پژوهش‌های آبخیزداری
(پژوهش و سازندگی)
دوره ۳۰، شـماره ۴، شماره پیاپی ۱۱۷، زمستان ۱۳۹۶، صفحات ۵۸-۴۷
شناسه‌ی دیجیتال: 10.22092/WMEJ.2018.117085

مقایسه‌ی دو مدل ناپارامتری کا-نزدیک‌ترین همسایه و درخت تصمیم ام۵ در پیش‌بینی آب‌دهی رود در حوزه‌ی آب‌خیز کرج

صفورا عرب
کارشناس ارشد مهندسی منابع آب دانشکده‌ی کشاورزی دانشگاه بیرجند
عباس خاشعی سیوکی*
(نویسنده‌ی مسئول)* دانشیار گروه علوم و مهندسی آب دانشکده‌ی کشاورزی دانشگاه بیرجند
محسن پوررضابیلندی
استادیار گروه علوم و مهندسی آب دانشکده‌ی کشاورزی دانشگاه بیرجند
سیدرضا هاشمی
استادیار گروه علوم و مهندسی آب دانشکده‌ی کشاورزی دانشگاه بیرجند

تاریخ دریافت: تیر ۹۶ تاریخ پذیرش: اردیبهشت ۹۷

* Corresponding Email:(abbaskhashei@birjand.ac.ir)

چکیده
اهمیت برنامه‌ریزی و مدیریت منابع آب، رشد روزافزون جمعیت و محدودیت منابع آب سطحی در کشور پیشبینی دقیق‌تر جریان رودها را با استفاده از ابزارها و روش‌های نوین مدل‌سازی به ضرورتی اجتناب‌ناپذیر تبدیل کرده است. برای پیش‌بینی جریان رودها در سال‌های گذشته روش‌های مختلفی ابداع شده است، و یکی از آن‌ها مدل‌های مبتنی برداده است. در این پژوهش با استفاده از دو مدل مبتنی برداده (ام۵، کاان‌ان) و داده‌های آب‌وهواشناسی (آب‌دهی، بارش، دما و تبخیر) در دوره‌ی ۱۳۸۸– ۱۳۸۱ آب‌دهی رود در حوزه‌ی آب‌خیز کرج پیش‌بینی، و کارآیی و دقت آنها بررسی و مقایسه شد. برای تعیین داده‌های مورد نیاز در آموزش و انتخاب ترکیب‌های بهینه از روش نوین آزمون گاما استفاده شد. ترکیب‌های مناسب برای ورودی مدل تعیین، و به دو مدل داده‌محور ام۵ و کاان‌ان وارد شد. نتایج نشان داد که در هر دو مدل به‌کاررفته برای ترکیب‌هایی که پارامتر آب‌دهی در آن بوده است پیش‌بینی دقیقی به‌دست آمد. علاوه بر این، مدل ام۵ دقت بیش‌تری از مدل کاان‌ان دارد، به‌طوری‌که آر۲ برای مدل‌های ام۵ و کاان‌ان به‌ترتیب ۰/۹۷۳۸، ۰/۹۳۴۵، مقدار آرام‌اس‌ای آنها به‌ترتیب ۰/۵۴۶۸، ۰/۸۶۷۶، و مقدار کاجی‌ای آنها ۰/۹۸۵۵ و ۰/۹۶۳۶ بود.

واژه‌های کلیدی: آب‌دهی رود، آزمون گاما، حوزه‌ی آب‌خیز کرج، درخت تصمیم ام۵، شاخص گوپتا، کا- نزدیک‌ترین همسایه

Comparison of Two Nonparametric Models, K- nearest neighbor and M5 Decision Tree in Forecasting the River Discharge in the Karaj Catchment

Safora Arab
M.Sc. Water Engineering Department, University of Birjand
Abbas Khashei-Siuki*
*(Corresponding Author) Associate Professor, Water Engineering Department, University of Birjand (abbaskhashei@birjand.ac.ir)
Mohsen Pourreza-Bilondi
Assistant Professor, Water Engineering Department, University of Birjand
SeyedReza Hashemi
Assistant Professor, Water Engineering Department, University of Birjand

Abstract
The importance of water resources planning and management, the fast growing population, and the limited surface water resources, have made the application of the new technology to forecasting of river flow. A necessity, various methods have been presented in recent years to forecast the river flow, and the data-based models are considered the most reliable for this purpose. The river flow in the Karaj Catchment has been simulated using the data based models (KNN and M5). Hydroclimatological data (discharge, precipitation, temperature and evaporation) for the 2002 to 2009 duration have been collected to carry out the simulation processes. The performance and accuracy of the models were examined and compared. The Gamma test was used to select appropriate compositions. Suitable compositions were determined as the model inputs (KNN and M5). These features were entered in to the two data-based models. Results showed that both models simulated reliable flow predictions, if the discharge had been entered as an input. The M5 model showed a better precision as compared with the KNN model. The Coefficient of determination (R^2) for the KNN and M5 models were 0.97 and 0.93, respectively. The RMSE were 0.55 and 0.87, for the same two models, respectively, and the value of the KGE were 0.99, 0.96, respectively.

Keywords: Decision tree M5, Gamma test, Gupta Index, k- nearest neighbor, river discharge

مقدمه

پیش‌بینی دقیق و به‌موقع منابع آب دردسترس را می‌توان نکته‌ی کلیدی در برنامه‌ریزی، مدیریت و بهره‌وری بهینه از منابع آب دانست. براین پایه، دست‌یابی به روش‌های مطمئن پیش‌بینی جریان رودها برای برنامه‌ریزی در بهره‌برداری به‌موقع از منابع آب از اهمیت روزافزونی دارد. در برنامه‌ریزی و مدیریت منابع آب مدل‌های پیش‌بینی جریان رود ابزاری کارآمد است، بنابراین، بررسی روش‌های نوین و تعیین الگوریتم‌های مناسب برای پیش‌بینی و مدل‌سازی اندازه‌ی آب دهی رودها ضروری است (فتاحی‌نفچی و همکاران ۲۰۱۲). مدل، درک ساده‌یی از یک سامانه‌ی واقعی است (مانند فرآیند پیچیده ی تبدیل بارش به روان‌آب). می‌توان مدل را نظریه، قانون یا اندیشه‌یی ساختاری دانست. با توسعه‌ی مدل امکان توضیح ساده‌ی سامانه‌یی پیچیده فراهم می‌شود (حامد انسانیت ۲۰۱۲).

عوامل بسیاری بر پدیده‌های آب‌شناختی مؤثراند که تعیین میزان تأثیرشان نامشخص است. دراین‌صورت تعیین بهترین ترکیب از عوامل ورودی به مدل‌های هوشمند وقت‌گیر است. کاربرد روش‌های نوین در این راستا از مراحل سعی‌وخطا می‌کاهد و به روند مدل‌سازی سرعت می‌بخشد. در این تحقیق از روش آزمون گاما در مرحله‌ی پیش‌پردازش مدل استفاده شد. نخستین بار کانسکر (۱۹۹۷) این آزمون را به‌صورت خلاصه معرفی کرد، و در سال‌های بعد دورانت (۲۰۰۱) و جونس و همکاران (۲۰۱۳) آن‌را با جزئیات بررسی کردند (رمسان و همکاران ۲۰۰۸). این روش ابزاری برای مدل‌سازی غیرخطی است که به‌کمک

آن می‌توان قبل از ایجاد مدل، ترکیب مناسب از عوامل ورودی را برای مدل‌سازی داده‌های خروجی و ایجاد مدلی هموار، بررسی نمود، به‌طوری‌که با بررسی خطا در هر ترکیب ورودی، معیاری مناسب برای قضاوت بر آن‌ها فراهم می‌شود. درنهایت ترکیبی که برای مدل‌سازی بهینه‌ی خروجی کم‌ترین مقدار خطا دارد، ترکیب بهینه‌ی ورودی می‌شود (مقدم‌نیا و همکاران ۲۰۰۹). آزمون گاما روشی غیرخطی و پیش‌رفته برای تحلیل داده‌ها است، که می‌تواند داده‌ها را برای ایجاد مدلی خطی تجزیه‌وتحلیل کند. در روند شناسایی مدل برای تجزیه‌وتحلیل داده‌ها، از این آزمون برای توسعه‌ی مدل شبکه‌ی عصبی استفاده شد (لین و همکاران ۲۰۰۸). از پیش پردازش آزمون گاما در مدل حداقل مربعات ماشین بردار پشتیبان و سامانه استنتاج تطبیقی عصبی- فازی[1] و شبکه‌ی عصبی مصنوعی[2] استفاده شد. نتایج نشان دادند که جی‌تی-ال‌اس‌اس‌وی‌ام دقت زیادی دارد و مراحل سعی‌وخطا در مدل‌سازی کم شده است (سیفی و ریاحی‌مدوار ۲۰۱۲).

