



شماره ۱۱۷، زمستان ۱۳۹۶

# پژوهش‌های آب‌نخرداری

(پژوهش و سازندگی)

## مقایسه‌ی دو مدل ناپارامتری کا-نزدیک‌ترین همسایه و درخت تصمیم ام ۵ 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 پوررضایلندی

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 بر داده است. در این پژوهش با استفاده از دو مدل مبتنی بر داده (ام ۵، کان‌ان) و داده‌های آب‌وهواشناسی (آب‌دهی، بارش، دما و تبخیر) در دوره‌ی ۱۳۸۸-۱۳۸۱ آب‌دهی رود در حوزه‌ی آب‌خیز کرج پیش‌بینی، و کارایی 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 ام ۵ و کان‌ان 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 ام ۵ دقت بیشتری از مدل کان‌ان دارد، به‌طوری‌که ۲ برای مدل‌های ام ۵ و کان‌ان به ترتیب ۰/۹۷۳۸، ۰/۹۳۴۵، مقدار آرام‌اس‌ای آنها به ترتیب 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 ام ۵، 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 در دسترس 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 (فتاحی‌نقچی و همکاران ۲۰۱۲). مدل، درک ساده‌ی از یک سامانه‌ی واقعی است (مانند فرآیند پیچیده‌ی تبدیل بارش به روان‌آب). می‌توان مدل را نظریه، قانون یا اندیشه‌ی ساختاری

دانست. با توسعه‌ی مدل امکان توضیح ساده‌ی سامانه‌ی پیچیده فراهم می‌شود (حامد انسانیت ۲۰۱۲). عوامل بسیاری بر پدیده‌های آب‌شناختی مؤثراند که تعیین میزان تأثیرشان نامشخص است. در این صورت تعیین بهترین ترکیب از عوامل ورودی به مدل‌های هوشمند وقت‌گیر است. کاربرد روش‌های نوین در این راستا از مراحل سعی و خطا 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) آن را با جزئیات بررسی کردند (رمسان و همکاران ۲۰۰۸). این روش ابزاری برای مدل‌سازی غیرخطی است که به‌کمک

شاخص بارش معیار<sup>۴</sup> مقایسه شدند و نشان داد که میان این دو روش تفاوت زیادی در طبقه بندی خشکسالی نیست (یورکلی ۲۰۱۲) برای پیش بینی جریان در آب راه های مرکب، دو روش برنامه نویسی ژنتیک خطی و مدل درخت ام ۵ مقایسه شدند. نتایج نشان داد که دقت مدل ام ۵ به دقت مدل ال جی پی نزدیک است؛ از این رو می توان برای پیش بینی جریان رود در رویدادهای سیل از مدل ام ۵ استفاده کرد (ظهیری و عظمت الله ۲۰۱۲)

روش کا-نزدیک ترین همسایه، زیرمجموعه ی روش های ناپارامتری خودگردان ساز است. این نام به دلیل روند باز-نمونه گیری با جایگزینی از مجموعه ی مشاهده، داده شده است. پژوهش های زیادی با این روش انجام شده است. عمل کرد هیدرولیکی دستگاه های آبیاری تحت فشار با استفاده از مدل های شبکه ی عصبی مصنوعی و رگرسیون ناپارامتری ارزیابی شدند. نتایج نشان دادند که مدل شبکه ی عصبی مصنوعی ام ال پی می تواند ضریب یکنواختی لوله ی فرعی های آبیاری تحت فشار را با استفاده از مشخصات هیدرولیکی و فیزیکی لوله ی فرعی با کم ترین خطا برآورد نماید. روش کا-نزدیک ترین همسایه میانگین اندازه های سی یو را با دقت مدل سازی می کند، اما نمی تواند این کار را با دقت برای انحراف معیار این مقادیر انجام دهد، و خطای آن در مرحله ی آزمون بسیار بیشتر از دو مدل دیگر برآورد شد (ابابابی و همکاران ۲۰۱۳). نشان داده شده است که نزدیک ترین همسایه روش مناسبی برای تعیین سطح هشدار (ضریب آستانه) و پیش بینی احتمال روی داد بهمن در محورهای بهمن خیز استان تهران است (سلیمانی مطلق و همکاران ۲۰۱۱). نتایج تهیه ی نقشه های کاربری با روش کانان و مجموعه ی داده ی آموزشی ۵۳۸ قطعه زمین در سراسر کانادا و ۱۸ متغیر پیش بینی کننده ی محیطی نشان دهنده ی برتری روش کانان از روش های قبلی خاک بود (مانسی و همکاران ۲۰۱۴). برای ریزمقیاس نمای بارش روزانه به کمک داده های مدل گردش عمومی در ایالت پنجااب از سه روش میدان تصادفی مشروط (سی آراف)، کا-نزدیک ترین همسایه، و ماشین بردار پشتیبان استفاده، و سه مدل با هم مقایسه شد. نتایج برتری مدل های سی آراف و کانان از مدل اس وی ام را نشان داد (دپاشری و موجدمدار ۲۰۱۱).

