

Prediction of Peak Velocity of Pollutant Transport in Rivers Using Group Method Data Handling (GMDH) and Intelligent Hybrid Method GMDH-HS

E. Golkar¹, M. M. Ahmadi², K. Qaderi³, M. Rahimpour³

1. MSc Student of Water Structure, Department of Water Engineering, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran
2. Assist. Prof., Department of Water Engineering, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran (Corresponding Author) ahmadi_mm@uk.ac.ir
3. Assoc. Prof., Department of Water Engineering, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

(Received Apr. 24, 2017 Accepted Feb. 7, 2018)

To cite this article :

Golkar, E., Ahmadi, M.M., Qaderi, K., Rahimpour, M., 2018, "Peak velocity of pollutant transport prediction in rivers using Group Method Data Handling (GMDH) and intelligent hybrid method (GMDH-HS)." *Journal of Water and Wastewater*, 30(1), 64-76. Doi: 10.22093/wwj.2018.87689.2389 (In Persian)

Abstract

Estimating of the peak velocity of pollutant in a flow using hydraulic and geometrical parameters is very important in predicting the pollution transport in rivers. Suitable empirical equations are developed, none of which is reliable enough in estimating the peak velocity of pollutant before a thorough calibration. So, in this research artificial intelligence methods are used for this purpose. The objective of this research was to predict the peak velocity of pollutant by Group Method Data Handling (GMDH) methods and an intelligent hybrid method (GMDH-HS). The result of these methods were compared to the best regression equation. The dimensionless relative discharge (Q/a), dimensionless drainage area, the ratio of discharge at the section at the time of measurement to drainage area (Q/Da) and the reached slope (S) were taken as input parameters to these models. These data were collected from several different rivers in the United States. Mackey-Glass standard function was used to evaluate the performance of the GMDH and GMDH-HS models. The results indicated that the proposed models predicted the peak velocity of pollutant precisely (CE GMDH = 0.9328, CE GMDH-HS = 0.9038 & CE Vp Equation = 0.3802) and these models are more accurate compared to the best nonlinear regression equation.

Keywords: Estimating Peak Velocity, Intelligence Methods, Group Method Data Handling, GMDH-HS Method.



پیش‌بینی سرعت پیک آلودگی در رودخانه با استفاده از روش GMDH و روش هوشمند تلفیقی GMDH-HS

الهام گل‌کار^۱، محمد مهدی احمدی^۲، کورش قادری^۳، مجید رحیم‌پور^۳

۱- دانشجوی کارشناسی ارشد سازه‌های آبی، بخش مهندسی آب، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران

۲- استادیار بخش مهندسی آب، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران
(نویسنده مسئول) ahmadi_mm@uk.ac.ir

۳- دانشیار بخش مهندسی آب، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران

(دریافت ۹۷۲/۴ پذیرش ۹۷۱/۱۸)

برای ارجاع به این مقاله به صورت زیر اقدام فرمایید:

گل‌کار، ا.، احمدی، م. م.، قادری، ک.، رحیم‌پور، م.، ۱۳۹۷. "پیش‌بینی سرعت پیک آلودگی در رودخانه با استفاده از روش هوشمند (GMDH) و روش هوشمند تلفیقی GMDH-HS" مجله آب و فاضلاب، ۳۰(۱)، ۶۴-۷۶. Doi: 10.22093/wwj.2018.83092.2389

چکیده

تخمین مشخصات ماده آلوده کننده با استفاده از پارامترهای هندسی و هیدرولیکی جریان، یکی از عوامل مؤثر در پیش‌بینی انتقال آلودگی در رودخانه‌ها است. روش‌های عددی و تحلیلی مناسبی برای شبیه‌سازی انتقال و پخش آلودگی توسعه داده شده، اما هیچ یک از آنها قبل از کالیبره کردن با اطمینان قابل استفاده نیستند. بنابراین ساخت مدلی که بتواند با سرعت بیشتر و با اطلاعات کلی، انتقال آلودگی را پیش‌بینی کند، ضروری به نظر می‌رسد. از جمله روش‌های جدیدی که برای مدل‌سازی پیش‌بینی رفتار پدیده‌ها استفاده می‌شوند، روش‌های هوشمند داده محور می‌باشند. در این پژوهش پیش‌بینی سرعت پیک آلودگی توسط دو مدل برخورد گروهی با داده‌ها (GMDH) و مدل هوشمند تلفیقی (GMDH-HS) صورت گرفت. بنابراین دو مدل با کد نویسی در نرم‌افزار MATLAB بر مبنای دو روش ساخته و به کار گرفته شدند. پارامترهای ورودی به این مدل‌ها شامل دبی نسبی بی‌بعد (Q'a)، سطح نفوذ بی‌بعد (D'a)، نسبت دبی به سطح نفوذ بی‌بعد و شیب رودخانه (S) بود. به منظور صحت‌سنجی مدل‌های استفاده شده از تابع استاندارد Mackey-Glass استفاده شد. نتایج کاربرد دو مدل حاکی از عملکرد مناسب، دقت بالای هر دو مدل به ویژه مدل GMDH و افزایش دقت محاسبه سرعت پیک آلودگی به میزان حدوداً ۵۲ درصد نسبت به بهترین معادله رگرسیونی بود.

واژه‌های کلیدی: تخمین سرعت پیک ماده آلوده کننده، روش‌های هوشمند، روش برخورد گروهی با داده‌ها (GMDH)، روش هوشمند تلفیقی (GMDH-HS)

۱- مقدمه

معادله دیفرانسیل با مشتقات جزئی است. ایجاد می‌شوند. توسعه مدل‌های ریاضی توانایی درک بهتر از پیش‌بینی و کنترل کیفیت آب در رودخانه‌ها را ایجاد می‌کند. هدف مدل‌های کیفیت آب رودخانه‌ای، شبیه‌سازی تغییرات غلظت آلودگی در مسیر حرکت جریان در رودخانه است.

آلودگی آب در نتیجه فعالیت اقتصادی، اجتماعی و توسعه جمعیت همواره از مسائل مهم در مدیریت محیط زیست بوده است. تکنیک‌های مدل‌سازی به‌عنوان یک ابزار برای درک جریان‌ات زیست محیطی مورد استفاده قرار می‌گیرند. در این بین مدل‌های کیفیت آب غالباً توسط مدل‌های ریاضی که شامل یک یا چند



مطالعات آزمایشگاهی و همچنین مطالعات میدانی بر اساس مشاهدات ماده ردیاب (Wiler and Messinger, 2013)، نیز برای مطالعه عوامل مؤثر بر این ضریب انجام شده است و معادلاتی نیز ارائه شده است. تمام روش‌های ارائه شده منجر به تخمینی از مقدار ضریب پخش خواهند شد که دارای خطا و عدم قطعیت است. بنابراین این ضریب به‌عنوان یک ضریب کالیبراسیون در مدل‌های شبیه‌سازی حرکت آلودگی مطرح است.

