

## تخمین ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی با استفاده از رویکرد هوشمند داده‌محور GMDH

کوروش قادری<sup>1\*</sup>، مریم حسین زاده<sup>2</sup>

تاریخ دریافت: 1394/3/19 تاریخ پذیرش: 1395/8/26

### چکیده

تخمین دقیق ضریب پراکندگی طولی آلاینده‌ها در اغلب مسائل هیدرولیکی و زیست‌محیطی رودخانه مانند مهندسی رودخانه، طراحی آبگیر، مدل‌سازی جریان در دلتاها و ارزیابی ریسک ناشی از تزریق آلاینده‌ها به جریان رودخانه دارای اهمیت خاصی می‌باشد. بررسی تحقیقات اخیر نشان می‌دهد که استفاده از روش‌های داده‌محور باعث افزایش دقت و صحت تخمین ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی شده است. در این تحقیق توانایی و عملکرد روش هوشمند داده‌محور برخورد گروهی با داده‌ها (GMDH) در تخمین ضریب پراکندگی طولی مورد بررسی قرار گرفته است. مجموعه‌ای از 71 داده مشاهداتی از رودخانه‌های مختلف گردآوری شده که 51 داده برای آموزش و 20 داده باقی‌مانده برای آزمون مدل استفاده شده است. متغیرهای هیدرولیکی و هندسی ورودی مدل شامل عمق متوسط جریان (H)، عرض آبراهه (W)، سرعت متوسط طولی جریان (U)، سینوسی آبراهه ( $\sigma$ ) و سرعت برشی متوسط بستر ( $U^*$ ) و متغیر خروجی ضریب پراکندگی طولی ( $K_L$ ) می‌باشند. یک برنامه کامپیوتری در محیط برنامه‌نویسی نرم‌افزار MATLAB برای مدل‌سازی ضریب پراکندگی طولی بر اساس رویکرد هوشمند داده‌محور GMDH تهیه شده است. بررسی عملکرد مدل توسعه داده شده با استفاده از معیارهای آماری  $R^2$ ، RMSE، CC و DR حاکی از عملکرد بسیار بالای مدل GMDH در تخمین ضریب پراکندگی طولی در رودخانه دارد. همچنین مقایسه نتایج روش GMDH با چند رابطه تجربی و چند روش داده‌محور دیگر مانند شبکه‌های عصبی (ANN)، ماشین بردار پشتیبان (SVM) و الگوریتم ژنتیک (GA) بیانگر شناسایی صحیح الگوی حاکم بر فرایند پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی توسط مدل توسعه داده شده می‌باشد. مقادیر شاخص‌های آماری  $R^2$ ، RMSE و CC در مدل GMDH به ترتیب برابر با 0/99، 0/995 به دست آمده است که حاکی از عملکرد مناسب مدل ارائه شده در تخمین ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی می‌باشد

**واژه‌های کلیدی:** روابط تجربی، روش‌های داده‌محور، ضریب پراکندگی، محاسبات نرم، مدل‌سازی، GMDH

### مقدمه

می‌باشد. زمانی که پساب کشاورزی یا فاضلاب صنعتی و خانگی وارد رودخانه‌ها می‌شود، تحت فرایند اختلاط با جریان رودخانه به سمت پایین‌دست انتقال می‌یابد. زمان حرکت آلاینده در رودخانه، نرخ پراکندگی آلودگی، میزان کاهش در غلظت پیک و الگوی نهایی غلظت آلاینده از مسائل مهمی است که می‌بایست پاسخ آن‌ها در حل مسئله انتقال و پراکندگی آلاینده مشخص شوند.

توانایی تخمین و شبیه‌سازی جریان، انتقال آلاینده و دیگر مواد محلول در سامانه رودخانه برای مدیریت و برنامه‌ریزی منابع آب اهمیت زیادی دارد. توانایی رودخانه‌ها در پراکندگی آلاینده‌ها در جریان را در جهت طولی، عرضی و عمودی به ترتیب می‌توان با ضرایب پراکندگی طولی ( $K_L$ )، پراکندگی عرضی ( $K_T$ ) و پراکندگی عمودی ( $K_V$ ) توصیف نمود. پس از اینکه اختلاط در مقطع عرضی کامل شد، فرایند پراکندگی طولی اهمیت بیشتری نسبت به پراکندگی عرضی و عمودی خواهد داشت (Seo and Cheong, .)

رودخانه‌ها نه تنها به عنوان اصلی‌ترین و بزرگ‌ترین منبع آب در دسترس، دارای اهمیت فراوانی برای مدیران و بهره‌برداران منابع آبی می‌باشند بلکه با استفاده از فرایند پخشیدگی، اختلاط و قدرت خودپالایی خود، باعث کاهش و دفع آلاینده‌ها و کاهش خطرات زیست‌محیطی می‌شوند (پورآبادی و همکاران، 1392). یکی از مسائل مهم در استفاده از منابع آب سطحی و بویژه رودخانه‌ها، آلودگی آب‌های سطحی توسط پساب کشاورزی و فاضلاب صنعتی

1- استادیار بخش مهندسی آب، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان

2- فارغ التحصیل کارشناسی ارشد مهندسی منابع آب، دانشگاه شهید باهنر کرمان، کرمان

\* - نویسنده مسئول: (Email: kourosqaderi@uk.ac.ir)

ضریب پراکندگی طولی توسط محققان مختلفی ارائه شده است. در صورت در دسترس بودن داده‌های واقعی، به سادگی می‌توان این ضریب را تعیین نمود. محاسبه دقیق این ضریب با استفاده از داده‌های واقعی به دلیل پیچیدگی‌های خاص اندازه‌گیری ضریب پراکندگی در فرایند انتقال آلودگی به ندرت انجام می‌شود. Seo and Baek., (2004) در صورت عدم وجود داده‌های واقعی می‌بایست از دیگر روش‌ها برای تعیین ضریب پراکندگی طولی استفاده نمود (Kashefipour and Falconer., 2002). در جدول 1 تعدادی از این روابط که بر مبنای مطالعات تجربی، رگرسیون‌های خطی و غیرخطی و آنالیز ابعادی متغیرهای تاثیرگذار بر ضریب پراکندگی طولی به دست آمده‌اند ارائه شده است. در این جدول H: عمق جریان، U: سرعت جریان،  $U^*$ : سرعت برشی، W: عرض کانال،  $S_i$ : سینوسیته مسیر جریان و P: عدد پیکلت می‌باشد.

(1998). بنابراین لازمه آگاهی از چگونگی انتقال و پراکندگی آلاینده‌ها، تعیین مقدار دقیق ضریب پراکندگی طولی ( $K_x$ )، در رودخانه‌ها می‌باشد (Cheong et al., 2007). تیلور در سال 1954، معادله یک بعدی انتقال و پراکندگی را به صورت رابطه 1 ارائه نمود:

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = K_x \frac{\partial^2 C}{\partial x^2} \quad (1)$$

که C: متوسط غلظت مواد محلول یا آلودگی مقطع، t: زمان، x: فاصله طولی از نقطه تزریق مواد محلول یا آلودگی در جهت جریان، U: سرعت متوسط طولی جریان و  $K_x$ : ضریب پراکندگی طولی می‌باشد. در استخراج این معادله فرض شده که ضریب پراکندگی عرضی و عمقی قابل صرف‌نظر کردن می‌باشد. همچنین مقدار ضریب پراکندگی طولی در طول زمان و مکان تغییر نمی‌کند (Toprak and Cigizoglu., 2008).

