




Dissolved oxygen content interval prediction based on auto regression recurrent neural network

Jiande Huang^{1,5,6} · Yue Huang¹ · Shahbaz Gul Hassan^{1,2,3,4,5,6} · Longqin Xu^{1,2,3,4,5,6} · Shuangyin Liu^{1,2,3,4,5,6} 

Received: 18 August 2021 / Accepted: 25 October 2021 / Published online: 5 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Dissolved oxygen content prediction plays an important role in the intelligence management of aquaculture systems. Compared with traditional point prediction, interval prediction can quantify the uncertainties effectively and more closes to the fact. Unfortunately, there is rarely study about it. In this paper, a novel interval prediction method based on a deep auto-regression recurrent neural network (DeepAR) is proposed to construct prediction intervals (PIs) directly. Besides, a variational mode decomposition (VMD) has been used to extract the frequency feature of the original data. Moreover, a multi-objective weighted optimization framework based on the sparrow swarm algorithm (SSA) was proposed to improve PIs accuracy. Finally, simulations with water quality datasets were conducted to show the effectiveness of the proposed model. The results demonstrated that the proposed model significantly improved PI quality and performance compared to the state-of-the-art method.

Keywords Water quality · Dissolved oxygen · Interval prediction · DeepAR · Multi-objective optimization

1 Introduction

The level of dissolved oxygen (DO) is one of the primary indicators that can reflect the degree of water pollution. It is also one of the vital measures for ensuring the health and development of aquaculture fish (Hu et al. 2015; Rahman et al. 2020; Xiao et al. 2019). A sufficient amount of information on the state of water quality is essential for effective management. Thus, appropriate modeling is required to forecast future events. As a result, while modeling, the uncertainty in water quality data should be carefully evaluated to make more reliable decisions.

Recently, artificial neural network (ANN) based models have been successfully applied in the point prediction of water quality. These models can obtain correlations between input and output without comprehending their basic principles. Wu used a backpropagation neural network based on particle swarm optimization algorithm (PSO) to forecast the DO of regional groundwater in Xilin Gol League (Wu et al. 2018). Li et al., (2021) compared the performance of the recurrent neural network (RNN), long-short time memory neural network (LSTM) and gated recurrent unit (GRU), determined that the GRU had higher performance and was more suited for dissolved oxygen prediction. Zhang et al., (2020) present a model to anticipate dissolved oxygen trends

✉ Shuangyin Liu
shuangyinliu@zhku.edu.cn

Shahbaz Gul Hassan
mhasan387@zkhu.edu.cn

¹ College of Information Science and Technology,
Zhongkai University of Agriculture and Engineering,
Guangzhou 510225, China

² Smart Agriculture Engineering Technology Research
Center of Guangdong Higher Education Institutes,
Zhongkai University of Agriculture and Engineering,
Guangzhou 510225, China

³ Guangzhou Key Laboratory of Agricultural Product
Quality, Safety Traceability Information Technology,
Zhongkai University of Agriculture and Engineering,
Guangzhou 510225, China

⁴ Guangdong Provincial Agricultural Products Safety
Big Data Engineering Technology Research Center,
Zhongkai University of Agriculture and Engineering,
Guangzhou 510225, China

⁵ Guangdong Province Key Laboratory of Waterfowl Healthy
Breeding, Guangzhou 510225, China

⁶ Academy of Smart Agricultural Engineering Innovations,
Zhongkai University of Agriculture and Engineering,
Guangzhou 510225, China

based on a combination of kernel principal component analysis (KPCA) and Recurrent Neural Network (RNN).

In most cases, point prediction of ANN-based models would be considered firstly by lots of studies. However, they may suffer from low reliability when the uncertainty level presents in the DO time series. Because the point prediction cannot capture the modeling uncertainty information, it could not explain the forecast accuracy (Huang et al. 2017). On the contrary, modeling for uncertainty information may be used to determine prediction error and the likelihood of correct forecasts, making decision-making easier.

Interval prediction (PI) is a powerful tool for estimation uncertainty to inform managers (Quan et al. 2015). As the confidence level has measured, PI offered a limit that captured the observed values. $((1-a) \%)$ (Chatfield 1993; Ma et al. 2020; Voyant et al. 2020). Some methods to compute the PIs include the delta method, Bayesian method, bootstrap and low upper bound estimation (LUBE). The delta approach involved using Taylor's series in an ANN non-linear regression model (Chryssolouris et al. 1996; Momotaz and Dohi 2016). This technique hypothesizes that the noise of original data is homogeneity and normal distribution, which is inconsistent with reality and the method questionable. In the Bayesian approach that ANN's parameters have their own probability distribution instead of a single value. Therefore the output of the Bayesian neural network will likewise have distributions depending on the observed training set (MacKay 1992). The Bayesian method needs great calculation force to computer Hessian matrix to construct PIs. Moreover, the accuracy of the model largely depends on prior knowledge. PIs are calculated using the bootstrap method, an easy and widely used way to calculate them (Efron and Tibshirani 1993; Lu et al. 2020). As a resample method, the bootstrap method needs to train several different ANNs. This method is easy to realize and does not need any assumption about the distribution of data. The only disadvantage of the method is its high computational cost for large datasets, and the whole performance over depends on a single ANN. The LUBE method directly outputs the lower and upper bound of PIs via an ANN (Lian et al. 2020). This technique includes a heuristic search algorithm to construct the architecture and parameters of the ANN due to LUBE cannot be expressed as a supervised learning problem, which consumes great computational force. In conclusion, each method has own advantage and disadvantage, how to select the suitable technique to quantify the uncertainty of dissolve oxygen need to be further researched.

