



Application of Acoustic Tomographic Data in Short-Term Forecasting of Streamflow Using Combinatorial GMDH Algorithm (CGA)

Y. Olfatmiri¹, E. Jabbari¹, M. Bahreinimotlagh^{2*}, H. Alizadeh¹, A. H. Hasanabadi¹

¹Department of Civil Engineering, Iran University of Science and Technology, Tehran, Iran

²Department of Water Resources Studies and Research, Water Research Institute, Tehran, Iran

Review History:

Received: Apr. 11, 2021

Revised: Jul. 13, 2021

Accepted: Aug. 02, 2021

Available Online: Sep. 16, 2021

Keywords:

GMDH Algorithm

Short-Term Forecasting

Streamflow Forecasting

FAT

ABSTRACT: Short-term forecasting of streamflow is one of the most important goals in water resources management and flood control. However, one of the problems that researchers always face in this type of prediction is the Lack of an accurate and high-resolution database. The Fluvial Acoustic Tomography (FAT) is an innovative technology that acquires streamflow data. Therefore, by using the data collected from this technology with a suitable forecast model, accurate short-term streamflow forecasting can be achieved. In this research, the effect of FAT data on short-term streamflow forecasting by the Combinatorial GMDH Algorithm (CGA) has been investigated and compared with one obtained from the Rating Curve method. The k-fold cross-validation criterion has been used to prevent over-fitting. The results showed that the FAT data increases the accuracy of short-term forecasting. As an example, the Nash-Sutcliffe coefficient (ENS) for the 1, 6, 12, 24, 48, and 72 hours forecast horizons were 0.98, 0.96, 0.94, 0.88, 0.73, and 0.54, respectively. While these values for the Rating Curve ones were 0.97, 0.84, 0.61, 0.27, 0.12, and 0.11, respectively.

1- Introduction

Predicting the parameters of the hydrological cycle is one of the most important factors for the successful management of water resources. One of these important parameters is streamflow. Streamflow Forecasting can help engineers in water supply, flood warning systems, and reservoir management [1]. Researchers have always sought to forecast streamflow. For this purpose, they have tried to estimate the streamflow using different hydrological models [2].

To forecast the streamflow of rivers, Hydrological models are generally divided into two categories: physical models and data-driven models. The complexity of the relationship between environmental variables and flow rate limits the use of physical models [3]. Also, as an advantage of data-driven models, these models require fewer input variables and can be developed using existing data records [4].

The main purpose of this study is to forecast the short-term streamflow of the Gono River in Hiroshima province, Japan, using data collected by the Fluvial Acoustic Tomography (FAT) method and compared with the Rating-Curves method. In this regard, an accurate database with a high time resolution (30 seconds) has been prepared for six months; Then, by one of the classical models of group method of data handling (GMDH), under the title of Combinatorial GMDH algorithm (CGA), short-term streamflow is forecasted by

FAT method and compared with Rating-Curves method (1-hour time resolution). Finally, the effect of the application of FAT technology on increasing the accuracy of short-term streamflow forecasting has been investigated. The innovation of this research is the study of the effect of FAT data on the accuracy of short-term forecasts as 1, 6, 12, 24, 48, and 72 hours-ahead.

2- Methodology

The group method of data handling (GMDH) is a self-organizing learning method presented by Ivakhnenko (1968) to overcome the dead-end problem of equation complexity and linear dependency present in standard regression equations. One of the classic methods of data group classification is the Combinatorial GMDH algorithm (CGA).

The combinatorial GMDH algorithm (CGA) is the main single-layer algorithm for GMDH, to prevent over-fitting. First, shuffle the input data then, from this shuffled data, 70% of the data will be selected as training data and the remaining 30% will be selected as testing data. According to the k-fold cross-validation criteria, the test data set is randomly divided into (K) sub-samples (In this study $K \leq 5$) of the same size. In each sub-sample, using the validation process, the number (K-1) of the sub-sample is considered as training data and one as the validation data set.

*Corresponding author's email: m.bahreini@wri.ac.ir



The input data matrix contains (N) observational data related to the (M) parameter. Training data are used to generate the optimal model. The optimal model is constructed by minimizing the following fitness function:

$$AR_{(B)} = \frac{1}{N_B} \sum_{i=1}^{N_B} (y_i - y_{i(B)})^2 \rightarrow \min \quad (1)$$

(AR) fitness function, (N_B) Training data, (y_i) output in each layer, and ($y_{i(B)}$) is Testing data.

In each layer, a new input variable is applied and controlled using the fitness function; the process of increasing accuracy is controlled. If no improvement in the performance of the algorithm is observed in the model with the addition of new variables, the model is stopped. The first layer is defined as the following equation.

$$y = a_0 + a_1x_i \quad i = 1, 2, \dots, M \quad (2)$$

Non-linear individuals can also be added to polynomials in subsequent layers. The coefficients in these equations for each layer are obtained by the least-squares method. The general state of the equations in layer (m) is defined as the following form:

$$y = a_0 + a_1x_i + a_2x_j + \dots + a_mx_l \quad (3)$$

$$i = j = l = 1, 2, \dots, M$$

Using the above algorithm for two different databases prepared by Rating-Curves and FAT, short-term forecasts are made for time intervals of 1, 6, 12, 24, 48, and 72 hours-ahead. To perform the above calculations, Wolfram Mathematica 12.1 software has been used.

3- Results and Discussion

In this study, the type of input data is the discharge collected in the river; therefore, the input data matrix is created by time delays on the discharge time series. The target data matrix also includes the discharge collected in the river. Then, with the combinatorial GMDH algorithm (CGA), short-term forecasts are made for the lead time of 1, 6, 12, 24, 48 and 72 hours-ahead. Finally, to investigate the effect of data collection on prediction accuracy, all processes related to the forecasting algorithm, in addition to the data extracted using FAT technology, were repeated on the data extracted by Rating-Curves. To evaluate the obtained information, different statistical criteria are used to evaluate the performance of each relationship; because statistical criteria alone cannot be a good criterion for the accuracy of different methods. In this

Table 1. Results of short-term forecasts performed on FAT and Rating-Curve data

ENS	RMSE (M ³ /S)	Lead Time (Hour)	Input Data Type
0.98	2.105	1	FAT
0.096	5.952	6	
0.94	9.207	12	
0.88	14.637	24	
0.73	22.375	48	
0.54	29.361	72	
0.97	4.12	1	Rating-Curve
0.84	19.677	6	
0.61	31.179	12	
0.27	41.661	24	
0.12	47.60	48	
0.11	45.96	72	

research, R, RMSE, MAE, MARE, and Nash-Sutcliffe (ENS) criteria have been used.

According to Table 1, the modeling performed by the CGA method showed that the forecast accuracy for FAT data is higher than the data prepared by the Rating-Curves data. For example, the RMSE for a 1h-ahead forecast of FAT data is 2.105 m³/s; while this criterion for 1-hour forecast on Rating-Curves data events has almost doubled to 4.12 m³/s. As the lead time increases, although the forecast error increases for both databases; But the prediction error on FAT data is still lower. For a 24-hour forecast, the RMSE difference between the two databases almost triples from 14.637 m³/s to 41.661 m³/s. This difference can be seen by other evaluation criteria. According to the Nash-Sutcliffe criterion, with increasing the lead time up to 12 h-ahead for FAT data, the forecast accuracy decreases slightly. For 24 h-ahead, the accuracy of the forecast decreases significantly. For Rating-Curve data with an increased lead time, the intensity of the accuracy decrease increases. So that the Nash-Sutcliffe coefficient for 12 and 24 hours-ahead, decreases from 0.94 and 0.88 for FAT data to 0.61 and 0.27 for Rating-Curve data, respectively.

According to the Nash-Sutcliffe criterion, with increasing the lead time up to 12 h-ahead for FAT data, the forecast accuracy decreases slightly. For 24 h-ahead, the accuracy of the forecast decreases significantly. For Rating-Curve data with an increased lead time, the intensity of the accuracy decrease increases. So that the Nash-Sutcliffe coefficient for 12 and 24 hours-ahead, decreases from 0.94 and 0.88 for FAT data to 0.61 and 0.27 for Rating-Curve data, respectively. The Nash-Sutcliffe (E_{NS}) figures for the 48 and 72 hours-ahead forecasts indicate that the forecasts for Rating-Curve data during this lead time are not accurate. As the Nash-Sutcliffe (E_{NS}) is (0.12 and 0.11), respectively, while this figures for the forecasts on the FAT are (0.73 and 0.54) and have relatively good accuracy.

4- Conclusion

The results of short-term forecasting performed on two different databases of FAT, and Rating-Curves showed that, in general, the combinatorial GMDH algorithm (CGA) algorithm. Has good accuracy in short-term forecasting. Forecast accuracy, regardless of the database used to predict reduced by increasing the lead time. The highest accuracy is related to 1-hour ahead and the lowest accuracy is related to 72-hour ahead forecasts. The accuracy of short-term forecasting performed on FAT data is higher than the Rating-Curves data. The reason for this high accuracy of prediction is the high time resolution and high accuracy of data collected. Therefore, it can be concluded that the application of FAT makes the algorithm used to predict better and more accurate results. The use of FAT can also help the forecasting model to increase the prediction accuracy for flood peaks.

References

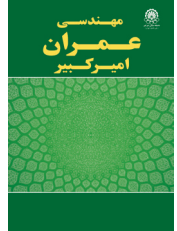
- [1] M. Abbasi, A. Farokhnia, M. Bahreinimotlagh, R. Roozbahani, A hybrid of Random Forest and Deep Auto-Encoder with support vector regression methods for accuracy improvement and uncertainty reduction of long-term streamflow prediction, *Journal of Hydrology*, 597 (2021) 125717.
- [2] M. Ehteram, F. Binti Othman, Z. Mundher Yaseen, H. Abdulmohsin Afan, M. Falah Allawi, A. Najah Ahmed, S. Shahid, V. P Singh, A. El-Shafie, Improving the Muskingum flood routing method using a hybrid of particle swarm optimization and bat algorithm, *Water*, 10(6) (2018) 807.
- [3] M.S. Khan, P. Coulibaly, Bayesian neural network for rainfall-runoff modeling, *Water Resources Research*, 42(7) (2006).
- [4] A. Mosavi, P. Ozturk, K.-w. Chau, Flood prediction using machine learning models: Literature review, *Water*, 10(11) (2018) 1536.

HOW TO CITE THIS ARTICLE

Y. Olfatmiri, E. Jabbari, M. Bahreinimotlagh, H. Alizadeh, A. H. Hasanabadi, *Application of Acoustic Tomographic Data in Short-Term Forecasting of Streamflow Using Combinatorial GMDH Algorithm (CGA)*, *Amirkabir J. Civil Eng.*, 54(5) (2022) 351-354.

