



## ارزیابی عملکرد توابع کرنل در تخمین جریان رودخانه‌ها با استفاده از ماشین بردار پشتیبان

\* محمد عیسی‌زاده<sup>۱</sup>، حجت احمدزاده<sup>۲</sup> و محمدعلی قربانی<sup>۳</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد گروه مهندسی آب، دانشگاه تبریز، <sup>۲</sup> دانشجوی دکتری گروه مهندسی آب، دانشگاه تبریز،

<sup>۳</sup> دانشیار گروه مهندسی آب، دانشگاه تبریز

تاریخ دریافت: ۹۳/۱۲/۲۲؛ تاریخ پذیرش: ۹۴/۱۰/۲۰

### چکیده

**سابقه و هدف:** پیش‌بینی دقیق رواناب رودخانه‌ها نقش مهمی در مدیریت بهینه منابع آب در دسترس دارد. در سال‌های اخیر، ماشین بردار پشتیبان (SVM) که یکی از مهم‌ترین مدل‌های داده‌کاوی است برای این منظور مورد توجه قرار گرفته است. این مدل یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد. همانند مدل‌های داده‌کاوی دیگر مدل SVM نیز می‌تواند در مواقعی که فقط داده‌های رواناب در دسترس می‌باشد جهت شبیه‌سازی رواناب مورد استفاده قرار گیرد (مدل‌سازی خودهمبسته). به‌طور معمول سه تابع کرنل پایه شعاعی (RBF)، چندجمله‌ای درجه  $d$  و خطی در ماشین بردار پشتیبان مورد استفاده قرار می‌گیرند که کاربرد هر یک از این توابع با پارامترهای مختلف در تخمین رواناب رودخانه‌ها ممکن است منجر به نتایج متفاوتی شود. بنابراین ارزیابی کارایی و دقت هر یک از این توابع و انتخاب تابع کرنل مناسب در پیش‌بینی جریان رودخانه ضروری است. همچنین از آنجا که مدل‌های سری زمانی AR، ARMA و ARIMA از مدل‌های اصلی در شبیه‌سازی خودهمبسته رواناب می‌باشند بنابراین می‌توان از طریق مقایسه عملکرد هر یک از توابع کرنل با این مدل‌ها، دقت نسبی این توابع در این زمینه را مورد بررسی قرار داد. بنابراین ارزیابی دقت هر یک از توابع کرنل در شبیه‌سازی رواناب ماهانه و مقایسه عملکرد آن‌ها با مدل‌های سری زمانی هدف اصلی این پژوهش را رقم می‌زند.

**مواد و روش‌ها:** در این پژوهش حوضه خرخره‌چای به‌عنوان منطقه مورد مطالعه انتخاب شده و جریان ماهانه مشاهداتی این حوضه در ایستگاه آب‌سنجی سسته جهت واسنجی و اعتبارسنجی مدل‌ها به‌کار گرفته شد. برای این منظور، در ابتدا ۷۵ درصد از داده‌های جریان ماهانه (۱۳۸۴-۱۳۶۷) برای واسنجی مدل‌ها انتخاب شده و ۲۵ درصد داده‌ها (۱۳۹۰-۱۳۸۵) جهت اعتبارسنجی مدل‌ها استفاده شد. سپس توزیع احتمالاتی داده‌های جریان ماهانه در ایستگاه آب‌سنجی سسته براساس آزمون‌های کلموگروف-اسمیرنوف و شاپیرو-ویلک مورد بررسی قرار گرفته و نرمال‌سازی توزیع داده‌ها انجام گرفت. پس از بهینه‌سازی پارامترهای مربوط به هر یک از توابع کرنل، مقادیر جریان ماهانه در ایستگاه آب‌سنجی سسته پیش‌بینی شده و عملکرد این توابع با استفاده از جذر میانگین مربعات خطا (RMSE) و ضریب همبستگی (CC) مورد ارزیابی قرار گرفت.

\* مسئول مکاتبه: [mohammadisazade@gmail.com](mailto:mohammadisazade@gmail.com)

**یافته‌ها:** بررسی‌های این پژوهش نشان داد که اگرچه تفاوت معنی‌دار بین نتایج سه تابع کرنل وجود ندارد، ولی تابع کرنل چندجمله‌ای درجه ۴ با مقادیر ضریب همبستگی و جذر میانگین مربعات خطا به ترتیب برابر با ۰/۸۶ و ۵/۸۸ (مترمکعب در ثانیه) در دوره تست، در مقایسه با توابع کرنل دیگر از دقت بالا و عملکرد بهتری در پیش‌بینی جریان ماهانه برخوردار است. همچنین نتایج نشان داد که مدل  $ARMA(6,2)$  با مقادیر ضریب همبستگی و مجذور میانگین مربعات خطا به ترتیب برابر با ۰/۸۲ و ۶/۴۷ (مترمکعب در ثانیه) در دوره تست، نسبت به سایر مدل‌های سری زمانی عملکرد خوبی را در پیش‌بینی جریان ماهانه حوضه خرخره‌چای دارا می‌باشد.

**نتیجه‌گیری:** در نهایت مقادیر جریان ماهانه پیش‌بینی شده با استفاده از تابع کرنل چندجمله‌ای درجه ۴ (به‌عنوان نماینده مدل SVM) با نتایج مدل  $ARMA(6,2)$  (به‌عنوان نماینده مدل‌های سری زمانی) مقایسه گردید و این نتیجه حاصل شد که مدل SVM از کارایی بهتری نسبت به مدل‌های سری زمانی در پیش‌بینی جریان ماهانه حوضه خرخره‌چای برخوردار است.

**واژه‌های کلیدی:** پیش‌بینی جریان رودخانه، توابع کرنل، حوضه خرخره‌چای، ماشین بردار پشتیبان، مدل‌های سری زمانی

#### مقدمه

کردن مدل و انجام صحیح محاسبات، ساختار مدل‌های سری زمانی می‌تواند با ساختار سری‌های هیدرولوژیکی سازگاری و مطابقت ویژه‌ای داشته باشد (۲۰). بیش‌تر مدل‌های سری زمانی رایج در هیدرولوژی و منابع آب از نوع خطی بوده که  $AR^3$ ،  $ARMA^4$  و  $ARIMA^5$  از متداول‌ترین نوع آن‌ها می‌باشند (۲۰). ماشین‌های بردار پشتیبان، سیستم‌های یادگیری خاصی هستند که از یک فضای فرضیه به نام فضای مشخصه<sup>۶</sup> شامل توابع خطی با بعد زیاد استفاده کرده و می‌توانند برای پیش‌بینی (دسته‌بندی و رگرسیون) مورد استفاده قرار گیرند. این سیستم‌ها توسط الگوریتم یادگیری که بر اساس تئوری بهینه‌سازی استوار است، واسنجی می‌شوند. این تئوری که توسط وپنیک و کورتیس (۱۹۹۵) به‌عنوان روشی قدرتمند و مهم معرفی گردیده از پیش‌قدر یادگیری<sup>۷</sup> که از تئوری یادگیری آماری استخراج شده استفاده می‌کند (۲۲).

مدل‌های مختلفی در تخمین جریان حوضه‌های آبخیز مورد استفاده قرار می‌گیرند که کارایی و همچنین موارد استفاده از آن‌ها براساس ساختار حوضه، داده‌های در دسترس و مورد نیاز و همچنین هدف شبیه‌سازی از همدیگر متفاوت است. انتخاب مدلی که بتواند میزان جریان رودخانه را با دقت بالا و خطای کم مدل‌سازی نماید کمک شایانی در مدیریت صحیح طرح‌های منابع آبی و افزایش عملکرد آن‌ها دارد. در دهه‌های اخیر مدل‌های سری زمانی و اخیراً مدل ماشین بردار پشتیبان<sup>۱</sup> (SVM) به‌دلیل ماهیت ریاضی و داده محور بودن آن، جهت تخمین جریان رودخانه‌ها مورد توجه پژوهشگران قرار گرفته‌اند.