DeepAR, proposed recently by Salinas et al., (2020), is an autoregressive recurrent neural network with probabilistic forecasting ability. DeepAR applies RNN and autoregression technique to predicts scalar time series, which learns such a model from historical data. Compared with the methods mentioned above, DeepAR does not assume Gaussian

noise but can incorporate many likelihood functions. As a supervised learning model, DeepAR can effectively train its parameters using the backpropagation algorithm and directly gives the lower and upper bound of PIs. Moreover, this method works with litter hyperparameter turning and is applied to small-size datasets. In recent years, DeepAR has achieved success in many research fields. For example (Dong et al. 2021), established a model based on DeepAR techniques for deformation trend prediction. Park et al., (2020) photovoltaic generation data captured at Hadong, Korea, to investigate the probabilistic prediction schemes of day-ahead photovoltaic generations with DeepAR. The simulation results show that DeepAR is helpful for efficient grid management. Although DeepAR has a solid ability to account for uncertainties, to the best of our knowledge, DeepAR has been rarely used in interval prediction for water quality in the existing literature.

In this paper, DeepAR is applied for water quality interval prediction. The data on water quality, on the other hand, was often subject to various sources of uncertainty, including measurement mistakes and random occurrences (Nourani et al. 2021). Extraction of crucial feature information, obtaining high-quality water quality data sets. Further, the Variational Mode Decomposition (VMD) technique and sparrow search algorithm (SSA) improved the model's performance. A novel water quality interval prediction framework VMD-DeepAR-SSA was proposed. VMD can deconstruct a complicated original time series signal into a set of band-limited intrinsic mode functions (IMFs) to better analyze their actual distinctive application (Dragomiretskiy and Zosso 2014). SSA is an efficient multi-objective swarm intelligence optimization algorithm with high-performance global search capability, better stability and convergence accuracy, fewer setting parameters and easy implementation (Xue and Shen 2020). Several independent DeepAR models were structured to fit the IMFs obtained by VMD, then the value of the prediction interval is weighted, and SSA searches the best combination of the weights to obtain the ultimate prediction interval.

2 Methodology

2.1 Variational mode decomposition

VMD is a non-recursion and self-adaptive multiresolution variational mode decomposition method with advantages such as high decomposition accuracy, few decomposition layers, and control mode mixing. VMD decomposes a signal to several modes with limited bandwidth and estimates the central frequency of corresponding mode compositions (Dragomiretskiy and Zosso 2014).

VMD process main includes variational model construction and variation model solution two parts. The variational model is to search k finite bandwidth mode function and minimize the sum of the estimated bandwidth of each mode function under the constraint. This constrained optimization problem can be expressed as:

$$\min_{\{u_k, w_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] c^{-jw_k t} \right\|_2^2 \right\}, \text{ s.t. } \sum_k u_k = f \tag{1}$$

where f is the original input signal, $u_k = (u_1, u_2, \dots, u_k)$ is mode function, $w_k = (w_1, w_2, \dots, w_k)$ is central frequency one-side spectrum of each mode function calculated by Hilbert transform. To solution the constraint problem, introducing quadratic penalty parameter α and augmented Lagrangian function:

$$L(u_k, w_k, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \tag{2}$$

The saddle point of the above-mentioned is obtained using an alternating direction multiplier algorithm, which is the optimal solution, to realize the adaptive decomposition of complex modulated signals and obtain K mode functions.

2.2 DeepAR model

DeepAR is a probabilistic forecasting RNN architecture that generates Monte Carlo samples for probabilistic forecasting. This paper researches the application of DeepAR model in DO interval prediction. Denoting z_t as the value of DO time series at time t , our target is to construct the conditional distribution P of future time series length T .

$$P(z_{t_0:T} | z_{1:t_0-1}) = \prod_{t=t_0}^T \ell(z_t | \theta(h_t, \Theta)) \tag{3}$$

$$h_t = h(h_{t-1}, z_{t-1}, \Theta) \tag{4}$$

where ℓ is likelihood function, θ is a parameter of ℓ , A multi-layer recurrent neural network with LSTM cells implements the function h , Θ are parameters of the model.

As show as the Fig. 1, in the training phase: the input to the network at ever time step t includes the preceding time step's target value z_{t-1} and the output of the preceding hidden layer h_{t-1} . The model parameter Θ constructed by RNN parameters $h(\bullet)$ and $\theta(\bullet)$, by maximizing the log-likelihood:

$$L = \sum_{t=1}^{t_0} \ln \ell(z_t | \theta(h_t)) \tag{5}$$

In the prediction phase: a sample $\tilde{z}_t \sim \ell(\bullet | \theta)$ is drawn by mento carlo simulation as the input to next time step until the end of prediction range $t = t_0 + T$ generating one sample trace. Repeating this prediction procedure produces a large number of traces reflecting the jointly predicted distribution. The quantile or expectation can be calculated via those traces.

The architecture of $\theta(h_t)$ only depend on likelihood function $\ell(z | \theta)$, in this paper gaussian likelihood for real-valued data is considered:

$$\ell_g(z | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{6}$$

$$\mu(h_t) = wh_t + b \tag{7}$$

$$\sigma(h_t) = \lg(1 + \exp(wh_t + b)) \tag{8}$$

where w is denoted as the weight matrix, b is denoted as the bias matrix, σ and μ are standard deviation and mean value respectively.

