



بررسی کارایی روش دسته‌بندی گروهی داده‌ها و تبدیل موجک در پیش‌بینی رواناب (مطالعه موردی: حوضه آبریز قره‌سو)

محمد رضا گودرزی^۱، حسام گودرزی^۲

تاریخ ارسال: ۱۳۹۷/۰۴/۲۳

تاریخ پذیرش: ۱۳۹۸/۰۳/۰۱

مقاله پژوهشی و کاربردی

چکیده

امروزه بکارگیری مدل‌های داده محور ابزارهای جدیدی برای شبیه سازی و مدل‌سازی در علوم مختلف می‌باشد. فرآیند بارش- رواناب از مهم‌ترین و پیچیده‌ترین پدیده‌ها در چرخه هیدرولوژی است. در این مطالعه ضمن معرفی مدل ترکیبی موجک-دسته‌بندی گروهی داده‌ها، کارایی آن جهت مدل سازی فرآیند بارش-رواناب حوضه آبریز قره سو مورد مطالعه قرار گرفت. در ابتدا سری‌های زمانی بارش و رواناب با استفاده از تبدیل موجک به چندین زیر سری تجزیه گشته تا بر نایب ستایی آن غلبه گردد. سپس این زیر سری‌های زمانی به عنوان ورودی روش دسته‌بندی گروهی داده‌ها برای پیش‌بینی رواناب روزانه در نظر گرفته شده است. کارایی مدل ترکیبی با شاخص‌های ضریب تبیین (DC) و ریشه میانگین مربعات خطا (RMSE) ارزیابی شدند. نتایج حاصل از مدل‌ها بیانگر آن است که بیشترین مقدار ضریب تبیین و مقدار ریشه میانگین خطا برای مدل منفرد GMDH به ترتیب ۰/۶۵ و ۰/۰۷ و برای مدل ترکیبی به ترتیب ۰/۹۱ و ۰/۰۵ است. دلیل برتری مدل ترکیبی نسبت به مدل منفرد ناشی از این است که مدل ترکیبی دسته‌بندی گروهی داده‌های موجکی، به جای استفاده از سری‌های زمانی داده‌های بارش و رواناب در یک مقیاس کلی، از چندین زیر سری پردازش شده زمانی با درجات تجزیه مختلف به عنوان ورودی در مدل استفاده می‌نماید. همچنین نتایج نشان داد که مدل ترکیبی Wavelet-GMDH در مقایسه با سایر مدل‌های ترکیبی مانند شبکه عصبی مصنوعی موجکی (WANN) به سبب عملکرد لایه‌ای مدل GMDH که دربرگیرنده ترکیبات دوتایی از متغیرهای ورودی است و با انتخاب تعداد نرون‌های بهینه در هر لایه حرکت به سمت داده‌های پیش‌بینی شده را جهت‌دهی می‌نماید، دارای کارایی و دقت بیشتری است.

واژه‌های کلیدی: روش دسته‌بندی گروهی داده‌ها، تبدیل موجک، تابع موجک، مدل‌سازی بارش-رواناب، حوضه آبریز قره‌سو.

^۱: استادیار دانشکده عمران، دانشگاه یزد، یزد، ایران، تلفن: ۰۹۸۹۱۲۲۱۷۰۷۹۵، پست الکترونیکی: Goodarzimr@yazd.ac.ir نویسنده

مسوول

^۲ دانشجو، کارشناسی ارشد مهندسی آب و سازه‌های هیدرولیکی، گروه مهندسی عمران، دانشگاه آیت اله العظمی بروجردی (ره)، بروجرد، ایران، تلفن: ۰۹۸۹۳۹۳۱۸۲۷۵۳، پست الکترونیکی: Hesam.swt@gmail.com



مقدمه

بحران آب و مسائل پیرامون آن نشان دهنده اهمیت مدیریت پایدار منابع آب می‌باشد. متعادل‌سازی منابع آب و ارزیابی آن و وضعیت تقاضا در منطقه و برنامه-ریزی مناسب از سیاست‌های مدیریت آب است که خواسته‌های فعلی و آینده را دربر خواهد داشت. افزایش روز افزون جمعیت جهان و کاهش منابع آبی نیازمند تلاش در بحث مدیریت منابع آب و سعی در برقراری توازن و تعادل میان بخش‌های مختلف آن می‌باشد که سرآغاز مهمی در ایجاد توسعه پایدار در همه زمینه‌ها بخصوص منابع آبی خواهد بود. اگر چه بررسی دقیق و جز به جز این فرآیند بیشتر از آنکه جنبه مهندسی داشته باشد دارای ساختار و جنبه علمی است اما بسیاری از پروژه‌ها و طرح‌های مهندسی نیازمند داشتن یک تخمین از کمیت و ویژگی‌های رواناب حاصله از بارش هستند. در این زمینه دیدگاه‌ها و روش‌های مختلفی را می‌توان یافت که در پی یافتن راه حلی دقیق برای برآورد رواناب حاصل از بارش در یک ناحیه مشخص می‌باشند که این مهم در واقع، همان هدف نهایی از مدل‌سازی بارش- رواناب در یک حوضه آبریز است. مشکل اصلی در این زمینه، دخالت و تاثیر پارامترهای مختلفی همانند تبخیر، نفوذ، دما و دیگر عواملی است که در علم هیدرولوژی جایگاه ویژه‌ای را به خود اختصاص داده‌اند (دوانلو تاجبخش و همکاران، ۱۳۹۸). از طرفی باید اشاره نمود که فرآیند مدل‌سازی بارش- رواناب یک پدیده غیرخطی بوده و دارای تغییرات زمانی و مکانی گسترده‌ای می‌باشد و در نتیجه به سادگی قابل مدل‌سازی نیست (Hsu et al., 1995). بنابراین جهت مدل‌سازی پدیده به دلیل الگو و ماهیت غیرخطی آن، دیدگاه‌های متفاوتی برای توسعه و بهبود مدل‌های پیش‌بینی کننده آن ارائه گردیده است. به طور کلی دو نگاه مفهومی و سیستمی در مدل‌سازی پدیده‌های هیدرولوژیکی وجود دارد. مدل‌های مفهومی بر مبنای روابط و معادلات ریاضی حاکم و پارامترهای فیزیکی موجود در پدیده ارائه می‌شوند و هدف این مدل‌ها تکیه بر انجام تحقیقات علمی در نحوه