مدل‌های سری زمانی از دو بخش اصلی، شامل مؤلفه تصادفی و مؤلفه جبری تشکیل شده‌اند که مؤلفه جبری مدل با استفاده از ارقام مشاهداتی و بخش تصادفی با استفاده از روش‌های مختلف استوکاستیک<sup>۲</sup> به‌دست می‌آید. بنابراین در صورت درست انتخاب

3- Auto Regressive  
4- Auto Regressive Moving Average  
5- Auto Regressive Integrated Moving Average  
6- Feature Space  
7- Learning Bias

1- Support Vector Machine  
2- Stochastic

عصبی مصنوعی را برای شبیه‌سازی بارش- رواناب در یک حوضه کوهستانی با داده‌های محدود در یوتارانچال هندوستان به‌کار گرفت. نتایج بررسی وی نشان از توانایی مدل ماشین بردار پشتیبان با تابع کرنل پایه شعاعی در پیش‌بینی رواناب مستقیم، جریان پایه و جریان کل این منطقه کوهستانی و پر از تپه داشت (۱).

کاکائی‌لفادانی و همکاران (۲۰۱۳) سه مدل ماشین بردار پشتیبان و نرو فازی (ANFIS) و مدل هیدرولوژیکی مفهومی Mike11/NA را در شبیه‌سازی جریان روزانه حوضه اسکندری به‌کار گرفتند. نتایج نشان داد که هر سه مدل دارای عملکرد مناسب در منطقه مورد مطالعه می‌باشند. در نهایت مدل ماشین بردار پشتیبان با تابع کرنل پایه شعاعی بالاترین ضریب همبستگی و کم‌ترین میانگین مربعات خطا را در شبیه‌سازی جریان رودخانه نسبت به دو مدل دیگر از خود نشان داد (۱۶). بنی‌حیب و ولی‌پور (۲۰۰۸) عملکرد سه مدل آرما، آریمما و شبکه عصبی مصنوعی را در پیش‌بینی ماهانه جریان ورودی به مخزن سد دز مورد مقایسه قرار دادند که نتایج بیانگر دقت بیش‌تر مدل شبکه عصبی مصنوعی نسبت به دو مدل دیگر در پیش‌بینی جریان ماهانه بوده است (۶). انتخاب مدلی که میزان جریان رودخانه را با دقت بالا و خطای کم‌تر پیش‌بینی نماید می‌تواند کمک شایانی در مدیریت صحیح طرح‌های منابع آبی و افزایش عملکرد این طرح‌ها داشته باشد. با توجه به آنچه آمد پژوهش حاضر سعی دارد تا کارایی مدل SVM را در کنار مدل‌های سری زمانی در پیش‌بینی رواناب ماهانه حوضه رودخانه خرخره‌چای مورد ارزیابی قرار دهد و دقت این مدل‌ها را در این پیش‌بینی نسبت به یکدیگر بررسی کند. با توجه به این‌که ترکیب ورودی‌ها و نوع ساختار به‌کار گرفته شده برای هر یک از این مدل‌ها تأثیر چشم‌گیری در نتایج پیش‌بینی و عملکرد آن‌ها

تاکنون پژوهش‌های زیادی در مورد ارزیابی و همچنین مقایسه عملکرد هر یک از این مدل‌ها با مدل‌های هوشمند از قبیل شبکه عصبی مصنوعی، فازی و نروفازی در تخمین جریان رودخانه انجام گرفته است. آسفا و همکاران (۲۰۰۵) مدل SVM با تابع کرنل خطی را برای پیش‌بینی جریان‌های فصلی و ساعتی در حوضه رودخانه سویتر در ایالات متحده به‌کار گرفتند. آن‌ها در این پژوهش با توجه به حجم جریان معادل برف و همچنین حجم جریان در دوره‌های پیشین، میزان حجم جریان در دوره‌های ۶ ماهه و ۲۴ ساعته را پیش‌بینی نمودند که نتایج مطالعه ایشان از دقت مطلوبی برخوردار بود (۳). در پژوهشی دیگر جاین و کومار (۲۰۰۷) کارایی مدل‌های شبکه عصبی مصنوعی و مدل‌های سری زمانی را در پیش‌بینی مقادیر جریان ماهانه رودخانه کلرادو تحت تأثیر سه سناریو مورد بررسی قرار دادند که در هر سه سناریو دقت مدل شبکه عصبی بهتر از مدل خود همبسته (AR) به‌دست آمد (۱۳).

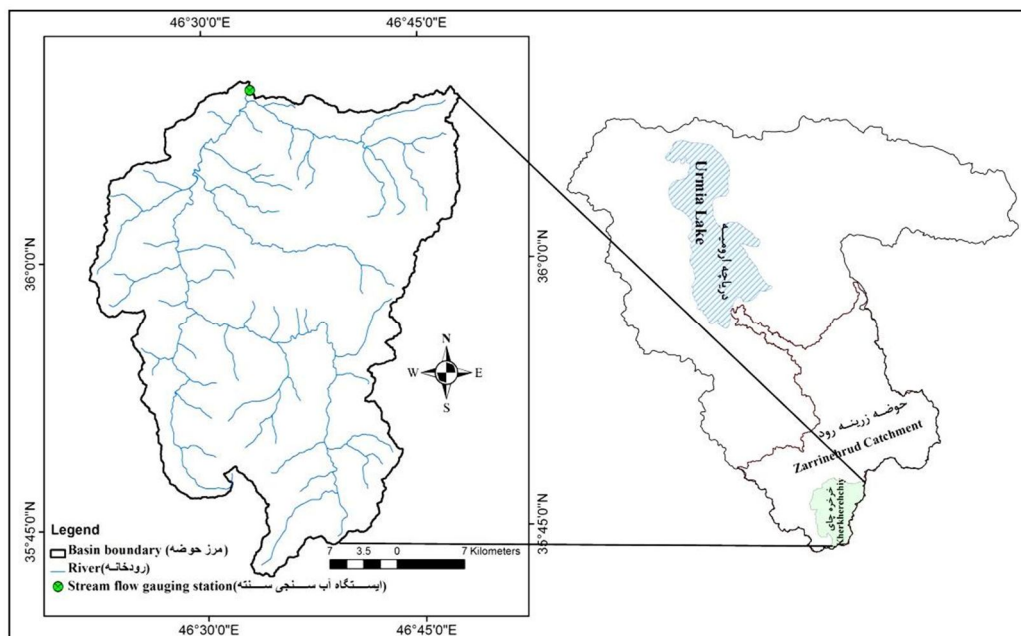
داملی و یالسن (۲۰۰۷) به پیش‌بینی جریان‌های سیلابی با استفاده از سری‌های زمانی در رودخانه می‌سی‌سی‌پی آمریکا پرداختند. نتایج مطالعه ایشان نیز بیانگر توانایی سری‌های زمانی در ساخت داده‌های جریان روزانه و صحت پیش‌بینی‌های حاصله است (۹). لیو (۲۰۱۱) در مطالعه خود عملکرد مدل‌های رگرسیون، ARIMA و شبکه عصبی مصنوعی را در پیش‌بینی جریان روزانه رودخانه وایت‌کلی‌کریک نزدیک ایالت پنسیلوانیا مورد ارزیابی قرار داده و دقت آن‌ها را در تخمین جریان رودخانه با یکدیگر مقایسه کردند. نتایج پژوهش آنان نشان داد زمانی که داده‌ها از توزیع نرمال تبعیت کرده و دارای طول دوره آماری کافی باشند، مدل‌های خطی کم‌ترین خطا را در پیش‌بینی جریان رودخانه دارا می‌باشند (۱۷). آداموفسکی (۲۰۱۳) روش مدل ماشین بردار پشتیبان و شبکه

این حوضه سهم قابل توجهی در حجم آب ورودی به سد بوکان و در نتیجه تأمین نیازهای آبی پایین دست این سد دارد. مساحت این حوضه در محل ایستگاه آب‌سنجی سنته بالغ بر ۱۲۳۲ کیلومتر مربع بوده و متوسط ارتفاع آن از سطح دریا ۲۰۳۵ متر است. همچنین متوسط بارش سالانه این حوضه در حدود ۴۰۲ میلی‌متر است. داده‌های مورد استفاده در این پژوهش شامل جریان ماهانه مشاهداتی در ایستگاه آب‌سنجی سنته طی دوره ۱۳۶۷ الی ۱۳۹۰ است. به طوری که داده‌های دوره آماری (بهمن ۱۳۸۴ - بهمن ۱۳۶۷) به عنوان داده‌های واسنجی و دوره (شهریور ۱۳۹۰ - اسفند ۱۳۸۴) به عنوان داده‌های صحت‌سنجی هر یک از مدل‌ها انتخاب گردیدند. به عبارت دیگر ۷۵ درصد از داده‌های مشاهداتی برای واسنجی و ۲۵ درصد دیگر جهت صحت‌سنجی مدل‌ها در نظر گرفته شدند (شکل ۲).

