

ارزیابی مدل شبکه عصبی توسعه یافته با فیلتر کالمن و برنامه ریزی بیان ژن در شبیه سازی بارش-رواناب، مطالعه موردی: دشت ملایر، استان همدان

سیده اسرین حسینی^۱، محمدرضا گلابی^۲، صفر معروفی^{۳*}، نسیم خالدیان^۴ و محمد سلطانی^۵
^{۱،۴} کارشناس ارشد مهندسی منابع آب، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان، ^۲ دانشجوی دکتری منابع آب، دانشکده مهندسی علوم آب، دانشگاه شهید چمران اهواز، ^۳ استاد گروه مهندسی آب، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان و ^۵ دانشجوی دکتری آبیاری و زهکشی، دانشکده کشاورزی، پردیس ایوریحان، دانشگاه تهران

تاریخ پذیرش: ۱۳۹۸/۰۶/۱۶

تاریخ دریافت: ۱۳۹۷/۱۲/۱۵

چکیده

شبیه سازی فرایند بارش-رواناب یکی از گام های مهم در مطالعات مهندسی آب و مدیریت منابع آب است. برنامه ریزی بهره برداری از منابع آب های سطحی و زیرزمینی، ساماندهی رودخانه ها و هشدار سیل نیاز به پیش بینی آبدی رودخانه و رواناب حوزه آبخیز دارد. در مطالعه حاضر، به منظور شبیه سازی فرایند بارش-رواناب از مدل شبکه عصبی توسعه یافته با فیلتر کالمن (EKFNN) استفاده شد و سپس نتایج با الگوریتم برنامه ریزی بیان ژن (GEP) که در بیشتر مطالعات اخیر از خود، کارایی بالایی در شبیه سازی بارش-رواناب نشان داده بود، مقایسه شد. داده های مورد استفاده در این مطالعه، مقادیر روزانه بارش و رواناب ایستگاه های پیهان، مرویل و نامیله دشت ملایر در استان همدان است که برای دوره آماری ۱۳۸۰ تا ۱۳۹۲ مورد استفاده قرار گرفت. نتایج نشان داد که در دشت ملایر دقت (R^2) مدل GEP به طور متوسط برابر ۰/۹۴ و کمترین ریشه میانگین مربعات خطا (RMSE) برابر ۰/۴۲ و دقت مدل EKFNN برابر ۰/۹۶ و کمترین ریشه میانگین مربعات خطا برابر (RMSE) ۰/۲۴ بود. همچنین نتایج نشان داد، با توجه به این که سرعت اجرای مدل GEP بیشتر و در زمان کوتاهی قادر به ارائه نتایج بود، اما مدل EKFNN که برای اولین بار برای شبیه سازی جریان روزانه رودخانه مورد استفاده قرار گرفته است، به عنوان مدل برتر برای دشت ملایر انتخاب شد.

واژه های کلیدی: پیش بینی آبدی رودخانه، مدیریت منابع آب، هشدار سیل، مدل EKFNN، مدل GEP

مقدمه

طرح های آبی است. خاصیت غیرخطی، نبود قطعیت ذاتی این فرایند، نیاز به اطلاعات وسیع و پیچیده بودن مدل های فیزیکی از دلایلی است که باعث شده محققان به سوی مدل های هوشمند نظیر شبکه های عصبی مصنوعی^۱ (ANNs)، برنامه ریزی بیان ژن^۲ (GEP) و مدل های جدیدی همچون شبکه های تصمیم بیزی^۳

شبیه سازی فرایند بارش-رواناب به عنوان مهمترین گام در مطالعات مهندسی آب و مدیریت منابع آب است. برنامه ریزی بهره برداری از منابع آب های سطحی و زیرزمینی، ساماندهی رودخانه ها و هشدار سیل، نیاز به پیش بینی آبدی رودخانه و رواناب حوزه آبخیز دارد. به طوری که، آگاهی از چگونگی ارتباط بین بارندگی و رواناب حوزه های آبخیز بخش جدا نشدنی مطالعات

¹ Artificial Neural Networks

² Gene Expression Programming

³ Bayesian Network

در مطالعه‌ای دیگر، Karimi و همکاران (۲۰۱۵) پیش‌بینی کوتاه‌مدت و بلندمدت جریان رودخانه فلیوس^۲ در ترکیه را با استفاده از مدل ترکیبی برنامه‌ریزی بیان ژن-موجک مورد بررسی قرار دادند. نتایج نشان داد که مدل ترکیبی عملکرد بهتری نسبت به مدل GEP دارد. آن‌ها همچنین، از مدل‌های ARMA^۳، ANN و ANFIS برای مقایسه استفاده کردند. نتایج نشان داد که عملکرد مدل ترکیبی از بقیه مدل‌ها بهتر بوده است. همچنین، Singi و همکاران (۲۰۱۵) به مدل‌سازی رواناب روزانه حاصل از حوضه کاپگری در شرق هندوستان پرداختند. آن‌ها با استفاده از روش نمونه‌گیری مجدد تعداد لایه‌های پنهان و تعداد نرون‌های بهینه را پیشنهاد دادند و نشان دادند، این روش حتی در مرحله آموزش سری‌های زمانی کوتاه‌مدت نتایج خوبی را ارائه می‌دهد. در پژوهشی Noori و Kalin (۲۰۱۶) از مدل شبکه عصبی مصنوعی و SWAT به‌منظور پیش‌بینی روزانه جریان رودخانه در ۲۹ حوزه آبخیز نزدیک و اطراف آتلانتا در جنوب شرقی ایالات متحده استفاده کردند. با توجه به نتایج به‌دست آمده از مدل‌ها و ارزیابی مقدار خطای آن‌ها، هر دو مدل شبکه عصبی مصنوعی و SWAT عملکرد خوبی برای این هدف داشته‌اند. مقایسه کارایی سه مدل شبکه عصبی بیزین، GEP و SVM در تخمین رواناب ماهانه رودخانه کاکا-رضا به‌وسیله Ghorbani و همکارانش (۲۰۱۶) انجام شد. از میان سه مدل ذکر شده، شبکه عصبی بیزین بهترین عملکرد را داشته است. همچنین، Naeimi و Kalourazi و همکاران (۲۰۱۷) سه مدل شامل مدل درختی، نزدیک‌ترین همسایگی و رگرسیون خطی چندمتغیره را برای برآورد آبدی در حوزه‌های آبخیز بدون آمار استان گلستان بررسی کردند و نتایج نشانگر برتری مدل نزدیک‌ترین همسایه نسبت به دو مدل دیگر بود. در پژوهشی، Motamednia و همکاران (۲۰۱۷) با استفاده از داده‌های هواشناسی و هیدرومتری طی دوره زمانی ۱۳۵۰-۱۳۴۹ تا ۱۳۹۱-۱۳۹۰ رواناب در حوضه امامه با استفاده از مدل‌های شبکه عصبی پرسپترون چند لایه، تابع پایه شعاعی و