مدل ام۵ مشابه توابع خطی جداشده[3] است، که ترکیبی است از مدل‌های رگرسیون خطی و رگرسیون درختی و کاربردهای زیادی در علوم مختلف دارد. برآورد تبخیر از سطح آزاد با کمک مدل درختی ام۵ انجام شد. نتایج نشان دادند که دو عامل میانگین روزانه‌ی دمای هوا و سرعت باد بیش‌ترین تأثیر را بر برآورد تبخیر از سطح آزاد دارند، و با داشتن این دو عامل، می‌توان با کمک مدل درختی ام۵، تبخیر از سطح آزاد را با دقتی پذیرفتنی تخمین زد (امامی و علی‌پور ۲۰۱۱). بار رسوب بستر با استفاده از روش درخت تصمیم‌گیری رگرسیونی و مقایسه با روش‌های تجربی برآورد شد. نتایج نشان دادند که درخت تصمیم در برآورد مواد معلق دقت بیش‌تری دارد (هجی بخش ۲۰۱۰). کارآیی مدل درختان تصمیم‌گیری در برآورد مواد معلق رودی را در حوزه‌ی آب‌خیز باباامان خراسان شمالی بررسی شد. نتایج نشان داد که به‌طورکلی و براساس معیارهای آماری استفاده‌شده (آر، آرام‌اس‌ای)، مدل درخت تصمیم‌گیری در مقایسه با روش منحنی سنجه‌ی رسوب تطابق بسیار بیش‌تری با مقادیر اندازه‌گیری شده داشته است و دقت و کارآیی بیشتری دارد (عشقی وهمکاران ۲۰۱۴). در حوزه‌ی کاسل هند، تجمع رسوب معلق با استفاده از شبکه‌ی عصبی مصنوعی، منطق فازی و الگوریتم درخت تصمیم مدل‌سازی شد. نتایج نشان دادندکه مدل درخت تصمیم ام۵ دقت و درستی بیش‌تری از شبکه‌ی عصبی و منطق فازی دارد (سنتیل کومار و همکاران ۲۰۱۲). طبقات خشک‌سالی در حوزه‌ی شکرک ترکیه با روش درخت تصمیم ارزیابی و پیش‌بینی شد. نتایج این روش با روش طبقه‌بندی و تحلیل خشک سالی بر اساس شاخص بارش معیار[4] مقایسه شدند و نشان داد که میان این دو روش تفاوت زیادی در طبقه‌بندی خشکسالی نیست (یورکلی ۲۰۱۲) برای پیش‌بینی جریان در آب راه‌های مرکب، دو روش برنامه‌نویسی ژنتیک خطی و مدل درخت ام۵ مقایسه شدند. نتایج نشان داد که دقت مدل ام۵ به دقت مدل ال‌جی‌پی نزدیک است؛ ازاین‌رو می‌توان برای پیش بینی جریان رود در رویدادهای سیل از مدل ام۵ استفاده کرد (ظهیری و عظمت الله ۲۰۱۲)

روش کا-نزدیک‌ترین همسایه، زیرمجموعه‌ی روش‌های ناپارامتری خودگردان‌ساز است. این نام به‌دلیل روند باز-نمونه‌گیری با جای گزینی از مجموعه‌ی مشاهده، داده شده‌است. پژوهش‌های زیادی با این روش انجام شده‌است. عمل‌کرد هیدرولیکی دستگاه های آبیاری تحت فشار با استفاده از مدل‌های شبکه‌ی عصبی مصنوعی و رگرسیون ناپارامتری ارزیابی شدند. نتایج نشان دادند که مدل شبکه‌ی عصبی مصنوعی ام‌ال‌پی می‌تواند ضریب یکنواختی لوله‌ی فرعی‌های آبیاری تحت فشار را با استفاده از مشخصات هیدرولیکی و فیزیکی لوله‌ی فرعی با کم ترین خطا برآورد نماید .روش کا-نزدیک‌ترین همسایه میانگین اندازه‌های سی‌یو را بادقت مدل‌سازی می‌کند، اما نمی‌تواند این کار را با دقت برای انحراف معیار این مقادیر انجام دهد، و خطای آن در مرحله‌ی آزمون بسیار بیشتر از دو مدل دیگر برآورد شد (اباباپی وهمکاران ۲۰۱۳). نشان داده شده است که نزدیک‌ترین همسایه روش مناسبی برای تعیین سطح هشدار (ضریب آستانه) و پیش‌بینی احتمال روی‌داد بهمن در محورهای بهمن‌خیز استان تهران است (سلیمانی مطلق وهمکاران ۲۰۱۱). نتایج تهیه‌ی نقشه‌های کاربری با روش کاان‌ان و مجموعه‌ی داده‌ی آموزشی ۵۳۸ قطعه زمین در سراسر کانادا و ۱۸ متغییر پیش‌بینی‌کننده‌ی محیطی نشان‌دهنده‌ی برتری روش کاان‌ان از روش‌های قبلی خاک بود (مانسی و همکاران ۲۰۱۴). برای ریزمقیاس‌نمایی بارش روزانه به‌کمک داده‌های مدل گردش عمومی در ایالت پنجاب از سه روش میدان تصادفی مشروط (سی‌آراف)، کا-نزدیک‌ترین همسایه، و ماشین بردار پشتیبان استفاده، و سه مدل با هم مقایسه شد. نتایج برتری مدل‌های سی‌آراف و کاان‌ان از مدل اس‌وی‌ام را نشان داد (دپاشری و موجومدار ۲۰۱۱).

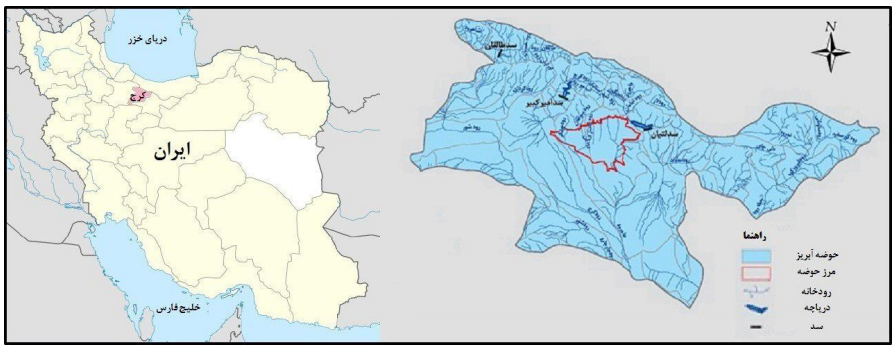
با توجه به پژوهش‌های معدودی که در پیش بینی آب‌دهی رود با استفاده از دو مدل ناپارامتری کا-نزدیک‌ترین همسایه و درخت تصمیم ام۵ انجام شده، دو مدل ناپارامتری فوق در این مقاله مقایسه می‌شود، اولین بار است که در ایران شاخص گوپتا (کاجی‌ای) برای ارزیابی این دو مدل استفاده شده است.

