Received February 28, 2020, accepted March 12, 2020, date of publication March 17, 2020, date of current version March 27, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.2981528

Multivariate Seawater Quality Prediction Based on PCA-RVM Supported by Edge Computing Towards Smart Ocean

XIAOCHUAN SUN[1], XIN WANG[1], DI CAI[1], ZHIGANG LI[1], YUANYUAN GAO[1], AND XUSHENG WANG[2]
[1]College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China
[2]College of Science, North China University of Science and Technology, Tangshan 063210, China
Corresponding author: Xusheng Wang (zdhwxs@ncst.edu.cn)

This work was supported in part by the Natural Science Foundation of Hebei Province under Grant F2018209181, in part by the Hebei Colleges and Universities Science Foundation under Grant QN2018115, and in part by the S&T Major Project of the Science and Technology Ministry of China under Grant 2017YFE0135700.

ABSTRACT Seawater quality prediction has a tremendous potential of enabling future smart ocean. However, this time-sensitive application puts forward a strict delay requirement, thus easily leading to overwhelmed networks. Edge computing is emerging as an effective means of solving network overload, due to its edge-based distributed processing. Therefore, we develop a hybrid multivariate prediction model for seawater quality assessment in an edge computing environment, considering the combination of principal component analysis (PCA) and relevance vector machine (RVM). The PCA method is employed for dimension reduction of ten seawater quality factors in advance. Six principal components are extracted from multiple features, used as input variables of the subsequent predictor. Finally, a RVM is developed to predict the future trends of dissolved oxygen and pH, measuring seawater quality. Experimental results on the real-world ocean sensor data show that our PCA-RVM based multivariate prediction model outperforms single RVM, SVM and its extended version in prediction accuracy and efficiency, meanwhile statistical testings confirm this finding.

INDEX TERMS Edge computing, PCA-RVM, water quality, multivariate prediction, smart ocean.

I. INTRODUCTION

With the rapid development of Internet of things, artificial intelligence, edge computing and 5G [1]–[5], ocean environmental monitoring is being developed into an intelligent computing platform supplied by various kinds of processing units, transmission devices, and sensors [6]. Hence, future ocean informatization should become smarter to provide better experiences for users [7]. Currently, human activities have devoted to exploring many fields of the ocean. It is generally known that ocean covers more than 70% of the earth, thus it is crucially important to the survival and development of humanity. Ocean environmental monitoring is regarded as the only way of oceanic information acquirement, including meteorological and hydrological conditions, sediments, life, ocean dynamics [8]–[10]. These information can be accessed by remote sensing, shore-based radar, and buoy systems. Many types of ocean information processing are the key means to capture valuable knowledge in ocean data, especially seawater quality prediction. It is greatly beneficial for ocean environmental protection, disaster warning, and fishery safety [11].

Massive data, generated from oceanic sensor devices, inevitably requires high-bandwidth to guarantee efficient transmission. Along with an explosive increase of such data, the communication network may be finally overwhelmed by the overload transmission. It throws a severe challenge to online data processing. In this situation, mobile edge computing (MEC) is viewed as a workable solution. Through deployment of cloud-like infrastructure near data sources, data can be preferentially handled at the edge, largely relaxing the whole network [12]. Given this, this paper aims at investigating the MEC-based seawater quality prediction, in order to make the quick response for potential ocean emergencies.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaofei Wang.

Machine learning has become the mainstream methodology for ocean-related data processing. In this scenario, support vector machine (SVM) has been broadly studied as a state-of-art supervised regression method. Its popularity is attributed largely to the good nonlinear approximation capacity, often beating other coeval alternatives such as neural networks and decision tree [13]. Tan *et al.* demonstrated the adaptability of SVM to predict real-time water quality data in the mode of small sample [14]. Liu *et al.* developed a hybrid prediction model based on the combination of wavelet analysis, SVM and extended Cauchy particle swarm optimization for dissolved oxygen content in intensive aquaculture [15]. Kisi *et al.* accessed the performance of SVM in modeling the river water pollution, measured by chemical oxygen demand, total coliform, fecal coliform and pH, respectively [16]. Singh *et al.* considered both classification and regression of SVM for the data processing related to surface water quality, effectively assisting the optimization of water monitoring program [17]. Kong *et al.* proposed an optimized SVM for real-time eutrophication status evaluation, considering four CDOM fluorescence components, chlorophyll a (Chl-a) and dissolved oxygen (DO) [18]. Despite these recent advancements, SVM also shows a series of ill-posed problems, i.e., incalculable posterior probability distribution, inestimable penalty factor C, and necessary Mercer's condition for kernel projecting. As another classical kernel ML model, relevant vector machine (RVM) is a Bayesian alternative to SVM, significantly alleviating the above defects [19]. It implies that RVM can implement probability distribution calculation for output, no cross-validation is required for hyper-parameters, and its kernel function can be specified arbitrarily without constraints of positive paradigm. To the best of our knowledge, there are few works on RVM-based seawater quality prediction. Thus in the paper, we intend to explore the possibility of RVM predictor in this task.