با توجه به پژوهش های معدودی که در پیش بینی آب دهی رود با استفاده از دو مدل ناپارامتری کا-نزدیک ترین همسایه و درخت تصمیم ام ۵ انجام شده، دو مدل ناپارامتری فوق در این مقاله مقایسه می شود، اولین بار است که در ایران شاخص گوپتا (کاجی ای) برای ارزیابی این دو مدل استفاده شده است.

آن می توان قبل از ایجاد مدل، ترکیب مناسب از عوامل ورودی را برای مدل سازی داده های خروجی و ایجاد مدلی هموار، بررسی نمود، به طوری که با بررسی خطا در هر ترکیب ورودی، معیاری مناسب برای قضاوت بر آن ها فراهم می شود. در نهایت ترکیبی که برای مدل سازی بهینه ی خروجی کم ترین مقدار خطا دارد، ترکیب بهینه ی ورودی می شود (مقدم نیا و همکاران ۲۰۰۹). آزمون گاما روشی غیرخطی و پیشرفته برای تحلیل داده ها است، که می تواند داده ها را برای ایجاد مدلی خطی تجزیه و تحلیل کند. در روند شناسایی مدل برای تجزیه و تحلیل داده ها، از این آزمون برای توسعه ی مدل شبکه ی 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<sup>۱</sup> و شبکه ی عصبی مصنوعی<sup>۲</sup> استفاده شد. نتایج نشان دادند که جی تی-ال اس اس وی ام دقت زیادی دارد و مراحل سعی و خطا در مدل سازی کم شده است (سیفی و ریاحی مدوار ۲۰۱۲).

مدل ام ۵ مشابه توابع خطی جدا شده<sup>۳</sup> است، که ترکیبی است از مدل های رگرسیون خطی و رگرسیون درختی و کاربردهای زیادی در علوم مختلف دارد. برآورد تبخیر از سطح آزاد با کمک مدل درختی ام ۵ انجام شد. نتایج نشان دادند که دو عامل میانگین روزانه ی دمای هوا و سرعت باد بیش ترین تأثیر را بر برآورد تبخیر از سطح آزاد دارند، و با داشتن این دو عامل، می توان با کمک مدل درختی ام ۵، تبخیر از سطح آزاد را با دقتی پذیرفتنی تخمین زد (امامی و علی پور ۲۰۱۱). بار رسوب بستر با استفاده از روش تصمیم گیری رگرسیونی و مقایسه با روش های تجربی برآورد شد. نتایج نشان دادند که درخت تصمیم در برآورد مواد معلق دقت بیشتری دارد (هجی بخش ۲۰۱۰). کارآیی مدل درختان تصمیم گیری در برآورد مواد معلق رودی را در حوزه ی آب خیز باباامان خراسان شمالی بررسی شد. نتایج نشان داد که به طور کلی و براساس معیارهای آماری استفاده شده (آر، آرام اس ای)، مدل درخت تصمیم گیری در مقایسه با روش منحنی سنجه ی رسوب تطابق بسیار بیشتری با مقادیر اندازه گیری شده داشته است و دقت و کارآیی بیشتری دارد (عشقی و همکاران ۲۰۱۴). در حوزه ی کاسل هند، تجمع رسوب معلق با استفاده از شبکه ی عصبی مصنوعی، منطق فازی و الگوریتم درخت تصمیم مدل سازی شد. نتایج نشان دادند که مدل درخت تصمیم ام ۵ دقت و درستی بیشتری از شبکه ی عصبی و منطق فازی دارد (سنتیل کومار و همکاران ۲۰۱۲). طبقات خشک سالی در حوزه ی شکرک ترکیه با روش درخت تصمیم ارزیابی و پیش بینی شد. نتایج این روش با روش طبقه بندی و تحلیل خشک سالی بر اساس

