

افزایش دقت پیش بینی بارکل رسوبی با استفاده از الگوریتم‌های تکاملی (مطالعه موردی: رودخانه قطورچای)

کیومرث روشنگر^{۱*} - محمد تقی اعلمی^۲ - فاطمه وجودی مهربانی^۳

تاریخ دریافت: ۱۳۹۲/۰۸/۲۰

تاریخ پذیرش: ۱۳۹۳/۰۲/۰۹

چکیده

تخمین درست میزان رسوبات حمل شده توسط رودخانه‌ها در پروژه‌های منابع آبی از اهمیت بالایی برخوردار است. بطور کلی پدیده فرسایش و انتقال رسوب از پیچیده ترین مسائل هیدرودینامیکی است که تعیین دقیق معادلات حاکم بر آن بدلیل تاثیرات پارامترهای مختلف، به آسانی میسر نیست. با وجود انجام تحقیقات بسیاری در زمینه کاربرد مدل‌های هوشمند نظیر شبکه‌های عصبی و برتری این مدل‌ها نسبت به روابط ریاضی و تجربی موجود نظیر منحنی سنجه رسوب، بدلیل غیر صریح بودن و پیچیدگی حاکم بر انتخاب و معماری شبکه مناسب، کاربرد این مدل‌ها توسعه کمتری نسبت به روش‌های صریحی نظیر برنامه ریزی ژنتیک داشته است. در این پژوهش، بمنظور توانمند سازی پیش بینی صریح بار رسوبی رودخانه قطورچای از یک سو الگوریتم‌های تکاملی نظیر برنامه ریزی ژنتیک (GP) و الگوریتم ژنتیک (GA) به کار گرفته شده و از سوی دیگر از مدل‌های نیمه تجربی تعیین بار کل رسوب و منحنی سنجه استفاده گردیده است. مقایسه و تجزیه تحلیل نتایج حاصل از روش‌های کلاسیک، منحنی سنجه بهینه و روش برنامه ریزی ژنتیک، کرائی بسیار بالای الگوریتم‌های تکاملی را ($DC=0/907$ و $RMSE=0/067$) بعنوان ابزاری قدرتمند در بهینه سازی و پیش بینی صریح بار رسوبی کل رودخانه نشان می دهد.

واژه‌های کلیدی: بار رسوبی، برنامه‌ریزی ژنتیک، بهینه‌سازی پیش بینی، منحنی سنجه

مقدمه

متنوعی نیازمند بوده و از طرفی نتایج حاصل از آن‌ها حتی در یک ایستگاه از یک رودخانه مشخص معمولاً یکسان و مشابه نبوده و انتخاب یکی از آنها مشکل می باشد، محققان در جستجوی راه‌های عملی برای پیش بینی رسوب برآمده اند و تلاش آنها باعث ایجاد روش دیگری موسوم به روش منحنی سنجه رسوب شده است. در این روش یک منحنی توانی مبتنی بر دو پارامتر متغیر Q_s (دبی رسوب) و Q_w (دبی آب) و نیز ضرایب ثابت a و b بر داده‌های آماری دبی جریان و دبی رسوبی برازش داده می شود. اما فرآیند بدست آوردن یک رابطه برای تخمین میزان رسوب، بطور اساسی یک مسأله نگاهت غیرخطی است و روش‌های هوش مصنوعی بعنوان ابزاری توانمند در حل اینگونه مسائل بشمار می آیند (۱). این مدل‌ها بعنوان یک جعبه سیاه مناسب که کمتر در قید و بند مسائل فیزیکی بوده و قادرند فرآیند غیرخطی و غیر ایستای جریان رودخانه را بدون نیاز به مدل سازی عامل‌های محیطی و ژئومتری موثر بر جریان رودخانه مدل سازی کنند، می باشند. از جمله تحقیقات صورت گرفته در این زمینه، می توان به کار سارانگی و بهاتاچارا (۹) اشاره کرد که درحوزه آنگیر بنها

به دلیل اهمیت بالای پدیده رسوب گذاری و تخمین بار رسوبی در پروژه‌های آبی و ارتباط تنگاتنگ آن با مسائل اقتصادی، تاکنون تحقیقات فراوانی در زمینه روش‌های برآورد بار رسوبی و تأثیر عوامل مختلف بر روی میزان رسوب انجام گرفته است. تمامی روش‌های مذکور جهت تخمین میزان بار رسوبی در یک ایستگاه در زمان‌های مختلف و تحت جریان‌های مختلف تمرکز داشته و در واقع این نوع تخمین مستلزم داشتن ایستگاه در مقطع مورد نظر و در دست بودن آمار مربوطه می باشد. روش‌های تخمین میزان رسوب را می توان در دو گروه دسته بندی کرد. رویکرد روش‌های سری اول، ایجاد مدل‌های ریاضی با توجه به مفاهیم فیزیکی انتقال ذرات و حل معادلات هیدرودینامیک می باشد. با توجه به اینکه این روش‌ها به داده‌های

۱، ۲ و ۳- به ترتیب دانشیار، استاد و کارشناس ارشد، گروه عمران- سازه‌های هیدرولیکی، دانشکده فنی و مهندسی، دانشگاه تبریز

*- نویسنده مسئول: Email: Kroshangar@yahoo.com

الگوریتم ژنتیک.

برنامه ریزی بیان ژن

برنامه ریزی بیان ژن (GEP) که براساس تئوری داروین ارائه شده توسط فریرا در سال ۱۹۹۹ ابداع شد. برنامه ریزی بیان ژن (GEP) نیز همانند برنامه ریزی ژنتیک (GP)، یک الگوریتم ژنتیکی است که از جمعیتی از افراد استفاده کرده و آنها را مطابق برازندگی انتخاب می کند و تغییرات ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی اعمال می نماید (۴).

مراحل اصلی در برنامه ریزی بیان ژن

اولین مرحله در الگوریتم GEP، تولید جمعیت اولیه از راه حل هاست. این موضوع می تواند بوسیله فرآیند تصادفی و یا با استفاده از مقدراری اطلاعات درباره مسئله انجام شود. سپس کروموزومها بصورت بیان درختی (ETS) نشان داده شده، که این هم مطابق با یک تابع برازش ارزیابی می گردد. معمولاً تابع برازش بوسیله پردازش تعدادی از مسئله هدف، که مورد برازش نیز نامیده می شوند، ارزیابی می گردد (۸).