(BN) و ماشین بردار پشتیبان^۱ (SVM) روی آورند. زیرا با استفاده از آن‌ها، عدم قطعیت‌های موجود و درک روابط ذاتی میان داده‌ها را که ریاضیات کلاسیک قادر به حل آن‌ها نیست را می‌توان شبیه‌سازی کرد. در رابطه با کاربرد مدل‌های اشاره‌شده در پیش‌بینی جریان رودخانه Ahmadi و همکاران (۲۰۱۵) به مقایسه روش‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان برای پیش‌بینی جریان روزانه رودخانه باراندوزچای پرداختند. مقایسه نتایج دو مدل نشان داد که اگرچه دقت روش برنامه‌ریزی ژنتیک نسبت به روش ماشین بردار پشتیبان اندکی بیشتر بود، اما روش ماشین بردار پشتیبان به مراتب ساده‌تر از روش برنامه‌ریزی ژنتیک بوده، می‌تواند به‌عنوان یک روش کاربردی برای پیش‌بینی جریان روزانه به‌کار رود. در مطالعه‌ای، Solgi و همکاران (۲۰۱۷) به بررسی عملکرد مدل GEP با روش‌های پیش‌پردازش داده‌ها برای مدل‌سازی جریان رودخانه گاماسیاب پرداختند. برای این منظور، از داده‌های بارش، دما، تبخیر و جریان رودخانه گاماسیاب در ایستگاه وراینه با یک دوره آماری ۴۳ ساله استفاده شد. بررسی ساختارهای مختلف برای مدل GEP نشان داد که عملکرد مدل در دوره روزانه خوب بوده، ولی در دوره ماهانه عملکرد کاهش یافته است. مقایسه مدل ترکیبی موجک با مدل GEP نشان داد که عملکرد مدل ترکیبی در هر دو دوره زمانی روزانه و ماهانه از مدل ساده بهتر بوده است. همچنین، Dehghani و همکاران (۲۰۱۵) به مقایسه و ارزیابی مدل‌های شبکه عصبی بیزین، GEP، ماشین بردار پشتیبان و رگرسیون خطی در تخمین دبی جریان در حوضه صوفی‌چای پرداختند. کارایی مدل‌ها با استفاده از چهار معیار: ارزیابی مدل، ضریب همبستگی، ریشه میانگین مربعات خطا، ضریب نش-ساتکلیف و بایاس محاسبه شد. مقایسه مدل‌ها نشان داد که شبکه عصبی بیزین با ضریب همبستگی (۰/۹۹۱) و جذر میانگین مربعات خطا ($0/31 \text{ m}^3 \text{ s}^{-1}$)، بهترین تخمین را از داده‌های روزانه جریان رودخانه نسبت به مدل‌های GEP، ماشین بردار پشتیبان و رگرسیون خطی داشته است.

² Filyos³ Auto Regressive Moving Average¹ Support Vector Machine

جایگزین مناسبی برای مدل‌های مفهومی باشد.

مواد و روش‌ها

در این مطالعه، برای شبیه‌سازی فرایند بارش-رواناب، ایستگاه‌های آب‌سنجی و باران‌سنجی در محدوده دشت ملایر استفاده شد. حوزه آبخیز ملایر از دیدگاه مدیریت منابع آب یکی از محدوده‌های مطالعاتی حوزه علیای رودخانه کرخه با وسعت ۲۹۸۴ کیلومتر مربع محسوب می‌شود. قسمت‌های محدودی از این حوزه (حاشیه شمال غربی) از نظر تقسیمات کشوری جزء دشت توپسرکان، گوشه جنوب شرقی حوزه، جزء استان مرکزی و قسمت اعظم حوزه را محدوده دشت ملایر در بر می‌گیرد. متوسط نزولات جوی سالانه بر اساس دوره آماری ۱۴ ساله در ایستگاه‌های اندازه‌گیری پیهان و نامیله و بر اساس دوره آماری ۲۳ ساله در ایستگاه خیرآباد به ترتیب برابر ۲۹۶/۴، ۲۹۴/۵ و ۳۵۴/۴ و به‌طور میانگین برای دشت ملایر ۳۱۵ میلی‌متر در سال محاسبه شده است. مختصات جغرافیایی ایستگاه‌های مورد مطالعه دشت ملایر در جدول‌های ۱ و ۲ آورده شده است. در این مطالعه داده‌های روزانه بارش و رواناب ایستگاه‌های باران‌سنجی و آب‌سنجی دشت ملایر در طول سال‌های ۱۳۸۰ تا ۱۳۹۲ مورد استفاده قرار گرفت و همگنی داده‌ها با نرم‌افزار Minitab 16 بررسی شد. موقعیت مناطق مورد مطالعه، در شکل ۱ نمایش داده شده است.

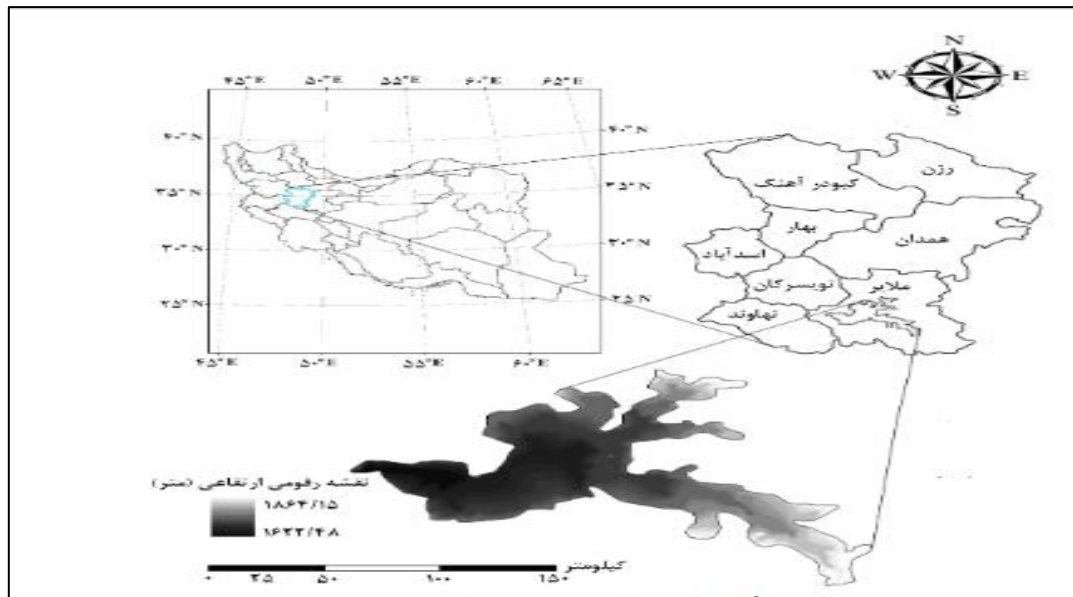
سامانه عصبی فازی تطبیقی تخمین زدند. نتایج مطالعه آن‌ها نشان داد، از بین مدل‌های یاد شده سامانه عصبی فازی تطبیقی عملکرد بسیار خوبی داشته است و به خوبی می‌تواند رواناب را پیش‌بینی کند. در پژوهشی دیگر، Solgi و همکاران (۲۰۱۸) از دو مدل GEP و ماشین بردار پشتیبان برای پیش‌بینی بارش ماهانه شهرستان نهاوند استفاده کردند. داده‌های مورد استفاده بارش، دما و رطوبت نسبی ماهانه ایستگاه ورآینه در یک دوره ۳۲ ساله ۱۳۹۲-۱۳۶۲ بود. نتایج به‌دست آمده نشان داد که عملکرد هر دو مدل خوب و مشابه بوده، ولی با توجه به بررسی معیارهای ارزیابی مختلف، مدل GEP عملکرد کمی بهتری داشته است. لذا هدف از این مطالعه، در ابتدا شبیه‌سازی جریان روزانه به‌وسیله شبکه عصبی فیلتر کالمن توسعه‌یافته (EKFNN) است که برای اولین بار برای این منظور مورد استفاده قرار گرفته است و به‌عنوان نوآوری این پژوهش مطرح است و سپس، مقایسه این رویکرد جدید در شبیه‌سازی بارش-رواناب با مدل GEP است که در بیشتر مطالعات، با توجه به تعداد زیاد مدل‌های موجود برای شبیه‌سازی، برتری خود را ثابت کرده است و سپس، مقایسه آن‌ها با یکدیگر برای تشخیص این‌که کدام یک از این مدل‌های شبیه‌سازی، نسبت به بقیه برتری دارد و سوال اساسی این پژوهش، این است که آیا مدل برتر در این مطالعه، با توجه به صحت در مرحله آموزش و تست می‌تواند در شرایط کمبود داده و اطلاعات،