Actually, seawater quality greatly depends on the fluctuations of physical, chemical and biological factors, such as temperature (TEM), conductivity (CON), turbidity(TUR), salinity (SAL), pH, total soluble solids (TSS), DO, dissolved oxygen saturation (DOS), Chl-a, and blue-green algae (BGA). Generally, DO and pH are used to measure the water quality [15]–[18]. Moreover, it has been proved that multivariate prediction, other than univariate mode, is more beneficial for this assessment [17], since it can provide more knowledge for approximation. So far, it is a still particularly challenging problem [20]–[22]. Inspired by it, a multivariate RVM-based prediction model is considered in our scenario.

Besides, a great deal of studies have indicated that data-driven single ML models, especially for high-dimensional data, don't have the competitive advantages in nonlinear approximation tasks. Principal component analysis (PCA) is viewed as a feasible solution for this issue [23]–[25]. In theory, it can achieve covariance-matrix-based dimensionality reduction for given correlated variables, thereby providing more favorable input for regression. The key idea behind PCA is to determine a series of orthogonal basis vectors projecting high-dimensional data into a low-dimensional space [26]. As a result, many interrelated variables can be converted into fewer uncorrelated variables, i.e., principal components. Therefore, our multivariate time series requires the PCA-enabled preprocessing for better seawater quality evaluation.

This paper investigates the combination of PCA and RVM in multivariate seawater quality prediction. In this structure, PCA is responsible for dimension reduction of multiple types of data, but preserving their variations as many as possible. Driven by selected principle components, RVM can achieve a nonlinear approximation for DO and pH. The case on real-world multivariate seawater data processing is implemented to access the suitability of the PCA-RVM model. The main contributions of this paper are shown below.

1) Our PCA-RVM is a successful attempt for ocean-related multivariate time series prediction. The multi-source and heterogeneous ocean data makes accurate prediction difficult. Our proposal can solve this modeling problem, contributing to the establishment of smart ocean.
2) PCA is used for dimensionality reduction on ten different types of ocean time series. It devotes to improving the computational efficiency of the RVM predictor.
3) The superior performance of our PCA-RVM over three state-of-the-art model are demonstrated on the basis of trend comparison, prediction accuracy and training time. In addition, the utility of PCA-RVM is accessed in terms of the pattern matching of two statistical measures.

The rest of this paper is organized as follows. In Section II, the framework and algorithm of PCA and RVM are introduced in detail. Section III verifies the applicability of the model through simulation experiments. Section IV summarizes the study, and puts forward the future study direction.

II. METHODOLOGIES

The overall prediction framework in edge computing environment is shown in the Fig 1. The edge computing node can gather data from oceanic collectors, such as buoy, surveying vessel, underwater vehicle and aircraft. Our seawater quality prediction can be achieved in this edge. Functionally, this framework consists of two key functional components, namely the PCA analyzer and the RVM predictor. The former lies in reasonably reducing the dimension of multivariate ocean data, to obtain good clustering quality, while the latter realize a good nonlinear approximation on this basis.

Firstly, various types of seawater quality data are collected through sensor devices deployed in ocean. Generally speaking, multivariate time series often contain more dynamic information of the seawater quality than the univariate time series. Then this multivariate series is injected into the PCA analyzer for data extraction, where redundant information hidden in multiple variables can be eliminated by linear transformation. Actually, this data preprocess can provide more excellent initialization input for predictor. Finally, we consider the multi-input-single-output RVM predictor for

