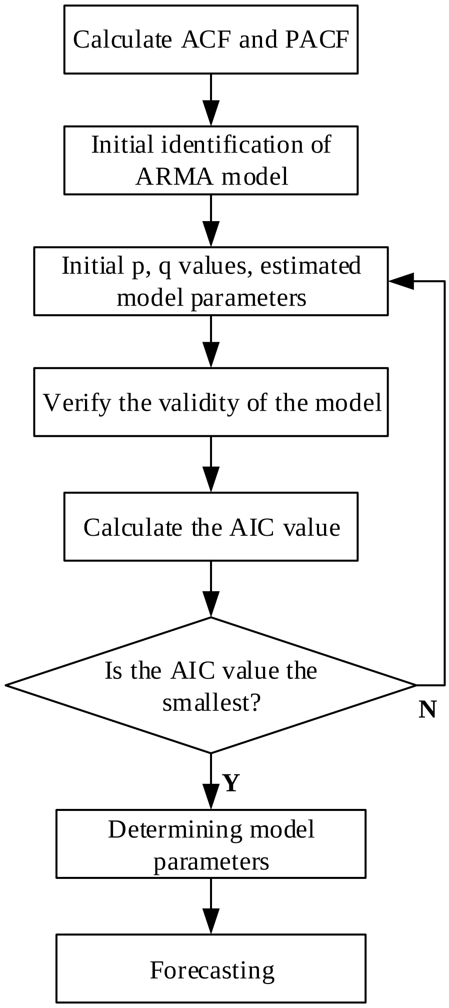


Figure 2. The programming chart of ARMA

The specific steps for the modeling process are as follows:

(1) Model identification: Since only the time series data are available to be obtained, the ARMA (*p, q*) model should be identified based on the two statistics parameters, autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF).

The parameter selection principle: The values of *p, q* are determined by the truncation and tailing characteristics of ACF and PACF. With the increase of lag order, if AC or PAC shows sinusoidal attenuation or exponential attenuation approaching zero, they have trailing property. If AC or PAC quickly approaches 0 from a certain lag period, it has truncation. By this method, only preliminary order determination can usually be carried out. For further precise order determination, it shall be tested from bottom to top. In this paper, the most widely used AIC method is used to determine the order determination of the model.

(2) Parameter estimation. The most commonly used methods, nonlinear least squares method (NLLS), is used to estimate the parameters in this paper.

(3) Model verification. Check whether the residual sequence of the fitted model is a white noise sequence. If the residual error meets the requirement of white noise sequence, the model selection is reasonable; otherwise, repeat the steps (1) ~ (2) until the appropriate model is determined.

(4) The model order determination. The AIC values of the verified model with different orders are then calculated based on the AIC method. The model order is determined when the smallest AIC value appears.

2.3. VAR algorithm based forecasting model

The VAR algorithm structures the model by using each endogenous variable as a function of the hysteresis value of all endogenous variables. The VAR algorithm is similar to the multivariate linear regression model that is widely used in multivariate statistical analysis. Therefore, many methods involved in multivariate linear regression with multiple dependent variables can be applied to the VAR model.

Proceed from the data; VAR algorithm does not contain exogenous variables. The mathematical form of the model with p-order is shown in Eq. (2) (Chan & Eisenstat, 2018):

$$Z_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_0 \times Z_{t-i} + a_t \quad (2)$$

Where: Z_t is a multivariate time series with one-dimensional endogenous variables, Φ is the one-dimensional constant vector. When Φ is not equal to 0, Z_t is a random vector sequence with independent and identical distribution. The mean value is 0.

This equation of model is convenient to analyze the dynamic relationship between endogenous variables. The dynamic relationship is the relation between the variable to be studied as well as the p-phase lag of itself and other variables. In view of the backward shift operator, the model is converted into Eq. (3):

$$\Phi(\mathrm{B}) \times Z_t = \Phi_0 + a_t \quad (3)$$

Where, $\Phi(\mathrm{B}) = I_l - \sum_{i=1}^{p} \Phi_i \times B^i$, it is a matrix polynomial with p-order.

After the preliminary model is determined, it needs to be tested. The residuals, which plays an important role in the modeling process, need to be tested. After the establishment of the model, it is more important to test the stability of the model. If the VAR based model is stable, it will not produce spurious regression and is trusted to be effective for practical forecasting. The content of the model test mainly includes two parts: (1) Ensure the stability of the model; (2) Give the direction of further improvement if necessary. In this paper, the residuals of the model are tested by the multivariate portmanteau test method. The null hypothesis of the test method is: H_0: $R_1 = \cdots = R_{m\text{-}0}$, the alternative hypothesis is: H_1: $R_j \neq 0$, $\exists j \in [1,m]$, where m is a predetermined positive integer. The sequence of residuals can be calculated by Eq. (4)

(Patilea & Raïssi, 2015):

$$Q_k(m) = T^2 \times \sum_{P=1}^{m} \frac{1}{T-P} \times tr(\widehat{C_p'} \times \widehat{C_0^{-1}} \times \widehat{C_p} \times C_0^{-1}) \sim \chi^2((m-p) \times k^2) \quad (4)$$

where, $Q_k(m)$ is a chi-square distribution with a progressive order of freedom $(m\text{-}p) \times k^2$.

Since the VAR based model is established based on time-series data, the VAR based model is a non-theoretical model in practical applications and the influence of the variables of the model is determined by Granger causality method. The details of this method are shown in the Appendix. In the meanwhile, another method, Impulse Response Function (IRF), is used to explore the relationship between variables. When calculating the impulse response, the model must be guaranteed to be stable. If the model is unstable, the impulse response of a changing model has not only the effects of disturbances, but also the effects of changes in the system itself in the calculation process. The impulse response can be used to describe the dynamic response of disturbances generated by one endogenous variable to other variables in the VAR based model. The variance decomposition of the error is also used to further evaluate the importance of different impacts by analyzing the contribution of endogenous variables. The specific modeling process is shown in Figure 3:

(1) Augmented Dickey-Fuller (ADF) test: Judge whether the sequence is stable by the ADF test. If it fails to pass the ADF test, the difference of the sequence is carried out until the sequence passes the ADF test.