جدول ۱- مختصات جغرافیایی ایستگاه‌های باران‌سنجی دشت ملایر

ایستگاه	طول جغرافیایی (شمالی)	عرض جغرافیایی (شرقی)	ارتفاع از سطح دریا (متر)
پیهان	۴۸° ۵۲' ۰۷"	۳۴° ۰۷' ۴۹"	۱۸۷۰
نامیله	۴۸° ۴۹' ۱۴"	۳۴° ۱۶' ۱۷"	۱۷۷۳

جدول ۲- مختصات جغرافیایی ایستگاه‌های آب‌سنجی دشت ملایر

ایستگاه	طول جغرافیایی (شمالی)	عرض جغرافیایی (شرقی)	ارتفاع از سطح دریا (متر)
پیهان	۴۸° ۵۲' ۲۰"	۳۴° ۰۷' ۵۸"	۱۸۳۰
مرویل	۴۸° ۵۳' ۱۴"	۳۴° ۱۰' ۰۱"	۱۸۴۱



شکل ۱- موقعیت منطقه مورد مطالعه

کمتری نسبت به الگوریتم استاندارد پس انتشار (SBP)^۲، می تواند شبکه ها را بهتر آموزش دهد. آن ها همچنین، نشان دادند که الگوریتم EKF نیازمند نمایش داده های آموزشی کمتری نسبت به SBP بوده است. اگرچه الگوریتم EKF می تواند با تکرارهای آموزشی کمتر، نتایج بهتری نسبت به SBP ارائه دهد، این روش لزوماً نیازمند زمان آموزشی کمتر نیست. Puskorius و Feldkamp (۱۹۹۱) نشان دادند که در واقع الگوریتم EKF جهانی در مقایسه با الگوریتم SBP نیاز به زمان آموزشی بیشتری دارد. این مطلب، به خاطر محاسبات اضافی مورد نیاز در هر تکرار آموزشی در الگوریتم EKF است. همچنین، آن ها نشان دادند که با اصلاح هوشمندانه الگوریتم EKF، وزن های شبکه عصبی می تواند در گروه های جدا مرتب شده و هر گروه می تواند به صورت جداگانه پردازش شود. به این ترتیب، با جدا کردن وزن ها، پیچیدگی محاسباتی الگوریتم EKF می توان تا حد زیادی کاهش یابد. یکی دیگر از پیشرفت های قابل توجه در الگوریتم آموزش شبکه عصبی EKF ارائه شده به وسیله Puskorius و Feldkamp (۱۹۹۷) ایده آموزش چند جریان است. با استفاده از تعریف الگوریتم EKF، با در نظر گرفتن تنها یک نمونه، به روزرسانی وزن انجام می شود. در مقابل، در جایی که به روزرسانی وزن بر مبنای بسیاری

استانداردسازی داده ها: به علت این که وارد کردن داده ها به صورت خام باعث کاهش سرعت و دقت مدل می شود، از روش استانداردسازی داده ها استفاده شده است. با روش استانداردسازی هر عدد تبدیل به عددی بین صفر تا یک می شود. با توجه به پیشنهاد Golabi و همکاران (۲۰۱۳) از معادله زیر برای استانداردسازی استفاده شده است.

$$y = 0.1 + 0.8 \times \left(\frac{x - \bar{x}}{x_{\max} - x_{\min}} \right) \quad (1)$$

که در آن، X داده مورد نظر، \bar{X} میانگین داده ها، X_{\max} بیشینه داده ها، X_{\min} کمینه داده ها و y داده استاندارد شده است. سپس، در این پژوهش برای ایجاد مدل ۷۰ درصد داده ها برای آموزش و ۳۰ درصد برای صحت سنجی مدل به کار رفت.

در این پژوهش، $R(t)$ ، $R(t-1)$ (به ترتیب بارندگی در زمان $t-1$ و t) و $Q(t-1)$ ، $Q(t-2)$ (به ترتیب دبی در زمان $t-2$ و $t-1$) به عنوان ورودی و $Q(t)$ (دبی در زمان t) به عنوان خروجی مدل در نظر گرفته شد.

الگوریتم فیلتر کالمن تعمیم یافته (EKF): الگوریتم فیلتر کالمن تعمیم یافته (EKF) و الگوریتم آموزش شبکه عصبی مصنوعی، اولین بار به وسیله Singhal و Wu (۱۹۸۹) معرفی شدند. آن ها نشان دادند که الگوریتم EKF با داشتن خطای آموزش

² Standard Back-Propagation¹ Extended Kalman Filter

نسل به جلو کیفیت جمعیت نیز به طور نسبی بهبود می‌یابد (Ferreira, ۲۰۰۶). مدل GEP نیز همانند الگوریتم ژنتیک و برنامه‌ریزی ژنتیک، یک الگوریتم ژنتیکی است (Alami و همکاران، ۲۰۱۴). به طوری که از جمعیتی از افراد استفاده می‌کند که آن‌ها را مطابق برانزنگی انتخاب می‌کند و تغییرات ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی اعمال می‌کند. تفاوت اساسی بین این سه الگوریتم، مربوط به ماهیت افراد آن‌ها است، به طوری که در الگوریتم ژنتیک، افراد رشته‌های خطی با طول ثابت (کروموزوم‌ها) و در برنامه‌ریزی ژنتیک، نهاده‌های غیرخطی با اندازه‌ها و اشکال متفاوت (درختان تجزیه) هستند، در حالی که در مدل GEP، افراد به صورت رشته‌های خطی با طول ثابت (ژنوم یا کروموزوم‌ها) کدگذاری شده، سپس به شکل نهاده‌های غیرخطی با اندازه‌ها و اشکال متفاوت (یعنی نمایش دیاگرام ساده یا ساختار درختی) بیان می‌شوند (Ferreira, ۲۰۰۱). برای کسب اطلاعات بیشتر با این مدل به پژوهش Alami و همکاران (۲۰۱۴) مراجعه شود. پارامترهای مورد استفاده و نرخ آن‌ها در استخراج مدل‌ها برای تخمین دبی با استفاده از مدل GEP به طور خلاصه در جدول ۳ ارائه شده است.

معیارهای ارزیابی مدل‌ها: برای ارزیابی مدل‌ها از سه معیار، ضریب تعیین^۱ (R^2)، ریشه میانگین مربعات خطا^۲ (RMSE)، پارامتر نش (NASH) استفاده شد. پارامتر آماری نش در سال ۱۹۷۰ توسط نش و ساتکلیف معرفی شد (Nash و Sutcliffe, ۱۹۷۰). در زیر معادله‌های این معیارها ارائه شده است.

$$RMSE = \sqrt{\frac{\sum (Q_{obs} - Q_{pre})^2}{N}} \quad (2)$$

$$R^2 = \left[\frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2}} \right]^2 \quad (3)$$

از نمونه‌ها باشد، یا حتی به روزرسانی در تمام مجموعه نمونه‌های آموزشی، انجام شود، الگوریتم پس‌انتشار می‌تواند در حالت دسته‌ای اجرا شود. در حالت دسته‌ای، قبل از انجام به روزرسانی وزن، شیب‌ها برای چندین نمونه آموزشی جمع می‌شوند. مفهوم چند جریان به الگوریتم EKF اجازه می‌دهد تا برای به روزرسانی هر وزن، بیش از یک نمونه آموزشی مورد پردازش قرار گیرد. برای مطالعه بیشتر در مورد این الگوریتم به Heimes (۱۹۹۸) مراجعه شود.