روش‌های تئوری و تجربی زیادی برای تخمین مقدار صحیح

جدول 1- روابط تخمین ضریب پراکندگی طولی ارائه شده توسط محققان مختلف

شماره	رابطه $K_x$	نام رابطه
(2)	$0.58 \left( \frac{H}{U_*} \right) UW$	کوئین و کیفر (1974)
(3)	$0.11 \left( \frac{U^2 W^2}{HU_*} \right)$	فیشر (1975)
(4)	$0.18 \left( \frac{U_*}{U} \right)^{1.5} \left( \frac{U^2 W^2}{HU_*} \right)$	لیو (1977)
(5)	$2 \left( \frac{W}{H} \right)^{1.5} HU_*$	ایواسا و آیا (1991)
(6)	$7.428 + 1.775 \left( \frac{U}{U_*} \right)^{1.572} \left( \frac{W}{H} \right)^{0.62} HU_*$	توکلی‌زاده و کاشفی‌پور (1385)
(7)	$0.55 \left( \frac{WU_*}{H^2} \right)$	لی و همکاران (1998)
(8)	$5.915 \left( \frac{U}{U_*} \right)^{1.428} \left( \frac{W}{H} \right)^{0.62} HU_*$	سئو و چانگ (1998)
(9)	$0.15 \left( \frac{U}{U_*} \right)^2 \left( \frac{W}{H} \right)^{1.66} \left( \frac{HU_*}{8 \varepsilon_t} \right)$	دنگ و همکاران (2001)
(10)	$10.612 \left( \frac{U}{U_*} \right) HU_*$	کاشفی‌پور و فالکنر (2002)
(11)	$2 \left( \frac{U}{U_*} \right)^{1.25} \left( \frac{W}{H} \right)^{0.96} HU_*$	سهای و دوتا (2009)
(12)	$0.583 \left( \frac{U}{U_*} \right)^{0.562} \left( \frac{W}{H} \right)^{1.287} HU_*$	چانگ و سئو (2003)
(13)	$43.928 \left( \frac{U}{U_*} \right)^{0.845} \left( \frac{W}{H} \right)^{0.217} P^{-0.042} S_i^{-1.6981} HU_*$	چانگ و سئو (2007)
(14)	$0.41 \left( \frac{U}{U_*} \right)^{0.98} \left( \frac{W}{H} \right)^{0.663} P^{-0.772} S_i^{2.226} HU_*$	سهای (2012)

اغلب روابط ارائه شده بر مبنای فرضیات و شرایط خاص **انهار** طبیعی مورد مطالعه به دست آمده‌اند. به همین دلیل روابط به دست آمده تنها برای همان مکان و شرایط اعتبار دارند (Sahay., 2011). بنابراین

برای تخمین مقدار ضریب پراکندگی طولی با استفاده از روابط ارائه شده در جدول 1، احتیاج به مشخصات هیدرولیکی و هندسی محدوده مورد مطالعه می‌باشد (Azamathulla and Ghani., 2011).

شده است. همچنین نتایج حاصل از روابط تجربی و روش GMDH با یکدیگر مقایسه شده و دقت روش‌های مورد استفاده در این تحقیق با استفاده از داده‌های اندازه‌گیری شده مورد ارزیابی قرار گرفته است. داده‌های مورد استفاده در این تحقیق مشابه با داده‌های استفاده شده توسط دیگر محققان می‌باشد (Sahay., 2011; Tayfur and Singh., 2005; Chong and Seo., 2003 and Adrash., 2010) در تحقیق ساهای، آدرش، چانگ و سئو و تیفور و سینگ می‌باشد. در ادامه روش GMDH، ساختار آن، الگوریتم و چگونگی آموزش پارامترهای مدل و سپس داده‌های مورد استفاده در این تحقیق و در نهایت نتایج مدل‌سازی ضریب پراکندگی طولی با استفاده از مدل GMDH و مقایسه نتایج به دست آمده با دیگر نتایج دیگر محققان ارائه شده است.

### مواد و روش‌ها

برای مدل کردن یک سیستم می‌بایست روابط ریاضی صریح بین متغیرها بطور دقیق معلوم باشد. در اغلب سیستم‌ها این روابط نامعلوم هستند و استخراج چنین مدل‌سازی صریحی بسیار مشکل است. در این حالت می‌توان از روش‌های محاسبات نرم<sup>2</sup> که داده محور هستند و محاسبات را در شرایط غیردقیق انجام می‌دهند، استفاده نمود (Sanchez et al., 1997). اجزای اصلی محاسبات نرم شامل منطق فازی، شبکه‌های عصبی و الگوریتم‌های تکامل‌گرا دارای توانایی‌های زیادی در شناسایی سیستم‌های غیرخطی و مسائل کنترلی دارند. یکی از انواع این روش‌ها، روش هوشمند داده‌محور GMDH می‌باشد که یک رویکرد خودسازماندهی<sup>3</sup> داده بوده که به تدریج مدل‌های پیچیده‌تری در طول ارزیابی عملکرد مجموعه داده‌های ورودی و خروجی تولید می‌کند. این روش برای اولین بار توسط ایواخنکو بصورت یک روش تجزیه و تحلیل چند متغیره برای مدل‌سازی و شناسایی سیستم‌های پیچیده مطرح شد (Ivakhnenko., 1968). ایده اصلی GMDH به صورت ساختن یک تابع تحلیلی مبتنی بر یک شبکه پیش‌رونده بر اساس توابع انتقالی دو جمله‌ای می‌باشد. ساختار آن به صورت یک شبکه چندلایه پیش‌رونده می‌باشد که ساختاری مشابه با شبکه‌های عصبی مصنوعی دارد با این تفاوت که ساختار این شبکه، تعداد نرون‌ها در هر لایه و تعداد لایه‌ها از ابتدا معلوم نبوده و در طول فرایند آموزش تعیین می‌شود. در روش GMDH، طراح نقش بسیار کمی در تعیین بهترین توپولوژی و ساختار شبکه دارد و این خود شبکه است که به مرور ساختار خود را بازسازی می‌کند. آموزش پارامترهای این شبکه براساس رویکرد تخمین حداقل مربعات انجام می‌شود (Ivakhnenko and Ivakhnenko., 1995). GMDH

استفاده از مدلی که محدود به شرایط خاصی نباشد و با دقت بالا بتواند در مسائل عمومی‌تر کاربرد داشته باشد اجتناب‌ناپذیر است. در چند دهه اخیر استفاده از روش‌های هوشمند داده‌محور در زمینه‌های مختلف علمی و تحقیقاتی و به ویژه در مهندسی منابع آب گسترش یافته است. در تحقیقی از روش شبکه‌های عصبی برای تخمین مقادیر Kx استفاده شده است (Tayfur and Singh., 2005; Toprak and Cigizoglu., 2008). نتایج حاکی از دقت بالای روش شبکه‌های عصبی در تخمین مقادیر Kx بوده است. همچنین دیگر محققان از شبکه‌های عصبی با الگوریتم‌های مختلف آموزش برای تخمین ضریب پراکندگی طولی بهره گرفتند (Rowinski and Piotrowski., 2008). ریاحی مدوار و ایوب‌زاده (1387) از سیستم استنتاج فازی-عصبی انطباقی برای تخمین ضریب پراکندگی طولی استفاده نمودند. استفاده از مدل استنتاج فازی-عصبی انطباقی دقت تخمین این ضریب را بهبود بخشیده است. در تحقیق دیگری مدلی بر مبنای رویکرد برنامه‌ریزی ژنتیک برای تخمین ضریب پراکندگی طولی در رودخانه‌ها ارائه شده است (Azamathulla and Ghani., 2011). مدلی بر مبنای دو روش ماشین بردار پشتیبان (SVM) و برنامه‌ریزی ژنتیک (GP) برای تخمین ضریب پراکندگی طولی توسعه داده شده است (Adrash., 2010). نتایج حاکی از توانمندی دو روش فوق در مقایسه با شبکه‌های عصبی برای تخمین ضریب پراکندگی طولی بوده است. در تحقیق دیگر، ساهای، مقادیر K را با استفاده از شبکه‌های عصبی محاسبه نمود. مقادیر معیارهای آماری این تحقیق در مقایسه با نتایج ارائه شده توسط دیگر محققان بیانگر عملکرد بالای روش مورد استفاده بوده است (Sahay., 2011). مجموعه داده‌های مورد استفاده در این تحقیق مشابه با تحقیقات دیگر محققان می‌باشد. (Cheong and Seo., 2003; Adrash., 2010; Sahay., 2011). متغیرهای ورودی در تحقیقات فوق شامل عمق متوسط جریان (H)، عرض آبراهه (W)، سرعت متوسط طولی جریان (U) و سرعت برشی متوسط بستر (U\*) بوده اما در تحقیق (Sahay., 2011)، متغیر سینوسیتی آبراهه ( $\sigma$ ) نیز لحاظ شده است. لحاظ کردن متغیر سینوسیتی آبراهه به عنوان متغیر ورودی، باعث ارتقای عملکرد شبکه‌های عصبی در تخمین مقادیر ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌ها شده است.

