Research Article

A Fuzzy ARIMA Correction Model for Transport Volume Forecast

Yumeng Xie [ID], Peilin Zhang [ID], and Yanyi Chen

School of Transportation, Wuhan University of Technology, Wuhan 430063, China

Correspondence should be addressed to Peilin Zhang; plzhanghq@126.com

In recent years, few water transportation forecasting studies conduct relative to transportation forecasting. As a neglected area, the inland waterway volume prediction is an important indicator for investment management and government policymaking. Considering the time-series forecasting, some researchers try to narrow the predicted value interval. However, certain limitations detract from their popularity. For instance, if the prediction length is more than ten, the result would not be acceptable. Therefore, we propose a hybrid model that combines both of their unique properties' advantages to provide more accurate traffic volume forecasts. Also, the forecasting process will be more straightforward. The empirical results present the proposed model and improve the long-term predictive accuracy at waterway traffic volume.

1. Introduction

A long-term prediction usually hardly predicts more than ten periods and follows the time pass, and the accumulated error will make the result barely acceptable. On the contrary, how to combine different influence factors and historical time-series data is also an interesting part for researchers. In this case, some scholars propose Fuzzy Time Series (FTS), Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANNs), etc., to deal with this problem.

Thus, this research focused on keeping the long-term prediction result in a narrow interval value range. In other words, the study attempts to combine different methods for exploiting its strong points to increase forecasting accuracy rates. This work will benefit the long-term forecasting area. Usually, the inland waterway volume (VOL) has a long transfer period, and it also contributes to the inland waterway investment management and government policymaking.

2. Related Works

When the fuzzy prediction methods were first introduced to deal with the forecasting issue, the historical data were linguistic values [1–3]. Since then, the researchers have generally accepted the standard form of Fuzzy Time Series (FTS) [4]. After that, numerous research studies in the forecasting area work on the FTS method. Some of them forecasted the pollution concentrations [5]. Duru adopted a fuzzy integrated logical forecasting model for dry bulk shipping index forecasting [6]. Chang and Chen proposed the temperature prediction method based on fuzzy clustering and fuzzy rule interpolation techniques [7]. Chen and Dang mentioned a method to construct with variable spreads [8]. However, both of them used fuzzy rule techniques. Aladag and Basaran proposed an approach that used feedforward neural networks to gain the high-order fuzzy time data [9]. Huang proposed a fuzzy time-series forecasting, which used a multivariable heuristic function to improve forecasting results by integrating various univariate models [10]. Karnik and Mendel designed a type-2 fuzzy logic system to deal with the Mackey-Grass chaotic time-series forecasting problem [11]. Yu and Huang applied neural networks to fuzzy time series forecasting and proposed bivariate models to improve forecasting quality [12]. Wang and Hsu an improved fuzzy time series model for forecasting short-term time-series data [13]. Lau proposed a method to forecast energy consumption change based on the fuzzy logic method [14].

Another common approach for prediction is Autoregressive Integrated Moving Average (ARIMA) model, which was a typical time-series prediction method. Cullinane applied the Box–Jenkins approach to forecast the shipping market and discuss the different forecasting stages in shipping [15]. Fanoodi et al. predicted the blood platelet demands based on Artificial Neural Networks (ANNs) and ARIMA to reduce the supply chain's uncertainty [16]. Pavlyuk devoted to a multivariate model including ARIMA and Vector Autoregressive Moving Average (VARMA), Error Correction Vector Autoregressive Moving Average (EC-VARMA), Space-time ARMA (STARMA), and Multivariate Autoregressive Space State (MARSS) models to explain short-term traffic flow forecasting [17]. Li and Parson introduced a nonlinear approach in neural networks and evaluated its short and long period forecasting performance relative to ARMA models [18]. Cullinane et al. forecaste the Baltic Freight Index (BFI) and investigate the impact of a change in the BFI composition in 1993, although the results indicated that the modification did not significantly affect the behavior of the BFI [19]. Wong generated the short- and long-term predictions of the Baltic Dry Index (BDI) and used Fuzzy heuristic modeling with Grey System and ARIMA, and the results indicated that the ARIMA has a better performance in the long-term forecast [20].

However, the ARIMA prediction interval was too broad to be precise for the forecasting purpose. In this case, some researchers combine Fuzzy Time Series and ARIMA model to solve this issue. Duru et al. presented a bivariate long-term fuzzy inference system for time-series forecasting tasks in the freight market by combining FTS and ARIMA models [21]. Wong et al. proposed a traditional ARIMA model and the Fuzzy Time Series Method for forecasting the amount of Taiwan export [22]. Torbat et al. proposed a fuzzy autoregressive integrated moving average model as an improved ARIMA version to narrow the forecast interval [23]. Li and Hu raised three examples to test the Neuro-Fuzzy System ARIMA (NFS-ARIMA) model for nonlinear problems forecasting ability [24]. Tseng integrated the time-series ARIMA model and fuzzy regression model as a Fuzzy ARIMA (FARIMA) model to forecast NT dollars' exchange rate of US dollars [25]. Tseng and Tseng propose a Fuzzy Seasonal ARIMA (FSARIMA) forecasting model to predict the total production value of the Taiwan machinery industry [26].

Compared with the previous works, the contributions of this paper are as follows.

Through original multivariate information, transform into integrated information by constructing a fuzzy membership set. Therefore, Correction Fuzzy Time Series Autoregressive Integrated Moving Average (CFTSARIMA) method dramatically simplifies calculating and efficiently processing large amounts of data.

