

بهبود عملکرد مدل‌های هوشمند بر پایه الگوریتم موجک و تبدیلات لگاریتمی در تخمین بار رسوب معلق

رضا حاجی‌آبادی¹ - سعید فرزین^{2*} - یوسف حسن‌زاده³

تاریخ دریافت: 1393/07/03

تاریخ پذیرش: 1393/11/04

چکیده

یکی از دلایل پیچیدگی تخمین و پیش‌بینی پدیده‌های هیدرولوژیکی و به خصوص سری‌های زمانی وجود ویژگی‌هایی نظیر روند، نویز و نوسانات با فرکانس بالا در آن‌ها می‌باشد که با استفاده از پیش‌پردازش داده‌ها به وسیله نویززدایی و تبدیلات لگاریتمی، می‌توان برخی عوامل پنهان و تاثیرگذار در این پیچیدگی را شناسایی و حذف نمود و یا درک این ویژگی‌ها را برای مدل‌های پیش‌بینی ساده‌تر نمود. در این تحقیق با استفاده از دو مدل هوشمند برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی تخمین بار رسوب معلق مورد بررسی قرار می‌گیرد، سپس میزان تاثیر دو رویکرد نویززدایی و تبدیلات لگاریتمی به عنوان پیش‌پردازشگر، در بهبود نتایج مورد ارزیابی و مقایسه قرار می‌گیرد. به منظور نویززدایی از تبدیلات موجک استفاده شده است. نتایج تحقیق نشان می‌دهد پس از نویززدایی، معیار نش - ساتکلیف در شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن به ترتیب 0/15 و 0/14 افزایش داشته و مقدار جذر میانگین مجذورات خطانیز در شبکه عصبی مصنوعی از 199/24 به 141/17 میلی‌گرم بر لیتر و در برنامه‌ریزی بیان ژن از 234/84 به 193/89 میلی‌گرم بر لیتر کاهش یافته است. تاثیر رویکرد تبدیلات لگاریتمی نیز در بهبود نتایج شبکه عصبی مصنوعی تا حدود زیادی مشابه با رویکرد نویززدایی می‌باشد. در حالی که در برنامه‌ریزی بیان ژن تاثیر نامطلوب داشته و پس از تبدیلات لگاریتمی Ln و Log، معیار نش - ساتکلیف از 0/57 به ترتیب به 0/31 و 0/21 کاهش یافته است و مقدار جذر میانگین مجذورات خط نیز از 234/84 میلی‌گرم بر لیتر به ترتیب به 298/41 میلی‌گرم بر لیتر و 318/72 میلی‌گرم بر لیتر افزایش یافته است.

واژه‌های کلیدی: برنامه‌ریزی بیان ژن، پدیده‌های هیدرولوژیکی، شبکه عصبی مصنوعی، نویززدایی

مقدمه

نروفازی و برنامه‌ریزی بیان ژن اشاره کرد که در سال‌های اخیر مطالعات زیادی در این زمینه صورت گرفته است. سلاجقه و فتح‌آبادی (20) از دو روش شبکه عصبی و منطق فازی برای برآورد رسوب معلق رودخانه کرج بهره بردند که نتایج حاکی از برتری روش منطق فازی نسبت به شبکه عصبی داشت. داننده مهر و همکاران (4) توانایی دو مدل برنامه‌ریزی ژنتیک و شبکه عصبی مصنوعی را در پیش‌بینی بار رسوب معلق مورد بررسی قرار دادند و نشان دادند که مدل برنامه‌ریزی ژنتیک از دقت بالاتری نسبت به شبکه عصبی مصنوعی برخوردار است. دستورانی و همکاران (5)، ناگی و همکاران (15) و آلپ و سی‌گی‌زوگلو (1) با استفاده از شبکه عصبی مصنوعی به پیش‌بینی و بررسی بار رسوب معلق پرداختند که نتایج آنها نشان دهنده توانایی بالای این مدل در پیش‌بینی‌ها می‌باشد. آیتک و کیشی (3) از برنامه‌ریزی ژنتیک (GP) برای پیش‌بینی بار رسوب معلق و استخراج رابطه آن با دبی جریان استفاده کردند و نتایج آنها حاکی از برتری این روش نسبت به روش‌های سنتی نظیر منحنی سنج و

یکی از مهم‌ترین مسائل هیدرولوژیکی تخمین بار رسوب معلق رودخانه‌ها می‌باشد. تخمین بار رسوبات معلق رودخانه‌ها نقش اساسی در مسائل مربوط به مدیریت منابع آب و طراحی سازه‌های هیدرولیکی دارد و ماهیت غیرخطی و همچنین دامنه تغییرات وسیع در غلظت بار رسوبی، ایجاب می‌کند که یک روش غیرخطی برای برآورد بار رسوبی به کار گرفته شود. از روش‌های مناسب برای پیش‌بینی و تخمین این پدیده می‌توان به روش‌های هوشمند نظیر شبکه عصبی مصنوعی،

1- دانشجوی دکتری گروه آب، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران

2- استادیار گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان

* - نویسنده مسئول: (Email: saeed.farzin@semnan.ac.ir)

3- استاد گروه آب، دانشکده مهندسی عمران، دانشگاه تبریز

مصنوعی به تنهایی دارد. کیشی و سیمن (11) نیز نشان دادند که ترکیب موجک و ماشین بردار پشتیبان در پیش‌بینی دبی جریان ماهانه منجر به نتایج با دقت بالاتری نسبت به ماشین بردار پشتیبان می‌شود. کیشی و شیری (12) با ترکیب موجک با دو مدل برنامه ریزی بیان ژن و نروفازی باعث افزایش دقت قابل توجه نتایج این دو روش در پیش‌بینی بارش روزانه شدند. حسن‌زاده و همکاران (8) نیز در مطالعه‌ای خشکسالی را بر اساس شاخص بارش استاندارد شده (SPI) در شهر تبریز با استفاده از ترکیب موجک و نظریه آشوب مورد بررسی و پیش‌بینی قرار دادند و به نتایج دقیق‌تری دست یافتند.

یکی دیگر از روش‌های مناسب در آموزش موثرتر مدل‌های هوشمند استفاده از داده‌های انتقال داده شده توسط توابع لگاریتمی است. لاک و همکاران (13) با استفاده از شبکه عصبی مصنوعی به پیش‌بینی کوتاه مدت بارش در یک حوضه شهری پرداختند سپس به منظور آموزش موثرتر شبکه عصبی از داده‌های انتقال داده شده توسط تابع لگاریتم استفاده کردند که نتایج آنها نشان دهنده دقت بهتر این مدل‌ها نسبت به مدل‌های آموزش داده شده با داده‌های خام بود. آکیل و همکاران (2) نیز از دو مدل شبکه عصبی مصنوعی و نروفازی در پیش‌بینی رواناب در مقیاس ساعتی و روزانه بهره بردند. آنها نیز از تابع لگاریتم به منظور آموزش موثرتر مدل‌های هوشمند استفاده کردند و نشان دادند که استفاده از تابع لگاریتم منجر به نتایجی با دقت بهتر می‌گردد.

هدف از تحقیق حاضر توسعه مدل‌ساز برای رسوب معلق‌روزانه با بهره‌گیری از روش‌های هوشمند نظیر برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی می‌باشد در ادامه، تاثیر نوپزدایی توسط تبدیلات موجک و پیش‌پردازش توسط تبدیلات لگاریتمی به منظور آموزش موثرتر مدل‌ها، در بهبود نتایج مطرح می‌شوند که تاکنون بحث در این مورد کمتر صورت گرفته است. همچنین به منظور نوپزدایی توسط تبدیلات موجک از شش نوع مختلف توابع خانواده دابچی⁴ و به منظور پیش‌پردازش داده‌ها از دو نوع تابع لگاریتمی با پایه مختلف استفاده شده است تا تابع مناسب و بهینه نیز در هر مدل هوشمند مشخص گردد.