علاوه بر مشکلات در برآورد ضریب پخش طولی، سرعت جریان نیز برای شبیه‌سازی حرکت آلودگی باید برآورد شود. سرعت جریان معمولاً با استفاده از مدل‌های جریان پیش‌بینی می‌شود و این مدل‌ها نیاز به جزئیات هندسی کانال و ضریب مقاومت جریان دارند که این اطلاعات نیز به ندرت در اختیار می‌باشند. در نتیجه در دسترس بودن اطلاعات ورودی قابل اعتماد، تقریباً همیشه ضعیف‌ترین اتصال در حلقه رویدادهای مورد نیاز برای پیش‌بینی حرکت، رقیق‌سازی و اختلاط آلاینده در رودخانه‌ها و جریان‌ها است (Jabson, 1995).

با توجه به این مطالب استفاده از روش‌های قابل اعتماد که بتواند با سرعت بیشتر و بدون نیاز به محاسبه پارامترهای پیچیده، مشخصات حرکت آلودگی و ضریب پخش را پیش‌بینی کند بسیار کاربردی و حائز اهمیت خواهد بود. معادلات رگرسیونی از جمله این روش‌ها می‌باشند اما با توجه به پژوهش‌های وایلی و همکاران این معادلات دقت پایینی دارند (Wiley and Messinger, 2013). روش دیگر که سودمندی و دقت آن در زمینه‌های مختلف مهندسی و مدیریت منابع آب اثبات شده است روش پیش‌بینی هوشمند^۳ یا محاسبات نرم^۴ است. این روش‌های محاسباتی الهام گرفته از طبیعت^۵ است و کاربرد آن‌ها در حل مسائل پیچیده دنیای واقعی است.

روش‌های محاسبات هوشمند به دو گروه اصلی روش‌های داده محور^۶ و روش‌های دانش محور^۷، تقسیم می‌شوند. روش‌های داده محور مبتنی بر اطلاعات و داده‌ها می‌باشند و فقط به داده‌های ورودی و خروجی بستگی دارد. از جمله روش‌های محاسبات نرم

سازمان‌های بین‌المللی همواره به حفاظت رودخانه‌ها و پاک‌سازی آب آن‌ها توجه ویژه‌ای دارند. به این دلیل توانایی تخمین و شبیه‌سازی جریان و انتقال آلودگی و رسوب برای برنامه‌ریزان منابع آب، اهمیت زیادی دارد. شبیه‌سازی حرکت آلودگی در جریان‌های طبیعی، اطلاعات مشخصی برای مدیریت کیفیت رودخانه و دیگر مسائل مرتبط فراهم خواهد کرد (Julinel and Riha, 2017).

این شبیه‌سازی‌ها غالباً با به‌کارگیری مدل‌های ریاضی که در آنها از روش‌های عددی برای حل معادلات دیفرانسیل استفاده شده، صورت می‌پذیرد. علاوه بر این، در چنین مدل‌هایی از روش‌های آماری (Razek and Abouelnour, 1972) و روش‌های مبتنی بر آنالیز ابعادی (Zalenakova et al., 2012) نیز برای پیش‌بینی آلودگی در رودخانه استفاده شده است. هدف اصلی این شبیه‌سازی‌ها برآورد میزان انتقال-پخش آلودگی و زمان حرکت آلودگی است. پیش‌بینی این پارامترها به‌منظور پیش‌بینی نرخ حرکت و میزان رقیق شدن آلودگی دارای اهمیت است. سرعت متوسط جریان و خصوصیات اختلاط از اطلاعات اساسی مورد نیاز برای این پیش‌بینی است (Harvey and Jobson, 1996).

در مواقعی که بار آلاینده به‌صورت ناگهانی به رودخانه وارد می‌شود^۱ برآورد سریع و دقیق از زمان و سرعت پیک آلودگی، برای برنامه‌ریزان و مهندسان منابع آب و محیط زیست بسیار ضروری است. مدل‌های عددی قابل قبولی برای شبیه‌سازی حرکت آلودگی تهیه شده است (Faramarzi et al., 2005). اساس این مدل‌ها حل معادلات حاکم با روش‌های عددی است اما هیچ‌کدام از آن‌ها قبل از کالیبره کردن و اعتبارسنجی روی رودخانه مورد نظر قابل استفاده نیستند. علاوه بر این، این مدل‌ها احتیاج به اطلاعاتی راجع به سرعت جریان و میزان اختلاط دارند درحالی‌که روش قابل اعتمادی برای پیش‌بینی ضرایب پخش و میزان اختلاط با توجه به اطلاعات هیدرولیکی وجود ندارد. از روش‌های پیش‌بینی هوشمند نیز به‌طور گسترده در سال‌های اخیر توسط محققان برای برآورد ضریب انتشار طولی استفاده شده است (Sattar and Gharabaghi, 2015).

³ Computational Intelligent

⁴ Soft Computing

⁵ Nature-inspired

⁶ Data Driven Method

⁷ Knowledge Based Method

¹ Travel Time

² Instantaneous Sources



یک رویکرد خود سازماندهی است و به تدریج مدل‌های پیچیده‌تری در طول ارزیابی عملکرد مجموعه داده‌های ورودی و خروجی تولید می‌کند (Ivakhnenko, 1968). در روش GMDH ارتباط بین متغیرهای ورودی - خروجی توسط سری توابع ولترا که مشابه با چند جمله‌ای گسسته شده کولموگروف-گابور می‌باشد، بیان می‌شود (Ivakhnenko, 1968).

$$y = w_0 + \sum_{i=1}^n w_i X_i + \sum_{i=1}^n \sum_{j=1}^n w_{ij} X_i X_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ijk} X_i X_j X_k + \dots \quad (1)$$

که در آن

y متغیری خروجی، $W=(w_0, w_1, \dots, w_m)$ بردار وزن‌ها و $X=(X_1, X_2, \dots, X_m)$ بردار متغیرهای ورودی است.

مدل GMDH بر مبنای خودسازمانی ذهنی^۷ توسعه داده شده تا بتواند به پیچیدگی مسائل چندبعدی فایق آید. این روش ابتدا متغیرها را به صورت دوتایی در نظر گرفته و سپس یک مقدار آستانه برای ارزیابی متغیرهایی که نمی‌توانند به تراز عملکرد برسند، مشخص می‌کند. این روال یک الگوریتم خودسازمانده را بیان می‌کند. در این پژوهش مدل GMDH با توجه به گام‌های زیر ساخته شد.

گام اول: تقسیم داده‌ها به دو مجموعه آموزش و تست داده‌های اصلی به دو دسته آموزش و آزمون تقسیم شدند. از داده‌های آموزش برای تخمین برخی ویژگی‌های سیستم غیرخطی استفاده شد و مجموعه آزمون برای برآورد مجموعه کامل ویژگی‌ها به‌کارگرفته شد.

گام دوم: ایجاد ترکیب دوتایی از متغیرها در هر لایه تمام ترکیبات دوتایی n متغیر ورودی برای لایه ساخته شد. تعداد این ترکیبات برابر است با

$$C_n^m = \frac{n!}{r!(m-r)!} \quad (2)$$

که در آن

n تعداد متغیرهای ورودی و r ضریبی است که برابر با ۲ در نظر گرفته شده است.

