

## پیش‌بینی جریان رودخانه کشکان با استفاده از ترکیب روش‌های شبکه عصبی مصنوعی، آنالیز

### موجک و K- نزدیک‌ترین همسایه

داریوش یاراحمدی<sup>۱</sup> - حمید میرهاشمی<sup>۲\*</sup>

تاریخ دریافت: ۱۳۹۶/۰۳/۰۲

تاریخ پذیرش: ۱۳۹۶/۰۸/۱۵

### چکیده

پیش‌بینی دقیق هیدرولوژی و منابع آب می‌تواند اطلاعاتی مفیدی برای برنامه‌ریزی شهری، آمایش زمین، طراحی پروژه‌های شهری و مدیریت منابع آب ارائه دهد. در این مطالعه با در نظر گرفتن اهمیت قابل توجهی که رودخانه کشکان در تأمین بخش مهمی از آب رودخانه کرخه و مشروب ساختن زمین‌های کشاورزی استان لرستان دارد مدل پیش‌بینی سری زمانی جریان این رودخانه با استفاده از روش‌های K- نزدیک‌ترین همسایه (K-NN)، شبکه عصبی مصنوعی (ANN) و ترکیب آنالیز موجک (WT) اجرا شد. در این خصوص ابتدا با استفاده از نمایه هرست، حافظه سری زمانی رودخانه یاد شده به مقدار  $0/6$  به دست آمد که نشان از حافظه بلندمدت و رفتار دینامیکی سیگنال سری زمانی آن داشت. در ادامه با در نظر گرفتن اینکه سری زمانی جریان رودخانه تابعی از سری‌های زمانی با تأخیر  $1-3-5-7-10$  و  $15$  روز است. فرآیند مدل‌سازی سیگنال رواناب با استفاده از دو روش K-NN و ANN انجام گرفت. در گامی دیگر سری زمانی سیگنال رواناب با استفاده از موجک مادر میر، به  $4$  زیر سیگنال تجزیه شد که با اتخاذ این زیرسیگنال‌ها به جای سیگنال اصلی، مدل‌های ترکیبی K-NN-WT و ANN-WT جهت شبیه‌سازی رواناب اجرا شدند. نتایج حاصل از سنجش‌های کارایی عملکرد مدل نشان دادند که مدل K-NN با خطای  $4/6$  درصد و ضریب همبستگی  $0/9$  از عملکرد مناسب‌تری نسبت به شبکه عصبی که متحمل خطاهای نامتقارنی شده بود برخوردار است. از سوی دیگر ترکیب آنالیز موجک عملکرد هر دو مدل بهبود پیدا کرد که در این خصوص مدل ANN-WT با خطای  $1/2$  درصد و ضریب همبستگی  $0/989$  شبیه‌سازی دقیق‌تری را نسبت به سه مدل دیگر انجام داد.

**واژه‌های کلیدی:** رودخانه کشکان، سیگنال سری زمانی، موجک میر، نمایه هرست

### مقدمه

عوارض کوهستانی و دشتی هستند و همچنین از کاربری‌های پیچیده و ساده برخوردارند، عملی بغرنج بوده، چراکه جریان‌های معمول و حدی رودخانه در هر بخش از این حوضه‌های وسیع، وابسته به متغیرهای زیادی هستند که وزن و نوع هر یک از این متغیرها در بخش‌های مختلف حوضه از مقادیر متفاوتی برخوردار است. از سویی، به سبب پیچیدگی و عدم دانش کافی در مورد فرآیندهای فیزیکی چرخه هیدرولوژی و عدم توانایی در دخالت دادن تمام متغیرهای مؤثر در مقدار جریان رودخانه، به نظر می‌رسد که ساخت مدل‌های پیش‌بینی مرسوم با درجه دقت بسیار بالا، بسی دشوار خواهد بود. از این گذشته، اساس بسیاری از تصمیم‌گیری‌ها در فرآیندهای هیدرولوژیکی و تصمیمات بهره‌برداری از منابع آب بر پایه‌ی پیش‌بینی و تحلیل سری‌های زمانی است. به‌عنوان نمونه، بهره‌برداری در زمان واقعی از مخازن سدها نیازمند آگاهی از آورد جریان ورودی به مخزن در آینده بوده که این پروسه به‌وسیله پیش‌بینی سری زمانی جریان رودخانه امکان‌پذیر است (۱۰). به عبارتی مدل‌سازی سری‌های زمانی

پیش‌بینی و مدل‌سازی جریان رودخانه، گامی اساسی در راستای برنامه‌ریزی، طراحی و بهره‌برداری از سیستم منابع آب (۱۸) و مدیریت مسائل منوط به آن همچون خشک‌سالی و سیلاب‌های مخرب در حوضه‌های آبریز است (۲۴)؛ زیرا کمبود و مازاد جریان رودخانه به ترتیب منجر به خسارات مالی (خشک‌سالی) و جانی - مالی (سیلاب) خواهد شد (۸). به‌طوری‌که چنین پیش‌بینی‌های از جریان رودخانه با فراهم کردن هشدارهای لازم در مورد رخداد سیل و کمک به تنظیم جریان خروجی منابع آب در هنگام پایین بودن تراز جریان رودخانه، به مدیریت منابع آب کمک می‌کنند (۲). لازم به ذکر است که فرآیند مدل‌سازی و پیش‌بینی در حوضه‌های وسیعی که حاصل درهم‌تنیدگی

۱- دانشیار آب و هواشناسی گروه جغرافیا، دانشگاه لرستان

۲- دکترای آب و هواشناسی و مدرس گروه جغرافیا، دانشگاه لرستان

(Email: climate90@yahoo.com)

\*- نویسنده مسئول:

نشان می‌دهد. کناس<sup>۱</sup> و همکاران (۵) با تجزیه سیگنال سری زمانی رواناب حوضه تیرسو (واقع در ساردینا، ایتالیا) با استفاده از تبدیلات موجک پیوسته و گسسته و مدل سازی رواناب از طریق شبکه عصبی مصنوعی - موجک نتیجه گرفتند که تجزیه با استفاده از موجک گسسته، رفتار هیدرولوژیکی جریان رودخانه را دقیق‌تر پیش‌بینی می‌کند. پرامنیک و پاندا<sup>۲</sup> (۲۰) با اتخاذ ۵ الگو با زمان‌های تأخیر متفاوت از رودخانه ماهندی در هندوستان به‌عنوان ورودی روش‌های شبکه عصبی و نروفازی (ANFIS) دریافتند که رواناب شبیه‌سازی شده توسط مدل نروفازی دارای تطابق بهتری با رواناب مشاهداتی نسبت به مورد شبکه عصبی است. وئو و چائو<sup>۳</sup> (۲۴) با مدل‌سازی جریان ماهانه چندین رودخانه در کشور چین نشان دادند که روش K-NN در پیش‌بینی جریان رودخانه از کارایی مناسب‌تری نسبت به روش آریمایا و شبکه عصبی برخوردار است. یانگ<sup>۴</sup> و همکاران (۲۹) با ترکیب مدل HEC-HMS با مدل‌های نروفازی (ANFIS) و نرو ژنتیک (GANN) در طی رخداد پدیده تیفون در حوضه لانونگ در جنوب تایوان نشان دادند که شبیه‌سازی‌های ایجاد شده توسط مدل‌های ترکیبی HEC-HMS-ANFIS و HEC-HMS-GANN قادرند دقت پیش‌بینی را به‌طور قابل توجهی بهبود بخشند. در صورتی که مدل HEC-HMS به تنهای قادر نیست پیش‌بینی رضایت‌بخشی از رواناب ساعتی حوضه در طی رخداد تیفون به عمل آورد.

در این راستا با توجه به اهمیتی که مقدار جریان رودخانه و نوسان‌های کوتاه‌مدت و بلندمدت آن در ابعاد مختلف زندگی بشر ایفا می‌کند، دانستن رفتار و عملکرد آن ضرورت دارد. چراکه با کشف دینامیک رفتار آن می‌توان به چگونگی عملکرد آن در آینده پی برد؛ بنابراین هدف از انجام این مطالعه کشف و شبیه‌سازی عملکرد رودخانه کشکان با استفاده از روش‌های آماری و هوشمند است تا بتوان با ارائه مدل‌هایی با عدم قطعیت کمتر گامی در راستای بهبود برنامه‌ریزی‌های مبتنی بر جریان این رودخانه برداشت.

## مواد و روش‌ها

### منطقه مورد مطالعه

حوضه آبریز رودخانه کشکان با وسعت ۹۵۰۲/۷ کیلومتر مربع در ناحیه جنوب غربی ایران واقع شده است. گستره جغرافیایی آن بین مختصات ۳۳ درجه و ۸ دقیقه تا ۳۴ درجه و ۲ دقیقه عرض شمالی و ۴۷ درجه و ۱۲ دقیقه تا ۴۸ درجه و ۵۹ دقیقه طول شرقی قرار دارد.

به‌منظور تولید داده و پیش‌بینی متغیرهای هیدرولوژیکی، یک گام مهم در طراحی و تحلیل حساسیت منابع آب است (۱۹). چنانکه بسیار از فعالیت‌های مرتبط با برنامه‌ریزی و اجرای مؤلفه‌های سیستم منابع آب نیازمند پیش‌بینی رخداد‌های آینده است (۱۲) تا بر این اساس با مدیریت خشک‌سالی و سیلاب، خسارات ناشی از آن‌ها تا حد ممکن کمینه شود (۲۲). در این راستا، پژوهش‌های گسترده‌ای به‌منظور مدل‌سازی و پیش‌بینی جریان رودخانه در قالب مدل‌های آماری، تجربی، احتمالاتی و اخیراً هوش مصنوعی صورت گرفته که عمده این پژوهش‌ها در راستای کاهش عدم قطعیت و خطای پیش‌بینی و همچنین ارزیابی میزان کارایی روش‌های اتخاذ شده بوده‌اند.