1-ANFIS
2-ANN
3-Piece- Wise linear functions
4-SPI

مواد و روش‌ها

منطقه‌ی بررسی شده

حوزه‌ی کرج در ۳۰–۶۰ کیلومتری شمال و شمال‌غربی تهران یکی از حوزه‌های برف‌گیر رشته کوه البرز است و به‌دلیل قرار داشتن سد امیرکبیر در خروجی آن، یکی از حوزه‌های آب‌خیز مهم در تأمین آب تهران است. این حوزه در محدوده‌ی بالادست سد کرج با مساحت ۸۴۰ کیلومترمربع در پهنه‌ی جنوبی رشته کوه‌های البرز مرکزی میان عرض جغرافیایی $^{/}۵۳\ ^{0}۳۵$ و $^{/}۱۰\ ^{0}۳۶$ و طول جغرافیایی $^{/}۰۳\ ^{0}۵۱$ و $^{/}۳۵\ ^{0}۵۱$ است. مرز جغرافیایی آن از شمال به حوزه‌ی چالوس، از شرق به حوزه‌ی لتیان، از غرب به حوزه‌ی کردان، و از جنوب به کوه‌های شمال تهران و دشت تهران و رباط کریم می‌رسد (شکل ۱).



شکل۱- موقعیت حوزه‌ی آب‌خیز کرج.

آزمون گاما

آزمون گاما ابزاری برای مدل‌سازی غیرخطی است که می‌توان با آن ترکیب مناسب عامل‌های ورودی برای مدل‌سازی داده‌های خروجی، و ایجاد مدلی هموار را بررسی نمود. با این آزمون مقدار میانگین مربعات خطای مدل پیش از استفاده محاسبه، و ترکیب بهینه‌ی متغیرهای ورودی شناسایی می‌شود. در واقع آزمون گاما تخمینی از کم ترین میانگین مربعات خطای آن بخش از داده‌ها است که با مدل خطی نمی‌توان آن‌ها را تعیین کرد. فرض بر این است که مجموعه‌ی داده‌های مشاهده یی به‌صورت زیر باشد:

$$\{(x_i , y_i), 1 \le i \le M\} \qquad (۱)$$

که در آن X_i بردار مشاهده‌یی ورودی (مانند آب‌دهی، بارش، دمای هوا با تأخیرهای مختلف و هر متغیر مؤثر دیگر)، y_i خروجی مشاهده یی (مانند میانگین آب‌دهی روان‌آب روزانه) و M تعداد کل مشاهده‌ها است. از بردار X_i (در فضای M بُعدی) برای مدل‌سازی y_i (در فضای R) استفاده می‌شود. بنابراین، باتوجه به فرض پیش‌گفته رابطه‌ی میان ورودی و خروجی مانند زیر است:

$$y = f(x) + r \qquad (۲)$$

که در آن f تابع هموار، و r متغیر تصادفی نشان‌دهنده‌ی خطا است. فرض می‌شود میانگین توزیعی که r از آن پیروی می‌کند صفر و واریانس خطا، var (r)، کران‌دار است. آزمون گاما بر پایه‌ی مجموعه‌ی N[i,k] است.

برای به دست آوردن مقدار بهینه‌ی بیش‌ترین مقدار نزدیک‌ترین همسایگی، ابتدا آماره‌ی گاما برای مقادیر مختلف نزدیک‌ترین همسایگی از رابطه‌ی (۲) محاسبه می‌شود. X N[i,k] به‌عنوان k امین نزدیک‌ترین همسایگی‌ها برای بردار مشاهده‌یی ورودی X_i تعریف می‌شود، به‌طوری‌که $1 \le$ $k \le p$ و $1 \le i \le M()$ است. برای به دست آوردن مقدار آماره‌ی گاما، ابتدا با توجه به داده‌های ورودی به شرح زیر، مقدار دلتا ($\delta M(k)$)، محاسبه می‌شود:

$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M} \left|X_{N[i,k]} - X_i\right|^2 \qquad (۳)$$

که در آن نماد |...| فاصله‌ی اقلیدسی بردار N[i,k]X از بردار همسایگی‌های آن است. مقدار آماره‌ی گامای کوچک، ($\gamma M(k)$) با استفاده از داده‌های خروجی به شرح زیر به دست می‌آید:

$$(k) = \frac{1}{2M}\sum_{i=1}^{M} (y_{N[i,k]} - y_i)^2\ \gamma_M \qquad (۴)$$

که در آن yN[i,k] مقدار خروجی نظیر k امین نزدیک ترین همسایگی برای بردار X_i است. برای محاسبه‌ی Γ خط رگرسیون خطی از p نقطه بر مقادیر $\delta M(k), \gamma M(k)$ برازش داده می‌شود. مقدار آماره‌ی Γ در واقع عرض از مبدأ و A شیب‌خط رگرسیون است.

$$\gamma = A\delta + \Gamma \qquad (۵)$$

مدل درخت تصمیم ام۵

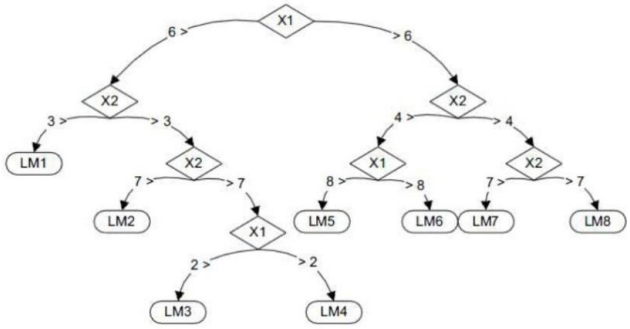
روش درخت تصمیم با استفاده از داده‌های ورودی ویژگی هدف را خروجی می گیرد، و بر اساس مدلی با ساختار درختی پیش‌بینی یا طبقه‌بندی می‌کند. کوینلان (۱۹۹۲) مدل درخت تصمیم ام۵ را برای پیش بینی داده های پیوسته پیشنهاد کردند. درخت تصمیم معمولاً از چهار بخش ریشه، شاخه، گره‌ها و برگ‌ها تشکیل می‌شود؛ گره‌ها با دایره نمایش داده می‌شوند و شاخه‌ها نشان‌دهنده‌ی اتصال میان گره‌ها است. در ریشه‌ی درخت تصمیم‌گیری، صفت هست و داده‌ها براساس آن صفت جداسازی می‌شوند. این تقسیم‌بندی تا جایی که داده‌ها دیگر جداشدنی نباشند، یا در فرآیند رسیدن به هدف نیازی به تقسیم‌بندی بیش‌تر احساس نشود ادامه می‌یابد (هند و همکاران۲۰۱۱). درخت تصمیم به صورت چپ به راست یا بالابه‌پایین کشیده می‌شود و گره‌ی اول که ریشه است در بالا گذاشته می‌شود. هر گره نشان‌دهنده صفتی معین است، و شاخه‌ها بازه‌ای از مقادیر معلوم برای صفت اند. از درختان تصمیم استقرایی براساس مجموعه‌یی از داده‌های معلوم درخت تصمیم استنتاجی به‌دست می‌آید، و این درخت می‌تواند به صورت مجموع قوانینی برای پیش‌بینی ویژگی‌های معلوم استفاده شود (عشقی و همکاران ۲۰۱۴). در تولید درخت تصمیم از الگوریتم استنتاجی یا معیار انشعاب (تقسیم) استفاده می‌شود. معیار تقسیم برای الگوریتم مدل ام۵ ارزیابی انحراف معیار مقادیر رده‌یی است که کمیتی از خطا است و به گره می‌رسد؛ و کاهش انتظارداشته دراین خطا را همچون نتیجه‌ی آزمون هرصفت در آن گره محاسبه می‌کند. کاهش انحراف معیار (SDR) به صورت زیر محاسبه می‌شود:

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (۶)$$

$$sd(T) = \sqrt{\frac{1}{N}(\sum_{i=1}^{N} {y_i}^2 - \frac{1}{N}(\sum_{i=1}^{N} y_i)^2)} \quad (۷)$$

که در آن T مجموعه‌یی از نمونه‌های واردشده به هر گره، Ti زیر مجموعه‌یی از نمونه‌هایی که i–امین خروجی آزمون بالقوه را دارند، sd انحراف معیار، yi مقدار عددی ویژگی هدف نمونه‌ی i، و N شماره‌ی داده‌ها است [۲].