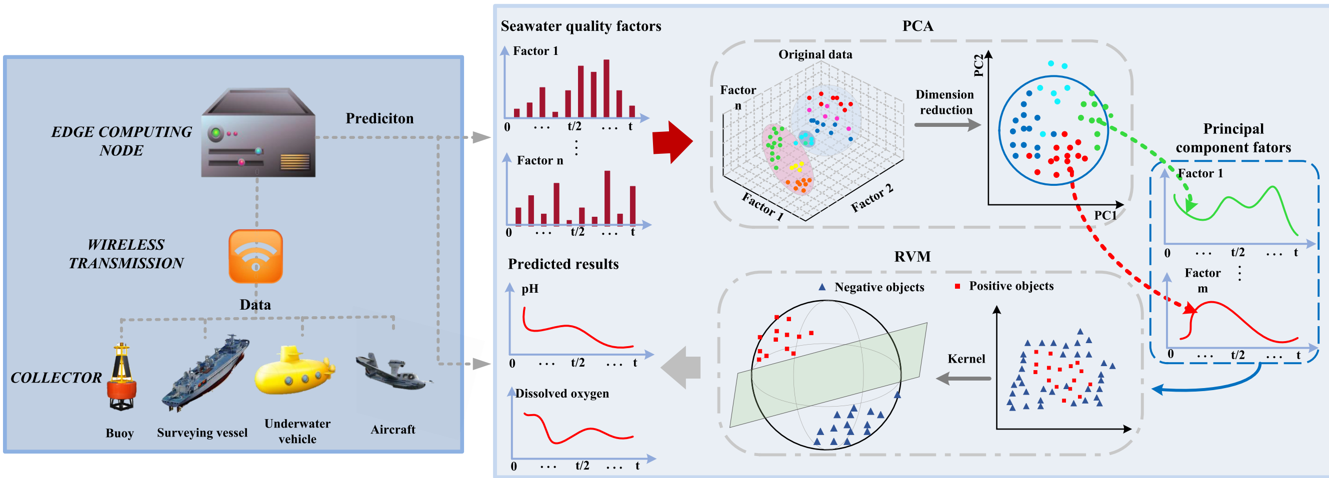


FIGURE 1. **Schematic diagram of PCA-RVM based seawater quality prediction in ocean edge computing scenario.**

seawater quality modeling, where the input signal is a combination of all seawater quality factors, and the model readout is either pH or DO. In order to better understand our proposal, we describe these characteristics of both PCA and RVM in detail.

A. DIMENSION REDUCTION

PCA is a variable reduction routine, which is very useful for processing a lot of relevant information in datasets. Principal components are linear transformations of input variables, and are sorted according to the standard of relevant information represented by the variance. The maximum amount of information in datasets is represented by the first component and the remaining information is optimized by the other ones under the restrictions of not related to other components. Basically, reducing the number of variables is the purpose of this method. The specific algorithm is described as follows.

Firstly, m samples with n observed attributes are taken into matrix $X_{m\times n}$. The above matrix is normalized, given by:

$$Y_{pq} = \frac{\left(X_{pq} - \overline{X_q}\right)}{s_q}, \quad p = 1, 2, \ldots, m;\; q = 1, 2, \ldots, n \quad (1)$$

where Y is expressed as a standardized sample matrix, $\overline{X_q}$ and s_q are the mean and standard deviation of the observed attributes in column q, respectively.

Based on the standardized matrix, a correlation coefficient matrix R is constructed as follows:

$$R = \left[r_{pq}\right]_{n\times n} = \frac{1}{m-1}\left(Y^T Y\right) \quad (2)$$

Then according to the matrix R, the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ can be calculated and sorted from large to small, which is $\lambda_1 > \lambda_2 > \ldots > \lambda_n$. Next, we calculate principal component contribution rate (CR), expressed as:

$$CR_p = \frac{\lambda_p}{\sum_{p=1}^{n} \lambda_p} \quad (3)$$

The number of principal components to be determined at last is determined by cumulative contribution rate (CCR), that is

$$CCR\,(j) = \frac{\sum_{p=1}^{j} \lambda_p}{\sum_{p=1}^{n} \lambda_p} \geq 95\% \quad (4)$$

Only these components representing information above 95% in the total variable are retained.

Finally, for each $\lambda_i (i = 1, 2, \ldots, j)$, we try to solve the system of equations $Ra = \lambda_i a$ to obtain the unit feature vector a_i. The j principal components are acquired as follows:

$$U_{pq} = Y_p^T a_i, \quad i = 1, 2, \ldots j \quad (5)$$

where $Y_p = \left(Y_{p1}, Y_{p2}, \ldots, Y_{pn}\right)^T$.

B. NONLINEAR APPROXIMATION

After PCA, a RVM approximator is used to predict the water quality parameters. Consider a set of training samples $\{x_n, t_n\}_{n=1}^{N}$, where x_n and t_n denote the input vector and the target vector, the regression model of RVM is defined as follows:

$$\begin{cases} t_n = y\,(x_n; w) + \varepsilon_n \\ y\,(x_n; w) = \sum_{n=1}^{N} w_n K\,(x, x_n) + w_0 \end{cases} \quad (6)$$

where w are weight vectors, ε_n is a Gaussian distribution satisfying $\varepsilon_n \sim N\left(0, \sigma^2\right)$, and $K\,(x, x_n)$ is a kernel function. Here, the Gaussian radial basis function is used as the kernel for the study of this subject, which is widely used because of its low complexity. The Gaussian likelihood distribution of the target vector t can be estimated:

$$p\left(t|w,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-N/2}\exp\left\{-\frac{\|t-\Phi w\|^2}{2\sigma^2}\right\} \tag{7}$$

where

$$\begin{cases} t=(t_1,t_2,\ldots,t_N)^T \\ w=(w_0,w_1,\ldots,w_N)^T \\ \Phi=(\phi(x_1),\phi(x_2),\ldots,\phi(x_N)) \\ \phi(x_n)=[1,K(x_n,x_1),K(x_n,x_2),\ldots,K(x_n,x_N)]^T \end{cases}$$

To avoid model over-fitting problems, a Gaussian prior probability is defined to constrain the weights w:

$$p(w|\alpha)=\prod_{i=0}^{N}N\left(w_i|0,\alpha_i^{-1}\right) \tag{8}$$

where $\alpha=(\alpha_0,\alpha_1,\ldots,\alpha_N)$ is the hyperparameter.