⁷ Heuristic self-organization

می‌توان روش‌های رگرسیونی، شبکه عصبی مصنوعی^۱، روش گروهی برخورد با داده‌ها^۲، روش هوشمند مختلط^۳، الگوریتم بهینه‌سازی جستجوی هارمونی^۴، تکامل رقابتی جوامع^۵، الگوریتم ژنتیک^۶ و ... را نام برد.

از لحاظ تئوری برای مدل کردن یک سیستم لازم است که روابط ریاضی صریح بین ورودی‌ها و خروجی‌ها به‌طور دقیق معلوم باشد. استخراج یک مدل‌سازی صریح بسیار مشکل است و در بسیاری از سیستم‌ها این روابط نامعلوم هستند. در این حالت روش‌های داده محور که محاسبات را بر مبنای داده‌های ورودی و خروجی انجام می‌دهند، مورد توجه قرار می‌گیرند (Hoseinzadeh et al., 2013).

در چند سال اخیر، کاربرد اصلی تعدادی از روش‌های فوق در زمینه‌های مختلف به‌ویژه مهندسی منابع آب گزارش شده است (Tsai and Yen, 2017). روش‌های داده محور، توانایی‌های زیادی در شناسایی سیستم‌های غیرخطی و مسائل کنترلی دارند. پژوهش‌های زیادی برای استفاده از این روش‌ها به‌عنوان ابزاری مؤثر برای شناسایی و مدل‌سازی سیستم‌های پیچیده صورت گرفته است (Sanchez et al., 1997).

در این پژوهش مدل پیش‌بینی سرعت حرکت پیک آلودگی در رودخانه‌ها با استفاده از روش برخورد گروهی با داده‌ها GMDH و روش هوشمند تلفیقی GMDH-HS توسعه داده شد. این مدل‌ها امکان پیش‌بینی زمان حرکت آلودگی را بدون نیاز به برآورد پارامترهای پیچیده و تنها با استفاده از اطلاعات کلی هیدرولیکی در زمان کوتاه فراهم می‌کند که می‌تواند به‌طور مؤثری مورد استفاده مسئولان، برنامه‌ریزان و بهره‌برداران منابع آب قرار گیرد.

۲- روش تحقیق

۲-۱- ساخت مدل پیش‌بینی با استفاده از الگوریتم برخورد گروهی با داده‌ها (GMDH)

در این پژوهش از روش هوشمند کنترل گروهی داده‌ها (GMDH) به‌منظور پیش‌بینی سرعت پیک آلودگی استفاده شد. روش GMDH

¹ Artificial Neural Network (ANN)

² Group Method Data Handling (GMDH)

³ Nero- Fuzzy

⁴ Harmony Search (HS)

⁵ Shuffled Complex Evolution (SCE)

⁶ Genetic Algorithm (GA)



گام پنجم: قانون توقف برای ایجاد ساختمان چندلایه با مقایسه مقدار شاخص لایه فعلی با لایه بعدی که باید ایجاد شود، از ایجاد لایه‌های بعدی در صورتی که مقدار شاخص بهبود نیابد یا کمتر از مقدار پیش‌فرض تعریف شده باشد، متوقف می‌شود. در غیر این صورت گام‌های دوم تا چهارم تکرار می‌شوند تا زمانی که دقت مقادیر پیش‌بینی شده به محدوده تعیین شده برسد. بعد از کامل شدن گام‌های اشاره شد، یک مدل هوشمند با ساختمان چند لایه ساخته خواهد شد که توانایی پیش‌بینی فرایند پیچیده انتقال آلودگی را خواهد داشت.

۲-۲- ساخت مدل بهینه‌سازی با استفاده الگوریتم جستجوی هارمونی^۳

الگوریتم جستجوی هارمونی، یک الگوریتم فرا اکتشافی است که در سال ۲۰۰۱ ارائه شده است و برای بهینه‌سازی مسائل از پدیده موسیقی در یافتن یک حالت خارق العاده از هارمونی الهام گرفته است (Geem et al., 2001, Alikhani et al. 2012). به منظور بهبود پیش‌بینی زمان حرکت آلودگی، الگوریتم GMDH با روش HS تلفیق شد تا امکان بهبود پیش‌بینی فراهم شود. بر این اساس یک کد کامپیوتری برای ساخت مدل بهینه‌سازی HS بر مبنای گام‌های زیر در نرم‌افزار MATLAB نوشته شد

گام اول: مقداردهی اولیه پارامترهای روش HS این گام شامل مقداردهی nvar: تعداد متغیرها (۶)، HMS: اندازه حافظه هارمونی (۱۰۰)، HMCR: سرعت بررسی حافظه هارمونی (۰/۹۶)، PAR: سرعت تنظیم قطعات (۰/۴۴)، BW: پهنای باند (۰/۲) و NI: تعداد تکرار (۱۰۰۰۰). مقادیر ذکر شده مربوط به متغیرهای استفاده شده در این پژوهش است.

گام دوم: تشکیل حافظه هارمونی (HM) در این مرحله به تعداد HMS بردار جواب در حافظه هارمونی قرار داده می‌شود به این منظور هر بردار جواب به صورت زیر ایجاد می‌شود (انتخابی تصادفی خواهد بود)

$$X_i^j = LBX_i + \left(\frac{UBX_i - LBX_i}{HMS} \right) \cdot rand(0,1) \quad (6)$$

³ Harmony Search (HS)

با توجه به تعداد نرون‌ها (ترکیبات دوتایی) در هر لایه به همان تعداد خروجی Z وجود دارد (Z_1, Z_2, \dots, Z_m)

$$Z_m = W_0 + w_1 X_1 + w_2 X_2 + \dots + w_r X_r \quad (1 < r < n) \quad (3)$$

معادله ۳ یک معادله جامع برای این روش است. در این پژوهش تعداد ورودی‌ها چهار بود. در الگوریتم GMDH این معادله، به معادله انتقال^۱ معروف است.

گام سوم: برآورد وزن‌ها در هر لایه تخمین ضرایب تابع انتقال به روش حداقل مربعات^۲ است که این ضرایب به صورت معادله ۴ است

$$w_i = (X_i^T X_i)^{-1} X_i Y_{obs} \quad (4)$$

در این معادله

$$Y_{obs} = \{Y_{obs1}, Y_{obs2}, \dots, Y_{obsM}\}^T$$

بردار ضرایب وزنی، $W = \{W_0, W_1, \dots, W_5\}$

بردار مقادیر خروجی مشاهداتی و X بردار جفت متغیرهای ورودی است.

گام چهارم: بهینه‌سازی المان‌ها در هر لایه بیان بهینه‌سازی یک سیستم غیرخطی با آنالیز رگرسیون به داده‌های آموزش صورت می‌گیرد.