1-ANFIS

2-ANN

3-Piece- Wise linear functions

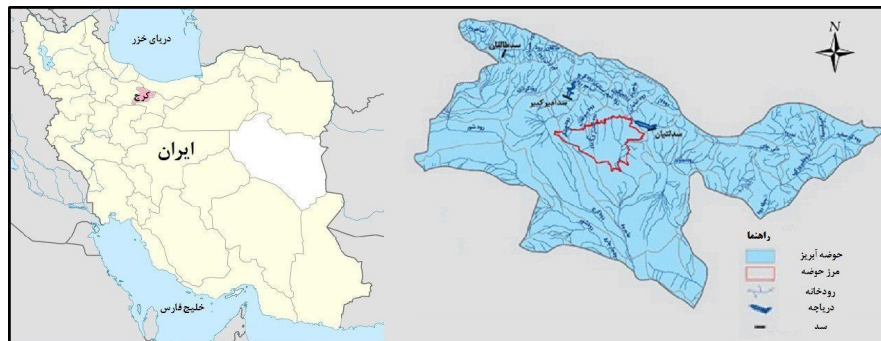
4-SPI

## 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 ۳۵° ۵۳' و ۳۶° ۱۰' و طول جغرافیایی ۵۱° ۰۳' و ۵۱° ۳۵' 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).



شکل ۱- 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 کم‌ترین میانگین مربعات خطای آن بخش از داده‌ها است که با مدل خطی نمی‌توان آن‌ها را تعیین کرد. فرض بر این است که مجموعه‌ی داده‌های مشاهده‌ی به‌صورت زیر باشد:

(۱)  $\{(x_i, y_i), 1 \leq i \leq M\}$   
 که در آن  $X_i$  بردار مشاهده‌ی ورودی (مانند آب‌دهی، بارش، دمای هوا با تأخیرهای مختلف و هر متغیر مؤثر دیگر)،  $y_i$  خروجی مشاهده‌ی (مانند میانگین آب‌دهی روان آب 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، با توجه به فرض پیش‌گفته رابطه‌ی میان ورودی و خروجی مانند زیر است:

$$y = f(x) + r \quad (2)$$

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،  $\text{var}(r)$ ، 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.  $X$  نزدیک‌ترین به‌عنوان  $k$  امین نزدیک‌ترین همسایگی‌ها برای بردار مشاهده‌ی ورودی  $X_i$  تعریف می‌شود، به‌طوری که  $1 \leq k \leq p$  و  $1 \leq i \leq M$  است. برای به دست آوردن مقدار آماره‌ی گاما، ابتدا با توجه به داده‌های ورودی به شرح زیر، مقدار دلتا  $(\delta M(k))$ ، محاسبه می‌شود:

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^M |X_{N[i, k]} - X_i|^2 \quad (3)$$

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

$$(\gamma M(k)) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i, k]} - y_i)^2 \gamma_M \quad (4)$$

که در آن  $y_{N[i, k]}$  مقدار خروجی نظیر  $k$  امین نزدیک‌ترین همسایگی برای بردار  $X_i$  است. برای محاسبه‌ی  $\Gamma$  خط رگرسیون خطی از  $p$  نقطه بر مقادیر  $\delta M(k), \gamma M(k)$  برازش داده می‌شود. مقدار آماره‌ی  $\Gamma$  در واقع عرض از مبدأ و شیب خط رگرسیون است.