DOI: [10.22060/ceej.2021.19850.7270](https://doi.org/10.22060/ceej.2021.19850.7270)





کاربرد داده‌های تیکه نگاری صوتی در پیش‌بینی کوتاه مدت نرخ جریان رودخانه‌ها با استفاده از مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها

یوسف الفت میری^۱، ابراهیم جباری^۱، مسعود بحرینی مطلق^{۲*}، حسین علیزاده^۱، امیرحسین حسن‌آبادی^۱

۱- دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
۲- پژوهشکده مطالعات و تحقیقات منابع آب، موسسه تحقیقات آب، تهران، ایران.

تاریخچه داوری:

دریافت: ۱۴۰۰/۰۱/۲۲
بازنگری: ۱۴۰۰/۰۴/۲۲
پذیرش: ۱۴۰۰/۰۵/۱۱
ارائه آنلاین: ۱۴۰۰/۰۶/۲۵

کلمات کلیدی:

الگوریتم دسته‌بندی گروهی داده‌ها
پیش‌بینی کوتاه مدت
پیش‌بینی نرخ جریان
تیکه نگاری صوتی

خلاصه: پیش‌بینی کوتاه مدت نرخ جریان رودخانه‌ها اهمیت فراوانی در مدیریت منابع آب و کنترل سیلاب دارد. یکی از مشکلاتی که محققین در این نوع از پیش‌بینی‌ها همواره با آن مواجه هستند؛ نبود یک بانک داده‌ای دقیق و با تفکیک‌پذیری زمانی بالا می‌باشد. فناوری تیکه نگاری صوتی یکی از روش‌های داده برداری نوین است که علاوه بر دقت بالای داده‌های برداشت شده دارای تفکیک‌پذیری زمانی بالایی نیز است؛ بنابراین با کاربرد داده‌های برداشت شده با استفاده از این فناوری به همراه یک مدل پیش‌بینی مناسب می‌توان به یک پیش‌بینی کوتاه مدت دقیق از نرخ جریان رودخانه‌ها دست یافت. در این تحقیق تأثیر کاربرد داده‌های برداشت شده توسط فناوری تیکه نگاری صوتی در پیش‌بینی کوتاه مدت نرخ جریان توسط مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها، بررسی شده و با داده‌های به دست آمده از روش دی-اشل مقایسه شده است. به منظور جلوگیری از بیش‌برازش شدن مدل پیش‌بینی از معیار اعتبارسنجی کافولد استفاده شده است. نتایج این تحقیق نشان داد که کاربرد داده‌های تیکه نگاری صوتی باعث افزایش دقت پیش‌بینی کوتاه مدت می‌شود؛ به طوری که ضریب نشت - ساتکلیف برای پیش‌بینی ۱، ۱۲، ۲۴، ۴۸ و ۷۲ ساعته دی بر روی داده‌های برداشت شده با روش تیکه نگاری صوتی به ترتیب (۰/۹۸، ۰/۹۶، ۰/۹۴، ۰/۸۸، ۰/۷۳ و ۰/۵۴) به دست آمد. در حالی که این مقادیر برای داده‌های برداشت شده با روش دی-اشل به ترتیب (۰/۹۷، ۰/۸۴، ۰/۶۱، ۰/۲۷، ۰/۱۲ و ۰/۱۱) محاسبه شد. از دلایل بالا بودن دقت پیش‌بینی انجام شده بر روی داده‌های تیکه نگاری صوتی نسبت به داده‌های دی-اشل می‌توان به تفکیک‌پذیری زمانی بالا و دقت بیشتر داده‌های برداشت شده توسط تیکه نگاری صوتی نسبت به دی-اشل اشاره کرد.

۱- مقدمه

پیش‌بینی و آگاهی از پارامترهای چرخه هیدرولوژیکی یکی از عوامل بسیار مهم برای مدیریت موفق منابع آب می‌باشد. یکی از این پارامترهای مهم، نرخ جریان رودخانه‌ها است؛ پیش‌بینی نرخ جریان رودخانه‌ها می‌تواند در سامانه‌های آبرسانی و هشدار سیلاب رودخانه‌ها و مدیریت مخازن به مهندسان کمک کند [۱]. محققین همواره به دنبال پیش‌بینی جریان‌های رودخانه بوده‌اند. بدین منظور آن‌ها با استفاده از مدل‌های هیدرولوژیکی مختلف، سعی در تخمین دبی رودخانه‌ها داشته‌اند [۲].

در بحث پیش‌بینی نرخ جریان رودخانه‌ها، مدل‌های هیدرولوژیکی عموماً به دو دسته کلی مدل‌های فیزیکی و مدل‌های تصمیم‌گیری داده محور تقسیم می‌شوند. پیچیدگی ارتباط بین متغیرهای محیطی و نرخ جریان باعث ایجاد محدودیت در به کار بردن مدل‌های فیزیکی می‌شود [۳]. همچنین

به عنوان یک مزیت از مدل‌های داده محور، این مدل‌ها نیاز به متغیرهای ورودی کمتری دارند و می‌توان با استفاده از سوابق داده‌های موجود این مدل‌ها را توسعه داد [۴].

سامسودین و همکاران^۱ (۲۰۱۰) برای دو رودخانه سلانگور و برنم واقع در مالزی که به ترتیب شامل ۴۷ و ۴۳ سال داده دبی ماهانه بودند اقدام به پیش‌بینی نرخ جریان ماهانه این دو رودخانه‌ها کردند؛ بدین منظور آن‌ها از روش‌های خود همبسته میانگین متحرک یکپارچه (ARIMA^۲)، شبکه عصبی مصنوعی (ANN^۳)، حداقل مربعات بردار ماشین پشتیبان (LSSVM^۴) و دسته‌بندی گروهی داده‌ها (GMDH^۵) بهره بردند. همچنین به عنوان یک نوآوری نیز با ترکیب دو مدل دسته‌بندی گروهی

- 1 Samsudin.et.al
- 2 Autoregressive Moving Average
- 3 Artificial Neural Networks
- 4 Least Squares Support Vector Machines
- 5 Group Method of Data Handling (GMDH)

* نویسنده عهده‌دار مکاتبات: m.bahreini@wri.ac.ir



[۸]. با استفاده از این روش، سرعت و دبی جریان [۹]، رسوبات معلق [۱۰]، انرژی جریان‌های جزر و مدی [۱۱] و دمای آب [۱۲] اندازه‌گیری می‌شود. مطالعات فراوانی به منظور ارزیابی صحت داده‌های برداشت شده توسط تیکه نگاری صوتی انجام شده است؛ که در آن‌ها روش تیکه نگاری صوتی را با دیگر روش‌های اندازه‌گیری مقایسه کرده‌اند.

کاوانیشی و همکاران (۲۰۱۸) برای اندازه‌گیری دبی آب در رودخانه گونو واقع در استان هیروشیما ژاپن، به مدت ۱۲۷ روز، نرخ جریان رودخانه را توسط دو روش دبی-اشل و تیکه نگاری صوتی اندازه‌گیری کردند و نتایج نشان داد که نرخ جریان محاسبه شده توسط روش دبی-اشل نسبت به روش تیکه نگاری صوتی دارای دقت کمتری بوده چرا که اریبی نسبی^{۱۰} مربوط به دبی-اشل در محدوده $\pm 20\%$ بوده است؛ همچنین تیکه نگاری صوتی توانسته است نامنظمی‌ها با طول موج کوتاه را در سری زمانی نرخ جریان ثبت کند [۱۳].

سوواف و کاوانیشی (۲۰۱۹) دو روش دبی-اشل و تیکه نگاری صوتی به منظور اندازه‌گیری نرخ جریان رودخانه در هنگام سیلاب را با یکدیگر مقایسه کردند و نتایج نشان داد که تیکه نگاری صوتی دقت بیشتری در اندازه‌گیری دبی جریان غیردائمی دارد [۱۴].

در ایران نیز مطالعاتی با استفاده از روش تیکه نگاری صوتی به منظور اندازه‌گیری سرعت جریان در رودخانه‌ها [۱۵] و دریاچه‌ها [۱۶]، دمای آب دریاچه‌ها [۱۷]، بار معلق [۱۸] و جریان‌های ساحلی [۱۹] انجام شده است. در نتیجه روش تیکه نگاری صوتی به عنوان روشی کارآمد جهت اندازه‌گیری به هنگام جریان رودخانه‌ها [۲۰] در شرایط نرمال و سیلابی [۲۱] قابل استفاده است.

هدف اصلی این تحقیق پیش‌بینی کوتاه مدت نرخ جریان رودخانه گونو واقع در استان هیروشیما ژاپن با استفاده از داده‌های روش تیکه نگاری صوتی و مقایسه آن با روش دبی-اشل است. در این راستا یک بانک اطلاعاتی دقیق با تفکیک زمانی بالا (۳۰ ثانیه) به مدت شش ماه تهیه شده؛ سپس توسط یکی از مدل‌های کلاسیک الگوریتم دسته‌بندی گروهی داده‌ها تحت عنوان مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها اقدام به پیش‌بینی کوتاه مدت نرخ جریان رودخانه با روش تیکه نگاری صوتی^{۱۱} و مقایسه آن با روش دبی-اشل (تفکیک زمانی ۱ ساعته) شده است و در نهایت تأثیر کاربرد فناوری تیکه نگاری صوتی بر افزایش دقت پیش‌بینی کوتاه مدت نرخ

داده‌ها و حداقل مربعات بردار ماشین پشتیبان مدل جدید GLSSVM را ارائه کردند. نتایج نشان داد که ترکیب مدل دسته‌بندی گروهی داده‌ها باعث افزایش دقت پیش‌بینی مدل حداقل مربعات بردار ماشین پشتیبان شد و مدل ترکیبی GLSSVM دارای دقت بالاتری در پیش‌بینی ماهانه دبی این رودخانه‌ها نسبت به سایر مدل‌های نام برده دارد [۵].

والتون و همکاران^۱ (۲۰۱۹) برای پیش‌بینی پیک سیلاب ۲ ساله در ایالت آیووا آمریکا از دو مدل کلاسیک دسته‌بندی گروهی داده‌ها به نام‌های مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها (CGA^۲) و مدل چندلایه الگوریتم دسته‌بندی گروهی داده‌ها (MGA^۳) و یک مدل ابتکاری ارائه شده توسط مؤلفین به نام الگوریتم ساختار تعمیم یافته دسته‌بندی گروهی داده‌ها (GSGA^۴) استفاده کردند. از مقایسه نتایج تحلیل توسط این سه روش به این نتیجه رسیدند که مدل ساختار تعمیم یافته دسته‌بندی گروهی داده‌ها می‌تواند ایراداتی را که در به کارگیری جداگانه مدل‌های کلاسیک دیده می‌شود رفع کند و نتایج دقیق‌تری را در بحث پیش‌بینی ارائه دهد [۶].