هدف از انجام این تحقیق تخمین صحیح و دقیق ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی با استفاده از یک رویکرد هوشمند داده‌محور به نام روش برخورد گروهی با داده‌ها (GMDH<sup>1</sup>) می‌باشد. به همین منظور یک کد کامپیوتری براساس روش GMDH در محیط برنامه‌نویسی نرم‌افزار MATLAB تهیه

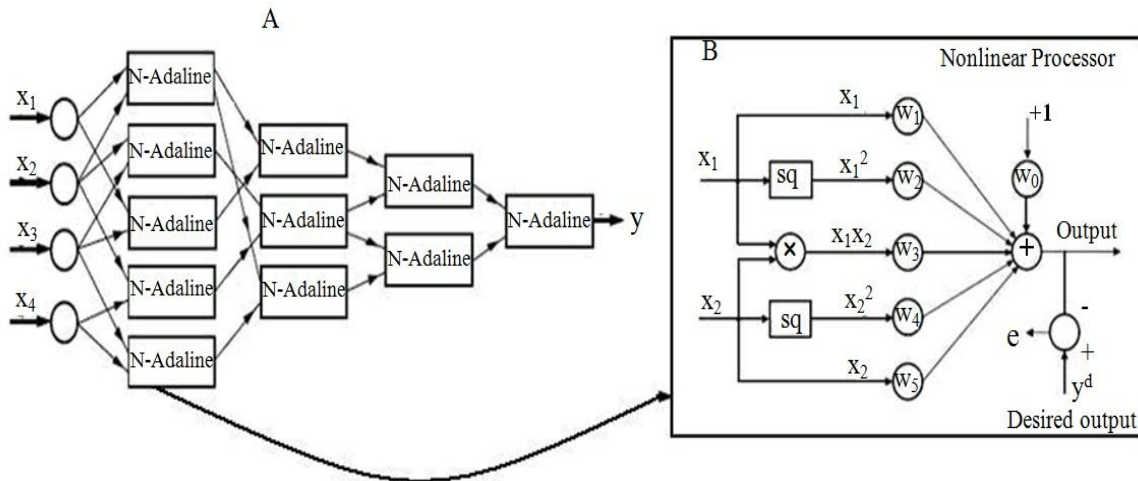
که  $y$  خروجی مطلوب،  $\hat{y}$  خروجی محاسباتی،  $\hat{y}$  میانگین خروجی های محاسباتی و  $\bar{y}$  میانگین خروجی های مطلوب است. برای تعیین ساختار شبکه ابتدا متغیرهای ورودی به مدل انتخاب می گردند. بر حسب نوع مسئله داده ها به صورت تصادفی یا به صورت متوالی به دو گروه داده های آموزش (واسنجی) و آزمون (تست) تقسیم می شوند. سپس برای تمام ترکیبات دوتایی از متغیرهای ورودی،  $N$ -آدالین (نرون) ساخته می شود. بر اساس طبیعت مسئله مورد بررسی نوع تابع محرک برای هر نرون انتخاب می گردد. در مرحله بعد تخمین ضرایب وزنی برای تمام  $N$ -آدالین ها با روش تخمین حداقل مربعات  $(LSE)$  و معیار فیلتراسیون انجام می شود. معیار فیلتراسیون محاسبه شده برای تمام  $N$ -آدالین ها با مقدار آستانه تعیین شده در هر لایه مقایسه می گردد. اگر مقدار آستانه در هر نرون بزرگ تر از معیار فیلتراسیون باشد، نرون مربوطه حذف می شود و در غیر این صورت برای تشکیل لایه بعدی مورد استفاده قرار می گیرد. هر نرون در ساختار GMDH یک تابع غیرخطی از ورودی ها را اجرا می کند. این تابع می تواند با توجه به طبیعت حاکم بر مسئله به صورت توابع غیرخطی درجه اول، غیر خطی درجه دوم، سینوسی های پربولیک و ... انتخاب شود. از تابع غیرخطی درجه دوم به صورت رابطه 16 به عنوان تابع محرک یا تابع انتقال در هر نرون استفاده شده است:

کاربردهای موفقیت آمیزی در مهندسی، علوم و اقتصاد شامل شناسایی قوانین فیزیکی، پیش بینی کوتاه مدت فرایندها، شناسایی الگوها در متغیرهای گسسته و پیوسته داشته است (Nariman Zadeh et al., 2002). از این روش برای مدل سازی سری های زمانی (Nikolaev and Hitoshi., 2003) برای پیش بینی قیمت فروش ارتباطات مخابراتی (Hwang., 2006)، قادری و همکاران (1389) برای بهره برداری بهینه از مخازن، حسین زاده (1392) برای مدل سازی رابطه دبی-اشل در رودخانه، ایوانی (1392) برای تخمین نرخ انتقال رسوب در رودخانه و پورنعمت رودسری (1393) برای شبیه سازی فرایند بارش-رواناب استفاده نمودند. به طور کلی کاربردهای زیادی از این روش در مهندسی منابع آب گزارش نشده است (Sumsudin et al., 2010).

### ساختار GMDH

در شکل 1 دیاگرام شماتیک شبکه GMDH همراه با نمایی از ساختار  $N$ -آدالین با تابع محرک چندجمله ای درجه دو نشان داده شده است. در این شکل  $x_1$  و  $x_2$  ورودی ها و  $y$  خروجی می باشد. معیار آستانه برای تعیین ساختار شبکه به صورت رابطه 15 تعریف می شود:

$$R^2 = \frac{\sum_{t=1}^N (\hat{y}(t) - \bar{y})^2}{\sum_{t=1}^N (y(t) - \bar{y})^2} \quad (15)$$



شکل 1- (A): شماتیک از ساختار GMDH و (B): ساختار هر N-adaline (Ivakhnenko and Ivakhnenko, 1995)

ضرایب شش گانه در هر نرون در شبکه از طریق رویکرد حداقل مربعات محاسبه می شوند.

$$Y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 \quad (16)$$

جدول 2- مجموعه داده‌های جمع‌آوری شده مرتبط با ضریب پراکندگی طولی در رودخانه‌های مختلف (Adrash., 2010)