انیس حسینی و ذاکر مشق (۳) به‌منظور پیش‌بینی رواناب رودخانه کشکان از منظر نظریه آشوب ابتدا با استفاده از روش‌های میانگین اطلاعات متقابل (لازم به ذکر است که این روش زمان تأخیر ۱۵ روز را به‌عنوان زمان بهینه تشخیص داد) و نزدیک‌ترین همسایگان کاذب، فضای فاز را بازسازی نموده و آنگاه با استفاده از بُعد همبستگی و توان لیاپانوف به ترتیب بُعد فراکتالی جاذب سیستم و حساسیت به شرایط اولیه را تخمین زدند. براین اساس شرایط یک سیستم آشوبناک را در سری زمانی دبی کشکان شناسایی کردند که در ادامه با استفاده از روش تقریب محلی در فضای فاز، فرآیند پیش‌بینی رواناب را با ضریب همبستگی ۰/۸۸ و جذر میانگین مربعات خطای ۳۳/۱۴ متر مکعب در ثانیه انجام دادند. شفائی و همکاران (۲۲) با تجزیه سیگنال سری زمانی جریان روزانه رودخانه آجی‌چای به ۱۱ زیر سری با استفاده از دو تابع موجک مادر هار و میر، مدل پیش‌بینی موجک - شبکه عصبی مصنوعی را برای جریان روزانه این رودخانه بنا نهادند. با مقایسه نتایج حاصل از تجزیه دو تابع موجک مادر یاد شده نتیجه گرفتند که تابع موجک مادر میر دارای قابلیت بهتری در تجزیه سیگنال سری زمانی بوده؛ همچنین مدل ترکیبی موجکی - عصبی نسبت به مدل شبکه عصبی از ضریب دقت بالاتری برخوردار است. عبداللهی اسدآبادی و همکاران (۱) به‌منظور ارائه مدل پیش‌بینی رواناب روزانه رودخانه حوضه آبخیز بهشت‌آباد بر اساس افق‌های ۱، ۲، ۳ و ۷ روز تأخیر زمانی، از سه روش موجک متقاطع، شبکه عصبی مصنوعی و موجک - شبکه عصبی مصنوعی استفاده کردند. مقایسه نتایج مترتب بر عملکرد این سه روش حاکی از دقت مدل ترکیبی موجکی - عصبی در قیاس با دو مدل دیگر بود. همچنین دقت مدل موجک متقاطع تنها در تأخیر زمانی ۱ روز، بهتر از مدل شبکه عصبی بود. ثانی‌خانی و همکاران (۲۱) با استفاده از روش‌های K-NN، ماشین بردار پشتیبان و فازی - عصبی تطبیقی به مدل‌سازی جریان ماهانه رودخانه باراندوزچای پرداختند. نتایج این مطالعه نشان داد که به‌رغم آنکه هر سه روش مورد استفاده، مقدار دبی بیشینه را کمتر از مقدار مشاهداتی برآورد می‌کنند با این حال، ماشین بردار پشتیبان عملکرد بهتری نسبت به دو روش دیگر در تخمین مقدار دبی بیشینه

1- Cannas

2- Pramanik and Panda

3-Wu and Chau

4- Young

همواره صفر خواهد بود؛ بنابراین دامنه تعدیل شده برابر خواهد بود با:

$$R_n = \max(Y_1, \dots, Y_n) - \min(Y_1, \dots, Y_n) \quad (3)$$

از آنجایی که میانگین  $Y$  صفر است، بیشینه آن همواره بزرگ‌تر یا مساوی صفر و کمینه آن کوچک‌تر یا مساوی صفر خواهد بود؛ بنابراین دامنه تعدیل شده همیشه غیر منفی است. هرست با استفاده از قاعده نصف در آمار رابطه ۴ را تعریف کرد:

$$\left(\frac{R}{S}\right)_n = a \cdot n^H \quad (4)$$

که در این رابطه،  $R$  دامنه تغییرات،  $S$  انحراف معیار سری زمانی،  $a$  عدد ثابت،  $n$  تعداد مشاهده‌ها و  $H$  نمای هرست است. این رابطه را می‌توان به صورت رابطه ۵ نیز نوشت

$$\log\left(\frac{R}{S}\right)_n = \log a + H \log(n) \quad (5)$$

در این راستا با رسم سمت چپ رابطه ۵ در مقابل  $\log(n)$ ، نمای هرست برآورد شد. برحسب یافته‌های هرست، اگر مقدار نمای هرست برابر با  $0/5$  شد، بر یک فرآیند مستقل نرمال دلالت دارد. اگر بین  $0/5$  و  $1$  قرار گرفت بر یک سری زمانی دوام‌دار با حافظه بلندمدت دلالت دارد که این به نوبه خود نشان‌دهنده ساختار غیرخطی سری زمانی نیز است. در نهایت اگر نمای هرست مثبت و کمتر از  $0/5$  شد، نشان از بی‌دوام بودن سری زمانی دارد.

وجود حافظه بلندمدت سری زمانی رواناب نشان از دینامیکی بودن این سری دارد بدین عنوان که حالت کنونی این سری وابسته به حالت‌های گذشته آن است؛ بنابراین زمان‌های تأخیر  $1, 3, 5, 7, 10$  و  $15$  روز قبل برای رواناب محاسبه که بدین ترتیب میزان رواناب به‌عنوان تابعی از این سری‌های زمانی قلمداد شدند. در مدل‌های پیش‌بینی سری‌های زمانی، سمت راست تابع رواناب (رابطه ۶) به‌عنوان ورودی و سری زمانی سمت چپ به‌عنوان سری زمانی هدف محسوب می‌شوند.

$$Q(t) = f \left( \begin{matrix} Q(t-1), Q(t-3), Q(t-5), \\ Q(t-7), Q(t-10), Q(t-15) \end{matrix} \right) \quad (6)$$

### الگوریتم انجام پیش‌بینی با استفاده از K-NN

روش K-نزدیک‌ترین همسایه از معروف‌ترین روش‌های رگرسیون ناپارامتری است. در این روش تابع توزیع مقادیر پیش‌بینی با استفاده از توزیع ناپارامتری تابع کرنل به دست می‌آید. این مدل بدین عنوان تدوین یافته که هرگاه شرایطی مشابه شرایط تاریخی مشاهده شده در زمان حال به وقوع پیوست، شرایطی مشابه در آینده مشاهده شرایطی خواهد بود که در آن تاریخ رخ داده است. مبنای مورد استفاده در این روش به این ترتیب است که با اندازه‌گیری و مشاهده متغیرهای مستقل در زمان واقعی، مدل به جستجوی حالت‌های مشابه

این حوضه بخش مهمی از سرشاخه‌های پرآب رودخانه کشکان را تشکیل می‌دهد، به طوری که حدود ۳۳٪ از خاک استان لرستان را در برمی‌گیرد. در تقسیم‌بندی کلی هیدرولوژیکی ایران، بخشی از حوضه آبریز خلیج فارس به شمار می‌رود. این حوضه  $18/5$ ٪ از مساحت حوضه آبریز بزرگ کرخه را شامل می‌شود.

برای رهیافت به این مطالعه، سری داده‌های دبی روزانه ایستگاه کشکان - پل دختر (واقع در پایاب رودخانه کشکان) در بازه زمانی ۱۳۷۰-۱۳۹۳ به‌عنوان ورودی اولیه مورد استفاده قرار گرفتند. روش‌های بکار گرفته شده در این مطالعه مبتنی بر بررسی حافظه سری زمانی رواناب، مدل‌سازی سری زمانی با استفاده از روش آماری K-نزدیک‌ترین همسایه (K-NN)، شبکه عصبی مصنوعی، ترکیب موجک - K-NN و ترکیب عصبی - موجک است که در ادامه مفاهیم تئوریک و چگونگی کاربرد این رویکردها شرح داده می‌شود.

اساس بسیاری از تصمیم‌گیری‌ها در فرآیندهای هیدرولوژیکی و تصمیمات بهره‌برداری از منابع آب مبتنی بر پیش‌بینی و تحلیل سری زمانی است. از سویی در فرآیند مدل‌سازی و پیش‌بینی، بررسی این مورد که آیا سری زمانی دارای حافظه بلندمدت است، ضرورت دارد؛ زیرا یک سری زمانی در صورتی از قابلیت پیش‌بینی برخوردار است که دارای حافظه بلندمدت باشد. این تفکر مبتنی بر آن است که آیا در سری زمانی در دسترس، تمامی پدیده‌های حدی قابل انتظار رخ داده‌اند (۱۰). در سری‌های زمانی هیدرولوژیکی (مثل رواناب) حافظه بلندمدت موجب وابستگی مقدار رواناب با مقادیر قبلی در گذشته شده که این نشان‌دهنده وجود پارامتری قابل پیش‌بینی در سری زمانی است.