2.3 Sparrow search algorithm

The primary purpose of the sparrow search algorithm is to simulate the foraging process of a sparrow swarm. The SSA process is one of the producer-scronger models with an early warning mechanism of investigation. Producers

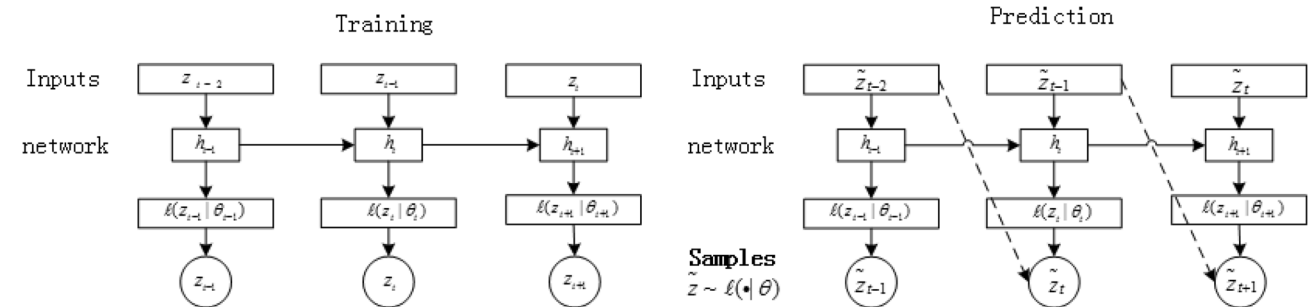


Fig. 1 Summary of DeepAR model training and prediction process

generally have enormous energy reserves, whereas scroungers are labeled as sparrows. In the meanwhile, certain sparrows are chosen to provide early notice of an inquiry. In the process of SSA, each sparrow owns a position stand for the position of food discovered by itself. There are three possible actions for a sparrow for a sparrow: as a producer to search food, like a scrounger to follow the producer and as an investigator to make an early warning. The position of each sparrow in a D -dimensional solution space with N sparrows is:

$$X_i = (x_{i,1}, x_{i,2}, x_{i,3}, \dots, x_{i,D}), i = 1, 2, \dots, N \quad (9)$$

And he fitness of the position:

$$f_i = f(X_i), i = 1, 2, \dots, N \quad (10)$$

In each generation, fitness top 20% sparrows are denoted as producers, the residue 80% sparrows are denoted as scroungers. The position of producer can be updated as follow:

$$x_{i,d}^{t+1} = \begin{cases} x_{i,d}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right), R_2 < ST \\ x_{i,d}^t + Q, R_2 \geq ST \end{cases} \quad (11)$$

where $x_{i,d}^{t+1}$ is the d th dimensional position of the i th sparrow in the t th generation, $\alpha \in (0, 1]$ is a random number between, Q is random number which obeys normal distribution. R_2 ($R_2 \in [0, 1]$) and ST ($ST \in [0.5, 1.0]$) are alarm value and the safety threshold respectively.

Following the producers, scroungers' position can be written as follow:

$$x_{i,d}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{xw_{i,d}^t - x_{i,d}^t}{i^2}\right), i > n/2 \\ xb_{i,d}^t + \frac{1}{D} \sum_{d=1}^D (rand[-1, 1] \cdot (xb_{i,d}^t - x_{i,d}^t)), i \leq n/2 \end{cases} \quad (12)$$

where xw is the worst position of sparrows and xb is the best one in t th generation. Some of sparrows will make early warning as well as foraging, they will abandon current food and remove to a new position when sense the danger, which can be described as follow:

$$x_{i,d}^{t+1} = \begin{cases} xb_{i,d}^t + \beta \cdot (x_{i,d}^t - xb_{i,d}^t), f_i \neq f_g \\ x_{i,d}^t + K \cdot \left(\frac{x_{i,d}^t - xw_{i,d}^t}{|f_i - f_w| + \varepsilon}\right), f_i = f_g \end{cases} \quad (13)$$

where β represents a random number for the normal distribution, $K \in [-1, 1]$ is a Uniform random number, ε is a small number that prevents the denominator from being unique.

3 The hybrid dissolve oxygen interval prediction model

As shown in Fig. 2, this paper combines VMD, DeepAR and SSA techniques to predict interval efficiency. For the proposed model, DO content data firstly be decomposed into several IMFs with different central frequencies by VMD. Then, each IMF is used as input of the DeepAR to construct the conditional distribution respectively. Finally, weighting combination the outputs of those DeepAR models to construct prediction intervals. Significantly, the combination of the optimal weight is searched by SSA by minimizing coverage width-based criterion (CWC). CWC can be calculated by prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW) as follows:

$$PICP = \frac{1}{N} \sum_{i=1}^N a_i \quad (14)$$

where N is the number of test data samples, and a_i is a binary value which can be calculated as follow:

$$a_i = \begin{cases} 0, y_i \notin [y_{i,l}, y_{i,u}] \\ 1, y_i \in [y_{i,l}, y_{i,u}] \end{cases} \quad (15)$$

where y_i is the value of test sample i , $y_{i,l}$ and $y_{i,u}$ are the estimated lower and upper bounds.

$$PINAW = \frac{1}{N * W} \sum_{i=1}^N (y_{i,u} - y_{i,l}) \quad (16)$$

where W is the range of PIs width.