ردیف	رودخانه	عرض W(m)	عمق H(m)	سرعت U(m/s)	سرعت برشی U <sub>s</sub> (m/s)	$\beta$	$\sigma$	$K_x$ واقعی (m <sup>2</sup> /s)
1	Amita River	37	0/81	0/29	0/070	3/82	1	23/2
2	Amita River	42	0/80	0/42	0/069	3/96	1	30/2
3	Antietam Creek, Md.	12/8	0/30	0/42	0/057	3/75	1/40	17/5
4	Antietam Creek, Md.	24/1	0/98	0/59	0/098	3/2	2/25	101/5
5	Antietam Creek, Md.*	11/9	0/66	0/43	0/085	2/89	2/25	20/9
6	Antietam Creek, Md.	21	0/48	0/62	0/069	3/78	1/26	25/9
7	Bayou Anacoco	20	0/42	0/29	0/045	3/66	1/41	13/9
8	Bayou Anacoco	17/5	0/45	0/32	0/024	3/32	1/41	5/8
9	Bayou Anacoco	25/9	0/94	0/34	0/067	3/69	1/41	32/5
10	Bayou Anacoco	36/6	0/91	0/40	0/067	3/86	1/41	39/5
11	Bayou Bartholomew, La	33/4	1/40	0/20	0/031	3/17	2/46	54/7
12	Bear Creek, Colo.*	13/7	0/85	1/29	0/553	2/78	1/08	2/9
13	Chattahoochee River, Ga*	75/6	1/95	0/74	0/138	3/66	1/27	88/9
14	Chattahoochee River, Ga	91/9	2/44	0/52	0/094	3/63	1/57	166/9
15	Clinch River, Va	48/5	1/16	0/21	0/069	3/73	1/25	14/8
16	Clinch River, Va*	28/7	0/61	0/35	0/069	3/85	1/14	10/7
17	Clinch River, Va	57/9	2/45	0/75	0/104	3/16	1/14	40/5
18	Clinch River, Va*	53/2	2/41	0/66	0/107	3/1	1/14	36/9
19	Comite River	13	0/26	0/31	0/044	3/91	1/31	7
20	Comite River	16	0/43	0/37	0/056	3/62	1/31	13/9
21	Comite River, La	15/7	0/23	0/36	0/039	4/22	1/31	69
22	Conococheague Creek, Md.	42/2	0/69	0/23	0/064	4/11	2/25	40/8
23	Conococheague Creek, Md.	49/7	0/41	0/15	0/081	4/8	2/25	29/3
24	Conococheague Creek, Md.*	43	1/13	0/63	0/081	3/64	1/31	53/3
25	Copper Creep, Va.	16/7	0/49	0/2	0/08	3/53	2/54	16/8
26	Copper Creep, Va.	18/3	0/38	0/15	0/116	3/88	2/54	20/7
27	Copper Creep, Va.	16/8	0/47	0/24	0/08	3/58	2/54	24/6
28	Copper Creep, Va.	19/6	0/84	0/49	0/101	3/15	1/26	20/8
29	Difficult Run, Va.	14/5	0/31	0/25	0/062	3/85	1/09	1/9
30	John Day River, Ore.*	25	0/58	1/01	0/14	3/76	1/08	13/9
31	John Day River, Ore.*	34/1	2/47	0/82	0/180	2/62	1/89	65
32	Little Pincy Creek, Md.	15/9	0/22	0/39	0/053	4/28	1/13	7/1
33	Minnesota River	80	2/74	0/03	0/002	3/37	1	22/3
34	Minnesota River	80	2/74	0/14	0/01	3/37	1	34/9
35	Mississippi River, La*	711/2	19/94	0/56	0/041	3/58	1/44	237/2
36	Mississippi River, Mo*	533/4	4/94	1/05	0/069	4/68	1/38	457/7
37	Mississippi River, Mo*	537/4	8/9	1/51	0/097	4/1	1/38	374/1
38	Missouri River	183	2/33	0/89	0/066	4/36	1/35	465
39	Missouri River*	197	3/11	1/53	0/078	4/15	1/35	892
40	Missouri River	201	3/56	1/28	0/084	4/03	1/35	837
41	Monocacy River, Md.*	48/7	0/55	0/26	0/052	4/48	1/28	37/8
42	Monocacy River, Md.*	93	0/71	0/16	0/046	4/88	1/28	41/4
43	Monocacy River, Md.	51/2	0/65	0/62	0/044	4/37	1/28	29/6
44	Monocacy River, Md.	97/5	1/15	0/32	0/058	4/44	1/61	119/8

بدست آمده محاسبه شده و تنها نرون‌هایی که دارای شاخص عملکرد بالاتر از مقدار آستانه هستند برای ادامه مسیر و تشکیل لایه بعدی انتخاب می‌شوند.

این ضرایب می‌بایست طوری محاسبه شوند که میانگین مربعات خطا بین خروجی‌های محاسباتی و واقعی حداقل باشد. بعد از بدست آمدن ضرایب بر اساس داده‌های آموزشی، شاخص عملکرد نرون‌های

## تجزیه و تحلیل مجموعه داده‌ها

برای تخمین میزان ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌ها با استفاده از مدل توسعه داده شده GMDH به مجموعه‌ای از داده‌ها که بیانگر خصوصیات هیدرولیکی و هندسی رودخانه باشند نیاز می‌باشد. بنابراین مجموعه‌ای از 71 سری داده‌های مشاهداتی ضریب پراکندگی طولی آلاینده‌ها از 32 رودخانه طبیعی مختلف جمع-آوری گردید (جدول 2).

به دو دلیل چنین مجموعه‌ای از داده‌های مشاهداتی انتخاب شده‌اند. 1: بیانگر محدوده‌ای وسیع از خصوصیات هیدرولیکی و هندسی رودخانه‌های مختلف هستند و 2: این مجموعه داده توسط دیگر محققان نیز مورد استفاده قرار گرفته‌اند. بنابراین امکان مقایسه نتایج حاصل از این تحقیق با دیگر نتایج ارائه شده امکان‌پذیر می‌باشد. مجموعه داده‌ها شامل  $U$ ،  $H$ ،  $W$ ،  $U^*$ ،  $\beta$ ،  $\sigma$  و  $K_x$ : ضریب

پراکندگی طولی می‌باشد.  $\beta$ : پارامتر شکل رودخانه است که به صورت لگاریتم نسبت عرض رودخانه به عمق جریان تعریف می‌شود. در جدول 3، خصوصیات آماری مجموعه داده‌های مورد استفاده ارائه شده است. همان‌طور که در این جدول مشاهده می‌گردد ضریب همبستگی بین متغیرهای وابسته و مستقل بسیار کم می‌باشد. همچنین تجزیه و تحلیل آماری داده‌ها نشان می‌دهد که هیچ رابطه خطی، غیرخطی یا لگاریتمی بین متغیرها در مجموعه داده‌های مورد بررسی وجود ندارد. مقادیر شاخص‌های آماری همراه با گستردگی زیاد داده‌ها بیانگر این است که هیچ روند خاصی بین متغیرهای ورودی و خروجی وجود ندارد. بنابراین ضروری به نظر می‌رسد که رویکردهای مدل‌سازی مبتنی بر محاسبات نرم جهت تخمین ضریب پخش طولی استفاده شود.

داده جدول 2

ردیف	رودخانه	عرض W(m)	عمق H(m)	سرعت U(m/s)	سرعت برشی U*(m/s)	$\beta$	$\sigma$	$K_x$ واقعی (m <sup>2</sup> /s)
45	Monocacy River, Md.	40/5	0/41	0/23	0/04	4/59	1/61	66/5
46	Muddy River	13	0/81	0/37	0/081	2/77	1	13/9
47	Muddy River	20	1/2	0/45	0/099	2/82	1	32/5
48	Nooksack River	86	2/93	1/2	0/53	3/38	1/3	153
49	Nooksack River	64	0/76	0/67	0/268	4/43	1/3	34/8
50	Powell River, Tenn. *	36/8	0/87	0/13	0/054	3/74	2/2	15/5
51	Red River, La	253/6	1/62	0/61	0/032	5/05	1/2	143/8
52	Red River, La	161/5	3/96	0/29	0/06	3/93	1/44	130/5
53	Red River, La	152/4	3/66	0/45	0/057	3/73	1/44	227/6
54	Red River, La	155/1	1/74	0/47	0/036	4/49	1/24	177/7
55	Sabina River, La	116/4	1/65	0/58	0/054	4/26	1/19	131/3
56	Sabina River, La *	160/3	2/32	1/06	0/054	4/24	1/17	308/9
57	Sabina River, Tex *	14/2	0/5	0/13	0/037	3/35	2/53	12/8
58	Sabina River, Tex	12/2	0/51	0/23	0/03	3/17	2/05	14/7
59	Sabina River, Tex	21/3	0/93	0/36	0/035	3/13	1/47	24/2
60	Salt Creek, Neber	32	0/5	0/24	0/038	4/16	1/38	52/2
61	Susquehanna River	203	1/35	0/39	0/065	5/01	1/13	92/9
62	Tangipahoa River, La	31/4	0/81	0/48	0/072	3/66	1/46	45/1
63	Tangipahoa River, La *	29/9	0/4	0/34	0/02	4/31	1/46	44
64	Tickfau River, La	15	0/59	0/27	0/08	3/23	1/75	10/3
65	White River *	67	0/59	0/35	0/044	4/73	1	30/2
66	Wind/Big River, Wyo	44/2	1/37	0/99	0/142	3/48	1/56	148/6
67	Wind/Big River, Wyo	85/3	2/38	1/74	0/153	3/58	1/56	464/6
68	Wind/Big River, Wyo*	59/4	1/1	0/88	0/119	3/99	1/18	41/8
69	Wind/Big River, Wyo	68/9	2/16	1/55	0/168	3/46	1/18	162/6
70	Yadkin River, N.C	70/1	2/35	0/43	0/101	3/39	2/17	111/5
71	Yadkin River, N.C	71/6	3/84	0/76	0/128	3/92	2/17	260/1