دارد؛ بنابراین در این پژوهش تلاش شد کارایی مدل‌های مذکور به‌ازای تأخیرهای ۱ تا ۱۵ ماهه جریان و همچنین انواع ساختارهای مختلف مدل‌ها مورد ارزیابی قرار گیرد. برای این منظور مدل‌های خطی ARMA، AR و ARIMA (به‌عنوان مدل‌های سری زمانی) و مدل SVM با سه تابع کرنل پایه شعاعی (RBF<sup>۱</sup>)، چندجمله‌ای درجه d و کرنل خطی جهت پیش‌بینی ماهانه رواناب به‌کار گرفته شدند.

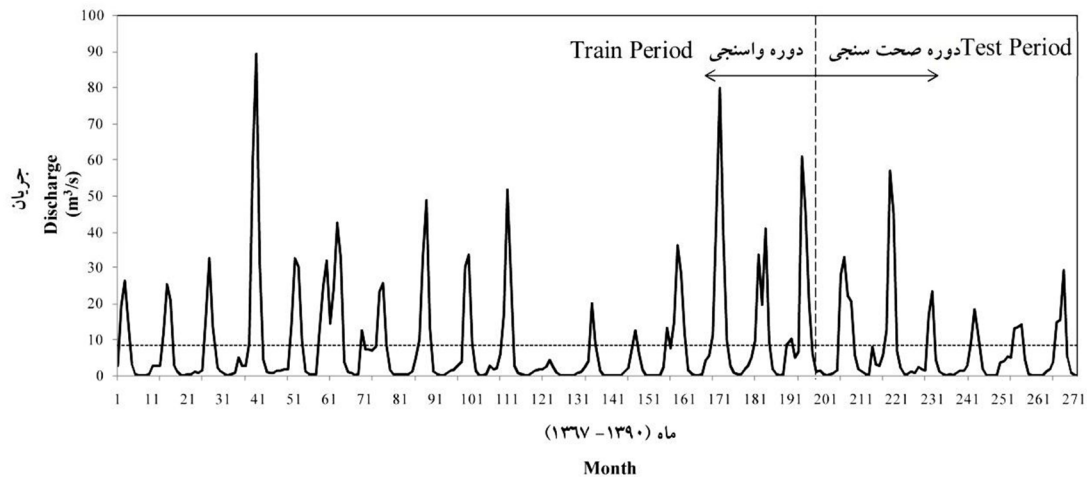
### مواد و روش‌ها

داده و منطقه مورد مطالعه: محدوده مورد مطالعه در این پژوهش حوضه رودخانه خرخره‌چای از شاخه‌های اصلی حوضه زرنه‌رود است که در طول جغرافیایی ۴۶ درجه و ۲۲ دقیقه تا ۴۶ درجه و ۴۸ دقیقه شرقی و عرض جغرافیایی ۳۵ درجه و ۴۲ دقیقه تا ۳۶ درجه و ۱۰ دقیقه شمالی قرار دارد (شکل ۱).



شکل ۱- موقعیت حوضه خرخره‌چای و ایستگاه آب‌سنجی سنته.

Figure 1. Location of Kherkherechiy river basin and Santeh hydrometric station.



شکل ۲- سری زمانی داده‌های جریان مشاهداتی رودخانه خرخره‌چای در ایستگاه سنته.

Figure 2. Observed time series of Kherkherehchiy river flow in the Santeh station.

$$W_{ij} = P_{ij} - \bar{Q}_j \quad (2)$$

$$K_{ij} = \frac{W_{ij}}{SD_W} \quad (3)$$

در این روابط  $n$  تعداد سال‌ها،  $Q_{ij}$  میزان جریان مشاهداتی در ماه  $j$  ام از سال  $i$  ام،  $\bar{Q}_j$  میانگین درازمدت جریان در ماه  $j$  ام،  $SD_W$  انحراف معیار متغیر  $W_{ij}$  و  $K_{ij}$  جریان نرمال شده در ماه  $j$  ام از سال  $i$  ام هستند. بعد از نرمال‌سازی داده‌ها، شاخص‌های آماری میانگین، مقادیر حداقل و حداکثر، انحراف معیار، چولگی و کشیدگی محاسبه گردیده و آزمون‌های کلموگروف-اسمیرنوف و شاپیرو-ویلک در سطح اطمینان ۰/۹۵ برای داده‌های مشاهداتی اولیه و نرمال شده انجام گرفت که نتایج آن‌ها در جدول ۱ آورده شده است. لازم به ذکر است که آزمون‌های آماری کلموگروف-اسمیرنوف و شاپیرو-ویلک با استفاده از نرم‌افزار SPSS 22 انجام گرفته است.

نرمال‌سازی داده‌ها: پیش‌شرط استفاده از مدل‌های سری زمانی تبعیت نمودن داده‌های مورد استفاده از توزیع نرمال است (۲). به‌علت تبعیت نکردن داده‌های دبی مشاهداتی ماهانه ایستگاه سنته از توزیع نرمال؛ از روش‌های متعددی هم‌چون باکس-کاکس<sup>۱</sup>، تبدیل جانستون<sup>۲</sup> و توابع لگاریتمی، نمایی، توانی، رادیکالی و معکوس هر یک از آن‌ها جهت نرمال‌سازی داده‌های جریان استفاده گردید. با به‌کارگیری هر یک از این روش‌ها بر روی داده‌های جریان و انجام آزمون‌های کلموگروف-اسمیرنوف و شاپیرو-ویلک در سطح اطمینان ۰/۹۵ بر روی آن‌ها، این نتیجه حاصل شد که هیچ‌یک از روش‌های مذکور در نرمال‌سازی داده‌های جریان کارساز نیستند. در نهایت در پژوهش حاضر از روابط ۱ تا ۳ که حاصل تجربه نویسندگان مقاله است، جهت نرمال‌سازی داده‌ها استفاده گردید.

$$P_{ij} = Ln(Q_{ij} + 0.052) \quad (1)$$

$$j = 1:12, \quad i = 1:n$$

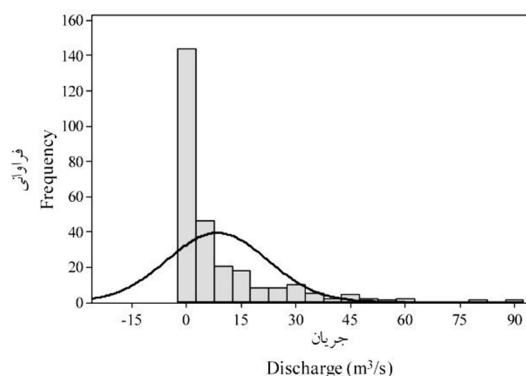
- 1- Box-Cox
- 2- Johnson-Transformation

جدول ۱- نتایج آزمون‌های کلموگروف-اسمیرنوف و شاپیرو-ویلک برای داده‌های مشاهداتی اولیه و نرمال‌شده.