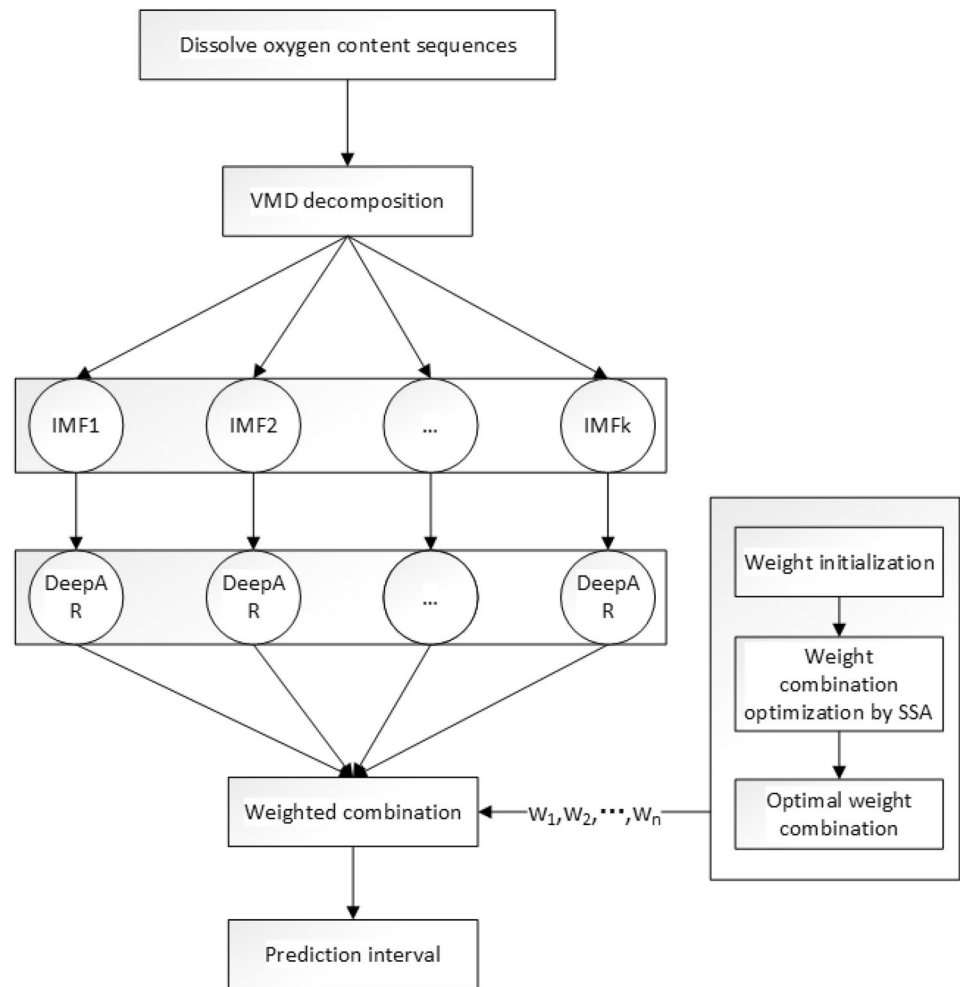
$$CWC = \beta * PINAW(1 + \gamma(PICP)e^{-\eta(PICP-\mu)}) \quad (17)$$

where γ is a binary value determined as:

$$\gamma = \begin{cases} 0, PICP \geq \mu \\ 1, PICP < \mu \end{cases} \quad (18)$$

In Eq. 18, γ and μ are two control parameters for CWC. μ reflects the coverage probability requirement of PIs, which can be calculated by the predefined confidence level $(1-\alpha)$. η and β is the penalty coefficient when the obtained PI dissatisfy the coverage probability requirement. In this paper, to gather high-quality PIs, penalty parameters (β , η , μ) of CWC are set to 1, 1 and 0.95 respectively.

Fig. 2 The implementation process of the proposed model



4 Results and discussions

4.1 Simulation data

Simulation DO content data was collected from river crab culture farms in Yixing city, Jiangsu Province, China. Thirty daily data sets with 10 min intervals were chosen as experiment data sets from the summer of 2019. Each data set includes 144 DO content points, and the first twenty-one days of data were as training data while the other nine data of data were used as testing data.

4.2 VMD processing of dissolved oxygen content data.
 To extract the characteristic hidden in the water quality data, the DO content data are decomposed by VMD into 5 IMFs with different central frequencies. The decomposition results based on VMD are shown in Fig. 3.

- (a) Original data
- (b) IMF1
- (c) IMF2
- (d) IMF3

- (e) IMF4
- (f) IMF5

4.2 Simulation results and discussion

After the data processing by VMD, obtaining five IMFs. Then, each IMF is trained by DeepAR. According to the advice from (Salinas et al. 2020), to facilitate learning time-dependent patterns, such as peak valley during noon, DeepAR automatically creates feature time series based on the frequency of the DO content time series. So, this paper provides the detailed custom feature time series of DO content as far as possible. The following table (Table 1) lists the custom time frequencies features.