(2)Initial model order selection. The information criterion values in different orders are calculated and ranked. The model order p referred to the smallest information criterion values will be selected. This paper calculates the information criterion values at different model orders by using the AIC method, Bayesian Information Criterion (BIC) method and Hannan-Quinn Criterion (HQC) method. The smallest p value is selected for model initialization.

(3) Granger causality test. The input variable of the VAR based model is selected in order to analyze the causality between influent COD and other influent variables by using Granger causality test method.

(4) Model order determination. The model order is determined by the IRF method. The preliminary several different orders are selected, and the VAR model is established to determine whether the effect of different single variables on other variables is consistent with the established VAR mode. The corresponding order is selected as the order of the final model to establish the VAR forecasting model.

(5) Model test and modification. The cross-correlation of the residual error for the preliminary model is tested by the multivariate portmanteau test method. When the residual error has no strong correlation or cross-correlation, the validity of the model is determined. Otherwise, repeat the steps (2) ~ (4).

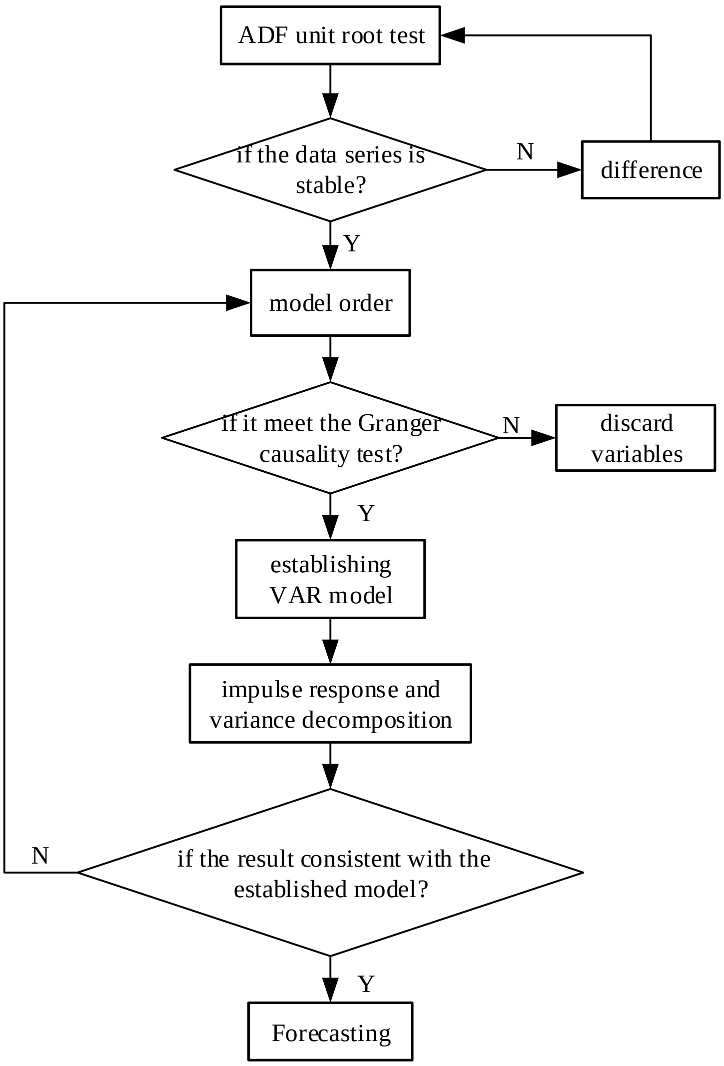


Figure 3. The programming picture of VAR

3. Results and discussion

3.1. The sewage inflow forecasting model

The original data of the sewage inflow from June 3, 2018 to June 23, 2018 with a sampling time of every 1 hour are collected from a municipal WWTP in Qingyuan. After preprocessing, a total of 481 sampling points are obtained for the cumulative sewage mass inflow. And the mass inflow of the sewage for every 1 hour is obtained by doing the first-order difference for the cumulative inflow data. Figure 4 shows the preprocessed data for sewage mass inflow, where the red part is the rainfall period released by the local meteorological department.