مدل برنامه‌ریزی بیان ژن (GEP): مدل GEP که در ادامه سیر تکاملی مدل‌های هوشمند به وجود آمده است، جزء روش‌های الگوریتم تکاملی محسوب می‌شود که مبنای تمامی آن‌ها بر اساس نظریه تکامل داروین استوار است (Danandeh Mehr و همکاران، ۲۰۱۰). مزیت مدل GEP نسبت به مدل‌های دیگر، از جمله شبکه عصبی مصنوعی این است که در مدل GEP، ابتدا ساختار (متغیرهای ورودی، هدف و مجموع توابع) تعریف شده و سپس، ساختار بهینه مدل و ضرایب، طی فرایند آموزش تعیین می‌شوند، در حالی که در شبکه‌های عصبی، ابتدا باید ساختار تعیین شود و فقط ضرایب مدل طی فرایند آموزش حاصل می‌شوند. همچنین، الگوریتم مدل GEP به طور خودکار می‌تواند متغیرهای ورودی که در مدل بیشترین تأثیر را دارند، انتخاب کند. در این روش، کروموزوم‌های خطی و ساده با طول ثابت، (مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود) و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، (مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک) ترکیب می‌شوند. اولین مرحله در الگوریتم مدل، تولید جمعیت اولیه از راه‌حل‌هاست. این امر می‌تواند به وسیله فرایند تصادفی و یا در نظر گرفتن اطلاعات ورودی درباره مسأله انجام شود. سپس، کروموزوم‌ها به صورت ساختار درختی ارائه شده و به وسیله تابع برآزش ارزیابی می‌شوند. در صورت دستیابی به راه‌حل مطلوب و یا رسیدن نسل‌ها به تعداد معین، تکامل متوقف شده و بهترین راه‌حل ارائه می‌شود. اگر شرایط توقف یافت نشود، نخبه‌گزینی انجام می‌شود و باقی راه‌حل‌ها به فرایندی گزینشی واگذار می‌شوند. این فرایند برای چندین نسل تکرار می‌شود و با پیش رفتن

¹ Coefficients of determination

² Root Mean Square Error

(R²) نشان می دهد که چند درصد تغییرات متغیر وابسته به وسیله متغیر مستقل تبیین می شود. RMSE ریشه میانگین مربع خطاهای داده های شبیه سازی شده و مشاهداتی را بیان می کند. واضح است که هر چه مقدار این عدد کمتر باشد، صحت نتایج به دست آمده بهتر خواهد بود. همچنین، هر چه مقدار پارامتر نش به یک نزدیک تر باشد، نشان دهنده صحت بیشتر مقادیر شبیه سازی شده با مقادیر مشاهداتی است.

$$NASH = 1 - \frac{\sum_{i=1}^N (Q_{obs} - Q_{pre})^2}{\sum_{i=1}^N (Q_{obs} - \bar{Q}_{obs})^2} \quad (4)$$

که در آن ها، n تعداد داده ها، Q_{obs} دبی جریان مشاهداتی، \bar{Q} میانگین دبی جریان مشاهداتی، Q_{pre} دبی جریان شبیه سازی شده، Q_{max} بیشینه دبی جریان و Q_{min} کمینه دبی جریان است. ضریب تعیین

جدول ۳- مقادیر پارامترهای مورد استفاده با استفاده از روش GEP

عملگرهای ژنتیکی Genetic operators		تنظیمات کلی General settings	
۰/۰۴۴	نرخ جهش	۳۰	تعداد کروموزومها
۰/۱	نرخ وارون سازی	۸	اندازه راس
۰/۱	نرخ ترانهش درج متوالی	۳	تعداد ژن ها در هر کروموزوم
۰/۱	نرخ ترانهش ریشه درج متوالی	۱۰۰۰	تعداد جمعیت تولیدی
۰/۳	نرخ ترکیب تک نقطه ای	جمع (+)	تابع پیوند
۰/۳	نرخ ترکیب دو نقطه ای		

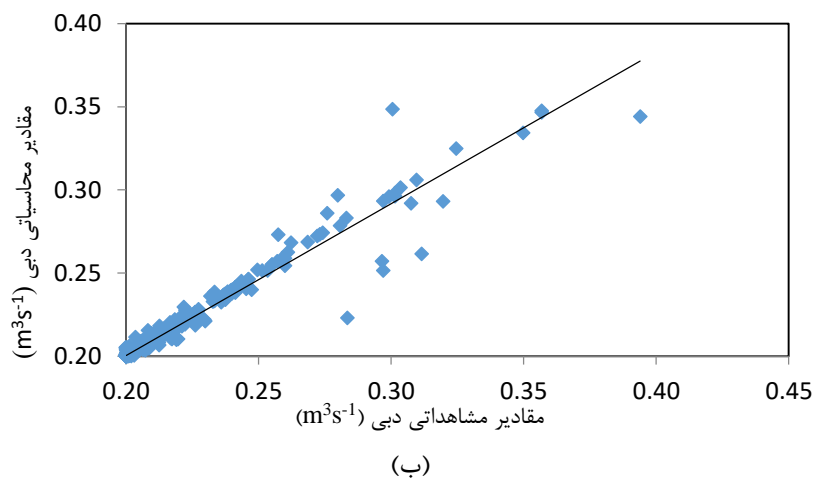
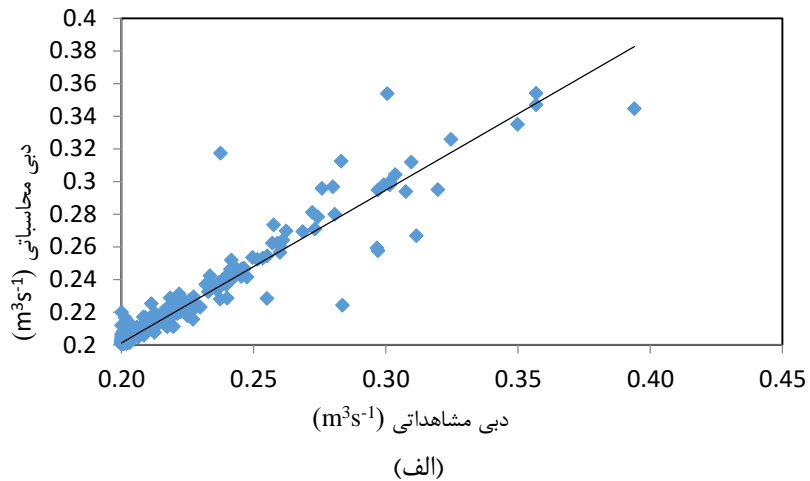
نتایج و بحث