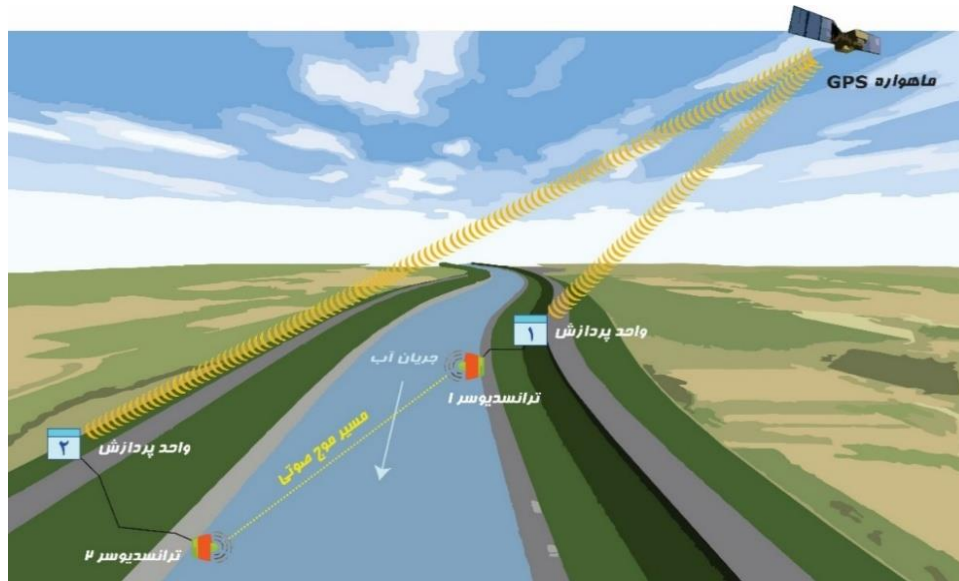
عادل پور و همکاران^۵ (۲۰۲۰) به منظور پیش‌بینی روزانه نرخ جریان رودخانه زلیکی رود در شمال ایران با استفاده از بانک دبی روزانه برداشت شده در ایستگاه‌های هیدرومتری، از دو مدل آماری روش‌های خود همبسته میانگین متحرک (ARMA^۶) و روش‌های خود همبسته میانگین متحرک یکپارچه (ARIMA^۷) به همراه سه مدل هوش مصنوعی الگوریتم دسته‌بندی گروهی داده‌ها، الگوریتم پرسپترون چندلایه (MLP^۸) و GRNN^۹ استفاده کردند. نتایج تحقیقات آن‌ها نشان داد که دو مدل الگوریتم پرسپترون چندلایه و دسته‌بندی گروهی داده‌ها از دیگر مدل‌های نام برده در پیش‌بینی روزانه نرخ جریان رودخانه دقت بالاتری داشتند و در بین این دو مدل پیش‌بینی نیز، مدل دسته‌بندی گروهی داده‌ها اندکی نسبت به مدل الگوریتم پرسپترون چندلایه دارای دقت پیش‌بینی بالاتری می‌باشد [۷].

کاوانیشی و همکاران (۲۰۱۰) فناوری تیکه نگاری صوتی رودخانه‌ای را جهت پایش پیوسته و بلند مدت خصوصیات جریان رودخانه‌ها ابداع کردند

- 1 Walton. et al
- 2 Combinatorial GMDH Algorithm
- 3 Multilayer GMDH Algorithm
- 4 Generalized Structure of GMDH Algorithm
- 5 Aghelpour. et. al
- 6 Autoregressive Moving Average
- 7 Autoregressive Integrated Moving Average
- 8 Multilayer Perceptron
- 9 General Regression Neural Network

10 Relative Biases

11 Fluvial Acoustic Tomography



شکل ۱. عملکرد فناوری تیکه نگاری صوتی [۹]

Fig. 1. Performance of FAT [9].

$$t_{down} = \frac{l}{C_m - U_m} \quad (1)$$

$$t_{up} = \frac{l}{C_m + U_m} \quad (2)$$

با حل دو معادله‌ی فوق، سرعت متوسط جریان آب و سرعت صوت حاصل خواهد شد:

$$C_m = \frac{l}{t_m} \quad (3)$$

$$U_m = \frac{C_m}{2l} \Delta t \quad (4)$$

که l فاصله افقی بین دو ایستگاه صوتی، $t_m = (t_{up} - t_{down}) / 2$ و $\Delta t = t_{up} - t_{down}$ است. در نتیجه با اندازه‌گیری سرعت متوسط در راستای امواج صوتی (U_m)، سرعت متوسط جریان رودخانه (V_m) به

جریان بررسی شده است؛ بنابراین نوآوری این تحقیق بررسی تأثیر داده‌های تیکه نگاری صوتی در دقت پیش‌بینی کوتاه مدت به صورت ۱، ۶، ۱۲، ۲۴، ۴۸ و ۷۲ ساعته می‌باشد.

۲- داده‌های تحقیق

۱-۲ روابط روش تیکه نگاری صوتی

در فناوری تیکه نگاری صوتی دو عدد ترانسدیوسر مطابق شکل ۱ در دو طرف رودخانه قرار می‌گیرند و سپس با ارسال هم‌زمان امواج صوتی که توسط ماهواره‌های GPS کنترل می‌شوند؛ گیرنده‌های GPS از نوع (u-blox LEA 6T) می‌باشد و از اختلاف زمان رفت و برگشت امواج صوتی در آب، سرعت جریان و سپس دبی جریان محاسبه می‌شود [۹].

روش محاسبه سرعت در روش تیکه نگاری صوتی، از طریق ارسال دو سویه امواج^۱ در محیط آبی می‌باشد. می‌دانیم که انتشار امواج صوتی در محیط آبی تابع سرعت صدا $C_{(x,y,z)}$ و سرعت جریان آب $U_{(x,y)}$ در آن محیط است؛ بنابراین زمان طی شده‌ی امواج صوتی بین دو ایستگاه بالادست و پایین‌دست که به صورت هم‌زمان ارسال گردیده‌اند، بدین صورت محاسبه می‌گردد [۸].

1 Reciprocal sound transmission



شکل ۲. محل قرارگیری ایستگاه های بالادست و پایین دست در سایت برداشت داده رودخانه گونو [۲۰].

Fig. 2. Location of upstream and downstream stations at the Gono River site [20].

این داده‌ها مربوط به رودخانه گونو^۱ واقع در شهر میوشی استان هیروشیما می‌باشد؛ شکل ۲ موقعیت و ایستگاه و دو سنجنده از یکدیگر را نشان می‌دهد [۲۰]. شهر میوشی با میانگین دمای متغیر بین ۸٫۵ تا ۱۹٫۴ درجه سانتی‌گراد از مناطق سردسیر ژاپن می‌باشد. میانگین بارش این شهر ۱۴۹۲ میلی‌متر است. طول رودخانه گونو ۱۹۴ کیلومتر و دارای دوشاخه اصلی ساجو^۲ و باسن^۳ می‌باشد. بستر این رودخانه ساختار سنگدانه‌ای و تخته سنگی دارد [۲۲].

به منظور ارزیابی نقش فناوری تیکه نگاری صوتی در افزایش دقت پیش‌بینی یک بانک داده‌ای دیگر نیز تهیه شده است که شامل دبی برداشت شده کوتاه مدت توسط روش دبی-اشل می‌باشد؛ مدت زمان برداشت هر دو نوع بانک داده‌ها برابر یکدیگر و بین دسامبر سال ۲۰۱۵ تا می ۲۰۱۶ است. شکل ۳ برای بازه زمانی مشخصی، دبی محاسبه شده توسط روش تیکه نگاری صوتی، دبی-اشل و مقایسه این دو را نشان می‌دهد. همچنین داده‌های دبی-اشل استفاده شده در این تحقیق به صورت ۱ ساعته برداشت شده است.

دست می‌آید. در نهایت با استفاده از رابطه پیوستگی، دبی جریان برابر است با حاصل ضرب سرعت متوسط جریان در مقطع عرضی رودخانه به دست می‌آید.

$$V_m = \frac{U_m}{\cos \theta} \quad (5)$$

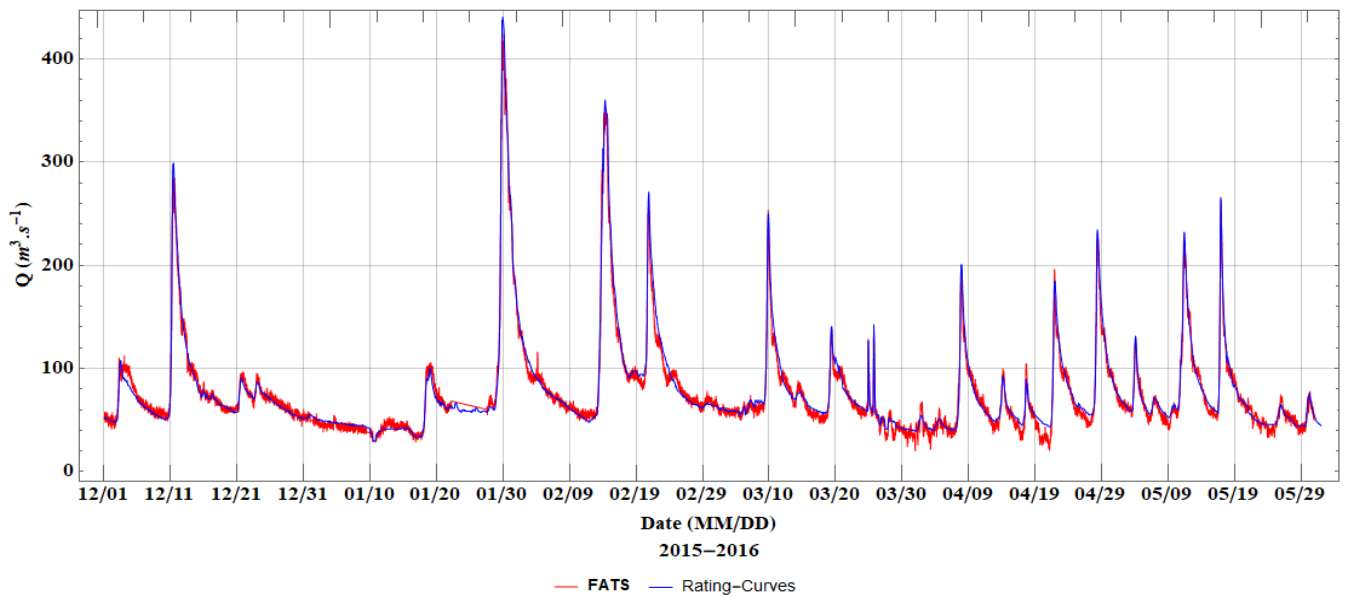
$$Q = A(H)V_m \sin \theta \quad (6)$$

که A مساحت مقطع عرضی بوده و تابعی از ارتفاع آب H است و θ زاویه بین راستای دو فرستنده و جهت جریان رودخانه می‌باشد.

۲-۲ منطقه مورد مطالعه

داده‌های دبی رودخانه که توسط روش تیکه نگاری صوتی در این تحقیق مورد استفاده قرار گرفته است شامل شش ماه برداری به فاصله زمانی ۳۰ ثانیه می‌باشد که برای افزایش دقت از میانگین متحرک مرتبه ۲۰ استفاده شده است؛ بنابراین فاصله زمانی هر داده ۱۰ دقیقه می‌باشد.

- 1 Gono River
- 2 Saijo
- 3 Basen



شکل ۳. مقایسه دبی اندازه‌گیری شده توسط روش دبی - اشل و روش تیکه نگاری صوتی.