عملگرهای برنامه ریزی بیان ژن

GEP از روش معروف چرخ-رولت، برای انتخاب افراد استفاده می کند. GEP برخلاف GP، چندین عملگر ژنتیکی برای تکثیر افراد با اصلاحات دارد (۸). هدف عملگر جهش، بهسازی تصادفی داخل کروموزوم های معین است. در GEP علاوه بر ترکیب تک نقطه ای و دو نقطه ای، نوع دیگری از ترکیب، بنام ترکیب ژنی نیز اجرا می شود که در آنها ژنها بطور کامل ترکیب می شوند. این عملگر، به طور تصادفی ژنها را در موقعیت یکسانی در دو کروموزوم والد انتخاب می کند تا دو فرزند جدید را تشکیل دهد.

تابع برازش

یکی از موارد مهم در (GEP)، تعیین تابع برازش است و هدف آن، یافتن راه حلی است که برای تمامی موارد برازش به اندازه یک خطای معین به خوبی عمل کند. از لحاظ ریاضی، برازش f_i از یک برنامه انفرادی i به صورت رابطه (۱) بیان می شود.

$$f_i = \sum_{j=1}^n \left(R - \left| 100 \times \frac{P_{ij} - T_j}{T_j} \right| \right) \quad (1)$$

که در آن محدوده انتخابی، P_{ij} مقدار پیش بینی شده بوسیله برنامه انفرادی i برای مورد برازش j (از میان n مورد برازش) و T_j مقدار هدف برای مورد برازش j است. قابل توجه است که عبارت داخل قدر مطلق متناظر است با درصد خطای نسبی. چنانچه $P_{ij}=T_j$ شود خطا

هندوستان با استفاده از الگوهای شبکه عصبی و رگرسیون دبی رسوب را برای هر دو الگو برآورد نموده و نتیجه گرفتند که الگوی شبکه عصبی به عنوان یک روش هوش مصنوعی از دقت بالاتری برخوردار است. مطالعات ذاکر مشفق و همکاران (۱۲)، کیشی (۶)، داوسن و ویلی (۲) بر روی جریان رودخانه ای حاکی از آنست که با کاربرد روش شبکه عصبی مصنوعی می توان دبی عبوری از رودخانه را با دقت قابل قبولی برآورد نمود. اگر چه در تمامی تحقیقات مذکور برتری دقت الگوهای متکی بر شبکه های عصبی مصنوعی نسبت به بر روابط تجربی و الگوهای سری زمانی نشان داده شده است لیکن بدلیل پیچیدگی در ایجاد ساختار و غیرصریح بودن این الگوها استفاده از اینگونه الگوها در عمل به طور مناسبی توسعه نیافته است. لذا توسعه یک الگوی صریح و آسان برای پیش بینی جریان رودخانه ها ضروری می باشد. مطالعات کتابخانه ای انجام شده نشان می دهد که در دهه اخیر روش های برنامه ریزی ژنتیک (GP) و الگوریتم ژنتیک (GA) به عنوان یک روش تناوبی موثر در پیش بینی داده ها در زمینه مهندسی آب مورد استفاده قرار گرفته اند. این روش ها جزء روش های الگوریتم تکاملی محسوب می شوند که مبنای تمامی آنها بر اساس نظریه تکامل داروین استوار است. الگوریتم های یاد شده اقدام به تعریف یک تابع هدف در قالب معیار های کیفی نموده و سپس تابع یاد شده را برای اندازه گیری و مقایسه روش های مختلف حل، در یک فرآیند گام به گام تصحیح ساختار داده ها به کار می گیرند و در نهایت، روش حل مناسب را ارائه می نمایند. کالرا و دئو (۵) برای بازیافت داده های گمشده امواج در طول ساحل غربی هندوستان از روش برنامه ریزی ژنتیک بهره جستند. لیونگ و همکاران (۷) و همچنین ویگهام و کراپر (۱۰) با الگوسازی بارش- رواناب به روش برنامه ریزی ژنتیک بدین نتیجه دست یافتند که پیش بینی رفتار بارش- رواناب در حوزه های آبریز به کمک برنامه ریزی ژنتیک سبب بروز خطای کمتری خواهد شد. هدف از تحقیق حاضر مقایسه کارایی روش های ریاضی پرکاربرد و روش مرسوم و رگرسیونی منحنی سنجه با بکارگیری روش برنامه ریزی ژنتیک و روش الگوریتم ژنتیک در پیش بینی میزان بار رسوبی در یک رودخانه طبیعی است.

مواد و روش ها

الگوریتم های تکاملی

الگوریتم های تکاملی روش های بر مبنای جستجوی تصادفی اند که از مدل سازی تکامل بیولوژیکی طبیعی الگوبرداری شده اند. آنها بر روی پاسخ های ممکن کار می کنند که از ویژگی برتری برخوردار و نیز بقای نسل بیشتری دارند، لذا تخمین نزدیک تری از پاسخ بهینه بدست می دهند. دو گونه پرکاربرد از الگوریتم های تکاملی که در این تحقیق نیز مورد استفاده هستند عبارتند از برنامه ریزی بیان ژن و

صفر می شود (۳).

شبکه ها به کمک رابطه زیر در بازه‌ی [۱ و ۰.۰۵] نرمال سازی شده‌اند:

$$X_{istd} = 0.05 + 0.95 \times \left(\frac{X_i - X_{i \min}}{X_{i \max} - X_{i \min}} \right) \quad (2)$$

که منظور از $X_{i \min}$ کمترین داده و از $X_{i \max}$ بیشترین داده از نظر مقدار است.

معیارهای ارزیابی

برای ارزیابی مدل‌های مورد نظر از شاخص‌های ضریب تعیین (DC) و جذر میانگین مربعات خطا (RMSE) بهره برده شده است که به ترتیب با استفاده از روابط زیر قابل محاسبه می باشند:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{oi} - Q_{ci})^2}{n}} \quad (3)$$

$$DC = 1 - \frac{\sum_{i=1}^n (Q_{oi} - Q_{ci})^2}{\sum_{i=1}^n (Q_{oi} - Q_{om})^2} \quad (4)$$

که در آنها Q_{oi} مقدار مشاهده شده در گام زمانی i ام، Q_{ci} مقدار محاسبه شده در همان زمان، n تعداد گام‌های زمانی و Q_{om} میانگین مقادیر مشاهداتی می باشد. کم بودن میزان RMSE و بالا بودن ضریب DC بیانگر دقت قابل قبول مدل و برتری آن نسبت به مدل‌های دیگر است.

