

Article

Prediction Model and Influencing Factors of CO₂ Micro/Nanobubble Release Based on ARIMA-BPNN

Bingbing Wang^{1,2}, Xiangjie Lu^{1,2}, Yanzhao Ren³, Sha Tao^{1,2,*} and Wanlin Gao^{1,2}

¹ College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; snowice310@cau.edu.cn (B.W.); lxj0528@cau.edu.cn (X.L.); gaowlin@cau.edu.cn (W.G.)

² Key Laboratory of Agricultural Information Standardization, Ministry of Agriculture and Rural Affairs, Beijing 100083, China

³ School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China; 20211003@btbu.edu.cn

* Correspondence: taos@cau.edu.cn

Abstract: The quantitative prediction of CO₂ concentration in the growth environment of crops is a key technology for CO₂ enrichment applications. The characteristics of micro/nanobubbles in water make CO₂ micro/nanobubble water potentially useful for enriching CO₂ during growth of crops. However, few studies have been conducted on the release characteristics and factors influencing CO₂ micro/nanobubbles. In this paper, the factors influencing CO₂ release and changes in CO₂ concentration in the environment are discussed. An autoregressive integrated moving average and backpropagation neural network (ARIMA-BPNN) model that maps the nonlinear relationship between the CO₂ concentration and various influencing factors within a time series is proposed to predict the released CO₂ concentration in the environment. Experimental results show that the mean absolute error and root-mean-square error of the combination prediction model in the test datasets were 9.31 and 17.48, respectively. The R² value between the predicted and measured values was 0.86. Additionally, the mean influence value (MIV) algorithm was used to evaluate the influence weights of each input influencing factor on the CO₂ micro/nanobubble release concentration, which were in the order of ambient temperature > spray pressure > spray amount > ambient humidity. This study provides a new research approach for the quantitative application of CO₂ micro/nanobubble water in agriculture.

Keywords: CO₂ prediction; CO₂ enrichment; CO₂ micro/nanobubble; combined prediction model



Citation: Wang, B.; Lu, X.; Ren, Y.; Tao, S.; Gao, W. Prediction Model and Influencing Factors of CO₂ Micro/Nanobubble Release Based on ARIMA-BPNN. *Agriculture* **2022**, *12*, 445. <https://doi.org/10.3390/agriculture12040445>

Academic Editors: Xiaoli Zhang, Dengsheng Lu, Xiujuan Chai, Guijun Yang and Langning Huo

Received: 17 February 2022

Accepted: 21 March 2022

Published: 23 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Carbon dioxide (CO₂) concentration is an important environmental factor that affects crop growth [1,2]. Traditional CO₂ gas fertilizers cannot be used in open environments because of their high diffusivity. CO₂ micro/nanobubbles have a measurable surface zeta potential, long residence time, slow release process, high gas mass transfer efficiency, and easy adsorption [3–6]. CO₂ nanobubbles have been widely used in gas-induced flotation technology and wastewater aeration [7,8]. In agriculture, they can be used in CO₂ slow-release and enrichment technology to increase agricultural yield.

In recent years, quantitative evaluations of the performance of artificially enriched CO₂ in a controlled environment, the regulation of the crop growth environment, and effective promotion of photosynthesis during crop growth have been applied. Zhang and Yasutake et al. proposed a 3D computational fluid dynamics (CFD) model to simulate gas distribution in a greenhouse after CO₂ enrichment [9]. Moon et al. proposed a method to predict the CO₂ concentration in a greenhouse using a long short-term memory model that realizes the quantitative regulation of CO₂ enrichment and promotes the growth of strawberries in a greenhouse [10]. To achieve the precise regulation of agricultural planting environments, machine learning algorithms, such as artificial neural networks (ANNs),

are increasingly used in agricultural environmental information prediction [11,12]. Several studies focused on the application of combination prediction ANN models in agricultural management. In recent years, there has also been a significant increase in their application in different scientific fields, such as financial and economic research, meteorology, and agro-ecological research studies [13–15]. The backpropagation neural network (BPNN), a widely used ANN model, has become an important fundamental model for constructing combination models. Zou et al. used a combination of BPNN and autoregressive integrated moving average (ARIMA) models to provide a useful method for predicting temporal changes in soil salt and water content in agriculture [16]. Cheng et al. [17] proposed an ARIMA-BPNN model, which is a new weighting method for a combined forecasting model, to predict grain production in China. The experimental results showed that the combined forecasting model obtained using the minimum sum of squared errors (MSEs) method can effectively improve forecasting accuracy. These studies indicate that combination prediction models can successfully model the complex relationship between independent and dependent variables in various scenarios.

Moreover, the stability of micro/nanobubbles is affected by factors, such as the gas type, bubble size, and liquid viscosity [18,19]. However, there has been no relevant research on CO₂ micro/nanobubbles and their release concentration characteristics in the environment. This study aims to clarify the spatiotemporal distribution characteristics and factors influencing the release of CO₂ from CO₂ micro/nanobubble water and construct a related prediction model. The specific objectives are to: (1) analyze the factors that affect the release of CO₂ from micro/nanobubble water and determine the temporal and spatial distribution characteristics of the CO₂ concentration after it is released; (2) construct the ARIMA-BPNN combination prediction model coupled with temporal and spatial characteristics to predict the CO₂ release concentration in the environment after spraying the CO₂ micro/nanobubble water; and (3) evaluate the performance of the combination model and analyze the importance of the influencing factors of ambient temperature, environmental humidity, spraying pressure, and spraying amount.

2. Materials and Methods

2.1. Preparation of CO₂ Micro/Nanobubble Water

In this experiment, the micro/nanobubble-generating equipment created by Hangzhou AiXiyue Technology Co., Ltd. (Hangzhou, China) was used. This generator uses the pressurized dissolved gas release method. Figure 1 illustrates a schematic of the equipment principle. The micro/nanobubble water prepared and used in this study takes CO₂ gas as the gas source, and the preparation procedure includes the following parameters that are the optimal preparation parameters of the equipment: a gas–liquid ratio of 2.76%, inlet water temperature of 25.5 °C, and cycle time of 30 min. The bubble content and gas solubility in the CO₂ micro/nanobubble water created according to the specifications were optimized at 7.76 mg/L and the average bubble particle size was 134.9 μm.

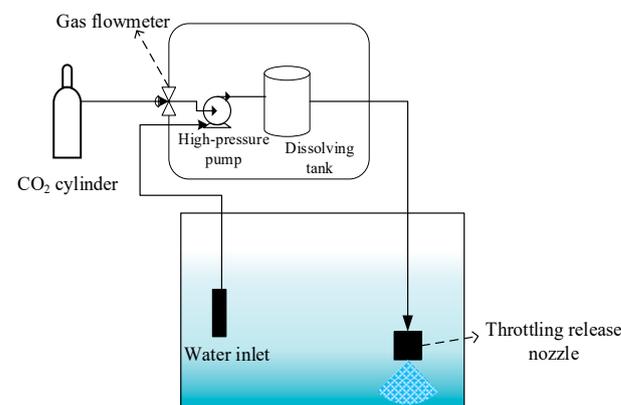


Figure 1. Schematic illustration of MNB generator.

2.2. Construction of Experimental Environment

To explore the release characteristics and concentration distribution of CO₂ micro/nano-bubbles in a particular spatial environment, we designed and developed a set of CO₂ concentration-monitoring devices for CO₂ micro/nanobubble release. The device consists of liquid supply, temperature control, and parameter monitoring assemblies. The temperature control assembly was composed of a temperature control box, temperature-regulating equipment, and an incubator. The internal spaces of the temperature control box and incubator were 180 cm × 120 cm × 150 cm and 60 cm × 40 cm × 100 cm (length × width × height), respectively. The role of the temperature control assembly was to avoid the impact of evaporation by micro/nanobubble water that can disturb data monitoring in the environment. The temperature control equipment was kept in a relatively constant experimental environment outside the incubator to avoid damage to the devices in the incubator. The liquid supply assembly was mainly composed of a micro/nanobubble generator and spray component. Figure 2 shows the spraying head coordinates as (30 cm, 20 cm, and 0 cm). The device enabled micro/nanobubble water to be ejected in the form of droplets of different particle sizes. Three sensors (Jingxun Changtong Electronic Technology Co., Ltd., Weihai, China) that could capture the CO₂ concentration, temperature, and humidity from the data monitoring assembly were installed at three monitoring points (20 cm, 40 cm, and 30 cm), (20 cm, 40 cm, and 60 cm), (20 cm, 40 cm, and 90 cm), referring to the coordinates in Figure 2. Thus, the released concentration of micro/nanobubbles could be measured at various temperatures. The impact of the temperature on the release of micro/nanobubbles could be studied by varying the temperature of the monitoring device. The details are shown in Figure 2.