$$\gamma = A\delta + \Gamma \quad (5)$$

## مدل درخت تصمیم ام ۵

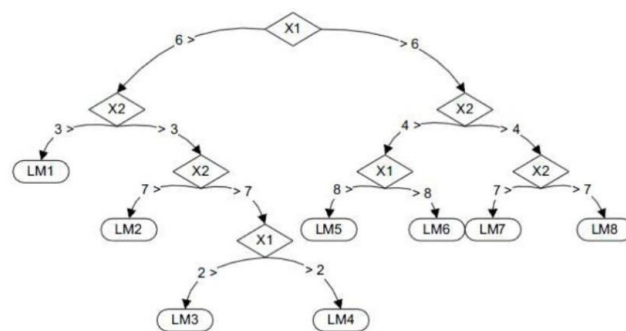
$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (۶)$$

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

که در آن T مجموعه‌یی از نمونه‌های وارد شده به هر گره،  $T_i$  زیر مجموعه‌یی از نمونه‌هایی که  $\bar{I}$  - امین خروجی آزمون بالقوه را دارند، sd انحراف معیار،  $y_i$  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 در نهایت 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 ام ۵ آورده شده است.

روش درخت تصمیم با استفاده از داده‌های ورودی ویژگی هدف را خروجی می‌گیرد، و بر اساس مدلی با ساختار درختی پیش‌بینی یا طبقه‌بندی می‌کند. کوینلان (۱۹۹۲) مدل درخت تصمیم ام ۵ را برای پیش‌بینی داده‌های 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 جداسازی نباشند، یا در فرآیند رسیدن به هدف نیازی به تقسیم‌بندی بیش‌تر احساس نشود ادامه می‌یابد (هند و همکاران ۲۰۱۱). درخت تصمیم به صورت چپ به راست یا بالا به پایین کشیده می‌شود و گره‌ی اول که ریشه است در بالا گذاشته می‌شود. هر گره نشان‌دهنده صفتی معین است، و شاخه‌ها بازه‌های از مقادیر معلوم برای صفت اند. از درختان تصمیم استنتاجی براساس مجموعه‌یی از داده‌های معلوم درخت تصمیم استنتاجی به دست 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 ام ۵ ارزیابی انحراف معیار مقادیر رده‌یی است که کمیتی از خطا است و به گره می‌رسد؛ و کاهش انتظار داشته در این خطا را همچون نتیجه‌ی آزمون هرفست در آن گره محاسبه می‌کند. کاهش انحراف معیار (SDR) به صورت زیر محاسبه می‌شود:



شکل ۲- ساختار سلسله‌مراتبی شبیه درختی ام ۵.

مجموعه‌ی داده‌یی مرجعی مانند مجموعه‌ی داده‌یی به کاررفته در آموزش و توسعه‌ی تابع‌های انتقالی سنتی برای یافتن نزدیک‌ترین (مشابه‌ترین) نمونه به نمونه‌ی 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، از روابط سنتی محاسبه‌ی فاصله‌ی اقلیدسی نمونه‌ی هدف تا هریک از

## 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 تابع‌های انتقالی سنتی از هیچ تابع ریاضیاتی از پیش تعریف‌شده‌یی برای تخمین متغیرهای مختلف استفاده نمی‌کند. در این رویکرد،

۳۰٪ داده‌ها (۷۶۷ گروه) تقسیم شد.

برای بررسی دقت برآورد مدل برای هر یک 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 به صورت زیر است:

$$R^2 = \left( \frac{\sum_{i=1}^n (Q_{oi} - \bar{Q}_o)(Q_{si} - \bar{Q}_s)}{\sqrt{\sum_{i=1}^n (Q_{oi} - \bar{Q}_o)^2} \sqrt{\sum_{i=1}^n (Q_{si} - \bar{Q}_s)^2}} \right)^2 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (Q_{si} - Q_{oi})^2} \quad (12)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (13)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{si} - Q_{oi})^2}{\sum_{i=1}^n (Q_{oi} - \bar{Q}_o)^2} \quad (14)$$

در این رابطه‌ها  $Q_{oi}$  آب‌دهی مشاهده‌ی،  $Q_{si}$  آب‌دهی تخمینی در زمان  $t$ ،  $i$  تعداد داده‌ها،  $Q_{oi}$  میانگین آب‌دهی مشاهده‌ی،  $r$  ضریب همبستگی خطی میان  $Q_s$  و  $Q_o$ ،  $\alpha$  نسبت انحراف معیار  $Q_s$  بر انحراف معیار  $Q_o$ ،  $\beta$  نسبت میانگین  $Q_s$  بر میانگین  $Q_o$  است. مقدار آر دو همیشه میان صفر و یک تغییر می‌کند.