تعیین الگوهای ورودی برای روش برنامه ریزی ژنتیک

همانگونه که پیشتر نیز ذکر شد، انتخاب جمعیت‌های اولیه تصادفی مختلف و تاثیرگذار در پدیده به منظور آموزش ماهیت ساز و کار حاکم بر پدیده نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت الگو نیز می شود. لذا در الگوسازی رسوب رودخانه‌ها نیز بایستی سعی نمود موثرترین داده‌های مشاهداتی را به عنوان داده‌های آموزشی انتخاب کرد. به عبارت دیگر برای تعیین بهترین الگو، الگوهای مختلفی به برنامه ریزی ژنتیک معرفی می شود و در نهایت از بین الگوهای مختلف، الگویی که در ارزیابی مقایسه‌ای، بهترین نتیجه را ایجاد کند به عنوان مناسب‌ترین الگو انتخاب می شود. در تحقیق حاضر پارامترهای موثر بر میزان رسوب می توانند به این صورت بیان شوند (۱۱):

$$C = f(Q_w, h, D, S, g, b, n, \mu) \quad (5)$$

در این مطالعه ترکیب‌هایی از دبی آب (Q_w)، عمق آب (h)، اندازه ذره (D)، شیب رودخانه (S) و عرض رودخانه (b) به عنوان ورودی‌های با

ساختار افراد GEP

در GEP، هر ژن بصورت بیان درختی کدگذاری می شود. در مورد کروموزوم‌های چند ژنی، تمامی بیان درختی‌ها با استفاده از تابع پیوند، از محل گره ریشه خود به یکدیگر متصل می شوند. هر ژن، ناحیه کدگذاری بنام ORF (چارچوب تفسیر باز) دارد، که بعد از کد برداری (رمز گشایی) بصورت ET بیان می شود و نشان دهنده یک راه حل کاندید، برای مسئله است.

الگوریتم ژنتیک

الگوریتم ژنتیک بوسیله هالند در دهه ۱۹۷۰ بصورت مفهوم الگوریتمیک بر مبنای استراتژی ادامه حیات موجودات دارای بهترین تطبیق با پیرامون از نوع داروین به طریق باز تولید جسمی ارائه شد که در آن هرچه مشخصات فردی هر راه حل قویتر باشد، انتخاب برتری ایجاد خواهد شد. الگوریتم ژنتیک معمولاً از سه عملگر اساسی تشکیل می‌شود: انتخاب، همپوشانی و جهش. این عملگرها جهت تصحیح راه‌حلهای مناسب و انتخاب مناسب‌ترین نوزاد برای تولیدهای موفقیت‌آمیز بکار برده می‌شوند. پس از انتخاب رشته‌های والدین بر اساس یک سری استراتژیها، عملگر همپوشانی برای هر کدام از این جفتها بکار برده می‌شود. دو فرد که از همپوشانی بدست آمده‌اند، اکنون با عملکرد جهش، در گام نهایی، برای تشکیل محصول جدید بکار می‌روند. به این ترتیب عملگر جهش نهایتاً الگوریتم ژنتیک را قادر می‌سازد تا به راه حل بهینه نزدیک شود.

معرفی منطقه و داده‌های مورد استفاده

تحقیق حاضر بر اساس داده‌های بارکل رسوب ثبت شده ایستگاه آبسنجی پل یزدکان واقع بر روی رودخانه قطورچای انجام پذیرفته است. رودخانه قطور واقع در شمال غرب حوزه آبریز ارس در مسیر روستای قره تپه با وسعتی معادل ۱۶۷۲ کیلومتر مربع و طول تقریبی ۱۴۷ کیلومتر در مختصات 13.2° $46'$ $44''$ طول شرقی و 33.3° $33'$ $24''$ عرض شمالی گسترده شده است. بازه مورد مطالعه برای سیلاب متوسط سالیانه ($Q=48.7 \text{ m}^3/\text{s}$) دارای عرض متوسط سطح آب ۲۲.۹۷ متر و عمق متوسط ۱.۰۸۶ متر می باشد.

یکی از مراحل آماده سازی و محاسبات عددی برای مدل سازی توسط روش‌های هوش مصنوعی، معیار نمودن یا نرمال سازی داده‌ها به منظور افزایش حساسیت به بزرگی داده‌ها و افزایش قدرت یادگیری می باشد که باعث افزایش توانایی پیش بینی می شود. روش‌های متنوعی جهت معیار نمودن داده‌ها وجود دارد که تفاوت اصلی آن‌ها در دامنه عمل آن‌هاست. در این تحقیق بردارهای ورودی و خروجی

نمایی به عنوان عملگرهای فرعی انتخاب شدند. در مرحله سوم در انتخاب ساختار کروموزوم‌ها، طول سر برابر با ۷ ($h=7$) انتخاب شده و با انجام آزمون و خطا از بین ۳ یا ۴ و ۵ ژن در هر کروموزوم در نهایت ۴ ژن در نظر گرفته شده است. همچنین در هر اجرا تعداد ۳۰ کروموزوم انتخاب گردیده است. در مرحله چهارم که انتخاب تابع پیوند است، ۴ ژن انتخابی از طریق تابع مجموع به هم پیوند داده می‌شوند. در مرحله آخر نیز از ترکیبی از کلیه عملگرهای بهسازی از قبیل جهش، وارون‌سازی، سه نوع ترانهش و ... استفاده شده است (جدول ۳).

شبیه‌سازی و بحث

مشخصات آماری حاصل از اجرای روش برنامه ریزی ژنتیک نیز برای هر دو گروه از الگوهای انتخاب شده معنادار فیزیکی و بی بعد برترتیب در جدول (۴) و (۵) ارائه شده است. با بررسی جدول ۴ قابل مشاهده است که الگوی ورودی شماره ۶ که دبی جریان را مدنظر قرار می‌دهد نسبت به الگوهای دیگر این جدول دارای بیشترین ضریب تعیین و کمترین خطای آزمون (صحت سنجی) است ($DC=0.907$ و $RMSE=0.067$).

معنی فیزیکی و ترکیب‌هایی از آن‌ها نیز بصورت بی بعد جهت مدل‌سازی برنامه‌ریزی ژنتیک انتخاب می‌شوند. الگوهای معنادار فیزیکی و بی بعد انتخاب شده برترتیب در جدول ۱ و جدول ۲ نشان داده شده‌اند.

که در جدول ۲ متغیر Fr نشانگر عدد فرود، Re نشانگر عدد رینولدز و $u_* = \sqrt{ghS}$ سرعت برشی است.