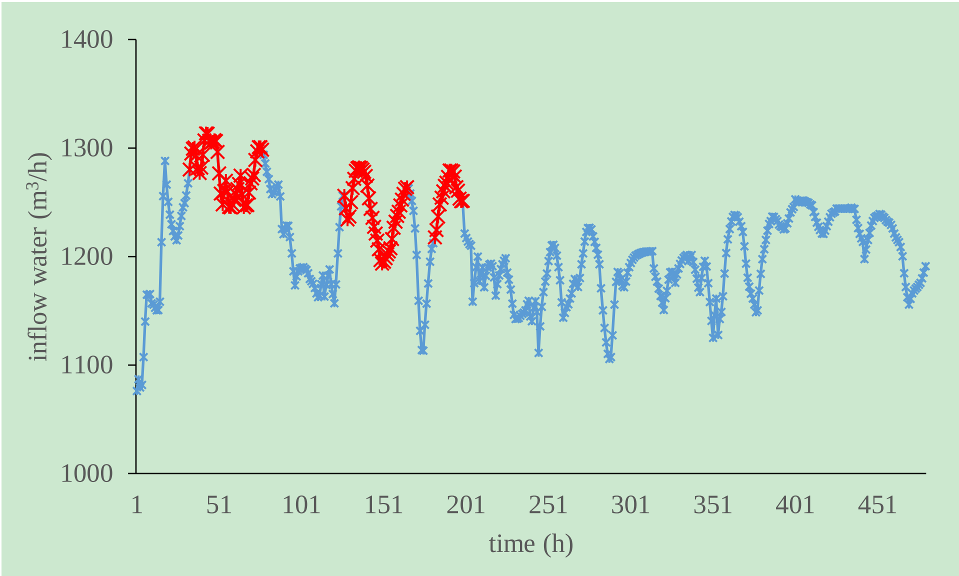


Figure 4. Preprocessed data for sewage mass inflow

The preprocessed data are divided into two parts. One part including the data of the first 19 days is used to train the model parameters. The other part including the data of the 20th day is used for model testing. Since the original data sequence fails to meet the stability requirement of the ARMA algorithm, the data sequence needs to be differentiated. Here, the first-order difference of the data sequence can be carried out to meet the sequence stationarity, and then the relevant ACF and PACF are solved to judge the tailing and truncation of the model to select the appropriate model order. The NLLS method is used to estimate the model parameters, and the model lag of the determined coefficients is obtained to test the model. The rationality of the model is judged by whether the residual error is the white noise sequence. Finally, the AIC method is used to determine the order of the model, as shown in Table 1. According to Table 1, it can be found that the AIC value is the smallest when p=5, and q=3. Therefore, the forecasting model of sewage inflow per unit time in the sewage treatment plant is ARMA (5, 3).

Table 1. AIC value table for different p and q orders

q value / p value	1	2	3	4	5	6
1	4.6678	4.6207	4.6252	4.6291	4.6318	4.6266
2	4.6237	4.6195	4.6190	4.6232	4.6261	4.5754
3	4.5520	4.6244	4.6288	4.5721	4.5755	4.5614
4	4.5579	4.6074	4.5669	4.5641	4.5977	4.6043
5	4.5617	4.5673	4.5135	4.5273	4.5951	4.5188
6	4.6152	4.5991	4.5202	4.5745	4.6036	4.5535

At the same time, the adaptive mechanism is used in the model for rolling forecasting. The collected real-time sewage inflow data during the forecasted time period will be added to the historical sewage database for re-calculating the parameters of the ARMA (5, 3). The updated forecasting model with new parameters is then used to forecast the next time sewage inflow data. In this way, the dynamic forecasting model is established by modifying the parameters in real time.

3.2. The influent COD forecasting model

The original data of influent COD is also collected from this municipal WWTP. The sampling time for influent COD is every 1 minute. The data preprocessing method of the influent COD is similar to the method for the sewage inflow. The preprocessed data of the influent COD is shown in Figure 5. The variation trend of four groups of correlation parameters (influent COD, pH, NH_3-N, and temperature) in 28 hours is also shown in Figure 5. The variation range of influent COD is between [146, 156]. The general trend is not evidently related to the human water using period. The influent NH_3-N fluctuates within the range of [72, 78]. Combined with the variation range of influent pH and temperature, it can be found that the correlation between the influent COD and other variables of the influent has a mutual influence. However, the appropriate influencing variables shall be selected in combination with the quantitative mathematical analysis.

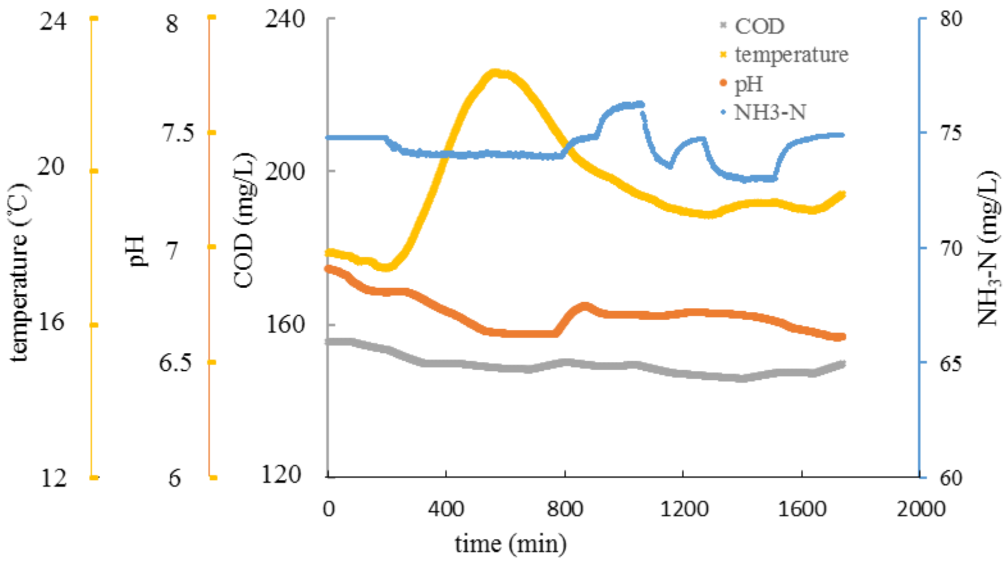


Figure 5. The variation range of COD and the related variables

3.2.1. ADF unit root test

In the ADF unit root test, the availability of intercept and time trend items has a significant impact on the results of the test. From Figure 5, it can be found that the four variables do not show the consistent trend, so the ADF test with an intercept but without trend is used. If the ADF test value is less than the critical values of 1%, 5%, and 10%, it indicates that the data sequence is stable. The test results are shown in Table 2, where: Z_{t1} is the influent COD in mg/L. Z_{t2} is influent pH, and Z_{t3} is influent NH_3-N in mg/L. $\triangle Z_{t3}$ is the influent NH_3-N with the first order difference. $\triangle Z_{t4}$ is the influent sewage temperature with the first order difference.

According to the ADF test results it can be found that the COD and pH are stable at different significance levels on time series, while the NH_3-N and temperature are unstable. However, these two variables are stable with a first-order difference. Therefore, $\triangle Z_{t3}$ and $\triangle Z_{t4}$ are selected as the input parameters together with Z_{t1} and Z_{t2}.