در این گام بهترین متغیرها و ترکیب دوتایی آنها در هر لایه شناسایی و انتخاب می‌شوند. بعد از محاسبه خروجی‌ها Z_i ، به منظور تعیین ساختار شبکه،

لازم است از یک معیار خارجی به منظور انتخاب نرون‌هایی که توصیف‌کننده هدف می‌باشند، استفاده نمود. این معیار به صورت معادله زیر محاسبه می‌شود

$$\Delta p^2 = \frac{\sum_{i=1}^N (V_{pca}(i) - \bar{V}_{pca})^2}{\sum_{i=1}^N (V_{pobs}(i) - \bar{V}_{pobs})^2} \quad (5)$$

که در آن

V_{pobs} خروجی مشاهداتی، V_{pca} خروجی محاسباتی، \bar{V}_{pobs} میانگین خروجی‌های مشاهداتی و \bar{V}_{pca} میانگین خروجی‌های محاسباتی است.

¹ Transfer Function (TF)

² Least Square Error (LSE)

ارزیابی می تواند از جنبه های مختلف نظیر حداکثر بازده، حداقل هزینه و حداقل خطا صورت بگیرد. اگر بردار جواب جدید، x^{new} از بدترین بردار جواب در حافظه هارمونی، از نظر مقدار تابع هدف بهتر باشد، هارمونی جدید به حافظه هارمونی اضافه شده و بدترین هارمونی از حافظه خارج می شود.

گام پنجم: بررسی شرط توقف

معمولاً در الگوریتم جستجوی هارمونی تعداد تکرار (جستجو) به عنوان شرط توقف در نظر گرفته می شود که پارامتر مربوط به آن (NI) در مرحله اول مقداردهی می شود.

اگر تعداد هارمونی های ایجاد شده به NI رسیده باشد محاسبات متوقف می شوند و در غیر این صورت مراحل سوم تا پنجم تکرار خواهند شد (شکل ۱).

که در آن X_i^j بیانگر متغیر i ام در بردار جواب j ام می باشد. پارامترها اعداد یکنواخت تصادفی بوده و مقداری بین صفر و یک خواهد داشت. مقدار تابع به ازای هر بردار جواب نیز محاسبه می شود تا حافظه هارمونی بر اساس آن مرتب شود.

گام سوم: ایجاد یک هارمونی جدید از حافظه هارمونی.

در مرحله تولید هارمونی جدید وقتی حافظه هارمونی آماده شد یک بردار جواب جدید (هارمونی)، $X = \{X_1^{new}, X_2^{new}, \dots, X_N^{new}\}$ با استفاده از حافظه هارمونی ایجاد می شود به عبارتی برای هر متغیر یک مقدار جدید انتخاب می شود.

گام چهارم: به روز رسانی حافظه هارمونی.

الگوریتم جستجوی هارمونی بردار جدید، x^{new} که در مرحله سوم ایجاد شده است را با استفاده از تابع هدف ارزیابی می کند. این