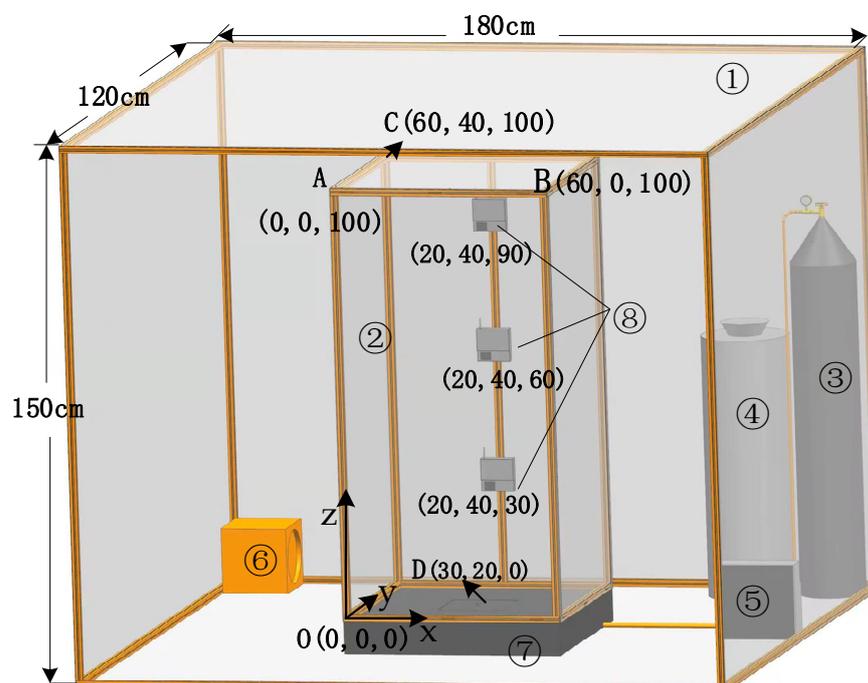


Figure 2. Schematic diagram of the experimental environment. ① Temperature control box; ② incubator; ③ CO₂ cylinder; ④ storage tank; ⑤ micro/nanobubble generator; ⑥ temperature regulating equipment; ⑦ spraying components; and ⑧ data monitoring assembly.

2.3. Design of CO₂ Gas-Release Experiment

To investigate the distribution characteristics and influencing factors of the CO₂ concentration in the environment following the release of CO₂ micro/nanobubbles, the following experiments were conducted. This experiment obtained the initial ambient temperature (25–35 °C) and initial ambient humidity (20–50% RH), and used the sensors to note the

temperature, humidity, and CO₂ concentration within 2 hours after spraying CO₂ micro/nanobubble water at the current site (30, 60, and 90 cm); the sampling frequency of the sensor was 1 min. The spray pressure and amount of micro/nanobubble water per unit area could be adjusted manually to 0.15–0.35 Mpa and 1.0–2.0 L, respectively.

2.4. Data Analysis Tools

The experimental data were stored in .xlsx format and the results were visualized using the Origin (version 2021b) software. All implementations of the programming code were performed using the Python programming language in the Pycharm (version 2019) integrated development environment. The ARIMA model was constructed using the Python third-party toolkit ARIMA module and the BP neural network model was constructed using the Python third-party library scikit-learn.

3. Fundamentals Analysis

3.1. ARIMA Model

The ARIMA model is a time-series modeling technique that can be used to capture the properties of linear data [20]. It is a method for forecasting nonstationary time-series data with a high-precision linear time-series forecasting approach. The model is divided into three components: autoregressive (AR (p)), moving average (MA (q)), and ARIMA (p, q) [21]. Equation (1) expresses this concept:

$$\begin{cases} \Phi(B)\nabla^d X_t = \Theta(B)a_t \\ E(a_t) = 0, \text{Var}(a_t) = \sigma_a^2, E(a_t a_s) = 0, s \neq t, \\ EX_s a_t = 0, \forall s < t \end{cases} \quad (1)$$

where $\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ is the autoregressive correlation coefficient polynomial of the time-series ARIMA model, $\nabla^d = (1 - B)^d$ is the higher-order difference, X_t is a time series, and $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ is a moving average coefficient polynomial.

The construction of the ARIMA model includes four stages: data stationarity testing, model establishment, model testing, and prediction [22]. The specific prediction steps are as follows:

Step 1: Determine the stationarity of the data according to the data autocorrelation function (ACF), partial autocorrelation function (PACF), variance and scatter plot, or unit root test method.

Step 2: Stationary processing of the nonstationary series. The autoregressive method is used to perform d-order differencing processing on the nonstationary series to stabilize the time series and extract effective information from the time-series data.

Step 3: Model identification and selection. The ACF and PACF are examined to select the model type.

Step 4: Select orders q and p from the ARIMA model. Several methods were developed based on the Akaike information criterion (AIC) [23], minimum description length (MDL) [24], AIC, Bayesian information criterion (BIC) [25], and Hannan–Quinn information criterion (HQIC) in the ARIMA model.

Step 5: Model checking. The model is tested according to the results of the model-residual ACF and PACF plots and the final ARIMA model-fitting diagram. When the model residual sequence lies within the confidence interval, the model passes the verification. Figure 3 shows the flowchart of the algorithm.

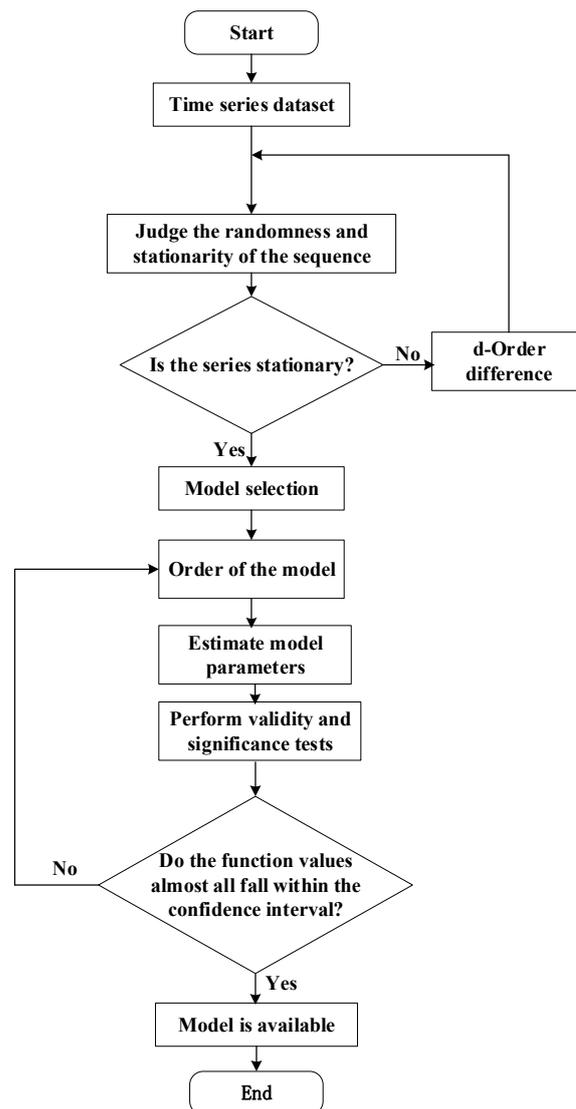


Figure 3. ARIMA model construction process.