Table 2. ADF unit root test results

Variable	ADF test value	1% threshold	5% threshold	10% threshold	*p* value	Conclusion
Z_{t1}	-10.6451	-2.5691	-1.9416	-1.6168	0	Stable
Z_{t2}	-11.1794	-2.5691	-1.9416	-1.6168	0	Stable
Z_{t3}	0.0518	-2.5691	-1.9416	-1.6168	0.349	unstable
$\triangle Z_{t3}$	-58.0250	-2.5691	-1.9416	-1.6168	0	Stable
Z_{t4}	3.9861	-2.5691	-1.9416	-1.6168	0.281	Unstable
$\triangle Z_{t4}$	-7.8256	-2.5691	-1.9416	-1.6168	0	Stable

3.2.2. Model order selection

Table 3 shows the test results of different information criterions with a maximum lag order of 13. When the lag order is 9, the BIC shows the minimal information content. When the lag order is 10, the HQC shows the minimal information content. The information content of AIC is decreased with the increasing lag orders. The results indicate that different information criterions have different emphases due to the different penalty factors of them. In the comparison of the results of AIC, BIC, and

HQC, it can be found that BIC and HQC are consistent to some extent: With the increasing of model order p, the trend of BIC and HQC is almost the same, they both show the trend of decreasing first and then increasing. From Table 3, the model order selection by AIC needs to be beyond 13th order, while 9th order by BIC, and 10th order by HQC. Therefore, p=9, namely VAR(9), is firstly selected. Since all the results of AIC, BIC, and HQC show a slow decreasing trend after 3rd order, p=3, namely VAR(3), is therefore selected.

Table 3. Statistical results of different information criteria for different lagged orders

P	AIC	BIC	HQ	*p*-value
0	22.8746	22.8746	22.8746	0
1	10.2092	10.2374	10.2196	0
2	3.7387	3.7593	3.7596	0
3	2.0151	2.0999	2.0464	0
4	1.9391	2.0521	1.9809	0
5	1.6998	1.8411	1.7521	0
6	1.5770	1.7673	1.6604	0
7	1.5164	1.7143	1.5896	0
8	1.4483	1.6744	1.5319	0
9	1.1692	1.4236	1.2633	0
10	1.1500	1.4326	1.2545	0
11	1.1495	1.4603	1.2644	0.0313
12	1.1399	1.4791	1.2654	0.0001
13	1.0989	1.4663	1.2348	0

3.2.3. Granger causality test

Granger causality test is used to analyze the causality between influent COD and other influent variables. The results of the Granger causality test is shown in Table 4. The sig value is the indicator of the credibility. It is the error probability of the results. The higher the sig value means less credibility. A sig value of 0.05 is generally considered to be acceptable at the wrong boundary level. That means if the sig is lower than 0.05, the original hypothesis needs to be rejected. Otherwise, it is acceptable.

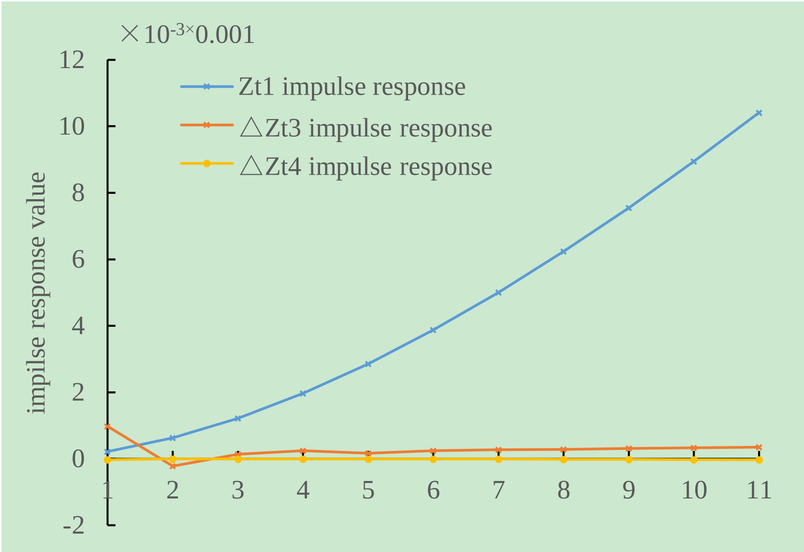
The test results show that influent pH is not the Granger cause of influent COD; in the meanwhile, the influent COD is not Grange cause of the influent pH, and there is no statistical causality between the two parameters. However, the NH_3-N with first order difference and the temperature first order difference are the Granger cause of influent COD. The three correlation parameters are interacted.

Table 4. Granger causality test results

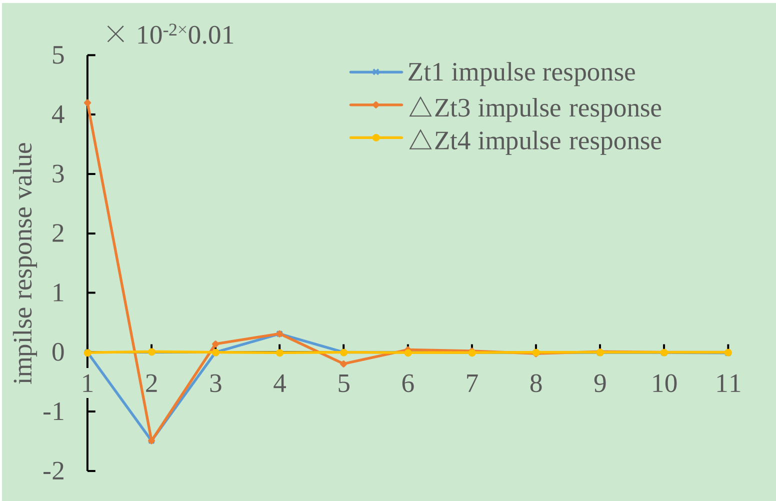
Null hypothesis	F-statistics	Sig value
Z_{t2} is not a reason for Z_{t1}	10.736	0
$\triangle Z_{t3}$ is not a reason for Z_{t1}	1.6099	0.2002
$\triangle Z_{t4}$ is not a reason for Z_{t1}	1.6815	0.1864
Z_{t1} is not a reason for Z_{t2}	10.736	0
Z_{t1} is not a reason for $\triangle Z_{t3}$	0.0124	0.9877
Z_{t1} is not a reason for $\triangle Z_{t4}$	5.1059	0.0062