\*مجموعه داده‌های اعتبارسنجی

جدول 3- خصوصیات آماری مجموعه داده‌ها

$K_x$	$\sigma$	$\beta$	$U^*(m/s)$	$U(m/s)$	$H(m)$	$W(m)$	شاخص‌های آماری
1	0/0039	0/0518	0/0009	0/454	0/165	0/301	ضریب همبستگی ( $R^2$ )
1/9	1	2/62	0/002	0/03	0/22	11/9	حداقل (MIN)
107/21	1/48	3/8	0/088	0/54	1/71	82/97	میانگین (MEAN)
892	2/54	5/05	0/553	1/74	19/94	711/2	حداکثر (MAX)
170/07	0/44	0/558	0/089	0/38	2/6	122/5	انحراف معیار (S)
3/03	1/22	0/165	4/02	1/37	5/38	3/49	چولگی ( $C_s$ )
1/58	0/3	0/147	1/01	0/71	1/52	1/47	ضریب تغییرات ( $C_v$ )

$$DR = \log \frac{Kx_{pred}}{Kx_{meas}} \quad (19)$$

که  $\bar{Kx}_{pred}$  و  $\bar{Kx}_{meas}$

به ترتیب میانگین ضریب پراکندگی طولی اندازه‌گیری شده و تخمین زده شده، N تعداد داده‌های مشاهداتی،  $Kx_{pred}$  و  $Kx_{meas}$  نیز بیانگر ضریب پراکندگی طولی اندازه‌گیری شده و تخمینی هستند. مقادیر نزدیک‌تر به صفر شاخص RMSE بیانگر دقت بیشتر مقادیر تخمین زده شده توسط مدل، مقادیر نزدیک به یک شاخص آماری CC بیانگر تطابق بیشتر مقادیر مشاهداتی و محاسباتی و مقدار DR برابر با صفر حاکی از تطبیق دقیق مقادیر محاسباتی و مشاهداتی می‌باشد. مقادیر مثبت DR بیانگر تخمین دست بالا و مقادیر منفی آن بیانگر تخمین دست پایین مقادیر محاسباتی است.

در شکل‌های 2 تا 4 نتایج مدل GMDH برای تخمین ضریب پراکندگی طولی رودخانه در مراحل آموزش و آزمون ارائه شده است. مشاهده این شکل‌ها نشان می‌دهد که مدل GMDH به خوبی و با دقت بسیار مطلوبی توانسته الگوی حاکم بر انتقال آلودگی و ضریب پراکندگی طولی را شناسایی نموده و این ضریب را در هر دو مرحله آموزش و آزمون به طور صحیح برآورد نماید. مقادیر شاخص‌های آماری نیز بیانگر دقت بالای نتایج به دست آمده توسط مدل GMDH می‌باشد. مقدار ضریب تبیین ( $R^2$ ) در مرحله آموزش و آزمون به ترتیب برابر با 0/999 و 0/99 و مقدار ریشه مربعات میانگین خطا (RMSE) به ترتیب در مرحله آموزش و آزمون برابر با 18/5 و 20/8 می‌باشد. مقادیر این شاخص‌های آماری در هر دو مرحله نیز حاکی از توانایی بالای مدل GMDH در مدل‌سازی انتقال آلودگی در رودخانه‌ها می‌باشد. بررسی بیشتر نشان می‌دهد که مقادیر نقاط اوج ضریب پراکندگی مشاهداتی در هر دو مرحله آموزش و آزمون نیز توسط مدل GMDH به خوبی تخمین زده شده‌اند. بزرگ‌ترین مقادیر مشاهداتی ضریب پراکندگی در مجموعه داده‌های آموزش برابر با 446/7 و 839/7 می‌باشد که به ترتیب برابر با 464/4 و 837/7 توسط مدل GMDH برآورد شده است. همچنین مقادیر اوج ضریب پراکندگی طولی 892 و 457/7 ( $m^2/s$ ) در دوره آزمون به

### توسعه مدل GMDH و نتایج

اولین گام در ساخت یک شبکه چندلایه GMDH انتخاب متغیرهای ورودی به مسئله می‌باشد. می‌توان تمام متغیرهایی را که احتمال تاثیرگذاری آن‌ها روی پدیده مورد بررسی وجود دارد انتخاب نمود. در لایه اول تمام ترکیبات دوتایی از متغیرهای ورودی ساخته شده و هر کدام از آن زوج متغیرها وارد یک آدالین می‌شوند. تعداد نرون‌ها در هر لایه از رابطه  $L_1 = \binom{2}{m}$  یا  $m(m-1)/2$  بدست می‌آید که m تعداد متغیرهای ورودی می‌باشد. معمولاً هر چه تعداد لایه‌ها بیشتر شود همگرایی مدل بیشتر شده و فاصله از خطا کم‌تر می‌گردد یعنی میزان اختلاف بین جریان خروجی محاسباتی و مشاهداتی کم‌تر می‌گردد. به همین دلیل و برای اینکه سریع‌تر به پاسخ نهایی برسیم میزان معیار در هر لایه نسبت به لایه قبلی بیشتر انتخاب می‌گردد. بنا به نظر محققان مختلف بهتر است معیار خطا طوری انتخاب گردد که تعداد نرون‌های انتخاب شده برای لایه بعدی در حدود 0/4 لایه قبلی باشند (Muller and Ivakhnenko., 1996).

قبل از استفاده از مدل GMDH، می‌بایست مجموعه داده‌های مورد نظر به دو گروه داده‌های آموزش و آزمون طبقه‌بندی گردند. از مجموع 71 داده مورد نظر، 51 داده در گروه آموزش و 20 داده باقی مانده در گروه زمون قرار می‌گیرند. داده‌هایی که با علامت \* در جدول 2 مشخص شده‌اند برای آزمون مدل استفاده شده‌اند.

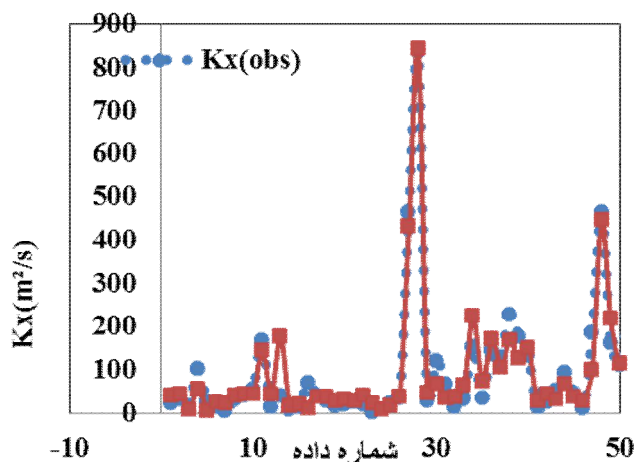
برای ارزیابی عملکرد مدل توسعه داده شده در این تحقیق و برای مقایسه نتایج به دست آمده با نتایج و روابط دیگر محققان از چند معیار آماری شامل مجذور خطای میانگین مربعات (RMSE)، ضریب همبستگی (CC) و اختلاف نسبی (DR) استفاده شده است. این معیارها به صورت روابط 17 تا 18 تعریف می‌شوند:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Kx_{pred} - Kx_{meas})^2}{N}} \quad (17)$$

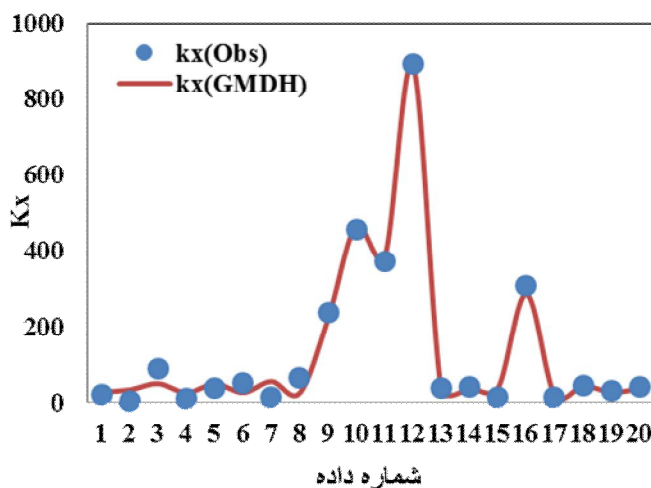
$$CC = \frac{\sum_{i=1}^N (Kx_{meas} - \bar{Kx}_{meas})(Kx_{pred} - \bar{Kx}_{pred})}{\sqrt{\sum_{i=1}^N (Kx_{meas} - \bar{Kx}_{meas})^2 \sum_{i=1}^N (Kx_{pred} - \bar{Kx}_{pred})^2}} \quad (18)$$

(رابطه 3)، لیو (رابطه 4)، ایواسا و آیا (رابطه 5)، سئو و چانگ (رابطه 8)، دنگ و همکاران (رابطه 9) و کاشفی پور و فالکونر (رابطه 10) نیز مقایسه شده است.