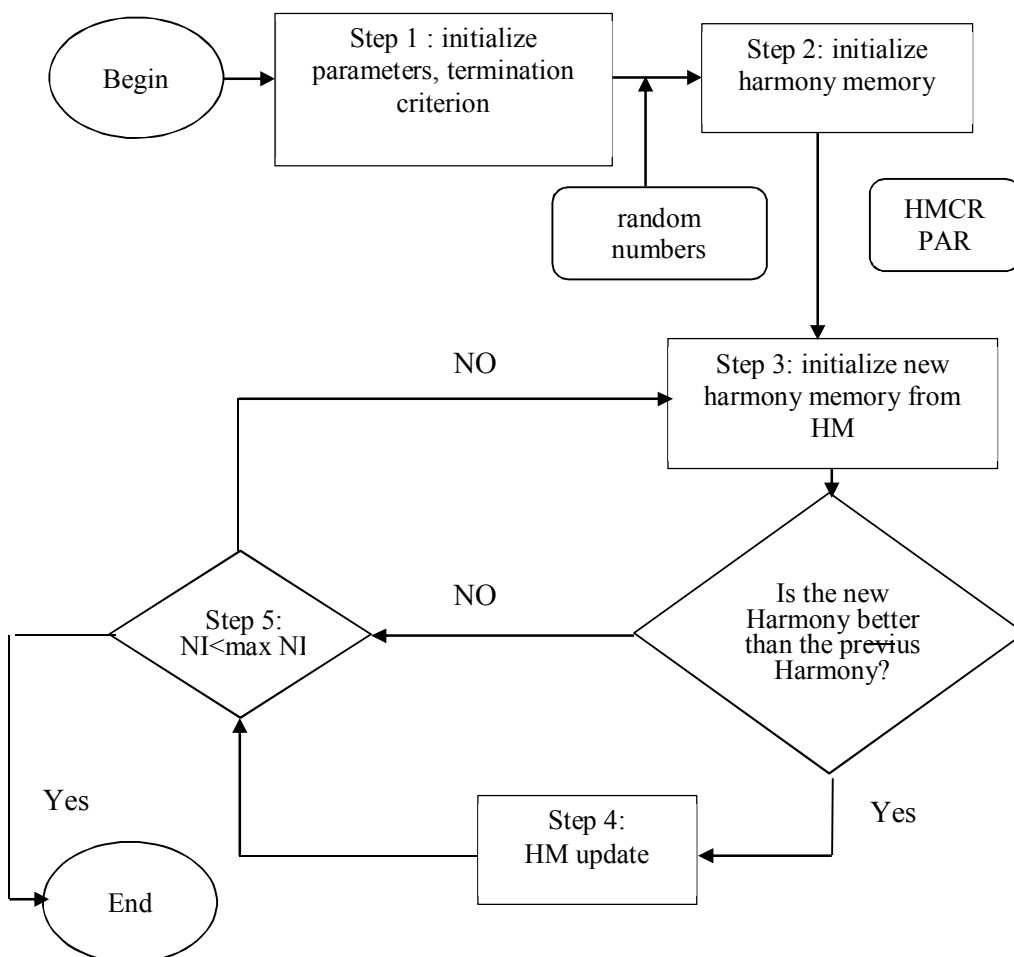


Fig. 1. Flowchart of intelligent HS model
شکل ۱- فلوچارت روش HS



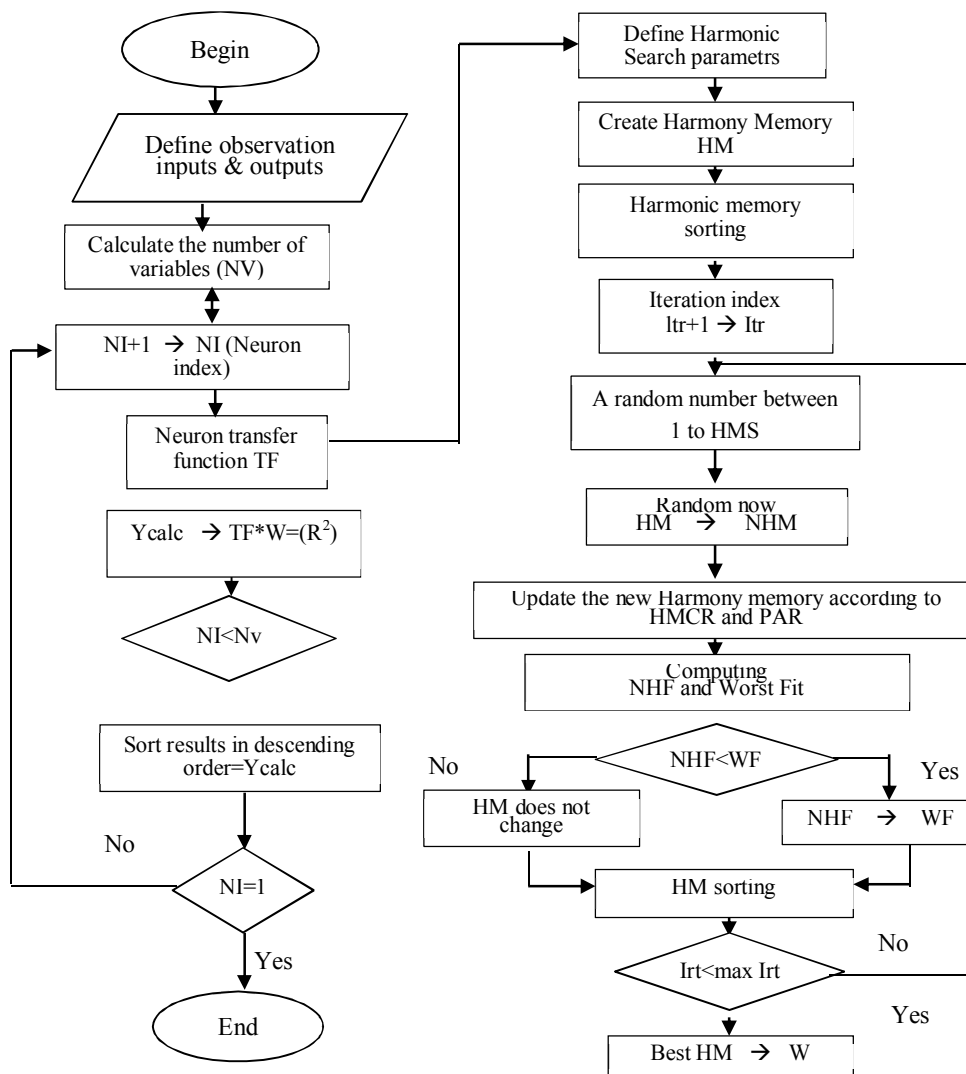


Fig. 2. Flowchart of intelligent GMDH-HS model
 شکل ۲- فلوچارت روش هوشمند تلفیقی GMDH-HS

۲-۴- داده‌های واقعی استفاده شده در مدل‌ها

در پژوهش حاضر از داده‌های برداشت شده در ۳۸۰ بازه مربوط به رودخانه مختلف در آمریکا که توسط سازمان زمین‌شناسی آمریکا^۱ جمع‌آوری شده است، استفاده شد. ۲۱۶ داده از مجموع داده‌ها توسط جابسون در سال برای استخراج رابطه رگرسیونی به‌منظور پیش‌بینی سرعت حرکت آلودگی استفاده شده است (Jobson, 2002) که در این پژوهش نیز همین داده‌ها به‌عنوان ورودی به مدل‌ها استفاده شد. این داده‌ها مربوط به رهاسازی

¹ U.S. Geological Survey

۲-۳- ساخت مدل هوشمند تلفیقی GMDH-HS

در این پژوهش به‌منظور ارتقای مدل GMDH و بهینه‌سازی اهداف، مدل ساخته شده GMDH با روش دانش‌محور جستجوی هارمونی (HS)، ترکیب شده و روش هوشمند تلفیقی GMDH-HS ساخته شد. به‌منظور تلفیق دو روش GMDH و HS، وزن‌های موجود در GMDH که قبلاً با روش حداقل مربعات به‌دست آمده‌اند، توسط روش جستجوی هارمونی برآورد و سپس وارد الگوریتم GMDH شد و سرعت پیک انتقال آلودگی پیش‌بینی شد. ساختار کد روش هوشمند تلفیقی GMDH-HS مطابق فلوچارت شکل ۲ است.



جدول ۱- رابطه پارامترهای آماری استفاده شده برای ارزیابی نتایج

Table 1. Evaluation criteria equations

Criteria	Range	Value	Equations*
MSRE	0 : ∞	0 < MSRE < 0.5	$\frac{1}{n} \sum_{i=1}^n \left(\frac{V_o - V_c}{V_o} \right)^2$
CE	-∞ : +1	CE > 0.8 CE > 0.9	$1 - \frac{\sum_{i=1}^n (V_o - V_c)^2}{\sum_{i=1}^n (V_o - \bar{V}_o)^2}$
MPRE	0 : ∞	0	$\frac{100}{n} \sum_{i=1}^n \text{abs} \left(\frac{V_o - V_c}{V_o} \right)$
RB	-∞ : +∞	0.5 < RB < 0.5	$\frac{1}{n} \sum_{i=1}^n \left(\frac{V_o - V_c}{V_o} \right)$
RMSE	0 : +∞	0 < RMSE < 0.5	$\sqrt{\frac{\sum_{i=1}^n (V_o - V_c)^2}{n}}$

* V_o: Observed values; V_s: Model data; \bar{V}_o : Average observed data

پژوهشگران بسیاری مورد استفاده قرار گرفته است. هدف از این مدل سازی، استفاده از مقادیر گذشته این سری زمانی تا زمان t، برای پیش بینی مقادیر این سری زمانی در گام زمانی t⁺ است.

۳-۲- شاخص های ارزیابی

برای ارزیابی نتایج به دست آمده از این مدل سازی از چندین معیار آماری استفاده شد. این معیارها شامل مربعات میانگین خطای نسبی^۱، درصد مجموع مطلق میانگین خطای نسبی^۲، شاخص تولید مناسب پاسخ ها (CE)، انحراف نسبی (RB) و مجذور میانگین خطای مربعات^۳ می باشند (جدول ۱).
که در آن

V_o داده های مشاهداتی، \bar{V}_o میانگین داده های مشاهداتی، V_c داده های مدل و n تعداد کل داده ها می باشند. MSRE و MPRE شاخصی برای بیان دقت نسبی مدل سازی است، ضریب CE بیان کننده شاخصی برای تولید مناسب پاسخ ها توسط مدل است و RB بیان می دارد که پاسخ های تولید شده نسبت به مقادیر مطلوب کمتر تخمین زده شده اند یا بیشتر. دامنه تغییرات MSRE بین صفر تا ∞ است که از این مقدار، مقادیر بین صفر تا ۰/۵ قابل قبول می باشند. دامنه تغییرات MPRE نیز بین صفر تا ∞ است و مقادیر

ماده ردیاب در بازه های طبیعی و اندازه گیری های مرتبط است. نتایج پیش بینی مدل های GMDH و GMDH-HS با نتایج حاصل از رابطه تجربی ارائه شده توسط جابسون، با هم مقایسه شد

$$V_p = 0.094 + 0.0143 (D_a')^{0.919} \times (Q_a')^{0.469} \times S^{0.159} \times \frac{Q}{D_a} \quad (۶)$$

که در این معادله

V_p سرعت پیک آلودگی، V_{mp} سرعت پیک محتمل آلودگی، D_a'
سطح نفوذ بی بعد، Q_a' دبی نسبی بی بعد، S شیب رودخانه و $\frac{Q}{D_a}$
نسبت دبی به سطح نفوذ است.