3.2.4. Model order determination

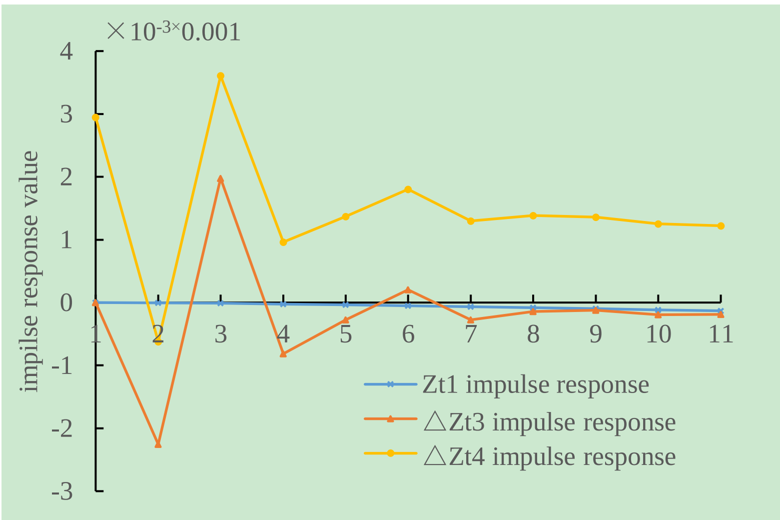
The impulse response results of different associated variables to the influent COD is shown in Figure 6. The impulse response of COD and influent temperature with the first-order difference to themselves is raising with the increasing of model forecasting period, as shown in Figure 6 (a) and (c). However, the impulse responses of the three variables to the other variables are quickly attenuated to zero, which indicates the variables have influence relationship.



(a) The impulse response function caused by Z_{t1}



(b) The impulse response function caused by $\varDelta Z_{t3}$



(c) The impulse response function caused by ΔZ_{t4}

Figure 6. Impulse response for different variables with different model forecasting period

In order to understand the contribution of each variable to the influent COD, the variance decomposition shall be carried out. The results obtained by the decomposition are shown in Table 5. It can be found that the main contribution of COD is relatively large in the first forecasting period. With the increasing of forecasting period, the influence of ΔZ_{t3} and the ΔZ_{t4} on COD increase gradually. The impulse response begins to decrease after the third forecasting period. It means the first three forecasting period has the highest influence on COD. Therefore, VAR(3) is finally selected as the influent COD forecasting model.

Table 5. Variance decomposition of model forecast period

Period	Forecast variance decomposition		
	Z_{t1} variance decomposition	ΔZ_{t3} variance decomposition	ΔZ_{t4} variance decomposition
1	1.0000	0	0
2	0.9999	0.9969	0.9986
3	0.99986	0.9949	0.9987
4	0.9997	0.9946	0.9987
5	0.9997	0.9945	0.9987

3.2.5. Model test and modification

Once the initial model has been obtained, the cross-correlation of residual error needs to be tested by the multivariate portmanteau test method. When $m>m_0$ (m_0: determined model order), $p<0.05$, there is no strong correlation or cross-correlation between the residual errors to determine the validity of the model. Otherwise, the model order determination shall be carried out again.

According to the obtained statistics results of the multivariate portmanteau test method, there are 9 parameters of VAR(3) model. As a result, the order of freedom of chi-square distribution of the test statistics $Q_k(m)$ is set as $9m$-9. For the VAR(9) model, there are 27 parameters. Thus the order of freedom of chi-square distribution of the test statistics $Q_k(m)$ for VAR(9) is $9m$-27. The p values of the two model test statistics $Q_k(m)$ are given in Table 6. For the VAR(3) model, when m>3, p<0.05. That means the residual errors of the established the VAR(3) model have no strong correlation or cross-correlation at a significant level of 5%. However, for the VAR (9) model when m=5 ($m<m_0$), p>0.05. That means the VAR(9) model have a strong correlation or cross-correlation and the forecasting result is not reliable. Therefore, the VAR(3) model is finally selected as the influent COD forecasting model. Here, the core equation of the influent COD forecasting model is shown in Eq. (5):

$$Z_{1,t} = 2.9028 \times Z_{1,t-1} - 0.001 \times \Delta Z_{2,t-1} - 0.00147 \times \Delta Z_{3,t-1} - 2.8073 \times Z_{1,t-2} + 0.000818 \times \Delta Z_{3,t-2} + 0.9045 \times Z_{1,t-3} + 0.000818 \times Z_{3,t-3} \tag{5}$$

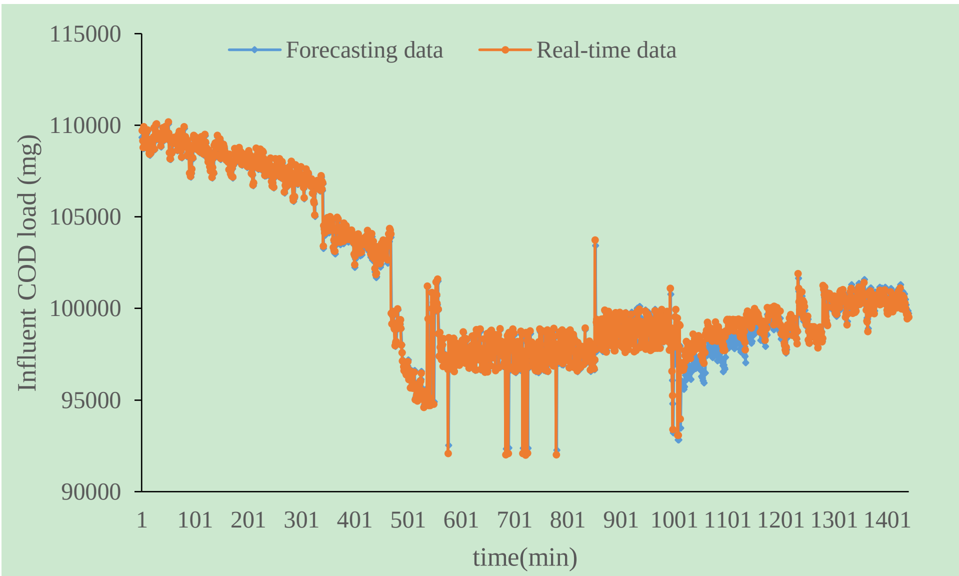
Table 6. The Q- statistic test value of different VAR models

