نشریه آبیاری و زهکشی ایران
شماره 4، جلد 10، مهر - آبان 1395، ص. 488-479

Iranian Journal of Irrigation and Drainage
No. 4, Vol. 10, Sept.-Oct. 2016, p. 479-488

پیش‌بینی جریان روزانه رودخانه با استفاده از مدل‌های داده محور

میثم سالاری‌جزی[1*]، خلیل قربانی[2]، الهه سهرابیان[3]، محمد عبدالحسینی[4]

تاریخ دریافت: 1395/3/10 تاریخ پذیرش: 1395/7/28

چکیده

پیش‌بینی صحیح جریان روزانه رودخانه یک ابزار مناسب برای برنامه‌ریزی و مدیریت منابع آب سطحی می‌باشد. به کارگیری مدل‌هایی مانند مدل درختی M5 و برنامه‌ریزی بیان ژن (GEP) که معادلات صریحی را برای پیش‌بینی ارایه می‌کنند موجب افزایش کارایی مدل‌های پیش‌بینی می‌شود. در این مطالعه جهت پیش‌بینی دبی جریان روزانه رودخانه، حوضه آبریزگالیکش از مناطق سیل‌خیز استان گلستان به عنوان منطقه مطالعاتی استفاده شد. داده‌های بارش و دبی جریان روزانه ایستگاه‌های هواشناسی و هیدرومتری گالیکش در یک دوره آماری 26 ساله (1388-1363)، استفاده و متغیرهای مستقلی از بارش و دبی جریان روزانه یک تا پنج گام زمانی قبل تشکیل شد و بر اساس آن‌ها پیش‌بینی دبی جریان روزانه با سه مدل درختی M5 و برنامه‌ریزی بیان ژن و مدل شبکه عصبی مصنوعی پرسپترون انجام شد. نتایج نشان داد که هر سه مدل دارای کارایی مناسب می‌باشند و هم‌چنین میزان جریان را بیش‌تر از مقادیر مشاهداتی برآورد می‌کنند. مقایسه نتایج مدل‌های مختلف نشان‌دهنده برتری نسبی مدل درختی M5 نسبت به مدل‌های دیگر می‌باشد. در حالت کلی می‌توان گفت که هر سه روش مذکور ضمن رقابت با یکدیگر نتایج نسبتاً دقیقی را جهت پیش‌بینی جریان روزانه در منطقه مورد نظر ارایه می‌کنند ولی به دلیل ارایه روابط خطی ساده و قابل فهم توسط مدل درختی M5، این روش می‌تواند به عنوان روشی کاربردی و جایگزین برای پیش‌بینی جریان روزانه مورد توجه قرار گیرد.

واژه‌های کلیدی: برنامه‌ریزی بیان ژن، پیش‌بینی جریان، حوضه گالیکش، شبکه عصبی مصنوعی، مدل درختی M5، مدل‌های داده محور

مقدمه

لزوم پیش‌بینی مناسب جریان رودخانه در کارهای عمرانی، ساماندهی رودخانه، سامانه‌های هشدار سیل و به‌خصوص برنامه‌ریزی جهت بهره‌برداری بهینه از مخازن سدها ضروری به نظر می‌رسد. تاکنون روابط و الگوهای گوناگون و پیچیده‌ای برای پیش‌بینی میزان آبدهی رودخانه‌ها مانند انواع الگوهای مفهومی بارش-رواناب، الگوهای سری زمانی و الگوهای ترکیبی (هیبرید) ارایه شده است، لیکن روابط ارایه شده بدلیل عدم شناخت دقیق و نیز پیچیدگی عوامل موثر بر آبدهی‌رودخانه‌ها، در بسیاری از موارد با مقادیر مشاهده شده تطابق نداشته و در مواقعی نیز میزان مقادیر محاسبه شده از روابط گوناگون تفاوت معنی‌داری با یکدیگر داشته‌اند (Liong et al., 2002). مطالعات کتابخانه‌ای انجام شده نشان می‌دهد که در دهه اخیر روش برنامه ریزی ژنتیک (GP[5]) به عنوان یک روش تناوبی موثر و نیز مدل‌های درختی به عنوان یکی از تکنیک‌های داده‌کاوی در پیش‌بینی پارامترهای هیدرولوژیکی و هواشناسی در زمینه مهندسی آب مورد استفاده قرار گرفته است (دانند مهر و مجدزاده طباطبایی، 1388). لیونگ و همکاران و هم‌چنین وایگهام و کراپر، با مدل‌سازی بارش-رواناب به روش برنامه‌ریزی ژنتیک بدین نتیجه دست یافتند که پیش‌بینی رفتار بارش-رواناب در حوضه‌های آبریز به کمک برنامه‌ریزی ژنتیک سبب بروز خطای کم‌تری خواهد شد (Liong et al., 2002; Whigham and Crapper., 2011). فرناندو و همکاران با استفاده از برنامه‌ریزی بیان ژن (GEP[6]) جریان رودخانه 4 حوضه Yanbian، Han، Brosna و Baihe را پیش‌بینی کردند و با توجه به شاخص‌های آماری بدست آمده نتیجه گرفتند مدل برنامه‌ریزی بیان ژن در پیش‌بینی جریان رودخانه در هر 4 حوضه از کارایی مناسبی برخوردار است (Fernando et al., 2012). دانند مهر و طباطبایی (1388) ضمن معرفی روش برنامه‌ریزی ژنتیک به عنوان یک روش

1- استادیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان
2- دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان
3-دانش آموخته کارشناسی ارشد مهندسی منابع آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان
4- استادیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان
(*-نویسنده مسئول: Email: meysam.salarijazi@gau.ac.ir)