طراحی برنامه ریزی بیان ژن

در پنج مرحله از عملکرد برنامه ریزی بیان ژن، مرحله نخست، شامل تعیین تابع برازش است. در این تحقیق طول محدوده انتخابی برابر با $(R=100)1000$ و مجموعه‌ای شامل ۱۰ برازش ($n=10$) انتخاب گردیده است، بنابراین مقدار تابع برازش برابر با ۱۰۰۰ خواهد بود ($f_{max}=1000$). مرحله دوم، شامل انتخاب مجموعه ترمینال‌ها و مجموعه توابع برای ایجاد کروموزوم‌ها است. مجموعه ترمینال‌ها همان متغیرهای مستقلند که شامل پارامترهای مختلف تاثیر گذار در برآورد رسوب می‌باشد. در انتخاب توابع نیز در الگوسازی‌های مختلف توابع متفاوتی در نظر گرفته شده است. در ادامه از چهار عملگر $\{+, -, \times, \div\}$ به عنوان عملگرهای اصلی انتخاب شده و عملگرهای دیگری همچون توان ۲ و ۳، ریشه سوم و دوم و تابع

جدول ۱- الگوهای انتخاب شده با معنای فیزیکی

Table 1- The selected models with physical conception

number of model; شماره الگو	selected models; الگوهای انتخاب شده
1	$Q_s = Q_s(Q_w, h, D, S, b)$
2	$Q_s = Q_s(Q_w, h, D, S)$
3	$Q_s = Q_s(Q_w, h, D, b)$
4	$Q_s = Q_s(Q_w, h, b, S)$
5	$Q_s = Q_s(Q_w, b, D, S)$
6	$Q_s = Q_s(Q_w)$

جدول ۲- الگوهای انتخاب شده بی بعد

Table 2- The selected none dimensional models

number of model; شماره الگو	selected models; الگوهای انتخاب شده
7	$Q_s = Q_s(h/D, h/b, v^2/gD, Fr, v/u_*, Re)$
8	$Q_s = Q_s(h/D, h/b, v^2/gD, Fr, Re)$
9	$Q_s = Q_s(h/D, h/b, v^2/gD, v/u_*, Re)$
10	$Q_s = Q_s(h/D, h/b, Fr, v/u_*, Re)$
11	$Q_s = Q_s(h/D, v^2/gD, Fr, v/u_*, Re)$
12	$Q_s = Q_s(h/b, v^2/gD, Fr, v/u_*, Re)$
13	$Q_s = Q_s(h/D, h/b, v^2/gD, Fr, v/u_*)$

جدول ۳ - مقادیر پارامترهای مورد استفاده در تخمین بار رسوبی با استفاده از GEP

Table 3- Parameter values used for predicting of sediment load by GEP

پارامتر؛ Parameter	مقدار؛ Value
اندازه سر؛ Head size	7
تعداد کروموزوم ها؛ Chromosomes numbers	30
تعداد ژن ها؛ Number of genes	3-4-5
نرخ جهش؛ Mutation rate	0.044
نرخ وارون سازی؛ Inversion rate	0.1
نرخ ترکیب تک؛ One-point recombination rate	0.3
نرخ ترکیب دو؛ Two-point recombination rate	0.3
نرخ ترکیب ژن؛ Gene recombination rate	0.1
نرخ ترانهش درج متوالی؛ Gene transposition rate	0.1
نرخ ترانهش ریشه درج؛ IS transposition rate	0.1
نرخ ترانهش ژن؛ RIS transposition rate	0.1
معیار خطای تابع؛ Fitness function error type	RMSE
تابع پیوند؛ Linking function	+

گیرد با توجه به نتایج آماری حاصله، قابل مشاهده است که الگوی ورودی شماره ۶ قابلیت پیش بینی بالاتری نسبت به الگوی شماره ۹ دارد، لذا در میان تمام الگوها، الگوی ورودی شماره ۶ به عنوان بهترین مدل پیشنهادی برای تخمین رسوب رودخانه قشورچای معرفی شده و جهت مقایسه با روش‌های دیگر تخمین رسوب بکار می رود.

این بدان معنی است که اگرچه در الگوهای دیگر تعداد پارامترهای بکار رفته بیشتر است ولی این الگوها پیچیده‌ترند و نتایج پیش بینی ضعیف تری دارند. مطابق جدول ۵ نیز الگوی شماره ۹ با دارا بودن ضریب تعیین بیشتر و خطای آزمون کمتر ($DC = 0.694$, $RMSE = 0.081$)، برتری نسبی به دیگر الگوهای بی بعد این جدول دارد. در نهایت در مقایسه‌ای که میان الگوهای برتر این دو گروه انجام می

جدول ۴ - مشخصات آماری نتایج برنامه ریزی ژنتیک برای ورودی‌های معنادار فیزیکی

Table 4- Statistical features of genetic programming results for the inputs by physical concepts

شماره؛ Number	آموزش؛ Train		آزمون؛ Test		الگوهای ورودی؛ imput group
	RMSE	DC	RMSE	DC	
1	0.033	0.877	0.08	0.706	$Q_s = Q_s(Q_w, h, D, S, b)$
2	0.030	0.901	0.072	0.798	$Q_s = Q_s(Q_w, h, D, S)$
3	0.031	0.901	0.0811	0.743	$Q_s = Q_s(Q_w, h, D, b)$
4	0.032	0.893	0.0698	0.775	$Q_s = Q_s(Q_w, h, b, S)$
5	0.0303	0.901	0.071	0.792	$Q_s = Q_s(Q_w, b, D, S)$
6	0.0311	0.812	0.067	0.907	$Q_s = Q_s(Q_w)$

جدول ۵ - مشخصات آماری نتایج برنامه ریزی ژنتیک برای ورودی‌های بی بعد

Table 5- Statistical features of genetic programming results for the none dimensional inputs

شماره؛ Number	آموزش؛ Train		آزمون و صحت سنجی؛ Test		الگوهای ورودی؛ Imput groups
	RMSE	DC	RMSE	DC	
7	0.099	0.636	0.095	0.622	$Q_s = Q_s(h/D, h/b, v^2/gD, Fr, v/u_*, Re)$
8	0.094	0.615	0.088	0.651	$Q_s = Q_s(h/D, h/b, v^2/gD, Fr, Re)$
9	0.092	0.631	0.081	0.694	$Q_s = Q_s(h/D, h/b, v^2/gD, v/u_*, Re)$
10	0.099	0.633	0.089	0.609	$Q_s = Q_s(h/D, h/b, Fr, v/u_*, Re)$
11	0.091	0.710	0.082	0.655	$Q_s = Q_s(h/D, v^2/gD, Fr, v/u_*, Re)$
12	0.094	0.650	0.085	0.617	$Q_s = Q_s(h/b, v^2/gD, Fr, v/u_*, Re)$
13	0.096	0.608	0.087	0.619	$Q_s = Q_s(h/D, h/b, v^2/gD, Fr, v/u_*)$

الگوی شماره ۶ را نشان می دهد.