According to prior distribution, the posterior distribution for sparse Bayesian learning can be written as:

$$p\left(w,\alpha,\sigma^2|t\right)=p\left(w|t,\alpha,\sigma^2\right)p\left(\alpha,\sigma^2|t\right) \tag{9}$$

The weight w posterior distribution is obtained by:

$$p\left(w|t,\alpha,\sigma^2\right)=\frac{p\left(t|w,\sigma^2\right)p(w|\alpha)}{p\left(t|\alpha,\sigma^2\right)} = N(w|\mu,\Sigma) \tag{10}$$

The posterior mean and covariance are described, given by:

$$\mu=\sigma^{-2}\Sigma\Phi^T t \tag{11}$$

$$\Sigma=\left(A+\sigma^{-2}\Phi^T\Phi\right)^{-1} \tag{12}$$

where $A=diag(\alpha)$.

An approximation is used for hyperparameter posteriors, and its purpose is to find the most likely values of α_{best} and σ_{best}^2. Hence, the search for the hyperparameter posterior replaces the relevance vector "learning".

$$p\left(t|\alpha,\sigma^2\right)=\int p\left(t|w,\sigma^2\right)p(w|\alpha)dw = N\left(t|0,\sigma^2 I+\phi A^{-1}\phi^T\right) \tag{13}$$

In the closed form, it is unable to obtain the value of α and σ^2 that maximize $p\left(t|\alpha,\sigma^2\right)$. So the optimization of $p\left(t|\alpha,\sigma^2\right)$ is performed by iterative re-estimation.

$$p(t_*|t)=\int p\left(t_*|w,\sigma_{best}^2\right)p\left(w|t,\alpha_{best},\sigma_{best}^2\right)dw = N\left(t_*|y_*,\sigma_*^2\right) \tag{14}$$

As can be seen from the computable Eq. (14), the model function weighted by the posterior mean is the predicted mean. The prediction variance includes two parts, namely the estimated noise on the data and the uncertainty of the weights. This means that our model can be used to make multivariate seawater quality prediction.

TABLE 1. CRs and CCRs of principal components.

Component	CR (%)	CCR (%)
1	36.8274	36.8274
2	21.7972	58.6246
3	20.5695	79.1941
4	11.5713	90.7654
5	3.6948	94.4602
6	3.3858	97.8460
7	1.6409	99.4869
8	0.4602	99.9471
9	0.0527	99.9998
10	0.0002	100

III. NUMERICAL SIMULATION

In the experiment, we consider the novel model combining PCA and RVM to predict seawater quality, which has great significance for water pollution treatment. Do and pH are important parameters of water quality, thus, this paper focuses on the prediction and analysis of them. In this section, the types and sources of experimental data are given firstly, followed by the PCA analysis of different seawater quality factors. Whereafter, RVM is used to make predictions with multidimensional inputs and single output, and its performance is compared with PCA-SVM, RVM and SVM. Specially by simulation verification, we determine that Gaussian kernel is the optimal projecting mode for this considered task. Finally, the performance of this model is further verified based on statistical methods.

A. DATA AND EVALUATION CRITERION

The experimental data are collected in a certain sea area in China by every half an hour for 20 consecutive days. It includes 10 chemical factors, i.e., SAL, CON, DO, Chl-a, TUR, BGA, TSS, DOS, TEM and pH. It is worth mentioning that DO and pH are important parameters for water quality assessment. Therefore in this paper, the effects of various water quality factors are integrated to predict the pH and DO through the proposed PCA-RVM model.

We take the PCA results of original seawater quality factors as input parameters, and the DO and pH as output parameters to configure PCA-RVM model. In this experiment, the first 800 data points are used as the training set, and the last 200 ones are used as the testing set. After 20 trials of simulation, we can make sure that kernel width $d = 14$ is the optimal parameter for PCA-RVM model through the grid search method. The prediction performance of the evaluated models is evaluated by mean absolute error (MAE), root mean

TABLE 2. Load matrix of principal components.