m	The p-value of Q-statistics of VAR(3)	The p-value Q-statistics of VAR(9)
1	1	1
2	1	1
3	1	1
4	0	0.02
5	0	0.16
6	0	0

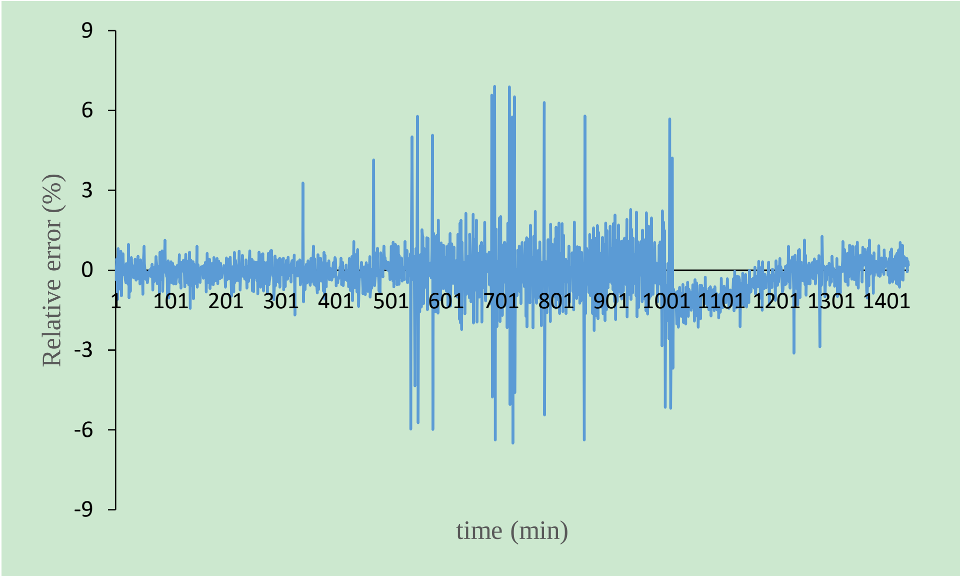
7	0	0
8	0	0
9	0	0

3.3. Verification of the influent COD load forecasting model

As mentioned before, the influent COD load is equal to the product of sewage mass inflow and influent COD. Therefore, the forecasting model of influent COD load can be obtained by forecasting the sewage inflow and influent COD. In order to test the forecasting performance, the real-time data of the influent COD load for 24 hours in another period of this municipal WWTP is used for verification. The comparison between the forecasting results and real-time measured data is shown in Figure 7.



(a) Forecasting results for influent COD load



(b) Relative error

Figure 7. Forecasting result

From Figure 7, no matter the influent COD load is in the stable period or in the large fluctuation period, the relative errors of forecasting results are within [-7%, 7%]. For more than 95% of the relative errors of the forecasting results are within [-5%, 5%], which is much less than industrial acceptable standard [-5%, 5%] for process control. The proposed influent COD load forecasting model has good reliability.

In order to objectively verify the feasibility of the model, the evaluation indicators are calculated as shown in Table 7. For the evaluation indicators: R^2 of the forecasting results is as high as 0.94, which shows high fitness between the forecasting results and the measured data. The mean absolute percentage error (MAPE) is 1.08%, which is far less than the judgment standard (the accuracy of the model is high if MAPE$<$10). The value of Theil inequality coefficient (TIC) is also close to zero. These evaluation indicators reveal that the influent COD load forecasting model for municipal WWTP proposed in this paper is reliable and has high accuracy.

Table 7. Evaluation indicator

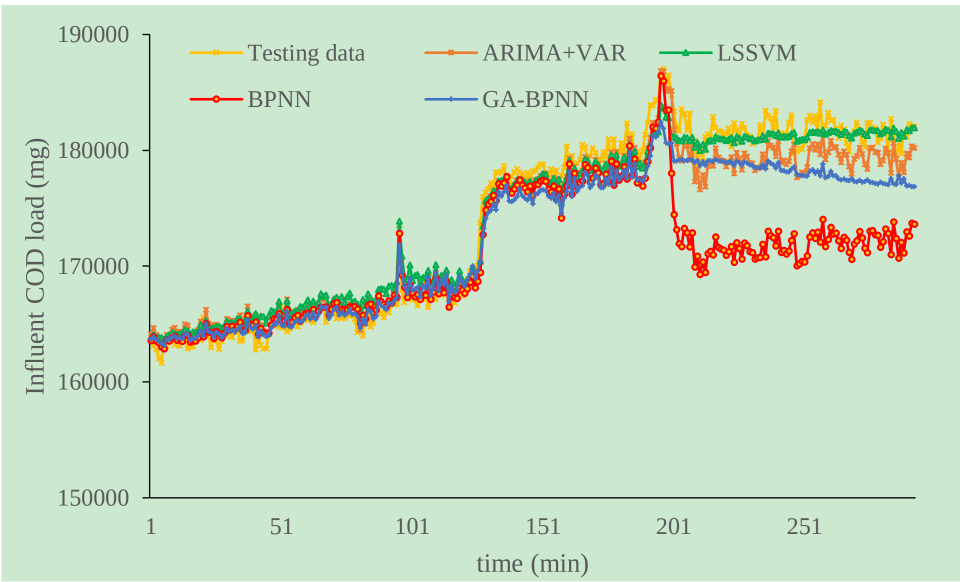
Evaluation index	R^2	MAPE (%)	TIC
COD load	0.94	0.68	0.00003