مواد و روش‌ها

در این مطالعه تخمین بار رسوبات معلق یک رودخانه به عنوان یک نمونه پدیده هیدرولوژیکی در نظر گرفته شده است. به منظور آموزش روش‌های هوشمند و صحت‌سنجی مدل‌ها، از داده‌های معتبر به مدت 5 سال (2004-2009)، شامل دبی جریان آب و بار رسوب معلق روزانه مربوط به یک ایستگاه هیدرومتری در رودخانه

رگرسون‌های چند متغیره می‌باشد. رجایی و همکاران (18) نیز از دو مدل شبکه عصبی مصنوعی و نروفازی برای پیش‌بینی بار رسوب معلق استفاده کردند که هر دو مدل نتایج مناسبی را ارائه کردند. ملسه و همکاران (14) نیز به مقایسه شبکه عصبی مصنوعی، رگرسون‌های چند متغیره و مدل خودرگرسو میانگین متحرک در پیش‌بینی بار رسوب معلق پرداختند که نتایج حاکی از برتری مدل شبکه عصبی داشت. کاکایی لعدانی و همکاران (9) از شبکه عصبی مصنوعی و ماشین بردار پشتیبان¹ در پیش‌بینی بار رسوب معلق بهره بردند و نتایج قابلیت بالای هر دو مدل را نشان داد.

با وجود توانایی بالای روش‌های هوشمند در تخمین پدیده‌های هیدرولوژیکی نظیر بار رسوب معلق برخی اوقات این روش‌ها منجر به نتیجه دلخواه و مناسب نمی‌شوند و نیاز به افزایش دقت این مدل‌ها می‌باشد. وجود ویژگی‌هایی نظیر روند²، نویز³، غیر خطی و غیر ایستا بودن در بسیاری از پدیده‌های هیدرولوژیکی به خصوص بار رسوب معلق موجب پیچیدگی بیشتر این پدیده‌ها گشته و از طرف دیگر با توجه به متکی بودن مدل‌های هوشمند به داده‌های مورد استفاده و یا به عبارتی داده‌گرا بودن این مدل‌ها، کیفیت و ویژگی خود این داده‌ها نیز از اهمیت بالایی برخوردار هستند.

تمام داده‌های اندازه‌گیری شده دارای مقداری نویز می‌باشند که قسمت ناخواسته داده‌ها می‌باشند. این نویزها را می‌توان به دو دسته کلی تقسیم کرد؛ نویزهای اندازه‌گیری که مربوط به نحوه اندازه‌گیری داده‌های مشاهداتی است و نویزهای دینامیکی که فرآیندی مربوط به ماهیت داده‌های مورد مطالعه می‌باشد و طی این فرآیند در هر گام زمانی یک مقدار تصادفی کوچک منجر به ایجاد نویز در سیستم می‌شود. در واقع این نویزها در اندازه‌گیری مقادیر خروجی یک سیستم تاثیر می‌گذارد، باعث بروز خطا در سیستم می‌شود و همچنین منجر می‌گردد که پیش‌بینی داده‌ها چندان دقیق نباشد. بنابراین استفاده از روش‌های نوپزدایی برای حذف نویز و پیش‌پردازش داده‌ها و یا روشی که بتواند این نویزها و نوسانات را حذف و یا درک آنها را برای مدل فراهم کند در بهبود کارایی این مدل‌ها تاثیرگذار است. به همین دلیل محققین همواره به دنبال راه‌حلی برای بهبود نتایج مدل‌های هوشمند بوده‌اند. مطالعات گذشته نشان می‌دهد که ترکیب روش‌های هوشمند و تبدیل موجک در بسیاری از پدیده‌های هیدرولوژیکی منجر به کسب نتایج بهتری می‌شود. پارتال و همکاران (17)، کیشی (10)، رجایی و همکاران (19) و نورانی و همکاران (16) به منظور برآورد و پیش‌بینی بار رسوب معلق روزانه از ترکیب موجک و شبکه عصبی مصنوعی استفاده کردند که نتایج آنها نشان از برتری مدل ترکیبی موجک و شبکه عصبی نسبت به مدل‌های شبکه عصبی

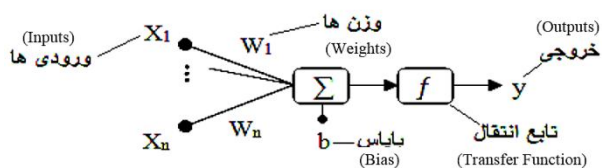
1- Support vector machine

2- Trend

3- Noise

4- Daubchies

شده است. در این تحقیق نیز یک شبکه پرسپترون سه لایه در محیط نرم افزار متلب کد نویسی شده است. الگوریتم یادگیری می تواند منجر به نتایج متفاوتی از شبکه می شود. منظور از الگوریتم یادگیری و یا آموزش نحوه اتصال وزن دار نرون ها و چگونگی تنظیم پارامترهای آنها می باشد. طی فرآیند آموزش، این وزن ها و مقادیر ثابتی که با آنها جمع می شوند و اصطلاحاً با یاس⁶ نامیده می شوند، بر اساس میل به خطای کمتر به طور پی در پی تغییر می کنند تا اینکه یکی از شروط خاتمه آموزش ارضا گردد. مدل عمومی شبکه های پرسپترون، شبکه های جلو رونده⁷ با روال آموزش انتشار به عقب⁸ است. شبکه های جلو رونده به شبکه هایی اطلاق می شود که در آنها تمام نرون های لایه اول، به لایه بعدی متصل بوده و در هر سطح این مسئله صادق بوده تا به لایه خروجی برسد. پس از مشخص شدن خروجی شبکه، ابتدا وزن های لایه آخر تصحیح شده و پس از آن و به ترتیب اوزان لایه های قبل تصحیح می شوند که به همین دلیل به آنها شبکه های با روال انتشار به عقب می گویند (21). شکل (1) یک مدل غیرخطی از یک نرون مصنوعی تنها را نشان می دهد.



شکل 1- مدل غیرخطی یک نرون مصنوعی
Figure 1- nonlinear model of a simple neuron

فرآیند صورت گرفته توسط این نرون را می توان از نظر ریاضی به صورت رابطه (1) بیان نمود.

$$y = f\left(\sum_{i=1}^n W_i X_i + b\right) \quad (1)$$

که در آن X_i ورودی (خروجی لایه نرون قبل) و W_i وزن متناظر با هر ورودی، b بایاس، f تابع انتقال و y خروجی نرون است. در این تحقیق نرمال سازی داده ها نیز با استفاده از رابطه (2) صورت گرفته است.

$$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

که در آن X_n داده نرمال شده و X_{\min} مینیمم داده در بازه مورد نظر و X_{\max} ماکزیمم داده در بازه مورد نظر و X_i داده نرمال نشده است.

Skunk واقع در ایالت آیوا آمریکا استفاده می گردد. ابتدا با استفاده از دو روش هوشمند برنامه ریزی بیان ژن و شبکه عصبی مصنوعی دقت تخمین بار رسوب معلق روزانه مورد بررسی قرار می گیرد سپس پس از نویززدایی به وسیله تبدیلات موجک و تبدیلات لگاریتمی به عنوان پیش پردازشگر تخمین مجدداً صورت گرفته و میزان تاثیر آنها در بهبود نتایج مدل های هوشمند بررسی می شود.