برای دست‌یابی به دقت بالاتر، فرآیند انشعاب در هر گره بارها تکرار می‌شود، تا درنهایت به گره پایانی که برگ است برسد. در برگ مجموع مجذور انحراف از میانگین داده‌ها نزدیک به صفر خواهد شد. انشعاب مکرر منجر به تولید درختی بزرگ با شاخه‌ها و گره‌های زیاد می‌شود و کار با این درخت را دشوار می‌سازد، بنابراین هرس شاخه‌های اضافی برای دست‌یابی به درختی بهینه و کارآمد ضروری است. درخت بهینه براساس کم‌ترین کردن خطای پیش‌بینی، انتخاب می‌شود (فلاحی و همکاران ۲۰۱۲). در شکل ۲ ساختار سلسله‌مراتبی مدل درختی ام۵ آورده شده است.



شکل ۲- ساختار سلسله مراتبی شبیه درختی ام۵.

روش کا-نزدیکترین همسایه

الگوریتم کا-نزدیک‌ترین همسایه نیز نوعی یادگیری مبتنی بر مثال یا یادگیری تنبل است. این الگوریتم از ساده‌ترین الگوریتم‌های یادگیری ماشینی است که هدف آن دسته‌بندی نمونه‌یی جدید براساس ویژگی نمونه‌های آموزش دهنده است. روش کا-نزدیک‌ترین همسایه برخلاف تابع های انتقالی سنتی از هیچ تابع ریاضیاتی ازپیش‌تعریف‌شده‌یی برای تخمین متغیرهای مختلف استفاده نمی‌کند. دراین روی‌کرد، مجموعه‌ی داده‌یی مرجعی مانند مجموعه‌ی داده‌یی به‌کاررفته در آموزش و توسعه‌ی تابع های انتقالی سنتی برای یافتن نزدیک‌ترین (مشابه‌ترین) نمونه به نمونه ی هدف، جست‌وجو می‌شود. نخستین گام در این زمینه، تعیین فاصله میان نمونه‌ی هدف با هریک از داده‌های مجموعه‌ی است. در بیش‌تر بررسی‌ها، برای اندازه‌گیری فاصله‌ی میان نمونه‌ی مجهول (هدف) و نمونه‌های مجموعه‌ی مرجع، ازروابط سنتیِ محاسبه‌ی فاصله‌ی اقلیدسی نمونه‌ی هدف تا هریک از

نمونه‌های مجموعه‌ی مرجع استفاده شده است. برای مشخص کردن فاصله‌ی میان داده‌های بانک مرجع با نمونه ی هدف از شکل کلی رابطه‌ی فیثاغورث استفاده می‌شود (جاکتاپ و همکاران ۲۰۰۴).

$$D(X,Y) = \sqrt{\sum_{i=1}^{nf}(X_i - Y_i)^2} \quad (۸)$$

که در آن X نماینده‌ی نمونه با چند عامل مشخص (x1 تا xn) از بانک داده ی مرجع و Y نمونه‌ی هدف با همان تعداد عامل (y۱ تا yn) است. به این ترتیب، نمونه داده‌های مجموعه‌ی مرجع به ترتیب صعودی از کم‌ترین (بیش‌ترین تشابه) تا بیش‌ترین فاصله (کم‌ترین تشابه) از نمونه‌ی خواسته شده دسته‌بندی و ارزش‌گذاری خواهد شد (جلالی و همایی ۲۰۱۰).

$$X = (x_1, x_2, x_3, \dots x_n) \quad (۹)$$

$$Y = (y_1, y_2, y_3, \dots y_n) \quad (۱۰)$$

میزان کارآیی این روش بسیار به کیفیت انتخاب نزدیک‌ترین(مشابه‌ترین) نمونه‌ها از مجموعه‌ی مرجع با نمونه‌ی هدف وابسته است. زیرا به‌راحتی نمی‌توان گفت که برای رسیدن به بیش ترین دقت در تخمین‌ها، استفاده از نزدیک‌ترین نمونه درمجموعه‌ی مرجع با کم‌ترین فاصله منجر به تخمینی با دقت پذیرفتنی خواهد شد. برای تعیین تعداد بهینه‌ی k درتخمین نمونه‌ی هدف، اعتبار سنجی تقاطعی با حذف تک‌تک عامل‌ها[1]، پیشنهاد شده است (جلالی و همایی ۱۳۸۹).

الگوریتم کا-نزدیک‌ترین همسایه (کاان‌ان) روشی ناپارامتری برای طبقه‌بندی و رگرسیون است. در هر دو مورد، ورودی شامل K نزدیک‌ترین نمونه‌های آموزشی در فضای ویژگی است، و خروجی به استفاده‌ی از کاان‌ان برای طبقه‌بندی و یا رگرسیون وابسته است. وزن‌دهی به همسایه‌ها در هر دو کاربرد این الگوریتم (طبقه‌بندی و رگرسیون) مفید است به‌طوری که همسایه‌های نزدیک‌تر به هدف، سهم بیش‌تری از همسایه‌های دورتر از هدف دارند. روش وزن‌دهی به هر همسایه به صورت d/۱ است، که d فاصله‌ی هدف تا همسایه است. وکا نرم‌افزار داده‌کاوی متن‌باز است که از بسیاری از الگوریتم‌های یادگیری ماشین پشتیبانی می‌کند. تمام قسمت‌های این نرم‌افزار به زبان جاوا نوشته شده است. در این تحقیق برای اجرای مدل کاان‌ان و درخت تصمیم ان۵ از این نرم‌افزار استفاده شده است. داده‌ها به دو گروه آموزش با به‌کارگیری ۷۰ ٪ داده‌ها (۱۷۸۵ گروه) و آزمون با ۳۰ ٪ داده‌ها (۷۶۷ گروه) تقسیم شد.

برای بررسی دقت برآورد مدل برای هریک از ترکیب‌ها از معیارهای ضریب تبیین آر دو، خطای جذر میانگین مربعات آرام‌اس‌ای، کلینگ گوپتا کاجی‌ای (گوپتا و همکارانش ۲۰۰۹)، و ناش- ساتکلیف ان‌اس‌ای (جعفرزاده و همکاران ۲۰۱۷) استفاده شد. رابطه‌های این معیارها به‌صورت زیر است:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(Q_{oi} - \bar{Q}_o)(Q_{si} - \bar{Q}_s)}{\sqrt{\sum_{i=1}^{n}(Q_{oi} - \bar{Q}_o)^2}\sqrt{\sum_{i}^{n}(Q_{si} - \bar{Q}_s)^2}}\right)^2 \quad (۱۱)$$

$$RMSE = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(Q_{si} - Q_{oi})^2} \quad (۱۲)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (۱۳)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_{si} - Q_{oi})^2}{\sum_{i=1}^{n}(Q_{oi} - \bar{Q}_o)^2} \quad (۱۴)$$

در این رابطه‌ها Qoi آب‌دهی مشاهده‌یی، Qsi آب‌دهی تخمینی در زمان i، t تعداد داده‌ها، Qoiمیانگین آب‌دهی مشاهده‌یی، r ضریب همبستگی خطی میان Qo و Qs،

α نسبت انحراف‌معیار Qs بر انحراف‌معیارQo،

βنسبت میانگین Qs بر میانگین Qo است. مقدار آردو همیشه میان صفر و یک تغییر می‌کند.