معادله (۶) رابطه‌ی حاصل از مدل GEP برای بهترین ساختار در

(۶)

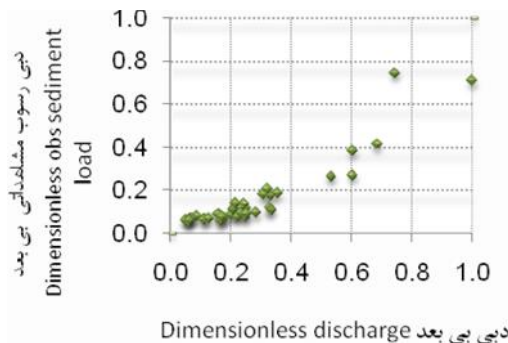
$$Q_s = \left(\frac{((Q/9.95) - 9.95) - (7.12/Q)}{(e^Q + 360.94)} + \frac{(1.32Q^{0.33})}{(57.6/Q)} + 9.4 / (18.8Q^2 - Q + 84.6) \right) + \left(\frac{(0.02Q)}{(7.59 - Q)} \right) / (1.26Q^{0.33})$$

موردنظر که در شکل ۱ نشان داده شده است، استفاده مستقیم از این روش نتایج قابل اعتمادی بدست نمی دهد و نیاز به روش هوشمند الگوریتم ژنتیک که توانایی اعمال شرایط محیطی و اولیه و همچنین قابلیت بهینه سازی دو پارامتر a و b و ارائه چندین منحنی پیش بینی شبیه سازی رسوب را دارد، آشکار می کند.

پس از برآزش منحنی توسط الگوریتم ژنتیک، نتایج آماری حاصله از تحلیل‌های مختلف و حصول چندین جفت از پارامترهای a و b و همچنین نتایج حاصل از منحنی سنجه به روش رگرسیونی در جدول ۶ نشان داده شده است.

مطابق جدول ۶، در میان جفت پارامترهای a و b، جفت پارامتر (a, b) = (۰.۷۵۹, ۱.۲۹۸) با دارا بودن RMSE کمتر (۰.۰۵) و DC بیشتر (۰.۸۸)، قابلیت پیش بینی دقیق تری نسبت به دیگر جفت پارامترهای روش منحنی ژنتیک و همچنین روش رگرسیونی دارد. شکل ۲ رابطه میزان رسوب پیش بینی توسط منحنی ژنتیک، منحنی سنجه و رسوب مشاهداتی را در برابر دبی جریان نشان می دهد.

با توجه الگوی برتر بدست آمده که ورودی در آن دبی جریان است، تشکیل منحنی سنجه ایستگاه مذکور منطقی بنظر می رسد. جهت پیاده سازی منحنی سنجه توسط روش‌های رگرسیونی، از داده های آموزشی برای برآزش این منحنی استفاده شده و سپس رابطه رگرسیونی بدست آمده بر روی داده های صحت سنجی اعمال می شود و در نهایت نتایج آماری بدست آمده از مرحله صحت سنجی برای مقایسه بکار می رود. رابطه رگرسیونی بدست آمده برای این ایستگاه، بصورت $Q_s = 0.11Q^{0.79}$ می باشد. پارامترهای a و b در این روش توسط روش کمترین مربعات بدست می آیند که در طول کل تحلیل ثابت هستند و در نتیجه این روش، فقط یک منحنی جهت تخمین رسوب پیشنهاد می کند. علاوه بر آن، با وجود اهمیت بالای شرایط محیطی و شرایط اولیه در میزان رسوب تولید شده، منحنی حاصل از این روش توانایی اعمال تاثیر این شرایط را در شبیه سازی رسوب مورد نظر ندارد. در نهایت باتوجه به این نکات و همچنین در نظر گرفتن پراکندگی نسبی میزان رسوب در برابر دبی ایستگاه



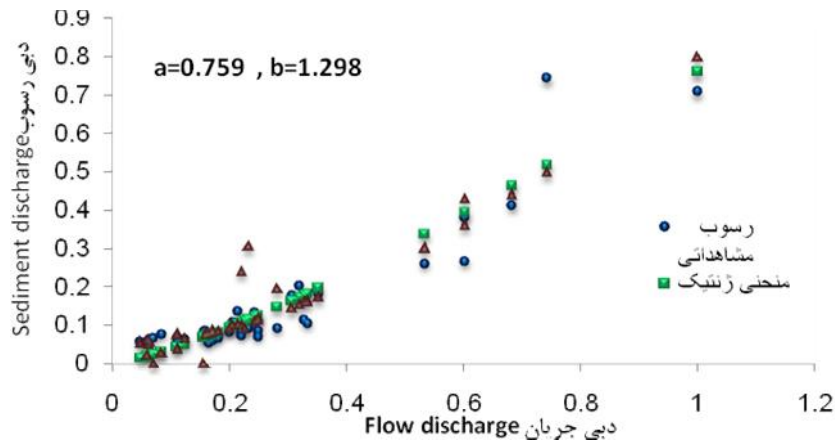
شکل ۱- رابطه دبی جریان با میزان رسوب مشاهداتی

Figure 1- The discharge relation by the observed sediment load

جدول ۶- نتایج آماری حاصل از منحنی سنجه رگرسیونی و منحنی ژنتیک پس از چندین تحلیل

Table 6- Statistical results of rating curve and genetic programming through several runs

روش مورد استفاده؛ Used method	a	b	DC	RMSE
روش رگرسیونی؛ Regression method	0.110	0.79	0.78	0.07
	0.33	0.552	0.52	0.10
	0.759	1.298	0.88	0.05
	0.538	0.957	0.79	0.07
	0.232	0.298	0.29	0.13
روش منحنی ژنتیک؛ Genetic curve method	0.343	0.574	0.55	0.10
	0.577	1.023	0.82	0.06
	0.419	0.738	0.67	0.09
	0.44	0.778	0.69	0.08



شکل ۲- مقادیر محاسباتی روش منحنی ژنتیک و روش منحنی سنجه در برابر مقادیر مشاهداتی
Figure 2- Computed values of genetic curve and rating curve methods against observed values

پراکندگی نیز شبیه سازی شده و برای مقایسه با دیگر روش‌های پیش بینی رسوب بکار رفته است. انتخاب این روابط به گونه ای بوده است که مشخصات ایستگاه موردنظر در محدوده کاربرد این روابط قرار بگیرد. روابط انتخاب شده که به ترتیب توسط انگلاند - هانسن، بگنولد و لارسن پیشنهاد شده اند در جدول ۷ نشان داده شده اند.