برنامه ریزی بیان ژن (GEP)¹

برنامه ریزی بیان ژن (GEP) اولین بار توسط فریرا در سال 1999 معرفی گردید. برنامه ریزی بیان ژن شبیه به الگوریتم ژنتیک² (GA) و برنامه ریزی ژنتیک³ (GP) است (7). به این دلیل شبیه به الگوریتم ژنتیک است زیرا هر دو از جمعیت افراد استفاده می کنند و افراد این جمعیت را بر اساس توابع برازش انتخاب می کنند و با استفاده از یک یا چند اپراتور تنوع ژنتیکی افراد جمعیت را تامین می کنند. تفاوت اساسی بین این سه الگوریتم در ماهیت افراد جمعیت نهفته شده است. در الگوریتم ژنتیک افراد به صورت رشته خطی با طول ثابت (کروموزوم ها) می باشند و در برنامه ریزی ژنتیک افراد به صورت نهادهای غیرخطی با اندازه و اشکال مختلف می باشند اما در برنامه ریزی بیان ژن افراد در ابتدا به صورت رشته های خطی با طول ثابت هستند که پس از آن به صورت نهادهای غیرخطی با اندازه و اشکال مختلف بیان می شوند (7). فرآیند پیش بینی در برنامه ریزی بیان ژن دارای مراحل است. قدم اول انتخاب تابع برازش است. دومین قدم انتخاب ترمینال ها (ورودی ها) است و مرحله آخر انتخاب مجموعه عملگرها می باشد. انتخاب مجموعه عملگرها در این تحقیق به صورت مجموعه $\{+, -, /, \times, \sqrt{\quad}, power\}$ در نظر گرفته شده است. در این تحقیق برای این مدل از جعبه ابزار GeneXproTools 4.0 استفاده شده است.

شبکه عصبی مصنوعی (ANN)⁴

شبکه های عصبی مصنوعی از عناصر عملیاتی ساده ای ساخته شده اند که به صورت موازی در کنار هم عمل می کنند و یک مدل ساده شده از مغز انسان هستند. یکی از ساده ترین و در عین حال کارآمدترین معماری پیشنهادی برای استفاده در مدل سازی عصب های واقعی، مدل پرسپترون چند لایه یا به اختصار MLP⁵ می باشد که از یک لایه ورودی، یک لایه خروجی و یک یا چند لایه پنهان تشکیل

- 1- Gene Expression Programming
- 2- Genetic Algorithm
- 3- Genetic Programming
- 4- Artificial Neural Network
- 5- Multi Layers Perceptron

6- Bias

7- Feed Forward

8- Back Propagation Training

تبدیل موجک¹

زیراستفاده کنیم:

$$\psi_{k,j}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - ks_0^j \tau_0}{s_0^j}\right) \quad (7)$$

که k و j مقادیر صحیح هستند و $s_0 > 1$ یک گام تاخیری ثابت است. فاکتور تبدیل τ_0 بستگی به گام تاخیری دارد. پس از ساده‌سازی رابطه فوق به رابطه زیر خواهیم رسید:

$$\psi_{k,j}(t) = s_0^{-j/2} \psi(s_0^{-j} t - k\tau_0) \quad (8)$$

با استفاده از روابط اشاره شده در نهایت ضرایب گسسته موجکی به صورت رابطه (9) به دست می‌آیند:

$$DWT_s^\psi(t) = \sum s(t) \psi_{k,j}^*(t) \quad (9)$$

تبدیلات لگاریتمی

همان‌طور که اشاره شد به دلیل متکی بودن مدل‌های هوشمند به داده‌ها و یا به عبارتی داده‌گرا بودن آن‌ها، خصوصیات و ویژگی‌های داده‌های مورد استفاده در دقت این مدل‌ها تاثیرگذار است. به منظور افزایش توانایی و دقت مدل‌های هوشمند در مرحله آموزش داده‌ها، استفاده از یک روش پیش پردازش مناسب برای آموزش موثرتر مورد نیاز می‌باشد. مطالعات گذشته نشان می‌دهد که شبکه‌های آموزش دیده با داده‌های پیش پردازش شده توسط تبدیلات مختلف از جمله تبدیلات لگاریتمی دارای دقت و عملکرد بهتری می‌باشند و معمولاً سریع‌تر همگرا می‌شوند (2 و 13). تبدیلات گوناگونی می‌تواند برای این منظور استفاده شود. در این تحقیق نیز برای آموزش موثرتر مدل‌های هوشمند از دو نوع تابع لگاریتمی با پایه‌های مختلف استفاده شده است تا تاثیر پایه تبدیل نیز بررسی شود. ابتدا تغییر مقیاس داده‌ها توسط روابط (10) و (11) انجام شده سپس مدل آموزش داده شده است.

$$Q, S \rightarrow Z_1 \rightarrow \begin{cases} Z_{1q} = \log(Q + b) \\ Z_{1s} = \log(S + b) \end{cases} \quad (10)$$

$$Q, S \rightarrow Z_2 \rightarrow \begin{cases} Z_{2q} = \ln(Q + b) \\ Z_{2s} = \ln(S + b) \end{cases} \quad (11)$$

در این روابط Z_{1q} ، Z_{1s} ، Z_{2q} و Z_{2s} مقادیر انتقال داده شده دبی (Q) و رسوب (S) توسط دو تابع لگاریتمی (یکی بر پایه ده و دیگری بر پایه عدد نپر) هستند و b نیز یک عدد ثابت برای حذف داده با مقدار صفر است زیرا این توابع در صفر تعریف نشده‌اند. در انتها و پس از آموزش مدل‌ها نیز نتایج آن به حالت اولیه برگردانده شده است.

معیارهای ارزیابی مدل

برای ارزیابی مدل بهینه از سه معیار خطا استفاده گردیده است:

در آنالیز و پردازش سیگنال‌ها به منظور تجزیه و کاهش نویز معمولاً از تبدیل فوریه استفاده می‌شود اما مشخصات زمانی و مکانی سیگنال ممکن است در طول انجام این تبدیل از بین رفته و تغییر کنند و پاره‌ای از خصوصیات سیگنال دچار مشکل شوند به طوری که تشخیص زمان و مکان یک پدیده خاص ناممکن باشد. تبدیل موجک اولین بار در سال 1984 توسط گراسمن² مطرح گردید. پس از آن در سال 1987 توسط داپیز³ جهت رفع مشکلات تبدیلات فوریه ارائه گردید (6). تبدیل موجک در شناسایی و تحلیل بسیاری از زوایای پنهان سیگنال‌ها نسبت به دیگر روش‌های تحلیل سیگنال از توانایی قابل قبولی برخوردار است. از کاربردهای اصلی تبدیل موجک در سری‌های زمانی می‌توان به نوینزدایی اشاره نمود. حذف نویز باید به صورتی باشد که سیگنال نوینزدایی شده تا حد زیادی به سیگنال اصلی شبیه باشد به طوری که خواص زمانی و مکانی اصلی سیگنال دچار تغییر نشود که این امر به کمک تبدیل موجک امکان‌پذیر گشته است.

$\psi(t)$ ، یک تابع موجک است اگر و تنها اگر تبدیل فوریه آن $\psi(t)$ در شرط صدق کند:

$$\int_{-\infty}^{+\infty} \frac{|\psi(\omega)|}{|\omega|^2} d\omega < +\infty \quad (3)$$

برای ارضای شرط بالا توسط موجک باید معادله زیر برقرار باشد:

$$\psi(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (4)$$

این خاصیت تابع با میانگین صفر می‌باشد. $\psi(t)$ یک تابع موجک مادر است که در تجزیه و تحلیل با دو عملگر ریاضی انتقال (τ) و مقیاس (s) در طول سیگنال مورد نظر، اندازه و مکان آن تغییر می‌کند.