عمل اصلی هر یک از اجزای چرخه هیدرولوژی به منظور درک کامل سیستم و نحوه عملکرد اجزا با یکدیگر است (کماسی و همکاران، ۱۳۹۵). در مدل‌های سیستمی، ارائه معادلات و روابط ریاضی در آن‌ها ممکن نیست و پارامترهای فیزیکی موثر بر آن به راحتی قابل تخمین نیستند. هدف این گروه، توسعه روابط معتبر بین پارامترهای اندازه‌گیری شده چرخه هیدرولوژی است، که از آن‌ها برای حل مسائل فنی و علمی استفاده می‌گردد. مدل‌های سیستمی با دریافت ورودی و انجام یک سری عملیات ریاضی، خروجی مد نظر را تخمین می‌زنند. انواع مدل‌های سیستمی دارای پارامترها و ضرایبی هستند که با توجه به داده‌های ورودی و خروجی مشاهداتی مورد تخمین قرار می‌گیرند. بنابراین مدل‌های سیستمی به داده‌های ورودی و خروجی از نظر کمیت و کیفیت داده‌ها وابسته است (Singh, 2008). به دلیل ماهیت غیرخطی اکثر فرآیندهای هیدرولوژیکی، امروزه شبکه عصبی مصنوعی (ANN^1)، به عنوان یک تابع تخمین‌گر خود سازمان‌ده، به طور گسترده‌ای در مدل‌سازی و پیش‌بینی سری‌های زمانی غیرخطی در فرآیندهای هیدرولوژیکی مورد استفاده قرار گرفته است. به طور کلی اصلی‌ترین و بارزترین مزایای مدل‌های ANN نسبت به سایر روش‌های آماری و مفهومی را می‌توان در قالب مواردی چون عدم نیاز به دانش قبلی، اعمال یک فیلتر غیر خطی تحت عنوان تابع محرک و توانایی پردازش ورودی‌های چند متغیره با خصوصیات متفاوت، دسته‌بندی نمود (Nourani et al., 2009). علی‌رغم ماهیت انعطاف‌پذیر ANN در مدل‌سازی فرآیندهای هیدرولوژیکی، این الگوریتم نقایصی را در مواجهه با سیگنال‌های هیدرولوژیکی، که عموماً ناپایستا بوده و محدوده وسیعی از مقیاس را شامل می‌شود، از خود نشان می‌دهد. از این رو، در چنین شرایطی پیش‌پردازش زمانی و یا مکانی داده‌ها قبل از اعمال آن‌ها به سیستم شبکه عصبی برای غلبه بر این مشکل ضروری می‌باشد (Nourani et al., 2009 (a)). (b). توانایی تبدیل موجدک در تجزیه هم‌زمان مکانی و زمانی سری‌های ناپایستا به زیرسری‌های با مقیاس‌های

همکاران (۲۰۱۳)، با استفاده از شبکه GMDH جریان جوی را توسط ادغام کردن داده‌های بارندگی و عوامل محیطی با ضریب تبیین ۰/۸۷ پیش‌بینی کردند. در این روش داده‌های حاصل از پیش‌بینی شبکه GMDH با شبکه عصبی RBF و شبکه ANFIS مقایسه شد که نتایج بدست آمده حاکی از عملکرد بهتر GMDH نسبت به دو مدل دیگر بود. در پژوهش دیگری Santin (۲۰۱۵) به پیش‌بینی پارامترهای بیولوژیکی پساب ناشی از فاضلاب با استفاده از شبکه عصبی مصنوعی پرداخته است. همچنین پارامترهای پیش‌بینی شده با حداکثر مقدار مجاز در استفاده از پساب‌های فاضلاب برای آبیاری گیاهان مقایسه گردیده است.

از آنجایی که سری زمانی فرایندهای هیدرولوژیکی بسیار پیچیده بوده بنابراین استفاده از تبدیل موجک و تجزیه سری زمانی به چند زیر سری موجب درک بهتر و در دست‌تر از رفتار کوتاه و بلند مدت سری زمانی داده‌ها می‌شود. با رویداد شت به نکات یاد شده پژوهش حاضر به بررسی عملکرد مدل ترکیبی Wavelet-GMDH برای پیش‌بینی روناب حاصل در حوضه آبریز قره‌سو می‌پردازد. همچنین با استفاده از تبدیل موجک، روند موجود در داده‌های سری زمانی مربوط به ایستگاه‌های مورد مطالعه را از بین برد، سپس با استفاده از روش دسته‌بندی گروهی داده‌ها روناب حاصل را پیش‌بینی می‌نماید.

مواد و روش

منطقه مورد مطالعه و داده

حوضه آبریز قره‌سو

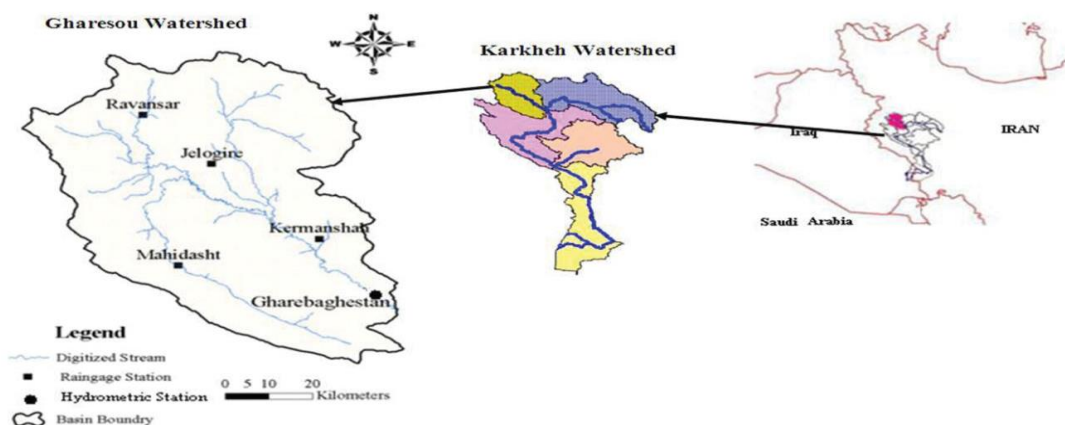
حوضه آبریز کرخه به مساحت ۵۰۷۶۴ کیلومتر مربع در غرب کشور، در مناطق میانی و جنوب غربی رشته کوه‌های زاگرس بین ۴۶ درجه، ۶ دقیقه و ۴۹ درجه ۱۰، دقیقه طول شرقی و ۳۰ درجه، ۵۸ دقیقه و ۳۴ درجه، ۵۶ دقیقه عرض شمالی قرار گرفته است. از مساحت بیان شده حدود ۲۷۶۴۵ کیلومتر مربع آن کوهستانی و حدود ۲۳۱۱۹ کیلومتر آن را دشت‌ها و

گوناگون، در تفسیر بهتر پدیده‌های هیدرولوژیکی موثر می‌باشد. تبدیل موجک، زیرسری‌های مفیدی از سری اصلی حاصل می‌نماید که این زیرسری‌ها با استخراج اطلاعات سودمند در مقیاس‌های مختلف، قابلیت مدل را در پیش‌بینی متغیر مورد نظر بالا می‌برند (کما سی و شرقی، ۱۳۹۶). تبدیل موجک با تجزیه سری‌های زمانی به کمک ضرایب اندکی به تفسیر ساختار سری و دریافت اطلاعات مفید در مورد تاریخچه آن کمک می‌کند. در مدل‌سازی بارش-روناب مدل ترکیبی موجک-شبکه عصبی (Wavelet-ANN) که به اختصار WANN نامیده می‌شود) روشی کارا بوده که از تبدیل موجک برای حصول فرکانس‌های مختلف فرآیند بارش-روناب و از ANN برای پیش‌بینی روناب آتی در مقیاس مورد نظر بهره می‌برد (Nakken, 1999). از این رو در طول سال‌های اخیر کاربرد مدل ANN به خصوص مدل هیبریدی Wavelet-ANN در مسائل مرتبط با مهندسی آب و هیدرولوژیکی به طور قابل ملاحظه‌ای افزایش یافته است (Nourani et al., 2014). با این وجود نقایصی را می‌توان به آن، از جمله عدم توجه کافی به داده‌های ورودی، پیچیده‌تر شدن محاسبات با افزایش داده‌های ورودی، احتمال وجود خطا و عدم همگرایی در اثر وجود داده‌های نامرتبط و نیز کندتر شدن روند آموزش تحت اثر افزایش ابعاد سری‌های زمانی ورودی، وارد دانست.