نتایج شبیه سازی دبی جریان ایستگاه آب سنجی پیهان: در شکل ۲، نمودار پراکنش میان مقادیر مشاهداتی و شبیه سازی شده دبی های روزانه به وسیله دو مدل EKFNNN و GEP برای داده های بخش صحت سنجی نشان داده شد. همان طور که از شکل ۲ مشخص است، بیشتر مقادیر شبیه سازی شده و مشاهداتی، به غیر از چند نقطه روی خط نیمساز قرار گرفته، این امر دلالت بر برابری مقادیر مشاهداتی و شبیه سازی شده بر اساس خط (Y=X) است. معادله خط رگرسیون برای مدل GEP به صورت $y = 0.9356x + 0.0142$ است که در اینجا مقدار

۰/۹۳۵۶ شیب خط رگرسیون و مقدار ۰/۰۱۴۲ عرض از مبدا می باشد. همچنین، معادله خط رگرسیون برای مدل EKFNNN به صورت $y = 0.9123x + 0.018$ است که در اینجا مقدار ۰/۹۱۲۳ شیب خط رگرسیون و مقدار ۰/۰۱۸ عرض از مبدا است. مقادیر سه معیار ارزیابی برای مقایسه کمی دو مدل در ایستگاه پیهان نیز در جدول ۴ ارائه شده است. با توجه به جدول ۴ و با مقایسه نتایج با یکدیگر مشخص می شود که هر دو مدل با دقت خوبی می توانند دبی جریان روزانه را در ایستگاه پیهان شبیه سازی کنند، اما از بین دو مدل به کار رفته مدل شبکه عصبی توسعه یافته با فیلتر کالمن عملکرد بهتری داشته است.

جدول ۴- نتایج مرحله صحت سنجی دو مدل GEP و EKFNN برای دبی های روزانه ایستگاه پیهان

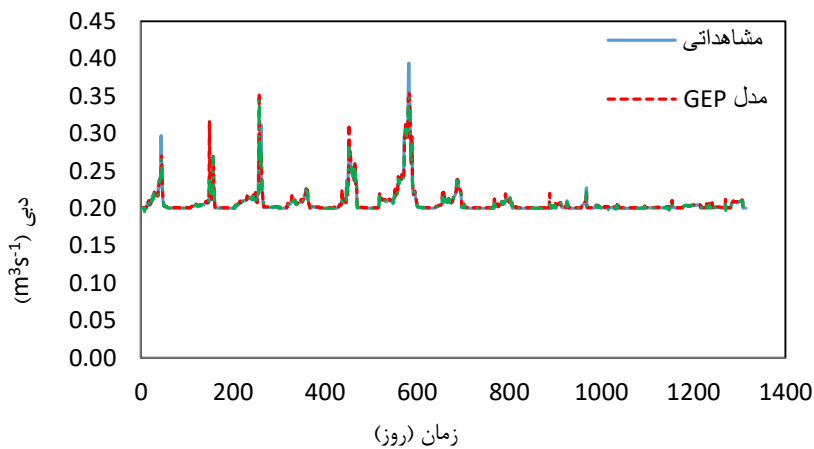
EKFNN			GEP			ایستگاه
RMSE	R ²	NASH	RMSE	R ²	NASH	
۰/۲۹۳	۰/۹۵۲	۰/۹۵	۰/۴۳۹	۰/۹۳۵	۰/۹۳	پیهان



شکل ۲- نمودار پراکنش میان مقادیر مشاهداتی و شبیه‌سازی شده دبی‌های روزانه در ایستگاه پیهان، الف) GEP و ب) EKFNN

می‌شود، تطابق مناسبی میان مقادیر مشاهده‌ای و شبیه‌سازی شده وجود دارد، ولی در برآورد مقادیر اوج هر دو مدل دارای عملکرد ضعیفی هستند.

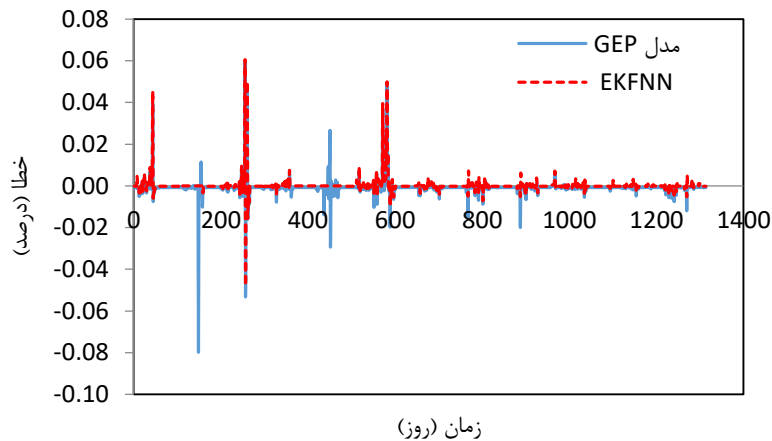
در شکل ۳، تغییرات مقادیر شبیه‌سازی شده و مشاهداتی دبی‌های روزانه نسبت به زمان برای هر دو مدل نشان داده شده است. همان‌طور که مشاهده



شکل ۳- سری‌های زمانی رواناب‌های مشاهداتی و شبیه‌سازی شده روش‌های GEP و EKFNN در ایستگاه پیهان

در این شکل نشان داده شده است، برای هر دو مدل بیشتر خطاها در نوار (± 0.1) درصد قرار گرفته است و بیشترین مقدار خطا برابر $7/97$ درصد و مربوط به مدل GEP است.

در شکل ۴، اختلاف مقادیر مشاهداتی و شبیه سازی شده دبی روزانه در مرحله صحت سنجی به وسیله هر دو مدل به صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به دوره آماری ترسیم شده است. همان طور که



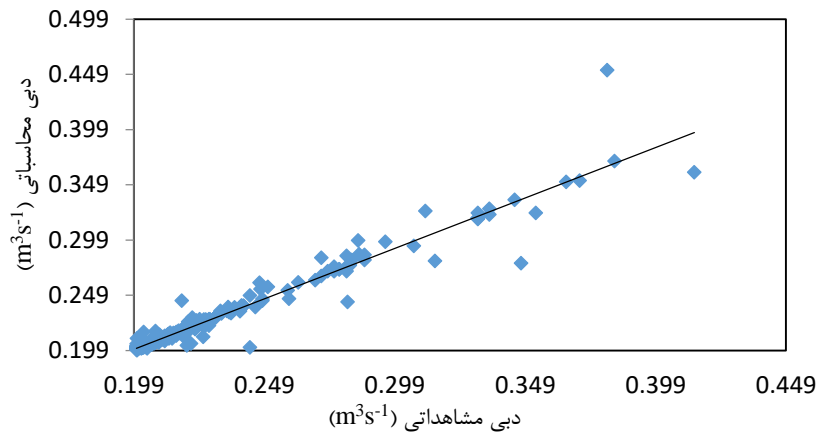
شکل ۴- نمودار خطای پهنه مدل های GEP و EKFNN به صورت درصدی از میانگین مقادیر مشاهداتی داده های صحت سنجی در ایستگاه پیهان

عرض از مبدا می باشد. همچنین، معادله خط رگرسیون برای مدل EKFNN به صورت $y = 0.9453x + 0.0114$ می باشد که در اینجا مقدار 0.9453 شیب خط رگرسیون و مقدار 0.0114 عرض از مبدا است. مقادیر سه معیار ارزیابی برای مقایسه کمی دو مدل در ایستگاه مرویل نیز در جدول ۵ ارائه شده است. با توجه به جدول ۵ و با مقایسه نتایج با یکدیگر مشخص می شود که هر دو مدل با دقت خوبی می توانند دبی جریان روزانه را در ایستگاه مرویل شبیه سازی کنند، اما از بین دو مدل به کار رفته مدل شبکه عصبی توسعه یافته با فیلتر کالمن عملکرد بهتری داشته است.