**Table 1. The results of Shapiro-Wilk and Kolmogorov-Smirnov test for observed and normalized data.**

شاپیرو-ویلک Shapiro-Wilk		کلموگروف-اسمیرنوف Kolmogorov-Smirnov		درجه آزادی degree of freedom	کشیدگی Kurtosis	چولگی Skewness	انحراف معیار Standard Deviation	جریان حداکثر Maximum Flow (m <sup>3</sup> /s)	جریان حداقل Minimum Flow (m <sup>3</sup> /s)	جریان متوسط Mean Flow (m <sup>3</sup> /s)	
سطح معنی داری Significance Level	آماره statistic	سطح معنی داری Significance Level	آماره statistic								
0.00	0.65	0.00	0.27	271	13.55	2.62	13.83	89.43	0.00	8.39	داده‌های مشاهداتی Observed Data
0.27	0.99	0.20	0.04	271	8.35	0.00	1.00	2.87	-2.86	0.00	داده‌های نرمال‌شده Normalized Data
0.21	0.99	0.20	0.05	203	0.45	-0.07	1.01	2.87	-2.86	0.06	داده‌های واسنجی Training Data
0.59	0.98	0.20	0.05	68	0.47	0.18	0.96	2.36	-2.86	-0.17	داده‌های صحت‌سنجی Testing Data

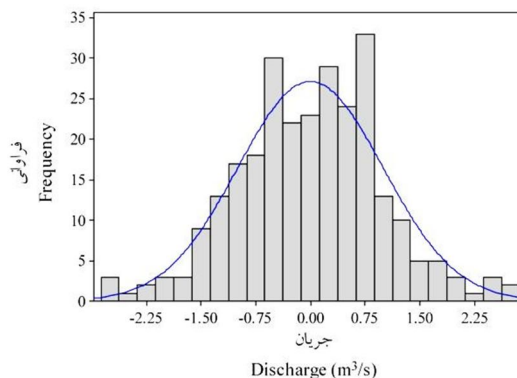
توجه به نتایج آمده در جدول ۱ می‌توان دریافت که توزیع احتمالاتی داده‌های جریان ماهانه مشاهداتی اولیه در ابتدا نرمال نبوده ولی پس از اعمال معادلات ۲۰ تا ۲۲ بر روی آن‌ها، توزیع داده‌ها به نرمال تبدیل شده است. شکل‌های ۳ و ۴ به ترتیب هیستوگرام فراوانی داده‌های اولیه و نرمال شده را نشان می‌دهند.



شکل ۳- هیستوگرام فراوانی داده‌های مشاهداتی اولیه.

**Figure 3. Frequency histogram of observed data.**

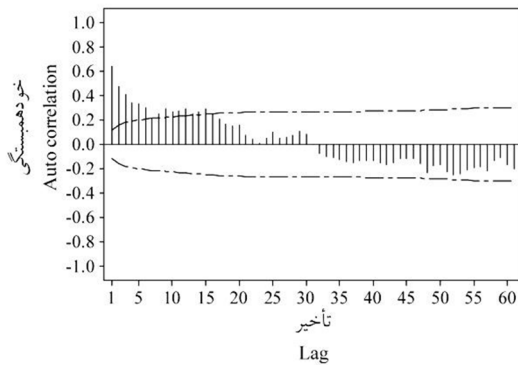
همچنین در این جدول نتایج مربوط به محاسبه شاخص‌ها و انجام آزمون‌های مذکور برای داده‌های اولیه (نرمال‌نشده) نشان داده شده است. لازم به ذکر است چنانچه طی آزمون‌های مذکور سطح معنی‌داری بیش‌تر از ۰/۰۵ به دست آید داده‌ها با سطح اطمینان ۰/۹۵ از توزیع نرمال تبعیت می‌کنند. بر این اساس با



شکل ۴- هیستوگرام فراوانی داده‌های نرمال شده.

**Figure 4. Frequency histogram of normalized data.**

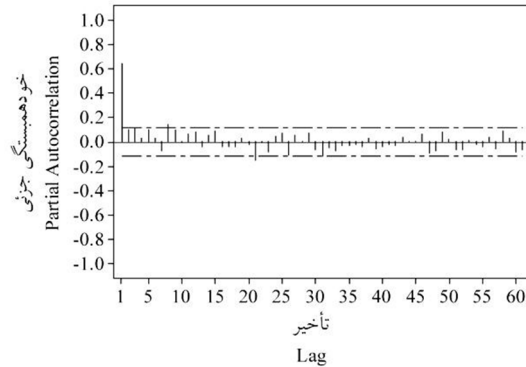
که بر جریان تأثیر معنی‌داری دارند را تعیین کرد (۱۴)، (۲۴). در شکل‌های ۵ و ۶ به ترتیب نمودار توابع ACF و PACF مربوط به داده‌های جریان ماهانه نرمال شده نشان داده شده است.



شکل ۵- نمودار تابع ACF برای داده‌های جریان ماهانه نرمال ایستگاه آب‌سنجی سته.

Figure 5. ACF diagram of normalized monthly river flow in the santeh station.

تعیین تعداد ورودی مدل‌ها: می‌توان با استفاده از توابع ACF و PACF میزان همبستگی جریان ماهانه با تأخیرهای ماهانه مختلف را مشخص نموده و از این طریق به‌طور تقریبی تعداد و درجه تأخیرهای ماهانه



شکل ۶- نمودار تابع PACF برای داده‌های جریان ماهانه نرمال ایستگاه آب‌سنجی سته.

Figure 6. PACF diagram of normalized monthly river flow in the santeh station.

مختلف مورد استفاده در هر یک از مدل‌ها آورده شده است.

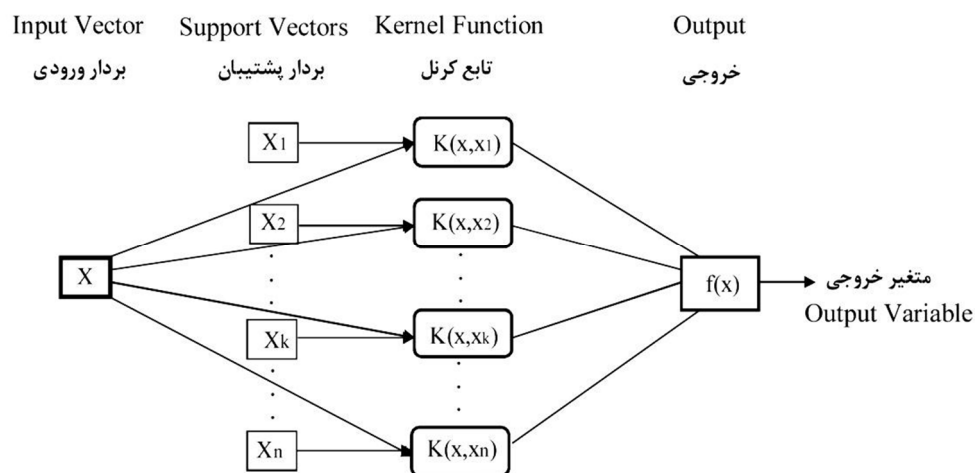
ماشین بردار پشتیبان (SVM): اولین کاربرد این روش در مسائل آب توسط دیباک و همکاران (۲۰۰۱) جهت شبیه‌سازی بارش رواناب ارائه شد (۱۰). ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد (۱۱). شکل ۷ ساختار ماشین بردار پشتیبان را نشان می‌دهد.

مطابق با نمودار PACF داده‌های نرمال‌شده جریان (شکل ۶)، اولین جریان تأخیر یافته (درجه تأخیر برابر با یک) بیش‌ترین تأثیر را در برآورد جریان دارد و همچنین براساس نمودار ACF جریان با تأخیرهای ۱ تا ۱۵ ماهه خارج از باند استقلال قرار گرفته‌اند. پس می‌توان نتیجه گرفت که تعداد تأخیرهای جریان ماهانه لازم جهت لحاظ نمودن در پیش‌بینی جریان توسط مدل‌های مذکور (تعداد متغیرهای ورودی لازم به مدل‌ها) برابر با ۱۵ ماه می‌باشد. در جدول ۲ ترکیب ورودی‌های

جدول ۲- ترکیب ورودی‌های مشخص شده بر اساس تابع‌های ACF و PACF برای هر یک از مدل‌ها.