3.2. BPNN

The BPNN model is a multilayer feedforward neural network trained by an error back-propagation algorithm [26]. Its main structure is composed of input, output, and hidden layers, and the different nodes in the same layer are independent and have no influence on each other. The output of each node only affects the output of its corresponding node in the next layer. It has been widely used in prediction, evaluation, language recognition, self-applicable control, and other fields [27–29]. The BPNN model is a supervised machine learning model that adopts an error-reverse algorithm and causes the output results to converge to the expected value by adjusting the weights and other parameters during training.

For a three-layer BP network, we assume that the number of neurons in the input, hidden, and output layers are n , m , and q , respectively, and the input to the neurons is $x_i^j = \sigma\left(\sum_{i=1}^n w_{ik}x_i + b_j\right)$, where $j = 1, 2, 3, \dots, p$; the output from the neurons in the output layer is $y_k = \sum_{j=1}^p w_{ik}x_j + b_k$, where $k = 1, 2, 3, \dots, m$. The network transmits the error between the predicted and actual values, and updates the weights between the neurons at each layer. The performance of the BPNN is mainly affected by the number of nodes and learning rate at each layer.

4. CO₂ Emission-Concentration Prediction with Spatiotemporal Coupled Properties Based on ARIMA-BPNN

4.1. Construction of the ARIMA-BPNN Hybrid Model

Linear regression can be considered if the periodic characteristics of a time series are stable over time and there is a correlation between different series within a certain timeframe. Equation (2) shows the linear relationship.

$$y_t = \alpha + \beta x_t + e_t, \tag{2}$$

where y_t and x_t are two different time series and e is the error series. If e_t has autocorrelation, the model is a regression model with time series errors [30]. Compared to the general situation, the ARIMA model considers the endogenous relationship of the sequence and quantifies the influence of external variables that can theoretically improve the prediction effect when the external environment changes [31]. The release process of CO₂ micro/nanobubbles is affected by external environmental factors and equipment parameters. These factors have a nonlinear relationship with the concentration of CO₂ released into the environment. Simultaneously, the released concentration of CO₂ exhibits periodicity in the time series under the same conditions. Therefore, to accurately predict the release concentration of CO₂ micro/nanobubbles, an ARIMA-BPNN combination prediction model is proposed. Figure 4 shows the flowchart of the process.

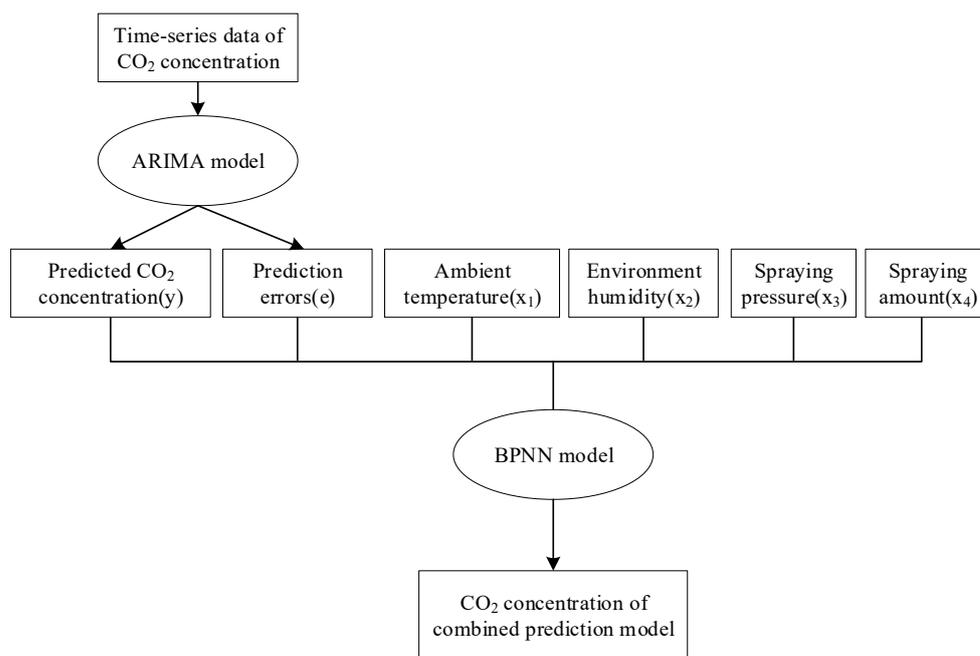


Figure 4. Flowchart of the ARIMA-BPNN hybrid model.

4.2. Calculation of CO₂ Concentration Spatiotemporal Coupling

An increase in CO₂ concentration in short-distance spaces affects the CO₂ concentration in long-distance spaces over time, owing to environmental factors, such as wind speed, temperature, and humidity. The CO₂ concentration in long-distance spaces reacts with the concentration in short-distance spaces, which is affected by the density of CO₂ gas [32,33]. To describe the temporal correlation of a single measurement point and the spatial correlation of multiple measurement points, a “spatiotemporal coupling coefficient” with spatiotemporal characteristics is proposed to describe the relationship between CO₂ release concentrations in the experimental environment and quantitatively describe the

intensity of spatial correlation at different times. The spatiotemporal coupling coefficient of the CO₂ concentration can be calculated using Equation (3).

$$R(X_A, X_B, \tau) = \frac{\text{cov}_\tau(X_A, X_B)}{\sigma_{X_A} \sigma_{X_B}}, \quad (3)$$

where X_A and X_B represent CO₂ concentration time-series data in two spatial points in the environment, $X_A = \{x_{A,t}\}_{t=1}^N$ and $X_B = \{x_{B,t}\}_{t=1}^N$; τ is the time-delay parameter that represents the time delay of the CO₂ concentration between spatial distances A and B . In this experiment, τ was equal to the time between the collection intervals of the CO₂ concentration. σ_{X_A} and σ_{X_B} represent the standard deviation of the time-series data on CO₂ concentrations at various spatial distances A and B . Additionally, $\text{cov}_\tau(X_A, X_B)$ is the covariance between X_A and X_B based on the time-delay parameter, and its calculation formula is as follows:

$$\text{cov}_\tau(X_A, X_B) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (x_{A,t} - \mu_{X_A})(x_{B,t+\tau} - \mu_{X_B}). \quad (4)$$

The measurement unit for the time-lag parameter was in minutes (min) because CO₂ micro/nanobubble release and gas diffusion characteristics refine the delay time of CO₂ concentration data into the minimum data collection interval. Thus, the delay time in the spatiotemporal coupling connection between the CO₂ concentration and temperature is only connected to the minimum collection interval of the concentration-collecting device, and is unrelated to the duration of the release of micro/nanobubble water. The parameters are more adaptable and, therefore, more capable of precisely describing the CO₂ spatiotemporal coupling coefficient.

4.3. Prediction of the Concentration of Released CO₂ Micro/Nanobubbles

The linear and nonlinear factors in the release process of CO₂ micro/nanobubbles were fully explored by constructing the ARIMA-BPNN hybrid model. Therefore, spatiotemporal coupling coefficients were used to measure the time-lag parameters of CO₂ concentration in different spaces, considering the temporal and spatial characteristics of CO₂-release concentration in the environment, for improving the prediction accuracy of the model. The prediction steps of the above combination model are as follows:

Step 1: Use the CO₂ concentration data with the time-series characteristics obtained in three different spaces to calculate the coupling coefficient using the Equation (3) coupling coefficient equation. The CO₂ concentration data of the corresponding time intervals are selected according to the size of the coupling coefficient for the construction and training of the ARIMA model.

Step 2: Select a suitable dataset of CO₂ concentration data collected within 2 h after spraying CO₂ micro/nanobubble water under different conditions and normalize these data to converge the value of the CO₂ concentration to the interval (0, 1).

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (5)$$

where x^* represents the CO₂ concentration after normalization, x_{\max} represents the maximum value in the dataset, and x_{\min} represents the minimum value in the dataset.