روش دسته‌بندی گروهی داده‌ها (GMDH^y) یک فناوری آموزش آماری شبکه‌ای و یک روش استقرایی است که با شبکه‌ها و تکنیک‌های قیاسی که معمولاً استفاده می‌شود تفاوت دارد. GMDH برای اولین بار در سال ۱۹۶۸ توسط Ivakmeko معرفی شد. استفاده از این روش برای مدل‌سازی داده‌های کم، پراکنده و یا نه چندان دقیق می‌تواند نتیجه‌ای بهتر، بهینه‌تر و با دقت بیشتری نسبت به مدل‌های فیزیکی داشته باشد (Ivakmeko, 1968). در سال‌های اخیر استفاده از شبکه‌های خود سازمانده عصبی در الگوریتم GMDH منجر به موفقیت این الگوریتم در عرصه‌های مختلف از جمله سری‌های زمانی شده است. Zhang و

تعریف شده است. داده‌های بارندگی از دو ایستگاه باران سنجی ماهیدشت و جلوگیره بدست آمده است. همچنین ایستگاه هیدرومتری قره باغستان واقع در خروجی حوضه به عنوان ایستگاه مبنا برای مشاهده تغییرات جریان حوضه انتخاب گردیده است. موقعیت حوضه آبریز و ایستگاه‌های مورد استفاده در شکل (۱) آورده شده است. همچنین جدول (۱) بیانگر مشخصات ایستگاه‌های مورد استفاده در پژوهش می‌باشد. (Zahabiyoun et al, 2013). جدول (۲) نیز بیانگر مشخصات آماری بارش و رواناب در دوره زمانی بیان شده در حوضه آبریز قره‌سو است.

کوه‌پایه‌ها تشکیل می‌دهند. مناطق کوهستانی این حوضه غالباً در بخش‌های شرقی و میانی متمرکزند و دشت‌ها عموماً در بخش‌های شمالی و جنوبی قرار دارند که ۴۵ درصد از مساحت حوضه را به خود اختصاص داده‌اند. زیرحوضه قره‌سو در شمال غربی حوضه کرخه و در غرب ایران واقع است. مساحت آن ۵۳۵۴ کیلومتر مربع و حداکثر و حداقل ارتفاع آن به ترتیب ۳۳۴۶ و ۱۱۸۰ متر می‌باشد. متوسط بارندگی سالانه در این بخش بین ۳۰۰ تا ۸۰۰ میلی‌متر متغیر است. اطلاعات هواشناسی مورد نیاز برای مطالعه عبارتند از بارش و رواناب (دوره ۱۹۷۰ تا ۲۰۰۰ میلادی) که برای سیستم



شکل (۱): منطقه مورد مطالعه، حوضه آبریز قره‌سو (Zahabiyoun et al, 2013)

جدول (۱): مشخصات ایستگاه‌های موجود در منطقه (Zahabiyoun et al, 2013)

(013)

طول جغرافیایی X	عرض جغرافیایی Y	ارتفاع (m) Height(m)	نوع ایستگاه Station type	نام ایستگاه
34 21	47 9	1318.6	سینوپتیک	کرمانشاه
34 43	46 39	1379.7	سینوپتیک	روانسر
34 35	46 51	1180	باران سنجی	جلوگیره
34 16	46 49	1415	باران سنجی	ماهیدشت
34 14	47 15	1230	هیدرومتری	قره باغستان

روش دسته‌بندی گروهی داده‌ها (GMDH)

روش دسته‌بندی گروهی داده‌ها یکی از رویکردهای استقرایی مبتنی بر تئوری پرسپترون بوده که برای شناسایی سیستم‌ها، مدل‌سازی و پیش‌بینی سیستم‌های پیچیده توسعه داده شده است. GMDH ترکیبی از N-آدلاین می‌باشد و نسبت به ساختار نوع پرسپترون دارای دقت بالاتری است چراکه از طبقه‌بندی اطلاعات بصورت مفید و غیرمفید استفاده کرده و به تعداد داده‌های مشاهداتی کمتر و در نتیجه به زمان اجرای

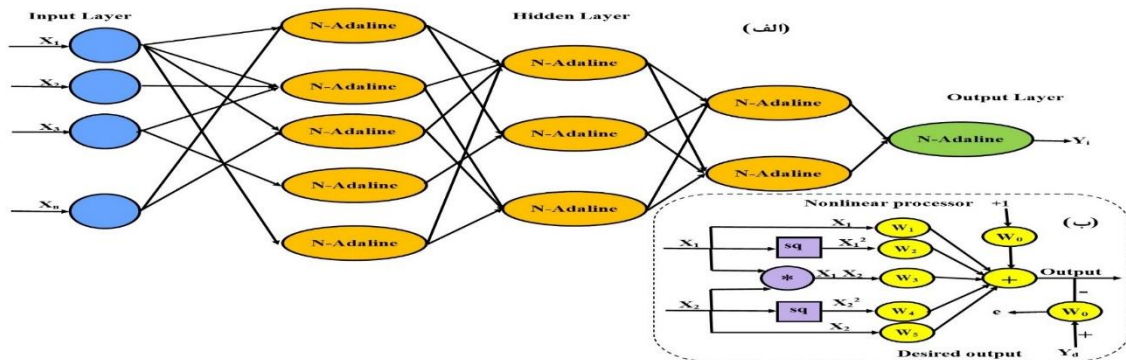
محاسبات کمتری نیازمند است. در تصویر (۲) دیاگرام شماتیک شبکه GMDH همراه با نمایی از ساختار N-آدلاین با تابع محرک چند جمله‌ای درجه دو نشان داده شده است. در شکل (۲)، Sq و x به ترتیب بیانگر مجذور و علامت ضرب، X_1 ، X_2 ، X_3 و X_4 بیانگر ورودی‌های مدل و Y خروجی می

باشد (ایوانی و همکاران، ۱۳۹۵). معیار خارجی برای تعیین ساختار شبکه بصورت زیر تعریف می‌شود:

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

جدول (۲): مشخصات آماری داده‌های روزانه و ماهانه بارش-رواناب

حوضه آبریز	سری زمانی	پارامترهای آماری
قره سو	بارش	کمینه
		پیشینه
		میانگین
رواناب	بارش	انحراف معیار
		کمینه
		پیشینه
		میانگین
		انحراف معیار



شکل (۲): الف- شماتیکی از ساختار GMDH و ب- ساختار هر N-Adaline

خروجی‌های واقعی داده‌های کنترل محاسبه می‌شوند و تنها نرون‌هایی که دارای شاخص عملکرد بالاتر از مقدار آستانه هستند برای ادامه مسیر انتخاب می‌شوند. ساختار شبکه GMDH از طریق الگوریتم زیر تعیین می‌شود (Nikolaev & Hitoshi, 2003):

۱. انتخاب متغیرهای ورودی به مدل
۲. جمع‌آوری مجموعه‌ای از داده‌های مشاهده‌ای مرتبط با موضوع
۳. ساخت N-آدلاین برای تمام ترکیبات ۲ تایی از متغیرهای ورودی
۴. انتخاب نوع تابع حرکت برای هر نرون
۵. تخمین ضرایب وزنی برای تمام N-آدلاین