Fig. 3. Comparison of discharge measured by the Rating-Curve and the FAT.

شده^۶ است و از بین این داده‌های برهم ریخته، ۷۰ درصد داده‌ها به عنوان داده آموزشی و ۳۰ درصد باقی مانده نیز به عنوان داده‌های ارزیابی انتخاب می‌شوند، سپس مطابق معیار اعتبارسنجی کا-فولد^۷ مجموعه داده‌های آموزشی به طور تصادفی به (K) زیر نمونه با حجم یکسان تفکیک می‌شود که در هر زیر نمونه با استفاده از فرآیند اعتبارسنجی تعداد $(K-1)$ زیر لایه به عنوان داده آموزشی و یکی به عنوان مجموعه داده اعتبار سنجی در نظر گرفته می‌شود. در این تحقیق $(K \leq 5)$ در نظر گرفته شده است.

ماتریس داده‌های ورودی شامل تعداد (N) داده مشاهداتی مربوط به (M) پارامتر است. داده‌های مشاهداتی به دو گروه داده‌های آموزشی (N_A) و داده‌های ارزیابی (N_B) تقسیم می‌شوند. از داده‌های آموزشی برای تولید مدل بهینه استفاده می‌شود، مدل بهینه با به حداقل رسیدن تابع تناسب^۸ زیر ساخته می‌شود:

$$AR_{(B)} = \frac{1}{N_B} \sum_{i=1}^{N_B} (y_i - y_{i(B)})^2 \rightarrow \min \quad (7)$$

۳- روش تحقیق

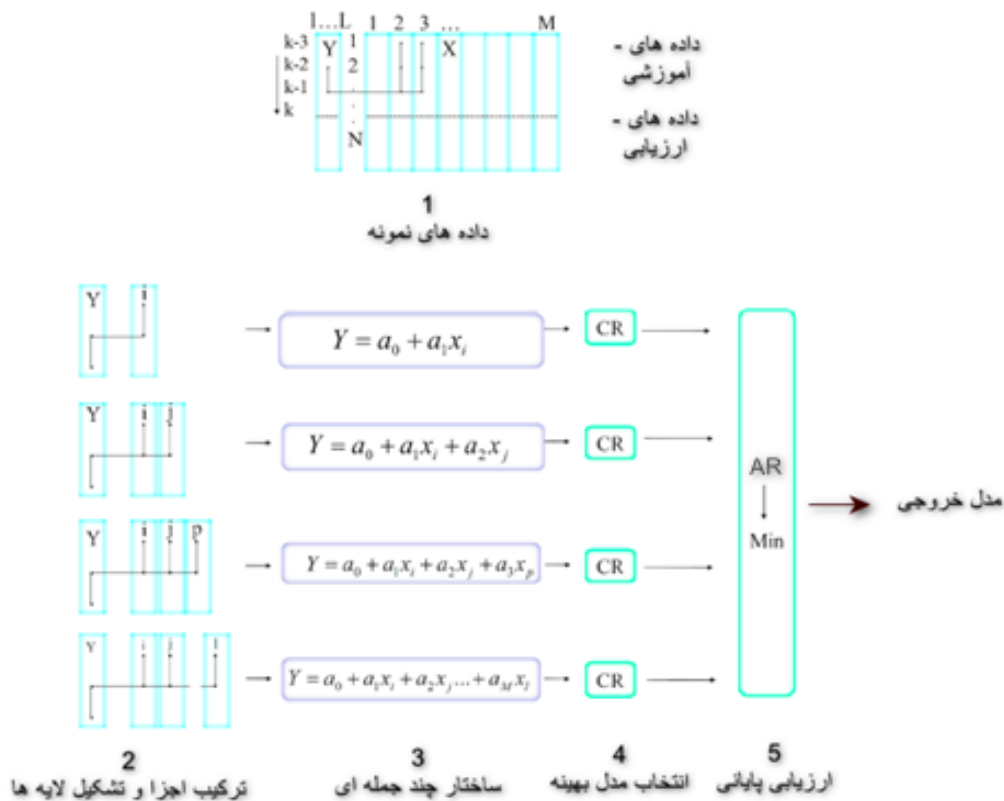
۳-۱- مدل پیش‌بینی ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها

روش دسته‌بندی گروهی داده‌ها یا $GMDH^1$ روش یادگیری خودسامان‌دهی است که مدل‌سازی دستگاه‌های غیرخطی را انجام می‌دهد. این روش اولین بار در سال ۱۹۶۸ توسط الکسی ایواکنو^۲ برای ساختن معادلات چندجمله‌ای مرتبه بالا ارائه شد که توانایی غلبه بر مشکلات مربوط به پیچیدگی معادله و وابستگی خطی در معادلات رگرسیون استاندارد را دارد. یکی از روش‌های کلاسیک دسته‌بندی گروهی داده‌ها، مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها (CGA^3) می‌باشد [۶]. الگوریتم دسته‌بندی گروهی داده‌ها را یک روش یادگیری ابتکاری^۴ می‌نامند؛ زیرا این روش قادر است مدل‌سازی سیستم‌های پیچیده از نوع رگرسیون با درجات بالا را انجام دهد.

مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها، یک مدل تک‌لایه از الگوریتم دسته‌بندی گروهی داده‌ها است. به منظور جلوگیری از بیش‌برازش^۵ مدل پیش‌بینی ابتدا چینش داده‌های مشاهداتی به صورت تصادفی برهم زده

6 Shuffle
7 k-fold cross-validation criteria
8 Fitness function

1 Group Method of Data Handling
2 Alexey Ivakhnenko
3 Combinatorial GMDH Algorithm
4 Heuristic
5 Overfitting



شکل ۴. ساختار کلی مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها [۶].

Fig. 4. The general structure of the CGA algorithm [6].

معادلات برای هر لایه توسط روش حداقل مربعات آ به دست می‌آید. حالت کلی معادلات در لایه (m) به صورت زیر تعریف می‌شود:

$$y = a_0 + a_1x_i + a_2x_j + \dots + a_mx_i \quad (9)$$

$$i = j = l = 1, 2, \dots, M$$

در روابط (۸ و ۹)، $\{a_0, a_1, a_2, \dots, a_m\}$ ضرایب وزنی و $\{x_i, x_j, \dots, x_l\}$ داده‌های ورودی در هر لایه می‌باشند.

با استفاده از الگوریتم فوق برای دو بانک داده‌ای متفاوت تهیه شده توسط دبی-اشل و تیکه نگاری صوتی، پیش‌بینی کوتاه مدت برای بازه‌های زمانی ۱، ۶، ۱۲، ۲۴، ۴۸ و ۷۲ ساعته انجام می‌شود. به منظور انجام محاسبات فوق از نرم‌افزار Wolfram Mathematica 12.1 استفاده شده است. در شکل ۴ ساختار کلی آن نشان داده شده است؛ مطابق با این

که (N_B) داده‌های ارزیابی، (y_i) خروجی در هر لایه و $(Y_{i(B)})$ داده‌های مربوط به بخش ارزیابی می‌باشد.

برای رسیدن به مدل خروجی از قانون خاتمه جستجوی جامع استفاده می‌شود. در هر لایه متغیر ورودی جدید، اعمال و با استفاده از تابع تناسب روند افزایش دقت، کنترل می‌شود. در صورتی که در مدل با افزایش متغیرهای جدید، بهبودی در عملکرد الگوریتم مشاهده نشود، مدل متوقف می‌شود. لایه اول که متغیرهای ورودی آن در لایه‌های بعدی نیز عضوی ثابت می‌باشد به صورت زیر تعریف می‌شود:

$$y = a_0 + a_1x_i \quad i = 1, 2, \dots, M \quad (8)$$

اعضای غیرخطی نیز می‌توانند در لایه‌های بعدی به چندجمله‌ای اضافه شده و باعث ایجاد پیچیدگی در مدل می‌شوند، ضرایب موجود در این

پیش‌بینی شده هستند همچنین نیز تعداد کل داده‌ها ورودی می‌باشد. MAE و RMSE دقت برازش و انحراف میان داده‌های مشاهداتی و مقادیر پیش‌بینی شده را نشان می‌دهند. شاخص RMSE برای نشان دادن توزیع نرمال خطای پیش‌بینی شده کاربرد دارد و برای همه مدل‌های پیش‌بینی مناسب نیست [۲۳]. از شاخص MEA برای محاسبه قدر مطلق اختلاف میان مقادیر مشاهداتی و پیش‌بینی شده استفاده می‌شود [۲۴]. MARE نشان دهنده خطای نسبی میان مقادیر پیش‌بینی شده نسبت به مقادیر مشاهداتی می‌باشد. شاخص R نیز مقداری بین [-۱ و +۱] می‌باشد که کوواریانس داده‌های پیش‌بینی شده را مشخص می‌کند؛ همچنین برای ارزیابی مدل بهینه از ضریب نش-ساتکلیف^۵ نیز استفاده شده است که رابطه آن به صورت زیر می‌باشد [۲۵].

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^N (Q_{O,i} - Q_{F,i})^2}{\sum_{i=1}^N (Q_{O,i} - \bar{Q}_{O,i})^2} \right] \quad -\infty \leq E_{NS} \leq 1 \quad (14)$$

در رابطه فوق $Q_{O,i}$ دبی مشاهداتی، $Q_{F,i}$ دبی پیش‌بینی شده و $\bar{Q}_{O,i}$ میانگین دبی مشاهده شده می‌باشد. هر چه معیار نش-ساتکلیف به یک نزدیک‌تر باشد مدل، دارای کارایی بهتری می‌باشد.

۴- نتایج و بحث

در این پژوهش با توجه به این که جنس داده‌های ورودی، دبی برداشت شده در رودخانه می‌باشد؛ بنابراین ماتریس داده‌های ورودی، با ایجاد تأخیرهای زمانی^۶ مورد نیاز بر روی سری زمانی دبی، ساخته می‌شود. ماتریس داده‌های هدف نیز شامل دبی برداشت شده در رودخانه می‌باشد، سپس با استفاده از مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها، اقدام به پیش‌بینی‌های کوتاه مدت، برای بازه‌های زمانی ۱، ۶، ۱۲، ۲۴، ۴۸ و ۷۲ ساعته می‌شود. در نهایت به منظور بررسی تأثیر نوع داده برداری بر روی دقت پیش‌بینی، کلیه فرایندهای ذکر شده علاوه بر داده‌های برداشت شده با استفاده از فناوری تیکه نگاری صوتی، بر روی داده‌های برداشت شده توسط دبی-اشل نیز تکرار شد. به منظور ارزیابی اطلاعات به دست آمده، از شاخص‌های آماری مختلفی جهت بررسی عملکرد هر یک از روابط استفاده می‌شود؛ زیرا یک

شکل در گام اول داده‌های ورودی به دو بخش داده‌های آموزشی و ارزیابی تقسیم می‌شوند. در گام دوم لایه‌های متفاوت برای توصیف پیچیدگی مدل‌ها تشکیل می‌شود. در گام سوم ساختار چندجمله‌ای مدل‌ها، ایجاد می‌شود و ضرایب هر چند جمله‌ای محاسبه می‌شود؛ سپس در گام چهارم مدل بهینه در هر لایه انتخاب می‌شود؛ در گام پنجم مدل اضافی برای لایه بعدی ساخته می‌شود و گام‌های ۲ تا ۵ مجدد تکرار می‌شوند، اگر مطابق با معیار AR مدل بهبود یافت، مجدد لایه‌های بعدی ساخته و مراحل ۲ تا ۵ طی می‌شوند؛ در غیر این صورت حلقه، به خاتمه جستجوی مدل بهینه می‌رسد.