همانطور که در شکل ۲ نیز قابل مشاهده است، منحنی ژنتیک با بهینه کردن پارامترهای منحنی سنجه از $a=0.11$ و $b=0.79$ به $a=0.759$ و $b=1.298$ قابلیت پیش بینی رسوب را خصوصاً در مقادیر پایین دبی جریان، بالا برده و نتایج نزدیکتری به مقادیر رسوب مشاهداتی بدست آورده است. در نهایت میزان رسوب تولید شده توسط چند رابطه ریاضی

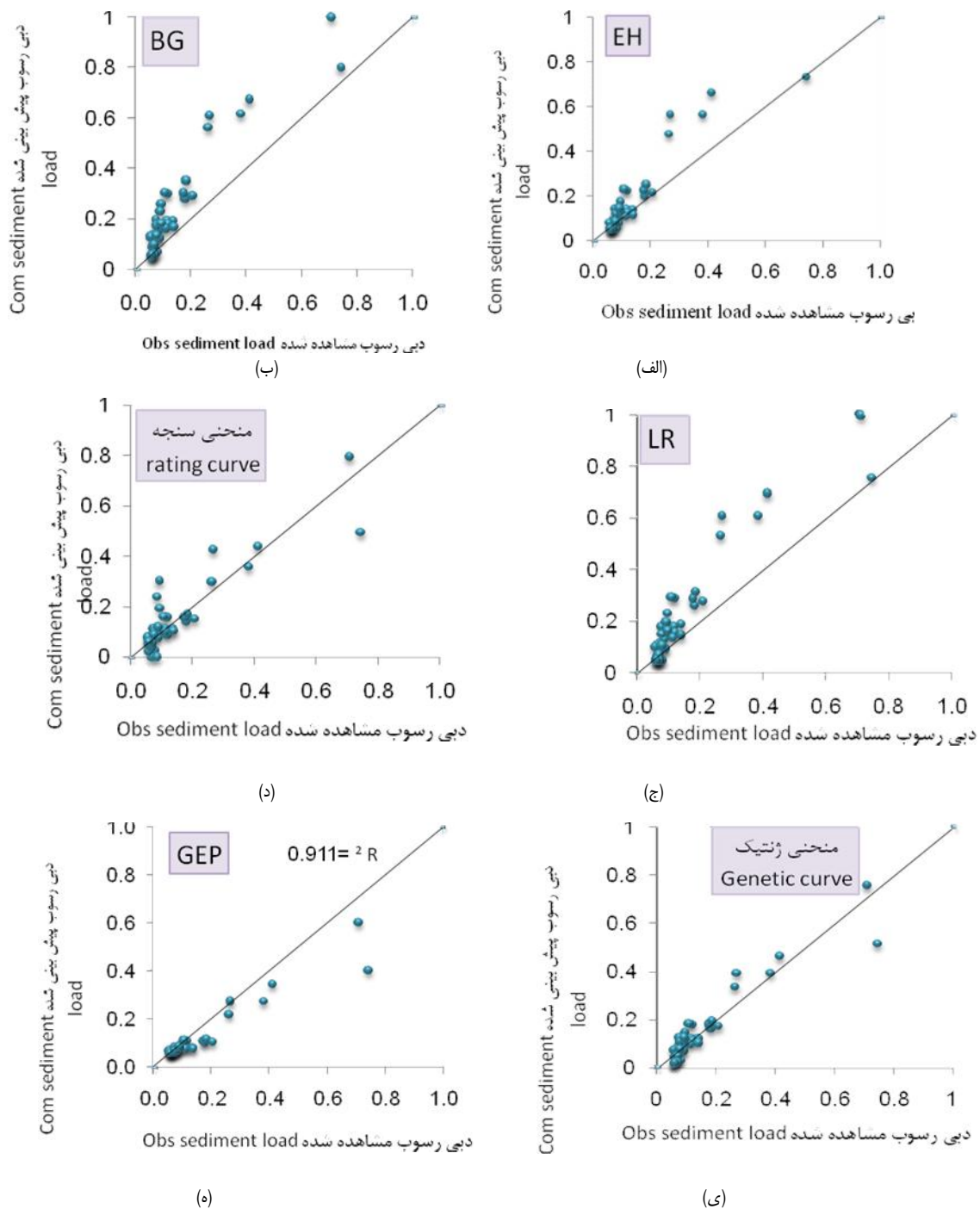
جدول ۷- روابط ریاضی انتخاب شده پر کاربرد

Table 7- The fully applicable mathematical equations

رابطه پیشنهادی offered equation	محقق Researcher
انگلاند-هانسن ۱۹۶۷: Engelund-Hansen 1967	
$Q_s = 0.05Q(G_s/G_s - 1)VS / (\sqrt{(G_s - 1)g D_s})(RS / (G_s - 1)D_s)^{0.5}$	
$Q_s = \dagger V (e_b / \tan \gamma + 0.0IV / \check{S}_s) / (\alpha_s - \alpha_w)$	بگنولد ۱۹۸۸ Bagnold 1988
$Q_s = \alpha Q (D_{50}/h)^{7/6} (\dagger_0 / \dagger_{cr} - 1) f(u_* / \check{S}_s)$ $\dagger_0 = 0.04762^2 \alpha V^2 (D_{50}/h)^{1/3}$	لارسن (۱۹۵۸) Lursen 1958

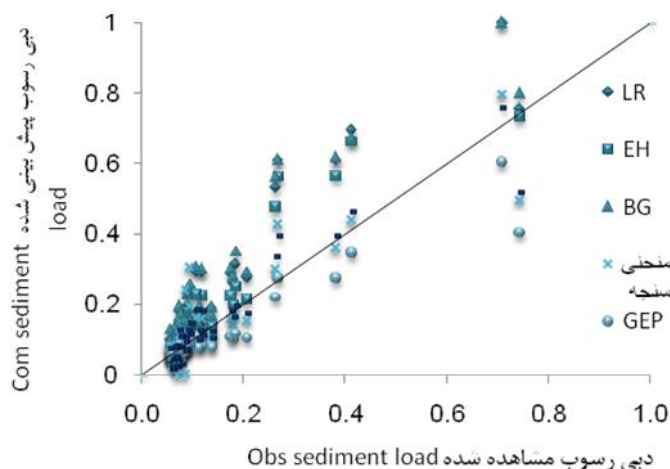
سنجی توسط نمودارهای پراکندگی شکل‌های (الف)، (ب)، (ج)، (د)، (ی) و (ه) از شکل (۳) قابل مشاهده است. شکل (۴) نیز نمودار مقایسه ای تمام روش‌های پیش بینی رسوب را نشان می دهد.