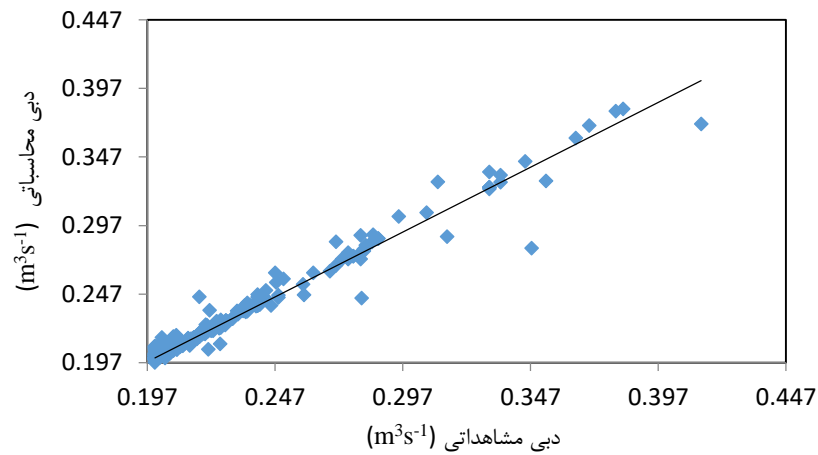
نتایج شبیه سازی دبی جریان ایستگاه آب سنجی مرویل: در شکل ۵، نمودار پراکنش میان مقادیر مشاهداتی و شبیه سازی شده دبی های روزانه به وسیله دو مدل EKFNN و GEP برای داده های بخش صحت سنجی نشان داده شد. همان طور که در شکل ۵ نشان داده شده است، بیشتر مقادیر شبیه سازی شده و مشاهداتی، به غیر از چند نقطه روی خط نیمساز قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و شبیه سازی شده بر اساس خط $(Y=X)$ است. معادله خط رگرسیون برای مدل GEP به صورت $y = 0.9134x + 0.0182$ می باشد که در اینجا مقدار 0.9134 شیب خط رگرسیون و مقدار 0.0182

جدول ۵- نتایج مرحله صحت سنجی دو مدل GEP و EKFNN برای دبی های روزانه ایستگاه مرویل

EKFNN			GEP			ایستگاه
RMSE	R ²	NASH	RMSE	R ²	NASH	
0.024	0.976	0.98	0.108	0.958	0.96	مرویل



(الف)

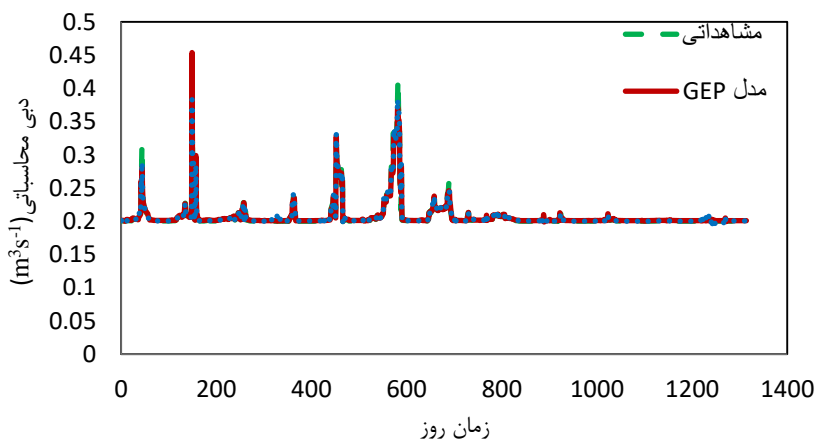


(ب)

شکل ۵- نمودار پراکنش میان مقادیر مشاهداتی و شبیه‌سازی شده دبی‌های روزانه در ایستگاه مرویل، الف) GEP و ب) EKFNN

شود، تطابق مناسبی میان مقادیر مشاهده‌ای و شبیه‌سازی شده وجود دارد، ولی در برآورد مقادیر اوج هر دو مدل دارای عملکرد ضعیفی هستند.

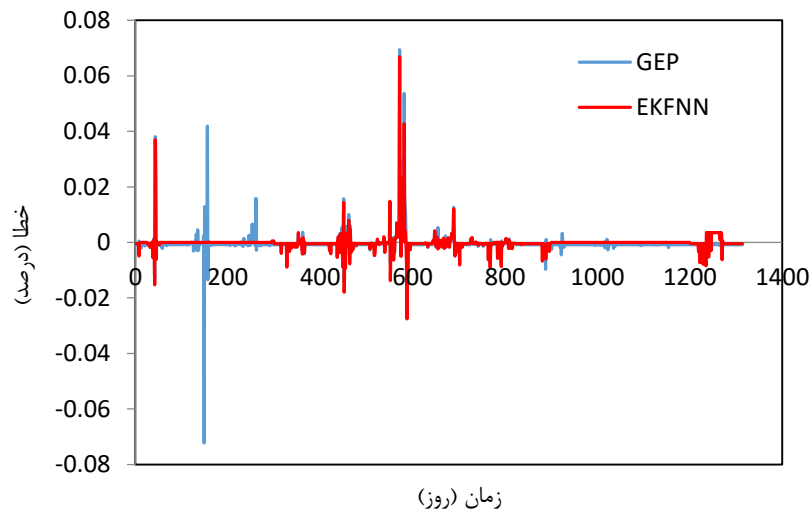
در شکل ۶ تغییرات مقادیر شبیه‌سازی شده و مشاهداتی دبی‌های روزانه نسبت به زمان برای هر دو مدل نشان داده شده است. همان‌طور که مشاهده می‌-



شکل ۶- سری‌های زمانی رواناب‌های مشاهداتی و شبیه‌سازی شده روش‌های GEP و EKFNN در ایستگاه مرویل

در این شکل نشان داده شده است، برای هر دو مدل بیشتر خطاها در نوار (± 0.1) درصد قرار گرفته است و بیشترین مقدار خطا برابر $7/22$ درصد و مربوط به مدل GEP می باشد.

در شکل ۷، اختلاف مقادیر مشاهداتی و شبیه سازی شده دبی روزانه در مرحله صحت سنجی به وسیله هر دو مدل به صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به دوره آماری ترسیم شده است. همان طور که



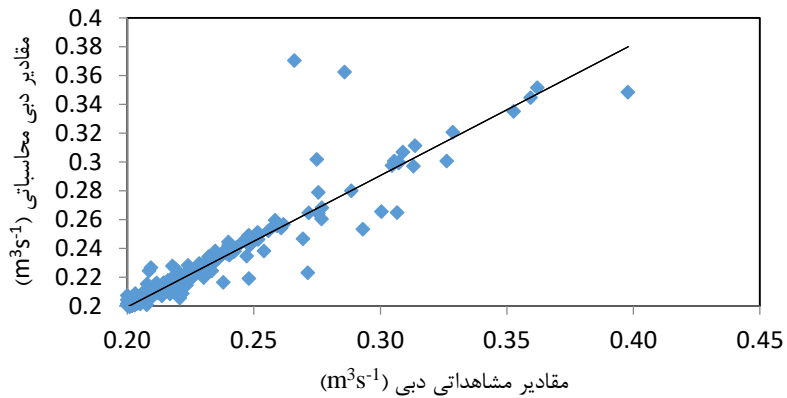
شکل ۷- نمودار خطای بهینه مدل های GEP و EKFNN به صورت درصدی از میانگین مقادیر مشاهداتی داده های صحت سنجی در ایستگاه مرویل

$0.9135x + 0.0167$ شیب خط رگرسیون و مقدار 0.167 عرض از مبدا است. همچنین، معادله خط رگرسیون برای مدل EKFNN به صورت $y = 0.9446x + 0.0104$ است که در اینجا مقدار 0.9446 شیب خط رگرسیون و مقدار 0.0104 عرض از مبدا می باشد. مقادیر سه معیار ارزیابی برای مقایسه کمی دو مدل در ایستگاه نامیله نیز در جدول ۶ ارائه شده است. با توجه به جدول ۶ و با مقایسه نتایج با یکدیگر مشخص می شود که هر دو مدل با دقت خوبی می توانند دبی جریان روزانه را در ایستگاه نامیله شبیه سازی کنند، اما از بین دو مدل به کار رفته مدل شبکه عصبی توسعه یافته با فیلتر کالمن عملکرد بهتری داشته است.