۳-۲- معیار ارزیابی عملکرد

برای ارزیابی عملکرد پیش‌بینی انجام شده از شاخص‌های آماری ضریب همبستگی (R)، میانگین مطلق خطا (MAE^*)، خطای جذر میانگین مربعات ($RMSE^*$) و میانگین خطای نسبی ($MARE^*$) مطابق آنچه در ذیل به آن‌ها اشاره شده است استفاده می‌شود:

$$R = \frac{\sum_{l=1}^n (\delta_l - \bar{\delta})(p_l - \bar{p})}{\left(\sum_{l=1}^n (o_l - \bar{o}) \right) \left(\sum_{l=1}^n (p_l - \bar{p}) \right)} \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |o_i - p_i| \quad (11)$$

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (o_i - p_i)^2} \quad (12)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|o_i - p_i|}{o_i} \right) \quad (13)$$

که در روابط فوق O_l و P_l به ترتیب دبی مشاهداتی و دبی پیش‌بینی شده هست؛ \bar{O} و \bar{P} به ترتیب میانگین دبی مشاهداتی و میانگین دبی

5 Nash-Sutcliffe
6 Lags

1 Correlation Coefficient
2 Mean absolute error
3 Root mean square error
4 Mean absolute relative error

جدول ۱. نتایج پیش بینی کوتاه مدت انجام شده بر روی داده‌های تیکه نگاری صوتی و دبی-اشل

Table 1. Results of short-term forecasts prediction performed on FAT and discharge data.

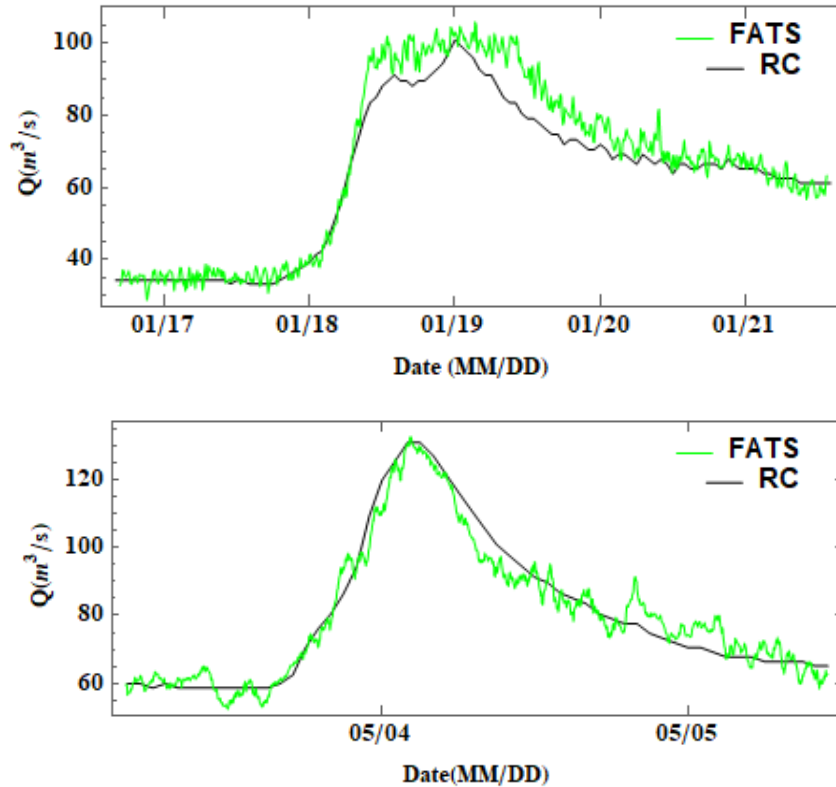
ENS	MARE	MAE	RMSE (M ³ /S)	R	بازه زمانی	
					پیش‌بینی (ساعت)	نوع داده‌های ورودی
<u>۰٫۹۸</u>	<u>۰٫۰۲۱</u>	<u>۱٫۵۱</u>	<u>۲٫۱۰۵</u>	<u>۰٫۹۸</u>	<u>۱</u>	تیکه نگاری صوتی
۰٫۹۶	۰٫۰۴۵	۳٫۳۴	۵٫۹۵۲	۰٫۹۷	۶	
۰٫۹۴	۰٫۰۵۵۱	۴٫۲۱۶	۹٫۲۰۷	۰٫۹۶	۱۲	
۰٫۸۸	۰٫۰۷۶۲	۶٫۰۰۹	۱۴٫۶۳۷	۰٫۹۴	۲۴	
۰٫۷۳	۰٫۱۱۹۷	۹٫۴۷	۲۲٫۳۷۵	۰٫۸۷	۴۸	
۰٫۵۴	۰٫۱۸۹۲	۱۴٫۰۹	۲۹٫۳۶۱	۰٫۵۴	۷۲	
<u>۰٫۹۷</u>	<u>۰٫۰۷</u>	<u>۱٫۵۵</u>	<u>۴٫۱۲</u>	<u>۰٫۹۸</u>	<u>۱</u>	دبی-اشل
۰٫۸۴	۰٫۰۹۶	۷٫۸۸	۱۹٫۶۷۷	۰٫۹۱	۶	
۰٫۶۱	۰٫۱۷۷	۱۳٫۹۷	۳۱٫۱۷۹	۰٫۷۸	۱۲	
۰٫۲۷	۰٫۲۸۴	۲۲٫۱۱	۴۱٫۶۶۱	۰٫۵۳	۲۴	
۰٫۱۲	۰٫۳۷۶۸	۲۶٫۰۸	۴۷٫۶۰	۰٫۳۴	۴۸	
۰٫۱۱	۰٫۳۸۹۱	۲۹٫۳۱۱	۴۵٫۹۶	۰٫۳۰	۷۲	

بازه زمانی پیش‌بینی گر چه خطای پیش‌بینی برای هر دو حالت افزایش می‌یابد؛ اما همچنان خطای پیش‌بینی بر روی داده‌های تیکه نگاری صوتی کمتر است؛ به طوری که برای پیش‌بینی ۲۴ ساعته مقدار اختلاف RMSE بین دو حالت تقریباً سه برابر رسیده و از مقدار ۱۴۶۳۷ متر مکعب بر ثانیه به مقدار ۴۱۶۶۱ می‌رسد. این اختلاف را می‌توان با بررسی دیگر معیارهای ارزیابی نیز مشاهده کرد.

همچنین مطابق با معیار E_{NS} - ساتکلیف با افزایش بازه زمانی پیش‌بینی تا ۱۲ ساعته برای داده‌های تیکه نگاری صوتی دقت پیش‌بینی کاهش ملایمی دارد. برای بازه زمانی ۲۴ ساعته دقت پیش‌بینی کاهش محسوسی پیدا می‌کند. برای داده‌های دبی-اشل با افزایش بازه زمانی

شاخص آماری به تنهایی نمی‌تواند بیانگر خوبی از دقت روش‌های مختلف باشد. در این پژوهش از معیارهای R ، $RMSE$ ، MAE ، $MARE$ و E_{NS} - ساتکلیف استفاده شده است.

مطابق با جدول ۱ نتیجه مدل‌سازی انجام شده توسط مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها نشان داد که دقت پیش‌بینی برای داده‌های تیکه نگاری صوتی بالاتر از داده‌های تهیه شده توسط روش دبی-اشل می‌باشد. به عنوان مثال $RMSE$ برای پیش‌بینی ۱ ساعته انجام شده بر روی داده‌های تیکه نگاری صوتی ۲٫۱۰۵ متر مکعب بر ثانیه است؛ در حالی که همین شاخص برای پیش‌بینی ۱ ساعته بر روی داده‌های دبی-اشل تقریباً دو برابر شده و به مقدار ۴٫۱۲ متر مکعب بر ثانیه می‌رسد. با افزایش



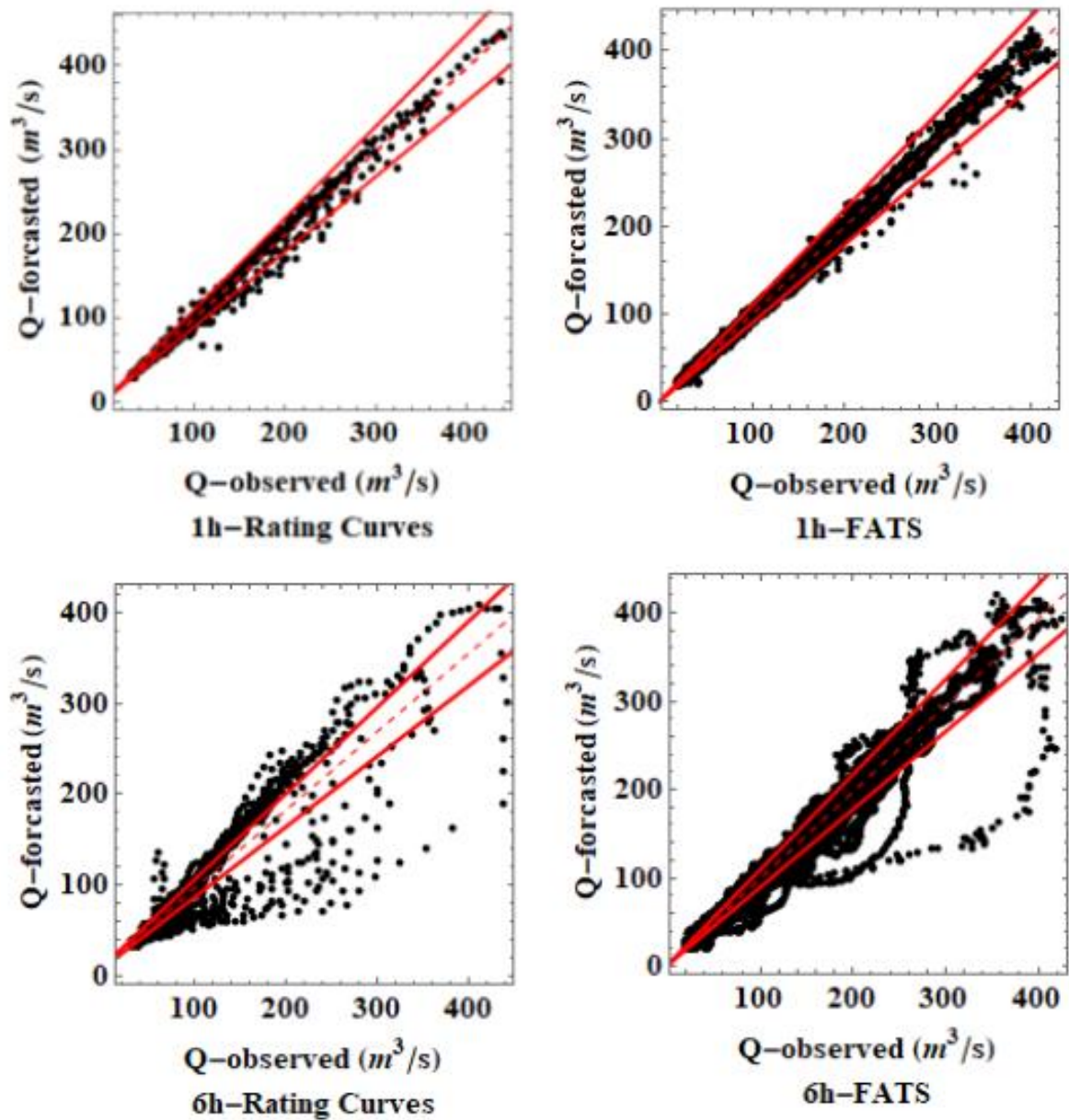
شکل ۵. مقایسه جزئیات نرخ جریان برداشت شده توسط تیکه نگاری صوتی و دبی-اشل برای دو پیک جریان

Fig. 5. Comparison of flow rate details collected by FAT and Rating-Curve methods for two flow peaks.