Factors	1	2	3	4	5	6
SAL	0.2916	0.1951	-0.3703	0.5038	0.0257	-0.1426
CON	0.4776	0.1508	0.1632	-0.0740	-0.0963	0.3585
DO	-0.3222	0.2137	0.3994	0.3820	-0.1529	-0.1612
Chl-a	-0.1930	0.4980	-0.2832	-0.1812	-0.0565	0.3406
TUR	-0.3831	-0.0904	-0.1713	0.3868	0.2673	0.6947
BGA	-0.2270	0.4153	-0.3215	-0.2158	-0.5581	-0.0814
TSS	0.3684	0.2004	-0.2776	0.4385	0.0298	-0.1410
DOS	-0.0603	0.3019	0.5597	0.3217	-0.2329	0.0660
TEM	0.4552	0.1271	0.2325	-0.1604	-0.1061	0.3993
pH	-0.0316	0.5634	0.1393	-0.2105	0.7171	-0.1903

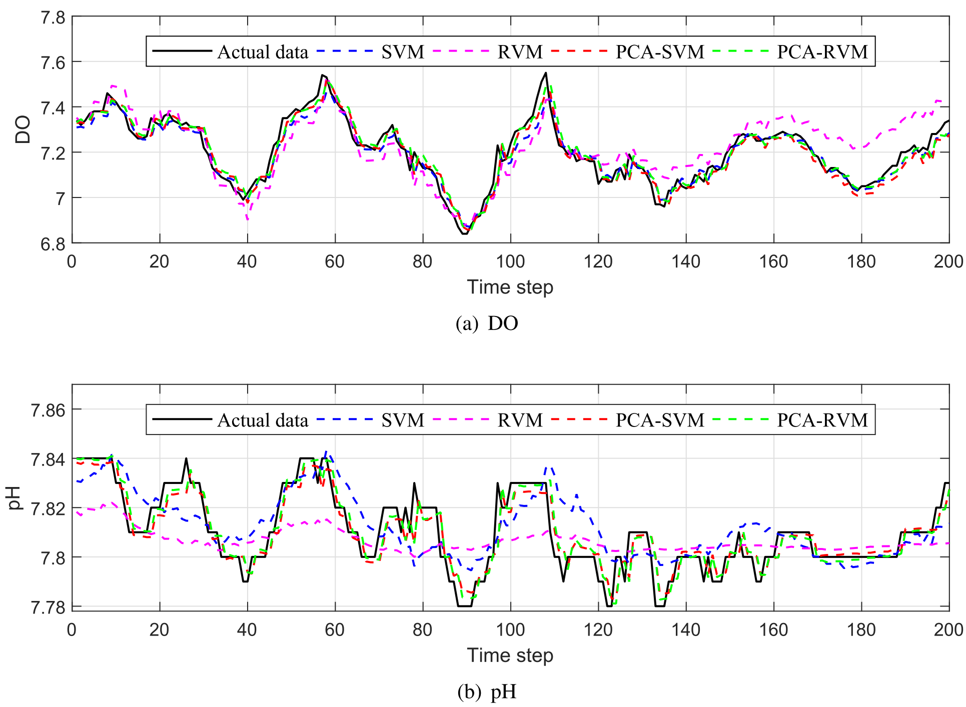


FIGURE 2. Outputs of trained PCA-RVM, PCA-SVM, RVM and SVM in multivariate seawater quality prediction problem.

square error (RMSE) and determination coefficient (CD), expressed as follows.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{16}$$

$$CD = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{17}$$

TABLE 3. Performance of the evaluated models in seawater quality prediction problem.

Model	Factor	MAE	RMSE	Time (s)
SVM	DO	0.0322	0.0437	0.0131
	pH	0.0095	0.0118	0.0150
RVM	DO	0.0740	0.0905	0.0070
	pH	0.0114	0.0139	0.0065
PCA-SVM	DO	0.0323	0.0431	0.0128
	pH	0.0048	0.0066	0.0127
PCA-RVM	DO	0.0315	0.0421	0.0056
	pH	0.0044	0.0066	0.0050

where n is the number of samples in the model test sets, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\hat{y_i}$ and y_i are predicted and real value respectively.

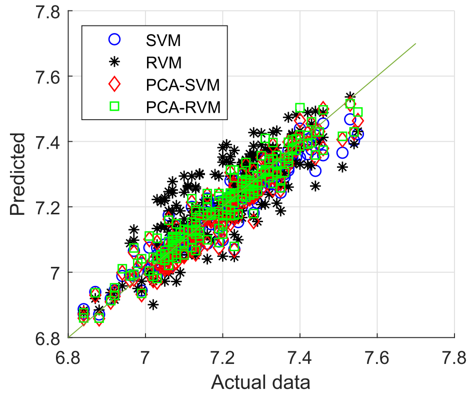
B. PCA ANALYSIS

As we all know, the factors affecting water quality are excessive and complex. PCA is used to analyze ocean water quality parameters. In this method, the key principal components affecting DO and pH are screened. Table 1 shows the results of the contribution rate (CR) and cumulative contribution rate (CCR) based on PCA respectively. Take the second component as an example, the CR of it is 21.7972%, while the value of CCR is 58.6246%. CCR is the sum of the CRs of the first component and the second component, *i.e.*, 36.8274% and 21.7972%. Analogously, the CCR of the third component is obtained by summing 36.8274%, 21.7972% and 20.5695%. In our scenario, the primary components are selected when the value of CCR exceeded 95%. Following this guidelines, the first six components are identified as principal components, which are used to replace the original variables.