هرست<sup>۱</sup> با مطالعه تراز آبی رودخانه نیل، آزمونی را برای بررسی حافظه بلندمدت ارائه داد (۹) که اینک به‌عنوان نمایه هرست محسوب می‌شود. مراحل انجام این آزمون که در مطالعه حاضر به‌عنوان ابزاری جهت بررسی حافظه بلندمدت سری زمانی رواناب رودخانه کشکان مورد استفاده قرار گرفته به شرح ذیل است. با اتخاذ یک سری زمانی رواناب با  $x = x_1, x_2, \dots, x_n$  ابتدا مقیاس داده‌ها با استفاده از رابطه ۱ نرمال شد.

$$k = 1, 2, \dots, n \quad z_k = (x_k - x_m) \quad (1)$$

که در آن  $x_k$  مقدار رواناب روزانه،  $x_m$  میانگین سری زمانی رواناب و  $z_k$  مقدار رواناب نرمال شده است. در مرحله بعد سری زمانی تجمعی رواناب محاسبه شد (رابطه ۲).

$$Y_i = \sum_{k=1}^i Z_k \quad (2)$$

نظر به اینکه میانگین  $Z$  صفر است، آخرین مقدار  $Y$  یعنی  $Y_n$

واقعی ( $t$ ) است. چنین برداری با متغیرهای مستقل مشاهده شده تاریخی به شکل زیر مقایسه می‌شود:

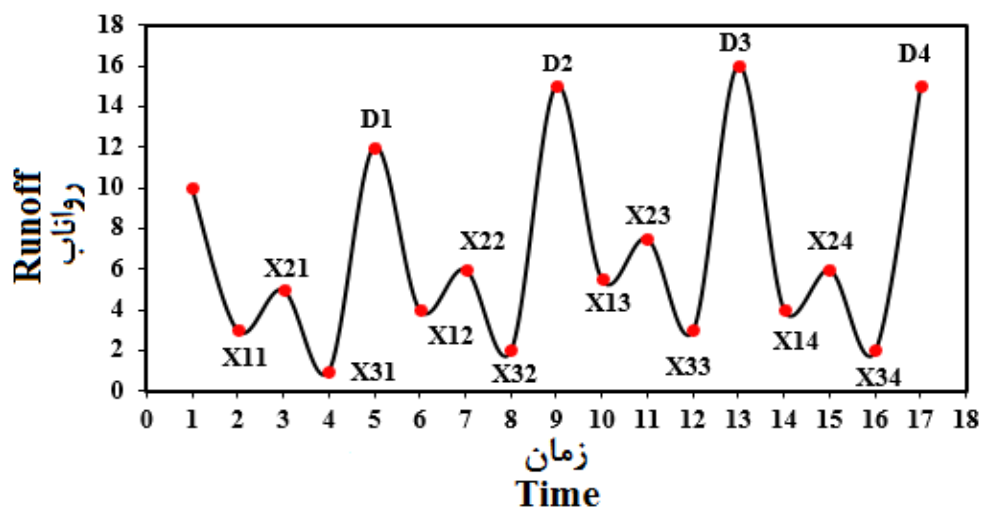
$$X_r = \{x_{1r}, x_{2r}, \dots, x_{nr}\} \quad (۸)$$

در این بردار  $x_{nr}$ ،  $n$  آمین متغیر مستقل مشاهده شده در زمان  $t$  در دوره مشاهده اطلاعات است. به ازای هر یک از بردارهای تاریخی تحت عنوان  $D_t$  وجود دارد. شکل ۱ شماتیک از نحوه در نظر گرفتن این متغیرها در یک سری تاریخی را نشان می‌دهد. با توجه به این شکل برای پیش‌بینی مقادیر اوج جریان ( $D$ ) از جریان مشاهده شده در ۳ بازه زمانی قبل ( $X$ ) استفاده می‌شود.

به حالت فعلی در سری‌های تاریخی می‌پردازد. وقایعی که در سری تاریخی بعد از این حالت‌ها پیش آمده‌اند می‌توانند به‌عنوان گزینه‌های محتمل در شرایط فعلی نیز در نظر گرفته شوند. احتمال رخداد هر یک از این حالت‌ها در شرایط کنونی بستگی به شباهت بردار مشاهداتی متغیرهای مستقل فعلی با بردار مستقل مشاهداتی در سری تاریخی دارد (۴ و ۱۰). الگوریتم انجام پیش‌بینی با استفاده از K-NN به شرح ذیل است. متغیرهای مشاهداتی در زمان واقعی را می‌توان به‌صورت برداری به شکل زیر نشان داد:

$$X_r = \{x_{1r}, x_{2r}, \dots, x_{nr}\} \quad (۷)$$

در این رابطه  $x_{nr}$  مبین متغیر مستقل مشاهده شده  $n$ ام در زمان



شکل ۱- شماتیک تعریف متغیرها در الگوریتم K-NN

Figure 1- Schematic definition of variables in the K-NN algorithm

$$Dis_t = W_t |X_{1r} - X_{1t}| + W_t |X_{2r} - X_{2t}| \quad (۱۰)$$

$$+ \dots + W_t |X_{nr} - X_{nt}|$$

در روابط ۹ و ۱۰،  $Dis_t$  فاصله بردار مشاهداتی متغیرهای مستقل فعلی از بردارهای تاریخی هستند. گزینه‌های که به‌عنوان گزینه برتر از میان سری تاریخی برای تعیین مقدار پیش‌بینی انتخاب می‌شوند گزینه‌هایی هستند که دارای کمترین فاصله از بردار مشاهداتی متغیرهای مستقل فعلی می‌باشند. چنانکه از رابطه‌های ۹ و ۱۰ برمی‌آید تعیین فاصله مشخص شده بستگی به وزنی دارد که برای هر کدام از متغیرهای مستقل در نظر گرفته می‌شود (۲۸). در این مطالعه برای تعیین وزن‌ها از روش اعتبارسنجی متقاطع<sup>۱</sup> استفاده شد. پس از محاسبه وزن‌ها مقدار عددی متغیر مطلوب  $D_r$  می‌تواند از رابطه زیر به دست آید.

تعداد سه متغیر مستقل برای پیش‌بینی و ۳ نزدیک‌ترین همسایه برای  $D_r$  منظور شده است. همچنین مقادیر  $X_r$  به‌صورت  $X_{14}$ ،  $X_{24}$  و  $X_{34}$  در این شکل قابل مشاهده‌اند.

جهت تشخیص همسایه‌ها ضرورت دارد که یک تابع فاصله مورد استفاده قرار گیرد (۲۶). مرسوم‌ترین نوع فاصله‌ها عبارت از فاصله اقلیدسی، مربع اقلیدسی، بلوک شهری (منهتن) و چپچف هستند (۲۳). از سویی در روش K-NN استفاده از فاصله اقلیدسی وزن‌دار شده بسیار مرسوم است (۴) (رابطه ۹).

$$Dis_t = \sqrt{W_t (X_{1r} - X_{1t})^2 + W_t (X_{2r} - X_{2t})^2 + \dots + W_t (X_{nr} - X_{nt})^2} \quad (۹)$$

در این مطالعه علاوه بر فاصله اقلیدسی از فاصله وزن‌دار بلوک شهری (رابطه ۱۰) نیز استفاده شد.

خاص به آن منجر به دریافت پاسخ خاص می‌شود، شبکه بر مبنای تطابق و هم‌سنجی بین ورودی و هدف سازگار می‌شود تا اینکه خروجی شبکه و هدف بر هم منطبق گردند. عموماً تعداد زیادی از این زوج‌های ورودی و خروجی به کار گرفته می‌شوند تا در این روند که از آن تحت عنوان یادگیری نظارت شده یاد می‌شود شبکه آموزش داده شود (۱۱). شبکه عصبی پرسپترون<sup>۲</sup> یکی از معروف‌ترین شبکه‌های عصبی است که در مدل چندلایه (MLP) دارای ۳ لایه شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی بوده که لایه پنهان خود می‌تواند بیش از یک لایه باشد و همچنین تعداد نورون‌های موجود در هر لایه متفاوت بوده و بر اساس ماهیت مسئله مورد بررسی با سعی و خطا تعیین می‌شود (۱۷)؛ بنابراین، تحت رویکردی دیگر، از مدل شبکه عصبی چندلایه با روش پس‌انتشار خطا (BP) با الگوریتم لوببرگ - مارکوئت<sup>۳</sup> و تابع محرک تان سیگموئید، مدل‌سازی و پیش‌بینی سری زمان جریان رودخانه کشکان انجام شد.

### آنالیز موجک

در گامی دیگر، جهت بهره‌گیری از کلیه اطلاعات سیگنال اصلی سری زمانی که از داده‌های خام در دسترس نیستند و همچنین جهت افزایش دقت مدل پیش‌بینی از تحلیل موجک جهت پیش‌پردازش و تجزیه سیگنال اصلی سری زمانی استفاده شد. چراکه تحلیل موجک قادر است بسیاری از منظرهای پنهان داده‌های که دیگر روش‌های تحلیل سیگنال در شناسایی آن‌ها ناتوان هستند را شناسایی کند. به عبارتی تبدیلات موجک قادر هستند تا اطلاعات مفیدی را که به‌سادگی از سیگنال اصلی قابل‌بازبایی نیستند را به دست آورند (۶). در این راستا چون تبدیل گسسته موجکی جهت پیش‌پردازش و تحلیل سری‌های زمانی مناسب‌تر از تبدیل پیوسته موجکی هست، بدین ترتیب که داده‌های تبدیل یافته با DWT<sup>۴</sup> اجزای اضافی نداشته و بنابراین تبدیل معکوس را می‌توان برای هر دسته از داده‌های زمان - بسامد به کاربرد (۸)، از تبدیلات موجک گسسته برای تجزیه سیگنال اصلی سری‌های زمانی رواناب استفاده شد.