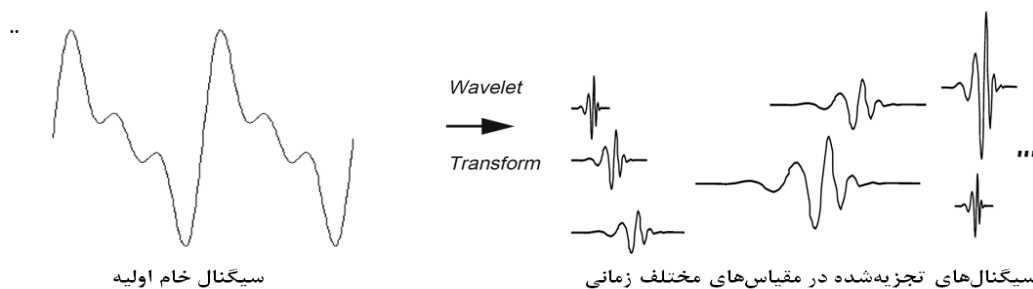
۶. مقایسه معیار محاسبه شده برای تمام N-آدلاین‌ها با مقدار آستانه تعیین شده. اگر مقدار آستانه بزرگتر باشد، N-آدلاین حذف می‌شود و دیگر N-آدلاین‌ها برای لایه بعد باقی می‌مانند. می‌توان برای تمام لایه‌ها یک مقدار آستانه ثابت یا برای هر لایه یک مقدار آستانه در نظر گرفت

در رابطه (۱)، $y(t)$ خروجی مطلوب، $\hat{y}(t)$ خروجی محاسباتی، میانگین خروجی‌های محاسباتی و \bar{y} میانگین خروجی‌های مطلوب است. شبکه عصبی GMDH دارای یک شبکه چند لایه و روبه جلو است و دربرگیرنده مجموعه‌ای از نرون‌ها است که از پیوند جفت‌های مختلف ورودی از طریق یک چند جمله‌ای درجه دوم به وجود می‌آیند. هر لایه در این شبکه از یک یا چند واجد پردازشگر تشکیل شده است که هر یک از آن‌ها دارای دو ورودی و یک خروجی می‌باشد. این واحدها عملاً نقش اجزای سازنده مدل را ایفا می‌نمایند و به شکل یک چند جمله‌ای درجه دوم فرض می‌شوند (رابطه ۱). پارامترهای مجهول الگوریتم GMDH ضرایب چند جمله‌ای رابطه (۲) می‌باشد. جهت محاسبه مقدار خروجی \hat{y}_n برای هر ورودی X براساس رابطه (۲)، باید مقدار میانگین مربعات خطا به حداقل برسد (ایوانی و همکاران، ۱۳۹۵).

(۲)

$$\hat{y}_n = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2$$

بعد از محاسبه ضرایب براساس داده‌های آموزشی، شاخص عملکرد نرون‌های بدست آمده از طریق همبستگی یا محاسبه میانگین خطای مربعات بین



شکل (۳): طرح شماتیک تبدیل موجک (کماسی و شرقی، ۱۳۹۶)

تبدیل موجک

این شرط با عنوان شرط پذیرفتگی موجک $\psi(x)$ شناخته می‌شود. رابطه فوق را می‌توان معادل با فرمول زیر دانست که باید ارضا شود (کماسی و شرقی، ۱۳۹۶):

$$\int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (2)$$

این ویژگی تابع با میانگین صفر، چندان محدود کننده نبوده و توابع بسیاری را می‌توان بر اساس آن تابع موجک نامید. $\psi(x)$ تابع موجک مادر است که توابع مورد استفاده در تحلیل با دو عمل ریاضی انتقال و مقیاس کردن، در طول سیگنال مورد تحلیل، تغییر اندازه و محل می‌یابند و در نهایت ضرایب موجک در هر نقطه از سیگنال (b) و برای هر مقدار از مقیاس (a) با رابطه زیر قابل محاسبه است.

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (3)$$

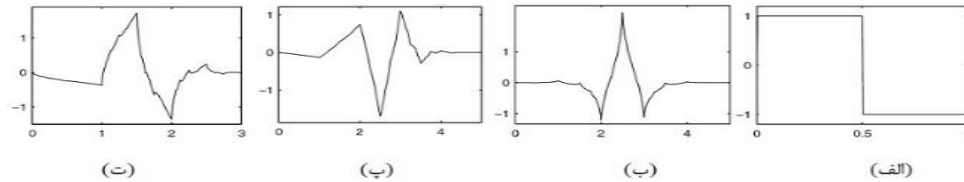
عمل مقیاس نمودن، به عنوان یک عمل گر ریاضی، سیگنال را بسط داده و یا فشرده می‌سازد. برای تابع مفروض $f(t)$ ، اگر $s < 1$ باشد، $f(st)$ حالت بسط داده شده $f(t)$ و اگر $s > 1$ باشد حالت فشرده شده تابع $f(t)$ می‌باشد. همان طور که در رابطه ۳ می‌بینیم، در تعریف تبدیل موجک، ترم مقیاس a در مخرج آمده است و لذا عکس آن چه گفته شد، اگر $a < 1$ باشد، سیگنال فشرده شده و اگر $a > 1$ باشد سیگنال بسط داده می‌شود. همچنین در رابطه فوق پارامتر b ، انتقال یا به زبان ساده‌تر تاخر و تقدم را روی یک تابع مدل می‌کند. در

تبدیل موجک یکی از تبدیلات ریاضی کارآمد در زمینه پردازش سیگنال می‌باشد. تبدیلات ریاضی برای به دست آوردن اطلاعات اضافی از سیگنال - که از خود سیگنال قابل دستیابی نیستند - استفاده می‌شوند. آنالیز موجک همانند آنالیز فوریه که یکی از مشهورترین تبدیلات ریاضی است، با بسط تابع‌ها سروکار داشته ولی این بسط بر حسب موجک‌ها صورت می‌گیرد. موجک، تابع مشخص مفروضی با میانگین صفر است و بر خلاف چند جمله‌ای‌های مثلثاتی، در فضا به صورت موضعی بررسی می‌گردند و به این ترتیب ارتباط نزدیک‌تری بین معوضی توابع و ضرایب آن‌ها امکان‌پذیر و پایداری عددی بیشتری در بازسازی و محاسبات فراهم می‌شود. هر کاربردی را که مبتنی بر تبدیل سریع فوریه است می‌توان با استفاده از موجک‌ها فرمول‌بندی کرد و اطلاعات فضایی (یا زمانی) موضعی بیشتری به دست آورد. تابع تبدیل موجک قابلیت تجزیه سری زمانی به چندین زیر سری زمانی با مقیاس‌های مختلف را داراست و با مطالعه زیر سری‌های زمانی حاصل از سری زمانی کلی، رفتار کوچک مقیاس و بزرگ مقیاس یک فرایند هیدرولوژیکی را مورد آنالیز قرار می‌دهد (شکل ۳). تابع موجک، تابعی است که دو ویژگی مهم را دارا باشد: نوسانی بودن و کوتاه مدت بودن. به عبارتی $\psi(x)$ تابع موجک است اگر و فقط اگر تبدیل فوریه آن $\psi(\omega)$ شرط زیر را ارضا کند (کماسی و شرقی، ۱۳۹۶):