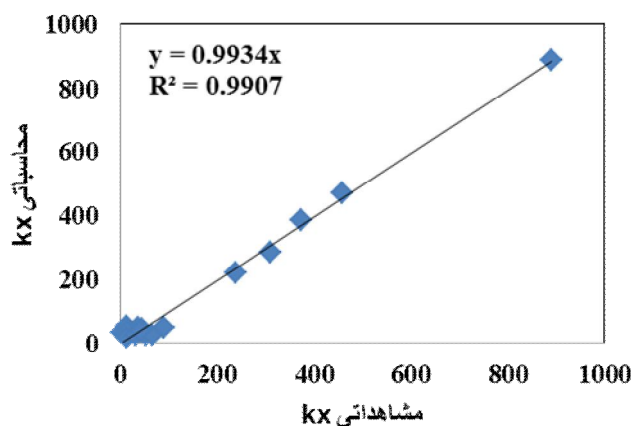
ترتیب برابر با مقادیر 889/2 و 470/6 به دست آمده است. در این تحقیق علاوه بر بررسی عملکرد نتایج حاصل از مدل توسعه داده شده، نتایج مدل با نتایج دیگر محققان مانند ساهای، آدرش، ساهای و دوتا و تیفور و سینگ و روابط ارائه شده توسط فیشر



شکل 2- عملکرد مدل GMDH در تخمین  $k_x$  (مرحله آموزش)



شکل 3- عملکرد مدل GMDH در تخمین  $k_x$  (مرحله آزمون)



شکل 4- مقایسه نتایج مدل GMDH با داده‌های واقعی (مرحله آزمون)



جدول ۴- مقایسه مقادیر ضرایب پراکندگی طولی واقعی و تخمین زده شده با استفاده از روش‌های مختلف

ردیف	نام رودخانه	مقادیر مشاهداتی (Kx)	GMDH	ANN (S)	ANN (T-S)	SVM (A)	GP (A)	GA (S-D)	Fischer رابطه ۳	Liu رابطه ۴	I-A رابطه ۵	S-C رابطه ۸	D-S-B رابطه ۱۱	K-F رابطه ۱۲
۱	Antietam Creek, Md.	۲۰/۹	۲۷/۹	۴۴/۵	۲۶/۸	۲۸/۱۲	۱۷/۱	۱۳/۷	۵/۱	۷/۴	۸/۶	*۲۰/۲	۱۵	۱۵/۲
۲	Bear Creek, Colo	۲/۹	۳۳/۶	۵۰/۰/۷	۳۹/۲	۳۰/۸	۵۱/۴	۳۹/۱	*۷/۳	۳۳/۶	۶۰/۷	۵۲/۲	۲۸/۱	۲۷/۱
۳	Chattahoochee River, Ca	۸۸/۹	۵۰/۰	*۹۵/۵	۷۷/۶	۱۴/۱	۱۰۰/۵	۱۴/۱	۱۲۷/۹	۱۲۷/۹	۱۲۹/۹	۱۶۹/۱	۱۶۸/۶	۸۲/۱
۴	Clinch River, Va	۱۰/۷	۲۵/۷	۱۳/۷	۲۶/۹	۲۴/۹	۲۴/۱	۲۵/۸	۲۶/۴	۳۷/۷	۲۷/۲	۲۷/۶	۲۸/۴	*۱۱/۵
۵	Clinch River, Va	۳۶/۹	۴۹/۷	*۴۸/۶	۷۶/۶	۶۶/۹	۷۷/۳	۹۷/۸	۵۲/۶	۵۲/۶	۵۳/۵	۱۳۹/۶	۱۱۸/۲	۱۰۴/۱
۶	Conococheague Creek	۵۲/۳	۲۶/۴	۶۵/۴	۴۳	۹۳/۲	۸۹/۸	۷۸/۲	۸۸/۲	۸۸/۲	۴۳	۹۶/۷	۹۲/۱	*۵۵/۸
۷	John Day River, Ore.	۱۳/۹	۵۵/۶	۹۴/۶	۴۵/۲	۱۳۷/۶	۵۱/۴	۷۱/۲	۸۶/۴	۷۲/۹	۴۶	۸۳/۳	۸۱/۷	*۴۴/۸
۸	John Day River, Ore.	۶۵	۲۵/۵	۸۲/۳	۷۷/۲	۲۸/۷	۸۲/۶	۶۷/۳	۱۹/۳	۳۲/۶	۴۵/۶	۱۱۶/۷	*۷/۱	۹۷/۹
۹	Mississippi River, La	۲۳۷/۲	*۲۲۱/۱	۳۲/۵	۸۳/۸	۵۲/۴	۵۸۳/۱	۱۳۳۷/۳	۲۱۳۴/۲	۶۹/۱/۹	۳۴/۸	۱۸۵۴/۵	۱۶۱/۷	۱۶۱/۸
۱۰	Mississippi River, Mo	۴۵۷/۷	*۴۷۰/۶	۵۰۴/۹	۸۲/۸	۵۲/۱	۳۷۹/۶	۱۸۳۴/۵	۱۰۱۲۲/۷	۲۷۹۰/۴	۶۴/۴	۱۷۹۳	۱۲۴/۵	۸۳/۶
۱۱	Mississippi River, Mo	۳۷۴/۱	*۳۸۶/۹	۴۲۰/۱	۸۳/۸	۵۲/۴	۵۱۸	۲۷۳/۶	۸۳۹۰/۴	۲۲۳۵/۴	۲۴۴/۶	۳۲۷/۵	۲۵۷۹/۳	۲۲۲۰/۱
۱۲	Missouri River	۸۹۲	*۸۸۹/۲	۸۸۰/۷	۷۳/۴	۹۷/۲	۱۰۱۷/۳	۱۰۷۴/۷	۴۱۱۹/۶	۷۷	۲۴۴/۶	۱۳۱۷/۳	۲۵۷۹/۳	۴۹۰/۵
۱۳	Monoacy River, Md.	۳۷/۸	*۳۲/۷	۲۲/۸	۲۷/۱	۶۳/۸	۳۳/۴	۳۱/۷	۶۱/۷	۹۰/۳	۴۷/۷	۲۷/۱	۲۸/۲	۷/۶
۱۴	Monoacy River, Md.	۴۱/۴	*۳۸/۵	۴۶/۱	۳۱/۴	۶۷/۱	۴۷/۵	۳۳/۴	۷۴/۶	۱۸۸/۱	۹۷/۹	۲۳/۵	۲۵/۸	۴/۲
۱۵	Powell River, Tenn.	۱۵/۵	۳۸/۷	۲۶/۳	۲۵/۳	۴۲/۳	*۱۴/۶	۱۰/۳	۵/۴	۲۳/۵	۲۵/۷	۹/۹	۹/۹	۲/۹
۱۶	Sabina River, La	۳۰۸/۹	۲۸۶/۷	*۳۰۳/۴	۳۴۶/۶	۴۰/۶	۳۲۰/۷	۶۰۳/۸	۲۵۳۵/۱	۴۷	۱۴۳/۹	۷۱۸/۷	۵۰۸/۴	۵۱۲/۳
۱۷	Sabina River, Tex	۱۲/۸	*۱۸/۹	۲۶/۹	۲۱/۹	۳۵/۴	۷/۱	۴/۴	۲	۵	۵/۶	۵/۲	۴/۶	۲/۴
۱۸	Tangipahoa River, La	۴۴	*۴۴/۳	۳۱/۴	۲۶/۵	۳۰/۲	۲۵/۹	۲۴/۷	۱۴۲/۱	۳۳/۲	۱۰/۳	۳۹/۲	۲۸/۷	۲۴/۵
۱۹	White River	۳۰/۲	۲۷/۲	۳۷/۱۷	*۳۱/۷	۹۳/۹	۱۸/۵	۶۵/۲	۲۲۹	۱۷۰	۶۲/۸	۵۴/۳	۴۶/۴	۱۷/۷
۲۰	Wind/Big. River, Wyo	۴۱/۸	*۳۶/۶	۱۱۲/۹	۵۹/۷	۱۳۲/۱	۱۶۹/۷	۱۴/۷	۲۲۹/۶	۱۸۶/۸	۱۰۳/۹	۱۵۹/۹	۱۵۶/۵	۷/۶

ANN(S):Sahay[21], ANN(T-S):Tayfur and Singh[27], SVM(A):Adrash[9], GP(A):Adrash[9], GA(S-D):Sahay and Dutta[21], I-A:Iwasa and Ayal[4], S-C:Seo and Ceong[25], D-S-B:Deng et al[21], K-F:Kashefipour and Falconer[16].