هر چه به یک نزدیک‌تر باشد، نشان‌دهنده‌ی تطابق بهتر اندازه‌های تخمینی و مقادیر اندازه‌گیری‌شده است. مقدار آرام‌اس‌ای بیانگر متوسط مربع خطا میان اندازه‌های مشاهده‌یی و پیش‌بینی‌شده است، و هر چه مقدار آن کم‌تر باشد نشان‌دهنده‌ی دقت بیش‌تر تخمین مدل است. شاخص کاجی‌ای رابطه‌ی میان اندازه‌های مشاهده‌یی و مدل‌سازی شده را نشان می‌دهد. این شاخص توسعه داده‌شد تا تجزیه‌ی مناسب‌تری از شاخص ناش-ساتکلیف به‌دست‌آید، تا تجزیه‌ی اهمیت نسبی اجزای متفاوت آن، مانند همبستگی، را در مدل‌سازی آب‌شناختی تسهیل کند (گوپتا و همکاران ۲۰۰۹). ضریب ناش-ساتکلیف از -∞ تا ۱ متغیر است (جعفرزاده و همکاران ۲۰۱۷). اگر مقدار آن صفر یا کم‌تر از آن شود بیانگر این است که میانگین مشاهده‌یی بهتر از مقدار برآوردی مدل است.

1-leave one out cross validation

هرچه مقدار این ضریب به عدد ۱ نزدیک‌تر شود عمل‌کرد مدل بهتر است (موریاسی و همکاران ۲۰۰۷).
در این تحقیق اغلب ترکیب‌های احتمالی از بارش، دما، تبخیر و آب‌دهی تهیه گردید. حرف Pبارش همان روز (میلی‌متر)، P-۱ بارش روز قبل، P-۲ بارش دو روز قبل، P-۳ بارش سه روز قبل، P-۴ بارش چهار روز قبل، P- بارش پنج روز قبل، q آب‌دهی همان روز (مترمکعب بر ثانیه)، q-۱ آب‌دهی روز قبل، q-۲ آب‌دهی دو روز قبل، q-۳ آب‌دهی سه روز قبل، q-۴ آب‌دهی چهار روز قبل، q-۵ آب‌دهی پنج روز قبل، T دمای همان روز (سانتی‌گراد)، T-۱ دمای روز قبل، T-۲ دمای دو روز قبل، E تبخیر همان روز (میلی‌متر بر روز)، E-۱ تبخیر روز قبل، و E-۲ تبخیر دو روز قبل گرفته شدند. ترکیب‌های به‌دست‌آمده با نرم‌افزار وین گاما آزموده شد. براساس آزمون گاما، بهترین ترکیب‌ها با اندازه‌های آماره‌ی گامای کم‌تر انتخاب شد (جدول۱).

نتایج و بحث

مبنای انتخاب ترکیب بهینه در این پژوهش کم‌ترین مقدار آماره‌ی گاما است. این آماره در حقیقت عرض ازمبدا عمودی خط رگرسیون است. آماره‌ی گاما واریانس خطاهای هر مدل خطی برازش داده شده بر داده‌ها را تخمین می‌زند. مقدار شیب خط رگرسیونی مفهومی از پیچیدگی مدل را می‌دهد. هرچه این شیب تندتر باشد نشان‌دهنده ی پیچیدگی بیش تر مدل در ترکیب‌های داده شده است. خطای معیار نیز بیانگر توان اطمینان مقدار گاما است، و مقادیر کوچک‌تر، قابلیت اطمینان کمتر مقدار گاما را نشان می‌دهد. نسبت V، نسبت آماره‌ی گاما به واریانس خروجی شده است. نزدیکی نسبت V به صفر نشان‌دهنده‌ی درجه‌ی زیادی از توان پیش‌بینی خروجی خاص است. اگر این نسبت‌به ۱نزدیک باشد، خروجی معادل نوسانات تصادفی است، تا جایی که نوسان‌های خطی گرفته شود. براساس نتیجه‌ی آزمون گاما ۱۸ ترکیب انتخاب شد. از هر گروه از ترکیب‌ها، ترکیب‌های با آماره‌ی گامای کم‌تر انتخاب شدند. برای بررسی تأثیر آب‌دهی روزهای قبل بر پیش‌بینی آب‌دهی تعدادی از ترکیب‌هایی که آب‌دهی روزهای قبل در آن‌ها وجود نداشت نیز انتخاب شدند. به همین علت در جدول ۱ترکیب‌هایی با آماره‌ی گامای بسیارزیاد (ترکیب نامناسب) نیز هست.

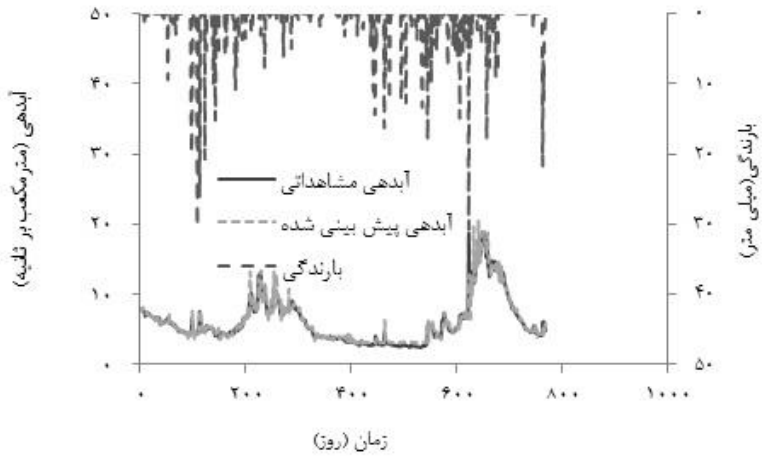
جدول ۱- نتایج آزمون گاما برای انتخاب ترکیب‌های مناسب.