شده، نشان می‌دهد که پیش‌بینی انجام شده بر روی داده‌های دبی-اشل در این بازه دارای دقت مناسبی نیستند؛ به طوری که عدد نش - ساتکلیف به ترتیب (۰٫۱۲ و ۰٫۱۱) می‌باشد در حالی همین عدد برای پیش‌بینی انجام شده بر روی تیکه نگاری صوتی (۰٫۷۳ و ۰٫۵۴) می‌باشد و دقت نسبتاً خوبی دارند. یکی از دلایلی که دقت پیش‌بینی انجام شده بر روی داده‌های تیکه نگاری صوتی بیشتر است؛ تفکیک زمانی بالای برداشت داده‌ها می‌باشد. با توجه به این که زمان برداشت داده‌ها در روش تیکه نگاری صوتی هر ۳۰ ثانیه و فاصله زمانی برداشت نرخ جریان در روش دبی-اشل هر ۱ ساعت یک بار می‌باشد بنابراین تغییرات نرخ جریان در روش تیکه نگاری صوتی نسبت به دبی-اشل به صورت جزئی‌تر در حال ثبت است و همین جزئیات بیشتر، مدل پیش‌بینی را برای یافتن روابط دقیق‌تر به منظور پیش‌بینی کوتاه مدت کمک می‌کند و این دقت بالای برداشت داده‌ها [۹ و ۱۳ و ۱۴] و فاصله زمانی کم برداشت آن‌ها باعث می‌شود با افزایش بازه زمانی نیز دقت پیش‌بینی نسبت به پیش‌بینی انجام شده بر روی داده‌های دبی-اشل کاهش شدیدی پیدا نکند. شکل ۵ نشان دهنده تفاوت میان برداشت داده‌ها، میان

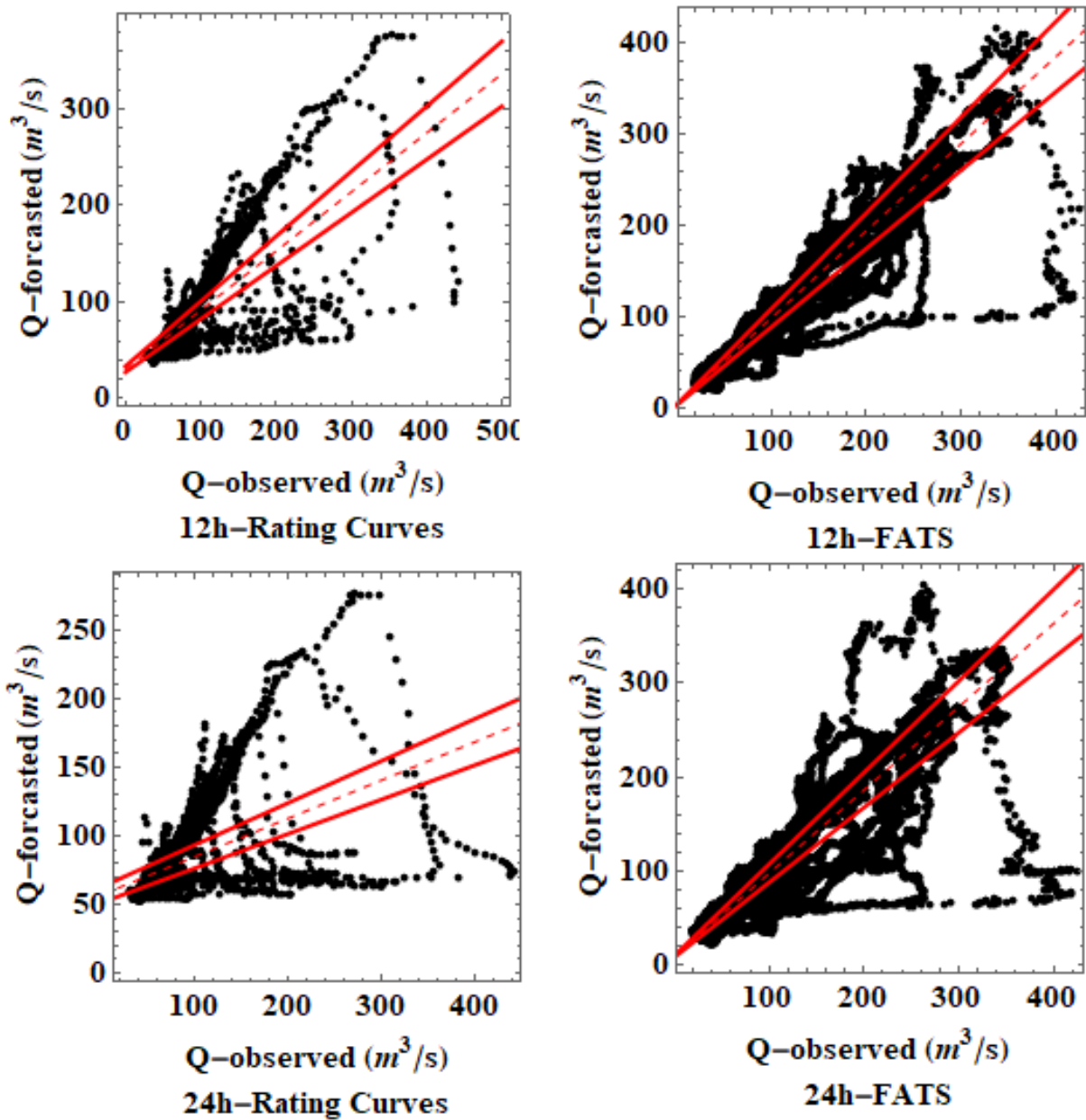
پیش‌بینی، شدت کاهش دقت افزایش می‌یابد؛ به طوری که ضریب نش - ساتکلیف بین بازه زمانی ۱۲ و ۲۴ ساعته به ترتیب از مقادیر ۰٫۹۴ و ۰٫۸۸ برای داده‌های تیکه نگاری صوتی به مقادیر ۰٫۶۱ و ۰٫۲۷ برای داده‌های دبی-اشل کاهش می‌یابد.

شکل ۶ نمودارهای نقطه‌ای مقایسه بین مقادیر پیش‌بینی شده و مشاهداتی را نشان می‌دهد. مطابق با این نمودارها بهترین برازش مربوط به پیش‌بینی ۱ ساعته داده‌های تیکه نگاری صوتی است. هر چه بازه زمانی افزایش پیدا می‌کند، تعداد داده‌های پیش‌بینی شده با محدوده خطای $\pm 10\%$ بیشتر می‌شوند. پراکندگی این داده‌ها برای پیش‌بینی انجام شده توسط داده‌های تیکه نگاری صوتی بین بازه‌های زمانی ۶ و ۱۲ ساعته از محدوده مذکور تفاوت چندانی ندارد در حالی که برای بازه زمانی ۲۴، ۴۸ و ۷۲ ساعته پراکندگی داده‌ها، با خطای بیش از $\pm 10\%$ افزایش چشم‌گیری پیدا می‌کند اما برای پیش‌بینی انجام شده بر روی داده‌های دبی-اشل این پراکندگی با افزایش بازه زمانی از ۱ ساعته تا ۷۲ ساعته مرتب در حال افزایش است. اعداد مربوط به نش - ساتکلیف (E_{NS})، برای پیش‌بینی‌های ۴۸ و ۷۲ ساعته انجام



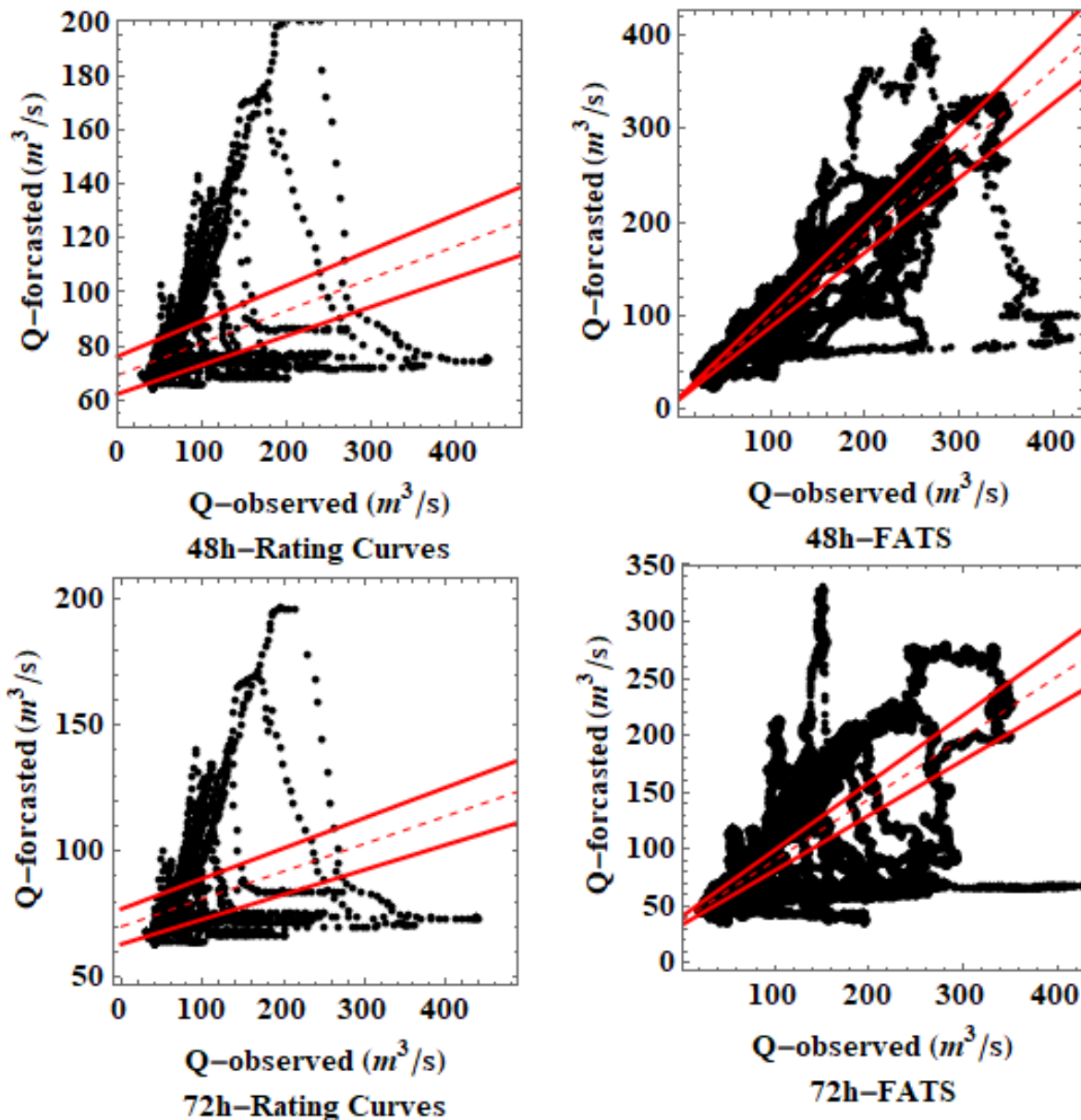
شکل ۶. الف) نمودار نقطه ای مقایسه ای بین مقادیر پیش‌بینی شده و مشاهداتی برای بازه های زمانی ۱ و ۶ ساعته با محدوده خطای $\pm 10\%$