اولین گام در تجزیه موجکی سیگنال سری زمانی، شناسایی موج مادر است. که به‌صورت رابطه ۱۳ تعریف می‌شود:

$$\Psi(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (13)$$

که  $\psi(t)$  تابع موجک مادر است که توابع مورد استفاده در تحلیل با دو عمل ریاضی انتقال و مقیاس در طول سیگنال مورد بررسی، تغییر اندازه و محل می‌یابند.

$$D_r = \sum_1^m \left[ \frac{1}{\sum_1^m \frac{Dis_t}{1}} \right] * D_t \quad (11)$$

در رابطه ۱۱، عبارت داخل کروشه یک تابع کرنل است و بیانگر مقدار احتمال وقوع یک پیش‌بینی برابر با میزان متغیر تاریخی ( $D_t$ ) در صورت مشاهده شرایط اقلیمی فعلی است. در این رابطه،  $m$  تعداد همسایه‌هایی است که برای پیش‌بینی اتخاذ شده‌اند. در الگوریتم پیش‌بینی با استفاده از نزدیک‌ترین همسایه، دو نوع پارامتر مهم وجود دارد که دقت پیش‌بینی را کنترل می‌کنند: اول تعداد همسایه‌ها ( $m$ ) و دوم وزن هر یک از متغیرها ( $w$ ) است که در نهایت رابطه پیش‌بینی به‌صورت رابطه ۱۲ توسط لال و شرما<sup>۱</sup> (۱۴) ارائه و بیان داشتند که مزیت کاربرد این روش بیش از رابطه ۱۱ است.

$$D_r = \sum_1^m \left[ \frac{\frac{1}{j}}{\sum_1^m \frac{1}{j}} \right] * D_t \quad (12)$$

در رابطه بالا  $j$  مرتبه هر یک از همسایه‌ها است. بدین ترتیب که هرچه فاصله مقادیر مستقل تاریخی از مقادیر مستقل مشاهداتی مطابق با رابطه ۱۲ کمتر باشد مرتبه آن بیشتر در نظر گرفته می‌شود.

### شبکه عصبی مصنوعی

مغز انسان حاوی میلیاردها نورون متصل به هم است که به سبب ساختاری که در آن نورون‌ها آرایش یافته و عمل می‌کنند، انسان‌ها قادرند که به سرعت الگوها را تشخیص داده و اطلاعات را پردازش کنند. یک شبکه عصبی مصنوعی طرح ریاضیاتی ساده شده‌ای از این شبکه عصبی زیستی است (۱۳). بنابراین در شبکه‌های عصبی مصنوعی سعی بر این است که ساختاری مشابه ساختار بیولوژیکی مغز انسان و شبکه اعصاب بدن ساخته شود تا همانند مغز، قدرت یادگیری، تعمیم‌دهی و تصمیم‌گیری داشته باشد و طرحی از یک فضای چند متغیره با استفاده از اطلاعات دریافتی تولید کند (۱۵). چنین شبکه‌های عصبی مصنوعی از عناصر ساده‌ای ترکیب شده‌اند که به‌طور موازی عمل می‌کنند. این عناصر از سیستم‌های عصبی زیستی الهام گرفته شده‌اند (۱۶). همچنانکه که در طبیعت، ساختار شبکه‌های عصبی از طریق نحوه اتصال بین عناصر (اجزاء) تعیین می‌شود، در شبکه‌های عصبی مصنوعی نیز با تنظیم مقادیر هر اتصال تحت عنوان وزن اتصال می‌توان نحوه ارتباط بین اجزای آن را تعیین کرد. پس از تنظیم یا همان آموزش شبکه عصبی، اعمال یک ورودی

2- Multi-Layer Perceptron(MLP)

3- Levenberg-Marquardt

4- Discrete Wavelet Transform

1- Lall and Sharma

$$R = \frac{\sum_{i=1}^n [(P_i - \bar{P})(O_i - \bar{O})]}{\sqrt{\sum_{i=1}^n [(P_i - \bar{P})^2 (O_i - \bar{O})^2]}} \quad (19)$$

$$CRM = \frac{\left[ \sum_{i=1}^n O_i - \sum_{i=1}^n P_i \right]}{\sum_{i=1}^n P_i} \quad (20)$$

در روابط بالا،  $O_i$  و  $\bar{O}$ : به ترتیب داده‌ها و میانگین داده‌های مشاهداتی،  $P_i$  و  $\bar{P}$  به ترتیب داده‌ها و میانگین داده‌های محاسباتی. در رابطه ۲۰ مقادیر مثبت ضریب CRM مبین گرایش مدل برای تخمین مقادیر کمتر از مقادیر مشاهداتی، مقدار منفی آن بیانگر گرایش مدل برای تخمین مقادیر بیشتر از مقادیر مشاهداتی و در این مدل، ایدئال مقدار این ضریب صفر است.

### نتایج و بحث

حافظه سیستم جریان رودخانه کشکان، با استفاده از نمایه هرست بافاصله ۱۰ روز و پایه میانه ۴۲۰۰ روز به مقدار ۰/۶ به دست آمد (شکل ۲). این مقدار گویایی رفتار غیرخطی و دینامیکی سیستم یاد شده بود و همچنین نشان از وجود حافظه بلندمدت در سری زمانی جریان رودخانه کشکان داشت. در ادامه، شبیه‌سازی جریان رودخانه کشکان با استفاده از روش‌های همچون K-NN و شبکه عصبی مصنوعی که مقید به پیش‌فرض‌های روش‌های خطی نیستند و از قابلیت کشف رفتار سیستم‌های غیرخطی برخوردارند انجام می‌گیرد. در گام نخست، با در نظرگیری سری‌های زمانی درج‌شده در سمت راست و چپ تابع سری زمانی (رابطه ۱) به ترتیب به‌عنوان ورودی و خروجی در مدل K-NN، ۸۰ درصد از حجم این داده‌ها برای آموزش مدل و ۲۰ درصد از آن‌ها برای آزمون مدل اختصاص پیدا کرد.

از سویی با اتخاذ دامنه‌ی از ۱ تا ۱۰ نزدیک‌ترین همسایه و دامنه‌ای از ۱۰۰۰ تا ۵۰۰۰۰ ذره (برای پوشش داده‌ها در بخش آموزش) این مدل تعبیه شد. آنگاه با ارزیابی معیارهای سنجش کارایی و دقت مدل در هر اجرا از دامنه‌های یادشده، بهترین مدل با ساختار ۶ همسایه و ۱۵۰۰۰ ذره به دست آمد. همچنین مدل ایجاد شده برحسب فاصله بلوک شهری از نتایج دقیق‌تری نسبت به فاصله اقلیدسی برخوردار بود. ارزیابی شناسه‌های سنجش عملکرد مدل نشان می‌دهد که K-NN مقادیر رواناب را اندکی بیشتر از رواناب مشاهداتی با میزان همبستگی ۰/۹۰ و خطای ۴/۶ درصد، شبیه‌سازی کرده است. همچنین نسبت تغییرات انحراف معیار داده‌ها نسبت به هر واحد از

$$\Psi_{s,t}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-t}{s}\right) \quad (14)$$

در رابطه ۱۴:  $t$  پارامتر انتقال و  $s$  پارامتر مقیاس است. برای محاسبه موجک گسسته بایستی به‌جای مقادیر  $t$  و  $s$  در رابطه ۱۴ مقدار گسسته آن‌ها به ترتیب  $kS_0^j \tau_0$  و  $S_0^j$  جایگذاری شود.

$$\Psi_{k,j}(t) = \frac{1}{\sqrt{S_0^j}} \Psi\left(\frac{t - kS_0^j \tau_0}{S_0^j}\right) \quad (15)$$

در رابطه ۱۵:  $j$  و  $k$  اعداد صحیح هستند.  $S_0 > 1$  یک گام تأخیری ثابت شده است.  $\tau_0$  فاکتور تبدیل است که به گام تأخیر وابسته است. با ساده کردن رابطه فوق رابطه ۱۶ به صورت زیر حاصل می‌شود.

$$\Psi_{k,j}(t) = S_0^{-j/2} \Psi(S_0^{-j} t - k \tau_0) \quad (16)$$

با کاربرد رابطه ۱۶، ضرایب گسسته موجک طبق رابطه ۱۷ به دست می‌آیند.

$$DWT_s^\Psi(t) = \sum s(t) \Psi_{k,j}(t) dt \quad (17)$$

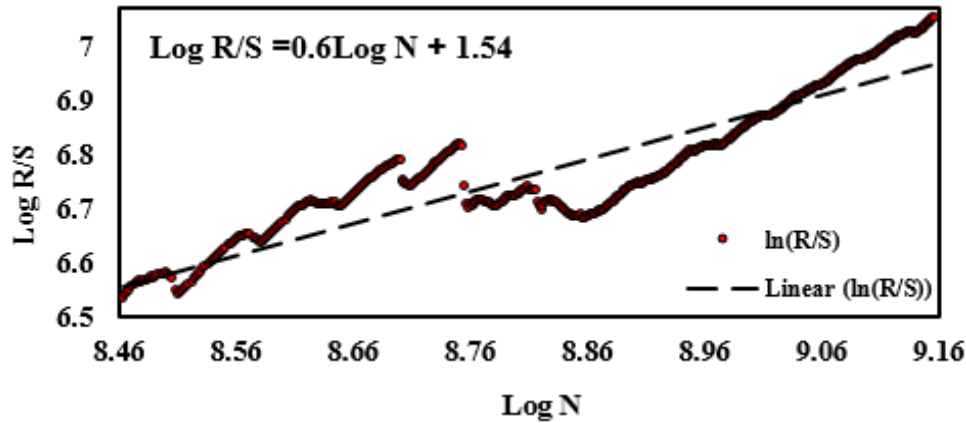
در تبدیل موجک گسسته، سیگنال از یک سری فیلترهای پایین‌گذر و بالاگذر عبور داده می‌شود که بخش به‌دست آمده از فیلتر بالاگذر حاوی اطلاعات فرکانس بالا است که جزئیات نامیده می‌شود. از سویی بخش حاصل از فیلتر پایین‌گذر حاوی اطلاعات فرکانس پایین و دربردارنده ویژگی‌های اصلی سیگنال است که کلیات نامیده می‌شود (۸). در ادامه به سبب شباهت موجک مادر میر (موجک پیچیده) به سیگنال سری زمانی رواناب، تجزیه این سیگنال‌ها با استفاده از تابع موجک میر انجام گرفت. برحسب رابطه ارائه شده توسط وانگ و دینگ  $L = \text{INT}[\log(N)]$  (L: درجه تجزیه، INT: تابع گردکننده و N: تعداد داده‌های سری زمانی) تعداد سطح تجزیه مناسب به مقدار ۴ سطح به دست آمد؛ بنابراین سیگنال رواناب به ۵ زیر سیگنال شامل ۴ سیگنال فرکانس بالا و یک سیگنال فرکانس پایین تجزیه شد که این ۵ سیگنال با تأخیرهای زمانی که در سمت راست تابع رواناب جای دارند به صورت ترکیبی با مدل K-NN و شبکه عصبی ترکیب شده و عمل شبیه‌سازی رواناب با احتساب به این مدل‌های ترکیبی نیز انجام گرفت. در آخر به منظور ارزیابی، مقایسه و گزینش دقیق‌ترین ساختار مدل رواناب از شاخص‌های ضریب همبستگی (R)، جذر میانگین مربعات خطا (RMSE) و ضریب CRM استفاده شد (رابطه‌های ۱۸، ۱۹ و ۲۰)