ترکیب	ترکیب پارامترها	آماره‌ی گاما	شیب	خطای معیار	نسبت V
۱	T, P, E	۰/۱۱۱۱۴	۲/۸۱۹۱	۰/۰۰۷۲۳	۰/۴۴۴۵۶
۲	T_{-2}, P_{-2}, E_{-2}	۰/۱۰۹۱۹	۳/۲۹۵۴	۰/۰۰۷۲۹	۰/۴۳۶۷۸
۳	$T, P, E, q_{-1}, q_{-2}, q_{-3}, q_{-4}$	۰/۰۰۴۶۶	۰/۱۰۲۱۴	۰/۰۰۰۹۵	۰/۰۱۸۶۵
۴	$T, P, E, q_{-1}, q_{-2}, q_{-3}, q_{-4}, q_{-5}$	۰/۰۰۳۸۹	۰/۰۹۵۴۹	۰/۰۰۰۹۷	۰/۰۱۵۵۷
۵	q_{-1}, q_{-2}, q_{-3}	۰/۰۰۸۳۷	۰/۱۳۰۴۳	۰/۰۰۰۷۴	۰/۰۳۳۴۸
۶	$q_{-1}, q_{-2}, q_{-3}, q_{-4}, q_{-5}$	۰/۰۰۸۲۹	۰/۰۶۰۱۳	۰/۰۰۰۷۵	۰/۰۳۳۱۵
۷	$T, T_{-1}, q_{-1}, q_{-2}, q_{-3}, q_{-4}$	۰/۰۰۶۱۰	۰/۱۲۸۰۹	۰/۰۰۱۰۳	۰/۰۲۴۴۲
۸	$T, T_{-1}, T_{-2}, q_{-1}, q_{-2}, q_{-3}, q_{-4}$	۰/۰۰۵۸۲	۰/۱۰۹۵۴	۰/۰۰۰۹۳	۰/۰۲۳۳۰
۹	$E, E_{-1}, q_{-1}, q_{-2}, q_{-3}, q_{-4}$	۰/۰۰۵۷۷	۰/۱۴۵۰۶	۰/۰۰۱۱۳	۰/۰۲۳۱۱
۱۰	$E, E_{-1}, E_{-2}, q_{-1}, q_{-2}, q_{-3}, q_{-4}$	۰/۰۰۵۵۳	۰/۱۲۴۲۲	۰/۰۰۰۸۶	۰/۰۲۲۱۱
۱۱	$P, P_{-1}, P_{-2}, q_{-1}, q_{-2}, q_{-3}, q_{-4}$	۰/۰۰۵۸۱	۰/۰۵۳۱۹	۰/۰۰۱۶۱	۰/۰۲۳۲۶
۱۲	$P, P_{-1}, P_{-2}, P_{-3}, q_{-1}, q_{-2}, q_{-3}$	۰/۰۰۵۸۲	۰/۰۴۵۳۳	۰/۰۰۰۹۰	۰/۰۲۳۲۹
۱۳	$P, P_{-1}, P_{-2}, P_{-3}, P_{-4}, q_{-1}$	۰/۰۰۰۵۳	۰/۰۸۶۱۰	۰/۰۰۱۰۶	۰/۰۱۲۲۵
۱۴	$P, P_{-1}, P_{-2}, P_{-3}, P_{-4}, q_{-1}, q_{-2}$	$-۲/۶۸۷\times۱۰^{-۶}$	۰/۰۵۳۵۷	۰/۰۰۰۹۶	۰/۰۲۹۱۵
۱۵	$P, P_{-1}, P_{-2}, P_{-3}, P_{-4}, q_{-1}, q_{-2}, q_{-3}, q_{-4}, q_{-5}$	۰/۰۰۱۷۰	۰/۰۳۲۷۲	۰/۰۰۰۷۵	۰/۰۰۶۷۹
۱۶	$P, P_{-1}, P_{-2}, P_{-3}, P_{-4}, P_{-5}, q_{-1}, q_{-2}, q_{-3}$	۰/۰۰۵۷۸	۰/۰۳۳۲۱	۰/۰۰۱۱۷	۰/۰۲۳۱۲
۱۷	P, P_{-1}, E, E_{-1}	۰/۱۶۳۳۴	۰/۴۷۶۲۲	۰/۰۰۶۰۲	۰/۶۵۳۳۸
۱۸	$P, P_{-1}, E, E_{-1}, E_{-2}$	۰/۱۵۸۶۵	۰/۴۰۲۹۳	۰/۰۰۶۱۴	۰/۶۳۴۶۱

جدول ۲ -شاخص‌های آماری دوره‌ی آزمون شبیه کا-نزدیک‌ترین همسایه.

	آموزش				آزمون			
ترکیب	R^2	RMSE	KGE	NSE	R^2	RMSE	KGE	NSE
۵	۱	۰	۱	۱	۰/۹۰۳۶	۱/۰۸۴۴	۰/۹۳۷۴	۰/۸۹۵۶
۶	۱	۰	۱	۱	۰/۹۳۴۵	۰/۸۶۷۶	۰/۹۶۳۶	۰/۹۳۳۲
۱۱	۱	۰	۱	۱	۰/۹۱۸۵	۰/۹۶۳۴	۰/۹۵۵۳	۰/۹۱۷۶
۱۴	۱	۰	۱	۱	۰/۹۱۷۸	۰/۹۶۶۸	۰/۹۵۲۰	۰/۹۱۷۰
۱۵	۱	۰	۱	۱	۰/۹۰۴۶	۱/۰۵۲۳	۰/۹۵۰۱	۰/۹۰۱۷

نتایج مدل کاان‌ان برای ترکیب‌های مختلف عامل ها در برآورد جریان (جدول ۲) نشان می‌دهند که هر پنج ترکیب عمل کرد مناسبی در برآورد جریان روزانه دارند. اختلاف در شاخص‌های آماری ترکیب‌ها با هم زیاد نیست، اما عمل کرد ترکیب شش شامل عامل آب‌دهی روز قبل تا پنج روز قبل با ضریب تبیین، مجذور میانگین مربعات خطا آرام‌اس‌ای، ضریب گوپتا (کاجی‌ای) و ضریب ناش- ساتکلیف (ان‌اس‌ای) به ترتیب یک، صفر، یک و یک در مرحله ی آموزش و ۰/۹۳۴۵، ۰/۸۶۷۶، ۰/۹۶۳۶ و ۰/۹۳۳۲، در مرحله‌ی آزمون در میان سایر ترکیب‌ها بهترین بوده است. نمودار اندازه‌های آب‌دهی مشاهده‌یی و پیش‌بینی شده برای ترکیب برتر مدل کاان‌ان (ترکیب شش) در شکل ۳ نشان داده شده است.

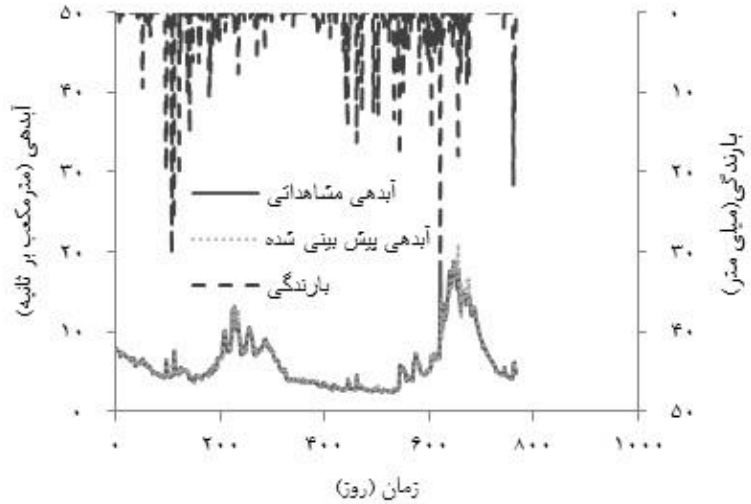


شکل ۳ -آب‌دهی روزانه مشاهده‌یی و پیش‌بینی شده‌ی شبیه کاان‌ان در دوره‌ی آزمون برای ترکیب ۶.

جدول ۳ -شاخص‌های آماری دوره‌ی آزمون شبیه ام۵.