3.4 Comparison and discussion for the forecasting performance of different models

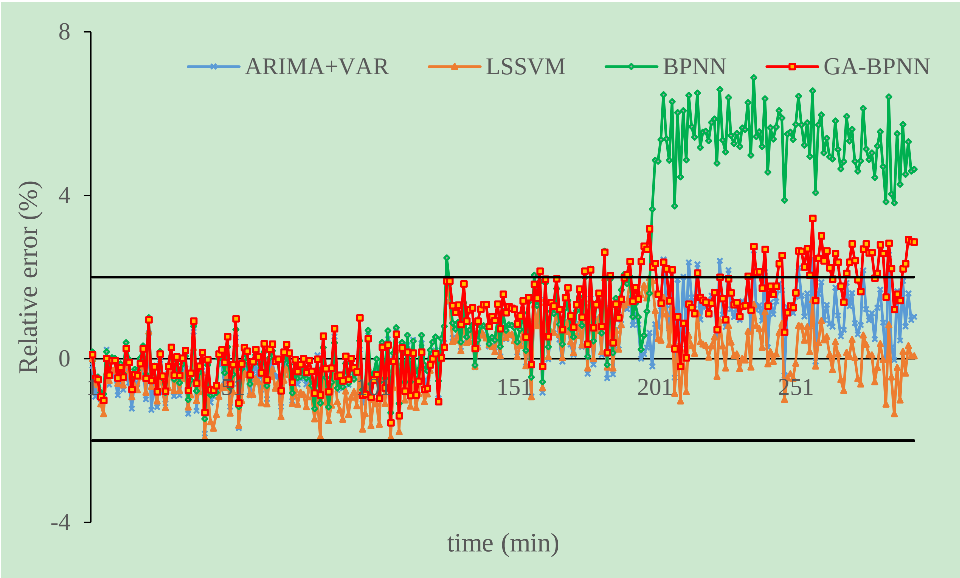
The comparative analysis of the forecasting performances of the ARMA+VAR algorithms based model, BPNN algorithm based model, LSSVM algorithm based model, and hybrid GA-BPNN algorithm based model are presented in this section. All the four forecasting models are developed under the same study case. In order to show the forecasting results of four models more clearly, this paper selected another 7.5 hours of data in the WWTP. The forecasting results as shown in Figure 8..

The forecasting performance of ARMA+VAR, BPNN, LSSVM, and GA-BPNN of the study case is shown in Fig. 8 (a) and the relative error is shown in Fig. 8 (b). Setting a benchmark of [-2%, 2%] in forecasting error makes it easy to find out the best consistent performer among the employed forecasting models. The discretized time points where the forecasting error lies within this benchmark are specifically shown in Fig. 8 (b). Considering the mentioned benchmark in forecasting error, the proposed ARMA+VAR model provides permissible forecasting error with 385 time points. On the contrary, for the BPNN model, the number of time points which lie within the forecasting error [-2%, 2%] is 221 for the BPNN model, 293 for the LSSVM model, and 248 for the hybrid GA-BPNN model. This reveals that the proposed ARMA+VAR model has the best consistent performance among all the employed models.

The mean absolute percent error (MAPE) and root mean square error (RMSE) of the forecasting performance for the BPNN, the LSSVM, the hybrid GA-BPNN, and proposed ARMA+VAR models are shown in Table 8. The MAPE of the ARMA+VAR model is 2 times less than that of BPNN. The MAPE of the ARMA+VAR model is reduced by 89.8% when compared with the hybrid GA-BPNN model and by 42.6% when compared with the LSSVM model. The verification results using industrial data show that the proposed ARMA+VAR model achieves the highest accuracy than the compared three models.



(a) Forecasting result



(b) Relative error

Fig.8. Forecasting results comparison of the four models

Table 8. The forecasting performance analysis

ARMA+VAR		BPNN		LSSVM		GA-BPNN	
MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
1.08	113.56	2.65	303.51	1.54	182.6	2.05	232.6

4. Conclusion

This paper proposed an influent COD load forecasting model based on hybrid artificial intelligence algorithms for municipal wastewater treatment plants. The real-time data are used for modeling and model verification. The influent COD load forecasting model consists of two parts, the sewage inflow forecasting model based on ARMA algorithm, and the influent COD forecasting model based on VAR algorithm. The forecasting model is established based on the historical data of sewage inflow and the correlation analysis of some key variables (include pH, NH_3-N, and temperature). The forecasting model is the basis of the feedforward-feedback control system for the aeration process in WWTP.

The proposed influent COD load forecasting model shows good reliability and high accuracy. The relative errors of the forecasting results are within [-7%, 7%], which meets the industrial acceptable standard for process control. Compared with three employed contrast forecasting models (BPNN, LSSVM, and hybrid GA-BPNN), the forecasting performance shows that the proposed ARMA+VAR model has the highest accuracy. It reveals that the ARMA+VAR model is superior to the other three forecasting models for future application in the papermaking process since its MAPE is only 1.08%. The forecasting model supplies the basis for the feedforward-feedback control system of the aeration process in WWTP and makes it possible for precise control for energy conservation for municipal WWTP.

Acknowledgments

This work is supported by the Fund of State Key Laboratory of Pulp and Paper Engineering (No. 201830), the Nature Science Funds of Guangdong Province (No.2017A030310562), and the Fundamental Research Funds for the Central Universities (No. 2017BQ023).