Table 2. Input combinations of each model specified based on ACF and PACF.

تعداد ورودی Input Number	ترکیب ورودی Input Combination	خروجی Output
۱	$Q_{t-1}$	$Q_t$
۲	$Q_{t-1}, Q_{t-2}$	$Q_t$
۳	$Q_{t-1}, Q_{t-2}, Q_{t-3}$	$Q_t$
۴	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}$	$Q_t$
۵	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}$	$Q_t$
۶	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}$	$Q_t$
۷	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}$	$Q_t$
۸	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}$	$Q_t$
۹	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}$	$Q_t$
۱۰	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}, Q_{t-10}$	$Q_t$
۱۱	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}, Q_{t-10}, Q_{t-11}$	$Q_t$
۱۲	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}, Q_{t-10}, Q_{t-11}, Q_{t-12}$	$Q_t$
۱۳	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}, Q_{t-10}, Q_{t-11}, Q_{t-12}, Q_{t-13}$	$Q_t$
۱۴	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}, Q_{t-10}, Q_{t-11}, Q_{t-12}, Q_{t-13}, Q_{t-14}$	$Q_t$
۱۵	$Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}, Q_{t-8}, Q_{t-9}, Q_{t-10}, Q_{t-11}, Q_{t-12}, Q_{t-13}, Q_{t-14}, Q_{t-15}$	$Q_t$



شکل ۷- ساختار ماشین بردار پشتیبان (۲۵).

Figure 7. Support vector machine structure (25).



$$W^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \varepsilon_i^* \quad (7)$$

$$y_i - W^T \cdot \phi(x_i) - b \leq \varepsilon + \varepsilon_i$$

$$\varepsilon_i, \varepsilon_i^* \geq 0, i = 1, 2, \dots, N$$

در معادلات بالا  $C$  عددی صحیح و مثبت است، که عامل تعیین جریمه در هنگام رخ دادن خطای واسنجی مدل می‌باشد،  $\phi$  تابع کرنل<sup>۲</sup>،  $N$  تعداد نمونه‌ها و دو مشخصه  $\xi_i^*$  و  $\xi_i$  متغیرهای کمبود<sup>۳</sup> هستند که حد بالا و پایین خطای آموزش مرتبط با مقدار خطای مجاز  $\varepsilon$  را مشخص می‌کنند (۱۱).

در مسائل پیش‌بینی می‌شود که داده‌ها، درون بازه مرزی  $\varepsilon$  قرار گیرند. حال اگر داده‌ای خارج از بازه  $\varepsilon$  قرار گرفت آن‌گاه یک خطا معادل  $\xi_i^*$  و  $\xi_i$  وجود خواهد داشت. ذکر این نکته نیز لازم است که مدل SVM مشکلات ناشی از کم‌تخمینی<sup>۴</sup> و فوق‌براشی<sup>۵</sup> را با کمینه کردن هم‌زمان دو ترم  $W^T \cdot W/2$  و خطای آموزشی، یعنی  $C \sum_{i=1}^N (\xi_i + \xi_i^*)$  را در رابطه ۶ حل می‌کند. بنابراین با معرفی ضریب لاگرانژ  $\alpha_i$  و  $\alpha_i^*$  مسأله بهینه‌سازی با حداکثرسازی عددی تابع درجه دوم زیر (رابطه ۸) با شرایط رابطه ۹ حل خواهد شد.

$$\sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \quad (8)$$

$$-0.5 \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \phi(x_i)^T \cdot \phi(x_j)$$

$$\sum_{i=1}^N (\alpha_i + \alpha_i^*) = 0 \quad 0 \leq \alpha_i \leq C \quad 0 \leq \alpha_i^* \leq C \quad (9)$$

$$i = 1, 2, \dots, N$$

تابع هدف رابطه ۸ تابع محدب است و بنابراین جواب رابطه ۸ یکتا و بهینه خواهد بود. پس از تعریف ضرایب لاگرانژ در رابطه ۸ مشخصه‌های  $w$  و

در مدل رگرسیون SVM تابعی مرتبط با متغیر وابسته  $y$  که خود تابعی از چند متغیر مستقل  $x$  است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند  $f(x)$  به‌علاوه مقداری اغتشاش<sup>۱</sup> (خطای مجاز  $\varepsilon$ ) مشخص شود.

$$f(x) = W^T \cdot \phi(x) + b \quad (4)$$

$$y = f(x) + noise \quad (5)$$

چنانچه  $W$  بردار ضرایب و  $b$  ثابت مشخصه‌های تابع رگرسیونی و  $\phi$  نیز تابع کرنل باشد، آن‌گاه هدف پیدا کردن فرم تابعی برای  $f(x)$  است. این مهم با واسنجی مدل SVM توسط مجموعه‌ای از نمونه‌ها (مجموعه واسنجی) محقق می‌شود.

این روند شامل بهینه‌سازی متوالی تابع خطاست. بسته به تعریف این تابع خطا دو نوع مدل SVM تعریف می‌شود:

SVM رگرسیونی نوع ۱ (به‌عنوان  $\varepsilon$ -SVM رگرسیونی نیز شناخته می‌شود)؛ SVM رگرسیونی نوع ۲ (به‌عنوان  $\nu$ -SVM رگرسیونی شناخته می‌شود). شایان ذکر است در این پژوهش از مدل  $\varepsilon$ -SVM رگرسیونی، به‌دلیل کاربرد گسترده آن در مطالعات رگرسیونی، برای پیش‌بینی جریان رودخانه استفاده شده است (۱۱).

بنابراین برای محاسبه  $W$  و  $b$  لازم است تابع خطا (رابطه ۶) در مدل  $\varepsilon$ -SVM با در نظر گرفتن شرایط مندرج (قیود) در رابطه ۷ بهینه شود.

$$\frac{1}{2} W^T w + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^* \quad (6)$$

1- Noise

2- Kernel Function

3- Slack Variable

4- Under Fitting

5- Over Fitting

$$f(x) = \sum_{i=1}^N \overline{\alpha}_i \cdot \phi(x_i)^T \cdot \phi(x) + b \quad (11)$$

در رابطه ۱۱،  $\overline{\alpha}_i$  میانگین ضرایب لاگرانژ می‌باشد. محاسبه  $\phi(x)$  در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل SVM رگرسیونی انتخاب یک تابع کرنل است (۱۱). ماشین بردار پشتیبان معمولاً با سه تابع کرنل پایه شعاعی<sup>۴</sup> (RBF)، چندجمله‌ای<sup>۵</sup> درجه  $d$  و خطی مورد استفاده قرار می‌گیرد که روابط مورد استفاده در هر کدام از آن‌ها به ترتیب در زیر آورده شده است (۵، ۱۵).

$$k(x, x_i) = \exp\left(-\|x - x_i\|^2 / \sigma^2\right) \quad (12)$$

$$k(x, x_i) = (t + \langle x, x_i \rangle)^d \quad (13)$$

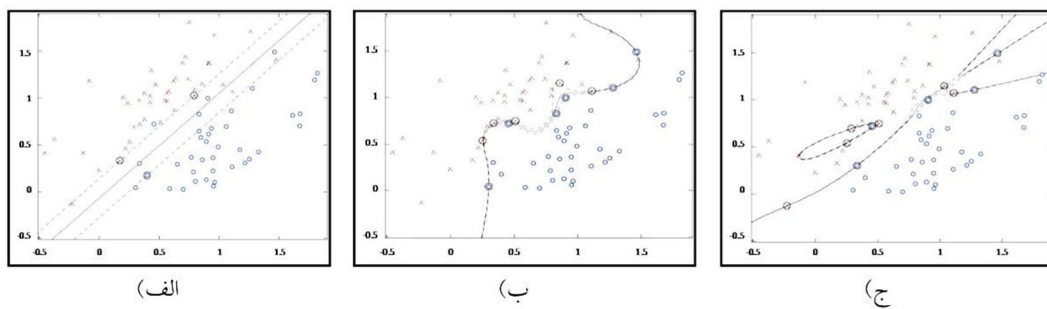
$$k(x, x_i) = x \cdot x_i \quad (14)$$

در شکل ۸ تابع کرنل خطی، پایه شعاعی و چندجمله‌ای درجه  $d$  در دسته‌بندی داده‌ها به کلاس‌های متمایز به صورت شماتیک نشان داده شده‌اند.