$$RMSE = \left[ n^{-1} \sum_{i=1}^n (O_i - P_i)^2 \right]^{0.5} \quad (18)$$

- 1- Details
- 2- Approximation

بهترین مدل برازشی شد. در این ساختار ۶، ۸ و ۱ به ترتیب به‌عنوان تعداد نورون ورودی، نورون لایه پنهان و نورون خروجی محسوب می‌شوند. معیارهای سنجش کارایی این مدل نیز حاکی از عملکرد نسبتاً قابل قبول آن در شبیه‌سازی رواناب است. به‌طوری‌که سری زمانی شبیه‌سازی شده دارای ضریب همبستگی ۰/۸۹ با سری زمانی مشاهداتی بوده و خطای پیش‌بینی آن در حدود ۵/۸ درصد است (شکل‌های ۵ و ۶).

میانگین آن‌ها در سری زمانی شبیه‌سازی شده همسان با سری زمانی داده‌های مشاهداتی است. بدین ترتیب که پراکندگی این دو سری نسبت به میانگین یکسان بوده و از توزیع برابری برخوردارند؛ بنابراین جداسازی داده‌ها در این مدل به‌خوبی انجام شده است (جدول ۱، شکل‌های ۳ و ۴). معماری شبکه عصبی مصنوعی جهت شبیه‌سازی رواناب با ۶ نورون ورودی و در نظرگیری ۳ تا ۲۰ نورون در یک لایه پنهان و یک نورون خروجی منتهی به ساختار ۶-۸-۱ به‌عنوان

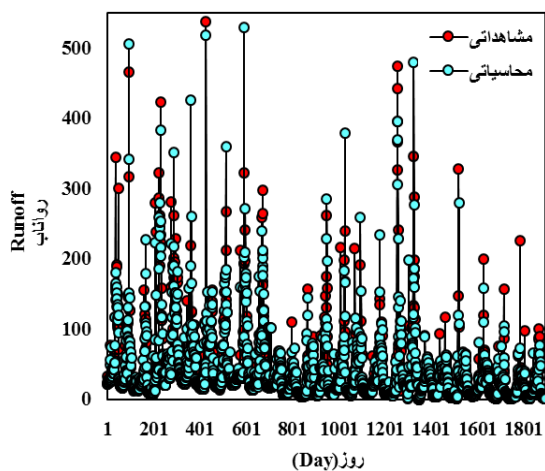


شکل ۲- نمودار هرست سری زمانی جریان رودخانه کشکان (۱۳۷۰-۱۳۹۳)  
Figure 2- Hurst plot of time series flow for the Kashkan river

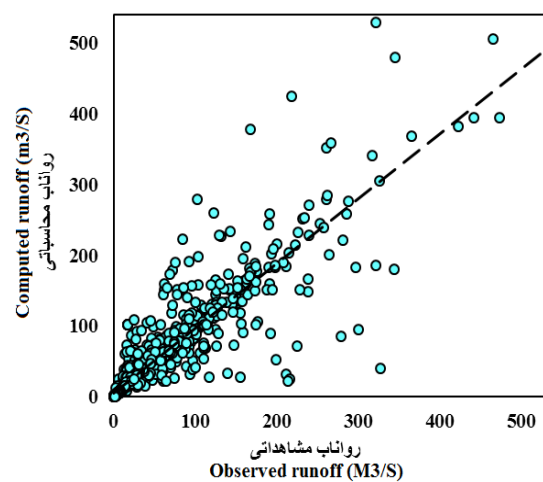
جدول ۱- نتایج عملکرد مدل K- نزدیک‌ترین همسایه (K-NN) و شبکه عصبی (ANN) در شبیه‌سازی رواناب کشکان

Table 1- Performance results for K-nearest neighbor (K-NN) and neural network (ANN) in Kashkan runoff modeling

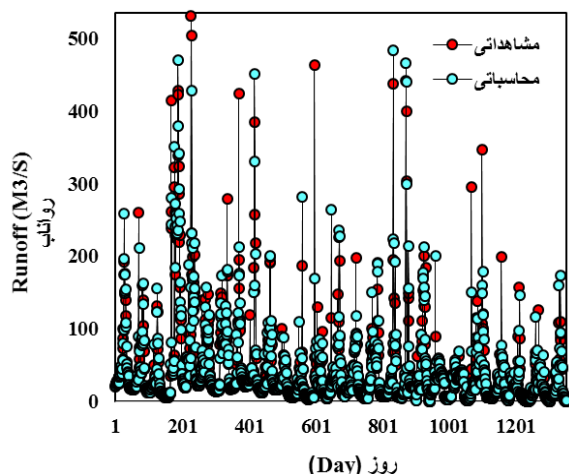
مدل Model	میانگین مشاهداتی Observed Mean (M <sup>3</sup> /s)	میانگین محاسباتی Computed Mean (M <sup>3</sup> /s)	انحراف معیار مشاهداتی Observed Std deviation	انحراف معیار محاسباتی Computed Std deviation	R	RMSE (M <sup>3</sup> /s)	CRM
K-NN	39.87	41.33	53.13	54.23	0.9	24.1	-0.037
ANN	42.03	42.63	60.38	57	0.89	27.88	-0.014



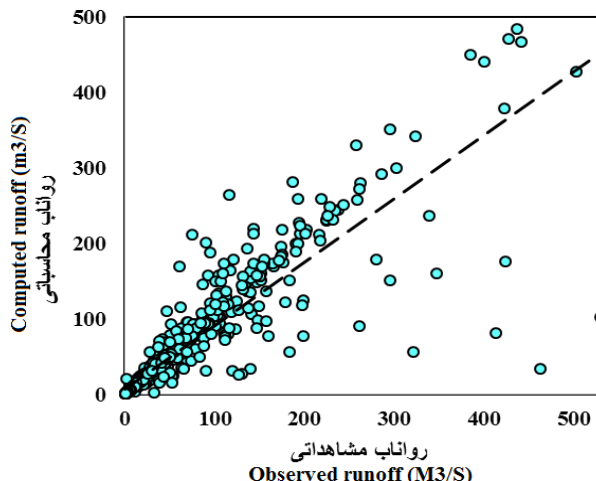
شکل ۴- نمودار مقایسه رواناب مشاهداتی و محاسباتی (K-NN)  
Figure 4- Comparison of the observed and computed runoff (K-NN)



شکل ۳- نمودار پراکندگی رواناب مشاهداتی و محاسباتی (K-NN)  
Figure 3- Scatter plot for observed and computed runoff (K-NN)



شکل ۶- نمودار مقایسه رواناب مشاهداتی و محاسباتی (ANN)  
Figure 6- Comparison of the observed and computed runoff (K-NN)



شکل ۵- نمودار پراکندگی رواناب مشاهداتی و محاسباتی (ANN)  
Figure 5- Scatter plot for observed and computed runoff (K-NN)

NN و ANN قلمداد شدند که در ادامه با قرارگیری در سمت راست تابع رواناب (رابطه ۶) منتهی به شکل گیری دو مدل ترکیبی تحت عنوان K-NN-WT و ANN-WT خواهند شد.

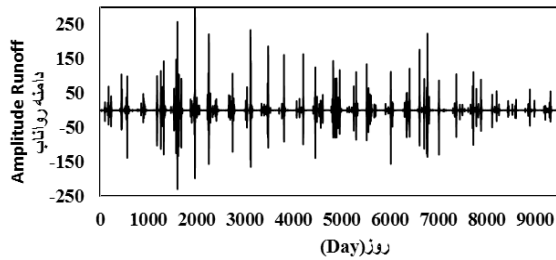
در این گام، با اختصاص ۱ تا ۲۰ همسایه و پوششی از ۱ تا ۱۰۰۰۰۰ ذره (در بخش آموزش مدل) رویکرد K-NN-WT در این بازه‌ها اجرا شد. همچنانکه از سنجش‌های کارایی عملکرد مشخص است مدل با ساختار ۶ همسایه، ۵۰۰۰۰ ذره و فاصله اقلیدسی بهترین عملکرد را در شبیه‌سازی رواناب در قیاس با سایر ساختارهای این مدل نشان می‌دهد (جدول ۲). در این خصوص، مدل یاد شده مقادیر رواناب را برخلاف مدل K-NN کمتر از مقادیر مشاهداتی و با خطای کمتری به مقدار ۲/۷ درصد شبیه‌سازی کرد.