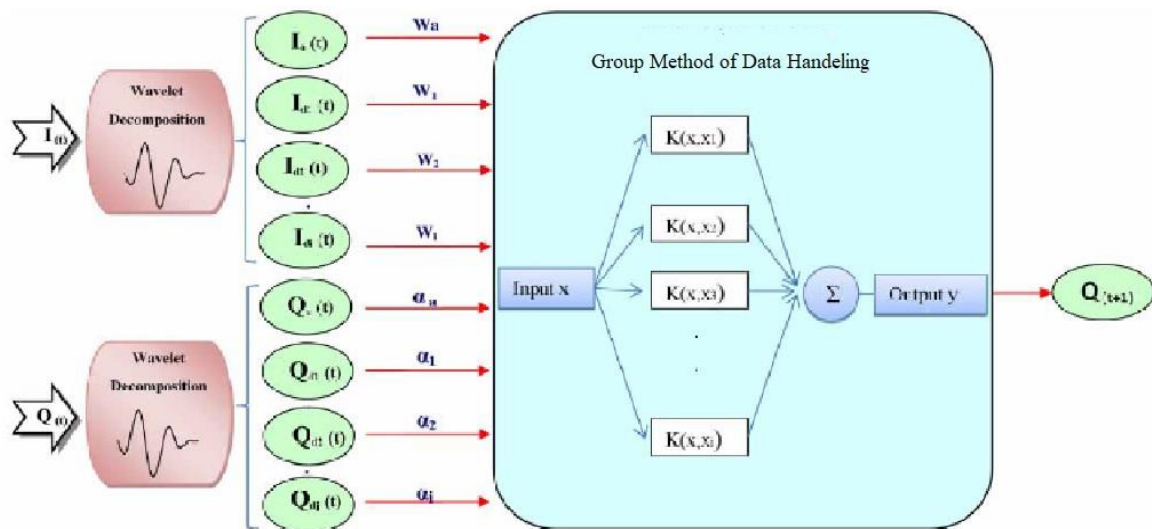
$$\int_{-\infty}^{+\infty} \frac{|\psi(\omega)|}{|\omega|^2} d\omega < +\infty \quad (1)$$

نهایت می‌توان حالت پیوسته تبدیل موجک (که به اختصار CWT^3 نامیده می‌شود) را به شکل زیر نوشت:

$$CWT(a, b) = Wf(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx = \int_{-\infty}^{+\infty} f(x) \psi_{a,b}(x) dx \quad (4)$$



شکل (۴): نمودار الف) تابع موجک Haar ب) تابع موجک Coif1 پ) تابع موجک Sym3 ت) تابع موجک db2 (کماسی و شرقی، ۱۳۹۶)



شکل (۵): ساختار مدل ترکیبی WGMDH

توابع موجک دارای انواع بسیاری هستند که تابع مهم‌ترین و پرکاربردترین آن‌ها در شکل ۴ نشان داده شده است:

مدل ترکیبی موجک-روش دسته‌بندی گروهی داده‌ها Wavelet-GMDH

گوناگون مکانی و زمانی داده‌ها، تغییراتی در فرآیندهای هیدرولوژیکی وجود داشته باشد که انجام یک پیش-پردازش مناسب می‌تواند باعث بهبود کارایی روش‌های

مدل ترکیبی پیشنهادی Wavelet-GMDH متشکل از دو مرحله می‌باشد. گام نخست، پیش‌پردازش داده‌ها می‌باشد چرا که ممکن است به دلیل توزیع‌های

به ترتیب مرتبه تجزیه سری بارش ($I(t)$) و سری رواناب ($Q(t)$) می‌باشند (Nourani et al., 2014).

آماره‌های ارزیابی

جهت ارزیابی نتایج از معیارهای ارزیابی ضریب تبیین (DC^5) و جذر میانگین مربعات خطا ($RMSE^6$) استفاده شده است، که روابط ۶ و ۷ بیانگر این مقادیر است. در این روابط N تعداد داده‌های مشاهداتی، S_{obs} داده‌های مشاهده‌ای، S_{pre} مقادیر محاسبه‌شده و \bar{S} میانگین داده‌های مشاهداتی هستند.

$$DC = 1 - \frac{\sum_{i=1}^N (S_{pre} - S_{obs})^2}{\sum_{i=1}^N (S_{obs} - \bar{S})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (S_{pre} - S_{obs})^2}{N}} \quad (7)$$

RMSE اختلاف مقادیر مشاهده شده و مقادیر محاسبه شده را نشان می‌دهد. کمترین مقدار RMSE بالاترین صحت پیش‌بینی را نشان می‌دهد. هرچه RMSE به صفر و DC به یک نزدیکتر شود، جواب بهتری برای مدل حاصل شده است (پورحقی و همکاران، ۱۳۹۷).

نتایج و بحث

برای ارزیابی مدل، از داده‌های حوضه قره‌سو در مدت ۲۰ سال و به صورت روزانه استفاده گردید. داده‌ها به دو قسمت دوازده ساله و هشت ساله تقسیم شدند. دوازده سال اول به عنوان دوره واسنجی مدل استفاده شده و هشت سال بعدی به عنوان دوره ارزیابی بکار گرفته شد. در این مدل بارش، دما و رواناب به عنوان عوامل ورودی به سیستم معرفی شده و پس از نرمال‌سازی برای تولید رابطه استفاده گردید. همان گونه که بیان شد عوامل گوناگونی نظیر بارش، تبخیر، پوشش گیاهی و ... در پدیده رواناب دخیل می‌باشند و انتخاب عوامل تاثیرگذار به عنوان متغیرهای ورودی در پیش‌بینی آن امری حائز اهمیت است. مطالعات پیشین نشان داده است که رواناب به عنوان پدیده‌ای مارکوف شناخته

داده‌محور گردد. تحلیل موجک یکی از روش‌های پیشنهادی جهت انجام پیش‌پردازش بر روی داده‌ها می‌باشد. در مرحله دوم با استفاده از روش GMDH مدل‌سازی فرآیند بارش-رواناب صورت گرفته و رواناب روزانه پیش‌بینی می‌گردد (شکل ۵). در روش پیشنهادی ابتدا سری‌های بارش ($I(t)$) و رواناب ($Q(t)$) به زیرسری‌هایی با درجه تجزیه مختلف تجزیه می‌شوند. اگر i و j به ترتیب درجه تجزیه سری‌های بارش و رواناب در نظر گرفته شوند، تعداد داده‌هایی ورودی برابر با $i+j+2$ خواهد بود چرا که مدل از دو سری زمانی استفاده می‌نماید. بنابراین هر سری زمانی به $i+1$ و $j+1$ زیرسری تجزیه می‌شود. در مرحله بعد داده‌های تجزیه شده برای مدل‌سازی به عنوان ورودی مدل GMDH به کار گرفته می‌شوند. مدل‌سازی صورت گرفته در این پژوهش در نرم‌افزار MATLAB انجام گردیده است.