Appendix

Granger causality test

Granger (1969) put forward the concept of causality, which is easy to deal with VAR algorithm based forecasting model. Granger causality test can be used to analyze the relationship between two time series variables. In general, for the variables Y and X, Granger causality requires the estimation (Farokhzadi et al, 2018):

$$Y_t = \sum_{i=1}^{m} a_i X_{t-i} + \sum_{i=1}^{m} \beta_i Y_{t-i} + u_{1t} \quad \text{(A.1)}$$

$$X_t = \sum_{i=1}^{m} \lambda_i Y_{t-i} + \sum_{i=1}^{m} \delta_i X_{t-i} + u_{2t} \quad \text{(A.2)}$$

Where: m is the number of time-lag term X, namely the number of parameters to be estimated in the constrained regression equation; t is the time in min; a_i and λ_i are parameter coefficients; u_{1t} and u_{2t} are irrelevant white noises.

The Granger causality test is completed by a constrained F test, as shown in Eq. (A.3) (Farokhzadi et al, 2018):

$$F = {(RSS_R - RSS_U)/m} \Big/ {RSS_U/(n-k)} \quad \text{(A.3)}$$

Where, RSS_R is the sum of residual errors obtained by a constrained regression that does not contain X time-lag terms; RSS_U is the sum of residual errors of unconstrained regression that contains X time-lag terms, and n is the sample size; k is the number of parameters to be estimated in the unconstrained regression.

If $F > F_\alpha(m, n-k)$, the null hypothesis is rejected, and X is considered to be the Granger cause of Y (Meng & Han, 2018).

References

Babu, C.N., Reddy, B.E., 2014. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. Appl. Soft Comput. J. 23, 27–38.

Barak, S., Sadegh, S.S., 2016. Forecasting energy consumption using ensemble ARIMA-ANFIS hybrid algorithm. Int. J. Electr. Power Energy Syst. 82, 92–104.

Chan J., Eisenstat E., 2018. Comparing hybrid time-varying parameter VARs. Ecol. Lett. 171, 1-5.

Chen, H.W., Yu, R.F., Ning, S.K., Huang, H.C., 2010. Forecasting effluent quality of an industry wastewater treatment plant by evolutionary grey dynamic model. Resour. Conserv. Recycl. 54, 235–241.

Deng, W., Wang, G., 2017. A novel water quality data analysis framework based on time-series data mining. J. Environ. Manage. 196, 365–375.

Deng, W., Wang, G., Zhang, X., 2015. A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting. Chemom. Intell. Lab. Syst. 149, 39–49.

Farokhzadi, M., Hossein-Zadeh, G.A., Soltanian-Zadeh, H., 2018. Nonlinear effective connectivity measure based on adaptive Neuro Fuzzy Inference System and Granger Causality. Neuroimage. 181, 382-394.

Gebler, D., Wiegleb, G., Szoszkiewicz, K., 2018. Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks. Water Res. 139, 395–405.

Li, W., Li, L., Qiu, G., 2017. Energy consumption and economic cost of typical wastewater treatment systems in Shenzhen, China. J. Clean. Prod. 163, 374–378.

López-Morales, C.A., Rodríguez-Tapia, L., 2019. On the economic analysis of wastewater treatment and reuse for designing strategies for water sustainability: Lessons from the Mexico Valley Basin. Resour. Conserv. Recycl. 140, 1–12.

Man, Y., Shen, W., Chen, X., Long, Z., Pons, M.N., 2017. Modeling and simulation of the industrial sequencing batch reactor wastewater treatment process for cleaner production in pulp and paper mills. J. Clean. Prod. 167, 643–652.

Man, Y., Shen, W., Chen, X., Long, Z., Corriou, J.P., 2018. Dissolved oxygen control strategies for the industrial sequencing batch reactor of the wastewater treatment process in the papermaking industry. Environ. Sci. Water Res. Technol. 4, 654–662.

Meng, X., Han, J., 2018. Roads, economy, population density, and CO_2: A city-scaled causality analysis. Resour. Conserv. Recycl. 128, 508–515.

Park, Y.S., Engel, B.A., 2015. Analysis for Regression Model Behavior by Sampling Strategy for Annual Pollutant Load Estimation. J. Environ. Qual. 44, 1843-1851.

Patilea, V., Raïssi, H., 2013. Corrected portmanteau tests for VAR models with time-varying variance. J. Multivar. Anal. 116, 190-207

Sen, P., Roy, M., Pal, P., 2016. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. Energy 116, 1031–1038.

Son, H., Kim, C., 2017. Short-term forecasting of electricity demand for the residential sector using weather and social variables. Resour. Conserv. Recycl. 123, 200–207.

Suchetana, B., Rajagopalan, B., Silverstein, J.A., 2019. Investigating regime shifts and the factors controlling Total Inorganic Nitrogen concentrations in treated wastewater using non-homogeneous Hidden Markov and multinomial logistic regression models. Sci. Total Environ. 646, 625–633.

Vrečko, D., Hvala, N., Stražar, M., 2011. The application of model predictive control of ammonia nitrogen in an activated sludge process. Water Sci. Technol. 64(5), 1115-1121.

Wang, J., Li, L., Li, F., Kharrazi, A., Bai, Y., 2018. Regional footprints and interregional interactions of chemical oxygen demand discharges in China. Resour. Conserv. Recycl. 132, 386–397.

Wang Q., Song X., Li R., 2018. A novel hybridization of nonlinear grey model and linear ARIMA residual correction for forecasting U.S. shale oil production. Energy. 165, 1320-1331.

Yang, T., Long, R., Cui, X., Zhu, D., Chen, H., 2017. Application of the public–private partnership model to urban sewage treatment. J. Clean. Prod. 142, 1065–1074.

Ye, H., Ren, Q., Hu, X., Lin, T., Shi, L., Zhang, G., Li, X., 2018. Modeling energy-related CO_2 emissions from office buildings using general regression neural network. Resour. Conserv. Recycl. 129, 168–174.

Yuan, C., Liu, S., Fang, Z., 2016. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. Energy 100, 384–390.