در جدول ۷ پارامترهای α_s ، $\tan \gamma$ و e_b به ترتیب نشانگر سرعت سقوط ذره، نسبت نیروی برشی مماسی به قائم و ضریب بازدهی که تابعی از سرعت جریان است، می باشند. میزان بار رسوبی پیش بینی شده توسط سه رابطه انگلاند - هانسن، بگنولد و لارسن، با نتایج حاصل از مقادیر محاسباتی روش منحنی سنجه، روش الگوریتم ژنتیک و روش برنامه ریزی ژنتیک با مقادیر مشاهداتی مقایسه گشته و نتایج حاصل برای مرحله صحت



شکل ۳- مقادیر پیش بینی رسوب روش های متفاوت (الف) رابطه انگلاند - هانسن (ب) بگنولد (ج) لارسن (د) منحنی سنجه (ی) منحنی ژنتیک (ه) برنامه ریزی ژنتیک که در آن R , EH , BG به ترتیب نمایانگر روش های بگنولد ، انگلاند - هانسن و لارسن هستند

Figure 3- The predicted sediment values by different methods of A) England -Hansen, B) Bagnold, C) Larsen, D) rating curve, E) Genetic curve, F) genetic Programming relations



شکل ۴- نمودار مقایسه‌ای همه روش‌های پیش‌بینی مورد استفاده

Figure 4- The comparing figure of all of the used predicting methods

نتیجه‌گیری

در تحقیق حاضر، نتایج حاصل از بکارگیری روش‌های کلاسیک، روش منحنی سنج و روش‌های الگوریتم تکاملی برنامه ریزی ژنتیک و الگوریتم ژنتیک در پیش‌بینی میزان رسوب رودخانه قطورچای تعیین‌گشته و با یکدیگر مقایسه گردید. نتایج حاکی از عملکرد بهتر روش برنامه‌ریزی ژنتیک و الگوریتم ژنتیک نسبت به روش‌های مذکور است. پس از روش‌های الگوریتم تکاملی، روش‌های منحنی سنج، روش انگلاند - هانسن، لارسن و بگنولد به ترتیب از نظر قابلیت پیش‌بینی رسوب ایستگاه موردنظر رتبه‌بندی می‌شوند. نتایج برتر روش برنامه‌ریزی ژنتیک با ورودی دبی جریان، حاکی از برقراری ارتباط مستقیم میزان رسوب این رودخانه با مشخصات جریان آن است. نتیجه حاصل، با عملکرد مناسب و قابل قبول روش منحنی ژنتیک نیز قابل تأیید است. طوری که روش منحنی ژنتیک قابلیت بالایی در بهینه‌سازی ضرایب منحنی سنج رودخانه موردنظر از خود نشان داده است.

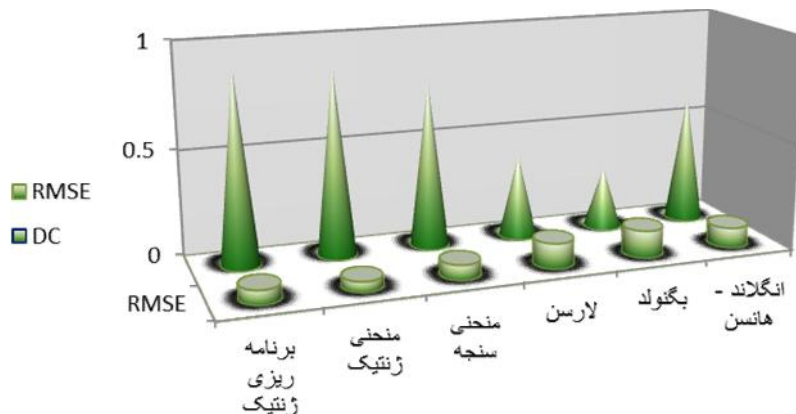
شاخص‌های آماری مربوط به نتایج حاصل از کاربرد روش برنامه‌ریزی ژنتیک، روش منحنی ژنتیک، روش منحنی سنج و روش‌های کلاسیک در پیش‌بینی میزان رسوب برای مرحله صحت‌سنجی در جدول (۸) ارائه شده‌اند. از مقایسه نتایج جدول یاد شده چنین بر می‌آید که در میان تمام روش‌های پیش‌بینی، روش برنامه‌ریزی ژنتیک با دارا بودن RMSE کمتر (۰.۰۶۷) و DC بیشتر (۰.۹۰۷) دقت بالاتری نسبت به دیگر روش‌ها از خود نشان می‌دهد. پس از روش برنامه‌ریزی ژنتیک، روش منحنی ژنتیک که بهینه‌شده روش منحنی سنج می‌باشد در مقایسه با دیگر روش‌ها نتایج نزدیک‌تری به مقادیر اندازه‌گیری شده بدست آورده است (DC=۰.۸۸ و RMSE=۰.۰۵). پس از آن، روش منحنی سنج در مقایسه با روابط کلاسیک، با تقریب کمتری مقادیر اندازه‌گیری شده را شبیه‌سازی کرده است (DC = ۰.۷۸) و (RMSE = ۰.۰۷۲).

در نهایت روش انگلاند - هانسن نیز با داشتن خطای کمتر و ضریب تعیین بیشتر به عنوان روش برتر در میان روش‌های کلاسیک مورد استفاده در این ایستگاه، معرفی می‌شود. جهت مقایسه بهتر، نتایج RMSE و DC حاصل از روش‌های مختلف پیش‌بینی بصورت نموداری در شکل (۵) نشان داده شده است.

جدول ۸- شاخص‌های آماری مربوط به نتایج حاصل از روش‌های مختلف برای مرحله آزمون

Table 8- Statistical indexes of results of different methods for the training stage

انگلاند - هانسن Engelund- Hansen		بگنولد Bagnold		لارسن Lursen		منحنی سنج Rating curve		منحنی ژنتیک Genetic curve		برنامه ریزی ژنتیک Genetic programming	
RMSE	DC	RMSE	DC	RMSE	DC	RMSE	DC	RMSE	DC	RMSE	DC
0.096	0.61	0.13	0.29	0.12	0.39	0.072	0.78	0.05	0.88	0.067	0.907



شکل ۵- مقایسه عملکرد GEP و روش منحنی ژنتیک در مقایسه با روش منحنی سنج و روش های لارسن، بگنولد و انگلاند-هانسن
Figure 5- Comparing of GEP and genetic curve methods performance against rating curve, Larsen, Bagnold and Engelund – Hansen methods

هوشمند در حل مسائل منابع آب توسط سایر محققین نظیر داوسن و ویلیبی (۲)، کالرا و دئو (۵)، کیشی (۶)، لییونگ و همکاران (۷) و بسیاری دیگر موید نتیجه بدست آمده از این تحقیق می باشد. در نهایت می توان نتیجه گرفت که با استفاده از الگوریتم های تکاملی، بررسی های دقیق تر بر روی جریان با میزان دبی بالا و پایین و همچنین استفاده از داده های بیشتر در زمانهای مختلف می تواند موجب تقویت شبیه سازی و ارائه رابطه عمومی تر برای رودخانه مذکور گردد.