مدل ترکیبی موجک-شبکه عصبی WANN

مدل موجک-شبکه عصبی از یک ساختار سه لایه تشکیل یافته است. لایه اول شامل نرون‌های موجکی می‌باشد که ورودی آن‌ها زیرسری‌هایی است که با استفاده از تبدیل موجک بر روی سری‌های زمانی بارش و رواناب حاصل می‌گردد. در مدل موجک-شبکه عصبی برای انجام محاسبات از شبکه عصبی جهت یافتن ضرایب وزن بارش در ساختار شبکه استفاده می‌گردد. در این مدل نخست سری‌های زمانی بارش و رواناب به زیرسری‌هایی با مقیاس‌های متفاوت تجزیه می‌شوند. به عنوان نمونه می‌توان سری زمانی را به یک مقیاس بلند مدت و چند مقیاس کوتاه مدت (برای رهگیری خواص گذرا و نوسانات کوتاه مدت سری) تجزیه نمود. در این صورت زیرسری‌های زمانی که تحت عنوان $a(t)$ (همانند $I_a(t)$ و $Q_a(t)$) نام‌گذاری می‌شوند، معرف زیرسری تقریبی (مقیاس بلند مدت)، و برای زیرسری‌های زمانی که تحت عنوان $d_i(t)$ و $d_j(t)$ با اندیس‌های i و j (همانند $I_{di}(t)$ و $Q_{dj}(t)$) نام‌گذاری می‌شوند، معرف زیرسری جزئی (مقیاس کوتاه مدت) هستند. پارامترهای i و j نیز

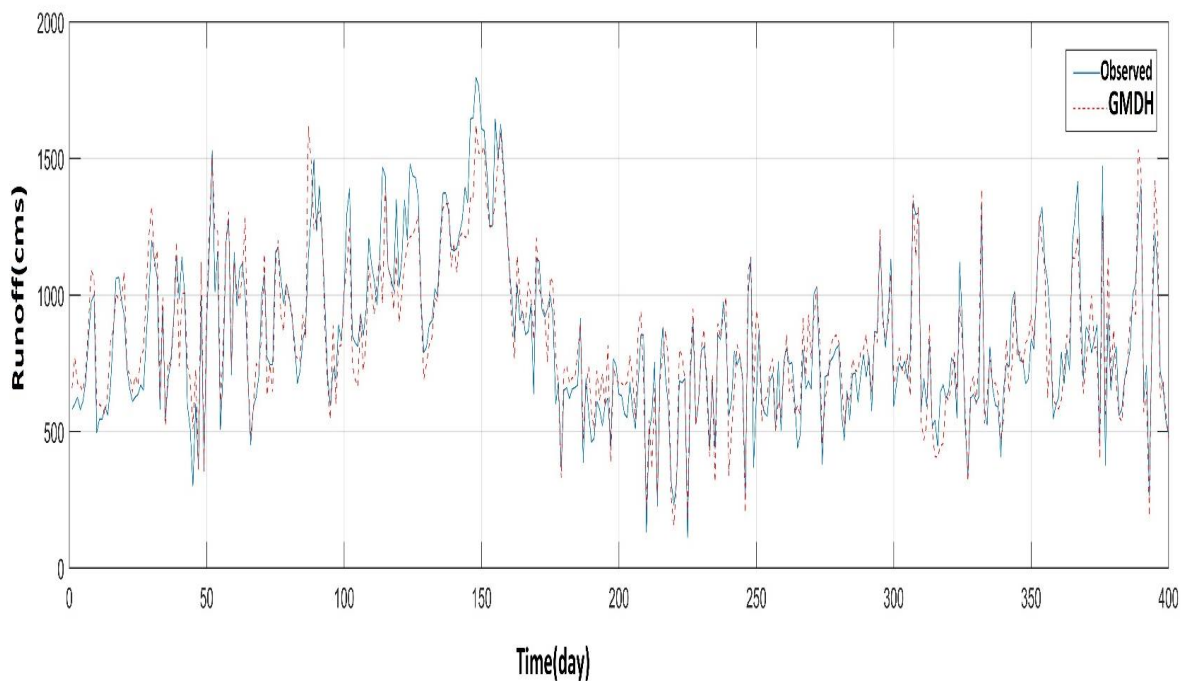
مدل WGMDH در پیش‌بینی رواناب بر اساس بارش می‌پردازد. همان گونه که مشخص است ماهیت فرآیند بارش-رواناب کاملاً غیرخطی و ناپایستا بوده تا جایی که کارایی مدل‌های خطی پیشین در امر شبیه‌سازی مورد قبول واقع نگردید. در ابتدا کارایی مدل GMDH و پس از آن کارایی مدل ترکیبی WGMDH مورد ارزیابی قرار گرفته است. عملکرد شبکه عصبی GMDH به صورت لایه ای بوده و دربرگیرنده تمام ترکیبات دوتایی از متغیرهای ورودی است. تعداد نرون‌ها در لایه اول برابر با ۶ می‌باشد. در صورتیکه تمام نرون‌های لایه اول انتخاب گردند تعداد نرون‌های لایه دوم برابر ۱۵ خواهد بود. به همین ترتیب پیش‌رفته و در هر لایه مقدار آستانه را افزایش می‌دهد تا مقدار خطا در هر لایه نسبت به لایه قبل کمتر شود و در نهایت یک نرون و مقدار خروجی نهایی محاسبه شود. در صورتیکه مقدار خطای مطلوب تامین نشود می‌توان نسبت داده‌های واسنجی و صحت‌سنجی و یا تعداد متغیرهای ورودی را افزایش یا کاهش داد که برای هر دو مدل این روش صادق است. جدول (۳)، نتایج حاصل از عملکرد مدل‌ها را در دو مقیاس زمانی روزانه و ماهانه نشان می‌دهد. همانگونه که از این جدول برمی‌آید مقدار DC و RMSE برای مدل WGMDH نسبت به سایر مدل‌ها بهتر می‌باشد. در واقع بهره‌گیری از آنالیز موجک در تجزیه سری‌های زمانی بلند و کوتاه و همچنین استفاده از مدل GMDH برای پیش‌بینی داده بصورت همزمان توانست نتایج مناسبی را ارائه دهد.

جدول (۳): نتایج حاصل از مدل‌های محاسباتی در مقیاس‌های روزانه در منطقه مورد مطالعه

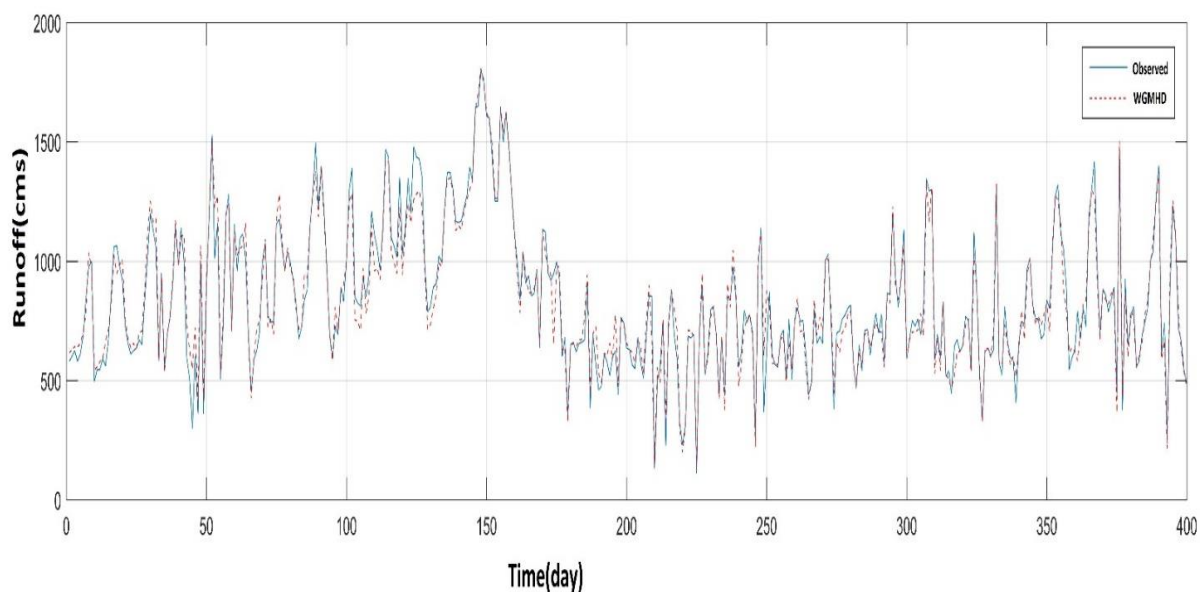
DC		RMSE		مدل	خروجی مدل	ورودی مدل	حوضه آبریز
Tr.*	Ver.*	Tr.	Ver.				
۰/۶۹	۰/۶۵	۰/۰۳	۰/۰۷	GMDH	Q_{t+1}	I_t, Q_t	قره سو
۰/۹۵	۰/۹۱	۰/۰۳	۰/۰۵	WGMDH			

شکل (۶)، عملکرد مدل‌های GMDH و WGMDH را در پیش‌بینی رواناب روزانه نشان می‌دهد.