Five DeepAR models have been structured to learn the conditional distribution of the predicted target of five IMFs. The upper and lower bounds of those models are sum simplify to establish the initial prediction intervals. As shown in Fig. 4, the initial PIs results present a good performance with high CICP (CICP=0.9007), which means PIs can cover over 90% of the observed value. However, the results have

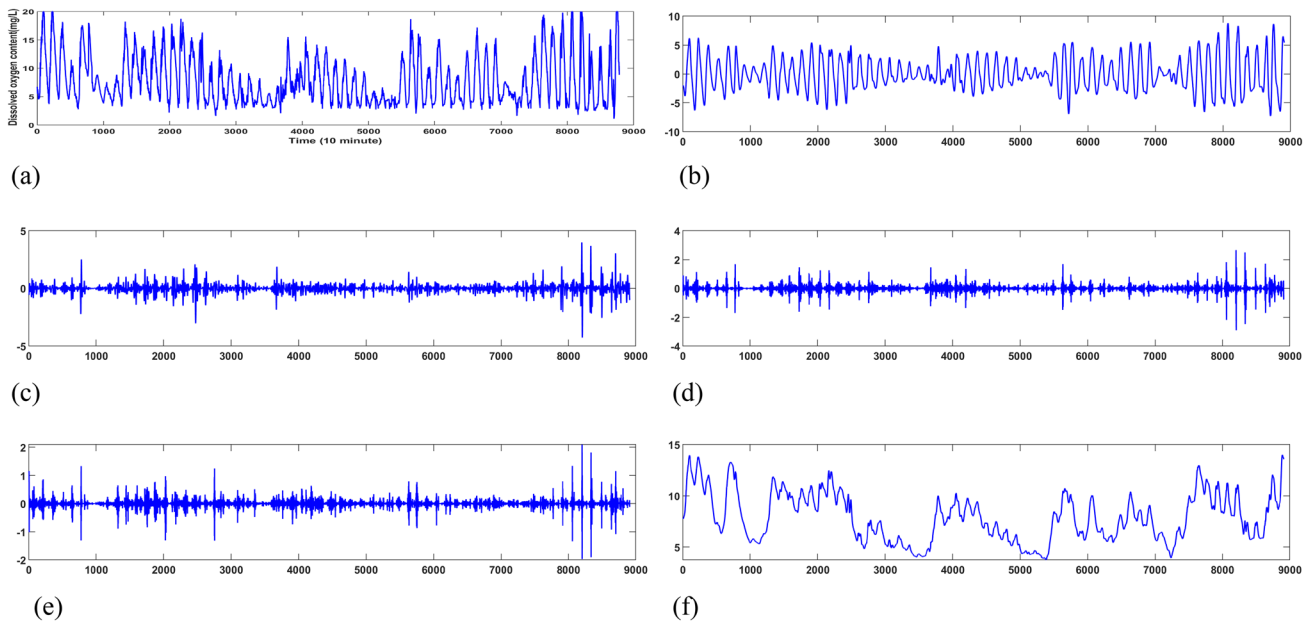


Fig. 3 The IMFs from high frequency to low frequency using the VMD method

Table 1 The custom time frequencies feature of DO content time series

Frequency of the Time Series	Derived Features
Minute	Minute-of-hour
Hour	Hour-of-day
Day	Day-of-week, day-of-month

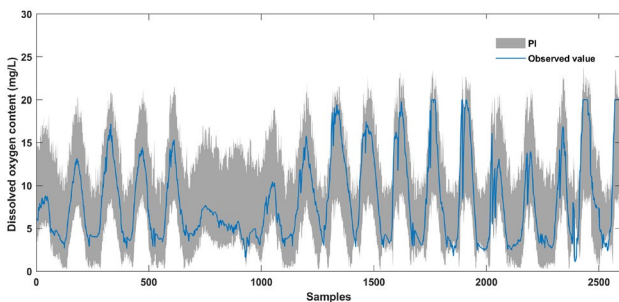


Fig. 4 The initial PIs of test sets

a high PINRW (PINRW = 0.1574) value and a cover rate of the lower bound (0.9165). In particular, information on the narrow interval width is valuable more than the wide one and the cover rate of the lower bound of the PIs should be more considered in DO interval prediction. To further improve the performance of PIs, the weighting method is used to reconstruct the PIs and SSA searches for the optimal weight combination..

For the SSA process, population size $N=80$, the producers account for 20% of the population size, the early warner account for 20% of the population size, and the number of iterations is 40. The trend of fitness value is shown in Fig. 5, which decreases with the number of iterations and finally be stable. Figure 5 shows the changing trend of weights in the optimization process, which to stable after 15 iterations. This shows that the SSA can quickly converge to the optimal solution. Meanwhile, the weight of lower frequency IMFs is seen to be bigger than the higher one, which means that the lower frequency IMFs contain more time-series features and represent the original DO content data instead of the higher frequency IMFs. After optimization and weighting reconstruction, the PIs result is shown in Fig. 6; compared with the initial PIs result, the proposed method improves not only the PICP (PICP = 0.9503) but also the critical index cover rate of the lower bound (0.9749) and decreases the PINRW (0.1324) effective.

Changing trend of fitness function.

Changing trend of IMF1's weight.

Changing trend of IMF2's weight.

Changing trend of IMF3's weight.

Changing trend of IMF4's weight.

Changing trend of IMF5's weight.