Step 3: Determine the parameters p , d , and q of the ARIMA model based on the data characteristics of CO₂ concentration. The ARIMA (p , d , q) model is used to predict the CO₂ concentration and obtain the preliminary prediction value of the CO₂ concentration, as shown in Algorithm 1.

Algorithm 1 ARIMA

```

Require: x
Ensure: y
1: for i = 0; i < 7; i++ do
2: if ad f(x) = true then
3: x ← Dif f
4: break
5: else
6: x ← Dif ference(x)
7: continue
8: p, q ← AIC (x), BIC(X), HQIC(x)
9: y ← ARIMA (x, p, d, q)

```

Step 4: Use the real value in the CO₂ concentration dataset in Step 2 to subtract the preliminary prediction value in Step 3 to obtain the residual value of the CO₂ concentration prediction.

Step 5: Determine the initialized weights of the BPNN and network training parameters.

The CO₂ concentration data predicted by the ARIMA model, coupled with spatiotemporal characteristics, were used as the actual values, and the other four parameters were used to construct a dataset for the neural network prediction model. The datasets were divided into training and test sets in an 8:1 ratio. The training set was used for model creation and training, whereas the test set was used to test the performance of the model. The pseudocode of the BPNN used for model training is shown in Algorithm 2:

Algorithm 2 BPNN

```

Require: y, x, net
Ensure: result
1: x[i] ← {tem[i], h[i], p[i], u[i], y[i], e[i]}
2: net.train(net, inputn, outputn)
3: inputntest ← mapminmax(inputntest)
4: BPsim ← sim(net, inputntest)
5: result ← mapminmax(reverse, BPsim)

```

5. Instance Simulation and Analysis of Results

5.1. Factors Involved in CO₂ Release and Dataset Selection

Considering the unique characteristics of CO₂ micro/nanobubble release, data were collected within 2 hours of spraying CO₂ micro/nanobubble water in the experimental environment. According to the experimental method described above, 12,600 datasets were obtained. Two methods were used to divide the datasets.

Dataset 1: Within 2 hours of spraying CO₂ micro/nanobubble water, the CO₂ concentration data were monitored by the sensors at 3 different spatial distances. Three parameters were included in the dataset: spatial distance, time after spraying, and CO₂ concentration. Subsequently, the time-series datasets of CO₂ concentrations were built based on the calculated spatiotemporal correlation coefficients. The first 80% CO₂ concentration time-series data were used as the training set to establish the ARIMA time-series model to predict the CO₂ concentration coupled with spatiotemporal characteristics. The second 20% CO₂ concentration data were used as the validation set to conduct the prediction-effect test.

Dataset 2: The CO₂ concentration data predicted by the ARIMA model coupled with spatiotemporal characteristics were used as the actual values, and the other four parameters were used to construct a dataset for the neural network prediction model. The datasets were divided into training and test sets in an 8:1 ratio. The training set was used for model creation and training, whereas the test set was used to test the performance of the model.

5.2. Simulation Parameters

The process of selecting the model parameters and experimental platform used in this study was as follows:

(1) ARIMA model

Step 1: Calculate the spatiotemporal correlation coefficient of the distance of 0.6 m and 0.9 m with a target space 0.3 m. The results are shown in Figure 5. It shows that, when the time-delay parameter of the 2 nontarget spaces is 60 s, the correlation coefficient with the target space is the largest, that is, 0.895 and 0.837, respectively. Therefore, the collection interval of the CO₂ concentration was 1 min, and the CO₂ concentration time-series dataset was constructed by coupling with the 0.3 m space.

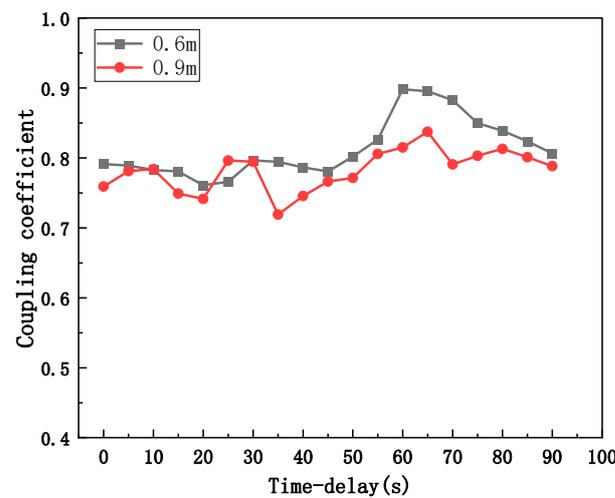


Figure 5. Space–time coupling coefficient of target space distance.

Step 2: Judge the stationarity of the sequence preliminarily by observing the diagram of CO₂-release concentration change and use the ADF test for the stationarity test. The ADF test results are shown in Table 1.

Table 1. Comparison of ADF test results.

Comparison Item	ADF	1% Significance Level	5% Significance Level	10% Significance Level
Before first-order difference test	−0.6987	−3.16	−2.89	−2.85
First-order difference test	−24.44	−3.44	−2.87	−2.57

From the table, the sequence is unstable before the differential operation, and the ADF value of the dataset is −24.44, which is evidently less than the 1%, 5%, and 10% significance level values after the first-order difference. The ADF test indicates that the data are stationary and reached the ARIMA model stabilization requirements after the first-order difference.

Step 3: Use the autocorrelation and partial correlation coefficients to estimate the model order. The results of data correlation detection are shown in Figures 6 and 7. To further identify the order of the model, numerous (p, q) combinations were set and the AIC, BIC, and HQIC values were compared under different combinations. Figure 8 shows that when the model order (p, q) combination was (4, 6), the AIC value was the smallest (635.66), the minimum BIC value was 631.89, and the model order was ARIMA (4, 1, 4); when the model order (p, q) combination was (3, 5), the HQIC value was the smallest. Considering the lowest AIC value as the premise, the value of (p, q) was (4, 6). Therefore, BIC and HQIC are 640.77 and 647.60, respectively, and the difference between them and the corresponding minimum value is the smallest. Therefore, the parameters of the experimental ARIMA model are ARIMA (4, 1, 6).

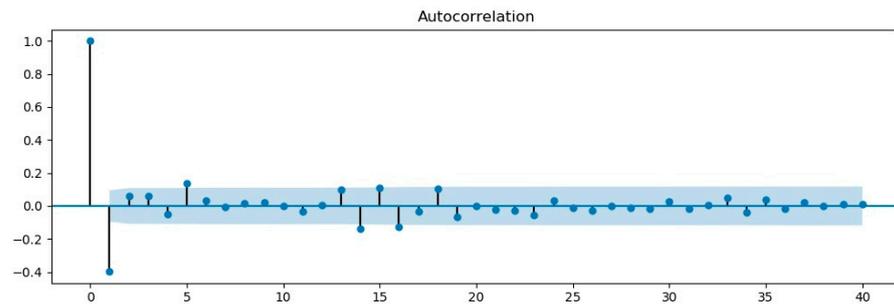


Figure 6. ACF diagram after first difference.

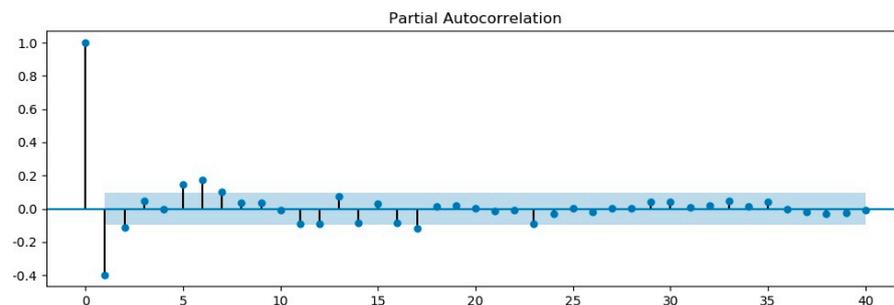


Figure 7. PACF diagram after first difference.