می‌شود. از این رو مقدار رواناب در زمان حال تابعی از مقادیر بارش و رواناب در زمان‌های پیشین خواهد بود. شبکه عصبی استفاده شده در مدل‌های ANN/WANN به منظور مدل‌سازی بارش-رواناب حوضه، پرسپترون سه لایه است که این نوع از مدل‌های شبکه عصبی با الگوریتم انتشار برگشتی، به طور گسترده در مدل‌سازی‌های هیدرولوژیکی استفاده شده است. شکل بهینه شبکه‌های عصبی مصنوعی به صورت آزمون و خطا به دست می‌آید. هدف از این آزمون و خطا، یافتن تعداد نرون‌های لایه پنهان و تعداد تکرار الگوریتم می‌باشد تا به واسطه آن‌ها مدل تعریف گردد. لازم به ذکر است که تعداد تکرارهای آموزشی کم می‌تواند منجر به آموزش ناقص و تعداد تکرارهای زیاد منجر به حفظ یا از برگردن توسط شبکه در مرحله آموزش گردد. لذا بایستی مقدار بهینه‌ای برای تعداد تکرارها در نظر گرفت تا کیفیت مدل برای هر دو مرحله آموزش و آزمایش قابل قبول باشد که با توجه به مطالعات پیشین این تعداد در بین ۱۵۰ تا ۲۰۰ در نظر گرفته می‌شود. به جهت قدرت همگرایی زیاد، برای آموزش شبکه‌های عصبی از الگوریتم لوببرگ-مارکوارت استفاده شد. فرایند آموزش شبکه زمانی متوقف می‌شود که میزان خطا در بین داده‌های آزمایش، شروع به افزایش می‌کند. در این پژوهش از تابع محرک تانژانت سیگموئید، به عنوان هسته غیرخطی شبکه‌های عصبی بهره گرفته شده است با توجه به نقایص برشمرده شده در خصوص مدل WANN، پژوهش کنونی به بررسی



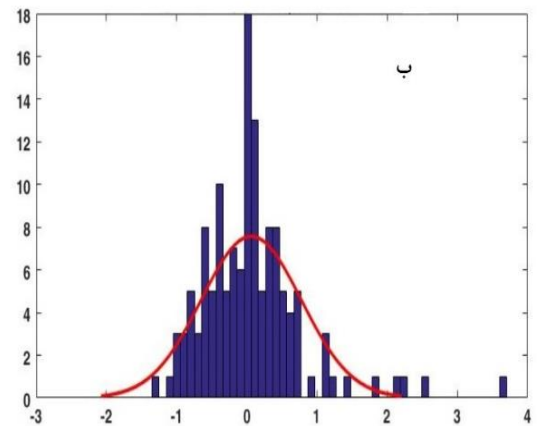
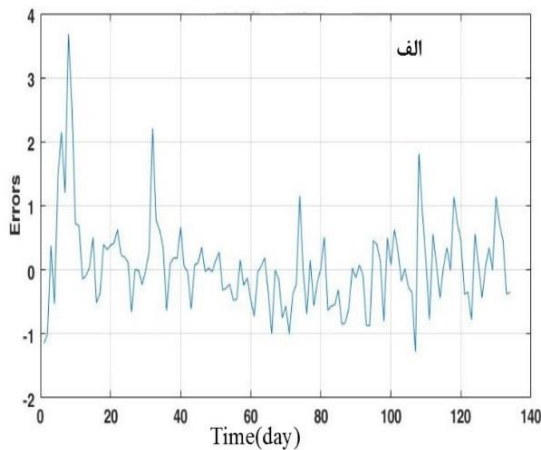
شکل (۶): مقایسه عملکرد مدل GMDH با داده‌های مشاهداتی



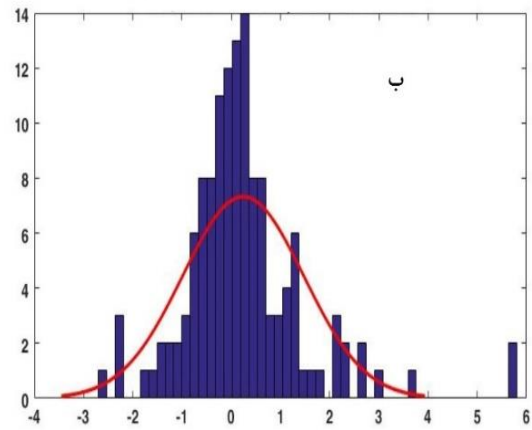
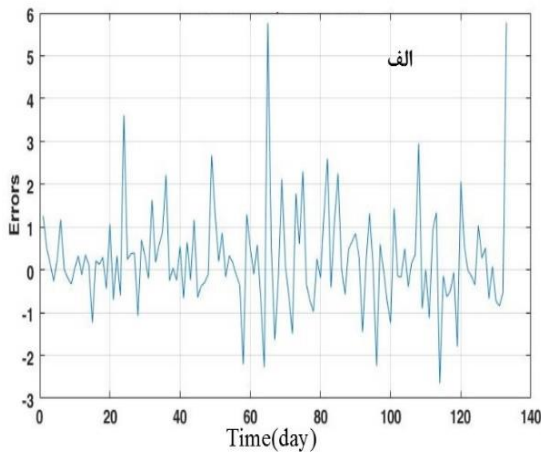
شکل (۷): مقایسه عملکرد مدل WGMDH با داده‌های مشاهداتی

مقیاس ماهانه نیز بیشتر نمود پیدا کرده و عملکرد بهتر مدل را سبب شده است (شکل (۷)).

همانگونه که از این دو شکل برمی‌آید، بهره‌گیری از تبدیل موجک توانسته در پیش‌بینی مقادیر بیشینه و کمینه رواناب موثر باشد. همچنین تبدیل موجک در



شکل(۸): الف-مقدار خطای اندازه‌گیری شده مدل WGMDH در پیش‌بینی رواناب ، ب-نمودار توزیع نرمال خطای داده‌های پیش‌بینی شده در مدل WGMDH در دوره صحت‌سنجی



شکل(۹): الف-مقدار خطای اندازه‌گیری شده مدل GMDH در پیش‌بینی رواناب ، ب-نمودار توزیع نرمال خطای داده‌های پیش‌بینی شده در مدل GMDH در دوره صحت‌سنجی

همچنین شکل‌های (۸) و (۹)، مقدار خطای اندازه‌گیری شده و توزیع نرمال خطا را برای دو مدل WGMDH و GMDH را برای یک بازه زمانی ۱۴۰ روزه نشان می‌دهد.