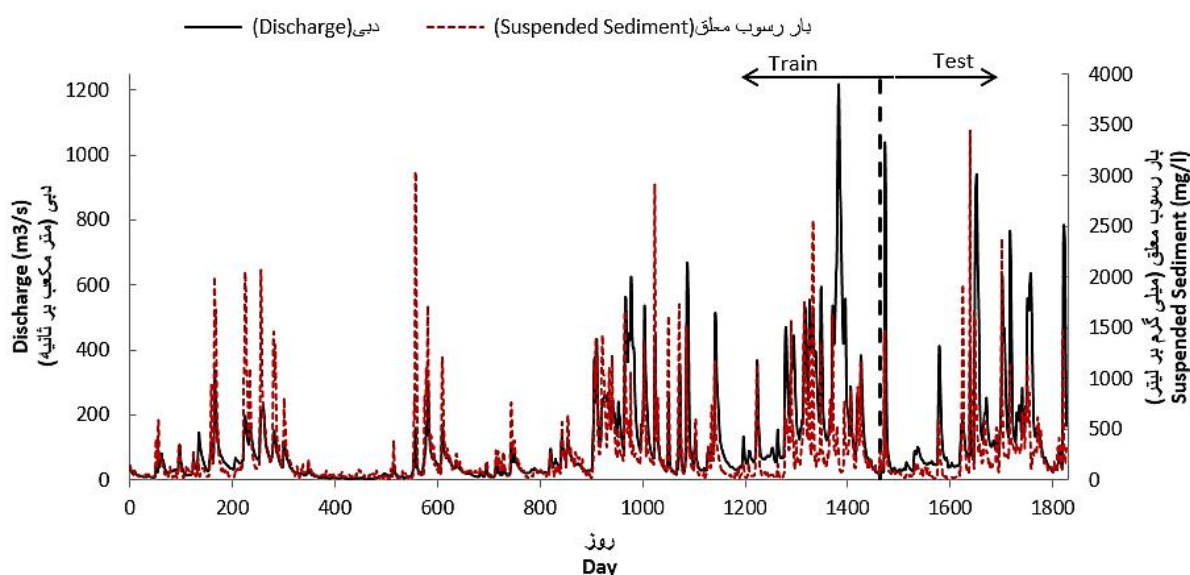
$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (5)$$

با استفاده از این تبدیل الگوریتم موجک پیوسته سیگنال طبق رابطه زیر به دست می‌آید:

$$C(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-\tau}{s}\right) dt = \int_{-\infty}^{+\infty} f(t) \psi_{s,\tau}(t) dt \quad (6)$$

با گسسته کردن پارامترهای مقیاس و انتقال، موجک پیوسته به موجک گسسته تبدیل خواهد شد. برای این منظور کافی است به جای مقادیر پارامترهای مقیاس و انتقال از مقادیر گسسته آنها طبق رابطه

- 1Wavelet Transform
- 2Grassman
- 3Daubechies



شکل 2- مقادیر دبی و رسوب ایستگاه مورد مطالعه در دوره‌های آموزش و صحت‌سنجی
Figure 2- Discharge and suspended sediment values in train and test periods

می‌باشد و از داده‌های روزانه معتبر به مدت 5 سال (2004-2009)، شامل دبی جریان آب و رسوب استفاده شده است. ایستگاه در طول جغرافیایی "37°16'91 غربی و عرض جغرافیایی "13°45'40 شمالی قرار گرفته است. برای آموزش مدل‌ها از داده‌های 4 سال (80 درصد کل داده‌ها) به صورت روزانه استفاده شده است و داده‌های یک سال (20 درصد کل داده‌ها) نیز به صورت روزانه برای صحت‌سنجی در نظر گرفته شده است. شکل (2) سری‌های زمانی دبی و بار رسوب معلق روزانه مربوط به دوره‌های آموزش و صحت‌سنجی را نمایش می‌دهد.

ساختار نوین‌زدایی

در این تحقیق به منظور تجزیه سیگنال‌ها (سری‌های زمانی دبی و رسوب) از تبدیل موجک گسسته تنها در سطح اول استفاده شده است به طوری‌که پس از عبور سیگنال‌ها از فیلترهای بالاگذر و پایین‌گذر، سیگنال تنها به دو بخش تجزیه خواهد شد. بخش اول حاصل از عبور سیگنال از فیلتر پایین‌گذر است که شامل اطلاعات فرکانس پایین و مشخصات اصلی و هویتی سیگنال است و کلیات نامیده می‌شود. بخش دوم در نتیجه عبور سیگنال از فیلتر بالاگذر است که شامل اطلاعات فرکانس بالا می‌باشد و جزئیات نام دارد. شکل (3) سری زمانی دبی مورد استفاده در مرحله آموزش را نشان می‌دهد که توسط تابع $Db2$ به دو بخش کلیات و جزئیات تقسیم شده است.

میانگین خطای مطلق¹:

$$MAE = \frac{1}{n} \sum_{i=1}^n |S_{oi} - S_{pi}| \quad (12)$$

جذر میانگین مجذورات خطا²:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_{oi} - S_{pi})^2} \quad (13)$$

معیار نش - ساتکلیف³:

$$NS = 1 - \left(\frac{\sum_{i=1}^n (S_{oi} - S_{pi})^2}{\sum_{i=1}^n (S_{oi} - \bar{S}_o)^2} \right) \quad (14)$$

در این روابط n تعداد داده‌ها، S_{oi} مقدار رسوب مشاهده شده، S_{pi} مقدار رسوب پیش‌بینی شده، \bar{S}_o میانگین رسوبات مشاهداتی است.

منطقه مورد مطالعه

داده‌های منطقه مورد مطالعه مربوط به یک ایستگاه هیدرومتری واقع در منطقه Iowa آمریکا است که از سازمان نقشه‌برداری زمین‌شناسی آمریکا⁴ تهیه شده است. این ایستگاه بر روی رودخانه Skunk قرار دارد. مساحت حوضه این رودخانه 11168 کیلومتر مربع