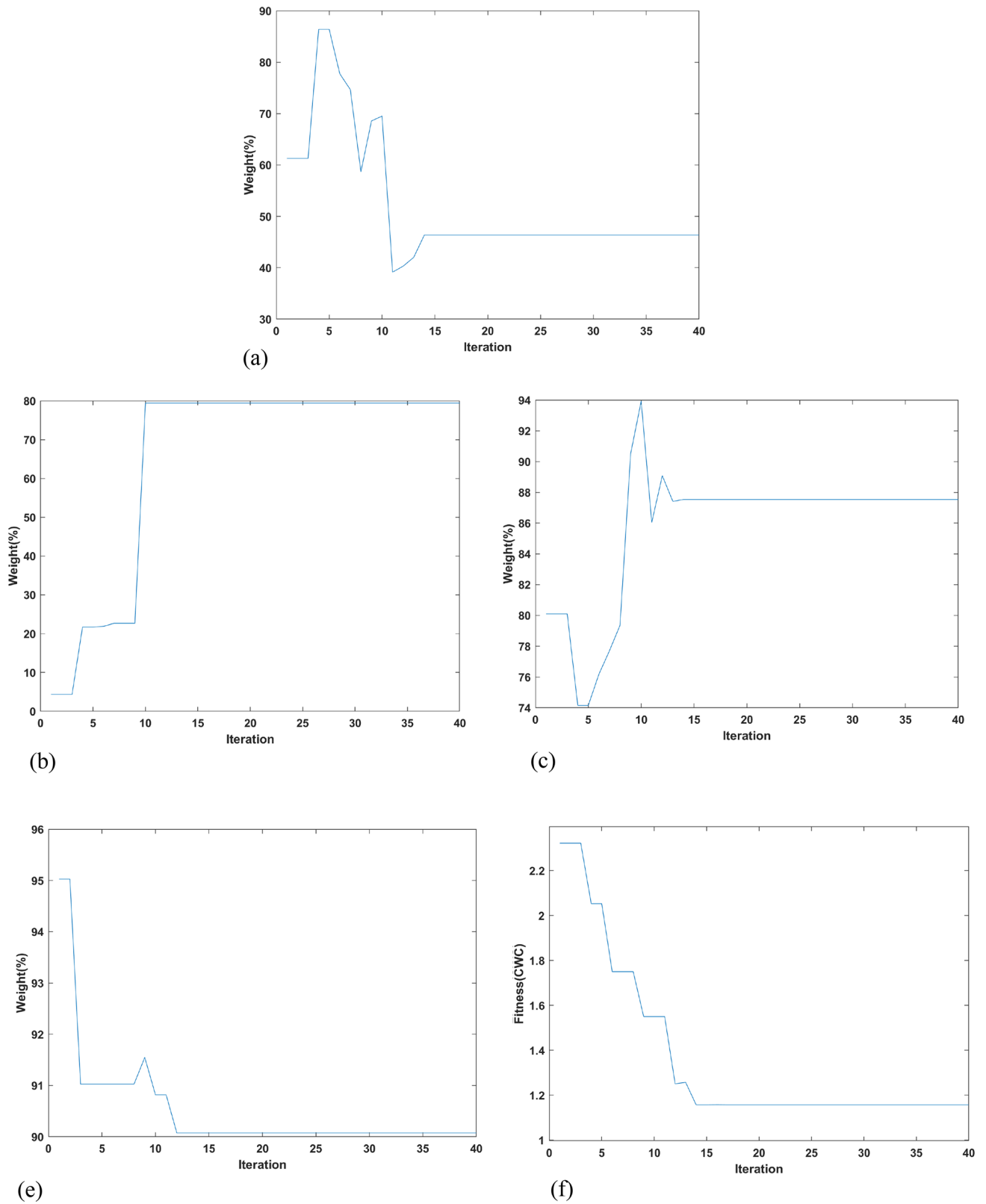


Fig. 5 The graph of SSA process

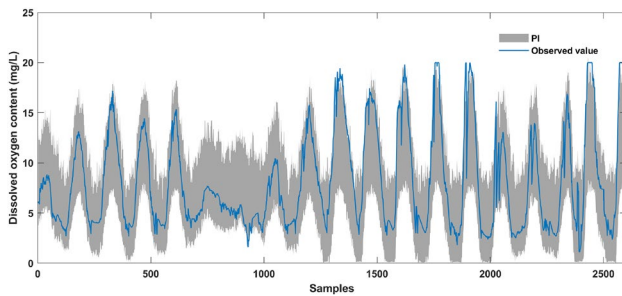


Fig. 6 The PIs of test sets by the proposed model

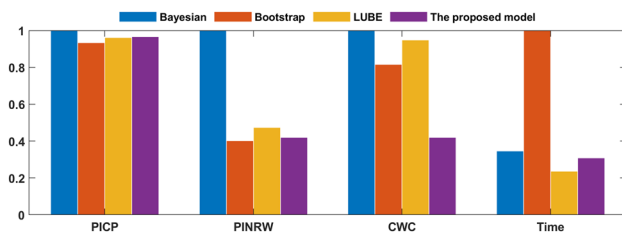


Fig. 7 Results of normalization index

Table 2 Comparison of different methods with various indexes

Model	PICP	PINRW	CWC	Run time (s)
Bayesian	0.9836	0.3750	0.3750	36.7096
Bootstrap	0.9192	0.1505	0.3057	106.0359
LUBE	0.9459	0.1776	0.3559	25.0401
The proposed model	0.9503	0.1574	0.1574	32.7923

4.3 Comparison with typical models

To verify the prediction effect of the proposed model, this section does the comparative analysis between the proposed model and the typical model used in interval prediction includes Bayesian, Bootstrap and LUBE. PICP, PINRW and CWC are the evaluating indicators, CWC is a comprehensive index for PICP and PINRW. To ensure the rigor of the paper, the same DO content training set and test set were used for experiments.