۳-۳ نتایج و بحث

۳-۱-۱- صحت سنجی مدل های GMDH و GMDH-HS

قبل از کاربرد مدل GMDH و GMDH-HS توسعه داده شده، صحت کارکرد آن توسط یک تابع استاندارد سری زمانی دیفرانسیلی Mackey-Glass که به شکل معادله ۷ است، بررسی شد

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1 + X^{10}(t-\tau)} - bx(t) \quad (۷)$$

که در آن

مقادیر a و b به ترتیب ۰/۲ و ۰/۱ می باشد. پیش بینی مقادیر آینده این سری زمانی، یک مسئله پایه در مدل سازی است که توسط

¹ Mean Squared Relative Error (MSRE)
² Mean Percent Reletive Error (MPRE)
³ Root Mean Square Error (RMSE)

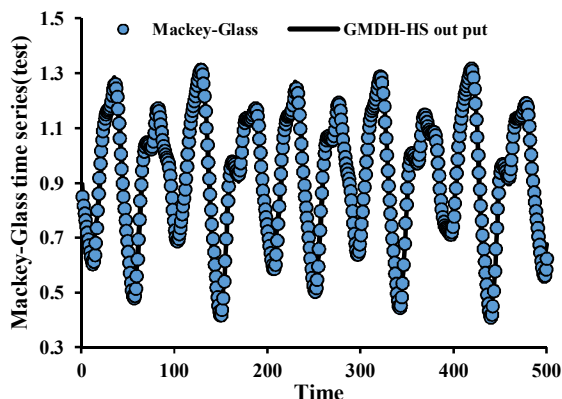


Fig. 3. Performance of GMDH for Mackey-Glass time series prediction; data training

شکل ۳- نمودار مقایسه مقادیر پیش‌بینی شده سری زمانی Mackey-Glass توسط مدل GMDH در مرحله آموزش با مقادیر واقعی

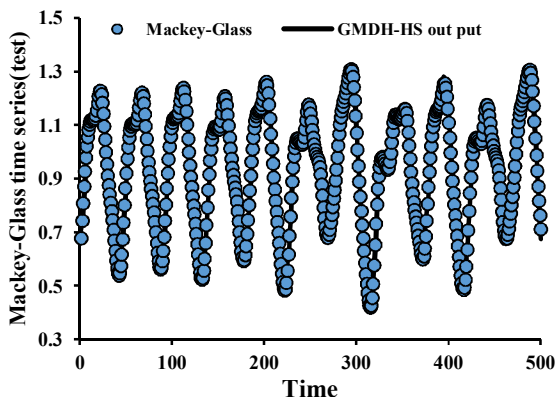


Fig. 4. Performance of GMDH for Mackey-Glass time series prediction; data testing

شکل ۴- نمودار مقایسه مقادیر پیش‌بینی شده سری زمانی Mackey-Glass توسط مدل GMDH در مرحله آزمون با مقادیر واقعی

استفاده از تابع محرک غیر خطی به ترتیب در مرحله آموزش و آزمون نمایش داده شده است. رویه مشابهی برای مدل GMDH-HS طی شد. نتایج مدل‌سازی پیش‌بینی سرعت پیک آلودگی توسط مدل GMDH-HS در شکل‌های ۸ و ۹ ارائه شده است. مقادیر شاخص ارزیابی مدل GMDH-HS نیز در جدول ۴ آورده شده است.

با توجه به نتایج جدول مدل ۴، GMDH-HS به‌خوبی قادر است سرعت پیک آلودگی در رودخانه‌های مختلف را برآورد نماید.

RB بین $-\infty$ تا ∞ است که مقادیر منفی بیانگر تخمین بیش از مقدار واقعی و مقادیر مثبت بیانگر تخمین کمتر از مقدار واقعی است. همچنین دامنه تغییرات ضریب CE بین $-\infty$ تا $+1$ بوده است که مقادیر بالای $0/8$ قابل قبول و مقادیر بالای $0/9$ بسیار رضایت‌بخش است.

۳-۳- صحت‌سنجی مدل GMDH و GMDH-HS

نتایج صحت‌سنجی مدل‌های توسعه داده شده GMDH و GMDH-HS توسط تابع Mackey-Glass در شکل‌های ۳ و ۴ و جدول ۲ ارائه شده است. با توجه به مقدار شاخص‌های ارزیابی ارائه شده در جدول ۲ کارکرد مدل توسعه داده شده در دامنه قابل قبول قرار دارد و قابل تأیید است.

۴-۳- نتایج کاربرد مدل‌های GMDH و GMDH-HS در

پیش‌بینی سرعت پیک آلودگی در رودخانه‌ها

به‌منظور اجرای مدل‌های ذکر شده، متغیرهای D'_a سطح نفوذ بی‌بعد، Q'_a دبی نسبی بی‌بعد، S شیب رودخانه و $\frac{Q}{D_a}$ نسبت دبی به سطح نفوذ به‌عنوان ورودی مدل‌های مورد استفاده قرار گرفته‌اند. در مدل‌سازی ۷۸ درصد داده‌ها برای آموزش شبکه و ۲۲ درصد داده‌ها برای آزمون مدل‌های توسعه داده شده مورد استفاده قرار گرفت. برای آموزش سیستم، مجموعه داده‌هایی به‌عنوان ورودی انتخاب شدند که دربرگیرنده تمام خصوصیات (بیشینه، کمینه و غیره) متغیرهای ورودی بودند.

در شکل ۵ شماتیکی از ساختار شبکه مدل پیش‌بینی GMDH این پژوهش ارائه شده است که در آن، نرون‌های منتخب به‌همراه CE آن‌ها در هر لایه قابل مشاهده است. نرون‌های تیره‌رنگ، نرون‌های منتخب و نرون‌های سفید، نرون‌های حذف شده در هر لایه می‌باشند.

پس از اینکه شبکه GMDH برای پیش‌بینی سرعت پیک آلودگی با استفاده از تابع محرک غیر خطی، توسعه و آموزش داده شد و ماتریس‌های وزنی در نرون‌های منتخب در هر لایه تهیه شد، با استفاده از داده‌های آزمون، کارکرد مدل GMDH بررسی شد. در جدول ۳ پارامترهای ارزیابی این مدل بیان شده است و در شکل‌های ۶ و ۷ نتایج حاصل از مدل GMDH توسعه داده شده با

جدول ۲- مقادیر شاخص‌های ارزیابی مدل‌های GMDH و GMDH-HS در مرحله صحت‌سنجی

Table 2. Performance indices of GMDH1 and GMDH-HS for Mackey-Glass time series prediction

Performance indices	CE	RMSE	MSRE	MPRE	RB
Training (GMDH)	0.9807	0.0320	0.0019	3.1557	-0.0016
Test (GMDH)	0.9844	0.0276	0.0013	2.5782	-0.0013
Training (GMDH-HS)	0.9809	0.0319	0.0017	3.1643	-0.0006
Training (GMDH-HS)	0.9786	0.0323	0.0016	3.0896	-0.00024