Generally, the FTSARIMA and FMLR models provide minimum widths' intervals for a given confidence degree. However, forecasting will significantly deviate from the real data in the long term. In this case, we propose a CFTSARIMA method, and the result could be acceptable. In equation (3), the h value level represents the value range

widths' influences. However, when the h value grows, the prediction interval's tendency is not changing, instead the space between Upper Control Limit (UCL) and Lower Control Limit (LCL) is increased. Therefore, it was necessary to recalculate a new center of the fuzzy numerical value a_i and the width c_i which was around the center fuzzy number.

Compared to railways and road deliveries, few research pieces focus on inland waterway transportations; however, precise long-term forecasting could benefit the economic aspect. The example demonstrated provides both the best- and worst-possible situations to decision-maker consideration.

3. Problem Statement

In this section, we first briefly review the coupled Fuzzy Clustering, Fuzzy Multiple Linear Regression method, and Fuzzy Time Series Autoregressive Integrated Moving Average method for long-term forecasting. However, the results diverge from observing data for solving the issue. We apply the Correction Fuzzy Time Series Autoregressive Integrated Moving Average (CFTSARIMA) method to solve the projection bias problems specifically. Table 1 summarizes the code of nomenclature for fuzzy ARIMA prediction and the correspondence between the symbol and the definition.

3.1. Fuzzy Clustering (See [7, 27]).
Different time-series groups have been chosen to participate in the prediction operation. For reducing the noise in the forecasting information, a fuzzy clustering method is useful.

To measure the cluster similarity, the Membership formula for different fuzzy sets is as follows.

When the predictors are positively correlated with the sample of eigenvalues, $r_{ij} = (x_{ij} - x_{j\min})/(x_{j\max} - x_{j\min})$.

When the predictors are negatively correlated with the sample of eigenvalues, $r_{ij} = (x_{i\max} - x_{ij})/(x_{i\max} - x_{i\min})$, in which x_{ij} are fuzzy sets, where $x_i, i = 1, 2, \ldots, j$, is the ith influence factor and j is the number of variable collected data sets. Let elements x_i and x_j belong to set r_{ij}. Then, the correlation coefficient r_{ij} lie in [0, 1].

ρ_i is the similarity between two fuzzy clusters y_{ij} and x_{ij} which are computed through

$$|\rho_i| = \frac{\sum_{j=1}^{n}(x_{ij} - \overline{x}_i)(y_{ij} - \overline{y}_i)}{\sqrt{\sum_{j=1}^{n}(x_{ij} - \overline{x}_i)^2 \sum_{j=1}^{n}(y_{ij} - \overline{y}_i)^2}}, \tag{1}$$

where $y_{ij}, i = 1, 2, \ldots, j$, is the ith predicted output, j is the number of data sets, And $\overline{y}_i$ and $\overline{x}_i$ are the average value of y_i and x_i, respectively.

In Table 2, the time series are divided into different groups, and Table 3 selects the high similarity data from fuzzy clusters and compares them to y.

3.2. Fuzzy Multiple Linear Regression (See [8, 28]).
The general form of a fuzzy multiple linear regression model can be expressed as follows:

TABLE 1: Code of nomenclature.

Symbol	Definition
r_{ij}	The correlation coefficient of x_{ij}, and value range is $[0, 1]$
x_{ij}	Fuzzy sets defined on x
y_{ij}	Fuzzy sets defined on y
ρ_i	The correlation coefficient between y_{ij} and x_{ij}
A_i	The fuzzy coefficient of the ith independent variable
S	The sum of individual spreads of the fuzzy parameters
H	The influences value used to adjust the value range of y_{ij}
a_j	The central value of the jth fuzzy coefficients
c_j	The spread value of the jth fuzzy coefficients
P	The order of autocorrelation
D	The number of differencing to achieve stationarity
Q	The order of the moving average
g_1 and g_2	Two clustering categories: group 1 and group 2
φ_i and θ_i	The parameters
ε_t	Error/disturbance terms
K	A constant number

TABLE 2: The values of x corresponding to fuzzy base clustering r_{ij}.

Series	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{ij}
x_{1j}	1	x_{12}	x_{13}	x_{14}	x_{15}	x_{1j}
x_{2j}	–	1	x_{23}	x_{24}	x_{25}	x_{2j}
x_{3j}	–	–	1	x_{34}	x_{35}	x_{3j}
x_{4j}	–	–	–	1	x_{45}	x_{4j}
x_{5j}	–	–	–	–	1	x_{5j}
x_{ij}	–	–	–	–	–	1

$$y_i = A_0 + A_1 x_1 + A_2 x_2 + \cdots + A_j x_j, \quad (2)$$

where y_i, $i = 1, 2, \ldots, j$, is the ith predicted output, j is the number of datasets, and x_j, $j = 0, 1, 2, \ldots, j$, is the jth variable of the collected data. The number of independent variables A_j is the fuzzy coefficient of the jth independent variable.

The criterion of minimizing the total vagueness S is defined as the sum of individual spreads of the fuzzy parameters of the model:

$$\text{minimize} \quad S = \sum_{j=1}^{k} c_j \left| x_{ij} \right|,$$

$$x_{ij} a_j + (1-h) c_j \left| x_{ij} \right| \geq y_i, \quad t = 1, 2, \ldots, k,$$

$$\text{Subject to} \quad x_{ij} a_j - (1-h) c_j \left| x_{ij} \right| \leq y_i, \quad t = 1, 2, \ldots, k,$$

$$c \geq 0,$$

$$Y(y_i) \geq h, \quad \text{for } i = 1, 2, \ldots, k. \quad (3)$$

TABLE 3: The values of y corresponding to fuzzy base clustering ρ_i.

	x_1	x_2	x_3	x_4	x_5	x_i
y	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_i

The membership value of each observation y_i is considering the condition that should be greater than an imposed threshold h, $h \in [0, 1]$. The choice of the h level value influences the widths c_j of the fuzzy parameters. This criterion simply expresses the fact that the fuzzy output of the model should "cover" all the data points $y_1, y_2, \ldots, y_k$ to a certain h level. a_j denotes a center of a fuzzy parameter, and c_j shows the fuzziness of its parameter. The fuzzy coefficients can be expressed as $A_1 = (a_1, c_1)$, $A_2 = (a_2, c_2)$, and $A_j = (a_j, c_j)$.