5- GP:Genetic Programming
6- GEP: Gene Expression Programming

صریح برای پیش‌بینی جریان رودخانه‌ها، از این روش به منظور بررسی تاثیر توالی دبی روزانه در پیش‌بینی جریان روزانه رودخانه آبسرده واقع در استان لرستان، استفاده کردند و دقت نتایج حاصله را با روش شبکه عصبی مصنوعی (ANN[1]) مورد مقایسه قرار دادند که نتایج حاکی از کارایی و دقت بالای برنامه‌ریزی ژنتیک در مقایسه با شبکه عصبی مصنوعی می‌باشد. فلاحی و همکاران (1390) از مدل رگرسیون درختی به منزله روشی کارآمد برای پیش‌بینی بارندگی در سه ایستگاه هواشناسی بندرانزلی، اراک و قم استفاده کردند که نتایج بیانگر توانایی و دقت بالای این مدل در پیش‌بینی بود. ظهیری و قربانی (1391) به عنوان یک راه حل جدید از درخت تصمیم M5 برای محاسبه دقیق دبی جریان در مقاطع مرکب استفاده کردند. نتایج نشان داد که مدل درخت تصمیم M5 ضمن سادگی محاسبه‌ها و معادله‌های ارایه شده قابلیت خوبی در تخمین دبی جریان در مقاطع مرکب آزمایشگاهی و صحرایی دارد. لاندهی و دیکسیت مدل درختی M5 را برای پیش‌بینی جریان رودخانه در یک روز قبل در دو ایستگاه رودخانه نارمادا و دیگری در حوضه رودخانه کریشنا در هند به کار بردند (Londhe and Dixit., 2010). ستاری و همکاران عملکرد مدل‌های درختیM5 و ماشین بردار پشتیبان (SVM[2]) را در برآورد جریان روزانه رودخانه سوهو، در حاشیه شهر آنکارا بررسی کردند (Sattari et al., 2013). نتایج حاصل از پژوهش آنان نشان داد که مدل درختی M5 نسبت به مدل SVM جریان روزانه رودخانه را بهتر پیش‌بینی می‌کند. هاتاچاریا و سولوماتین از روش‌های شبکه عصبی مصنوعی و مدل درختی M5 برای الگوسازی رابطه بین سطح آب و دبی جریان بهره بردند که بر اساس نتایج حاصل، دقت روش‌های فوق بیش‌تر از روش‌های معمول برآورد گردید (Bhattacharya and Solomatine., 2005). با توجه به مطالب ذکر شده، مدل‌های درختی M5 و GEP می‌توانند ابزار مناسبی جهت پیش‌بینی جریان روزانه در نظر گرفته شوند. بنابراین هدف از پژوهش حاضر پیش‌بینی دبی جریان روزانه با استفاده از این دو روش و به عنوان روش‌هایی که روابط صریحی را ارایه می‌کنند و مقایسه نتایج آن‌ها با روش ANN به عنوان یک روش جعبه سیاه است. بدین منظور از داده‌های دبی و بارش روزانه ثبت شده ایستگاه گالیکش در حوضه آبریز گالیکش در استان گلستان استفاده شد.

مواد و روش‌ها

منطقه مطالعاتی

حوضه آبریز گالیکش، واقع در استان گلستان از زیر حوضه‌های گرگان‌رود می‌باشد که دارای مساحتی برابر با 404/8 کیلومتر مربع و محیطی برابر با 88/6 کیلومتر است (شکل1). حداکثر ارتفاع حوضه 2461 متر، حداقل ارتفاع 378 متر و ارتفاع متوسط حوضه برابر با 1295 متر از سطح دریا و شیب متوسط حوضه 23/3 درصد می‌باشد. رودخانه اوغان در این حوضه قرار دارد که رواناب این رودخانه به رودخانه گرگان‌رود متصل شده و در نهایت در سد گلستان ذخیره می‌شود. طول شاخه اصلی رودخانه 26/2 کیلومتر و شیب خالص شاخه اصلی 3/5 درصد می‌باشد. زمان تمرکز این حوضه بر اساس روش کرپیچ برابر 3/9 ساعت و زمان تاخیر آن برابر 2/3 ساعت است. ایستگاه هیدرومتری گالیکش در ارتفاع 250 متری و در طول جغرافیایی ´28 °55 و عرض جغرافیایی ´ 16 °37 قرار دارد.

داده‌های مورد استفاده

در این پژوهش از دو سری داده بارش و دبی روزانه مربوط به ایستگاه‌های هواشناسی و هیدرومتری طی یک دوره آماری 26 ساله (1388-1363) استفاده شد که سال‌های 1370-1363 برای دوره واسنجی و سال‌های 1386-1382 برای دوره صحت‌سنجی انتخاب شدند. سری‌های زمانی بارش و دبی روزانه با تاخیر تا پنج گام زمانی نیز استخراج و به عنوان متغیرهای مستقل و پیش‌بینی کننده دبی در نظر گرفته شدند.

برنامه‌ریزی بیان ژن (GEP)

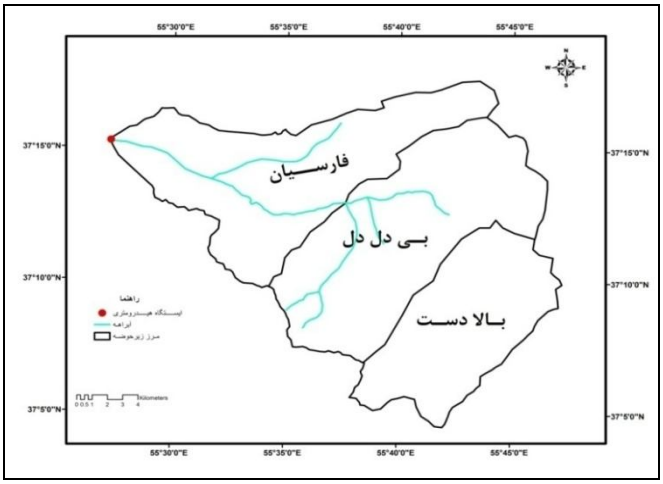
برنامه‌ریزی بیان ژن تعمیم یافته الگوریتم ژنتیک می‌باشد که برای اولین بار توسط جان کزا[3] در سال 1992 بر اساس تئوری داروین ارایه شد. این روش تکنیکی برای برنامه‌ریزی خودکار جهت حل مساله به صورت برنامه کامپیوتری ارایه می‌نماید و قابلیت بهینه سازی ساختار مدل و مولفه‌های آن را دارد. در برنامه‌ریزی بیان ژن از ساختار درختی استفاده می‌شود ولی الگوریتم ژنتیک بر پایه سیستم دودویی است که تفاوت اساسی ساختاری این دو سیستم محسوب می‌شود (Ferreira., 2004). در شکل 2 طرح کلی گام‌های مقدماتی برنامه‌ریزی بیان ژن نشان داده شده است.