b در مدل SVM رگرسیونی با استفاده از شرایط تئوری کرانش - کوهن - تاکر<sup>۱</sup> محاسبه می‌شوند (۱۲) که در آن  $W = \sum_{i=1}^N (\alpha_i + \alpha_i^*) \phi(x_i)$  است. در نتیجه برای مدل SVM رگرسیونی خواهیم داشت:

$$W = \sum_{i=1}^N (\alpha_i + \alpha_i^*) \phi(x_i)^T \cdot \phi(x) + b \quad (10)$$

باید توجه داشت که ترم‌های لاگرانژ  $(\alpha_i + \alpha_i^*)$  می‌تواند صفر و یا غیرصفر باشند. بنابراین فقط مجموعه داده‌هایی که ضرایب  $\overline{\alpha}_i$  آن‌ها غیرصفر است در معادله رگرسیون نهایی وارد می‌شوند و این مجموعه داده‌ها به‌عنوان بردارهای پشتیبان شناخته می‌شوند. به‌طور ساده، بردارهای پشتیبان آن داده‌هایی هستند که به ساختار شدن تابع رگرسیونی کمک می‌کنند. در میان بردارهای مذکور آن‌هایی که مقدار  $|\overline{\alpha}_i|$  آن‌ها کم‌تر از C باشد بردارهای پشتیبان حاشیه‌ای<sup>۲</sup> نامیده می‌شوند. هنگامی که مقدار  $|\overline{\alpha}_i|$  بردارهای پشتیبان برابر مقدار C باشد، به‌عنوان بردار پشتیبان خطی<sup>۳</sup>، یا بردار پشتیبان کراندار شناخته می‌شود. بردارهای پشتیبان حاشیه‌ای در حاشیه مرز غیرحساس یافت می‌شوند، در حالی که بردارهای پشتیبان خطی خارج از بازه هستند. در نهایت تابع SVM رگرسیونی را می‌توان به فرم زیر بازنویسی کرد:



شکل ۸- طبقه‌بندی داده‌ها توسط ماشین بردار پشتیبان با سه تابع کرنل خطی (الف)، پایه شعاعی (ب) و چندجمله‌ای (ج).

Figure 8. Data classification using support vector machine with three kernel functions a) linear b) radial basis c) polynomial.

- 1- Karush-Kuhn-Tucker
- 2- Margin Support Vector
- 3- Error Support Vector
- 4- Radial Basis Function

دهه ۱۹۶۰ توسط توماس<sup>۱</sup>، فیرینگ<sup>۲</sup> و یوجویچ<sup>۳</sup> آغاز گردیده و در دهه ۱۹۷۰ توسط باکس<sup>۴</sup> و جنکینس<sup>۵</sup> توسعه یافته است (۸). در پیش‌بینی با دقت بالا، افق زمانی مشخص و محدود بوده در حالی که در تولید داده‌ها، افق زمانی خاصی را نمی‌توان در نظر گرفت (۸). در این پژوهش از مدل‌های AR، ARMA و ARIMA جهت پیش‌بینی سری‌های زمانی استفاده گردیده که در جدول ۳ به‌طور مختصر به روابط هر کدام از آن‌ها اشاره شده است (۲۳، ۲۶).

با توجه به این‌که از پرکاربردترین توابع کرنل، کرنل خطی، پایه شعاعی و چندجمله‌ای می‌باشند (۷، ۱۷، ۲۱) در این پژوهش نیز از این سه تابع کرنل استفاده شده است. لازم به ذکر است فرآیند محاسبات ماشین بردار پشتیبان، بر اساس کدنویسی در محیط متلب انجام و پارامترهای توابع کرنل از طریق سعی و خطا بهینه گردیدند.

**مدل‌های سری زمانی:** شبیه‌سازی سری‌های زمانی از جمله رویکردهای مطرح در زمینه پیش‌بینی و تولید داده‌های جدید می‌باشد که در هیدرولوژی از اوایل

جدول ۳- مشخصات مدل‌های سری زمانی مورد استفاده در پژوهش.

Table 3. Characteristics of time series models used in the study.

توابع Functions	مدل سری زمانی Time series Model
$AR(p): X_t = \mu + \sum_{j=1}^p \phi_j (X_{t-j} - \mu) + \varepsilon_t$	رابطه ۱۵ (AR)
$ARMA(p, q): Z_t = \sum_{i=1}^p \phi_i Z_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}$	رابطه ۱۶ (ARMA)
$Z_t = X_t - \mu$	رابطه ۱۷
$u_t = \sum_{k=1}^p \phi_k u_{t-k} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}$	رابطه ۱۸ (ARIMA)
$X_t = u_t + u_{t-1} + u_{t-2} + \dots$	رابطه ۱۹

- 1- Thomas
- 2- Fiering
- 3- Yevjevich
- 4- Box
- 5- Jenkins

شده‌اند. دقیق‌ترین مدل با توجه به این معیارها، مدلی خواهد بود که مقدار این سه معیار برای آن به ترتیب نزدیک به یک، صفر و یک باشد.

$$CC = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})^2}{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (20)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (21)$$

$$NS = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (22)$$

در روابط بالا  $x_i$  و  $\bar{x}$  به ترتیب مقادیر مشاهداتی جریان و میانگین مقادیر مشاهداتی،  $y_i$  مقادیر محاسباتی جریان،  $\bar{y}$  میانگین مقادیر محاسباتی و  $N$  تعداد داده‌ها می‌باشد (۱۸).

مراحل پیش‌بینی در هر یک از مدل‌ها: در پژوهش حاضر پس از نرمال‌سازی داده‌های جریان، در ابتدا هر کدام از مدل‌های مورد استفاده به‌ازای هر یک از ۱۵ ترکیب ورودی معرفی شده در جدول ۲، با توجه به داده‌های دوره واسنجی (جدول ۱) آموزش داده شدند. سپس با استفاده از مدل‌های واسنجی شده، پیش‌بینی یک ماه بعد جریان رودخانه در دوره واسنجی برای ایستگاه سنته صورت گرفت. لازم به ذکر است که این مراحل در مدل SVM به‌ازای هر ۳ تابع کرنل (۴۵) (مدل) و برای مدل ARIMA به‌ازای میانگین متحرک و تفاضل‌های مختلف (۸۱ مدل) انجام گرفت. همچنین با استفاده از معکوس روشی که برای نرمال‌سازی داده‌ها توضیح داده شد، جریان پیش‌بینی شده نرمال به جریان پیش‌بینی شده واقعی تبدیل گردید. سپس مقادیر شاخص‌های آماری ذکر شده در بخش قبل، جهت ارزیابی عملکرد مدل‌ها در مراحل واسنجی و

در روابط فوق  $X_t$  و  $Z_t$  سری‌های زمانی وابسته،  $\mu$  میانگین سری زمانی داده‌ها،  $\phi_j$  ضرایب مدل AR و  $\varepsilon_t$  سری مستقل با توزیع نرمال، میانگین صفر و انحراف معیار  $\sigma_\varepsilon$ ،  $\theta_i$  و  $\phi_i$  ضرایب مدل ARMA،  $\theta_j$  و  $\phi_k$  ضرایب مدل ARIMA،  $n$  تعداد داده‌ها،  $\rho_j$  همبستگی جریان با تأخیر  $j$  و  $p$  مرتبه مدل‌ها و  $\sigma^2$  واریانس داده‌ها می‌باشند. همچنین سری  $u_t$  توسط تفاضل‌گیری  $d$  ام از سری  $X_t$  حاصل می‌شود. لازم به ذکر است که در پژوهش حاضر از روش حداکثر درست‌نمایی در تخمین پارامترهای مدل‌های فوق استفاده گردیده است. همچنین فرآیند محاسبات مدل‌های سری زمانی از طریق کدنویسی در محیط متلب انجام شده است.