In the following, take $A_1, A_2, \ldots, A_j$ in equation (2), and the result is as follows:

$$\begin{cases} y_{r(\text{UCL})} = (a_1 + c_1)x_1 + (a_2 + c_2)x_2 + \cdots + (a_j + c_j)x_j, \\ y_{r(\text{LCL})} = (a_1 - c_1)x_1 + (a_2 - c_2)x_2 + \cdots + (a_j - c_j)x_j. \end{cases} \quad (4)$$

3.3. FTSARIMA (See [23, 25, 26, 29]).

The basis of this technique is the Autoregressive Integrated Moving Average ARIMA (p, d, q). p is the order of autocorrelation, d is the number of differencing to achieve stationarity, and q is the order of the moving average. The mathematic formula (5) is given as

$$y_t = K + X_{t-1} + \sum_{i=1}^{p} \varphi_i X_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \quad (5)$$

where y_t is a time series, and let y denote the differenced (stationarized) version of x, e.g., $y_t = X_t - X_{t-1}$. K is a constant, the random variables ε_t and ε_{t-i} are error/disturbance terms, and φ_i and θ_i present correlative parameters. The formula gives the mathematic representation of an ARMA (p, d, q) model. The index t refers to the number of nonfuzzy data used in constructing the model. Transfer the fuzzy regression parameters problem as a linear programming problem [30].

The fuzzy model interval is as

$$\begin{cases} y_{s(\text{UCL})} = K + y_{t-1} + (a_1 + c_1)y_1 + \cdots + (a_i + c_i)y_{t-i} - (a_i + c_i)\varepsilon_{t-1} - \cdots - (\alpha_{t+i} + c_{t+i})\varepsilon_{t-i}, \\ y_{s(\text{LCL})} = K + y_{t-1} + (a_1 - c_1)y_1 + \cdots + (a_i - c_i)y_{t-i} - (a_i - c_i)\varepsilon_{t-1} - \cdots - (\alpha_{t+i} - c_{t+i})\varepsilon_{t-i}, \end{cases} \quad (6)$$

where a_i is the center of the fuzzy numbers and c_i presents the extension value.

Despite the advantages of the fuzzy time series autoregressive integrated moving average models (FTSARIMA), forecasted intervals are too wide when data are fluctuations or outliers in the datasets. Besides, the results also include the maximum and minimum value of y_s. The result is an interval between $y_{s(UCL)}$ and $y_{s(LCL)}$, obtained by the FTSARIMA and a confidence degree of a.

3.4. Combined Model. As mentioned above, the weight factors could participate in the combined model and the

variance analysis, which presents the impact of different factors through data analysis. Variance decomposition here is used as a weight to help combine the FMLR and FTSARIMA results. The new upper bound is y_{ucl} and the new lower bound is y_{lcl}:

$$\begin{cases} y_{ucl} = \alpha y_{r(UCL)} + \beta y_{s(UCL)}, \\ y_{lcl} = \alpha y_{r(LCL)} + \beta y_{s(LCL)}, \end{cases} \tag{7}$$

where α is the FMLR's weight and β is the FTSARIMA's weight. Through equation (3),

$$\begin{aligned} \text{minimize} \quad & S = \sum_{i=1}^{k} \sum_{j=1}^{k} c' \left| x_{ij} \right|, \\ & x_t \alpha' + (1-h) c' \left| x_{ij} \right| \geq y_{ucl}, \quad i = 1, 2, \ldots, k; j = 1, 2, \ldots, k, \\ \text{Subject to} \quad & x_t \alpha' - (1-h) c' \left| x_{ij} \right| \leq y_{lcl}, \quad i = 1, 2, \ldots, k; j = 1, 2, \ldots, k, \\ & c' \geq 0, \end{aligned} \tag{8}$$

where $\alpha' = (\alpha'_1, \alpha'_2, \ldots, \alpha'_j)$ and $c' = (c'_1, c'_2, \ldots, c'_j)$ are vectors of unknown variables. An interval with new width is obtained by the FTSARIMA and FMLR. This combined model considers the history data and other influence factors. The results of this model for long-term forecasting are more accurate than the others.

4. The Empirical Results and Validation

In general, the Mississippi shipping volume y_i appears to be influenced mainly by the local urbanization rate x_{i1}, gross domestic product x_{i2}, rate of industrialization x_{i3}, proportion of railway transportation x_{i4}, proportion of road transportation x_{i5}, proportion of waterway transportation x_{i6}, and rate of agriculture x_{i7}.

The information used consists of 47 annual observations of Mississippi River shipping volume from 1970 to 2016, 30 observations are initially used to formulate the model, and the last 17 observations are used to verify the results (see Supplementary Material for data analysis).

The research step is shown in Figure 1.

From Figure 1, there were four steps to process the FTSARIMA correction model, and the details are as follows.