فرآیند اجرایی گام به گام برنامه‌ریزی بیان ژن به این صورت است: 1- تولید یک جمعیت اولیه از فرمول‌ها که این فرمول‌ها از ترکیب تصادفی مجموعه توابع (عمل‌گرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند. 2- هر یک از افراد جمعیت مذکور با استفاده از توابع برازش مورد ارزیابی قرار می‌گیرند. 3- تولید یک جمعیت جدید از روابط(شکل3).

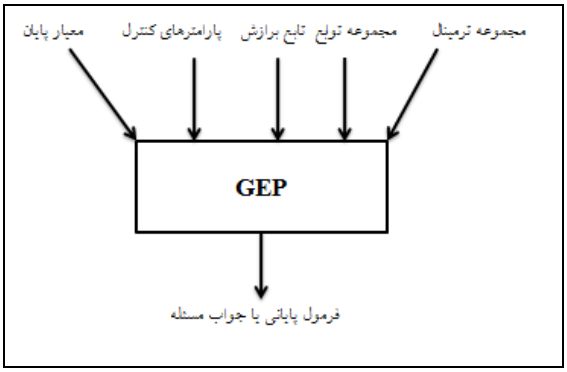
الگوی GEP اگر چه الگوی هوشمندی محسوب می‌شود، اما قابلیت در ارایه مدل ریاضی بین متغیر خروجی با متغیرهای ورودی، آن را از سایر الگوهای هوشمند متمایز نموده که این مساله در فرآیند

1- ANN:Artificial Neural Network
2- SVM: Support Vector Machine
3- Koza

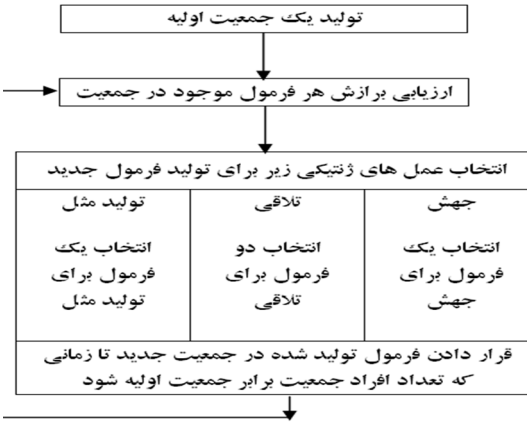
پیش‌بینی اهمیت بسزایی دارد. لازم به ذکر است که کلیه پارامترها با سعی و خطا تا دست‌یابی به بهترین نتیجه توسط کاربر انتخاب می‌شود. به طور خلاصه پارامترهای استفاده شده در هر بار اجرای مدل برنامه‌ریزی بیان ژن در جدول 1 ارایه شده است.



شکل 1- موقعیت جغرافیایی حوضه گالیکش



شکل 2- طرح کلی گام‌های مقدماتی برنامه‌ریزی بیان ژن



شکل 3- گام‌های اجرایی برنامه‌ریزی بیان ژن

جدول 1- مقادیر پارامترهای به‌کارگرفته شده در برنامه‌ریزی بیان ژن

تعداد کروموزوم‌ها	30	نرخ برگشت	0/1	نرخ ترکیب تک نقطه‌ای	0/3
طول هر راس	8	نرخ ترانهش ژنی	0/1	نرخ ترکیب دو نقطه‌ای	0/3
تعداد ژن‌ها	5	نرخ جهش	0/04	نرخ ترکیب ژنی	0/1

مدل درخت تصمیم M5

برای اولین بار کوایلان (1992)، مدل درخت تصمیم موسوم به M5 را برای پیش‌بینی داده‌های پیوسته ارایه نمود. این مدل، بر خلاف مدل‌های درخت تصمیم معمول که کلاس یا رده‌های گسسته را به عنوان خروجی ارایه می‌نمایند، یک مدل خطی چند متغیره را برای داده‌ها در هر گره از مدل درختی می‌سازد. تشکیل ساختار مدل-های درخت تصمیم‌گیری شامل مراحل ایجاد درخت و هرس کردن آن است (Quinlan., 1992; Witten and Frank., 2005). در مرحله ساختن درخت، از یک الگوریتم استنتاجی یا معیار تقسیم (انشعاب) برای تولید یک درخت تصمیم استفاده می‌شود. معیار تقسیم برای الگوریتم مدل M5، ارزیابی انحراف معیار مقادیر کلاسی است که به عنوان کمیتی از خطا به یک گره می‌رسد و کاهش مورد انتظار در این خطا را به عنوان نتیجه آزمون هر صفت در آن گره محاسبه می‌نماید.

کاهش انحراف معیار[1](SDR) از رابطه 1 بدست می‌آید:

$$SDR = sd(T) - \Sigma \frac{|T_i|}{|T|} sd(T_i) \quad (1)$$

که در آن، T بیان‌گر یک سری نمونه‌هایی است که به گره می-رسد، Ti: بیان‌گر نمونه‌هایی است که iامین خروجی سری پتانسیلی را دارند و sd: بیان‌گر انحراف معیار است. به دلیل فرآیند انشعاب، داده-های قرار گرفته در گره‌های فرزند، انحراف معیار کم‌تری نسبت به گره مادر داشته و بنابراین خالص‌تر هستند. پس از حداکثرسازی تمامی انشعاب‌های ممکن، M5 صفتی را انتخاب می‌کند که کاهش مورد انتظار را بیشینه نماید. این تقسیم بیش‌تر ساختار شبه درختی بزرگی را تشکیل می‌دهد که باعث بیش برازش می‌گردد. برای غلبه بر مساله بیش برازش، درخت تشکیل شده باید هرس شود. این کار با جایگزینی یک درخت فرعی با یک برگ انجام می شود. بنابراین مرحله دوم در طراحی مدل درختی شامل هرس نمودن درخت رشد یافته و جایگزینی درختان فرعی با توابع رگرسیونی خطی است. این تکنیک تولید مدل درختی، فضای پارامترهای ورودی را به نواحی یا زیرفضاهای کوچک‌تر تقسیم نموده و در هر کدام از آن‌ها، یک مدل رگرسیون خطی برازش می‌دهد. پس از هرس کردن، پروسه‌ی صاف کردن برای جبران گسیختگی‌های تند که به طور اجتناب ناپذیر بین مدل‌های خطی هم‌جوار در برگ‌های درخت هرس شده، مخصوصاً برای مدل‌هایی که از مقادیر کم‌تری از نمونه‌ها ساخته شده‌اند، انجام می‌شود (Bhattacharya and Solomatine., 2005). بعد از این‌که مدل خطی بدست آمد برای کمینه کردن خطای تخمین با حذف کردن پارامترها، ساده‌سازی مدل انجام می‌شود. در مدلM5 از یک جست‌وجوی حریصانه برای حذف متغیرهایی که مشارکت کمی در مدل دارند، استفاده می‌شود. البته گاهی اوقات همه متغیرها حذف شده و فقط یک مقدار ثابت باقی می‌ماند. در شکل 4 نحوه عملکرد مدل درخت تصمیم M5 برای یک مساله فرضی هر مدل نشان‌دهنده یک معادله رگرسیونی خطی می‌باشد. به عنوان مثال اگر X_1>2/5 و X_2>2 باشند، آن گاه مدل موسوم به فرم $Y= a_0+ a_1X_1+ a_2X_2$مورد استفاده قرار می‌گیرد.

شبکه عصبی مصنوعی

شبکه‌های عصبی مصنوعی در حقیقت مدل ساده شده‌ای از سیستم عصبی مرکزی می‌باشند و کارکردی شبیه به مغز انسان دارند. مغز به عنوان یک سیستم پردازش اطلاعات از عناصر اصلی ساختاری به نام نرون تشکیل شده است. اجتماعی از نرون‌های به هم مرتبط، بافت‌هایی که عصب نامیده می‌شوند را می‌سازند، که اطلاعات و پیام-ها را از یک قسمت بدن به قسمت دیگر بدن منتقل می‌کند (منهاج، 1393).

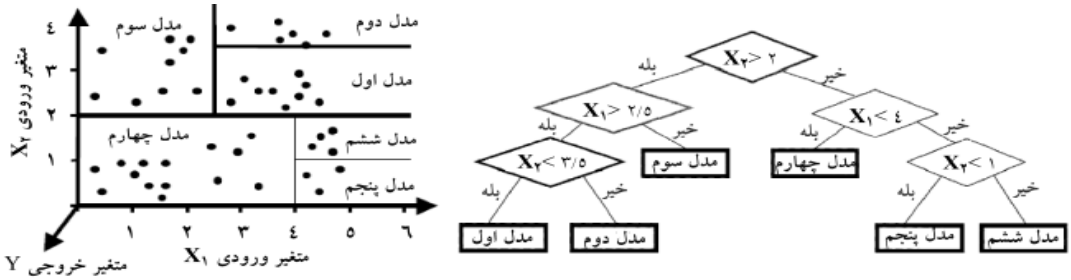
شبکه عصبی مصنوعی بر اساس یادگیری فرآیند حل مسئله و به عبارتی رسیدن به خروجی از طریق یافتن رابطه نهفته در فرآیند مورد نظر کار می‌کند.

بدین منظور الگو با دسته‌ای از داده‌ها آموزش داده شده تا در مورد ورودی‌های جدید با توجه به رابطه پیدا شده در مرحله آموزش، خروجی مناسب را محاسبه نماید. شبکه عصبی مصنوعی سعی می‌کند پس از دریافت ورودی‌ها، آن را به خروجی مطلوب برساند. این کار با استفاده از یک فعالیت انجام می‌شود. به این کار آموزش شبکه می-گویند. پس از انجام این کار، مدل شبکه عصبی خود را تست می‌کند و نهایتا مدل مورد تایید قرار می‌گیرد (شایان نژاد، 1385). در مدل شبکه عصبی از مدل شبکه‌های عصبی پرسپترون چندلایه[2] و از میان روش‌های مختلف آموزش به روش الگوریتم پس انتشار خطا[3]، به دلیل همگرایی سریع‌تر در آموزش شبکه استفاده شد.

1- Standard Deviation Reduction

2 -Multi Layer Perceotron

3 -Back Propagation Algorithm



شکل 4- عملکرد مدل درختی تصمیم M5. الف) تقسیم فضای پارامترهای ورودی (X1*X2) به 6 ناحیه. ب) بیان معیار تقسیم فضای پارامترهای ورودی به صورت درختی (Solomatine and Xue., 2004).

تبیین مدل‌ها

بررسی نتایج پژوهش‌های پیشین در ارتباط با پیش‌بینی جریان روزانه حاکی از آن است که بهترین نتایج از میان مدل‌های مختلف تعریف شده از داده‌های ورودی، تا پنج تاخیر زمانی حاصل شده است. در تحقیق حاضر تاثیر توالی دبی و بارش روزهای قبل در پیش‌بینی دبی روز مدنظر بوده، بنابراین از داده‌های دبی جریان و بارش‌های با توالی برگشتی تا 5 روز به عنوان داده‌های آموزشی به صورت ترکیب-های مختلف مطابق جدول 2 استفاده شده است که از این پس از آن-ها به عنوان الگو نام برده خواهد شد. الگوهای پیشنهادی در این تحقیق عبارتند از:

جدول 2- ترکیب ورودی مدل‌های M5،GEP و ANN

الگو	داده‌های ورودی
1	Q_{t-1},P_{t-1},P_t
2	$Q_{t-1},Q_{t-2},P_{t-1},P_{t-2},P_t$
3	$Q_{t-1},Q_{t-2},Q_{t-3},P_{t-1},P_{t-2},P_{t-3},P_t$
4	$Q_{t-1},Q_{t-2},Q_{t-3},Q_{t-4},P_{t-1},P_{t-2},P_{t-3},P_{t-4},P_t$
5	$Q_{t-1},Q_{t-2},Q_{t-3},Q_{t-4},Q_{t-5},P_{t-1},P_{t-2},P_{t-3},P_{t-4},P_{t-5},P_t$

که در آن Q_t بیان‌گر دبی جریان روز و Q_{t-1}، Q_{t-2}، P_{t-1} و P_{t-2} به ترتیب دبی جریان و بارش در یک روز قبل و دو روز قبل می‌باشد و به همین ترتیب الگوهای ورودی تا 5 روز قبل در نظرگرفته شده است.