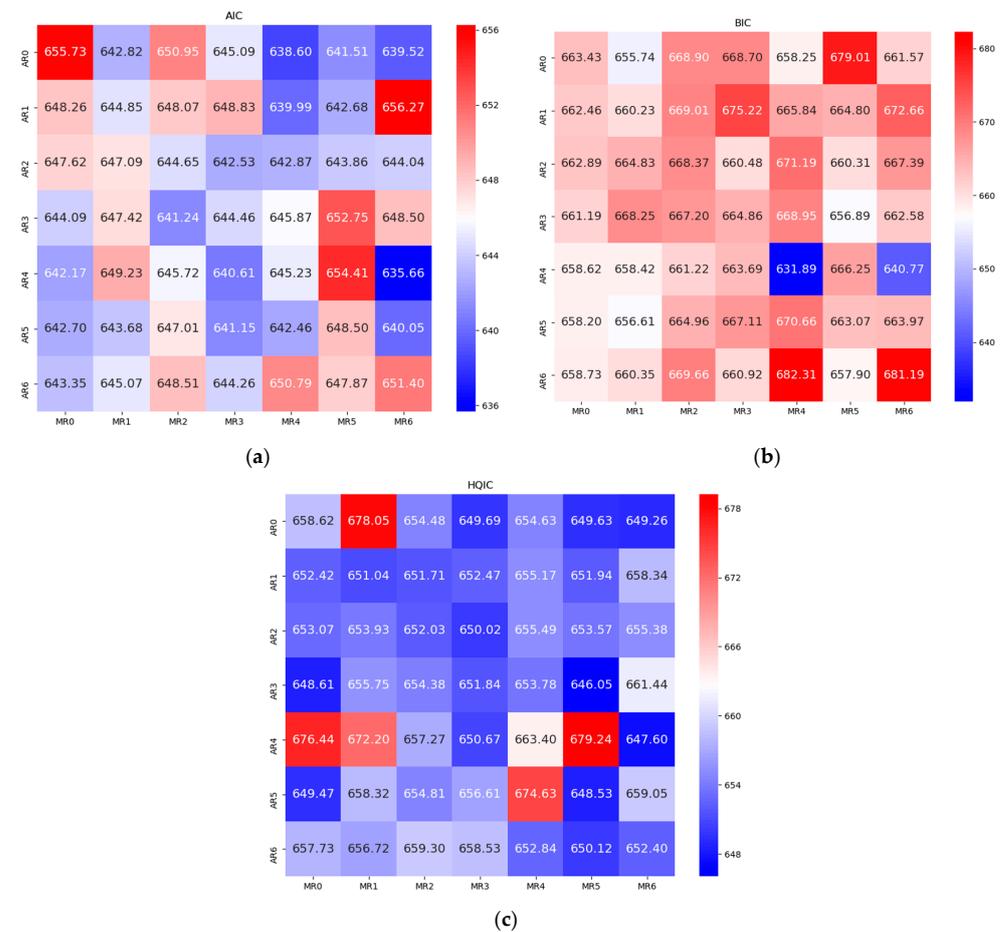


Figure 8. AIC, BIC, and HQIC values under different combinations. (a) AIC value, (b) BIC value, and (c) HQIC value.

Step 4: Perform ARIMA model checking and use residuals to test the model quality. The Durbin–Watson statistic can be used to test a model when the regression model contains an intercept term, explanatory variables are non-random, or the random disturbance term is a first-order linear autocorrelation [34]. Based on this, white noise was used to assess the residual sequence, that is, to determine whether the residual sequence autocorrelation function graph fell within the confidence interval. Figures 9 and 10 show the test results. The image shows that the residual sequence is almost entirely within the confidence interval, proving that the data sequence is white noise and that the ARIMA regression model is effective.

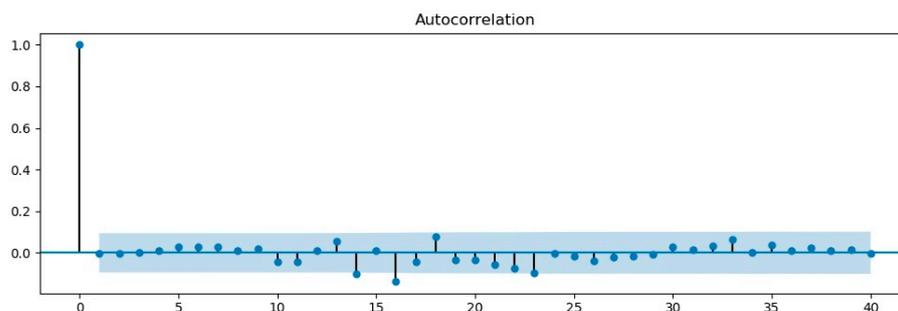


Figure 9. ACF diagram of the residual.

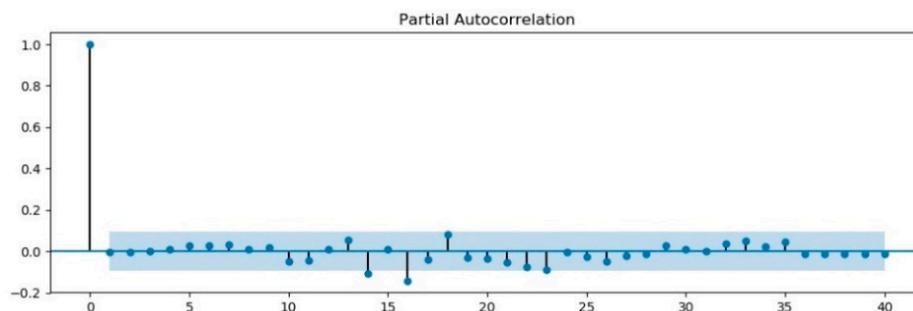


Figure 10. PACF diagram of the residual.

(2) BPNN model

In this hybrid model, the BPNN model describes the nonlinear relationship between the ambient temperature, humidity, equipment pressure, amount of bubble water sprayed, and residual CO₂ concentration predicted by ARIMA. Therefore, these parameters were taken as the input values, and the CO₂ concentration in a specific space coupled with space–time properties as the output values were used to train the network. The neural network had four layers: the input layer, output layer, and two hidden layers. The number of neurons in the input layer was equal to the number of model input parameters, that is, five, and the number of neurons in the output layer was one. The selection of the number of neurons in the hidden layer was obtained according to empirical formula 6, and the number of nodes in the hidden layer with the best fitting result was obtained by repeatedly testing the number of neurons within the value range.

$$H = \sqrt{M + N} + c, \tag{6}$$

where M and N represent the number of neurons in the input and output layers, respectively, and c is an integer in the range (1, 10). The comparison and analysis of each training result of the model show that when the number of neurons in the hidden layer is (7, 5), the training mean squared error reaches a minimum. Figure 11 shows the structure of the neural network used in this experiment.

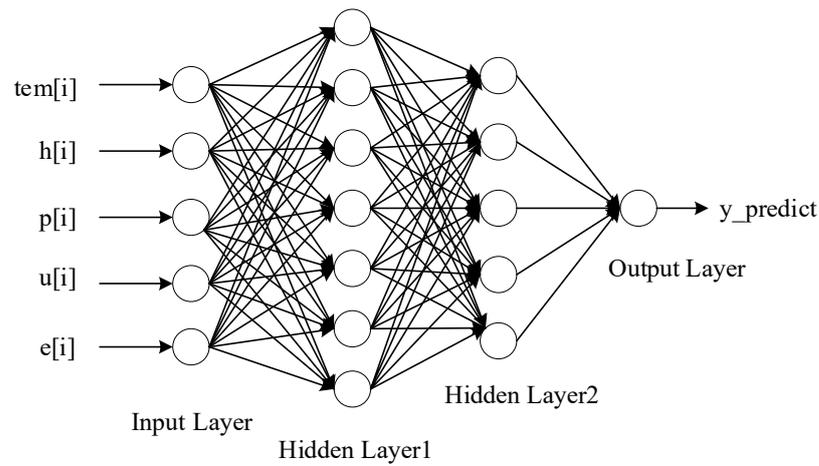


Figure 11. Architecture of the BPNN prediction model coupled with spatiotemporal characteristics.