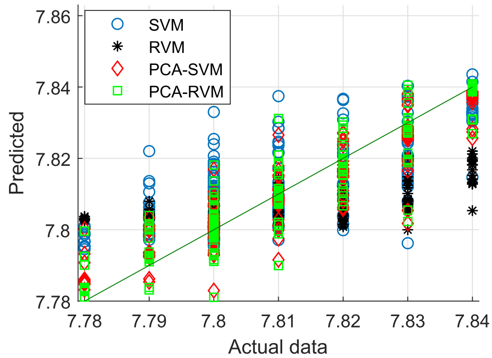
In order to further reveal intrinsic characteristics of principal components, Table 2 shows the influence degree of seawater quality factors on each principal component by calculating principal component load. Generally, the absolute value of each number is used as an object of comparison. It can be seen from Table 2 that in the first component, the absolute value of the conductivity appears to be maximal among these considered seawater quality factors, reflecting the preeminent influence. Besides, other factors that have a greater impact are TEM and TUR. Likewise, pH also shows the significant influence degree in the second component, followed by Chl-a, BGA and so on.

C. PREDICTION PERFORMANCE

Fig. 2 show the comparative curves of actual signal and predicted outputs obtained by PCA-RVM and its competitors for seawater quality factor prediction. As is intuitively seen from this figure, PCA-RVM and PCA-SVM have relatively better fitting effects for DO and pH than the other two models. It indicates that PCA is greatly beneficial to enhance prediction performance of single RVM and SVM.



(a) DO



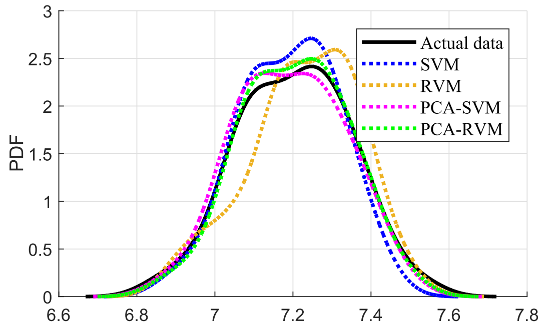
(b) pH

FIGURE 3. Scatter plot of the observed versus the predicted values for PCA-RVM, PCA-SVM, RVM and SVM in multivariate seawater quality prediction problem.

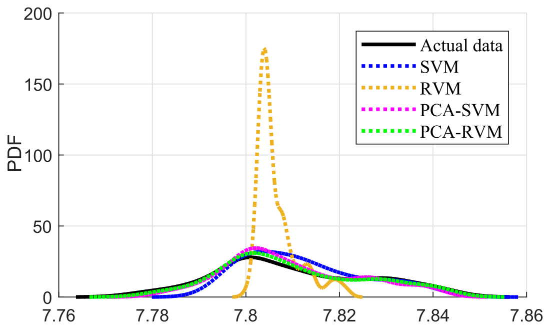
Table 3 lists corresponding prediction capacities of all evaluated models. It covers two aspects of quantifications, including accuracy measures based on MAE and RMSE, as well as computational cost embodied by training time. It should be specially explained that training time refers to time consumption of the PCA-based data dimension reduction and the RVM learning. From this table, we can see that our PCA-RVM is slightly better than PCA-SVM in the prediction tasks of DO and pH, whereas significantly outperforming RVM and SVM. Moreover, it is worth noting that the prediction accuracy of PCA-RVM are orders of magnitude higher than the ones of SVM and RVM. Besides, Table 3 also measures the training time for each model in this case study. On the whole, PCA-RVM shows extremely computationally efficient compared to other models, mainly due to the dimension reduction of multivariate series and sparser RVM learning.

D. STATISTICAL VALIDATION

Here, we validate the prediction performance of the PCA-RVM from the view of statistical validation,



(a) DO



(b) pH

FIGURE 4. **PDF comparisons between the actual signal and the predicted outputs produced by PCA-RVM, PCA-SVM, RVM and SVM in multivariate seawater quality prediction problem.**

considering scatter plot and probability density function (PDF). Scatter plot is an intuitive tool for visualizing data distribution. In our task, it is used to roughly observe the correlation between actual signal and predicted output. Scatter plot can also identify outliers effectively based on the density of point domain. PDF is another important reliability measure, describing the probability variation of input samples related to seawater quality factors. For an interval of the same length, the probability of certain sample falls in this interval is greater as its PDF gets larger.