4.1. U.S. Mississippi Inland Waterway Influencing Factors' Fuzzy Clustering. The purpose of using the fuzzy clustering method is to reduce the computational complexity and disturbance term, and a simple algorithm is introduced. Then, an equation is proposed to generate fuzzy membership matrices, which are the basis for the final clustering results. Finally, the closer the relationship between the sample and the factor, the greater the influence of the predictive factor on the phenomenon. Therefore, we can consider removing some effect factors.

The similarity of fuzzy clusters is computed through equation (1).

Let x_{ij} be fuzzy sets defined on x_i and x_j with membership functions r_{ij}. Let an element $x_i \in X$ and $x_j \in X$ belong to set x_{ij}.

Four different groups are selected and divided from the similarity coefficient which is greater than or equal to 0.9 (see Table 4).

Group 1: x_1

Group 2: x_2, x_3, and x_7

Group 3: x_4 and x_5

Group 4: x_6

Through the analysis of pertinence between y and x_1, specially x_4, x_5, and x_6 are less correlated with the y set (see Table 5). Therefore, group1 and group2 are retained to rearrange as time series g_1 and g_2.

4.2. Fuzzy Multiple Linear Regression. This fuzzy multiple linear regression model (FMLR) is applied to the Mississippi River's shipping volume. The input data are obtained from the fuzzy clusters and the output data (see Figure 2). In the column x_i, two clustering categories are indicated: group 1 and group 2. From these data, the fuzzy linear system is $y_i = A_0 + A_1 g_1 + A_2 g_2$ and the fitting model for the data is given.

The results of fuzzy parameters A_1 and A_2 are given as ($a_1 = 0.24, c_1 = 0; a_2 = 3.44, c_2 = 0.91$), where $h = 0$, a_j denotes a center of a fuzzy parameter, and c_j shows the fuzziness of its parameter:

$$y_r = 2.23 + <0.24, 0.00> g_1 + <3.44, 0.91> g_2. \tag{9}$$

In this case, the fuzziness of A_0 is 2.23. A_1 and A_2 are positive values and rely on the correlations variables g_1 and
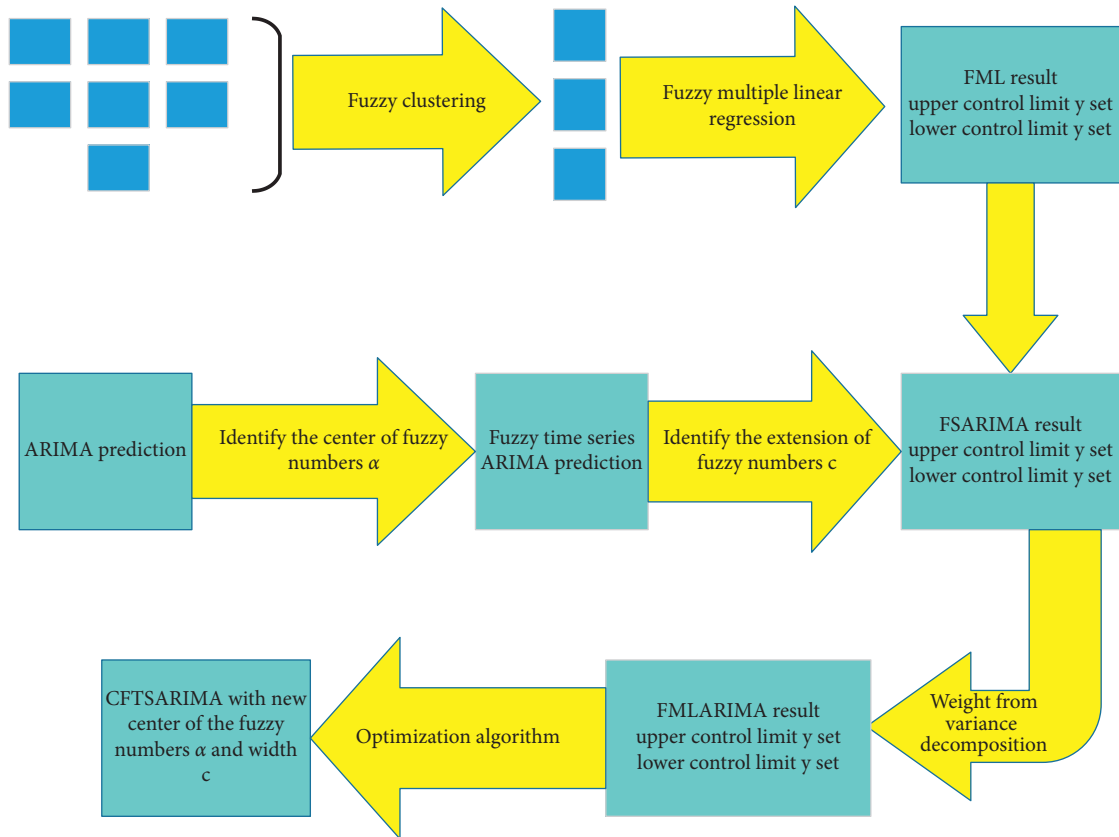
FIGURE 1: CFTSARIMA training procedure.

TABLE 4: Values of x corresponding to fuzzy base clustering r_{ij}.

Series	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{ij}
x_{1j}	1	0.77	0.73	0.19	0.13	0.16
x_{2j}	–	1	0.97	0.27	0.22	0.11
x_{3j}	–	–	1	0.31	0.27	0.1
x_{4j}	–	–	–	1	0.94	0.13
x_{5j}	–	–	–	–	1	0.22
x_{ij}	–	–	–	–	–	1

TABLE 5: Values of y corresponding to fuzzy base clustering ρ_i.

	x_1	x_2	x_3	x_4	x_5	x_i
y	0.69	0.88	0.86	0.28	0.23	0.15

FIGURE 2: Results of fuzzy linear regression (30 historical data).

g_2. The conventional regression model's confidential interval seems to estimate the upper and lower observation errors' limits. Setting $(a_1, a_2) = (0.24, 3.44)$ and $h = 0$. The following linear interval model has shown its result in Figure 2.