معیارهای ارزیابی

لگیتز و مک کیب، بیان کردند که شاخص ضریب همبستگی به تنهایی به عنوان معیاری جهت عملکرد مدل مورد استفاده قرار نمی-گیرد (Legates and McCabe., 1999). بنابراین به منظور سنجش کارایی مدل‌های مورد بحث در برآورد پارامتر خروجی مدل‌سازی، از چندین شاخص آماری استفاده شد.

ضریب همبستگی

$$r=\frac{\sum_{i=1}^{N}(Qobs-\overline{Q}_{obs})(Qsim-\overline{Q}_{sim})}{\sqrt{\sum_{i=1}^{N}(Qobs-\overline{Q}_{obs})^2\sum_{i=1}^{N}(Qsim-\overline{Q}_{sim})^2}} \quad (2)$$

ریشه میانگین مربعات خطا

$$RMSE=\sqrt{\frac{\sum_{i=1}^{N}(Q_{sim}-Q_{obs})^2}{N}} \quad (3)$$

میانگین خطای اریب

$$MBE=\frac{1}{n}\sum_{i=1}^{N}(Q_{sim}-Q_{obs}) \quad (4)$$

ناش-ساتکلیف

$$NS=1-\frac{\sum_{i=1}^{n}(Q_{sim}-Q_{obs})^2}{\sum_{i=1}^{n}(Q_{obs}-\overline{Q}_{obs})^2} \quad (5)$$

در روابط فوق Q_{sim} داده‌های شبیه‌سازی شده، Q_{obs} داده‌های مشاهداتی ، $\overline{Q}_{obs}$ و $\overline{Q}_{sim}$ بترتیب‌میانگین داده‌های مشاهداتی و محاسباتی و n تعداد مشاهدات است.مقدار RMSE بیان‌گر متوسط مربع خطا بین مقادیر مشاهداتی و پیش‌بینی شده است. این معیار بزرگی خطا را نشان می‌دهد و هرچه مقدار آن کم‌تر باشد نشان‌دهنده دقت بیش‌تر تخمین مدل است.

نتایج و بحث

پیش‌بینی جریان رودخانه‌ها به لحاظ اهمیت فراوانی که در بهره-برداری از منابع آب دارد، از دیرباز مورد توجه مهندسان علوم مهندسی منابع آب و هیدرولیک قرار گرفته است. در این تحقیق از داده‌های دبی و بارش روزانه حوضه آبریز گالیکش به منظور مدل‌سازی جریان با مدل های GEP و M5 و ANN استفاده شد. نتایج بدست آمده از هر سه مدل مورد تحلیل قرار گرفت.

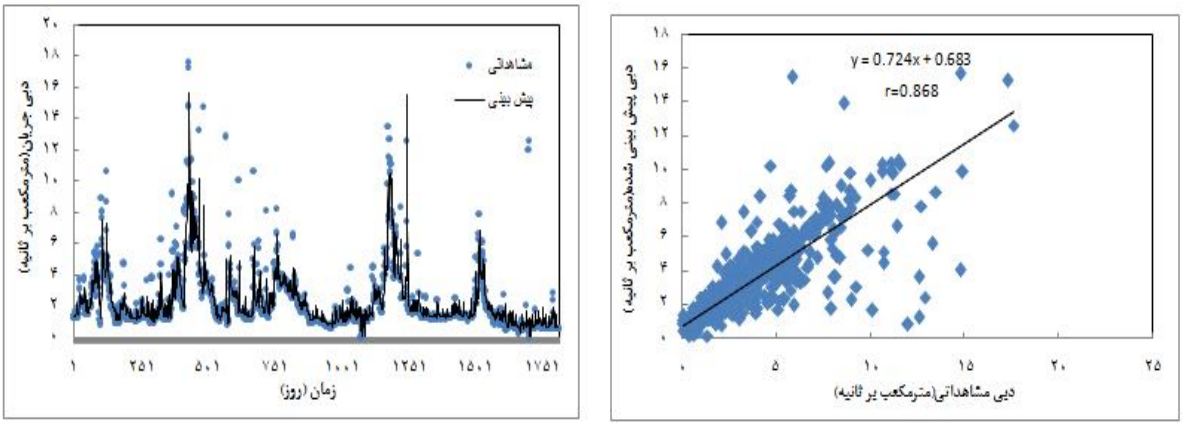
پیش‌بینی دبی با استفاده از مدل درختی برنامه‌ریزی بیان ژن

با توجه به جدول 3 که نتایج مدل GEP را برای ترکیب‌های مختلف ورودی در مدل‌سازی جریان نشان می‌دهد، ملاحظه می‌شود که تمامی مدل‌ها عملکرد مناسبی در برآورد جریان روزانه دارند و تاخیرهای زمانی به آن صورت در دقت مدل اثرگذار نمی‌باشد. الگوی ورودی 4 که دبی جریان را تا چهار روز قبل مدنظر قرار می‌دهد با ضریب همبستگی (0/868) و مقدار جذر میانگین مربعات خطا (1/1082)، در مرحله آزمون به عنوان بهترین مدل انتخاب شد. پراکنش داده‌های مشاهداتی و شبیه‌سازی شده برای بهترین الگوی GEP (الگوی چهارم) در شکل 5 (الف) نشان داده شده است. معادله برازش داده شده بر این داده‌ها نیز در شکل 5 (الف) نشان داده شده است. شیب خط برازش داده شده بین داده‌ها همواره کم‌تر از یک می‌باشد و این نشان می‌دهد که GEP عموما داده‌ها را کم‌تر از داده-های مشاهده شده شبیه‌سازی و پیش‌بینی می‌کند. شکل 5 (ب)، مقادیر دبی‌های محاسباتی(برای هر روز) را با مقادیر دبی مشاهداتی همان روز مقایسه می‌کند. معادله ارایه شده توسط مدل برنامه‌ریزی بیان ژن برای پیش‌بینی دبی جریان روزانه به ازای ورودی‌های مدل در رابطه 8 ارایه شده است.