جدول 5- مقایسه عملکرد شاخص‌های آماری در روش‌های مختلف (مرحله آزمون)

مدل	RMSE	R <sup>2</sup>	CC	RMSE Kx < 100	W/H < 50	محدوده DR	دقت مدل (%)
GMDH	21/5	0/99	0/995	19/45	19/7	0/4 تا 1/06	65
ANN(S)	37/6	0/98	0/99	32/6	30	0/22 تا 1/24	65
ANN(T-S)	192/9	0/69	0/83	16/7	135/2	0/22 تا 1/13	45
SVM(A)	107/8	0/76	0/87	58/5	48/4	0/26 تا 1/65	20
GP(A)	96/5	0/89	0/945	34/6	79/5	0/46 تا 1/25	50
GA(S-D)	663/6	0/48	0/69	35/5	245	0/46 تا 1/13	30
Fischer	2972	0/52	0/72	69/7	424/8	0/80 تا 1/35	10
Liu	679/5	0/40	0/63	62/6	104/9	0/45 تا 1/06	15
S-C	812	0/47	0/68	47	363/6	0/39 تا 1/25	20
K-F	525/5	0/44	0/66	23/5	309/4	0/99 تا 0/97	25
D-S-B	611	0/42	0/64	41/4	310/3	0/44 تا 0/98	20
I-A	195	0/38	0/61	29	31/4	0/63 تا 1/32	30

که مقادیر تخمین زده شده برای ضریب پراکندگی در محدوده 67 تا 150 درصدی مقادیر اندازه‌گیری شده قرار دارند. از جدول 5 مشاهده می‌گردد که بیش‌ترین دقت مدل‌سازی مربوط به مدل‌های GMDH و ANN(S) و رابطه فیشر دارای کم‌ترین دقت مدل‌سازی است.

### نتیجه‌گیری

اندازه‌گیری صحیح ضریب پراکندگی آلاینده‌ها در رودخانه‌های طبیعی امری پیچیده، زمان‌بر و مستلزم صرف هزینه‌های زیادی می‌باشد. بنابراین استفاده از مدل‌هایی که بتوانند به طور صحیح و دقیقی مقدار این ضریب را تخمین بزنند اجتناب‌ناپذیر است. هدف اصلی در این تحقیق ارائه مدلی بر مبنای رویکرد استقرایی داده‌محور GMDH برای تخمین صحیح ضریب پراکندگی طولی آلاینده‌ها با استفاده از مجموعه داده‌های گردآوری شده در رودخانه‌های طبیعی بوده است. با در نظر گرفتن مجموعه‌ای از متغیرهای ورودی که در برگیرنده خصوصیات هندسی رودخانه‌ها (عرض و سینوسیته رودخانه) و جریان (سرعت متوسط، عمق و سرعت برشی) بوده‌اند، مدل ارائه شده توانسته با دقت بسیار مناسبی مقادیر ضریب پراکندگی طولی را تخمین بزند. مقادیر شاخص‌های آماری RMSE، R<sup>2</sup> و CC در مدل GMDH به ترتیب برابر با 21/5، 0/99 و 0/995 به دست آمده است که حاکی از عملکرد مناسب مدل ارائه شده در تخمین ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی می‌باشد. همچنین مدل ارائه شده مقادیر اوج مشاهداتی ضریب پراکندگی طولی را که برابر با 892/7 و 457/7 و 374/1 بوده‌اند توانسته به خوبی به ترتیب با مقادیر 889/2، 470/6 و 386/9 تخمین بزند. همچنین عملکرد مدل GMDH با دیگر معادلات تجربی و روش‌های ارائه شده توسط دیگر

در جدول 4 مقادیر ضریب پراکندگی طولی به دست آمده از اجرای مدل توسعه داده شده GMDH، مقادیر اندازه‌گیری شده و نتایج حاصل از تحقیقات دیگر محققان با مجموعه داده‌های مورد نظر در این تحقیق ارائه شده است.

همان‌گونه که ملاحظه می‌گردد کم‌ترین تغییر در مقادیر RMSE به ترتیب مربوط به مدل‌های GMDH، ANN(S) و GP(A) می‌باشد که حاکی از توانایی بالای مدل‌های مذکور در تخمین ضریب پراکندگی طولی و شناسایی الگوی حاکم بر انتقال آلودگی چه در رودخانه‌های کوچک و چه در رودخانه‌های بزرگ می‌باشد. با حذف مقادیر حدی، بیش‌ترین بهبود در مقدار RMSE در رابطه فیشر مشاهده می‌شود که بیانگر قابلیت پایین مدل‌سازی این رابطه در رودخانه‌های بزرگ می‌باشد. محدوده شاخص DR بیانگر چولگی مقادیر محاسباتی ضریب پراکندگی طولی نسبت به مقادیر مشاهداتی می‌باشد. مقادیر مثبت شاخص DR نشان دهنده تخمین دست بالا و مقادیر منفی بیانگر تخمین دست پایین می‌باشد. ممکن است در نگاه اول محدوده 0/4- تا 1/06 شاخص DR مدل GMDH نشان دهنده تخمین دست بالای مدل GMDH باشد اما شایان ذکر است که حد بالای شاخص DR در مدل GMDH مربوط به دومین مقدار مشاهداتی ضریب پراکندگی طولی در جدول 4 یعنی ضریب پراکندگی طولی 2/9 می‌باشد که یک عدد خارج از محدوده می‌باشد. با فرض در نظر نگرفتن این مقدار، کران بالای شاخص DR در مدل GMDH به 0/6 کاهش می‌یابد. مطالعه دیگر محققان نیز این امر را تایید می‌کند (Adraash., 2010). در این صورت محدوده شاخص DR از 0/4- تا 1/06 به 0/4- تا 0/6 تغییر می‌کند که حاکی از توزیع مناسب مقادیر تخمینی می‌باشد. اگر دقت مدل به صورت درصد مقادیر DR بین 0/175- تا 0/175 تعریف شود که بیانگر این است

قادری، ک.، عرب، د.، تشنه‌لب، م. و قزاق، آ. 1389. مدل‌سازی بهره- برداری هوشمند از مخازن با استفاده از روش برخورد گروهی با داده‌ها (GMDH). مجله علمی پژوهشی تحقیقات منابع آب ایران. 6,3: 67-55.

Adrash, S. 2010. Prediction of longitudinal dispersion coefficient in natural channels using soft computing techniques. *Scientia Iranica, Transaction A: civil engineering*. 17.5:363-371.

Azamathulla, H.M and Ghani, A.A. 2011. Genetic programming for predicting longitudinal dispersion coefficients in streams. *Water Resources Management*. 25:1537-1544.

Cheong, T.S and Seo, I.W. 2003. Parameter estimation of transient storage model by a routing method for river mixing processes. *water Resource Research*. 39.4:1074-1084.

Cheong, T.S., Younis, B.A and Seo, I. 2007. Estimation of key parameters in model for solute transport in rivers and streams. *Water Resour Manage*. 1:1165-1186.

Hwang, H.S. 2006. Fuzzy GMDH-type Neural Network Model and its Application to Forecasting of Mobile Communication. *Computers and Industrial Engineering*. 50:450-457.

Ivakhnenko, A.G. 1968. The group method of data handling - a rival of the method of stochastic approximation. *Soviet Automatic Control c/c of Avtomatika*. 1.3:43-55.