The parameters for the model training are listed in Table 2. Figure 12 shows the variation curve of the root-mean-square error (RMSE) with the number of iterations during the learning process of the training and test datasets. As shown, when the number of iterations was 1000, the RMSE of the model was stable and reached the optimum value. The RMSEs of the training and test sets were 3.58×10^{-5} and 3.07×10^{-4} , respectively.

Table 2. Training parameter settings of ARIMA-BPNN.

Parameter	Value
Activation function	tan-sigmoid
Training function	traingdx
Loss function	L ₂ loss
Optimizer	SGD (stochastic gradient descent)
Learning rate	0.01
Iterations	1000

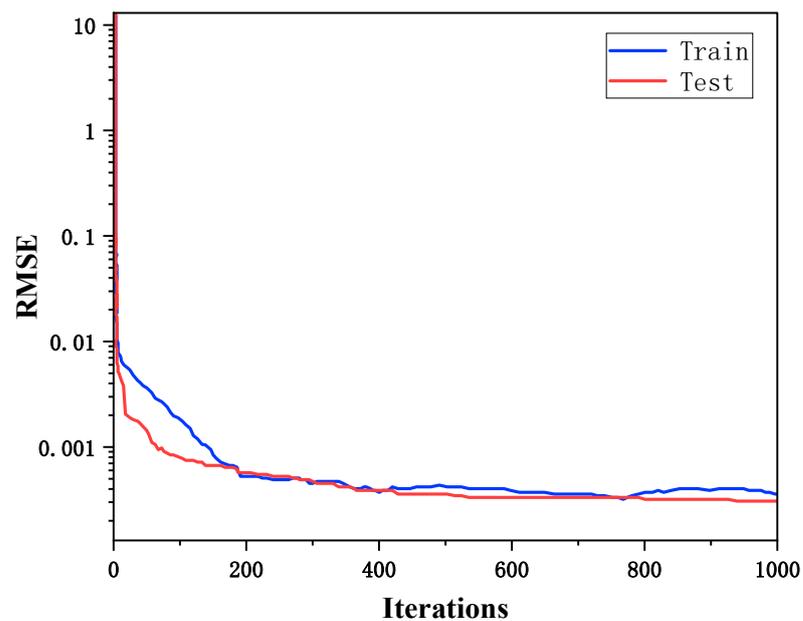


Figure 12. Prediction accuracy of the model with the number of iterations.

5.3. Model Evaluation Index

This study adopted three commonly used standard statistical measures, the *RMSE*, mean absolute error (*MAE*), and correlation coefficient (R^2), to evaluate the predictive ability of the combined model, which is the deviation between the prediction result and actual value. The specific calculation formulas are as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (7)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| MAE, \quad (8)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}. \quad (9)$$

In these three formulas, m is the quantity of sample data, y_i and \hat{y}_i are the measured and model-predicted values, respectively, and \bar{y} is the mean value of the sample data. The smaller the *RMSE* and *MAE* values of the three model evaluation indices, the higher is the accuracy of the prediction model and the better its prediction effect. R^2 represents the goodness of fit between the predicted results and measured values; the closer R^2 is to 1, the better is the interpretation of the independent variable to the dependent variable in the regression model [25].

5.4. CO₂ Release Prediction and Analysis in Micro/Nanobubble Water

5.4.1. Model Prediction Results and Analysis

To verify the effect of the combined prediction model ARIMA–BPNN, the prediction results of the single models, ARIMA and BPNN, were compared and analyzed with the prediction results of the ARIMA–BPNN model, and the results are presented in Table 3.

Table 3. Evaluation index values of each model.

Model	RMSE	MAE
BPNN	38.77	29.51
ARIMA	42.82	33.58
ARIMA-BPNN	17.48	9.31

Table 3 shows that there are obvious differences in the *RMSE* and *MAE* values of the single and combined prediction models, and the accuracy of the combined model is significantly higher than that of the single models. This indicates that the combined model with coupled spatiotemporal characteristics is more feasible for predicting CO₂ concentrations in the environment.

Figure 13 shows scatter plots of the prediction results of the three models for the test dataset. Through comparative experiments, it was concluded that the fitting result based on the ARIMA–BPNN hybrid model was the best, with the correlation coefficient R^2 between the measured and predicted values reaching 0.86. The fitting results of the other two models were not significantly different from each other. The combined prediction model compensates for the individual limitations of the single prediction models. Simultaneously, it can completely incorporate the effects of various environmental conditions on the release of CO₂ micro/nanobubbles, while maintaining the time-series prediction effect, allowing the model to better adapt to external factors.

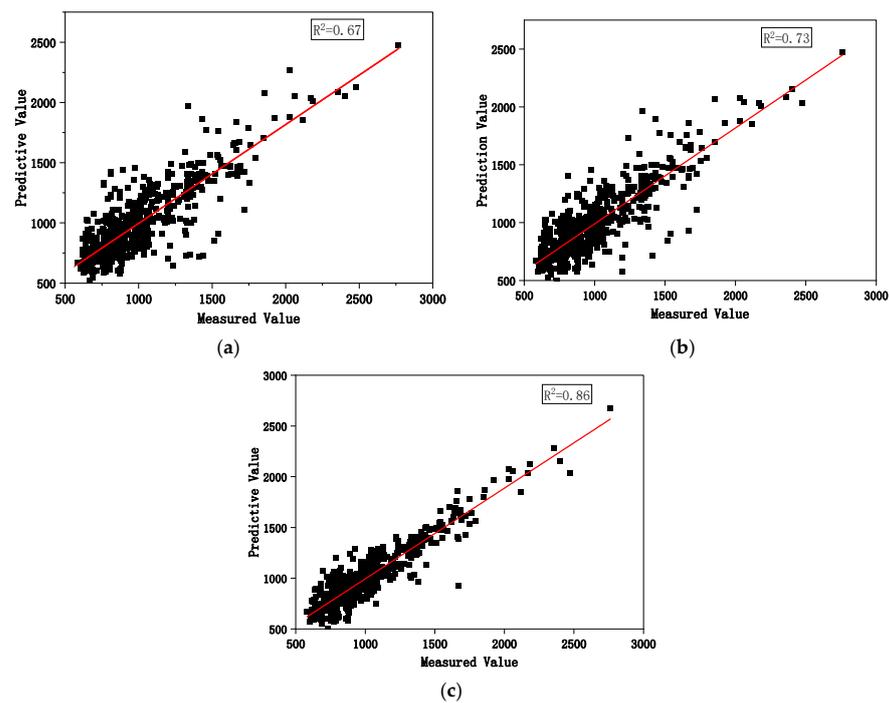


Figure 13. Scatter plot of predicted and measured values of three models. (a) BPNN model result, (b) ARIMA model result, and (c) ARIMA-BPNN model result.

5.4.2. Analysis on Factors Affecting CO₂ Release in Micro/Nanobubble Water

To analyze the importance and relationship among the factors affecting the release of CO₂ micro/nanobubbles, the mean influence value (MIV) algorithm [35] was used to evaluate the influence of the input neurons (environmental influence factor variables: temperature, humidity, spraying pressure, and spraying amount) of the hybrid prediction model on the output value of CO₂ concentration. The steps of the MIV influence factor algorithm are shown in Algorithm 3:

Algorithm 3 MIV

Input: k_i, P_i

Output: $|MIV_i|, y_{prediction}$

1: set adjustment rate k_i of MIV, $k_1 = 10\%$, $k_2 = 15\%$, $k_3 = 20\%$, $k_4 = 25\%$;

2: generate a new sample dataset P_{i-max}, P_{i-min} ;

3: use ARIMA-BPNN model to predict the new data set P_{i-max}, P_{i-min} , obtain the predicted results R_{i-max}, R_{i-min} ;

4: $IV_i = R_{i-max} - R_{i-min}$;

5: $|MIV_i| = \text{abs}(\text{mean}(IV_i))$.

The results of MIV influence factor algorithm measure the relative importance and weights of the input factors influencing the output variable. This algorithm can be used to optimize the input variables to reconstruct the training model, thereby reducing the dimensions of the network input parameters and further minimizing the complexity of the model.