Scatter plots of the predicted and observed values of DO and pH are illustrated in Fig.3. Specially, the green line $y = x$ in this plot represents a perfect prediction, indicating that there exists an exact match between the predicted and observed values. It is obvious that the more predicted points are concentrated along the datum line for PCA-RVM and PCA-SVM compared to RVM and SVM. It can further be quantified by the CD measure, as depicted in Table 4. As is seen, the most highest CD can be yielded by PCA-RVM in the pH prediction task, suggesting its most powerful nonlinear approximation. Nevertheless for DO, the CD value of PCA-RVM is slightly worse better than the one of PCA-SVM. It seems that the result is inconsistent with the finding of the best-behaved PCA-RVM in Table 3. This may be attributed to the presence of few significant outliers.

Fig. 4 shows the comparative PDF plots on the desired signal and the predicted outputs for verifying the utility of

TABLE 4. **CDs of the evaluated models in seawater quality prediction problem.**

Factor	SVM	RVM	PCA-SVM	PCA-RVM
DO	0.8917	0.7151	0.9113	0.9111
pH	0.5328	0.1412	0.8127	0.8371

the considered models in the seawater prediction problem. As is seen from this figure, the probability density curve of PCA-RVM is most consistent with the actual one, followed by PCA-SVM and SVM. Doubtlessly, RVM is still a worst-performing representative in this task.

Given these, these statistical visualizations again confirm conclusions in Section III-C. That is to say, our PCA-RVM is the best multivariate prediction model in seawater quality assessment.

IV. CONCLUSION

In this paper, we develop a PCA-RVM based seawater quality prediction model. Different from univariate prediction, our PCA-RVM can perform a multivariate prediction mode, whose input consists with ten different types of ocean factors, and readout is the group of two determination factors measuring seawater quality. The PCA method aims at reasonably reducing the dimension of multivariate ocean data, to obtain good clustering quality, while the RVM can realize a good nonlinear approximation on this basis. Numerical analyses confirm that the proposed model has a higher prediction ability but lower time consumption than other implementation approaches. For the future, it would be interesting to exploring the influence of multiple PCA versions on the RVM-based multivariate prediction.

REFERENCES

[1] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.

[2] N. J. Nilsson, *Principles of Artificial Intelligence*. San Mateo, CA, USA: Morgan Kaufmann, 2014.

[3] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, to be published, doi: 10.1109/COMST.2020.2970550.

[4] H. Zhang, M. Feng, K. Long, G. Karagiannidis, and A. Nallanathan, "Artificial intelligence based resource allocation in ultra dense networks," *IEEE Veh. Technol. Mag.*, vol. 14, no. 4, pp. 56–63, Oct. 2019.

[5] H. Zhang, H. Zhang, W. Liu, K. Long, J. Dong, and V. C. M. Leung, "Energy efficient user clustering, hybrid precoding and power optimization in terahertz MIMO-NOMA systems," *IEEE J. Sel. Areas Commun.*, to be published.

[6] J. Yang, J. Wen, Y. Wang, B. Jiang, H. Wang, and H. Song, "Fog-based marine environmental information monitoring towards ocean of things," *IEEE Internet Things J.*, to be published, doi: 10.1109/jiot.2019.2946269.

[7] S. Shen, Y. Han, X. Wang, and Y. Wang, "Computation offloading with multiple agents in edge-computing–supported IoT," *ACM Trans. Sensor Netw.*, vol. 16, no. 1, pp. 1–27, Feb. 2020.

[8] M. Ertogan, G. T. Tayyar, P. A. Wilson, and S. Ertugrul, "Marine measurement and real-time control systems with case studies," *Ocean Eng.*, vol. 159, pp. 457–469, Jul. 2018.