Following the FMLR 17 predictions, see Figure 3, the actual value is lower than estimated, which means the FMLR method for shipping volume forecasting still needs improvement.
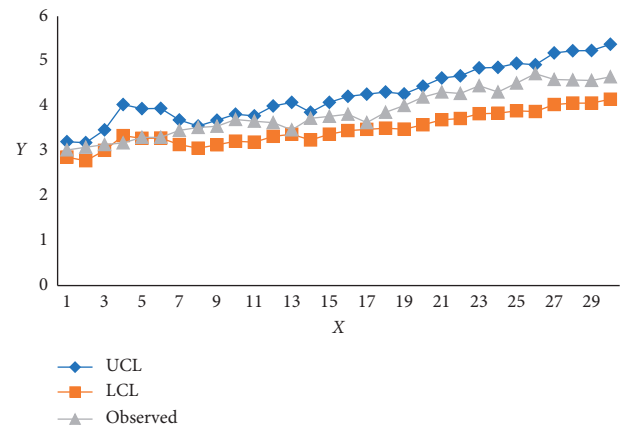
4.3. Fuzzy Time Series ARIMA Building the ARIMA Model.
The time-series data were preprocessed using the first-order regular differencing to stabilize the variance and remove the growth trend. The derived model is ARIMA (1, 1, 1), and the equation is

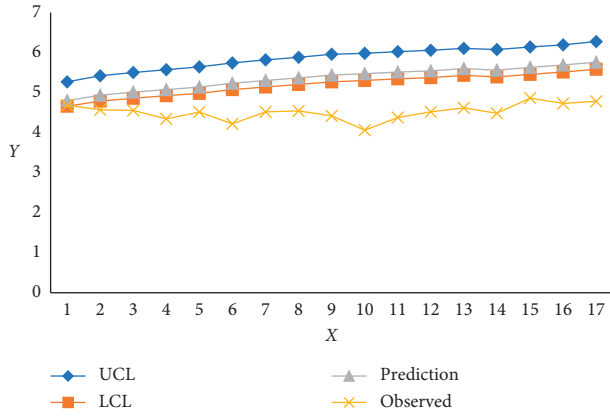$$y = 0.058 + y_{t-1} + 0.536 \cdot \left[y_{t-1} - y_{t-2} \right] - \varepsilon_{t-1}. \tag{10}$$

FIGURE 3: The result of the estimated value of the linear regression model (17 predicted value).

After determining the minimal fuzziness equation (3), the FTSARIMA model is

$$y = A_0 + A_{t-1} + (a_1, c_1) \cdot [y_{t-1} - y_{t-2}] - (a_2, c_2)\varepsilon_{t-1}. \tag{11}$$

Setting $(a_1, a_2) = (0.536, 1)$ and $h = 0$, the linear interval model is obtained, and the results are demonstrated in Figure 4.

The actual values locate under the prediction interval value set (see Figure 5) below even the most pessimistic forecasts. To improve the accuracy of model prediction in the long term, the influence of the other factors x_i is necessary to be discussed.

4.4. Correction Fuzzy Time Series ARIMA Model. From FMLR and FTSARIMA, two different estimated equations (10) and (13) are obtained. To restructure y_{ucl} and y_{lcl}, the variance decomposition is introduced to assign a weight to them.

Summing up the results of the variance decomposition, the 10th period (see Table 6) could be considered as the weight value that 0.88 belong to y_s and (1–0.88), is the allocation to y_r. According to the FMLR model and FTSARIMA method, there are two different upper control limit (UCL) and lower control limit (LCL) lines. Based on the table weights 0.88 and 0.12, respectively, the new series are as follows:

$$\begin{cases} y_{LCL} = 0.88y_{s(LCL)} + 0.12y_{r(LCL)}, \\ y_{UCL} = 0.88y_{s(UCL)} + 0.12y_{r(UCL)}, \end{cases} \tag{12}$$

in which y_s is the result of the FTSARIMA model and y_r are the FMLR's model result. From equation (13), we can obtain equation (14), and Figure 6 reveals the tendency:

$$y_s = 0.058 + y_{t-1} + <0.536,$$
$$0.63> [y_{t-1} - y_{t-2}] - <1, \tag{13}$$
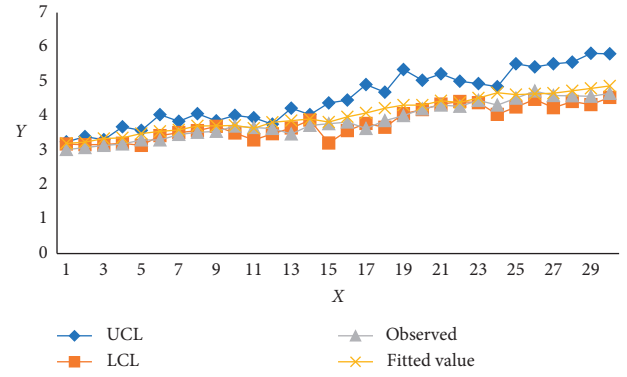$$1.57> \varepsilon_{t-1}.$$

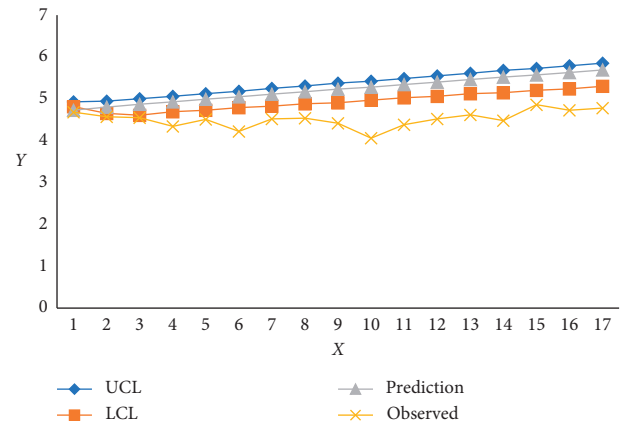FIGURE 4: Results of fuzzy ARIMA (30 historical data).