جدول ۳- شاخص‌های ارزیابی در دوره آموزش و آزمون در مدل GMDH

Table 3. Performance indices of GMDH, in training and testing the data

Performance indices	CE	RMSE	MSRE	MPRE	RB
Training period	0.8809	0.0723	0.6986	51.1416	-0.2960
Testing period	0.9328	0.0145	0.0116	7.6995	-0.0111

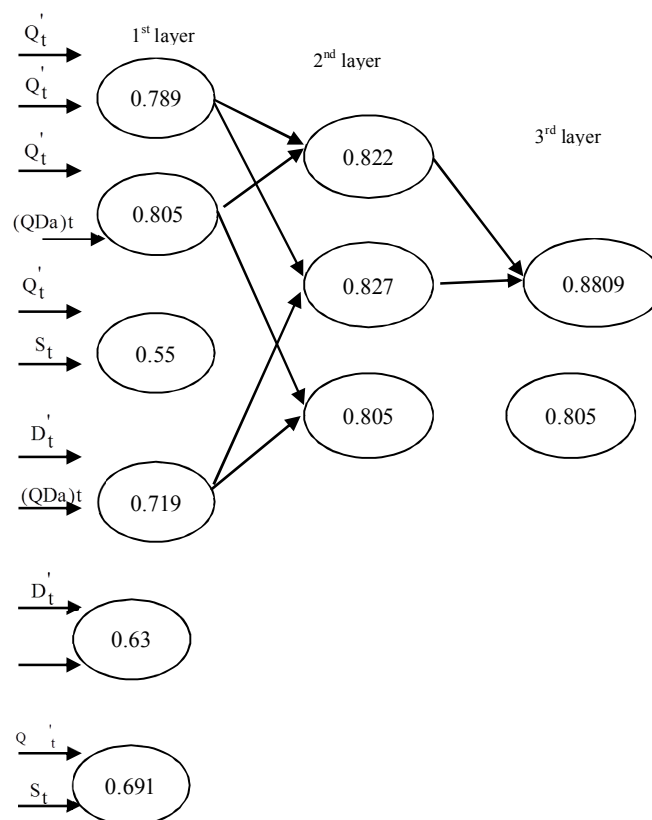


Fig. 5. GMDH structure for peak velocity of pollutant forecasting model construction

شکل ۵- شماتیکی از ساختار شبکه مدل GMDH توسعه داده شده برای پیش‌بینی سرعت پیک آلودگی



جدول ۴- شاخص‌های آماری دوره آموزش و آزمون مدل GMDH-HS

Table 4. Performance indices of GMDH-HS during training and testing period

Performance Indices	CE	RMSE	MSRE	MPRE	RB
Testing period	0.8406	0.0837	0.9612	59.9723	-0.4027
Training Period	0.9033	0.0175	0.0187	9.7868	-0.0484

کلیه شاخص‌های ارزیابی در دامنه قابل قبولی قرار دارند اما در مقایسه مدل‌های GMDH و GMDH-HS دقت نتایج GMDH-HS پایین‌تر است. این کاهش دقت ممکن است به دلایل پیچیدگی فرایند آموزش در هنگام ادغام دو شبکه به‌وجود آمده

همانطور که در شکل ۶ و ۷ نیز نشان داده شده است، اختلاف میان سرعت پیک‌های محاسباتی و مشاهداتی ناچیز است که این میزان خطای کم در پارامترهای ارزیابی محاسبه شده نیز به وضوح مشخص است.

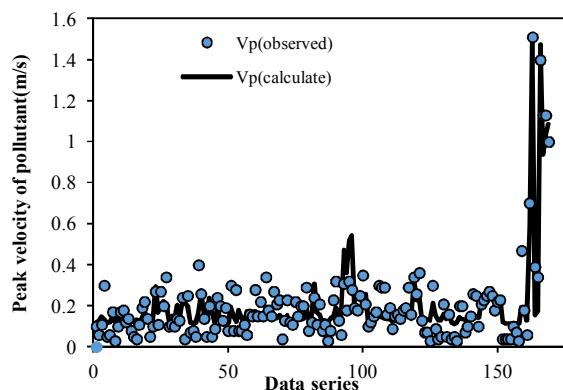


Fig. 8. Measured versus predicted peak velocity of pollutant using GMDH-HS for training period

شکل ۸- مقایسه سرعت پیک حرکت آلودگی محاسباتی (مدل GMDH-HS) و مشاهداتی در دوره آموزش

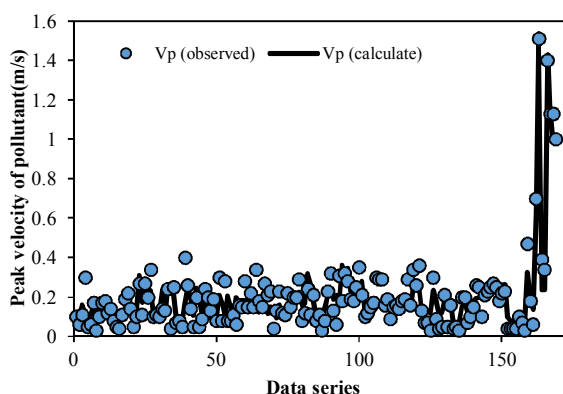


Fig.6. Measured versus predicted peak velocity of pollutant using GMDH for training period

شکل ۶- مقایسه سرعت پیک حرکت آلودگی محاسباتی (مدل GMDH) و مشاهداتی در دوره آموزش

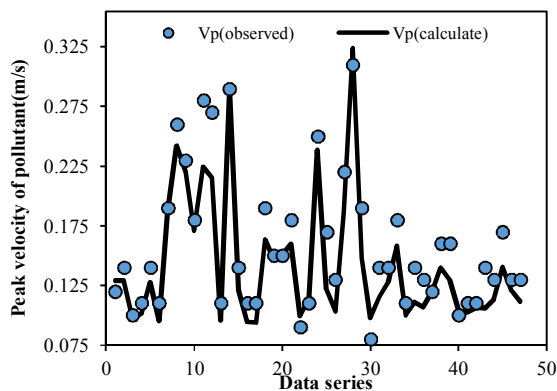


Fig. 9. Measured versus predicted peak velocity of pollutant using GMDH-HS for testing period

شکل ۹- مقایسه سرعت پیک حرکت آلودگی محاسباتی (مدل GMDH-HS) و مشاهداتی در دوره آزمون

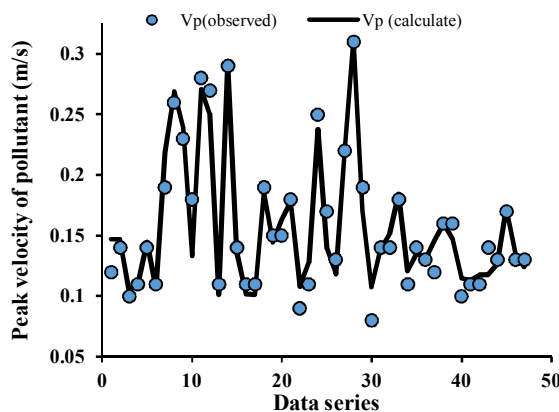


Fig. 7. Measured versus predicted peak velocity of pollutant using GMDH for testing period