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

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

$$D(X, Y) = \sqrt{\sum_{i=1}^{nf} (X_i - Y_i)^2} \quad (8)$$

که در آن  $X$  نماینده‌ی نمونه با چند عامل مشخص ( $x_1$  تا  $x_n$ ) از بانک داده‌ی مرجع و  $Y$  نمونه‌ی هدف با همان تعداد عامل ( $y_1$  تا  $y_n$ ) 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).

$$X = (x_1, x_2, x_3, \dots, x_n) \quad (9)$$

$$Y = (y_1, y_2, y_3, \dots, y_n) \quad (10)$$

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

الگوریتم کا-نزدیک‌ترین همسایه (کاان) 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 کاان برای طبقه‌بندی و یا رگرسیون وابسته است. وزن‌دهی به همسایه‌ها در هر دو کاربرد این الگوریتم (طبقه‌بندی و رگرسیون) مفید است به طوری که همسایه‌های نزدیک‌تر به هدف، سهم بیش‌تری از همسایه‌های دورتر از هدف دارند. روش وزن‌دهی به هر همسایه به صورت  $d/1$  است، که  $d$  فاصله‌ی هدف تا همسایه است. و کا 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 کاان و درخت تصمیم آن ۵ از این نرم‌افزار استفاده شده است. داده‌ها به دو گروه آموزش با به کارگیری ۷۰٪ داده‌ها (۱۷۸۵ گروه) و آزمون با

گاما است. این آماره در حقیقت عرض از مبدا 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 نشان‌دهنده‌ی 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 نسبت به ۱ نزدیک 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 ۱ ترکیب‌هایی با آماره‌ی گامای بسیار زیاد (ترکیب نامناسب) نیز هست.

هرچه مقدار این ضریب به عدد ۱ نزدیک‌تر شود عمل کرد مدل بهتر است (موریاسی و همکاران ۲۰۰۷). در این تحقیق اغلب ترکیب‌های احتمالی از بارش، دما، تبخیر و آبدهی تهیه گردید. حرف P بارش 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 (جدول ۱).

### نتایج و بحث

مبنای انتخاب ترکیب بهینه در این پژوهش کم‌ترین مقدار آماره‌ی

جدول ۱- نتایج آزمون گاما برای انتخاب ترکیب‌های مناسب.

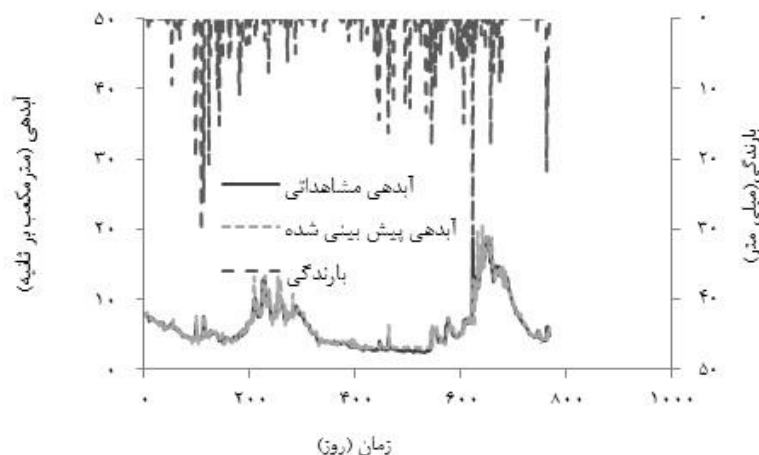
ترکیب	ترکیب پارامترها	آماره‌ی گاما	شیب	خطای معیار	نسبت V
۱	T, P, E	۰/۱۱۱۱۴	۲/۸۱۹۱	۰/۰۰۷۲۳	۰/۴۴۴۵۶
۲	T-2, P-2, E-2	۰/۱۰۹۱۹	۳/۲۹۵۴	۰/۰۰۷۲۹	۰/۴۳۶۷۸
۳	T, P, E, q-1, q-2, q-3, q-4	۰/۰۰۴۶۶	۰/۱۰۲۱۴	۰/۰۰۰۹۵	۰/۰۱۸۶۵
۴	T, P, E, q-1, q-2, q-3, q-4, q-5	۰/۰۰۳۸۹	۰/۰۹۵۴۹	۰/۰۰۰۹۷	۰/۰۱۵۵۷
۵	q-1, q-2, q-3	۰/۰۰۸۳۷	۰/۱۳۰۴۳	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	$-۲/۶۸۷ \times ۱۰^{-۶}$	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.