1- Mean Absolute Error

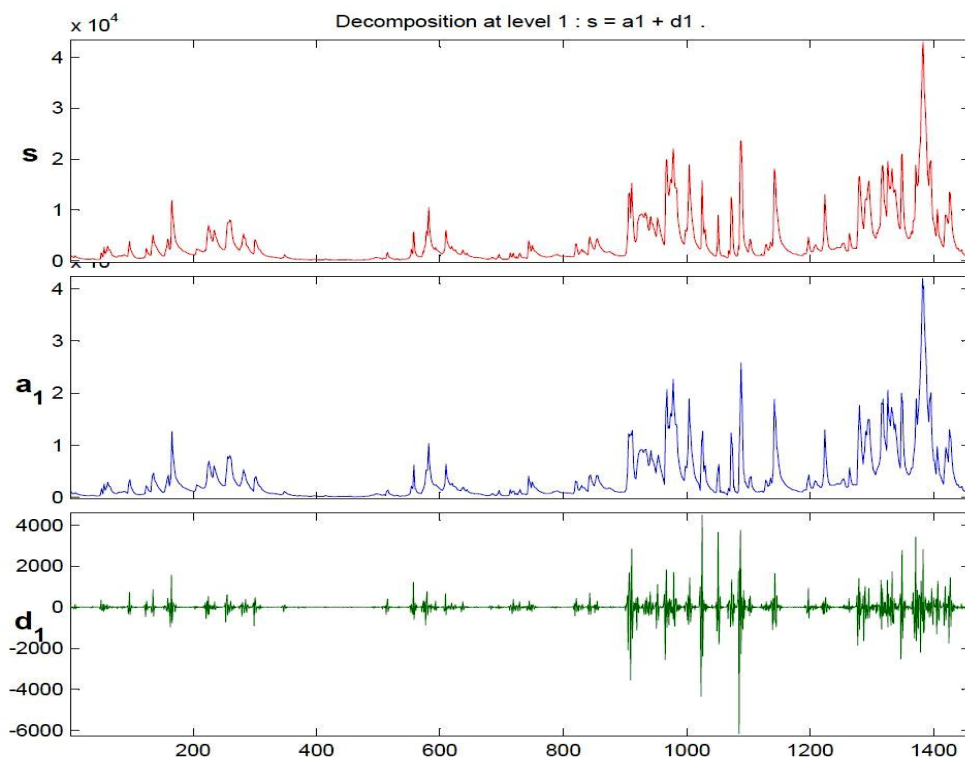
2- Root Mean Squared Error

3- Nash-Sutcliffe

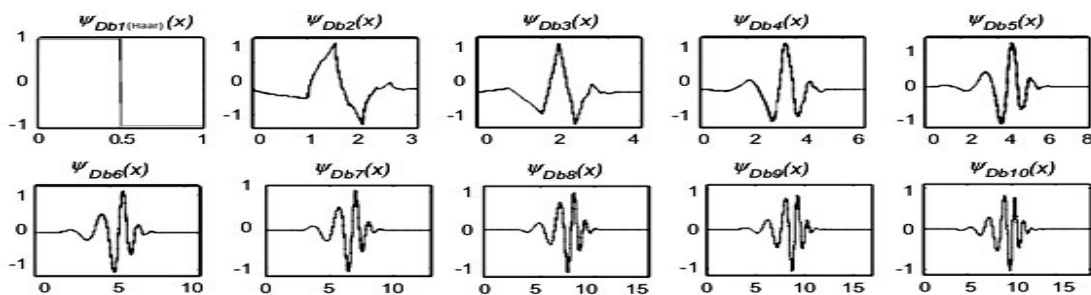
4- United States Geological Survey

5- Approximation

6- Detail



شکل 3- سری زمانی دبی تجزیه شده در سطح اول به وسیله Db2
Figure 3- Decomposition of Discharge time series at level 1 by Db2



شکل 4- توابع موجک خانواده دابچیز
Figure 4- Daubechies wavelet family

است تا بهترین تابع نیز برای هر روش هوشمند مشخص گردد (شکل 4).

نتایج و بحث

برای پیش‌بینی بار رسوب معلق در مدل‌های هوشمند از ترکیبات ورودی مختلف استفاده شده است و پارامترهای خطا برای هر ترکیب محاسبه گردیده است (جدول 1). این ورودی‌ها بر اساس همبستگی بین بار رسوب معلق، دبی جریان و تاخیرهای آنها انتخاب شده‌اند و شامل سه تاخیر روزانه بار رسوب معلق (رسوب روز قبل، دو روز قبل و

در شکل (3)، S نشان دهنده سیگنال اصلی، a_1 سیگنال کلیات و d_1 سیگنال جزئیات است. پس از تجزیه تمام سیگنال‌های ورودی، قسمت جزئیات آنها حذف و تنها قسمت کلیات برای آموزش مدل‌های هوشمند استفاده می‌گردد. در نهایت تاثیر نویززدایی سیگنال‌ها در نتایج روش‌های هوشمند به وسیله سه معیار ارزیابی معرفی شده مورد بررسی قرار گرفته است. در این تحقیق به منظور نویززدایی توسط تبدیلات موجک از شش نوع اول توابع خانواده دابچیز¹ که به اختصار با (Db) نمایش داده می‌شوند استفاده شده

1- Daubchies

سه روز قبل)، جریان دبی و دو تاخیر روزانه آن (دبی روز قبل و دو روز قبل) می‌باشند.

جدول 1- برآورد خطای ترکیبات ورودی مختلف در مدل‌های هوشمند

Table 1- Error of different input sets at intelligent models

مدل Model	ترکیبات ورودی Input Set	NS	MAE(mg/lit)	RMSE(mg/lit)
ANN	1. Q_t	-0.38	319.02	424.38
	2. Q_t, S_{t-1}	0.48	154.4	259.05
	3. Q_t, Q_{t-1}, S_{t-1}	0.42	128.05	273.66
	4. $Q_t, Q_{t-1}, S_{t-1}, S_{t-2}$	0.61	106.94	223.08
	5. $Q_t, Q_{t-1}, Q_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}$	0.69	99.42	199.24
GEP	1. Q_t	-0.54	335.68	447.56
	2. Q_t, S_{t-1}	0.41	134.34	275.33
	3. Q_t, Q_{t-1}, S_{t-1}	0.47	117.32	260.72
	4. $Q_t, Q_{t-1}, S_{t-1}, S_{t-2}$	0.57	97.9	234.84
	5. $Q_t, Q_{t-1}, Q_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}$	0.41	148.46	275.72

ژن تا حدودی دارای عملکرد مطلوب‌تری می‌باشد. در ادامه با اعمال نویززدایی و تبدیلات لگاریتمی به عنوان پیش‌پردازشگر بر روی بهترین ترکیب ورودی در هر مدل هوشمند، میزان تاثیر در بهبود نتایج مورد بررسی قرار گرفته است (جدول 2).

همانطور که جدول (1) نشان می‌دهد بهترین ترکیب در هر مدل مشخص شده است. در شبکه عصبی مصنوعی ترکیب پنجم و در برنامه‌ریزی بیان ژن ترکیب چهارم کمترین خطا را داشته‌اند. همچنین نتایج نشان می‌دهد شبکه عصبی مصنوعی نسبت به برنامه‌ریزی بیان

جدول 2- برآورد خطا در بهترین ترکیب مدل‌های هوشمند پس از نویززدایی توسط توابع مختلف دابچیز

Table 2- Error of best input set at intelligent models after denoising by Daubechies wavelet family

توابع موجک Wavelet	مدل Model	NS	MAE(mg/lit)	RMSE(mg/lit)
Db1(Harr)	ANN-5	0.72	98.91	190.12
	GEP-4	0.71	86.21	193.89
Db2	ANN-5	0.84	81.56	141.17
	GEP-4	0.65	90.16	212.03
Db3	ANN-5	0.81	81.3	155.25
	GEP-4	0.6	111.18	227.94
Db4	ANN-5	0.78	92.91	167.98
	GEP-4	0.49	125.43	255.82
Db5	ANN-5	0.75	96.25	179.52
	GEP-4	0.33	132.25	293.85
Db6	ANN-5	0.72	88.97	187.86
	GEP-4	0.68	84	201.15

جدول 3- برآورد خطا در بهترین ترکیب مدل‌های هوشمند پس از

آموزش توسط داده‌های انتقالی

Table 3- Error of best input set at intelligent models after train by logarithmic transformations

مدل (Model)	NS	MAE(mg/lit)	RMSE(mg/lit)
ANN-5-Log	0.77	70.35	170.96
ANN-5-Ln	0.82	65.34	150.08
GEP-4-Log	0.21	101.39	318.72
GEP-4-Ln	0.31	93.94	298.41