The following adjusted rates were set in the experiment: $k_1 = 10\%$, $k_2 = 15\%$, $k_3 = 20\%$, and $k_4 = 25\%$. The $|MIV_i|$ of each variable under each adjusted rate was calculated using several experiments, and the results are shown in Figure 14. It shows that the weights of the different input parameters were essentially the same at different adjusted rates. Ambient temperature (AT) is the most important element that influences the release concentration of CO₂ micro/nanobubbles among the four input variables; it was presented with the highest weight in the prediction model. The spray pressure and amount of spray had a large weight proportion, whereas the weight of ambient humidity had the smallest proportion. The results demonstrate that the release of CO₂ micro/nanobubbles is influenced by various

factors, and the correlation between these factors should be considered while building the prediction model.

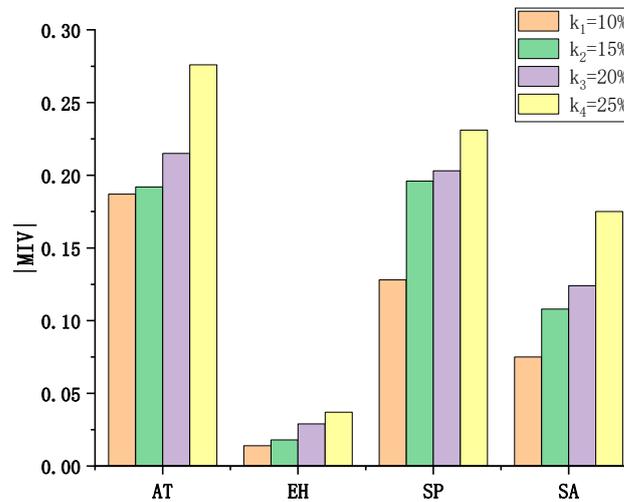


Figure 14. Change chart of $|MIV|$ for four impact factors.

Furthermore, the CO₂ concentration in the environment 2 hours after spraying CO₂ micro/nanobubble water was collected to investigate the effects of ambient temperature, humidity, spray pressure, spray amount, and other variables on the ambient CO₂ concentration. Figure 15 presents the results.

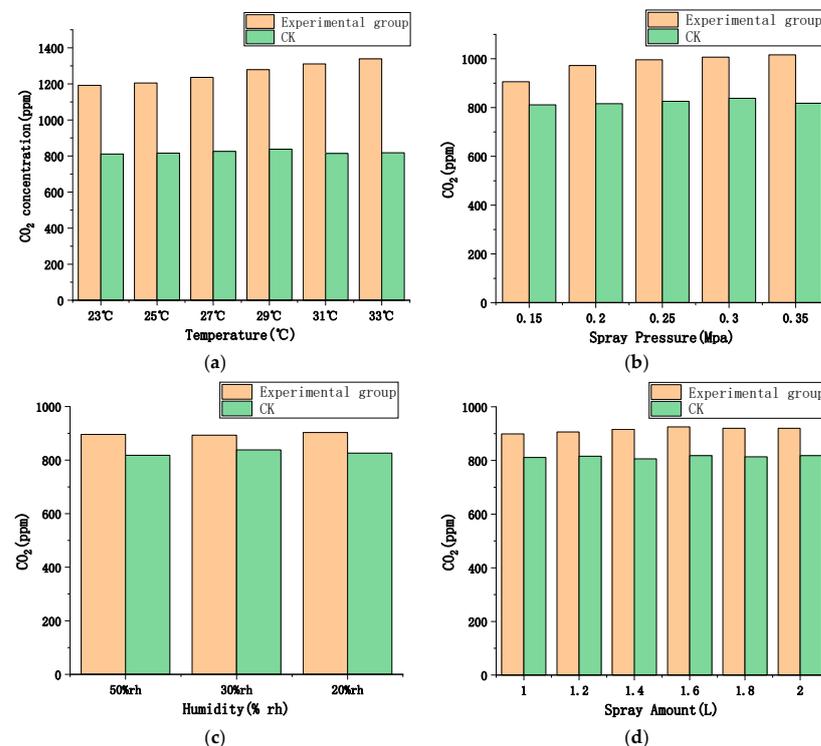


Figure 15. Histogram of the effect of different influencing factors on the concentration of CO₂ micro/nanobubbles. (a), (b), (c), and (d) represent the bar charts of CO₂ concentration in the experimental environment with the change of ambient temperature, spraying pressure, ambient humidity, and spraying amount, respectively.

Figure 15 shows that the ambient temperature, ambient humidity, spray pressure, and spray amount have an impact on the release of CO₂ micro/nanobubbles. Temperature had

the highest influence on the release of CO₂ micro/nanobubbles, and the CO₂ concentration increased with temperature. The influence of humidity and spray amount on the release of CO₂ was less than that of spray pressure. With an increase in spray pressure, the CO₂ concentration increased in the environment; however, the increasing trend was not obvious. These results are basically consistent with the results of the MIV algorithm. It indicates that the evaluation method for the influencing factors of CO₂ micro/nanobubble release based on the MIV algorithm is reliable.

6. Conclusions

In this study, the factors affecting the release of CO₂ from micro/nanobubble water were investigated and a hybrid prediction model coupled with spatiotemporal characteristics was proposed. The detailed experimental results are as follows:

- (1) Considering the linear and nonlinear properties of the gas release process, a hybrid prediction model based on the ARIMA-BPNN was constructed and compared to the prediction results of both the ARIMA and BPNN models. The results show that the fitting result based on the hybrid prediction model is the best, with R² reaching 0.86. The RMSE and MAE values are 17.48% and 9.31%, respectively. The ARIMA-BPNN model has good prediction accuracy and could accurately fit the complex mapping relationship between the influencing elements and CO₂ micro/nanobubble release concentration.
- (2) Based on the constructed hybrid model, the MIV algorithm was used to quantitatively analyze the influence weights of the input factors on the CO₂ concentration. The experimental results show that within the range of model input variables, ambient temperature has the highest weight in the prediction model as a key factor affecting the release of CO₂ micro/nanobubbles, followed by spray pressure and spray amount. The ambient humidity has the lowest weight with no significant effect.

In conclusion, the prediction model proposed in this paper provides guidance for the use of CO₂ micro/nanobubble water to enrich CO₂ for crops and provides a new research idea for the quantitative application of CO₂ micro/nanobubble water in agriculture. However, currently, the model has limitations in terms of the input influencing factors or other possible physical factors that affect the release of bubbles in micro/nanobubble water. Therefore, the influence of a higher release of CO₂ micro/nanobubbles should be explored in future research to further improve the prediction accuracy of the model. Simultaneously, the model was only used to predict the release concentration of CO₂ micro/nanobubbles and no relevant experiments were conducted on the applicability of the release characteristics of other gas sources.

Author Contributions: Data curation, B.W.; methodology, B.W.; software, B.W.; validation, B.W. and X.L.; formal analysis, B.W.; writing—original draft preparation, B.W.; writing—review and editing, S.T., Y.R. and B.W.; supervision, S.T. and Y.R.; project administration, S.T.; funding acquisition, S.T. and W.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of the Ministry of Science and Technology, Research on Field Planting Online Monitoring Technology and System Standards (2018YFF0213602).