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

(اناس‌ای) به ترتیب یک، صفر، یک و یک در مرحله‌ی آموزش (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

نتایج مدل کان‌ان برای ترکیب‌های مختلف عامل‌ها 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



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

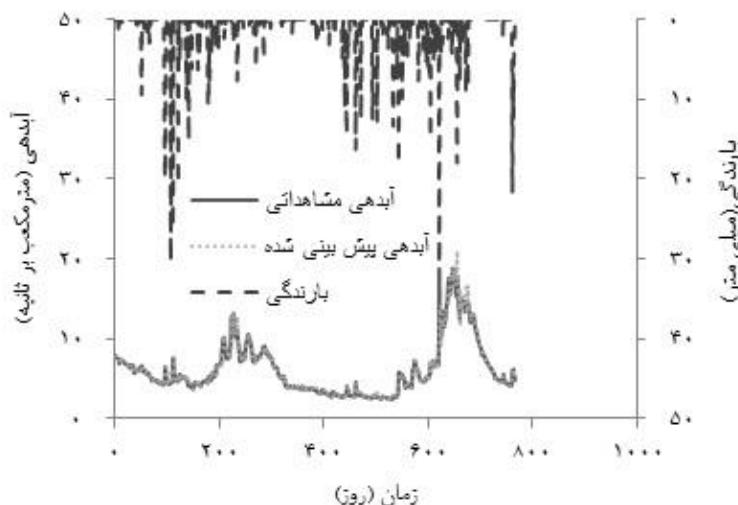
جدول ۳- شاخص‌های آماری دوره‌ی آزمون شبیه ام.۵.

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



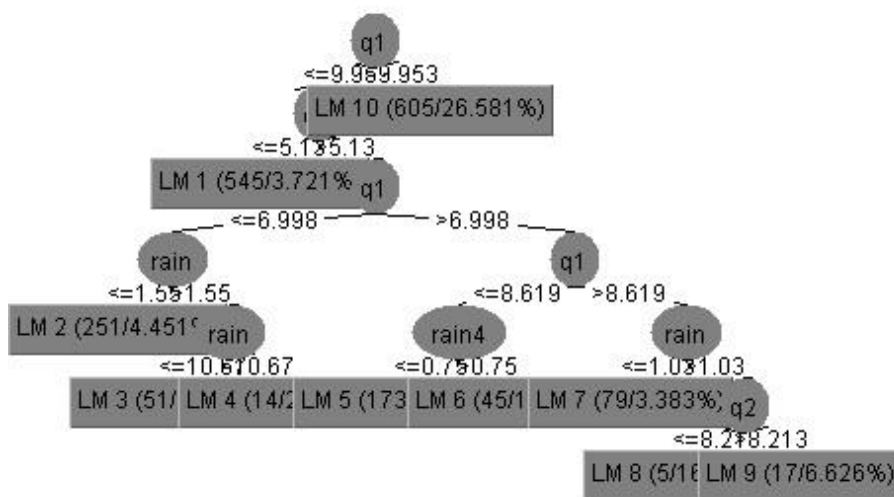
شش و پنج توانستند با دقت زیادی جریان رود کرج را برآورد کنند. براساس این ترتیب، نتیجه گرفته می‌شود که تأخیرهای زمانی باعث افزایش دقت نسبی ترکیب‌ها در برآورد جریان شده‌اند، به طوری که با افزایش تأخیرهای زمانی نتایج بهتری به دست آمده است. به همین دلیل، ترکیب ۱۵ که تأخیر زمانی آن برای بارش چهار روز و برای آب‌دهی پنج روز است بهترین ترکیب برای مدل ۵ام است. شکل ۴ نمودار اندازه‌ی آب‌دهی مشاهده‌ای و پیش‌بینی شده برای ترکیب برتر مدل ۵ام (ترکیب ۱۵) است.