اینگونه مسائل ارایه دهد. امروزه استفاده از تکنیک‌های هوش مصنوعی جایگاه ویژه‌ای در میان پژوهشگران یافته و نتایج مناسبی را ارایه داده است. یکی از این تکنیک‌ها استفاده از تبدیل موجک می‌باشد. استفاده از تابع تبدیل موجک رهیافت نوینی در مطالعه و مدلسازی منابع آب است که با توجه به ساختار و عملکرد آن نسبت به سایر روش‌های هوش مصنوعی از دقت

نتیجه‌گیری

افزایش روزافزون جمعیت و کاهش منابع آبی در دنیا سبب بروز چالش آبی گشته است. ایران به عنوان یک کشور در حال توسعه نیز با این مسئله روبرو بوده و مدیریت مناسب در این بخش نیازمند برنامه‌ریزی جامع و کاربردی می‌باشد. در سال‌های اخیر پژوهش‌های بین‌رشته‌ای توانسته است رهیافت‌های مفیدی را برای حل



خطا ۰/۰۷ بدست آمد. در این مرحله برای مدل WGMDH هم مقادیر ضریب DC و مقدار جذر میانگین خطا به ترتیب برابر ۰/۹۱ و ۰/۰۵ می باشد که خود تاییدی بر صحت این روش است.

مطلوب تری برخوردار است از آنجا که مدیریت منابع آب و تصمیم گیری در این زمینه بسته به عوامل متعدد است. همانگونه که بحث شد، برای منطقه مورد مطالعه در مرحله صحت سنجی داده ها برای مدل GMDH مقدار ۰/۶۵ برای ضریب DC و مقدار جذر میانگین

پی نوشت ها

1. Artificial Neural Network
2. Waikato Environment for Knowledge Analysis
3. Continuous Wavelet Transform
4. Standard Deviation Reduction
5. Determination Coefficient
6. Root Mean Square Error
7. Grope Methode of Data Handling

منابع

- ایوانی، ز.، احمدی، م.، قادری، ک.، ۱۳۹۵، برآورد بار معلق رودخانه ای با استفاده از روش گروهی کنترل داده ها، پژوهشنامه مدیریت حوزه آبخیز، ۲(۱۳): ۶۱-۷۱.
- پورحقی، ا.، سلگی، ا.، رائمنش، ف.، شهنی دارابی، م.، ۱۳۹۷، استفاده ترکیبی از تبدیل موجک و مدل های هوشمند در شبیه سازی جریان رودخانه (مطالعه موردی: رودخانه های کاکارضا و سراب صیدعلی)، نشریه علمی پژوهشی مهندسی آبیاری و آب ایران، شماره ۳۲، ۱۷-۱.
- دوانلو تاجبخش، ع.، نورانی، و.، مولاجو، ا.، ۱۳۹۸، بررسی کارایی مدل هیبریدی Wavelet-M5 در پیش بینی فرآیند بارش-رواناب (مطالعه موردی: حوضه آبی چای)، نشریه تحقیقات منابع آب، شماره ۱۵، ۱۰-۱.
- کماسی، م.، شرقی، س.، ۱۳۹۶، رندیابی عوامل موثر بر کاهش تراز آب زیرزمینی با بهره گیری از تبدیلات موجک متقابل و ارتباطی، نشریه علمی پژوهشی مهندسی آبیاری و آب ایران، شماره ۲۸، ۱۵۱-۱۳۸.
- کماسی، م.، نودری، ح.، قشلاقی، ن.، ۱۳۹۵، پیش بینی تراز آب دریاچه ارومیه با استفاده از روش های سری زمانی، شبکه عصبی مصنوعی و شبکه عصبی-موجکی، نشریه علمی پژوهشی مهندسی آبیاری و آب ایران، شماره ۲۴، ۷۷-۶۴.
- Hsu, K.L., Gupta, H.V. and Sorooshian, S., 1995. Artificialneural network modeling of the rainfall-runoff process. *Water resources research*, 31(10), pp.2517-2530.

Ivakhnenko, A.G., 1968. The Group Method of Data of Handling; A rival of the method of stochastic approximation. *Soviet Automatic Control*, 13, pp.43-55.

Nakken, M., 1999. Wavelet analysis of rainfall-runoff variability isolating climatic from anthropogenic patterns. *Environmental Modelling & Software*, 14(4), pp.283-295.



- Nikolaev, N.Y. and Iba, H., 2003. Polynomial harmonic GMDH learning networks for time series modeling. *Neural Networks*, 16(10), pp.1527-1540.
- a) Nourani, V., Alami, M.T. and Aminfar, M.H., 2009. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence*, 22(3), pp.466-472.
- b) Nourani, V., Komasi, M. and Mano, A., 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water resources management*, 23(14), p.2877.
- Nourani, V., Baghanam, A.H., Adamowski, J. and Kisi, O., 2014. Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *Journal of Hydrology*, 514, pp.358-377.
- Santin, I., 2015. Effluent Predictions in Wastewater Treatment Plants for the Control Strategies Selection, *Journal of Bilbao*, 2: 1009-1016.
- Singh, V.P., 1989. *Hydrologic systems: watershed modeling*. Prentice Hall, University of California, 320p
- Zhang, H., Liu, X., Cai, E., Huang, G. and Ding, C., 2013. Integration of dynamic rainfall data with environmental factors to forecast debris flow using an improved GMDH model. *Computers & geosciences*, 56, pp.23-31.
- Zahabiyoun, B., Goodarzi, M.R., Bavani, A.M. and Azamathulla, H.M., 2013. Assessment of climate change impact on the Gharesou River Basin using SWAT hydrological model. *CLEAN-Soil, Air, Water*, 41(6), pp.601-609.



Investigation of the Efficiency of Group Method of Data Handling and Wavelet Transform in Runoff Forecasting (Case Study: Gharehsoo Watershed)

Mohammadreza Goodarzi¹, Hesam Goodarzi²

Abstract

Rainfall-runoff process is one of the most important and complex phenomena in the hydrological cycle. Therefore, in its modeling, different perspectives for the development and improvement of predictive models have been presented. In this study, while introducing a combination of wavelet-group classification of data, its effectiveness for modeling the rainfall-run-off process in the Ghara-eos watershed was studied. At first, the rainfall and runoff time series were decomposed using a wavelet transform to several sub-basins to overcome its non-state. Then, these time subcircuits are considered as inputs of the grouped data collection method for predicting daily runoff. The efficiency of the combined model with DC and root mean square error (RMSE) were evaluated. The results of the validation of the models indicate that the highest amount of explanation coefficient and the lowest root mean of error for the single GMDH model were 0.65 and 0.07, respectively, and for the combined model of grouping the wavelet data The order is 0.91 and 0.05. The reason for the hybrid model's superiority to the single model is that the combination model of the grouping of wavelet data, instead of using the time series of rainfall and runoff data on a general scale, uses several time- Different decompositions are used as inputs in the model. Also, the results showed that the Wavelet-GMDH combination model compared to other composite models such as Waveline Artificial Neural Network (WANN) due to the GMDH model layer function, which includes binary combinations of input variables, and by selecting the number of optimal neurons in Each layer directs motion to the predicted data, has more efficiency and accuracy.

Keywords: Group method of data handling, wavelet transform, wavelet function, rainfall-runoff modeling, Gharehsoo watershed.

¹. Assistant Professor, Department of Civil Engineering, Yazd University, Yazd, Iran. Goodarzimr@yazd.ac.ir

². M.Sc. Student of Civil Engineering - Water and Hydraulic Structures, Grand Ayatollah Borujerdi University, Borujerd, Iran.