چنانکه از شکل‌های ۱۳ و ۱۴ محرز است جداسازی داده‌ها و دقت رواناب شبیه‌سازی شده در این مدل بهبود یافته و پراکنش مقادیر رواناب نسبت به میانگین در سیگنال شبیه‌سازی شده کمتر از همین مورد در سیگنال مشاهداتی رواناب حاصل شده است. از سویی تقارن خطاها و مقدار مطلق آن‌ها در مدل K-NN-WT نسبت به مدل K-NN اندکی کاهش پیدا کرده بدین مفهوم که مدل ترکیبی توانسته است مقادیر فراوانی از خطاهای مطلق را کاهش دهد اما در برخورد با فرکانس‌های متفاوت سری زمانی رفتار همگنی مانند مدل K-NN نشان نمی‌دهد. به‌واقع این فرآیند نشان از ضعف نسبی این مدل ترکیبی در قیاس با مدل K-NN نداشته زیرا این مدل توانسته مقادیر مطلق خطای که در تمامی فرکانس‌های رواناب ایجاد شده بود را کاهش دهد؛ بنابراین چنین عدم تقارنی به سبب مقادیر کمینه خطای رواناب شبیه‌سازی شده از توجه ویژه‌ای برخوردار نیست هرچند که در نظرگیری آن جهت بیشتر عدم قطعیت ضرورت دارد.

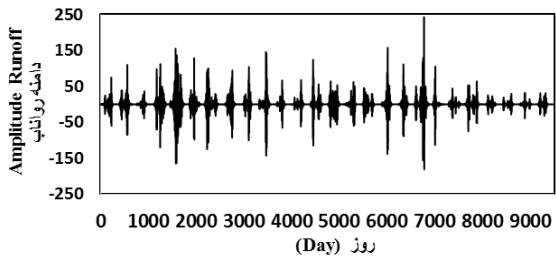
مدل ANN نیز مقدار رواناب را اندکی بیش از رواناب مشاهداتی و با خطای بیشتری نسبت به مدل K-NN تخمین زده است. از سویی رواناب شبیه‌سازی شده توسط مدل ANN نسبت به میانگین دارای پراکنش بیشتری در قیاس با سری زمانی رواناب مشاهداتی است؛ بنابراین چنین پراکنشی موجب افزایش خطای این مدل و عدم تقارن شدید خطاها شده است. مقایسه نمودارهای پراکنش و معیارهای سنجش عملکرد دو مدل یادشده نشان می‌دهد که مدل K-NN فراوانی رواناب‌های کمتر از میانگین را اندکی افزایش داده در صورتی مدل ANN فراوانی رواناب‌های کمتر از میانگین را کاهش داده است. بدین عنوان که مدل K-NN واکنش نسبتاً همگن تری در برابر رواناب‌های بیشینه و کمینه نشان می‌دهد. در صورتی که مدل ANN رواناب‌های کمینه را بسیار بیشتر و رواناب‌های بیشینه را بسیار کمتر از مدل مذکور شبیه‌سازی کرده است.

به‌طور کلی این دو رویکرد به‌رغم ارائه شبیه‌سازی نسبتاً قابل قبول رواناب روزانه رودخانه کشکان اما در برخورد با داده‌های فرکانس بالا و فرین عملکرد چندان مناسبی نشان ندادند. چراکه در چنین مواردی دستخوش خطاهای بالای می‌شدند. به‌واقع، این فرآیند به ماهیت داده‌های روزانه جریان رودخانه برمی‌گردد که سیگنال سری زمانی آن از رفتارهای روزانه، ماهانه، فصلی و سالانه برخوردار بوده و هرکدام از این رفتارها با توجه به دینامیکی که دارند وزن متفاوتی در تعیین رفتار آینده رواناب بر عهده دارند؛ بنابراین به‌منظور فائق آمدن به چنین مسئله‌ای، سیگنال اصلی سری زمانی رواناب (شکل ۷) با استفاده از تابع موجک میر به ۴ سطح شامل ۸ زیرسیگنال فرکانس بالا و فرکانس پایین تجزیه شد. بدین ترتیب که ۴ زیرسیگنال فرکانس بالا (شکل‌های ۸، ۹، ۱۰ و ۱۱) و یک سیگنال فرکانس پایین سطح ۴ (شکل ۱۲) به‌جای سیگنال اصلی رواناب به‌عنوان ورودی مدل K-

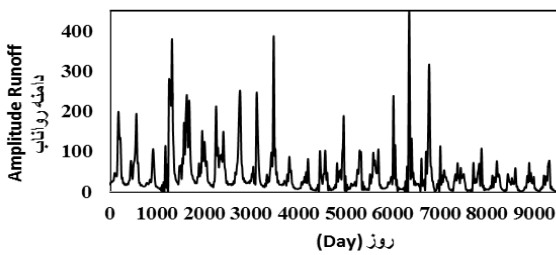




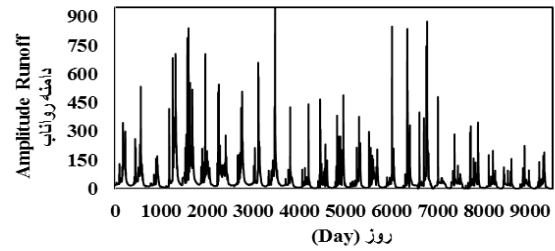
شکل ۸- سیگنال فرکانس بالای رواناب در سطح ۱  
Figure 8- Detail signal of runoff in level 1



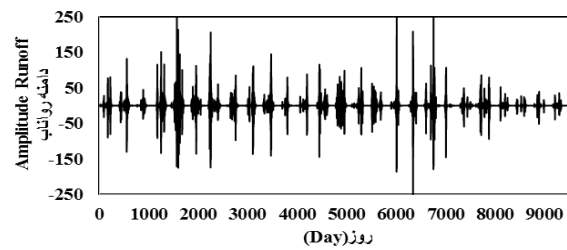
شکل ۱۰- سیگنال فرکانس بالای رواناب در سطح ۳  
Figure 10- Detail signal of runoff in level 3



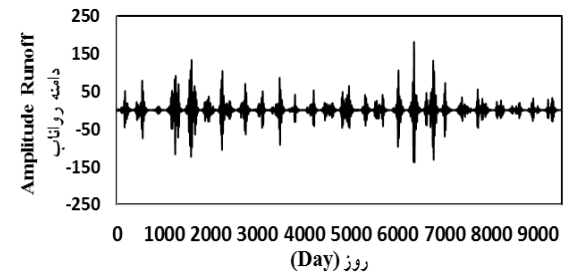
شکل ۱۲- سیگنال فرکانس پایین رواناب در سطح ۴  
Figure 12- Approximation signal of runoff in level 4



شکل ۷- سیگنال اصلی رواناب  
Figure 7- Original signal of runoff



شکل ۹- سیگنال فرکانس بالای رواناب در سطح ۲  
Figure 9- Detail signal of runoff in level 2



شکل ۱۱- سیگنال فرکانس بالای رواناب در سطح ۴  
Figure 11- Detail signal of runoff in level 4

رواناب قادر نبود واکنش مناسبی را نشان دهد در مدل ANN-WT مرتفع شد.

بنابراین ترکیب آنالیز موجک با شبکه عصبی در شبیه‌سازی سیگنال رواناب موجب انعطاف‌پذیری بیشتر شبکه عصبی شده است. چنانکه مشخص شد سیگنال سری زمانی رواناب کشکان از ماهیت دینامیکی برخوردار بوده و دارای کمینه‌ها و بیشینه‌های محلی است که این فرآیند باعث شده که مدل ANN قادر به تشخیص مناسب چنین کمینه و بیشینه‌های محلی در مقدار رواناب نباشد در صورتی که با جداسازی این سیگنال‌ها از سیگنال اصلی رواناب، عملکرد و کشف اطلاعات زودگذر و بلندمدت رواناب با وزن مناسب در دسترس شبکه عصبی قرار گرفتند.

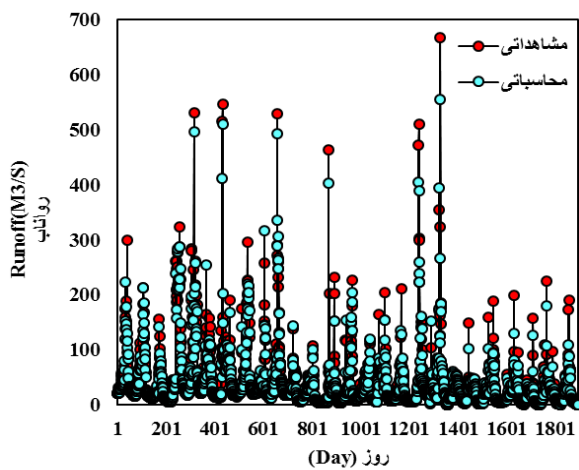
معماری مدل ترکیبی شبکه عصبی - آنالیز موجک نیز با احتساب به ۳۰ نورون ورودی و سه لایه پنهان انجام گرفت که ارزیابی سنج‌های کارایی عملکرد مدل نشان دادند که ساختار ۳۰-۷-۱ با ۱/۲ درصد خطای پیش‌بینی و ضریب همبستگی ۰/۹۸۹ بهترین شبیه‌سازی رواناب را در قیاس با سایر ساختارها این مدل ترکیبی بر عهده دارد (شکل‌های ۱۵ و ۱۶). چنانکه از پراکنش داده‌های رواناب مشخص است دقیق‌ترین جداسازی داده‌ها در این مدل اتفاق افتاده است. از سویی این مدل نیز مقادیر رواناب شبیه‌سازی شده را بیشتر از مقادیر رواناب مشاهداتی برآورد کرده است.