نتایج مدل ۵ام (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 تبیین، مجذور میانگین مربعات خطا آرام‌اس‌ای، ضریب گوپتا کاجی‌ای و ضریب ناش-ساتکلیف ان‌اس‌ای به ترتیب ۰/۹۷۳۶، ۰/۳۴۶۳، ۰/۹۸۰۸، ۰/۹۷۳۶ و ۰/۹۷۳۴ در مرحله‌ی آموزش، و ۰/۹۷۳۸، ۰/۵۴۶۸، ۰/۹۸۵۵ و ۰/۹۷۳۴ در مرحله‌ی آزمون، و ترکیب برتر مدل ۵ام بود. پس از ترکیب ۱۵ به ترتیب ترکیب‌های ۱۱، ۱۴،



شکل ۴- آب‌دهی روزانه‌ی مشاهده‌ی و پیش‌بینی شده‌ی مدل ۵ام 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 کانان و درخت تصمیم ۵م 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 کم‌تر بودن آماره‌ی گاما انتخاب شد. اگرچه، به‌طورکلی آماره‌ی گامای ترکیب‌هایی که شامل بارش و آب‌دهی و تأخیرهای این دو است کم‌تر بود؛ 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 پیش‌بینی آب‌دهی کم است و این ترکیب‌ها ورودی مدل نشدند. باتوجه به این که در مدل کانان از همسایگی‌های نزدیک نمونه‌ی مجهول در پیش‌بینی مجهول استفاده می‌شود، هرچه از تعداد آب‌دهی روزهای قبل بیش‌تری در پیش‌بینی استفاده شود نتیجه تطابق بیش‌تری با واقعیت دارد. به همین دلیل، ترکیب شش که آب‌دهی روز قبل تا ۵ روز قبل را شامل می‌شود علاوه بر مقدار آ۲ زیاد، مقدار آرام‌اس‌ای کم‌تری میان سایر ترکیب‌ها داشته است. با توجه به هر پنج

## 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. Dašturani 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,

ترکیب برتر می‌توان گفت که مدل کانان هنگامی می‌تواند آب‌دهی را درست پیش‌بینی کند که هر دو عامل آب‌دهی (تأخیرهای آب‌دهی) و بارش و تأخیرهای آن ورودی به مدل باشد، اگر نه مدل نمی‌تواند به درستی پیش‌بینی کند. در مدل ۵م 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 نپدیرفتنی بود، و برای هر پنج ترکیبی که این دو عامل در آن بود، شاخص‌های آماره‌ی بالای به دست آمد. هرچه تأخیرهای زمانی برای این دو پارامتر بیشتر شد نتیجه‌ی بهتری به دست آمد. اگرچه، نتایج شاخص‌های آماره‌ی هر پنج ترکیب بسیار به هم نزدیک است و اختلاف میان آن‌ها زیاد نیست، به‌طوری‌که اختلاف شاخص آ۲ 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 کانان و ۵م میزان جریان برآوردشده با اندازه‌های مشاهده‌ی آن هم‌خوانی زیادی داشته است، و هر دو مدل توانایی زیادی در برآورد جریان روزانه‌ی رود کرج دارند. هر ۵ ترکیب انتخاب‌شده عمل کرد مناسبی در برآورد آب‌دهی روزانه‌ی رود کرج دارند و تأخیرهای زمانی در دقت مدل اثر چندانی نداشته است. اندازه‌های هر چهار شاخص آماره‌ی استفاده‌شده در این تحقیق، در تمامی ترکیب‌ها نشان‌دهنده‌ی برتری نسبی مدل ۵م از مدل کانان 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.

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. 1377: 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 empirical models. MSc thesis, Civil Faculty, Yazd University: 72 p.

Jafarzadeh A. Pourreza Bilondi M. Khashei-Siuki 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. Gijssman 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. Hınıs 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.