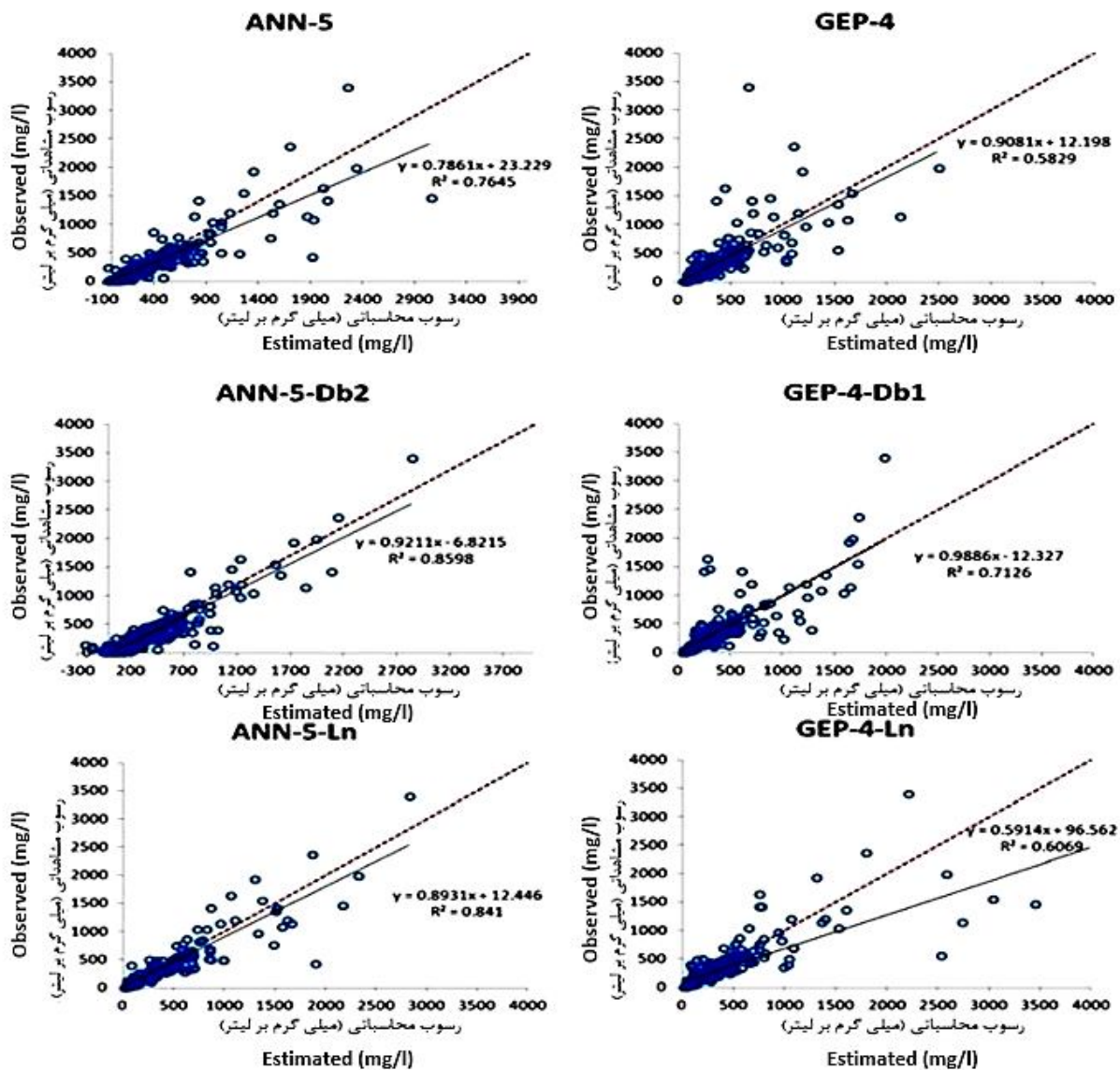
نتایج جدول (3) حاکی از آن است که در مدل شبکه عصبی مصنوعی پیش‌پردازش توسط تبدیلات لگاریتمی تاثیر قابل توجهی

نتایج جدول (2) نشان می‌دهد که نویززدایی تاثیر قابل توجهی در بهبود نتایج داشته است به طوری که بهترین نتیجه در شبکه عصبی مصنوعی مربوط به تابع (Db2) و در برنامه‌ریزی بیان ژن مربوط به تابع (Db1) می‌باشد.

در پیش‌پردازش توسط توابع لگاریتمی برخلاف نویززدایی سری زمانی تجزیه نمی‌شود بلکه تنها داده‌ها از فضای به فضایی دیگر انتقال داده خواهند شد. پس از آموزش شبکه با داده‌های پیش‌پردازش شده توسط توابع لگاریتمی، در انتها نتایج شبکه مجدداً به فضای اولیه برگردانده خواهد شد. همان‌طور که اشاره شد در این تحقیق این پیش‌پردازش توسط دو تابع لگاریتمی صورت گرفته است (جدول 3).

بهترین تابع نویززدایی در شبکه عصبی مصنوعی ($Db2$) می‌باشد. همچنین از نتایج می‌توان دریافت که تبدیل لگاریتمی بر پایه عدد نپر نسبت به تبدیل لگاریتمی معمولی دقت مطلوب‌تری ارائه می‌دهد. شکل (5) پراکنش مقادیر محاسباتی توسط مدل‌های مختلف را در مقابل مقادیر مشاهداتی نمایش می‌دهد.

در بهبود نتایج داشته است در حالی که در مدل برنامه‌ریزی بیان ژن تبدیلات لگاریتمی نه تنها تاثیر مطلوب نداشته است بلکه تا حدود اندکی منجر به افزایش خطا گردیده است. علاوه بر این خطای مدل شبکه عصبی مصنوعی آموزش داده شده با داده‌های انتقالی با وجود سادگی روش، تنها اندکی بیشتر از خطای نتایج مربوط به



شکل 5- پراکنش مقادیر محاسباتی در مقابل مقادیر مشاهداتی در مدل‌های مختلف
Figure 5- Scatter plot of estimated values versus observed values in the different models

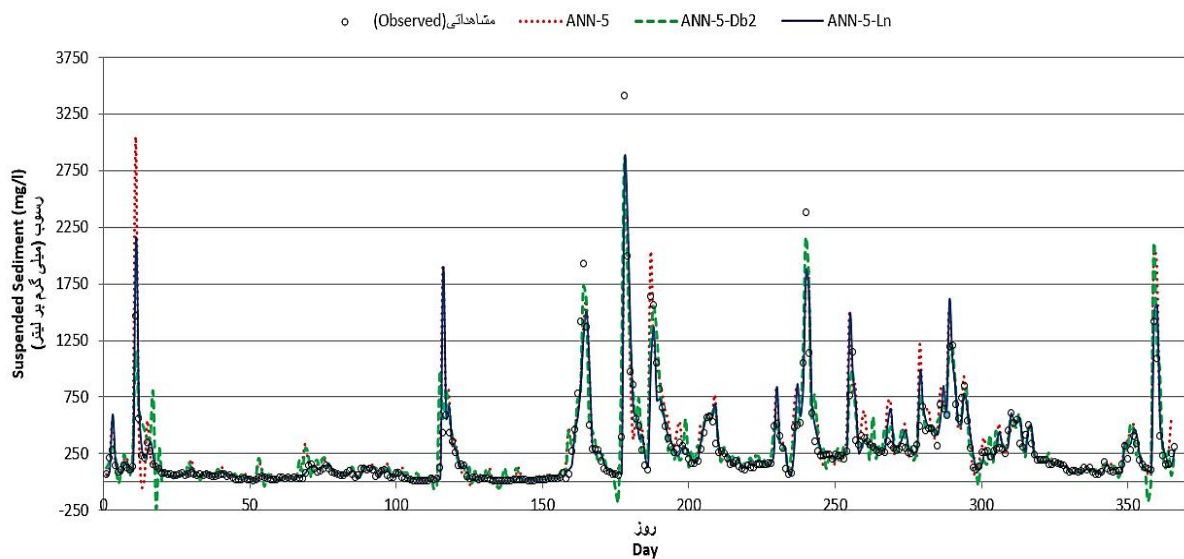
مختلف را نمایش می‌دهند. شکل (6) و شکل (7) نیز مقادیر پیش‌بینی شده توسط مدل‌های مختلف را نمایش می‌دهند.