به این ترتیب در مجموع مطالعه انجام شده، روش های الگوریتم تکاملی را به عنوان یک روش دقیق و عملی برای پیش بینی میزان رسوب رودخانه موردنظر پیشنهاد می نماید. دقت بالای نتایج حاصل از این روش با توجه به استخراج معادله صریح حاکم بر الگوی منتخب، از یک سو نشانگر سادگی استفاده از این روش و عدم نیاز به دانش بالای مبانی ریاضی الگوریتم، از سوی دیگر نشانگر قابلیت این روش به عنوان یک ابزار مفید و سریع در حل مسائل منابع آب همچون پیش بینی میزان رسوب کل رودخانه می باشد. توانمندی روش های

منابع

- 1- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000a. " Artificial Neural Networks in hydrology. I: Preliminary concepts." J. Hydrologic Eng. ASCE, 5(2), 115-123.
- 2- Dawson C. W. and Wilby. R. 1998. " A comparison of artificial neural network used for river flow forecasting. ", J. of Hydrology and Earth System Sciences, 3(4), 529-540.
- 3- Ferreira C. 2006. Automatically Defined Functions in Gene Expression Programming. In N. Nedjah, L. de M. Mourelle, A. Abraham, eds., Genetic Systems Programming: Theory and Experiences, Studies in Computational Intelligence, Vol. 13, pp. 21-56, Springer-Verlag.
- 4- Ferreira C. 2001. Gene expression programming: A new adaptive algorithm for solving problems, Complex Systems 13(2):87-129.
- 5- Kalra R., and Deo M.C. 2007. Genetic programming to retrieve missing information in wave records along the west coast of India., Applied Ocean Research 29(3), 99-111
- 6- Kisi O. 1999. River flow modeling using artificial network., J. of Hydrology, Vol.214, 32-48.
- 7- Liong S.Y., Gautam T.R., Khu S.T., Babovic V., Keijzer M., and Muttil N. (2002). Genetic programming, A new paradigm in rainfall runoff modeling., J. Am. Water Res. Assoc. 38(3), 705-718.
- 8- Lopes H. S.; and Weinert W. R. 2004. "EGIPSYS: An enhanced gene expression programming approach for symbolic regression problems, International Journal of Applied Mathematics and Computer Science, 14(3):375-384.
- 9- Sarangi A., and Bhattacharya A.K. 2005. Comparison of artificial neural network and regression models for sediment loss prediction from Banha watershed in India, Water technology Center, IARI, Pusa Campus, New Dehli 110012, India.
- 10- Whigham P.A., and Crapper P.F. 2001. Modeling rainfall runoff using Genetic programming. Mathematical and Computer Modeling, 33, 707-721.
- 11- Yang C.T. 1996. Sediment transport theory and practice. McGraw-Hill, USA.
- 12- Zaker Moshfeg M., Ghodsian M., and Montazer Gh. A. 2004. River flow forecasting using artificial neural networks., Proceeding of the hydraulics of dams and river structures. Yazdandoost and Attari (eds). London.



Enhancing Accuracy of Sediment Total Load Prediction Using Evolutionary Algorithms (Case Study: Gotoorchay River)

K. Roshangar^{1*}- M.T. Aalami²- F. Vojoudi Mehrabani³

Received: 11-11-2013

Accepted: 04-29-2014

Introduction: Exact prediction of transported sediment rate by rivers in water resources projects is of utmost importance. Basically erosion and sediment transport process is one of the most complexes hydrodynamic. Although different studies have been developed on the application of intelligent models based on neural, they are not widely used because of lacking explicitness and complexity governing on choosing and architecting of proper network. In this study, a Genetic expression programming model (as an important branches of evolutionary algorithms) for predicting of sediment load is selected and investigated as an intelligent approach along with other known classical and imperial methods such as Larsen's equation, Engelund-Hansen's equation and Bagnold's equation.

Materials and Methods: In this study, in order to improve explicit prediction of sediment load of Gotoorchay, located in Aras catchment, Northwestern Iran latitude: $38^{\circ}24'33.3''$ and longitude: $44^{\circ}46'13.2''$), genetic programming (GP) and Genetic Algorithm (GA) were applied. Moreover, the semi-empirical models for predicting of total sediment load and rating curve have been used. Finally all the methods were compared and the best ones were introduced. Two statistical measures were used to compare the performance of the different models, namely root mean square error (RMSE) and determination coefficient (DC). RMSE and DC indicate the discrepancy between the observed and computed values.

Results and Discussions: The statistical characteristics results obtained from the analysis of genetic programming method for both selected model groups indicated that the model 4 including the only discharge of the river, relative to other studied models had the highest DC and the least RMSE in the testing stage (DC= 0.907, RMSE= 0.067). Although there were several parameters applied in other models, these models were complicated and had weak results of prediction. Our results showed that the model 9, with the most DC and the least RMSE (DC=0.694, RMSE= 0.081), had the relative advantage to the other none dimensional models. Finally it is clear that the model 6 had more predicting capability rather than the model 9, so among all the models, model 6 was the best referring model for estimation of sediment load of the Gotoorchay river.

Conclusion: It was observed that the model including only the discharge of the Gotoorchay river the best model for estimation of sediment load and it was applied for comparing all the other sediment predicting models such as some classic methods that includes Larsen's equation, Engelund-Hansen's equation and Bagnold's equation and optimized rating curve. Among all methods, it was concluded that the genetic programming was superior to other methods in predicting sediment load of the mentioned river. Therefore, genetic programming that is a branch of evolutionary algorithms, with high prediction capability was offered as a powerful tool for optimizing and explicit predicting of total sediment load of the Gotoorchay River.

Keywords: Genetic Programming, Optimization of Prediction, Rating Curve, Sediment Load

1, 2 and 3- Associate Professor, Professor and M.Sc. Graduated of Civil Hydraulics Engineering, Faculty of Civil Hydraulics Engineering, University of Tabriz, Tabriz, Iran, Respectively

(* Corresponding Author Email: Kroshangar@yahoo.com)