به‌واقع مسئله‌ای که در رویکرد ANN اتفاق افتاده بود بدین ترتیب که خطای رواناب شبیه‌سازی شده دستخوش عدم تقارن شدیدی شده بود یعنی اینکه مدل در برخورد با فرکانس‌های متفاوت

جدول ۲- نتایج عملکرد مدل ترکیبی K- نزدیک ترین همسایه - آنالیز موجک (K-NN-WT) و مدل ترکیبی شبکه عصبی - آنالیز موجک (ANN-WT) در مدل سازی رواناب کشکان

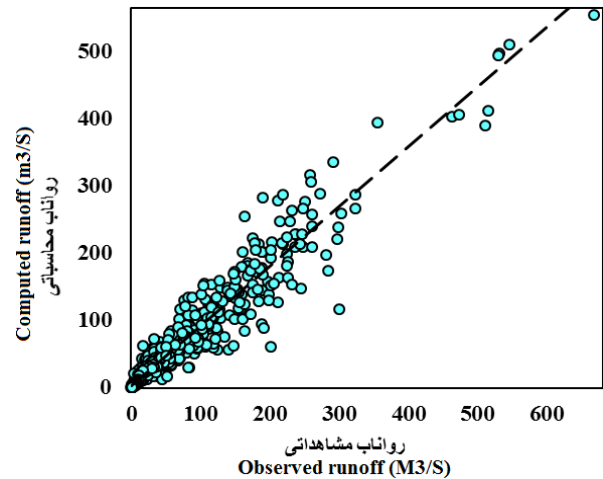
Table 2- Performance results for K-nearest neighbor and wavelet hybrid model (K-NN-WT) and neural network and wavelet hybrid model (ANN-WT) in Kashkan runoff modeling

مدل Model	میانگین مشاهداتی Observed Mean (M <sup>3</sup> /s)	میانگین محاسباتی Computed Mean (M <sup>3</sup> /s)	انحراف معیار مشاهداتی Observed Std deviation	انحراف معیار محاسباتی Computed Std deviation	R	RMSE (M <sup>3</sup> /s)	CRM
K-NN- WT	41.44	39.86	58.24	53.5	0.967	15.2	0.038
ANN-WT	43.8	45.3	65.27	63.34	0.989	9.9	0.035



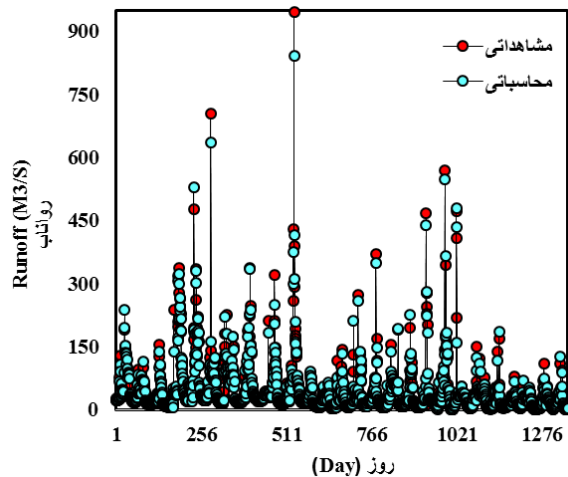
شکل ۱۴- نمودار مقایسه رواناب مشاهداتی و محاسباتی (K-NN-WT)

Figure 14- Comparison of the observed and computed runoff (K-NN-WT)



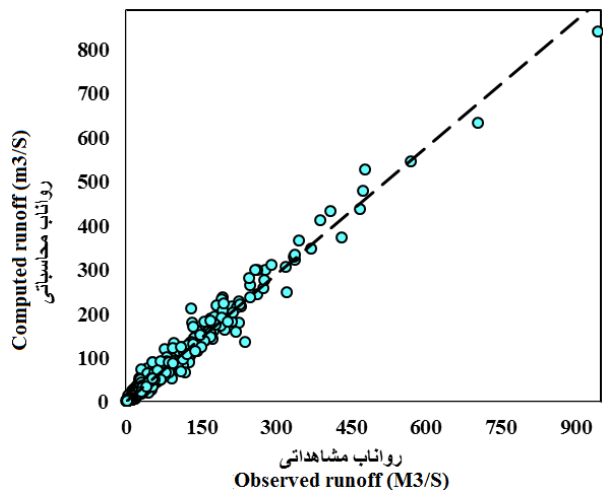
شکل ۱۳- نمودار پراکندگی رواناب مشاهداتی و محاسباتی (K-NN-WT)

Figure 13- Scatter plot of observed and computed runoff (K-NN-WT)



شکل ۱۶- نمودار مقایسه رواناب مشاهداتی و محاسباتی (ANN-WT)

Figure 16- Comparison observed and computed runoff (ANN-WT)



شکل ۱۵- نمودار پراکندگی رواناب مشاهداتی و محاسباتی (ANN-WT)

Figure 15- Scatter plot observed and computed runoff (ANN-WT)

## نتیجه‌گیری

چنین خطاهای نبوده ولی این مدل، عملکرد همگنی مقابل داده‌ها داشته و پراکندگی مقادیر رواناب را به‌درستی تشخیص داده بود. بنابراین به‌منظور افزایش دقت دو مدل یادشده و استخراج اطلاعات پنهان سیگنال رواناب که در فرکانس‌های متفاوت آن نهفته‌اند، ابتدا سیگنال رواناب در ۴ سطح تجزیه شد که این ۴ سطح به‌نوعی بیانگر فرکانس‌های روزانه، ماهانه و سالانه رواناب کشکان بودند. به این ترتیب، رفتارهای متفاوت سیگنال سری زمانی این رودخانه در دسترس مدل‌های یاد شده قرار گرفت و دو مدل ترکیبی ANN-WT و K-NN-WT اجرا شدند. نتایج به‌دست آمده از رواناب شبیه‌سازی شده توسط این مدل‌های ترکیبی حکایت از افزایش ضریب دقت و بهبود خروجی آن‌ها نسبت به مدل‌های منفرد داشت. به‌واقع تجزیه سیگنال اصلی رواناب توسط آنالیز موجک باعث شده بود که در فرآیند شبیه‌سازی، وزن‌دهی مناسبی به دینامیک‌های بلندمدت، میان‌مدت و کوتاه‌مدت سیگنال رواناب اختصاص پیدا کند و خطای ناشی از مدل‌سازی به نحو برجسته‌ای کاهش پیدا کند. به‌طوری‌که این دو مدل ترکیبی به ترتیب رواناب را به مقدار خطای ۱/۲ و ۲/۷ درصد شبیه‌سازی کردند.

به‌طورکلی مدل ترکیبی ANN-WT بهترین عملکرد را در قیاس با سایر مدل‌ها نشان داد. به‌واقع مسئله‌ای که مدل ANN با آن روبه‌رو بود با ترکیب آنالیز موجک حل شد. به‌طورکلی این نتایج حاکی از آن هستند که چون سیگنال روزانه رواناب رودخانه کشکان از حالت غیر ایستا و دینامیک پیچیده‌ای برخوردار بوده و همچنین دارای دینامیکی با فرکانس‌های متعدد زمانی است لذا استخراج این فرکانس‌ها می‌تواند به‌طور مؤثری منجر به بهبود مدل‌های شبیه‌سازی و کاهش عدم قطعیت شود.

سری‌های زمانی داده‌های روزانه جریان آب رودخانه اغلب از دینامیک پیچیده‌ای برخوردارند بدین مفهوم که فرکانس‌های متفاوتی در رفتار آن‌ها نهفته است که روش‌های آماری و هوشمند به‌تنهایی قادر به کشف این فرکانس‌های نیستند. این مطالعه به ارزیابی چگونگی عملکرد ترکیب دو مدل شبکه عصبی مصنوعی (ANN) و K-NN نزدیک‌ترین همسایه (K-NN) با آنالیز موجک در شبیه‌سازی سیگنال سری زمانی رواناب رودخانه کشکان تمرکز یافته بود. آزمون حافظه و رفتار سیگنال اصلی این سری زمانی با استفاده از نمایه هرست به مقدار ۰/۶ نشان داد که رواناب رودخانه کشکان دارای رفتاری دینامیکی بوده و از حافظه بلندمدت جهت به‌کارگیری مدل‌های سری زمانی برخوردار است. در این راستا با در نظرگیری تابعی که سیگنال سری زمانی رواناب را به ۶ سیگنال از سری‌های زمانی این رواناب با تأخیرهای ۱، ۳، ۵، ۷، ۱۰ و ۱۵ وابسته کرده بود، فرآیند شبیه‌سازی رواناب با اتکای به دو مدل یادشده انجام گرفت. ارزیابی سنج‌های کارایی عملکرد مدل حاکی از این بود که هر دو مدل در مجموع، مقدار رواناب را بیشتر از مقادیر مشاهداتی شبیه‌سازی کرده‌اند.