As shown in Table 2, it is evident that compared with Bootstrap, LUBE, and the proposed models, the PICP of the Bayesian model respectively 7.006%, 3.996%, and 3.504% beyond. However, the Bayesian method has the highest PICP but causes wider PIs and bigger PINRW. This may be caused by overestimating the variance of the outputs, which leads to a lower actual worth of PIs information. Moreover, the proposed model, LUBE and Bootstrap method have a close PINRW value, But the Bootstrap method has the lowest PICP value.

Considered that the PICP and PINRW are contradictory to each other indexes and the situation that models have similar PICP and PINRW values, the comprehensive index CWC is used to further assess the model's performance. As shown in Table 2 and Fig. 7, the Bayesian method has the highest PICP and the Bootstrap method has the lowest PINRW, but the bad performance of their other index leads to a low CWC value. The LUBE and the proposed model have close PICP and PINRW, but the CWC value of LUBE is much less than the proposed model because punishment factors of CWC keep the balance between PICP and PINRW. Although the proposed model is not the best in PICP and PINRW, the proposed model still be the best method for DO content interval prediction through evaluating the comprehensive index.

Various methods of PIs are established with different levels of complexity, computation burden, reliability and required time. One cannot claim that a certain method is senior to the others, but each method has its advantages and disadvantages when constructing the PIs of DO content. Therefore, a brief comparison of each technique is showed in Table 3 based on the optimal value of each index from Table 2 and Fig. 7. The proposed method is the most superior method in this case. The Bootstrap method is a point prediction model with Monte Carlo simulation, which often cannot accurately reflect the output concept of the probability model. Moreover, the Bootstrap method requires several different neural networks and largely depends on a single neural network's performance, making it a large computational burden and less reliable. The Bayesian method needs a high calculation power and the scale of the training dataset to improve its accuracy. Putting aside those requirements, the Bayesian method may be acceptable in this case.

Table 3 The brief comparison of different models

Model	Training time	Number of required networks	Computational burden	Reliability
Bayesian	Medium	One	High	High
Bootstrap	High	More than one	High	Medium
LUBE	High	One	Low	High
The proposed model	Medium	One	Low	High

Compared with other methods, LUBE is the most straightforward method to construct the PIs directly and requires a huge power to adjust the construction and parameters, which is a challenge to equipment.

5 Conclusions

Forecasting DO content is essential for aquaculture water quality management because it offers decisions for water quality hybrid systems. This paper first propose adding an interval prediction module for hybrid water quality management systems to improve water quality management ability. In addition, we conducted a detailed and comprehensive analysis, comparison, and summary of prior work about interval prediction, and put forward a novel PIs model developed for dissolved oxygen content prediction. The variance mode decomposition technique is used to extract periodic characteristics of the original water quality data. The SSA is used to calculate the weights of various DeepAR outputs. The results demonstrate that the proposed method has better forecasting performance than typical Bayesian, Bootstrap, and LUBE methods as measured by PICP, PINRW and CWC. Furthermore, after weighing the benefits and drawbacks of various approaches, it is concluded that the suggested model adequately captures the uncertainty of DO content time series and is a reliable tool for the development of water quality PIs. Prediction is the most popular task of time series analysis Because of its importance in industrial, social, and scientific applications. The proposed approach is used to solute the PIs problem about time series. It can also be used in other domains such as sales volume forecast for commodity, weather like wind speed forecast, even the stock market forecast.

6 Future directions

Time series prediction is a vital tool for DO content. This paper proposed a novel and complex frame to accurately obtain PIs of DO, which is a tedious trial and error process of manual neural architecture design. With the development of neural architecture search (NAS), artificial intelligence is expected to design a network structure (Ren et al. 2021), and we also hoped that it can design a more effective and simply neural architecture for interval prediction. In the meanwhile, in addition to the past values of DO content, they are many correlated factors such as the position of the farm without knowing in advance how they interact, may influence the trend of DO content. We need to explore more robust features by dimensionality reduction techniques like

linear discriminant analysis (LDA) or a spectral clustering technique (Li et al. 2019, 2018a, b; Yan et al. 2021).

Funding Finally, this work was supported in part by the National Natural Science Foundation of China under Grant 61871475, 61471-131, 61571444, in part by the special project of laboratory construction of Guangzhou Innovation Platform Construction Plan under Grant 201905010006, the Guangzhou Innovation Platform Construction Plan under Grant 2017B0101260016, the Foundation for High-level Talents in Higher Education of Guangdong Province under Grant 2017GCZX00014, 2016KZDXM0013, 2017KTSCX094, 2018LM2168, and the Beijing Natural Science Foundation under Grant 4182023.

Availability of data and material The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

Code availability The code during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Chatfield C (1993) Calculating interval forecasts. *J Bus Econ Stat*. <https://doi.org/10.1080/07350015.1993.10509938>
- Chryssolouris G, Lee M, Ramsey A (1996) Confidence interval prediction for neural network models. *IEEE Trans Neural Netw*. <https://doi.org/10.1109/72.478409>
- Dong M, Wu H, Hu H, Azzam R, Zhang L, Zheng Z, Gong X (2021) Deformation prediction of unstable slopes based on real-time monitoring and deepar model. *Sensors (Switzerl)*. <https://doi.org/10.3390/s21010014>
- Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. *IEEE Trans Signal Process*. <https://doi.org/10.1109/TSP.2013.2288675>
- Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. An Introduction to the Bootstrap. <https://doi.org/10.1007/978-1-4899-4541-9>
- Hu J, Wang J, Zhang X, Fu Z (2015) Research status and development trends of information technologies in aquacultures. *Nongye Jixie Xuebao/Transactions Chinese Soc. Agric Mach* 46:251–263. <https://doi.org/10.6041/j.issn.1000-1298.2015.07.037>
- Huang F, Huang J, Jiang S, Zhou C (2017) Landslide displacement prediction based on multivariate chaotic model and extreme learning machine. *Eng Geol*. <https://doi.org/10.1016/j.enggeo.2017.01.016>