شکل ۷- مقایسه سرعت پیک حرکت آلودگی محاسباتی (مدل GMDH) و مشاهداتی در دوره آزمون



جدول ۵- مقایسه شاخص‌های آماری مدل‌های توسعه داده شده GMDH و GMDH-HS
Table 5. Performance indices of GMDH, GMDH-HS and empirical equation

Performance Indices	CE	RMSE	MSRE	MPRE	RB
GMDH	0.9328	0.0145	0.0116	7.6995	-0.0111
GMDH-HS	0.9033	0.0175	0.0187	9.7868	-0.0484
Empirical equation	0.3802	0.0788	0.0136	0.5412	-0.0318

آلودگی مشاهداتی و برآورد شده توسط رابطه رگرسیونی Jobson و مقادیر پارامترهای ارزیابی این رابطه را نشان می‌دهد. با توجه به مقادیر شاخص‌های ارزیابی و مقایسه آنها برتری روش هوشمند GMDH در پیش‌بینی سرعت انتقال آلودگی نسبت به روش رگرسیونی غیرخطی چند متغیره که برای استخراج معادله ۶ استفاده شد، مشاهده می‌شود.

۴- نتیجه‌گیری

در این پژوهش مدل پیش‌بینی سرعت انتشار آلودگی که به صورت ناگهانی و اتفاقی به رودخانه وارد شده است، بر مبنای روش هوشمند کنترل گروهی داده‌ها (GMDH) و روش هوشمند تلفیقی (GMDH-HS) توسعه داده شد. مرحله آموزش و ارزیابی این پژوهش با استفاده از مجموعه داده‌های برداشت شده در آبراه‌های طبیعی صورت گرفت. با توجه به نتایج به دست آمده، دقت کارکرد هر دو مدل بسیار نزدیک به یکدیگر بود. اما در همه موارد، دقت نتایج به دست آمده از مدل GMDH-HS بیشتر از دقت نتایج به دست آمده از مدل GMDH بود. همچنین سرعت اجرای کد GMDH بسیار بیشتر از کد GMDH-HS بود. در نهایت مشاهده شد که دقت روش‌های هوشمند بسیار بالاتر از معادله رگرسیونی است که این امر نشان‌گر بهبود دقت محاسبه سرعت پیک آلودگی در رودخانه‌ها به میزان ۵۲ درصد تا ۵۵ درصد با استفاده از روش‌های هوشمند است.

References

- Alikhani, K., Qaderi, K. & Ahmadi, M.M. 2012. Predicting the distribution of deposited sediment in Karaj reservoir using various improved harmony search algorithms. *Journal of Soil and Water Conservation Research*, 21(3), 165-184. (In Persian)
- Faramarz, M., Ayubzadeh, S. & Mohammadi, K. 2005. Numerical simulation of distribution and transmission levels of pollution in the river floodplain. *The 5th Iranian Hydraulic Conference*. Shahid Bahonar University of Kerman, Kerman, Iran. (In Persian)

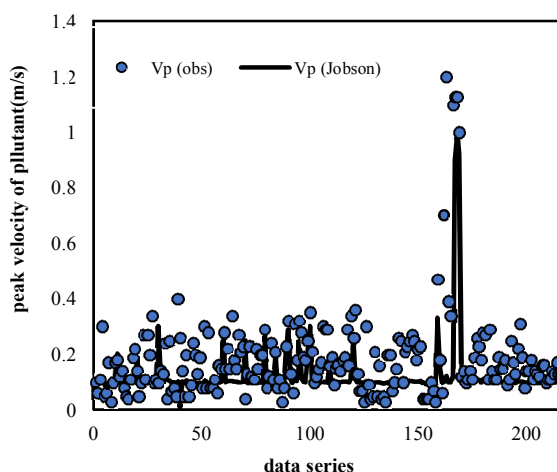


Fig. 10. Jobson equation result versus predicted for peak velocity of pollutant using GMDH-HS method

شکل ۱۰- مقایسه سرعت پیک مشاهداتی و سرعت پیک معادله رگرسیونی ارائه شده توسط جابسون (معادله ۶)

باشد. بنابراین این فرض که تلفیق مدل بهینه‌سازی هوشمند HS به الگوریتم GMDH منجر به افزایش دقت احتمالی پیش‌بینی می‌شود، در مورد برآورد سرعت انتقال آلودگی اثبات نشد و با توجه به زمان اجرای کمتر مدل GMDH نسبت به مدل GMDH-HS، این مدل پیشنهاد می‌شود. در ادامه نتایج خروجی مدل GMDH و GMDH-HS با معادله رگرسیونی ارائه شده توسط جابسون مقایسه شد (Jobson 1995). شکل ۱۰ و سطر سوم جدول ۵ به ترتیب مقایسه سرعت انتقال

- Harvey, E. & Jobson, H. E. 1996. *Prediction of traveltime and longitudinal dispersion in rivers and streams*. U.S. Geological Survey Water-Resources Investigations Report 84-4203, USA.
- Hoseinzadeh, M., Qaderi, K. & Ahmadi, M. M. 2013. Stage-discharge relationship modeling in the river using intelligent methods (GMDH, NF-GMDH). *National Conference Optimization of Water Use*, Gorgan, Iran. (In Persian)
- Ivakhnenko, A.G. 1968. Group method of data handling-rival of method of stochastic approximation. *Sov. Autom. Control*, 13(1), 43-55.
- Jabson, H. E. 1995. *Prediction of traveltime and longitudinal dispersion in rivers and streams*. U.S. Geological Survey Open-File Report 71-0096, USA.
- Julínek, T. & Říha, M. 2017. Longitudinal dispersion in an open channel determined from a tracer study. *Environmental Earth Science*, 76:592. <http://doi.org/10.1007/s12665-017-6913-1>
- Razek, M. & Abouelnour, A. 1972. A statistical methodology for predicting pollutants in a river. *Water Resources Bulletin*, 8(1), 15-23.
- Sanchez, E., Shibata T. & Zadeh, L. A. 1997. *Genetic algorithms and fuzzy logic systems*. World Scientific Publishing Co., Inc. River Edge, N.J., USA.
- Sattar, A. M. A. & Gharabaghi, B. 2015. Gene expression models for prediction of longitudinal dispersion coefficient in streams. *Journal of Hydrology*, 524, 587-596.
- Tsai, T. & Yen, T. 2017. GMDH algorithms applied to turbidity forecasting. *Appl. Water Sci.*, 7 (3), 1151-1160.
- Wiley, J. B. & Messinger, T. 2013. *Estimation of traveltime and longitudinal dispersion in streams in West Virginia*, West Virginia Bureau for Public Health, Office of Environmental Health Services, Scientific Investigations Report 2013-5182, Virginia, USA.
- ZelenaKov, M., Carnogurska, M. & Gargar, I., 2012. *Prediction of pollutant concentration in laborec river station, Slovak Republic*. *Procedia Engineering*, 42, 1474-1473.