$$Q=Q_{t-1}+(((Q_{t-4}/6.9)^{(1/9)}+((p/5.5)^{(0.5)}))+((Q_{t-1})/(\exp(-5.2)+(-5.2)) \quad (6)$$

جدول 3- مشخصات آماری نتایج مدل درختی GEP

الگوی ورودی	آموزش			آزمون		
	RMSE	r	MBE	RMSE	R	MBE
1	1/088	0/922	0/0009	1/089	0/860	0/042
2	1/109	0/925	-0/0195	1/123	0/862	-0/002
3	1/118	0/916	-0/0123	1/106	0/854	0/044
4	1/074	0/927	0/0005	1/1082	0/868	0/035
5	1/183	0/911	-0/00012	1/192	0/837	0/048



(الف) (ب)

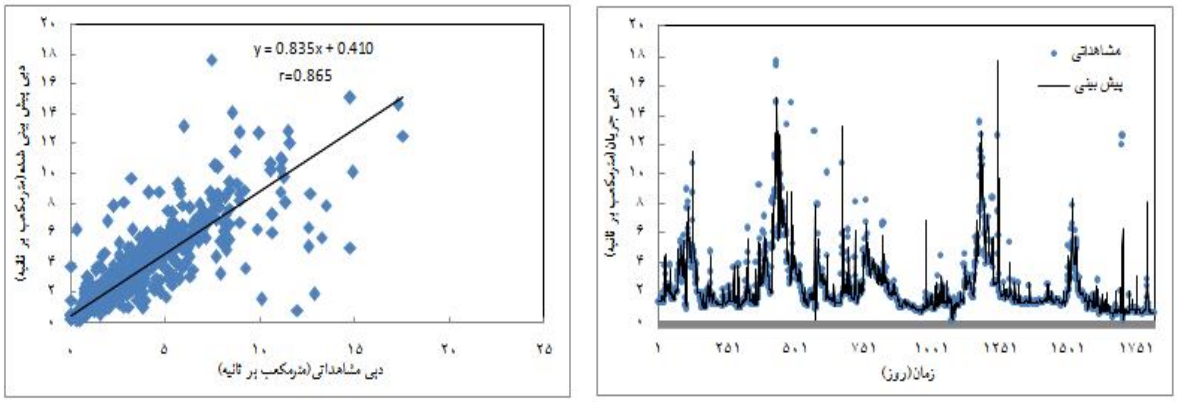
شکل 5- الف) نمودار پراکنش مقادیر دبی مشاهداتی و پیش بینی شده در بهترین مدل GEP، ب)آنمود مقادیر دبی مشاهداتی و پیش بینی شده در بهترین مدلGEP

پیش‌بینی دبی با استفاده از مدل درختی M5

با توجه به نتایج بدست آمده از مدل M5که در جدول 5 آمده است، ملاحظه می‌شود که این روش نیز عملکرد مناسبی در برآورد دبی جریان حوضه آبریز گالیکش داشته است. الگوی ورودی 3 که دبی جریان را تا سه روز قبل مدنظر قرار می‌دهد با ضریب همبستگی (0/865) و مقدار جذر میانگین مربعات خطا (1/059) در مرحله آزمون به عنوان بهترین مدل انتخاب شد. پراکنش داده‌های مشاهداتی و شبیه‌سازی شده برای بهترین الگوی M5 (الگوی سوم) در شکل 6 (الف) نشان داده شده است. شکل 6 (ب) مقادیر دبی‌های محاسباتی (برای هر روز) را با مقادیر دبی مشاهداتی همان روز مقایسه می‌کند.

جدول 4- مشخصات آماری نتایج مدل درختیM5

الگوی ورودی	آموزش			آزمون		
	RMSE	r	MBE	RMSE	R	MBE
1	0/9983	0/935	-0/001	1/104	0/856	0/044
2	0/9025	0/942	-0/002	1/101	0/862	0/036
3	0/8824	0/949	0/000	1/059	0/865	0/022
4	1/0247	0/932	-0/003	1/062	0/852	0/027
5	1/0035	0/934	-0/0003	1/068	0/851	0/031



(ب) (الف)

شکل 6-الف) نمودار پراکنش مقادیر دبی مشاهداتی و پیش بینی شده در بهترین مدلM5، ب)آبنمود مقادیر دبی مشاهداتی و پیش بینی شده در بهترین مدلM5

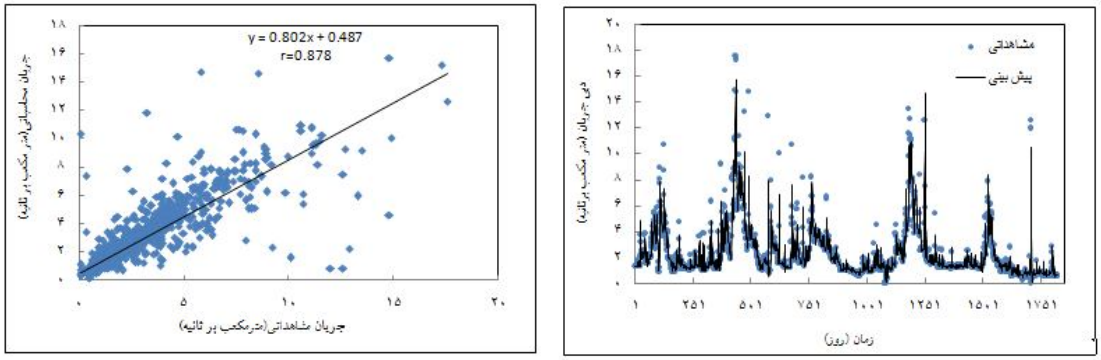
پیش‌بینی دبی با استفاده از مدل شبکه عصبی مصنوعی