- Li Z, Nie F, Chang X, Nie L, Zhang H, Yang Y (2018a) Rank-constrained spectral clustering with flexible embedding. *IEEE Trans Neural Networks Learn Syst.* <https://doi.org/10.1109/TNNLS.2018.2817538>
- Li Z, Nie F, Chang X, Yang Y, Zhang C, Sebe N (2018b) Dynamic affinity graph construction for spectral clustering using multiple features. *IEEE Trans Neural Netw Learn Syst.* <https://doi.org/10.1109/TNNLS.2018.2829867>
- Li Z, Yao L, Chang X, Zhan K, Sun J, Zhang H (2019) Zero-shot event detection via event-adaptive concept relevance mining. *Pattern Recognit.* <https://doi.org/10.1016/j.patcog.2018.12.010>
- Li W, Wu H, Zhu N, Jiang Y, Tan J, Guo Y (2021) Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Inf Process Agric.* <https://doi.org/10.1016/j.inpa.2020.02.002>
- Lian C, Zeng Z, Wang X, Yao W, Su Y, Tang H (2020) Landslide displacement interval prediction using lower upper bound estimation method with pre-trained random vector functional link network initialization. *Neural Netw.* <https://doi.org/10.1016/j.neunet.2020.07.020>
- Lu J, Ding J, Dai X, Chai T (2020) Ensemble stochastic configuration networks for estimating prediction intervals: a simultaneous robust training algorithm and its application. *IEEE Trans Neural Netw Learn Syst.* <https://doi.org/10.1109/TNNLS.2020.2967816>
- Ma J, Niu X, Tang H, Wang Y, Wen T, Zhang J (2020) Displacement prediction of a complex landslide in the three gorges reservoir area (China) using a hybrid computational intelligence approach. *Complexity.* <https://doi.org/10.1155/2020/2624547>
- MacKay DJC (1992) A practical bayesian framework for backpropagation networks. *Neural Comput.* <https://doi.org/10.1162/neco.1992.4.3.448>
- Momotaz B, Dohi T (2016) Prediction interval of cumulative number of software faults using multilayer perceptron. *Stud Comput Intell.* https://doi.org/10.1007/978-3-319-26396-0_4
- Nourani V, Paknezhad NJ, Tanaka H (2021) Prediction interval estimation methods for artificial neural network (Ann)-based modeling of the hydro-climatic processes, a review. *Sustain.* <https://doi.org/10.3390/su13041633>
- Park, Soyeong, Park, Sunme, Hwang, E., 2020. Normalized residue analysis for deep learning based probabilistic forecasting of photovoltaic generations. In: *Proceedings—2020 IEEE International Conference on Big Data and Smart Computing. BigComp.* <https://doi.org/10.1109/BigComp48618.2020.00-20>
- Quan H, Srinivasan D, Khosravi A (2015) Incorporating wind power forecast uncertainties into stochastic unit commitment using neural network-based prediction intervals. *IEEE Trans Neural Netw Learn Syst.* <https://doi.org/10.1109/TNNLS.2014.2376696>
- Rahman A, Dabrowski J, McCulloch J (2020) Dissolved oxygen prediction in prawn ponds from a group of one step predictors. *Inf Process Agric* 7:307–317. <https://doi.org/10.1016/j.inpa.2019.08.002>
- Ren P, Xiao Y, Chang X, Huang PY, Li Z, Chen X, Wang X (2021) A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Comput Surv.* <https://doi.org/10.1145/3447582>
- Salinas D, Flunkert V, Gasthaus J, Januschowski T (2020) DeepAR: probabilistic forecasting with autoregressive recurrent networks. *Int J Forecast.* <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- Voyant C, Notton G, Duchaud JL, Almorox J, Yaseen ZM (2020) Solar irradiation prediction intervals based on Box-Cox transformation and univariate representation of periodic autoregressive model. *Energy Focus Renew.* <https://doi.org/10.1016/j.ref.2020.04.001>
- Wu J, Li Z, Zhu L, Li G, Niu B, Peng F (2018) Optimized BP neural network for dissolved oxygen prediction. *IFAC-PapersOnLine* 51:596–601. <https://doi.org/10.1016/j.ifacol.2018.08.132>
- Xiao R, Wei Y, An D, Li D, Ta X, Wu Y, Ren Q (2019) A review on the research status and development trend of equipment in water treatment processes of recirculating aquaculture systems. *Rev Aquac.* <https://doi.org/10.1111/raq.12270>
- Xue J, Shen B (2020) A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst Sci Control Eng.* <https://doi.org/10.1080/21642583.2019.1708830>
- Yan C, Chang X, Luo M, Zheng Q, Zhang X, Li Z, Nie F (2021) Self-weighted robust LDA for multiclass classification with edge classes. *ACM Trans Intell Syst Technol.* <https://doi.org/10.1145/3418284>
- Zhang YF, Fitch P, Thorburn PJ (2020) Predicting the trend of dissolved oxygen based on the kPCA-RNN model. *Water (switzerl).* <https://doi.org/10.3390/w12020585>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.