FIGURE 5: The result of the estimated value of the fuzzy ARIMA model (17 predicted value).

TABLE 6: Variance decomposition of VOL.

Period	S.E.	VOL	g_1	g_2
10	0.25	88.42	0.13	11.45

By substituting $(a_1, a_2) = (0.536, 1)$, y_{ucl}, and y_{lcl} into equation (9), the new c value has been determined, and the following linear interval model is given in equation (14) and plotted in Figure 7:

$$y = 0.058 + y_{t-1} + <0.536,$$
$$0.86> \cdot [y_{t-1} - y_{t-2}] - <1, \tag{14}$$
$$3.68> \varepsilon_{t-1}.$$

Usually, three metrics are used to evaluate prediction intervals: coverage rate, calibration, and sharpness [31, 32]. The coverage refers to the statistical consistency between the forecasts and the observations, and it measures how many observations are inside the prediction interval. The properties of sharpness and resolution refer to the concentration of the predictive distribution, or how wide and variable are the intervals, and refer uniquely to the forecasts [33].
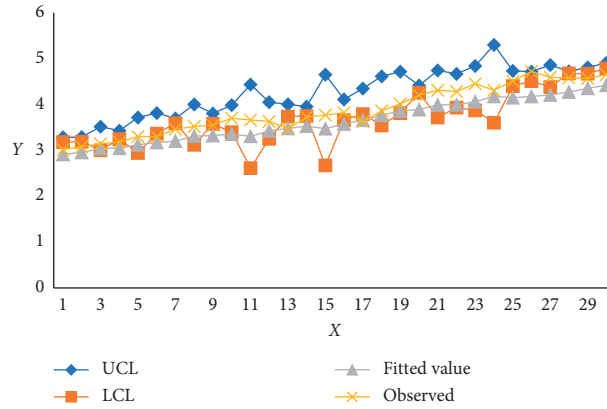
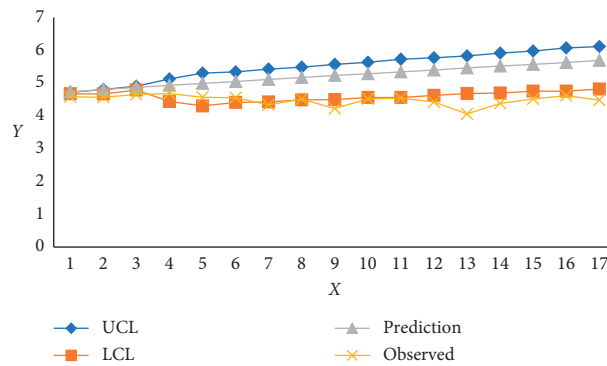Figure 6: The FTSARIMA and FMLR model (30 historical data).

Figure 7: The estimated value of the CFTSARIMA model (17 predicted value).

Compared to FMLR, FTSARIMA, and CFTSARIMA with 17 length predictions, only the CFTSARIMA conclusion has covered half of the observed values, and other prediction intervals are much higher than observed values. On the contrary, CFTSARIMA's prediction intervals are wide-ranging than FMLR and FTSARIMA, and the results are narrow than the traditional ARIMA method.

5. Results and Discussion

Theses examples' empirical results expose that the possible interval of the fuzzy ARIMA is narrower than the interval of classical ARIMA.

The fuzzy ARIMA usually is indicated as formulating the model. The output is fuzziness, leading to the assumption of white noise; in this case, FTSARIMA requires fewer observations than ARIMA.

In Table 7, UCL and LCL denote the upper control limit and the lower control limit and PV indicates the predicted values from different forecasting methods.

The fuzzy linear regression (FMLR) model is based on the probability distribution of statistics. The relationship between them can only manage the input and output information; therefore, massive information is required. This method is conducive to observing information with trend

growth, and the short-term prediction accuracy is only acceptable. On the contrary, long-term forecasts still need to be advanced. Table 7 is compared with four different results. The FTSARIMA is learned from the Fuzzy Linear Regression method and the ARIMA model; compared to both of them, the prediction interval has significantly narrowed. A four- or five-period forecasting result can be acceptable. If the prediction period is more than five, the actual value will be under the prediction interval; neglected other influence factors may cause this.

According to Figure 7 and Table 7, CFTSARIMA showed that the prediction interval contains the most observed value than FMLR and FTSARIMA. Although this method has also enlarged the range of forecasting value, the LCL line keeps a steady trend, and the prediction interval is still narrow than others.

To compare the proposed method's performance with the existing techniques, we apply the proposed approach to long-time forecasting by using 30 observations and 7 influence indexes. Table 7 shows a comparison of the upper control limit (UCL), lower control limit (LCL), and predicted values (PV) from different methods. The proposed method comes with prediction intervals that have significantly narrowed, and the predicted values are more precise than the existing methods. In other words, the proposed

TABLE 7: Forecasted production value of the traffic volume by using the fuzzy linear regression model (FMLR), fuzzy time series ARIMA model (FTSARIMA), and correction fuzzy time series ARIMA model (CFTSARIMA).