با توجه به نتایج بدست آمده از مدل ANN که در جدول 5 آمده است، به لحاظ دبی و بارش مربوط به سه روز قبل، الگوی سوم با جذر میانگین مربعات خطای (1/1082) مترمکعب بر ثانیه و ضریب همبستگی(0/878) نتایج قابل قبولی داشته است. همچنین اجرای ANN با الگوی ورودی چهارم و پنجم که دبی جریان روزانه را به ترتیب تا چهار و پنج روز قبل وابسته می‌سازد، نشان می‌دهد که نتایج از روز سوم به بعد رفته

رفته ضعیف‌تر شده و توالی بیش از حد رو به عقب در پیش‌بینی دبی روزانه تاثیر منفی می‌گذارد. به عبارت دیگر اگرچه از روز سوم به بعد تعداد روزهای دخیل (و به تبع آن پارامترهای دخیل) در پیش‌بینی جریان افزایش می‌یابد ولی این افزایش صرفا بر پیچیدگی الگو افزوده و نتایج پیش‌بینی را رفته رفته ضعیف‌تر می‌سازد. اگرچه تغییرات ضریب همبستگی در الگوها بسیار ناچیز بوده و نمی‌توان به صورت قاطع تصمیم گرفت ولی بدلیل پایین بودن ضریب RMSE در الگوی ورودی شماره 3 این الگو به عنوان بهترین الگو انتخاب شد. پراکنش داده‌های مشاهداتی و شبیه‌سازی شده برای بهترین الگوی ANN (الگوی سوم) در شکل 7 (الف) نشان داده شده است. شکل 7 (ب) مقادیر دبی‌های محاسباتی (برای هر روز) را با مقادیر دبی مشاهداتی همان روز مقایسه می‌کند.

جدول 5- مشخصات آماری نتایج مدلANN

الگوی ورودی	نرون پنهان	واسنجی			صحت سنجی		
		RMSE	r	MBE	RMSE	R	MBE
1	4	0/996	0/938	0/181	1/151	0/854	0/219
2	5	0/962	0/942	0/219	1/102	0/865	0/248
3	4	0/921	0/948	0/021	1/082	0/878	0/018
4	9	0/931	0/935	0/025	1/125	0/847	0/022
5	11	0/926	0/928	-0/021	1/138	0/839	-0/027



(ب) (الف)

شکل 7-الف) نمودار پراکنش مقادیر دبی مشاهداتی و پیش بینی شده در بهترین مدلANN، ب)آبنمود مقادیر دبی مشاهداتی و پیش بینی شده در بهترین مدلANN

آزمون مقایسه میانگین‌ها

برای ارزیابی میزان هم‌قوارگی یا یکسان بودن و نبودن میانگین، نمونه‌ای با میانگین جامعه در حالتی بکار می‌رود که انحراف معیار جامعه مجهول باشد. چون توزیع t در مورد نمونه‌های کوچک با استفاده از درجات آزادی تعدیل می‌شود، می‌توان از این آزمون برای نمونه‌های بسیار کوچک استفاده نمود. همچنین این آزمون مواقعی که خطای استاندارد جامعه نامعلوم و خطای استاندارد نمونه معلوم باشد، کاربرد دارد. برای بکار بردن این آزمون، متغیر مورد مطالعه باید در مقیاس فاصله‌ای باشد، شکل توزیع آن نرمال باشد. نتایج آماری حاصل از SPSS برای مقایسه ANN، GEP و M5 در جدول 6 آمده است.

با توجه به نتایج بدست آمده در مقایسه ANN با GEP قدرمطلق مقدار t محاسبه شده کوچک‌تر از مقدار t بحرانی (1/96) در سطح اطمینان 95 درصد می‌باشد، پس فرض برابری میانگین‌ها به احتمال 95 درصد مورد تایید نیست. مفهوم دیگر این است که اختلاف روش ANN با GEP معنی‌دار است اما در مقایسه ANN با M5 و GEP با M5 قدرمطلق مقدار t محاسبه شده بزرگ‌تر از مقدار t بحرانی در سطح اطمینان 95 درصد می‌باشد، پس فرض برابری میانگین‌ها به احتمال 95 درصد مورد تایید است. مفهوم دیگر این است که اختلاف روش ANNبا M5 و GEP با M5 نتوانسته آن‌چنان تاثیرگذار باشد و اختلاف آن‌ها معنی‌دار نیست.

جدول 6- نتایج t تست برای نمونه‌های جفت شده

	درجه آزادی	آماره آزمون	سطح معنی داری	اختلاف میانگین	خطای معیار
ANN-GEP	1821	-1/803	0/072	0/01135	0/48421
ANN-M5	1821	7/027	0/00	0/01020	0/43511
GEP-M5	1821	4/197	0/00	0/01220	0/52055

جدول 7- معیارهای ارزیابی مدل‌ها در دو مرحله واسنجی و صحت سنجی

	r	RMSE	MBE	NS	NSL	NSH
واسنجی(1370-1363)						
ANN	0/948	0/921	0/021	0/893	0/899	0/872
GEP	0/927	1/074	0/0005	0/858	0/646	0/792
M5	0/949	0/882	0/000	0/902	0/917	0/872
صحت‌سنجی(1386-1382)						
ANN	0/878	1/082	0/018	0/749	0/810	0/755
GEP	0/868	1/108	0/035	0/732	0/659	0/687
M5	0/865	1/059	0/022	0/763	0/791	0/721

نتیجه‌گیری

شاخص‌های آماری مربوط به نتایج حاصل از کاربرد روش‌های برنامه‌ریزی بیان ژن، مدل درختی M5 و مدل ANN در بررسی تاثیر توالی دبی و بارش روزانه در پیش‌بینی دبی در جداول 3، 4 و 5 ارایه شده است. در مقایسه روش‌ها در تمامی الگوها روش M5 بدلیل داشتن ضریب همبستگی بالاتر و مقدار کم‌تر در جذر میانگین مربعات خطا از دقت بالاتری برخوردار است. این موضوع در شکل‌های 5، 6 و 7 با مقایسه نتایج

حاصل از پیش‌بینی جریان رودخانه با هر سه روش در مقایسه با جریان روزانه مشاهده شده در دوره صحت‌سنجی نشان داده شده است. با توجه به نتایج بدست آمده از سه مدل به وضوح روشن است که مدل‌ها توانایی خوبی در برآورد جریان روزانه حوضه آبریز گالیکش دارند. هر چند نتایج شاخص‌های آماری حاکی از برتری نسبی مدلM5 نسبت به GEP و ANN است اما استفاده از هر سه مدل به منظور مدل‌سازی جریان توصیه می‌شود. در هر دو دوره واسنجی و صحت‌سنجی هر سه مدل جریان بیش‌تری نسبت به مقادیر مشاهداتی پیش‌بینی کرده‌اند.