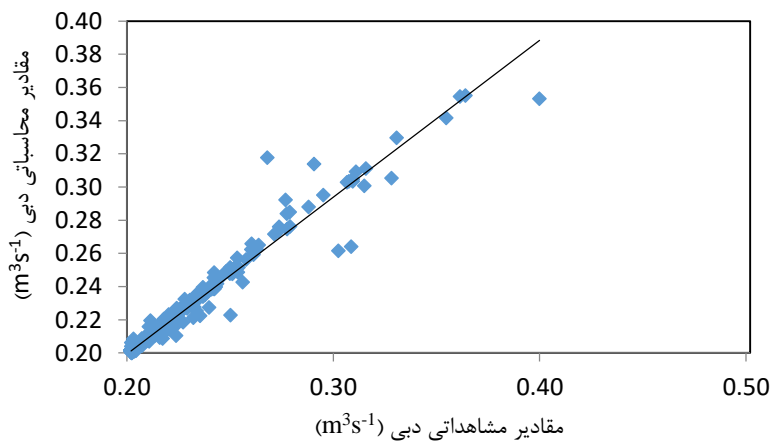
ایستگاه باران سنجی نامیله: در این ایستگاه با توجه به این که در انتهای حوضه ایستگاه آب سنجی وجود نداشت، از مجموع داده های ایستگاه های پیهان و مرویل استفاده شد. در شکل ۸، نمودار پراکنش میان مقادیر مشاهداتی و شبیه سازی شده دبی های روزانه به وسیله دو مدل EKFNN و GEP برای داده های بخش صحت سنجی نشان داده شد. همان طور که از شکل ۸ مشخص است، بیشتر مقادیر شبیه سازی شده و مشاهداتی، به غیر از چند نقطه روی خط نیمساز قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و شبیه سازی شده بر اساس خط $(Y=X)$ است. معادله خط رگرسیون برای مدل GEP به صورت $y = 0.9135x + 0.0167$ است که در اینجا مقدار

جدول ۶- نتایج مرحله صحت سنجی دو مدل GEP و EKFNN برای دبی های روزانه ایستگاه نامیله

ایستگاه	EKFNN			GEP		
	RMSE	R ²	NASH	RMSE	R ²	NASH
نامیله	۰/۴۱۳	۰/۹۶۸	۰/۹۶	۰/۷۱۷	۰/۹۱۶	۰/۹۱



(الف)

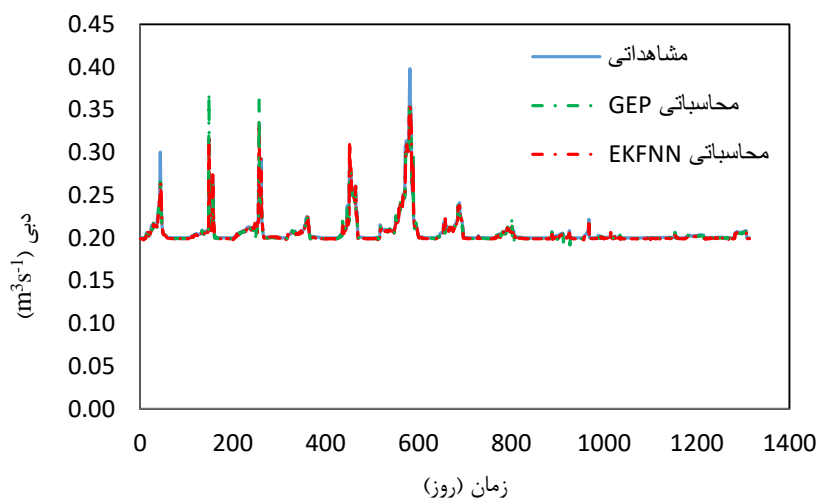


(ب)

شکل ۸- نمودار پراکنش میان مقادیر مشاهداتی و شبیه‌سازی شده دبی‌های روزانه در ایستگاه نامیله، الف) GEP (ب) EKFNN

شبیه‌سازی شده وجود دارد، ولی مثل دو ایستگاه قبل هر دو مدل در برآورد مقادیر اوج دارای عملکرد ضعیفی هستند.

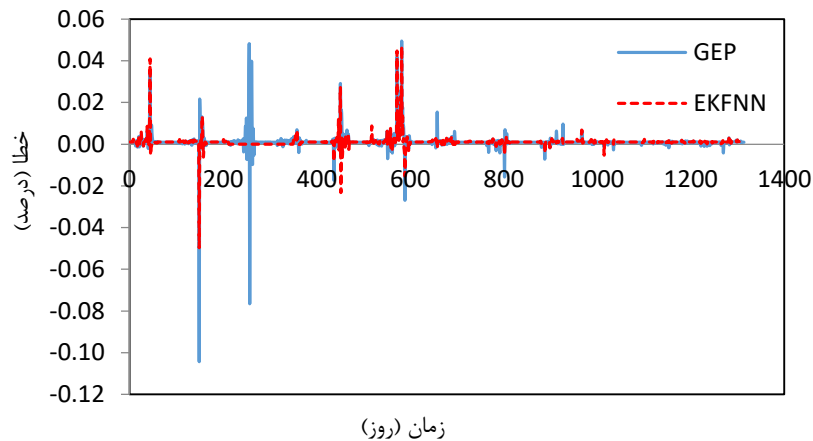
در شکل ۹، تغییرات مقادیر شبیه‌سازی شده و مشاهداتی دبی‌های روزانه نسبت به زمان برای هر دو مدل نشان داده شده است. همان‌طور که مشاهده می‌شود، تطابق مناسبی میان مقادیر مشاهده‌ای و



شکل ۹- سری‌های زمانی رواناب‌های مشاهداتی و شبیه‌سازی شده روش‌های GEP و EKFNN در ایستگاه نامیله

در این شکل نشان داده شده است، برای هر دو مدل بیشتر خطاها در نوار (± 0.1) درصد قرار گرفته است. بیشترین مقدار خطا برابر $10/43$ درصد و مربوط به مدل GEP است.