Institutional Review Board Statement: Not applicable in this study.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to express our gratitude to all those who helped me in the writing of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, C.; Hu, Z.H.; Yu, L.F.; Chen, S.T.; Liu, X.M. Responses of photosynthetic characteristics and growth in rice and winter wheat to different elevated CO₂ concentrations. *Photosynthetica* **2020**, *58*, 1130–1140. [[CrossRef](#)]
2. Hussin, S.; Geissler, N.; El-Far, M.M.; Koyro, H. Effects of salinity and short-term elevated atmospheric CO₂ on the chemical equilibrium between CO₂ fixation and photosynthetic electron transport of *Stevia rebaudiana* Bertoni. *Plant Physiol. Biochem.* **2017**, *118*, 178–186. [[CrossRef](#)] [[PubMed](#)]
3. Temesgen, T.; Bui, T.T.; Han, M.; Kim, T.; Park, H. Micro and nanobubble technologies as a new horizon for water-treatment techniques: A review. *Adv. Colloid. Interface Sci.* **2017**, *246*, 40–51. [[CrossRef](#)] [[PubMed](#)]
4. Takahashi, M. ζ potential of microbubbles in aqueous solutions: Electrical properties of the gas-water interface. *J. Phys. Chem. B* **2005**, *109*, 21858–21864. [[CrossRef](#)] [[PubMed](#)]
5. Kulkarni, A.A.; Joshi, J.B. Bubble formation and bubble rise velocity in gas-liquid systems: A review. *Ind. Eng. Chem. Res.* **2005**, *44*, 5873–5931. [[CrossRef](#)]
6. Parkinson, L.; Sedev, R.; Fornasiero, D.; Ralston, J. The terminal rise velocity of 10–100 μm diameter bubbles in water. *J. Colloid. Interface Sci.* **2008**, *322*, 168–172. [[CrossRef](#)]
7. Zimmerman, W.B.; Tesaf, V.; Bandulasena, H.H. Towards energy efficient nanobubble generation with fluidic oscillation. *Curr. Opin. Colloid. Interface Sci.* **2011**, *16*, 350–356. [[CrossRef](#)]
8. Cerrón-Calle, G.A.; Magdaleno, A.L.; Graf, J.C.; Apul, O.G.; Garcia-Segura, S. Elucidating CO₂ nanobubble interfacial reactivity and impacts on water chemistry. *J. Colloid. Interface Sci.* **2022**, *607*, 720–728. [[CrossRef](#)]
9. Zhang, Y.; Yasutake, D.; Hidaka, K.; Kitano, M.; Okayasu, T. CFD analysis for evaluating and optimizing spatial distribution of CO₂ concentration in a strawberry greenhouse under different CO₂ enrichment methods. *Comput. Electron. Agric.* **2020**, *179*, 105811. [[CrossRef](#)]
10. Moon, T.; Choi, H.Y.; Jung, D.H.; Chang, S.H.; Son, J.E. Prediction of CO₂ Concentration via Long Short-Term Memory Using Environmental Factors in Greenhouses. *Hortic. Sci. Technol.* **2020**, *38*, 201–209. [[CrossRef](#)]
11. Hamrani, A.; Akbarzadeh, A.; Madramootoo, C.A. Machine learning for predicting greenhouse gas emissions from agricultural soils. *Sci. Total Environ.* **2020**, *741*, 140338. [[CrossRef](#)] [[PubMed](#)]
12. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine learning in agriculture: A comprehensive updated review. *Sensors* **2021**, *21*, 3758. [[CrossRef](#)] [[PubMed](#)]
13. Jha, G.K.; Sinha, K. Agricultural price forecasting using neural network model: An innovative information delivery system. *Agric. Econ. Res. Rev.* **2013**, *26*, 229–239. [[CrossRef](#)]
14. Zhou, T.; Wang, F.; Yang, Z. Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. *Water* **2017**, *9*, 781. [[CrossRef](#)]
15. Xiang, Y.; Gou, L.; He, L.; Xia, S.; Wang, W. A SVR-ANN combined model based on ensemble EMD for rainfall prediction. *Appl. Soft Comput.* **2018**, *73*, 874–883. [[CrossRef](#)]
16. Zou, P.; Yang, J.; Fu, J.; Liu, G.; Li, D. Artificial neural network and time series models for predicting soil salt and water content. *Agric. Water Manag.* **2010**, *97*, 2009–2019. [[CrossRef](#)]
17. Cheng, W.; Zhou, Y.; Guo, Y.; Hui, Z.; Cheng, W. Research on prediction method based on ARIMA-BP combination model. In Proceedings of the 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 18–20 October 2019; pp. 663–666.
18. Phan, K.K.T.; Truong, T.; Wang, Y.; Bhandari, B. Formation and Stability of Carbon Dioxide Nanobubbles for Potential Applications in Food Processing. *Food Eng. Rev.* **2021**, *13*, 3–14. [[CrossRef](#)]
19. Tomiyama, A.; Celata, G.P.; Hosokawa, S.; Yoshida, S. Terminal velocity of single bubbles in surface tension force dominant regime. *Int. J. Multiph. Flow* **2002**, *28*, 1497–1519. [[CrossRef](#)]
20. Yang, H.; Li, X.; Qiang, W.; Zhao, Y.; Zhang, W.; Tang, C. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. *Comput. Netw.* **2021**, *193*, 108102. [[CrossRef](#)]
21. Fan, D.; Sun, H.; Yao, J.; Zhang, K.; Yan, X.; Sun, Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy* **2021**, *220*, 119708. [[CrossRef](#)]
22. Wang, F.; Zou, Y.; Zhang, H.; Shi, H. House price prediction approach based on deep learning and ARIMA model. In Proceedings of the 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 19–20 October 2019; pp. 303–307.
23. Zhai, M.; Li, W.; Tie, P.; Wang, X.; Xie, T.; Ren, H.; Zhang, Z.; Song, W.; Quan, D.; Li, M.; et al. Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: A time series predictive analysis. *BMC Infect. Dis.* **2021**, *21*, 280. [[CrossRef](#)] [[PubMed](#)]
24. Wang, Z.; Ding, Y. Research on Signal-to-Noise Ratio in Order Selection of AR Model. *Acta Math. Sci. (Ser. A)* **2020**, *40*, 811–823.
25. Bierens, H.J. *Information Criteria and Model Selection*; Pennsylvania State University: State College, PA, USA, 2004.
26. Wang, Z.; Liu, S.; Feng, L.; Xu, Y. BNNmix: A new approach for predicting the mixture toxicity of multiple components based on the back-propagation neural network. *Sci. Total Environ.* **2020**, *738*, 140317. [[CrossRef](#)] [[PubMed](#)]
27. Jiang, B.; Liu, H.; Xing, Q.; Cai, J.; Zheng, X.; Li, L.; Liu, S.; Zheng, Z.; Xu, H.; Meng, L. Evaluating traditional empirical models and BPNN models in monitoring the concentrations of chlorophyll-A and total suspended particulate of eutrophic and turbid waters. *Water* **2021**, *13*, 650. [[CrossRef](#)]

28. Zhao, J.; Yang, D.; Wu, J.; Meng, X.; Li, X.; Wu, G.; Miao, Z.; Chu, R.; Yu, S. Prediction of temperature and CO concentration fields based on BPNN in low-temperature coal oxidation. *Thermochim. Acta* **2021**, *695*, 178820. [[CrossRef](#)]
29. Kumari, N.; Belwal, R. Hybridized approach of image segmentation in classification of fruit mango using BPNN and discriminant analyzer. *Multimed. Tools Appl.* **2021**, *80*, 4943–4973. [[CrossRef](#)]
30. Wang, J.; Fang, J.; Zhao, Y. Visual prediction of gas diffusion concentration based on regression analysis and BP neural network. *J. Eng.* **2019**, *2019*, 19–23. [[CrossRef](#)]
31. Liu, M.; Ding, L.; Bai, Y. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. *Energy Convers. Manag.* **2021**, *233*, 113917. [[CrossRef](#)]
32. Qi, Y.; Li, Q.; Karimian, H.; Liu, D. A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* **2019**, *664*, 1–10. [[CrossRef](#)]
33. Wang, W.; Du, Y.; Chau, K.; Chen, H.; Liu, C.; Ma, Q. A Comparison of BPNN, GMDH, and ARIMA for Monthly Rainfall Forecasting Based on Wavelet Packet Decomposition. *Water* **2021**, *13*, 2871. [[CrossRef](#)]
34. Berkhin, P. *A Survey of Clustering Data Mining Techniques*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 25–71.
35. Wu, T.; Zhan, J.; Yan, W. Research on influence factors of real estate price based on MIV-BP neural network test. *Math. Pract. Theory* **2015**, *18*, 45–52.