[9] C. Trasviña-Moreno, R. Blasco, Á. Marco, R. Casas, and A. Trasviña-Castro, "Unmanned aerial vehicle based wireless sensor network for marine-coastal environment monitoring," *Sensors*, vol. 17, no. 3, p. 460, 2017.
[10] Z. Zhang, Y. Shao, W. Tian, Q. Wei, Y. Zhang, and Q. Zhang, "Application potential of GF-4 images for dynamic ship monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 911–915, Jun. 2017.
[11] G. Levy, S. Vignudelli, and J. Gower, "Enabling Earth observations in support of global, coastal, ocean, and climate change research and monitoring," *Int. J. Remote Sens.*, vol. 39, no. 13, pp. 4287–4292, Jul. 2018.
[12] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Sep. 2019.
[13] M. Pal and G. M. Foody, "Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1344–1355, Oct. 2012.
[14] G. Tan, J. Yan, C. Gao, and S. Yang, "Prediction of water quality time series data based on least squares support vector machine," *Procedia Eng.*, vol. 31, pp. 1194–1199, 2012.
[15] S. Liu, L. Xu, Y. Jiang, D. Li, Y. Chen, and Z. Li, "A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture," *Eng. Appl. Artif. Intell.*, vol. 29, pp. 114–124, Mar. 2014.
[16] O. Kisi and K. S. Parmar, "Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution," *J. Hydrol.*, vol. 534, pp. 104–112, Mar. 2016.
[17] K. P. Singh, N. Basant, and S. Gupta, "Support vector machines in water quality management," *Analytica Chim. Acta*, vol. 703, no. 2, pp. 152–162, Oct. 2011.
[18] X. Kong, Y. Sun, R. Su, and X. Shi, "Real-time eutrophication status evaluation of coastal waters using support vector machine with grid search algorithm," *Mar. Pollut. Bull.*, vol. 119, no. 1, pp. 307–319, Jun. 2017.
[19] C. Chen, Y. Wang, Y. Zhang, and Y. Zhai, "Indoor positioning algorithm based on nonlinear PLS integrated with RVM," *IEEE Sensors J.*, vol. 18, no. 2, pp. 660–668, Jan. 2018.
[20] M. Han and M. Xu, "Laplacian echo state network for multivariate time series prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 238–244, Jan. 2018.
[21] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 6085.
[22] M. Han, R. Zhang, T. Qiu, M. Xu, and W. Ren, "Multivariate chaotic time series prediction based on improved grey relational analysis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 10, pp. 2144–2154, Oct. 2019.
[23] Y. Luo and B. Wang, "Prediction of negative conversion days of childhood nephrotic syndrome based on PCA and BP-AdaBoost neural network," *IEEE Access*, vol. 7, pp. 151579–151586, 2019.
[24] Z. Geng, J. Chen, and Y. Han, "Energy efficiency prediction based on PCA-FRBF model: A case study of ethylene industries," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 8, pp. 1763–1773, Aug. 2017.
[25] I. L. Mallika, D. V. Ratnam, Y. Ostuka, G. Sivavaraprasad, and S. Raman, "Implementation of hybrid ionospheric TEC forecasting algorithm using PCA-NN method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 371–381, Jan. 2019.
[26] S. Duan, K. Chen, X. Yu, and M. Qian, "Automatic multicarrier waveform classification via PCA and convolutional neural networks," *IEEE Access*, vol. 6, pp. 51365–51373, 2018.

XIAOCHUAN SUN received the M.S. degree from the Guilin University of Electronic Technology, Guilin, China, in 2010, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2013.

He is currently with the College of Artificial Intelligence, North China University of Science and Technology, Tangshan, China. He is also a Postdoctoral Fellow with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include reservoir computing, neural networks, and deep learning architectures with applications in the domains of time series prediction, classification, and nonlinear system identification.

XIN WANG received the B.S. degree from the Taiyuan University of Technology, Taiyuan, China, in 2017. She is currently pursuing the master's degree with the College of Artificial Intelligence, North China University of Science and Technology, Tangshan, China. Her current research interests include support vector machine, relevance vector machine, and neural networks.

DI CAI received the B.S. degree from the School of Information Science and Technology, Agricultural University of Hebei, Baoding, China, in 2018. She is currently pursuing the master's degree with the College of Artificial Intelligence, North China University of Science and Technology, Tangshan, China. Her current research interests include neural networks and deep learning.

ZHIGANG LI received the M.S. degree from the Shanxi Agricultural University, Shanxi, China, in 1993, and the Ph.D. degree from the China University of Mining and Technology, Beijing, China, in 2007.

From 2010 to 2012, he joined the Research Station of Beijing Jiaotong University, as a Postdoctoral Research Fellow. He has been a Professor with the North China University of Science and Technology, Tangshan, China, since 2003, where, he is the Dean of the College of Artificial Intelligence. He is currently a Vice President of the Hebei Computer Society and Hebei Electronics Society. His current research content focuses on the directions of data mining and intelligent control theory.

YUANYUAN GAO received the B.S. degree from the Qinggong College, North China University of Science and Technology, Tangshan, China, in 2018, where she is currently pursuing the master's degree with the College of Artificial Intelligence. Her current research interests include neural networks and intelligent optimization algorithm.

XUSHENG WANG received the B.S. and M.S. degrees from the North China University of Science and Technology, Tangshan, China, in 1996 and 2009, respectively. He has been with the College of Information Engineering, North China University of Science and Technology, Tangshan, since 1996. He is currently the Secretary of the Party Committee of the College of Science. His current research content focuses on the artificial intelligence and pattern recognition.