نتایج مدل برنامه‌ریزی بیان ژن در نهایت منجر به ارائه یک رابطه ریاضی بین ورودی‌ها و خروجی نیز می‌گردد. روابط (15) و (16) و (17) رابطه ریاضی به دست آمده از برنامه‌ریزی بیان ژن در حالات

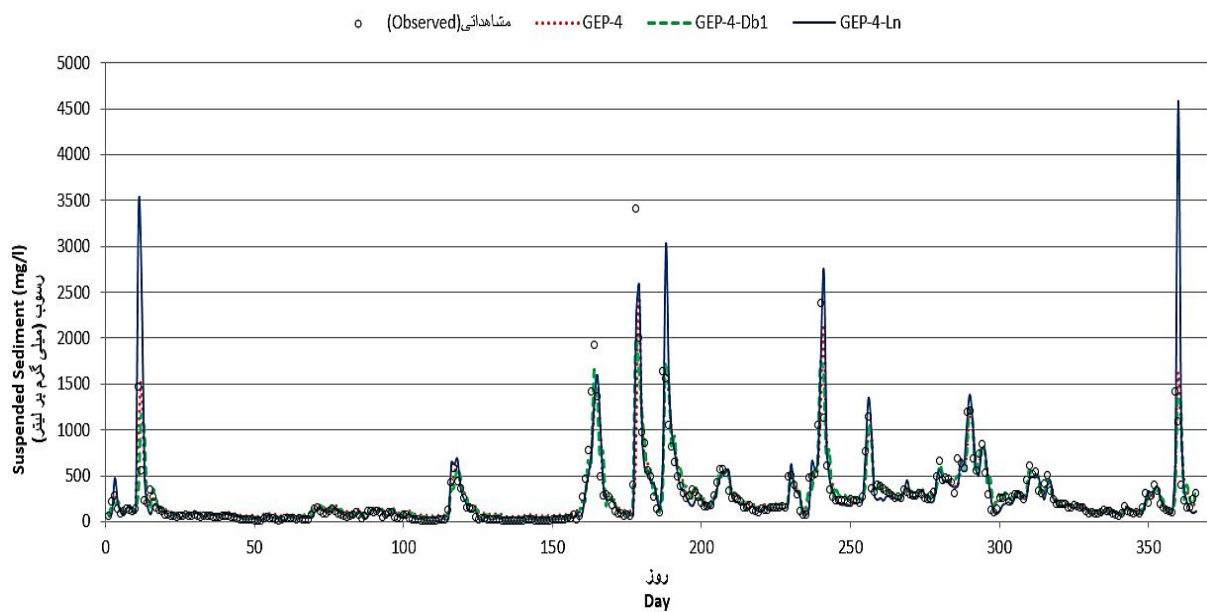
$$S_t(GEP - 4) = \frac{Q_t}{\sqrt{Q_{t-1}}} + 9.083465 - \left[\left(\frac{S_{t-1} \times Q_{t-1}}{Q_t} \right) \times (4.15) \right] + S_{t-1} \quad (15)$$

$$S_t(GEP - 4 - Db1) = S_{t-1} - \left[\left(\frac{S_{t-2}}{5.458} \right) \times \left(\frac{Q_t}{Q_{t-1}} \right) \right] + \sqrt{Q_t} \quad (16)$$

$$S_t(GEP - 4 - Ln) = 0.284515 + S_{t-1} - (0.057678 \times S_{t-2}) + Q_t - Q_{t-1} \quad (17)$$



شکل 6-مقادیر رسوب پیش‌بینی شده توسط مدل‌های مختلف شبکه عصبی مصنوعی
Figure 6- Estimated suspended sediment values in the different ANN models



شکل 7- مقادیر رسوب پیش‌بینی شده توسط مدل‌های مختلف برنامه‌ریزی بیان ژن
Figure 7- Estimated suspended sediment values in the different GEP models

نتیجه‌گیری کلی

عصبی مصنوعی آموزش دیده با سری زمانی پیش پردازش شده توسط تبدیلات لگاریتمی نیز نشان می‌دهد که نه تنها با وجود سادگی عملکرد توابع لگاریتمی نسبت به نویززدایی توسط تبدیلات موجک نتایج تا حدود زیادی (اندکی کمتر از نویززدایی) بهبود یافته است بلکه بر خلاف حالت نویززدایی، داده پیش‌بینی شده منفی نیز وجود ندارد. این در حالی است که پیش پردازش توسط تبدیلات لگاریتمی در مدل برنامه‌ریزی بیان ژن بهبودی در نتایج ارائه نکرده است و تا حدودی باعث افزایش خطا نیز گردیده است به طوریکه معیار نش-ساتکلیف از 0/57 در بهترین ترکیب برنامه‌ریزی بیان ژن پس از تبدیلات لگاریتمی Ln و Log به ترتیب به 0/31 و 0/21 کاهش یافته است و مقدار $RMSE$ نیز از 234/84 میلی‌گرم بر لیتر پس از تبدیلات لگاریتمی Ln و Log به ترتیب به 298/41 میلی‌گرم بر لیتر و 318/72 میلی‌گرم بر لیتر افزایش یافته است. در این تحقیق تاثیر نویززدایی توسط الگوریتم موجک و پیش‌پردازش توسط تبدیلات لگاریتمی در بهبود نتایج دو روش شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن بررسی گردید اما به منظور مطالعات تکمیلی پیشنهاد می‌گردد تاثیر روش‌های پیش‌پردازش دیگر و همچنین تاثیر پیش‌پردازش بر دیگر روش‌های هوشمند با ماهیت متفاوت‌تری نظیر ماشین بردار پشتیبان مورد بررسی قرار گیرد. همچنین می‌توان این مطالعات را برای پدیده‌های هیدرولوژیکی دیگر نظیر پیش‌بینی دبی و تبخیر و تعرق انجام داد.

به دلیل داده‌گرا بودن روش‌های هوشمند، خصوصیات داده‌ها در دقت این روشها تاثیرگذار است و منجر به پیچیدگی این مدل‌ها شده است. با استفاده از پیش‌پردازش داده‌ها توسط تبدیلات لگاریتمی و نویززدایی، می‌توان برخی عوامل پنهان و تاثیرگذار در این پیچیدگی را شناسایی و حذف نمود. در این تحقیق مدل‌سازی غلظت رسوب معلق رودخانه *Skunk* واقع در ایالت آیوا آمریکا با بهره‌گیری از دو مدل هوشمند شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن مورد بررسی قرار گرفت و در ادامه، تاثیر نویززدایی توسط تبدیلات موجک و پیش‌پردازش توسط تبدیلات لگاریتمی، در بهبود نتایج مطرح گردید. نتایج نشان داد که دقت هر دو مدل هوشمند شبکه عصبی و برنامه‌ریزی بیان ژن که با سری زمانی نویززدایی شده توسط تبدیلات موجک آموزش داده شده‌اند نسبت به حالت بدون نویززدایی افزایش قابل توجهی داشته است و ترکیب این دو مدل هوشمند با الگوریتم موجک منجر به افزایش کارایی این مدل‌ها گردیده است به طوریکه مقدار معیار نش-ساتکلیف در بهترین ترکیب در مدل شبکه عصبی مصنوعی 0/15 و در مدل برنامه‌ریزی بیان ژن 0/14 افزایش داشته است. همچنین معیار خطای $RMSE$ در شبکه عصبی مصنوعی پس از نویززدایی از 199/24 به 141/17 میلی‌گرم بر لیتر و در برنامه‌ریزی بیان ژن از 234/84 به 193/89 میلی‌گرم بر لیتر کاهش یافته است. معیار MAE نیز رفتاری مشابه با $RMSE$ دارد. نتایج مربوط به شبکه