تعیین زمان تأخیر و معیارهای ارزیابی: توابع خودهمبستگی<sup>۱</sup> (ACF) و خودهمبستگی جزئی<sup>۲</sup> (PACF) از جمله معیارهای مطرح در تعیین تعداد ورودی‌های مؤثر در پیش‌بینی متغیر مورد نظر می‌باشند (۱۴، ۲۴). در پژوهش حاضر نیز از این توابع جهت تعیین تعداد تأخیرهای جریان ماهانه لازم که باید در ارزیابی کارایی مدل‌های ماشین بردار و سری‌های زمانی مورد بررسی قرار گیرند استفاده گردید.

با استفاده از روش‌های مختلفی می‌توان عملکرد مدل‌ها را مورد ارزیابی و مقایسه قرار داد، یکی از این روش‌ها استفاده از معیارهای ارزیابی می‌باشد. از جمله معیارهای ارزیابی پرکاربرد در علوم مهندسی آب، ضرایب همبستگی<sup>۳</sup> (CC)، ریشه میانگین مربعات خطا<sup>۴</sup> (RMSE) و معیار نش- ساتکلیف<sup>۵</sup> (NS) می‌باشند که به ترتیب در روابط ۲۰، ۲۱ و ۲۲ آورده

- 1- Autocorrelation Function
- 2- Partial Autocorrelation Function
- 3- Correlation Coefficient
- 4- Root Mean Square Error
- 5- Nash -Sutcliffe

همان‌طور که در بخش مواد و روش‌ها ذکر شد در پژوهش حاضر به ازای هر یک از توابع کرنل چندجمله‌ای، پایه شعاعی و خطی جریان رودخانه در ایستگاه سنته مورد پیش‌بینی قرار گرفته است. در نهایت عملکرد هر یک از سه تابع کرنل همراه با ورودی‌های ۱ تا ۱۵ در مدل SVM با شاخص‌های آماری مورد بررسی قرار گرفت. در جدول ۴ بهترین نتیجه مربوط به هر تابع کرنل نشان داده شده است.

صحت‌سنجی محاسبه شد. در ادامه با استفاده از این شاخص‌ها بهترین مدل برای پیش‌بینی جریان ماهانه ایستگاه سنته انتخاب گردید.

### نتایج و بحث

نتایج ماشین بردار پشتیبان: در این مدل نیز از جریان با تأخیرهای ماهانه ۱ تا ۱۵ جهت پیش‌بینی جریان ماهانه حوضه استفاده گردید و عملکرد آن به‌ازای هر یک از تعداد ورودی‌ها مورد ارزیابی قرار گرفت.

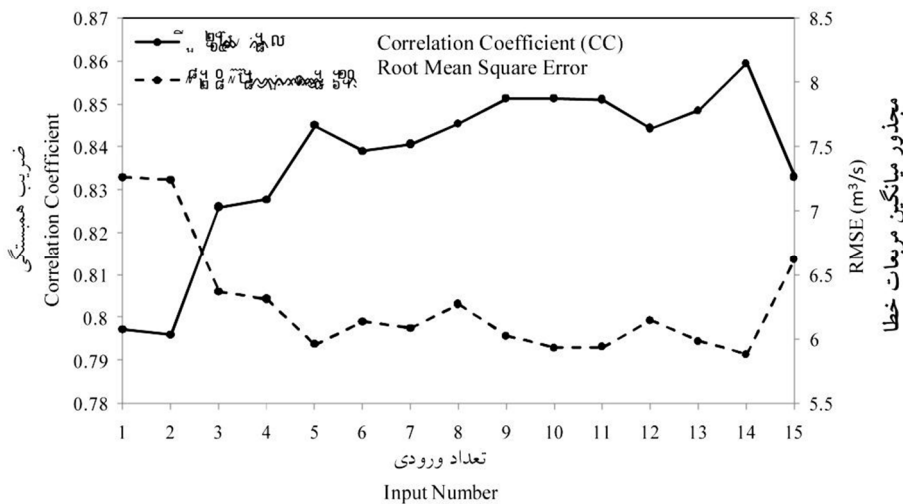
جدول ۴- ارزیابی عملکرد مدل SVM در تخمین جریان ماهانه حوضه خرخره‌چای در ایستگاه سنته.

**Table 4. Evaluation of SVM model performance in prediction of Kherkherehchiy monthly river flow in the Santeh station.**

مقادیر بهینه پارامترها Parameters Optimal Values					دوره صحت‌سنجی Testing Period			دوره واسنجی Training Period			بهترین تعداد ورودی Best of Input Number	تابع کرنل Kernel Function
$\epsilon$	t	d	$\sigma$	c	NS	RMSE (m <sup>3</sup> /s)	CC	NS	RMSE (m <sup>3</sup> /s)	CC		
1.7	304.94	4	-	15.96	0.73	5.88	0.86	0.51	8.82	0.81	14	چندجمله‌ای Polynomial
1.44	-	-	1598.91	116.66	0.73	5.89	0.86	0.51	8.77	0.82	14	پایه شعاعی Radial Basis
0.44	-	-	-	0.07	0.70	5.91	0.85	0.49	9.14	0.80	11	خطی Linear

ورودی‌های مختلف برای تابع کرنل چندجمله‌ای در دوره صحت‌سنجی نشان می‌دهد. مطابق با این شکل با افزایش تعداد ورودی‌ها به مدل SVM با تابع کرنل چندجمله‌ای درجه ۴، مقادیر شاخص‌های CC و RMSE به‌ترتیب روند صعودی و نزولی داشته و به‌ازای تعداد ورودی ۱۴ شاخص‌های مذکور به‌ترتیب به بیش‌ترین و کم‌ترین مقدار خود می‌رسند. بنابراین مدل SVM با تابع کرنل چندجمله‌ای و ۱۴ ورودی دارای دقت بیشتری در پیش‌بینی جریان ماهانه رودخانه خرخره‌چای نسبت به سایر مدل‌های SVM با تابع کرنل چندجمله‌ای و ورودی‌های مختلف می‌باشد.

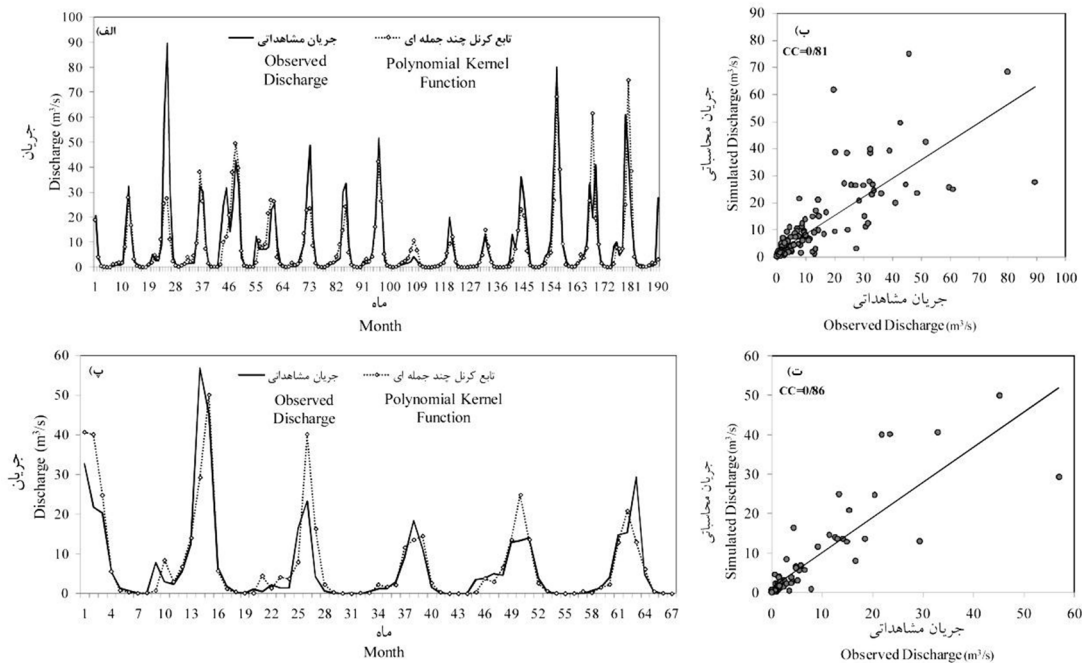
مطابق با این جدول مدل SVM با تابع کرنل چندجمله‌ای درجه ۴ و ۱۴ متغیر ورودی با مقادیر شاخص‌های CC و RMSE به‌ترتیب برابر با ۰/۸۶ و ۵/۸۸ (m<sup>3</sup>/s) مترمکعب بر ثانیه در دوره صحت‌سنجی نسبت به مدل‌های SVM دیگر بهترین کارایی را داشته است. همچنین پیش‌بینی‌ها نشان داد که مدل‌های SVM با توابع کرنل پایه شعاعی و چندجمله‌ای تقریباً از دقت یکسانی در تخمین رواناب برخوردارند و همچنین مدل SVM با تابع کرنل خطی نسبت به دیگر مدل‌های SVM عملکرد ضعیف‌تری دارد. در شکل ۹ مقادیر ضریب همبستگی (CC) و مجذور میانگین مربعات خطا (RMSE) را به‌ازای