منابع

دانند مهر،ع و مجدزاده طباطبایی،م.ر. 1388. بررسی تاثیر توالی دبی روزانه در پیش بینی جریان رودخانه‌ها با استفاده از برنامه‌ریزی ژنتیک. نشریه آب و خاک (علوم و صنایع کشاورزی). 24. 2: 325-333

شایان نژاد،م. 1385. مقایسه دقت روش‌های شبکه‌های عصبی مصنوعی و پنمن-مانتیس در محاسبه تبخیر و تعرق پتانسیل. اولین همایش ملی مدیریت شبکه‌های آبیاری و زهکشی، دانشگاه شهید چمران اهواز.

ظهیری،ع و قربانی،خ. 1391. شبیه‌سازی دبی جریان در مقاطع مرکب به کمک مدل درخت تصمیم M5. مجله پژوهش‌های حفاظت آب و خاک. 20، 3: 113-132

فلاحی،م.ر.، وروانی،ه و گلیان،س. 1390. پیش‌بینی بارش با استفاده از مدل رگرسیون درختی به منظور کنترل سیل. پنجمین کنفرانس سراسری آبخیزداری و مدیریت منابع آب و خاک کشور. انجمن مهندسی آبیاری و آب ایران. کرمان.

منهاج،م.ب. 1393. مبانی شبکه‌های عصبی. انشارات دانشگاه صنعتی امیرکبیر. 718 صفحه.

Bhattacharya,B and Solomatine,D.P. 2005. Neural networks and M5 model trees in water level-discharge relationship. J. Neurocomputing.63, 381-396.

Ferreira,C. 2001. Gene expression programming a new adaptive algorithm for solving problems. Complex Systems. 13:2.87–129

Ferreira,C. 2004. Gene Expression Programming and the Evolution of Computer Programs. Recent Developments in Biologically Inspired Computing, pages 82-103, Idea Group Publishing.

Fernando,K.A., Shamseldin,A.Y., Abrahart,R.J. 2012. River Flow Forecsting Using Gene Expression Programming Models. 10th International Conference on Hydro informatics, HIC 2012, Hamburg, Germany.

Koza,J.R. 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: The MIT Press.

Karayiannis,N. and Venetsanopoulos,A.N. 1993. Artifical Neural Network: Learning Alogorithms, Performance Evaluation, and Application. Kluwer Academic Publisher, Boston. 440p.

Liong,S.Y., Gautam,T.R., Khu,S.T., Babovic,V., Keijzer,M and Muttil,N. 2002. Genetic programming, A new paradigm in rainfall runoff modeling. Journal of the American Water Resources Association. 38.3: 705-718.

Londhe,S.N and Dixit,P.R. 2010. Forecasting Stream Flow Using Model Trees. International Journal of Earth Sciences and Engineering. 4.6: 282-285.

Legates,D.R and McCabe,G.J. 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydro climatic model validation. Water Resources Research. 35:1.233–241.

Quinlan,J.R. 1992. Learning with continuous classes. P 343-348, In: Proceedings of Fifth Australian joint conference on artificial intelligence, Singapore.

Sattari,M.T., Pal,M., Apaydin,H.,Ozturk,F. 2013. M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey, Water Resources. 40.3: 223-242.

Solomatine,D.P., Xue,Y. 2004. M5 model trees and neural networks: Application to flood Forecasting in the upper reach of the Huai river in China. Journal of Hydrologic Engineering. 9. 6: 491-501.

Witten,I.H and Frank,E. 2005. Data mining: practical machine learning toolsand techniques with Java implementations. Morgan Kaufmann: San Francisco.

Whigham,P.A and Crapper,P.F. 2001. Modeling rainfall runoff using Genetic programming. Mathematical and Computer Modeling. 33: 707-721.

Iranian Journal of Irrigation and Drainage
No. 4, Vol. 10, Sept.-Oct. 2016, p. 479-488

نشریه آبیاری و زهکشی ایران
شماره 4، جلد 10، مهر - آبان 1395، ص. 488-479

Prediction of Daily Stream-flow Using Data Driven Models

M. Salarijazi[1]*, K. Ghorbani[2], E. Sohrabian[3], M. Abdolhosseini[4]
Recived: May.30, 2016 Accepted: Oct.19, 2016

Abstract

Accurate prediction of river daily discharge is a suitable tool for water resources planning and management. Using models that present explicit equation, such as M5 model trees and Genetic expression programming, causes increase efficiency of these models. In this study, the Galikesh basin as one of most flood prone basins in Gloestan Province is considered for the prediction of river daily discharge. Data series used in this study are long term 26 years daily rainfall and river discharge series belong to Galikesh meteorology and hydrometry station. Daily rainfall and river discharge data from 1 to 5 days ahead are used as inputs for prediction by M5 model trees, genetic expression programming and artificial neural network models. The results indicate very good efficiency of the investigated models beside overestimation of the models to predict daily river discharge. Comparison of results of different models leads to selection of M5 model trees as best model among investigated models.

Keywords: Artificial neural network, Data driven models, Galikesh basin, Genetic expression programming, River discharge prediction, M5 model trees

1- Assistant Professor, Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources
2- Associate Professor, Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources
3- Graduated Student of M.Sc. of Water Resources Engineering, Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources
4- Assistant Professor, Department of Water Engineering, Gorgan University of Agricultural Sciences and Natural Resources
(*- Corresponding Author Email: meysam.salarijazi@gau.ac.ir)