Ivakhnenko, A.G and Ivakhnenko, G.A. 1995. The review of problems solvable by algorithms of the Group Method of Data Handling (GMDH), *Pattern recognition and image analysis*. 5.4: 527-535.

Kashefipour, M.S and Falconer, R.A. 2002. Longitudinal dispersion coefficients in natural channels. *Water Resources Reseach*. 36.6:1596-1608.

Muller, J.A and Ivakhnenko, A.G. 1996. Self-organizing modeling in analysis and prediction of stock market. in *Proceedings of the Second International Conference on Application of Fuzzy Systems and Soft Computing-ICAFS*. 491-500.

Nariman Zadeh, N., Darvizhe, A and Gharabaghi, H. 2002. Modeling of Explosive Cutting Process of Plates Using GMDH Neural Networks and Singular Value Decomposition. *Journal Mater process technol*. 128.1:80-87.

Nikolaev, Y and Hitoshi, I. 2003. Polynomial Harmonic GMDH Learning Networks for Time Series Modeling. *Neural Networks*. 16:1527-1540.

Rowinski, P.M and Piotrowski, A. 2008. Estimation of parameters of the transient storage model by means of multi-layer perceptron neural networks. *Hydrological Sciences-Journal-des Sciences*

محققین و مدل‌های ارائه شده توسط دیگر محققان شامل شبکه‌های عصبی، روش ماشین بردار پشتیبان و برنامه‌ریزی ژنتیک و الگوریتم ژنتیک نیز مقایسه شده است. مدل ارائه شده در این تحقیق با 65 درصد دقت در بین مدل‌ها و معادلات مقایسه شده، دارای بیش‌ترین دقت در مدل‌سازی می‌باشد. تقریباً هیچ کدام از روابط تجربی دقت قابل قبولی نداشته و نمی‌توانند در طرح‌های زیست‌محیطی و مدیریت آلودگی در رودخانه‌ها با اطمینان قابل قبولی مورد استفاده قرار گیرند. مدل ارائه شده شبکه‌های عصبی توسط ساهای، با ضریب تبیین برابر با 0/98 و دقت 65 درصد دارای نزدیک‌ترین عملکرد به مدل GMDH بوده است. مقادیر شاخص‌های آماری برای مدل ارائه شده در این تحقیق و همچنین مقایسه نتایج با دیگر مدل‌ها و روابط ارائه شده، حاکی از توانایی بالای مدل GMDH در تخمین ضریب پراکندگی طولی آلاینده‌ها در رودخانه‌های طبیعی می‌باشد.

## منابع

ایوانی، ز. 1392. برآورد نرخ انتقال رسوب در رودخانه کرج با استفاده از روش‌های داده‌محور GMDH و NF-GMDH. دانشگاه شهید باهنر کرمان، پایان نامه کارشناسی ارشد.

پورآبادی، م.، تکلدانی، م. و لیاقت، ع. 1386. بررسی تاثیر پارامترهای جریان بر ضریب پراکندگی آلودگی در کانال مستطیلی. ششمین کنفرانس هیدرولیک، شهرکرد، ایران.

پورنعمت رودسری، ع.، قادری، ک. و کریمی، ش. 1393. شبیه‌سازی فرایند بارش-رواناب با استفاده از روش کنترل گروهی داده‌ها (GMDH) در حوضه آبریز پل‌رود. پژوهشنامه مدیریت حوزه آبخیز. 5. 1: 68-84.

توکلی زاده، ا.ع. و کاشفی پور، س.م. 1385. تاثیر ضریب پخشیدگی بر مدل‌سازی کیفی آب‌های سطحی. هفتمین سمینار بین‌المللی مهندسی رودخانه، دانشگاه شهید چمران اهواز، اهواز.

حسین‌زاده، م. 1392. مدل‌سازی رابطه دبی-اشل با استفاده از روش‌های محاسبات نرم GMDH و NF-GMDH. دانشگاه شهید باهنر کرمان، کرمان، پایان نامه کارشناسی ارشد.

ریاحی مدوار، ح. و ایوب‌زاده، ع. 1387. تخمین ضریب پراکندگی طولی با استفاده از سیستم استنتاج فازی-عصبی انطباقی. مجله آب و فاضلاب. 67, 19: 34-46.

صدیق‌نژاد، ح. و کاشفی پور، س.م. 1385. مقایسه نتایج مدل‌های انتشار و انتقال رسوب برای آبگیر ماراد و ارائه مدل مناسب برای منطقه با استفاده از مدل ریاضی FASTER. هفتمین سمینار بین‌المللی مهندسی رودخانه، دانشگاه شهید چمران، اهواز.

- Seo, I.W., Cheong, T.S. 1998. Predicting longitudinal dispersion coefficient in natural streams. *Journal of Hydraulic Engineering*. 124.1:25-32.
- Sumsudin, R., Saad, P and Shabri, A. 2010. A hybrid least square support vector machines and GMDH approach for river flow forecasting. *Hydrology and earth system Science Discussions*, 7:3691-3731.
- Tayfur, G and Singh, V.P. 2005. Predicting longitudinal dispersion coefficient in natural streams by artificial neural network. *Journal of Hydraulic Engineering*. 131.11:991-1000.
- Toprak, Z.F and Cigizoglu, H.K. 2008. Predicting longitudinal dispersion coefficient in natural streams by artificial intelligence methods. *Hydrologic Process*. 22:4106-4129.
- Hydrologiques. 53.1: 165-178.
- Sahay, R.R. 2011. Prediction of longitudinal dispersion coefficients in natural rivers using artificial neural network. *Environ Fluid Mechanics*. 11:247-261.
- Sahay, R.R. 2012. Predicting Transient Storage Model Parameters of Rivers by Genetic Algorithm. *Water Resources Management*. 26:3667-3685.
- Sanchez, E., Shibata, T and Zadeh, L.A. 1997. Genetic algorithms and fuzzy logic systems. *Reveredge: word scientific*.
- Seo, I.W., Baek, K.O. 2004. Estimation of longitudinal dispersion coefficient in natural streams by artificial neural network. *Journal of Hydraulic Engineering*. 130.3:227-236.

## Prediction of Longitudinal Dispersion coefficient in Natural River Using GMDH Data Driven Approach

K. Qaderi<sup>1\*</sup>, M. Hoseynzadeh<sup>2</sup>

Received: Jum.09, 2015

Accepted: Nov.16, 2016

### Abstract

Accurate estimate of longitudinal dispersion coefficient is important in many hydraulic and environmental problems in rivers such as river engineering, intake designs, modeling flow in estuaries and risk assessments of pollutants into river flows. To accurate investigation of water quality using one dimensional model, the precise estimation of longitudinal dispersion coefficient is required. Direct measurements of longitudinal dispersion coefficients, with the aid of concentration samples taken in upstream and downstream of rivers is rather seldom. Recent research works indicate that, using the data driven method can improve the precise estimation of longitudinal dispersion coefficient in natural rivers. In this research, the usefulness and performance of Group Method of Data Handling (GMDH) approach are examined for predicting longitudinal dispersion coefficient in natural channels. A set of 71 data sets from different river has been gathered so that 51 sets of whole data were used for training and 20 remaining sets were used for test data sets. The hydraulic and geometric variables such as mean flow depth (H), width of channel (W), mean flow velocity (U), channel sinuosity ( $\sigma$ ) and shear velocity ( $U^*$ ) are used as input variables to predict longitudinal dispersion coefficient ( $K_x$ ). A computer program based on GMDH approach is written in MATLAB software for  $K_x$  modeling. Based on the values of various performance indices,  $R^2$ , RMSE, CC and DR, it is concluded that GMDH model in both training and validation period predicts the longitudinal dispersion coefficient more accurately. Comparison of GMDH model with empirical approach and another data driven method such as ANN, SVM and GA confirm that GMDH shows remarkably good performance in capturing governing pattern in longitudinal dispersion phenomena in natural rivers. Hence GMDH can be used as an efficient computational paradigm in the estimation of longitudinal dispersion coefficient in natural channel.

**Keywords:** Data driven, Dispersion coefficient method, Empirical relation, Environmental, GMDH, Modeling

1 -Assistant Professor. Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

2 -M.Sc, Water Resources Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

(\* - Corresponding Author: Kouroshqaderi@uk.ac.ir)