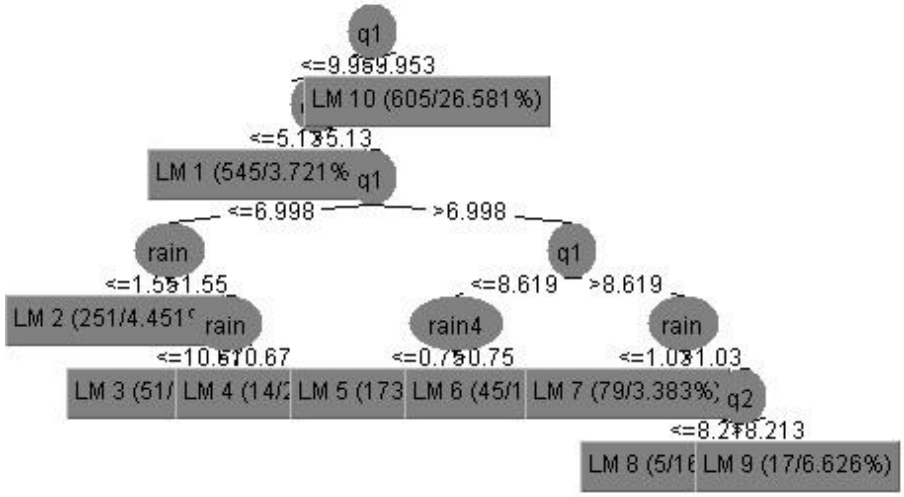
	آموزش				آزمون			
ترکیب	R^2	RMSE	KGE	NSE	R^2	RMSE	KGE	NSE
۵	۰/۹۵۸۴	۱/۶۹۰۸	۰/۹۷۰۲	۰/۹۵۸۴	۰/۹۵۶۱	۰/۷۱۳۳	۰/۹۷۳۹	۰/۹۵۴۸
۶	۰/۹۵۷۲	۱/۷۱۱۵	۰/۹۶۹۵	۰/۹۵۷۳	۰/۹۶۵۷	۰/۶۲۶۸	۰/۹۷۰۶	۰/۹۶۵۱
۱۱	۰/۹۷۳۴	۱/۳۵۱۸	۰/۹۸۰۷	۰/۹۷۳۴	۰/۹۷۳۸	۰/۵۴۸۰	۰/۹۸۵۱	۰/۹۷۳۳
۱۴	۰/۹۷۷۳	۱/۲۴۷۹	۰/۹۸۰۶	۰/۹۸۲۲	۰/۹۶۹۲	۰/۵۹۳۵	۰/۹۸۳۶	۰/۹۶۸۷
۱۵	۰/۹۷۳۶	۱/۳۴۶۳	۰/۹۸۰۸	۰/۹۷۳۶	۰/۹۷۳۸	۰/۵۴۶۸	۰/۹۸۵۵	۰/۹۷۳۴

نتایج مدل ام۵ (جدول ۳) نشان می‌دهند که این مدل در برآورد جریان روزانه‌ی رود کرج عمل کرد مناسبی داشته است. ترکیب ۱۵ با آب‌دهی روز قبل تا پنج روز قبل، و بارش همان روز و روز قبل تا چهار روز قبل، و ضریب تبین، مجذور میانگین مربعات خطا آرام‌اس‌ای، ضریب گوپتا کاجی‌ای و ضریب ناش-ساتکلیف ان‌اس‌ای به ترتیب ۰/۹۷۳۶، ۱/۳۴۶۳، ۰/۹۸۰۸ و ۰/۹۷۳۶ در مرحله‌ی آموزش، و ۰/۹۷۳۸، ۰/۵۴۶۸، ۰/۹۸۵۵ و ۰/۹۷۳۴ در مرحله‌ی آزمون، و ترکیب برتر مدل ام۵ بود. پس از ترکیب ۱۵ به‌ترتیب ترکیب‌های ۱۱، ۱۴، شش و پنج توانستند با دقت زیادی جریان رود کرج را برآورد کنند. براساس این ترتیب، نتیجه گرفته می‌شود که تأخیرهای زمانی باعث افزایش دقت نسبی ترکیب‌ها در برآورد جریان شده‌اند، به‌طوری‌که با افزایش تأخیرهای زمانی نتایج بهتری به‌دست‌آمده است. به‌همین دلیل، ترکیب ۱۵ که تأخیر زمانی آن برای بارش چهار روز و برای آب‌دهی پنج روز است بهترین ترکیب برای مدل ام۵ است. شکل ۴ نمودار اندازه‌ی آب‌دهی مشاهده‌ای و پیش‌بینی‌شده برای ترکیب برتر مدل ام۵ (ترکیب ۱۵) است.



شکل -۴ آب‌دهی روزانه‌ی مشاهده‌یی و پیش‌بینی‌شده‌ی مدل ام۵ در دوره‌ی آزمون برای ترکیب ۱۵.

درخت تصمیم برای ترکیب ۱۵ (ترکیب برتر) که به صورت قوانین-اگر و آن-گاه ارائه شده در شکل ۵ آورده شده است.



شکل ۵- درخت تصمیم برای ترکیب ۱۵ در مرحله‌ی آزمون.

نتیجه‌گیری

اهمیت برنامه‌ریزی و مدیریت منابع آب و رشد روزافزون جمعیت و محدودیت منابع آب سطحی در کشور، پیش‌بینی دقیق‌تر جریان رود را با استفاده از ابزارها و روش‌های نوین مدل‌سازی به‌ضرورتی گریزناپذیر تبدیل کرده است. در این پژوهش با کمک آزمون گاما از میان ترکیب‌های احتمالی بارش، تبخیر، دما و آب‌دهی، به‌همراه تأخیرهای زمانی آن‌ها، ترکیب‌های با آماره‌ی گامای کم‌تر (ترکیب‌های مؤثرتر) انتخاب و آب‌دهی حوزه‌ی آب‌خیز کرج با دو مدل کاان‌ان و درخت تصمیم ام۵ برآورد شد. هدف از استفاده‌ی از آزمون گاما جلوگیری از ورود تمام ترکیب‌های احتمالی تهیه‌شده به مدل‌های ناپارامتری، و تشخیص ترکیب ورودی مناسب به مدل‌ها بود. به همین علت پنج ترکیب پنج، شش، ۱۱، ۱۴ و ۱۵ به‌علت کم‌تربودن آماره‌ی گاما انتخاب شد. اگرچه، به‌طورکلی آماره‌ی گامای ترکیب‌هایی که شامل بارش و آب‌دهی و تأخیرهای این دو است کم‌تربود؛ به عبارت دیگر آن‌ها ترکیب‌های مناسب‌تر یا مؤثرتری در پیش‌بینی آب‌دهی اند. بیشترین اندازه‌ی آماره‌ی گاما (ترکیب‌های نامناسب) در ترکیب‌های یک، دو، ۱۷ و ۱۸، به علت نبودن عامل آب‌دهی بود. این ترکیب شامل دو عامل مؤثر در ذوب برف یعنی دما و تبخیر بود. از طرفی، حوزه‌ی بررسی شده برف‌گیر نیست، بنابراین اثر این دو عامل در پیش بینی آب‌دهی کم است و این ترکیب‌ها ورودی مدل نشدند. باتوجه به این که در مدل کاان‌ان از همسایگی‌های نزدیک نمونه‌ی مجهول در پیش‌بینی مجهول استفاده می‌شود، هرچه از تعداد آب‌دهیِ روزهای‌قبل بیش‌تری در پیش‌بینی استفاده شود نتیجه تطابقِ بیش تری با واقعیت دارد. به همین دلیل، ترکیب شش که آب‌دهی روز قبل تا ۵ روز قبل را شامل می‌شود علاوه بر مقدار آر۲ زیاد، مقدار آرام‌اس‌ای کم تری میان سایر ترکیب ها داشته است. با توجه به هر پنج ترکیب برتر می‌توان گفت که مدل کاان‌ان هنگامی می‌تواند آب‌دهی را درست پیش‌بینی کند که هر دو عامل آب‌دهی (تأخیرهای آب‌دهی) و بارش و تأخیرهای آن ورودی به مدل باشد، اگرنه مدل نمی تواند به درستی پیش‌بینی کند. در مدل ام۵ نیز دو عامل آب‌دهی و بارش در برآورد مناسب جریان تأثیرگذار است، به‌نحوی که ترکیب‌هایی که این دو پارامتر در آن‌ها نبود شاخص‌های آماریشان کم و نپذیرفتنی بود، و برای هر پنج ترکیبی که این دو عامل در آن بود، شاخص‌های آماری بالای به دست آمد. هرچه تأخیرهای زمانی برای این دو پارامتر بیشتر شد نتیجه‌ی بهتری به‌دست آمد. اگرچه، نتایج شاخص‌های آماری هر پنج ترکیب بسیار به هم نزدیک است و اختلاف میان آن‌ها زیاد نیست، به‌طوری‌که اختلاف شاخص آر۲ برای بهترین و بدترین ترکیب کم‌تر از ۰/۰۲ است.