در شکل ۱۰، اختلاف مقادیر مشاهداتی و شبیه سازی شده دبی روزانه در مرحله صحت سنجی به وسیله هر دو مدل به صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به دوره آماری ترسیم شده است. همان طور که



شکل ۱۰- نمودار خطای بهینه مدل های GEP و EKFNN به صورت درصدی از میانگین مقادیر مشاهداتی داده های صحت سنجی در ایستگاه نامیله

مربعات خطا (RMSE) برابر $0/42$ و پارامتر NASH برابر $0/93$ و دقت (R^2) مدل شبکه عصبی توسعه یافته با فیلتر کالمن $0/96$ و کمترین ریشه میانگین مربعات خطا برابر (RMSE) $0/24$ و پارامتر NASH برابر $0/96$ شده است. بنابراین، از بین دو مدل به کار رفته مدل شبکه عصبی توسعه یافته با فیلتر کالمن عملکرد بهتری داشته است. هر دو مدل در شبیه سازی مقادیر اوج دقت کمتری داشته اند، به طوری که با افزایش میزان جریان، دقت مدل کاهش می یابد. نتایج هر دو مدل در ایستگاه های ملایر از دقت خوبی برخوردار بوده است و نتایج مربوط به مدل برنامه ریزی بیان ژن با مطالعات مختلفی همچون Dehghani و همکاران (۲۰۱۵) و Solgi و همکاران (۲۰۱۷) همخوانی دارد ولی چون با مدل شبکه عصبی توسعه یافته با فیلتر کالمن تا کنون پژوهشی در این زمینه انجام نگرفته است، پیشنهاد می شود که در مطالعات آینده این مدل جدید در کنار سایر مدل ها برای شبیه سازی پارامترهای مختلف هواشناسی و هیدرولوژی مورد استفاده قرار گیرد و کارایی آن بررسی شود.

نتیجه گیری

شبیه سازی دقیق جریان و پیش بینی آن در مدیریت و برنامه ریزی منابع آب از اهمیت بسیار بالایی برخوردار است و می تواند یک مبنای اساسی برای پیش بینی منابع آب موجود فراهم کرده و منجر به کاهش خطر و جلوگیری از وقوع سیل ها شود. در این پژوهش برای اولین بار از شبکه عصبی توسعه یافته با فیلتر کالمن به منظور شبیه سازی فرایند بارش-رواناب دشت ملایر استفاده شد و نتایج این مدل با مدل GEP که در بیشتر مطالعات، با توجه به تعداد زیاد مدل های موجود برای شبیه سازی، برتری خود را ثابت کرده است، مقایسه شد. مقادیر دبی روزانه مشاهداتی با دبی روزانه تخمین زده شده در مدل های مذکور با استفاده از معیارهای ارزیابی مورد مقایسه قرار گرفت. همان طور که مشاهده شد، با مقایسه جواب ها با یکدیگر مشخص می شود که هر دو مدل با دقت خوبی می توانند دبی جریان روزانه را شبیه سازی کنند. در دشت ملایر برای مدل برنامه ریزی بیان ژن دقت (R^2) به طور متوسط برابر $0/94$ و کمترین ریشه میانگین

منابع مورد استفاده

1. Ahmadi, F., F. Radmanesh and R. Mirabbasi. 2015. Comparison between genetic programming and support vector machine methods for daily river flow forecasting, case study: Barandoozchay River. *Journal of Water and Soil*, 28(6): 1162-1171 (in Persian).
2. Alami, M.T., S. Sadeghfam, M.H. Fazelifard and L. Naghipour. 2014. Modeling the data series. University of Tabriz, 304 pages (in Persian).
3. Danandeh Mehr, A. and M.R. Majdzadeh Tabatabai. 2010. Prediction of daily discharge trend of river flow based on genetic programming. *Journal of Water and Soil*, 24(2): 325-333 (in Persian).
4. Dehghani, R., M.A. Ghorbani, M. Teshnelab and A. Rikhtehgar Gheasi. 2015. Comparison and evaluation of bayesian neural network, gene expression programming, support vector machine and linear regression in river discharge estimation, case study: Sufi Chay Basin. *The Iranian Society of Irrigation and Water Engineering*, 5(20): 66-85 (in Persian).
5. Feldkamp, L.A. and G.V. Puskorius. 1994. Training controllers for robustness: multi-stream DEW. In *Proceedings of the IEEE International Conference on Neural Networks*, Orlando, 6: 2377-2382.
6. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex System*, 13:87-129.
7. Ferreira, C. 2006. *Gene expression programming: mathematical modeling by an artificial intelligence (studies in computational intelligence)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
8. Ghorbani, M.A. and R. Dehghani. 2016. Application of bayesian neural networks, support vector machines and gene expression programming analysis of rainfall-runoff monthly, case study: Kakarza River. *Irrigation Science and Engineering*, 39(2): 125-138 (in Persian).
9. Golabi, M.R., A.M. Akhondali and F. Radmanesh. 2013. Comparison of the performance of different artificial neural network algorithms in seasonal rainfall modeling, case study: selected stations in Khuzestan Province. *Journal of Applied Geosciences Research*, 13(30): 151-169 (in Persian).
10. Heimes, F. 1998. Extended Kalman Filter neural network training: experimental results and algorithm improvements. *Systems, Man and Cybernetics*, IEEE International Conference on San Diego, CA, USA.
11. Karimi, S., J. Shiri, O. Kisi and A.A. Shiri. 2015. Short-term and long-term streamflow prediction by using 'wavelet-gene expression' programming approach. *ISH Journal of Hydraulic Engineering*, 3: 1-15 (in Persian).
12. Motamednia, M., A. Nohegar, A. Malekian, M. Saberi and K. Karimi. 2017. Runoff prediction using intelligent models. *Journal of Ecohydrology*, 4(4): 955-968 (in Persian).
13. Naeimi Kalourazi, Z., Kh. Ghorbani, M. Salarijazi and A.A. Dehghani. 2017. Investigation of effect of basin's physiographic and climatic parameters in seasonal river flow simulation. *Ecohydrology*, 3(4): 545-555 (in Persian).
14. Nash, J.E. and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part I, a discussion of principles. *Journal of Hydrology*, 10(3): 282-290.
15. Noori, R., M.A. Abdoli, A. Ameri and M. Jalili-Ghazizade. 2008. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis, a case study of Mashhad. *Environmental Progress and Sustainable Energy*, 28(2): 249-258.
16. Noori, N., L. Kalin. 2016. Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, 533: 141-151.
17. Noori, R., A. Karbassi, A. Farokhnia and M. Dehghani. 2009. Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques. *Environmental Engineering Science*, 26(10): 1503-1510.
18. Puskorius, G.V. and L.A. Feldkamp. 1991. Decoupled Extended Kalman Filter training of feed forward layered networks. In *Proceedings of the International Joint Conference on Neural Networks*, Seattle.
19. Puskorius, G.V. and L.A. Feldkamp. 1997. Multi-stream Extended Kalman Filter training for static and dynamic neural networks. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetic*, Orland.
20. Singh, G., R.K. Panda and M. Lamers. 2015. Modeling of daily runoff from a small agricultural watershed using artificial neural network with resampling techniques. *Journal of Hydroinformatics*, 17(1): 56-74.
21. Singhal, S. and L. Wu. 1989. Training multilayer perceptions the Extended Kalman Algorithm. *Advances in Neural Information Processing Systems*, San Meteo, CA: Morgan Kaufmann, 133-140.
22. Solgi, A., H. Zarei and M.R. Golabi. 2017. Performance assessment of gene expression programming model using data preprocessing methods to modeling river flow. *Journal of Water and Soil Conservation*, 24(2): 185-201 (in Persian).

23. Solgi, A., H. Zarei, M. Shehndarabi and S. Alidadi. 2018. Monthly precipitation forecast using gene expression planning models and support vector machine. *Journal of Applied Geosciences Research*, 50: 91-103 (in Persian).