	FMLR			ARIMA			FTSARIMA			CFTSARIMA			Observed value
	UCL	LCL	PV	UCL	LCL	PV	UCL	LCL	PV	UCL	LCL	PV	
1	5.27	4.66	4.80	4.96	4.51	4.73	4.93	4.81	4.73	4.72	4.68	4.27	4.68
2	5.42	4.78	4.93	5.06	4.55	4.80	4.95	4.65	4.80	4.81	4.67	4.35	4.57
3	5.50	4.86	5.01	5.14	4.60	4.87	5.00	4.61	4.87	4.90	4.79	4.42	4.55
4	5.57	4.92	5.08	5.2	4.66	4.93	5.06	4.69	4.93	5.13	4.44	4.47	4.34
5	5.64	4.98	5.14	5.26	4.72	4.99	5.12	4.73	4.99	5.31	4.31	4.53	4.51
6	5.74	5.07	5.23	5.32	4.78	5.05	5.18	4.79	5.05	5.34	4.41	4.59	4.22
7	5.81	5.14	5.30	5.38	4.84	5.11	5.25	4.82	5.11	5.43	4.43	4.65	4.52
8	5.88	5.20	5.37	5.44	4.9	5.17	5.31	4.88	5.17	5.49	4.49	4.70	4.54
9	5.96	5.27	5.44	5.50	4.95	5.23	5.38	4.91	5.23	5.57	4.50	4.76	4.42
10	5.98	5.30	5.47	5.56	5.01	5.28	5.42	4.97	5.28	5.63	4.56	4.80	4.06
11	6.02	5.34	5.51	5.61	5.07	5.34	5.48	5.03	5.34	5.72	4.56	4.85	4.38
12	6.05	5.37	5.54	5.67	5.13	5.40	5.55	5.06	5.40	5.77	4.62	4.90	4.52
13	6.10	5.43	5.60	5.73	5.19	5.46	5.61	5.12	5.46	5.83	4.68	4.95	4.62
14	6.07	5.39	5.56	5.79	5.24	5.52	5.68	5.14	5.52	5.92	4.70	4.98	4.48
15	6.14	5.46	5.63	5.85	5.30	5.57	5.73	5.21	5.57	5.98	4.76	5.03	4.86
16	6.19	5.51	5.69	5.90	5.36	5.63	5.79	5.24	5.63	6.07	4.76	5.08	4.73
17	6.27	5.58	5.76	5.96	5.42	5.69	5.85	5.30	5.69	6.11	4.82	5.14	4.78

method produces better forecasting results than the strategies presented in [8, 23, 25, 26, 28, 29] for long-time forecasting.

6. Conclusions

In this study, the primary objective is to find an accurate prediction method for long-term waterway traffic volume prediction, and the proposed algorithm consists of four parts.

In the first part, a fuzzy cluster reduces the algorithm design difficulty, and nonessential double-counting raises the counting yield. After cluster, seven influence factors are classified into 4 groups, and only two of them satisfy the criteria.

According to the fuzzy clustering, two group influence factors are converted to two different time series sets. A fuzzy multiple linear regression is used to estimate the upper bound $y_{r(\text{UCL})}$ and lower bound $y_{r(\text{LCL})}$.

In the third part, a typically fuzzy time series ARIMA method is adopted to define the center of fuzzy number a_i and the width around the center of the fuzzy number c_i. After that, FTSARIMA prediction intervals of $y_{s(\text{UCL})}$ and $y_{s(\text{LCL})}$ could be obtained.

In the final part, by utilizing the variance decomposition to identify the weight coefficient, then combine them into the FMLR interval value and FTSARIMA interval value to obtain a new result. Repetition of the FTSARIMA is carried out, using new y_{LCL} and y_{UCL}, to obtain the CFTSARIMA model. When the influence factors participate in forecasting processing, a long-term prediction becomes possible.

For future work, the prediction could improve in the following aspects:

The fuzzy clustering still needs to increase the method's robustness to nonstationarity and concept-drift in this case, and incremental learning may be helpful.

Although most of the actual observations in the long-term prediction have entered the prediction range, there is still a gap compared with the accurate prediction. Consider combining the Markov chain in fuzzy clustering to improve accuracy.

Supposing the influence factors have nonlinearity by employing the adaptive filtering algorithm combined with the fuzzy method to forecast the chaotic time series may solve this issue.

Data Availability

The data used to support this findings of the study are included within the supplementary information file.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of the paper.

Supplementary Materials

In the Supplementary Materials, most data came from "WATERBORNE COMMERCE OF THE United States" and the first line was given the timeline, from the year 1970 to 2016. The first part from lines 2 to 10 refers to the Mississippi River (river system), coal, petro and petro products, chemicals, crude materials, manufactured goods, food and farm, and other. The second part from lines 11 to 19 refers to the Mississippi River (mainstream), coal, petro

and petro products, chemicals, crude materials, manufactured goods, food and farm, and other. Both the first and second part unit of measurement is one hundred million tons. The third part from lines 20 to 26 refers to population (100 million), urban population (100 million), urbanization rate (percentage), GDP (USD 1 billion), per capita GDP (USD 10 million), US manufacturing ($1 billion), and US rate of industrialization (percentage). (*Supplementary Materials*)

References

[1] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series-Part I," *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1–9, 1993.

[2] Q. Song and B. S. Chissom, "Fuzzy time series and its models," *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1–9, 1993.