روش مدل درختی به‌دلیل سادگی، فهمیدنی بودن، و دادن روابط خطی ساده در محدوده‌ی خاصی از داده‌های ورودی (که به صورت قوانین-اگر-آنگاه- داده می‌شود) می‌تواند روشی جایگزین برای روش‌های دیگر و آن‌ها باشد. باتوجه به نتایج به دست آمده از دو مدل کاان‌ان و ام۵ میزان جریان برآوردشده با اندازه‌های مشاهده‌یی آن هم‌خوانی زیادی داشته است، و هر دو مدل توانایی زیادی در برآورد جریان روزانه‌ی رود کرج دارند. هر ۵ ترکیب انتخاب‌شده عمل‌کرد مناسبی در برآورد آب‌دهی روزانه‌ی رود کرج دارند و تأخیرهای زمانی در دقت مدل اثر چندانی نداشته است. اندازه‌های هر چهار شاخص آماری استفاده‌شده در این تحقیق، در تمامی ترکیب‌ها نشان‌دهنده‌ی برتری نسبی مدل ام۵ از مدل کاان‌ان است، که با نتایج ستاری و همکاران (۲۰۱۳)، ظهیری و قربانی (۲۰۱۲) و عظیمی و همکاران (۲۰۱۴) مطابقت دارد؛ این برتری معنی‌دار نبود، هر دو مدل برای برآورد جریان رود به‌کار می‌رود.

منابع

Ababaei B. VerdiyNejad R. 2013. Estimation of hydraulic performance of pressure irrigation systems using artificial neural networks and nonparametric regression. Water and Soil Journal (Agricultural Sciences and Technology). 27 (4): 769–779.

Azimi V. Vakilifar A. Asadi A. 2014 . Evaluation of gene expression programming and m5 Model for daily drain estimation (Case study: Lighvan River in the 2002-2002 period), International Journal of Water Resources and Development. 3 (3): 134–142.

Deepashree R. Mujumdar P. 2011. A comparison of three methods for downscaling daily precipitation in the Punjab Region, Hydrological Processes. 25(23):3575–3589.

Durrant PJ. 2001. WinGamma: a non-linear data analysis and modelling tool with application to flood prediction, Ph.D. thesis, Department of Computer Science, Cardiff University, Wales, UK.

Emami Far S. Alipour A. 2011. Using the M5 tree estimation to estimate evaporation from the exposed surface (Case Study: Qom Area). Eleventh General Seminar on Irrigation and Evaporation Reduction. Kerman Shahid Bahonar University.

Eshghi P. Farzadmehr J. Dasturani MT. Arabs Asadi Z. 2014. The Effectiveness of Intelligent Models in Estimating the River Suspended Sediments (Case Study: Babaaman Basin, Northern Khorasan) Journal of Watershed Management Research Vol. 7,

No. 14, Autumn and Winter 2016.
Fallahi MR. Varvani H. Goliyan S. 2012. Precipitation forecasting using regression tree model to flood control. 5th International watershed and water and soil resources management, Kerman, Iran.
Fatahi Nafchi RA. Mousavi SF. Kargar A. 2012. Application of artificial neural network method developed in a basin to estimate monthly discharge in sub-basins (Case study: BeheshtAbad Basin). Iran Water Research Journal. 6 (11): 193-197.
Gupta HV. Kling H. Yilmaz K. Martinez GF. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. Journal of Hydrology. l 377: 80–91.
Hamed Ensaniat N. 2012. Simulation of daily runoff using PSO algorithm in optimization of basin models. Master's thesis, Faculty of Engineering Islamic Azad University, Science Research Branch.
Hand, DJ, Mannila, H., Smyth, P 2001.Principles of data mining, Cambridge, Mass: The MIT Press.
Hejje Bakhsh P. 2010. Predicting the suspended load by regression decision making trees and comparing by emprical models.MSc thesis, Civil Faculty, Yazd University: 72 p.
Jafarzadeh A. Pourreza Bilondi M. Khashei-Siuki1 A. Aghakhani Afshar A Yaghoobzadeh M. 2017. Reliability estimation of rainwater catchment system using future GCM output data (case study: Birjand city) journal of European Water. 59: 169-175.
Jagtap SS. Lall U. Jones, JW. Gijsman AJ. Ritchie JT. 2004. Dynamic nearest-neighbor method for estimating soil water parameters.Trans. ASAE. 47:1437–1444.
Jalali VR. Homayi M. 2010. A nonparametric model is proposed using the -k closest neighboring technique to estimate the bulk density of the soil. Journal of Agricultural Science and Technology, Water and Soil Science. 56: 181- 190.
Jones J. Evans A. Margetts D. Durrant SJP. 2002. The GAMMA test. Idea Group Publishing. 26 p.
Lin B. Syed M. Falconer RA. 2008. Predicting faecal indicator levels in estuarine receiving waters eAn integrated hydrodynamic and ANN modelling approach. Environmental Modelling & Software. 23: 729– 740.
Mansuy N. Thiffault E. Pare D. Bernier P. Guindon L. Villemaire P. Poirier V. Beaudoin A. 2014. Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbour method. Geoderma. 59-73.
Moghaddamnia A. Ghafari Gousheh M. Piri J. Amin S. Han D. 2009. Evaporation estimation using artificial neural networks and adaptive neurofuzzy inference system techniques, Advanced, Journal of Water Res. 32: 88–97.
Moriasi DN. Arnold JG. Van Liew, MW. Bingner RL. Harmel RD. Veith TL. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations,Transactions of the ASABE. 50: 885–900.
Quinlan JR. 1992. Learning with continuous classes. In proceedings AI, 90 (Adams & Sterling, Eds), 343-348.
Remsan R. Shamim MA. Han D. Mathew J. 2008. ANFIS and NNARX based rainfall runoff modeling, International Conference on Systems, Man and Cybernetics. . Miyazaki, Japan.
Sattari MT. Nahrin F. Azimi V. 2013. Daily reference evapotranspiration prediction using artificial neural network model and M5 tree model (Case study: Bonab station). Irrigation and Drainage Journal of Iran. 1 (7): 104–113.
Seifi A. Riahi-Madvar H. 2012. Input variable selection in expert systems based on hybrid Gamma Test-Least Square Support Vector Machine ANFIS and ANN models. Provisional Chapter, INTECH. United kingdom.
Senthil Kumar A. Ojha C. Goyal M. Singh R. Swamee P. 2012.and Modeling of suspended sediment concentration at Kasol in India using ANN, Fuzzy logic, and decision tree algorithms. Journal of Hydrol. Eng. 17(3): 394–404.
Soleimani Motlagh M. Talebi A. Akrami M.

2011. Possibility of predicting the probability of occurrence of avalanches using the nearest neighbor method in the GIS software environment (Case study: Avalanche axes of Tehran Province). Iran Watershed Science and Engineering. 5 (16): 33–38.
Yurekli. K. Taghi Sattari MT. Anli AS. Hinis MA. 2012. Seasonal and annual regional drought prediction by using data-mining approach. Atmosfera. 25 (1): 85-105.
Zahiri A. Azamathullah HM. 2012. Comparison between linear genetic programming and M5 tree models to predict flow discharge in compound channels. Journal of Neural Comput & Applic. 24:413–420.
Zahiri A. Qourbani Kh. 2012. Flow simulation in compound sections using M5 decision tree model. Journal of Soil and Water Protection Research. 20 (3): 113–132.