Fig. 6. a. Comparative Scatter diagram between forecasted and observed values for 1 and 6 hours-ahead with an error range of 10%.



شکل ۶. ب) نمودار نقطه ای مقایسه ای بین مقادیر پیش‌بینی شده و مشاهداتی برای بازه های زمانی ۱۲ و ۲۴ ساعته با محدوده خطای $\pm 10\%$

Fig. 6. b. Comparative Scatter diagram between forecasted and observed values for 12 and 24 hours-ahead with an error range of 10%.

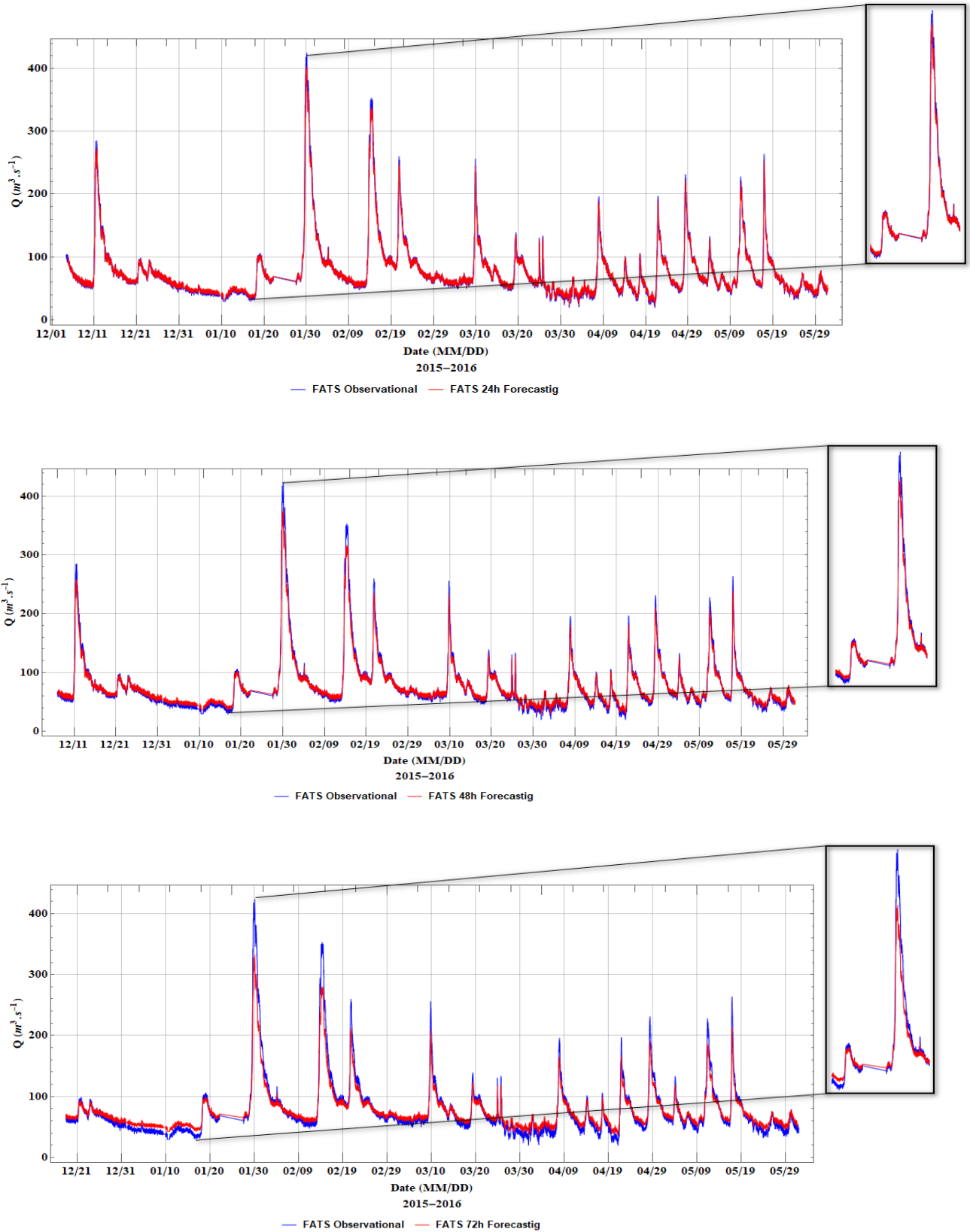


شکل ۶. ج) نمودار نقطه ای مقایسه ای بین مقادیر پیش‌بینی شده و مشاهداتی برای بازه های زمانی ۴۸ و ۷۲ ساعته با محدوده خطای $\pm 10\%$

Fig. 6. c. Comparative Scatter diagram between forecasted and observed values for 48 and 72 hours-ahead with an error range of 10%.

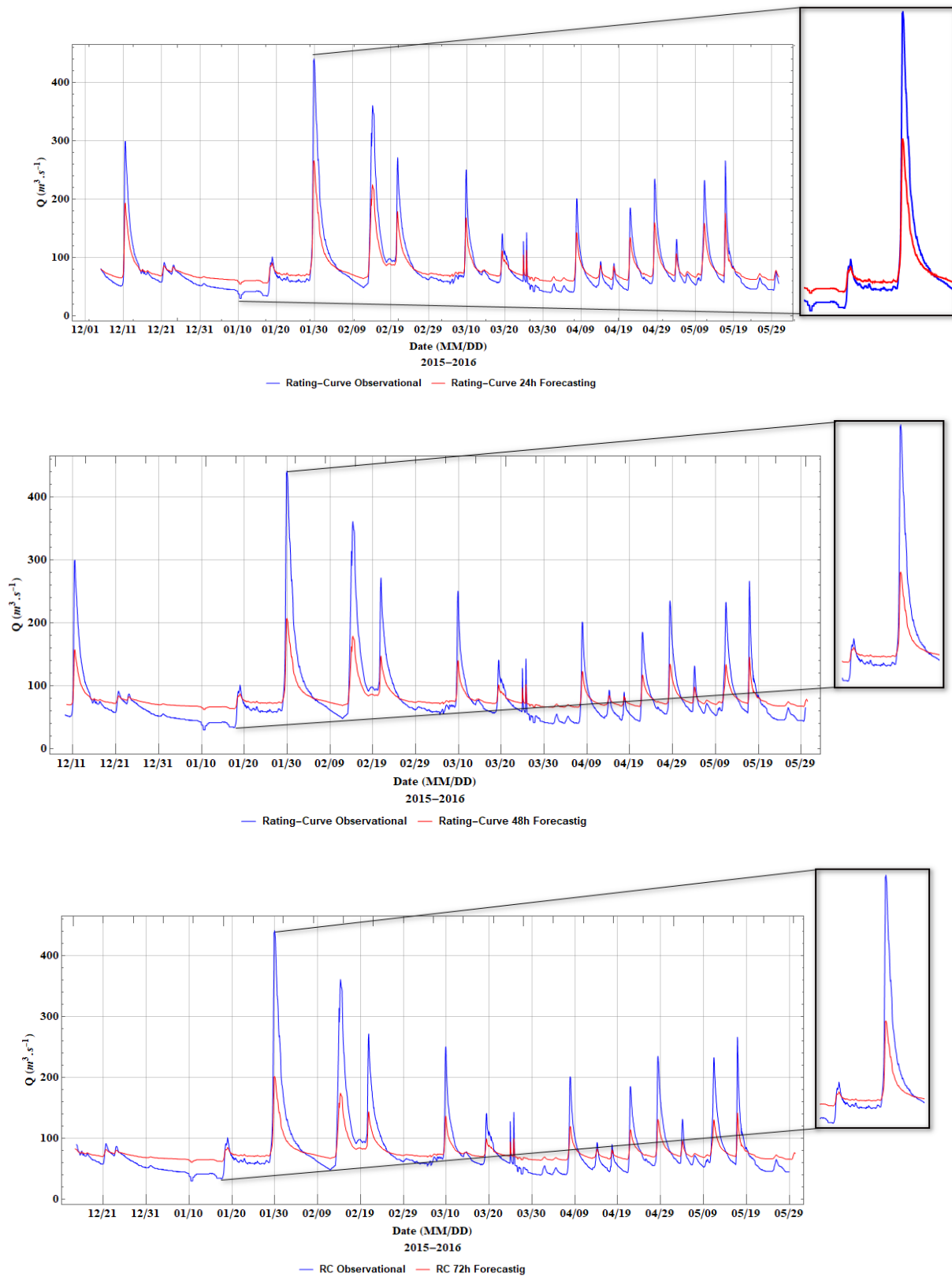
صوتی در پیش‌بینی کوتاه مدت نرخ جریان باعث می‌شود که الگوریتم پیش‌بینی کننده، تغییرات ناگهانی دبی رودخانه را پیش‌بینی کند. به عنوان مثال در تاریخ ۲۰۱۶/۰۱/۲۰ برای رودخانه مورد مطالعه، یک پیک سیلاب ناگهانی مشاهده می‌شود. استفاده از مدل ترکیبی الگوریتم دسته‌بندی گروهی داده‌ها در این تحقیق نشان داد که این مدل پیش‌بینی، پیک سیلاب را برای

تیکه نگاری صوتی و دبی-اشل می‌باشد؛ همان‌طور که در شکل ۵ مشخص است داده‌های تیکه نگاری صوتی برای دو پیک متفاوت به خوبی جزئیات تغییرات نرخ جریان را نسبت به دبی-اشل نشان می‌دهند؛ بنابراین تیکه نگاری صوتی جزئیات بیشتری از نوسانات دبی را نسبت به دبی-اشل نشان می‌دهد. همان‌طور که شکل ۷ نشان می‌دهد؛ استفاده از داده‌های تیکه نگاری



شکل ۷. الف) نمودار مقایسه بین دبی مشاهداتی و پیش‌بینی شده برای داده های تیکه نگاری صوتی.