[3] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series-part II," *Fuzzy Sets and Systems*, vol. 62, no. 1, pp. 1–8, 1994.

[4] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*, vol. 81, no. 3, pp. 311–319, 1996.

[5] D. Domańska and M. Wojtylak, "Application of fuzzy time series models for forecasting pollution concentrations," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7673–7679, 2012.

[6] O. Duru, "A fuzzy integrated logical forecasting model for dry bulk shipping index forecasting: an improved fuzzy time series approach," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5372–5380, 2010.

[7] Y. C. Chang and S. M. Chen, "Temperature prediction based on fuzzy clustering and fuzzy rules interpolation techniques," in *Conference Proceedings-IEEE International Conference on Systems, Man and Cybernetics*, pp. 3444–3449, San Antonio, TX, USA, October 2009.

[8] S.-P. Chen and J.-F. Dang, "A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy," *Information Sciences*, vol. 178, no. 20, pp. 3973–3988, 2008.

[9] C. H. Aladag, M. A. Basaran, E. Egrioglu, U. Yolcu, and V. R. Uslu, "Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4228–4231, 2009.

[10] K. H. Huarng, T. H. K. Yu, and Y. W. Hsu, "A multivariate heuristic model for fuzzy time-series forecasting," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 4, pp. 836–846, 2007.

[11] N. N. Karnik and J. M. Mendel, "Applications of type-2 fuzzy logic systems to forecasting of time-series," *Information Sciences*, vol. 120, no. 1–4, pp. 89–111, 1999.

[12] T. Yu and K. Huang, "A bivariate fuzzy time series model to forecast the TAIEX," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2945–2952, 2008.

[13] C. Wang and L. Hsu, "Constructing and applying an improved fuzzy time series model: taking the tourism industry for example," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2732–2738, 2008.

[14] H. C. W. Lau, E. N. M. Cheng, C. K. M. Lee, and G. T. S. Ho, "A fuzzy logic approach to forecast energy consumption change in a manufacturing system," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1813–1824, 2008.

[15] K. Cullinane, "A short-term adaptive forecasting model for biffex speculation: a box-jenkins approach," *Maritime Transport*, vol. 19, pp. 74–97, 2002.

[16] B. Fanoodi, B. Malmir, and F. F. Jahantigh, "Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models," *Computers in Biology and Medicine*, vol. 113, p. 103415, 2019.

[17] D. Pavlyuk, "Short-term traffic forecasting using multivariate autoregressive models," *Procedia Engineering*, vol. 178, no. Complete, pp. 57–66, 2017.

[18] J. Li and M. G. Parsons, "Forecasting tanker freight rate using neural networks," *Maritime Policy & Management*, vol. 24, no. 1, pp. 9–30, 1997.

[19] K. P. B. Cullinane, K. J. Mason, and M. Cape, "A comparison of models for forecasting the baltic freight index: box-jenkins revisited," *International Journal of Maritime Economics*, vol. 1, no. 2, pp. 15–39, 1999.

[20] H.-L. Wong, "BDI forecasting based on fuzzy set theory, grey system and ARIMA," *Modern Advances in Applied Intelligence*, vol. 8482, pp. 140–149, 2014.

[21] O. Duru, E. Bulut, and S. Yoshid, "Bivariate long term fuzzy time series forecasting of dry cargo freight rates," *The Asian Journal of Shipping and Logistics*, vol. 26, no. 2, pp. 205–223, 2010.

[22] H.-L. Wong, Y.-H. Tu, and C.-C. Wang, "Application of fuzzy time series models for forecasting the amount of Taiwan export," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1465–1470, 2010.

[23] S. Torbat, M. Khashei, and M. Bijari, "A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity markets," *Economic Analysis and Policy*, vol. 58, no. 2, pp. 22–31, 2018.

[24] C. Li and J.-W. Hu, "A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 2, pp. 295–308, 2012.

[25] F. M. Tseng, "Fuzzy ARIMA model for forecasting the foreign exchange market," *Fuzzy Sets and Systems*, vol. 118, no. 1, pp. 9–19, 2001.

[26] F.-M. Tseng and G.-H. Tzeng, "A fuzzy seasonal ARIMA model for forecasting," *Fuzzy Sets and Systems*, vol. 126, no. 3, pp. 367–376, 2002.

[27] A. Celikyilmaz and I. Burhan Turksen, "Enhanced fuzzy system models with improved fuzzy clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 3, pp. 779–794, 2008.

[28] P. Pandit, P. Dey, and K. N. Krishnamurthy, "Comparative assessment of multiple linear regression and fuzzy linear regression models," *SN Computer Science*, vol. 2, no. 2, pp. 1–8, 2021.

[29] A. Sahed, M. Mekidiche, and H. Kahoui, "Fuzzy auto-regressive integrated moving average (FARIMA) model for forecasting the gold prices," *Journal of Smart Economic Growth*, vol. 5, no. 1, pp. 1–13, 2020.

[30] H. Ishibuchi and H. Tanaka, "Interval regression analysis based on mixed 0-1 integer programming problem," *Journal of Japan Industrial Management Association*, vol. 40, no. 5, pp. 312–319, 1988.

[31] P. Pinson, "Properties of quantile and interval forecasts of wind generation and their evaluation," in *Proceedings of the European Wind Energy Conference and Exhibition 2006, EWEC*, Athens, Greece, March 2006.

[32] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.

[33] P. C. De Lima Silva, H. J. Sadaei, R. Ballini, and F. G. Guimaraes, "Probabilistic forecasting with fuzzy time series," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 8, pp. 1771–1784, 2020.