منابع

- 1- Alp M., and Cigizoglu, H.K. 2007. Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. *Environmental Modelling & Software*, 22: 2-13.
- 2- Aqil M., Kita I., Yano A., and Nishiyama, S. 2007. A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. *Journal of Hydrology*, 337: 22-34.
- 3- Aytek A., and Kisi, O. 2008. A genetic programming approach to suspended sediment modelling. *Journal of Hydrology*, 351: 288-298.
- 4- Danandehmehr A., Oliaie E., Ghorbani M.A. 2010. Suspended sediment load prediction based on river discharge and genetic programming method. *Watershed Management Researches Journal (Pajouhesh & Sazandegi)*, 88: 44-54. (in Persian with English abstract)
- 5- Dastorani M.T., Azimi Fashi Kh., Talebi A., Ekhtesasi M.R. 2012. Estimation of suspended sediment using artificial neural network (case study: Jamishan Watershed in kermanshah). *Journal of Watershed Management Research*, 6: 66-74. (in Persian with English abstract)
- 6- Daubechies I. 1992. Ten lectures on wavelets. Society for Industrial Mathematics.
- 7- Ferreira C. 2001. Gene expression programming a new adaptive algorithm for solving problems. *Complex Systems*, 13(2): 87-129.
- 8- Hassanzadeh Y., Lotfollahi-Yaghin M.A., Shahverdi S., Farzin S., Farzin N. 2013. De-noising and prediction of time series based on wavelet algorithm and chaos theory (case study: SPI drought monitoring index of Tabriz city). *Iran-water resources Research*, 3: 1-13. (in Persian with English abstract)
- 9- Kakaie Lafdani E., Moghaddam Nia A., and Ahmadi A. 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of Hydrology*, 478: 50-62.
- 10- Kisi O. 2010. Daily suspended sediment estimation using neuro-wavelet models. *International Journal of Earth Sciences*, 99: 1471-1482.
- 11- Kisi O., and Cimen M. 2011. A wavelet-support vector machine conjunction model for monthly streamflow

- forecasting. *Journal of Hydrology*, 399: 132 –140.
- 12- Kisi O., and Shiri J. 2011. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resource management*, 25: 3135 –3152.
- 13- Luk K.C., Ball J.E., and Sharma A. 2000. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology*, 227:56-65.
- 14- Melesse A.M., Ahmad S., McClain M.E., Wang X., and Lim Y.H. 2011. Suspended sediment load prediction of river systems: An artificial neural network approach. *Agricultural Water Management*, 98: 855-866.
- 15- Nagy H.M., Watanabe K., and Hirano M. 2002. Prediction of Sediment Load Concentration in Rivers using Artificial Neural Network Model. *Journal of Hydraulic Engineering*, 128: 588-595.
- 16- Nourani V., Yahyavi Rahimi A., and Hassan Nejad F. 2013. Conjunction of ANN and threshold based wavelet denoising approach for forecasting suspended sediment load. *International Journal of Management & Information Technology*, 3(1): 9 –26.
- 17- Partal T., and Cigizoglu H.K. 2008. Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. *Journal of Hydrology*, 358: 317 –331.
- 18- Rajae T., Mirbagheri S.A., Zounemat-Kermani M., and Nourani V. 2009. Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the Total Environment*, 407: 4916-4927.
- 19- Rajae T., Nourani V., Zounemat-Kermani M., and Kisi O. 2011. River suspended sediment load prediction: Application of ANN and Wavelet conjunction model. *Journal of Hydrologic Engineering*, 16(8): 613-627.
- 20- Salajegheh A., Fathabadi A. 2008. Estimation of the suspended sediment load of Karaj River using fuzzy logic and neural networks. *Journal of Range and Watershed Management*, 62: 271-282. (in Persian with English abstract)
- 21- Yu H.H., and Jenq N.H. 2002. *Handbook of Neural Network Signal Processing*. CRC Press.

Intelligent Models Performance Improvement Based on Wavelet Algorithm and Logarithmic Transformations in Suspended Sediment Estimation

R. Hajiabadi¹- S. Farzin^{2*}- Y. Hassanzadeh³

Received:25-09-2014

Accepted:24-01-2015

Introduction One reason for the complexity of hydrological phenomena prediction, especially time series is existence of features such as trend, noise and high-frequency oscillations. These complex features, especially noise, can be detected or removed by preprocessing. Appropriate preprocessing causes estimation of these phenomena become easier. Preprocessing in the data driven models such as artificial neural network, gene expression programming, support vector machine, is more effective because the quality of data in these models is important. Present study, by considering diagnosing and data transformation as two different preprocessing, tries to improve the results of intelligent models. In this study two different intelligent models, Artificial Neural Network and Gene Expression Programming, are applied to estimation of daily suspended sediment load. Wavelet transforms and logarithmic transformation is used for diagnosing and data transformation, respectively. Finally, the impacts of preprocessing on the results of intelligent models are evaluated.

Materials and Methods In this study, Gene Expression Programming and Artificial Neural Network are used as intelligent models for suspended sediment load estimation, then the impacts of diagnosing and logarithmic transformations approaches as data preprocessor are evaluated and compared to the result improvement. Two different logarithmic transforms are considered in this research, LN and LOG. Wavelet transformation is used to time series denoising. In order to denoising by wavelet transforms, first, time series can be decomposed at one level (Approximation part and detail part) and second, high-frequency part (detail) will be removed as noise. According to the ability of gene expression programming and artificial neural network to analysis nonlinear systems; daily values of suspended sediment load of the Skunk River in USA, during a 5-year period, are investigated and then estimated. 4 years of data are applied to models training and one year is estimated by each model. Accuracy of models is evaluated by three indexes. These three indexes are mean absolute error (MAE), root mean squared error (RMSE) and Nash-Sutcliffe coefficient (NS).

Results and Discussion In order to suspended sediment load estimation by intelligent models, different input combination for model training evaluated. Then the best combination of input for each intelligent model is determined and preprocessing is done only for the best combination. Two logarithmic transforms, LN and LOG, considered to data transformation. Daubechies wavelet family is used as wavelet transforms. Results indicate that diagnosing causes Nash Sutcliffe criteria in ANN and GEP increases 0.15 and 0.14, respectively. Furthermore, RMSE value has been reduced from 199.24 to 141.17 (mg/lit) in ANN and from 234.84 to 193.89 (mg/lit) in GEP. The impact of the logarithmic transformation approach on the ANN result improvement is similar to diagnosing approach. While the logarithmic transformation approach has an adverse impact on GEP. Nash Sutcliffe criteria, after Ln and Log transformations as preprocessing in GEP model, has been reduced from 0.57 to 0.31 and 0.21, respectively, and RMSE value increases from 234.84 to 298.41 (mg/lit) and 318.72 (mg/lit) respectively. Results show that data denoising by wavelet transform is effective for improvement of two intelligent model accuracy, while data transformation by logarithmic transformation causes improvement only in artificial neural network. Results of the ANN model reveal that data transformation by LN transfer is better than LOG transfer, however both transfer function cause improvement in ANN results. Also denoising by different

1-PhD Student, Department of Water Resources, Faculty of Civil Engineering, Iran University of Science and Technology

2- Assistant Professor, Department of Water and Hydraulic Structures, Faculty of Civil Engineering, University of Semnan

(*-Corresponding Author Email: saeed.farzin@semnan.ac.ir)

3- Professor, Department of Water Resources, Faculty of Civil Engineering, University of Tabriz

wavelet transforms (Daubechies family) indicates that in ANN models the wavelet function Db2 is more effective and causes more improvement while on GEP models the wavelet function Db1 (Harr) is better.

Conclusions: In the present study, two different intelligent models, Gene Expression Programming and Artificial Neural Network, have been considered to estimation of daily suspended sediment load in the Skunk river in the USA. Also, two different procedures, denoising and data transformation have been used as preprocessing to improve results of intelligent models. Wavelet transforms are used for diagnosing and logarithmic transformations are used for data transformation. The results of this research indicate that data denoising by wavelet transforms is effective for improvement of two intelligent model accuracy, while data transformation by logarithmic transformation causes improvement only in artificial neural network. Data transformation by logarithmic transforms not only does not improve results of GEP model, but also reduces GEP accuracy.

Keywords: Artificial neural network, Gene expression programming, Logarithmic transformations, Suspended sediment load, Wavelet transformation