Fig. 7. a. Comparison diagram between observed and forecasted streamflow for FAT data.



شکل ۷. ب) نمودار مقایسه بین دبی مشاهداتی و پیش‌بینی شده برای داده های دبی - اشل.

Fig. 7. b. Comparison diagram between observed and forecasted streamflow for Rating-Curve data.

O	دبی مشاهداتی، m^3/s
P	دبی پیش‌بینی شده، m^3/s
Q_o	دبی مشاهداتی، m^3/s
Q_p	دبی پیش‌بینی شده، m^3/s
R	ضریب همبستگی
$RMSE$	خطای جذر میانگین مربعات، m^3/s
C_m	سرعت متناسب صوت، m^2/s
V_m	سرعت متوسط آب، m^2/s
x	داده‌های دبی ورودی به مدل پیش‌بینی، m^3/s
y	خروجی مدل پیش‌بینی، m^3/s

علائم یونانی

θ	زاویه، درجه
----------	-------------

منابع

- [1] M. Abbasi, A. Farokhnia, M. Bahreinimotlagh, R. Roozbahani, A hybrid of Random Forest and Deep Auto-Encoder with support vector regression methods for accuracy improvement and uncertainty reduction of long-term streamflow prediction, *Journal of Hydrology*, (2020) 125717.
- [2] M. Ehteram, F. Binti Othman, Z. Mundher Yaseen, H. Abdulmohsin Afan, M. Falah Allawi, A. Najah Ahmed, S. Shahid, V. P Singh, A. El-Shafie, Improving the Muskingum flood routing method using a hybrid of particle swarm optimization and bat algorithm, *Water*, 10(6) (2018) 807.
- [3] M.S. Khan, P. Coulibaly, Bayesian neural network for rainfall-runoff modeling, *Water Resources Research*, 42(7) (2006).
- [4] A. Mosavi, P. Ozturk, K.-w. Chau, Flood prediction using machine learning models: Literature review, *Water*, 10(11) (2018) 1536.
- [5] R. Samsudin, P. Saad, A. Shabri, A hybrid least squares support vector machines and GMDH approach for river flow forecasting, *Hydrology and Earth System Sciences Discussions*, 7(3) (2010) 3691-3731.

پیش‌بینی انجام شده بر روی داده‌های تیکه نگاری صوتی بهتر و دقیق‌تر از پیش‌بینی انجام شده بر روی داده‌های دبی-اشل پیش‌بینی کرده است. این موضوع را با مقایسه دیگر پیک‌های سیلاب نیز می‌توان مجدد مشاهده کرد. با توجه به این که یکی از اهداف اصلی پیش‌بینی کوتاه مدت، کنترل سیلاب می‌باشد؛ بنابراین استفاده از داده‌های تیکه نگاری صوتی در بحث پیش‌بینی می‌تواند اطمینان بیشتری برای رسیدن به یک پیش‌بینی دقیق‌تر برای مهندسين ایجاد کند.

۵- نتیجه‌گیری

نتایج پیش‌بینی کوتاه مدت انجام شده بر روی دو نوع بانک داده‌ای متفاوت تیکه نگاری صوتی و دبی-اشل نشان داد که به طور کلی الگوریتم مدل ترکیبی دسته‌بندی گروهی داده‌ها که یکی از مدل‌های الگوریتم دسته‌بندی گروهی داده‌های کلاسیک هست دارای دقت مناسبی در پیش‌بینی کوتاه مدت می‌باشد. دقت پیش‌بینی انجام شده فارغ از بانک داده مورد استفاده با افزایش بازه زمانی پیش‌بینی کاهش می‌یابد. به طوری که بیشترین دقت مربوط به پیش‌بینی‌های ۱ ساعته و کمترین دقت مربوط به پیش‌بینی‌های ۷۲ ساعته است.

دقت پیش‌بینی کوتاه مدت انجام شده بر روی داده‌های تیکه نگاری صوتی بیشتر از داده‌های دبی-اشل می‌باشد. علت بالا بودن این دقت پیش‌بینی نیز، تفکیک زمانی بالای برداشت داده‌ها و صحت بالای داده‌های برداشت شده می‌باشد؛ بنابراین می‌توان نتیجه گرفت که کاربرد تیکه نگاری صوتی باعث می‌شود الگوریتم استفاده شده برای پیش‌بینی به نتایج بهتر و دقیق‌تری برسد. همچنین کاربرد داده‌های تیکه نگاری صوتی می‌تواند در افزایش دقت پیش‌بینی برای پیک‌های سیلاب به مدل پیش‌بینی کننده کمک کند.

۶- فهرست علائم

علائم انگلیسی

A	مساحت، m^2
AR	تابع تناسب
E_{NS}	ضریب نش-ساتکلیف
t	زمان، S
MAE	میانگین مطلق خطا
$MARE$	میانگین خطای نسبی

- M.H. KARDAN, S.A. Hasanli, Continuous Monitoring of Tidal Bores Using Acoustic Tomography Technique, (2019). (in Persian)
- [16] M. Bahreinimotlagh, R. Roozbahani, M. Eftekhari, A.K. Heydari, S. Abolhosseini, Investigation of Current Status in Haftbarm Lake Using Acoustic Tomography Technology, *Journal of Water and Soil*, 33 (2019). (in Persian)
- [17] M. Bahreinimotlagh, R. Roozbahani, M.J. Zareian, H. Kardan Moghadam, K. Mohtasham, The continuous water temperature monitoring by using Acoustic Tomography Technology, *Amirkabir Journal of Civil Engineering*, 51(5) (2018) 18-18. (in Persian)
- [18] M. Bahreinimotlagh, K. Kawanisi, A. Kavousi, R. Roozbahani, M. Abbasi, M.B. Al Sawaf, Influence of Suspended Sediment Concentration and Particle Sizes on the Sound Attenuation of the Fluvial Acoustic Tomography Technique, *Journal of Water and Environment Technology*, 18(5) (2020) 338-348.
- [19] M. Bahreinimotlagh, I. Khaki, R. Roozbahani, Y. Zohrabi, H. Kardan Moghaddam, Range-and depth averaged temperature measurement of the coastal sea using Acoustic Tomography technique, *Journal of Natural Environment*, 73(4) (2021) 637-647. (in Persian)
- [20] M. Bahreinimotlagh, R. Roozbahani, Y. Zohrabi, H. Kardanmoghadam, H. Dehban, K. Mohtasham, Feasibility study of real-time and automated monitoring of Iranian Rivers using 50-kHz fluvial acoustic tomography system, *Journal of Acoustical Engineering Society of Iran*, 8(1) (2020) 14-21. (in Persian)
- [21] M. Bahreinimotlagh, R. Roozbahani, M. Eftekhari, H. Kardanmoghadam, M. Abbasi, K. Mohtasham, Feasibility study of Fluvial Acoustic Tomography System for flood monitoring and determination of the measurement accuracy, minimum and maximum measurement ranges, *Iranian journal of Ecohydrology*, 6(3) (2019) 585-592. (in Persian)
- [22] M. Bahreinimotlagh, K. Kawanisi, M.M. Danial, M.B. Al Sawaf, J. Kagami, Application of shallow-water acoustic tomography to measure flow direction and river
- [6] R. Walton, A. Binns, H. Bonakdari, I. Ebtehaj, B. Gharabaghi, Estimating 2-year flood flows using the generalized structure of the Group Method of Data Handling, *Journal of Hydrology*, 575 (2019) 671-689.
- [7] P. Aghelpour, V. Varshavian, Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series, *Stochastic Environmental Research and Risk Assessment*, 34(1) (2020) 33-50.
- [8] K. Kawanisi, M. Razaz, A. Kaneko, S. Watanabe, Long-term measurement of stream flow and salinity in a tidal river by the use of the fluvial acoustic tomography system, *Journal of Hydrology*, 380(1-2) (2010) 74-81.
- [9] M. Bahreinimotlagh, K. Kawanisi, M.B. Al Sawaf, R. Roozbahani, M. Eftekhari, A.K. Khoshuie, Continuous streamflow monitoring in shared watersheds using advanced underwater acoustic tomography system: a case study on Zayanderud River, *Environmental monitoring and assessment*, 191(11) (2019) 1-9.
- [10] M. Bahreinimotlagh, K. Kawanisi, X.-H. Zhu, Acoustic Investigations of Tidal Bores in Qiantang Rive, 59 (2015) I_139-144.
- [11] K. Kawanisi, M. Bahreinimotlagh, M. Razaz, Energy flux measurement of tidal stream in a strait using two crossing ultrasonic transmission lines, 36th World Congr. Int. Assoc. Hydro-Environment Eng. Res.(IAHR-APD 2015), Hague, Netherlands, (2015) 1-4.
- [12] K. Kawanisi, A. Kaneko, S. Nigo, M. Soltaniasl, M.F. Maghrebi, New acoustic system for continuous measurement of river discharge and water temperature, *Water Science and Engineering*, 3(1) (2010) 47-55.
- [13] K. Kawanisi, M.B. Al Sawaf, M.M. Danial, Automated real-time streamflow acquisition in a mountainous river using acoustic tomography, *Journal of Hydrologic Engineering*, 23(2) (2018) 04017059.
- [14] M.B. Al Sawaf, K. Kawanisi, Novel high-frequency acoustic monitoring of streamflow-turbidity dynamics in a gravel-bed river during artificial dam flush, *Catena*, 172 (2019) 738-752.
- [15] M. Bahreinimotlagh, R. Roozbahani, M. Eftakhari,

- Singh, Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model, *Journal of Hydrology*, 554 (2017) 263-276.
- [25] J.E. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models part I—A discussion of principles, *Journal of hydrology*, 10(3) (1970) 282-290.
- discharge, *Flow Measurement and Instrumentation*, 51 (2016) 30-39.
- [23] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE), *Geoscientific Model Development Discussions*, 7(1) (2014) 1525-1534.
- [24] Z.M. Yaseen, I. Ebtahaj, H. Bonakdari, R.C. Deo, A.D. Mehr, W.H.M.W. Mohtar, L. Diop, A. El-Shafie, V.P.

چگونه به این مقاله ارجاع دهیم

Y. Olfatmiri, E. Jabbari, M. Bahreinimotlagh, H. Alizadeh, A. H. Hasanabadi, Application of Acoustic Tomographic Data in Short-Term Forecasting of Streamflow Using Combinatorial GMDH Algorithm (CGA), Amirkabir J. Civil Eng., 54(5) (2022) 1755-1772.

DOI: 10.22060/ceej.2021.19850.7270