از سویی مدل K-NN با خطای ۴/۶ درصد عملکرد بهتری نسبت به مدل ANN با خطای ۵/۸ درصد نشان داد. به‌واقع مدل دومی از واکنش همگنی در مقابل مقادیر محلی واقع در سیگنال رواناب برخوردار نبود یعنی اینکه فرکانس‌های روزانه، ماهانه، دوره‌ای و سالانه‌ای که در سیگنال رواناب رودخانه کشکان قرار دارند باعث به وجود آمدن این رفتار غیر همگن در شبکه عصبی شده‌اند. در این راستا، شبکه عصبی در مواجهه با چنین رفتارهای غیر سینوسی و محلی که جزء ماهیت سیگنال رواناب کشکان هستند متحمل خطاهای بسیار شدید شده بود. هرچند که مدل K-NN نیز مبرا از

## منابع

- 1- Abdollahi Asadabadi S., Dinpashoh Y., and Mirabbasi R. 2014. Forecasting of mean daily runoff discharge of behesht-abad River using wavelet analysis. *Journal of Water and Soil*, 28(3):534-545. (in Persian with English abstract)
- 2- Akhtar M.K., Corzo G.A., Van Andel S.J., and Jonoski A. 2009. River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin. *Hydrol. Earth Syst. Sci*, 13:1607-1618.
- 3- Anis Hosseini M., and Zaker Mashgh M. 2013. Analysis and forecasting of river flow kashkan using chaos theory. *Journal of Hydrolic*, 8(3):45-61. (in Persian with English abstract)
- 4- Azmi M., and Araghinejad. 2012. Development of K-Nearest Neighbour regression method in forecasting river stream flow. *J. of Water and Wastewater*, 2:108-119. (in Persian with English abstract)
- 5- Cannas B., Fanni A., See L., and Sias G. 2006. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Phys. Chem. Earth*, 31(18): 1164-1171.
- 6- Daubechies I. 1992. Ten lectures on wavelets. Society for Industrial Mathematics.
- 7- Haghizadeh A., Mohammadlo M., and Nouri F. 2015. Modeling rainfall – runoff process using artificial neural

- network and Neuro-Fuzzy Computing and multiple regression (case study: watershed of Korramabad). *Journal of Eco hydrology*, 2:233-243. (in Persian with English abstract)
- 8- Hassanzadeh Y., Lotfollahi M.A., Shahverdi S., Farzin S., and Farzin N. 2013. De-noising and prediction of time series based on the wavelet algorithm and chaos theory (Case Study: SPI drought monitoring index of tabriz city), *Iran-Water Resources Research*, 8(3):1-13. (in Persian with English abstract)
  - 9- Hurst H.E. 1951. Long-term storage capacity of reservoirs (with discussion). *Transactions of the American Society of Civil Engineers*, 116: 770-808.
  - 10- Karamuz M., and Araghinejad Sh. 2014. *Advanced hydrology*. AmirKabir University. Iran
  - 11- Kia, M. 2010. *Neural networks in matlab*. Qian academic publishing.
  - 12- Kisi O. 2007. Streamflow forecasting using different artificial neural network algorithms. *ASCE Journal of Hydrologic Engineering*, 12(5):532-539.
  - 13- Kisi O. 2005. Daily river flow forecasting using artificial neural networks and auto-regressive models. *Turkish J. Eng. Env. Sci*, 29:9-20.
  - 14- Lall U., and Sharma A. 1996. A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, 32(3):679-694.
  - 15- Lee S., Ryu J.H., Lee M.J., and Won J.S. 2006. The Application of artificial neural networks to landslide susceptibility mapping at Janghung, Korea. *Mathematical Geology*, 38(2):199-220.
  - 16- Mark H.B., Martin T.H., and Haward B.D. 2016. *Neural network toolbox™ getting started guide*. The MathWorks, Inc.
  - 17- Menhaj M. 2002. *Neural networks and artificial intelligent basic*. First edition AmirKabiruniversity. Press, 350p.
  - 18- Montaseri M., and Zamanzad Ghavidel S. 2014. River Flow Forecasting by Using Soft computing *Journal of Water and Soil*, 28 (2):394-405. (in Persian with English abstract)
  - 19- Nayak P.C., Sudheer K.P., Rangan D.M., and Ramasastri K.S. 2004. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1):52-66.
  - 20- Pramanik N., and Panda R.K. 2009. Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. *Hydrological sciences journal*, 54(2): 247-260.
  - 21- Sanikhani H., Dinpashoh Y., and Ghorbani M.A. 2014. Baranduz-chay river flow modeling using the K-nearest neighbor and intelligent methods. *Water and Soil Science*, 25(1):219-233.
  - 22- Shafaei M., Fakheifard A., Darbandi S., and Ghorbani M.A. 2013. prediction daily flow of vanyar station using ANN and wavelet hybrid procedure. *Irrigation & Water Engineering*, 14:144-128. (in Persian with English abstract)
  - 23- Shataee Sh., Kalbi S., Fallah A., and Pelz D. 2012. Forest attributes imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *International Journal of Remote Sensing*, 33(19):6254-6280. (in Persian with English abstract)
  - 24- Veiga V.B., Hassan Q.K., and He J. 2015. Development of Flow Forecasting Models in the Bow River at Calgary, Alberta, Canada. *Journal Water*, 7:99-115.
  - 25- Wang W., and Ding J. 2003. Wavelet network model and its application to the prediction of hydrology. *Nature and Science*, 1(1):67-71
  - 26- Wilson D. R., and Martinez T. R. 2000. Reduction techniques for exemplar-based learning algorithms. *Machine Learning*, 38(3): 257-286.
  - 27- Wu C.L., and Chau K.W. 2010. Data-driven models for monthly streamflow time series prediction. *Engineering Applications of Artificial Intelligence* 23:1350-1367.
  - 28- Yates D., Gangopadhyay S., Rajagopalan B., and Strzepek K. 2003. A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resoures Research*, 39 (7): 1114- 1121.
  - 29- Young C.C., Liu W.C., and Chung C.E. 2015. Genetic algorithm and fuzzy neural networks combined with the hydrological modeling system for forecasting watershed runoff discharge. *The Natural Computing Applications*, 1-13.



## Forecasting Kashkan River Flow using a Combination of Artificial Neural Network, Wavelet Analysis and K - Nearest Neighbor

D. Yarahmadi<sup>1</sup> - H. Mirhashemi<sup>2\*</sup>

Received: 23-05-2017

Accepted: 06-11-2017

**Introduction:** Forecasting and modeling of river flow is an essential step towards planning, designing and utilizing water resources management system which is subject to issues such as droughts and destructive floods in river basins. The river flow deficit and excess could result in financial and human losses. Such predictions of river flow not only provide the necessary warning signals about the flood risk, but also help to adjust the water outflows during low level of water flows which help to the water resource management. Due to the importance of river flows and its fluctuations in short and long term on different aspects of human lives, understanding its behavior and performance is crucial (necessary). Thus, with discovering its dynamic behavior, it is possible to predict its future performance. The aim of this study is to explore and simulation of Kashkan River's performance using the statistical intelligent methods to provide models with lower uncertainty in order to improve the planes based on Kashkan's River flows.

**Materials and Methods:** For this study, the series of daily discharge data from Poldokhtar- Kashkan station (located in the coastal river) over 1370-1393 were used as the primary input. Methods used in this study were based on memory uses the Hurst exponent of long memory time series. Runoff is the dynamics of the series. The current state of these series is dependent on its historical states. The delay time (lag time) of 1, 3, 5, 7, 10 and 15 days before the runoff were calculated. The amount of runoff was seen as a function of the time series. Considering the above-mentioned six time series as input signals, time series modeling using statistical methods K- nearest neighbor (K-NN), and artificial neural network, combined wavelet - K-NN and combining the wavelet nervous.

**Results and Discussion:** Kashkan's Memory river flow system, using the Hurst exponent within 10 days and mid-4200 based on the amount of 0.6 was obtained (Figure 2). This amount indicates a non-linearly behavior and a dynamic learning system. In addition, it shows the presence of long memory in the river flow time series. Then, by allocating 80% of the data for training and the remaining 20 percent for testing the model and adopting ranges from 1 to 10 nearest neighbor and a range of 1,000 to 50,000 particles (for data on education) Model K-NN were prepared. Using the criteria to assess the efficiency and accuracy of a model in each performance of the mentioned domains, the best model with the 6 neighbors structure and 15,000, was obtained. In this model stimulated the runoff with the correlation of 0.90 and a 4.6 error was obtained. On the other hand, artificial neural network architecture to simulate runoff with 6 input neurons in a hidden layer neurons and considering 3 to 20 and an output neurons leading to the 6-8-1 structure as the best model was fitted. This model has a correlation of 0.89 and the forecast error of 5.8 in the process of runoff simulation. Then using wavelet function, mortality, time-series signal runoff into 4 levels, including 8 under high frequency and low frequency signal was decomposed where high-frequency signals and low-frequency signal of 4 level were considered as the original signal for the input surface runoff. In this regard, the hybrid model K-NN-WT with runoff time series prediction error of 2.7 percent and the hybrid model ANN-WT with the correlation of 0.99 the estimation error of 1.2 were simulated.

**Conclusion:** Running 4 Artificial Neural Network (ANN), K-nearest neighbor (K-NN) and combining the wavelet analysis of the two models (ANN-WT and K-NN-WT) to predict the time series of runoff river showed that due to the existence of multiple time frequencies in the time series of the river signals, its decomposition it using wavelet analysis results in extraction of hidden information that are not available through the original signal. This information is the daily, monthly, quarterly and annual fluctuations. The hybrid models performance

1- Assistant Professor in Climatology, Geography Sciences Department, Lorestan University

2- Ph.D. of Climatology and Lecturer of Geography Sciences Department, Lorestan University

(\*- Corresponding Author Email: climate90@yahoo.com)

indicated higher accuracy and improved outcomes relative to individual models. In fact, the analysis of the original runoff signal by wavelet analysis in the process of simulation results in an appropriate weighting given to long-term and short term dynamic of runoff which led to significant lower error in modeling.

**Keywords:** Kashkan River, Hurst exponent, Myer wavelet, Time series signal