شکل ۹- روند تغییرات ضریب همبستگی و مجدور میانگین مربعات خطا با توجه به افزایش تعداد ورودی‌ها در مدل SVM با تابع کرنل چندجمله‌ای.

Figure 9. Correlation coefficient (CC) and root mean square errors (RMSE) changes according to the increase in number of inputs to the SVM model with polynomial kernel function.

در شکل ۱۰ جریان پیش‌بینی شده توسط مدل SVM با تابع کرنل چندجمله‌ای درجه ۴ به‌ازای ۱۴ داده شده است. ورودی برای دوره‌های واسنجی و صحت‌سنجی نشان داده شده است.



شکل ۱۰- جریان مشاهداتی و پیش‌بینی شده توسط مدل SVM در دوره‌های واسنجی (الف و ب) و صحت‌سنجی (پ و ت).

Figure 10. Observed and predicted river flow by the SVM model in the training (a and b) and testing periods (p and t).

نتایج مدل‌های سری زمانی: جهت حصول اطمینان از نتایج توابع ACF و PACF مدل‌های سری زمانی AR(p) با تأخیرهای ۱ تا ۱۵، مدل ARMA(p,q) با مقادیر مختلف  $p$  و  $q$  و مدل ARIMA(p,d,q) با تأخیر، میانگین متحرک و تفاضل‌های مختلف به کار گرفته شدند. بر این اساس جریان رودخانه با ۸۱ مدل مختلف سری زمانی پیش‌بینی شده و عملکرد هر یک از آنها در دوره‌های واسنجی و صحت‌سنجی با استفاده از شاخص‌های آماری NS، CC، RMSE(m<sup>3</sup>/s) مورد ارزیابی قرار گرفت. قابل ذکر است که همانند مدل SVM، ۷۵ درصد ابتدایی داده‌های نرمال شده جریان جهت واسنجی مدل‌ها و ۲۵ درصد باقی‌مانده جهت صحت‌سنجی آنها انتخاب گردید و همچنین شاخص‌های آماری آورده شده در جداول پس از بازگردانی داده‌های جریان نرمال پیش‌بینی شده به داده‌های جریان واقعی پیش‌بینی شده محاسبه شده‌اند. نتایج نشان داد مدل ARMA(۶,۲) از بین ۸۱ مدل سری زمانی به کار گرفته شده با مقادیر شاخص‌های آماری NS، CC، RMSE به ترتیب برابر با ۰,۷۷، ۰,۷۶ و ۷/۹۶(m<sup>3</sup>/s) با یک ورودی از دقت قابل‌قبولی در پیش‌بینی جریان ماهانه برخوردار است که این موضوع نشان از توانایی تابع PACF در تعیین مؤثرترین ورودی‌ها دارد. لازم به ذکر است که CC و RMSE(m<sup>3</sup>/s) برای اکثر مدل‌های AR، ARMA و ARIMA در دوره واسنجی نزدیک به هم بوده ولی معیار اصلی انتخاب بهترین مدل جهت پیش‌بینی روناب ایستگاه سته نتایج به دست آمده در دوره صحت‌سنجی می‌باشد. عملکرد و روابط به دست آمده برای مدل‌های AR(۱) و ARMA(۶,۲) به ترتیب در جدول ۵ و رابطه‌های ۲۳ و ۲۴ آورده شده‌اند.

نتایج مدل‌های سری زمانی: جهت حصول اطمینان از نتایج توابع ACF و PACF مدل‌های سری زمانی AR(p) با تأخیرهای ۱ تا ۱۵، مدل ARMA(p,q) با مقادیر مختلف  $p$  و  $q$  و مدل ARIMA(p,d,q) با تأخیر، میانگین متحرک و تفاضل‌های مختلف به کار گرفته شدند. بر این اساس جریان رودخانه با ۸۱ مدل مختلف سری زمانی پیش‌بینی شده و عملکرد هر یک از آنها در دوره‌های واسنجی و صحت‌سنجی با استفاده از شاخص‌های آماری NS، CC، RMSE(m<sup>3</sup>/s) مورد ارزیابی قرار گرفت. قابل ذکر است که همانند مدل SVM، ۷۵ درصد ابتدایی داده‌های نرمال شده جریان جهت واسنجی مدل‌ها و ۲۵ درصد باقی‌مانده جهت صحت‌سنجی آنها انتخاب گردید و همچنین شاخص‌های آماری آورده شده در جداول پس از بازگردانی داده‌های جریان نرمال پیش‌بینی شده به داده‌های جریان واقعی پیش‌بینی شده محاسبه شده‌اند. نتایج نشان داد مدل ARMA(۶,۲) از بین ۸۱ مدل سری زمانی به کار گرفته شده با مقادیر شاخص‌های آماری NS، CC، RMSE به ترتیب برابر با ۰,۷۷، ۰,۷۶ و ۷/۹۶(m<sup>3</sup>/s) با یک ورودی از دقت قابل‌قبولی در پیش‌بینی جریان ماهانه برخوردار است که این موضوع نشان از توانایی تابع PACF در تعیین مؤثرترین ورودی‌ها دارد. لازم به ذکر است که CC و RMSE(m<sup>3</sup>/s) برای اکثر مدل‌های AR، ARMA و ARIMA در دوره واسنجی نزدیک به هم بوده ولی معیار اصلی انتخاب بهترین مدل جهت پیش‌بینی روناب ایستگاه سته نتایج به دست آمده در دوره صحت‌سنجی می‌باشد. عملکرد و روابط به دست آمده برای مدل‌های AR(۱) و ARMA(۶,۲) به ترتیب در جدول ۵ و رابطه‌های ۲۳ و ۲۴ آورده شده‌اند.

جدول ۵- ارزیابی عملکرد مدل‌های AR(۱) و ARMA(۶,۲) در تخمین جریان ماهانه حوضه خرخره‌چای در ایستگاه سته.

Table 5. Evaluation of AR (1) and ARMA(6,2) models performance in prediction of Kherkherechihy monthly river flow in the Santeh station.

دوره صحت‌سنجی Testing Period			دوره واسنجی Training Period			مدل Model
NS	RMSE(m <sup>3</sup> /s)	CC	NS	RMSE(m <sup>3</sup> /s)	CC	
0.54	7.95	0.76	0.52	8.41	0.82	AR(۱)
0.62	6.47	0.82	0.52	9.82	0.77	ARMA(۶,۲)

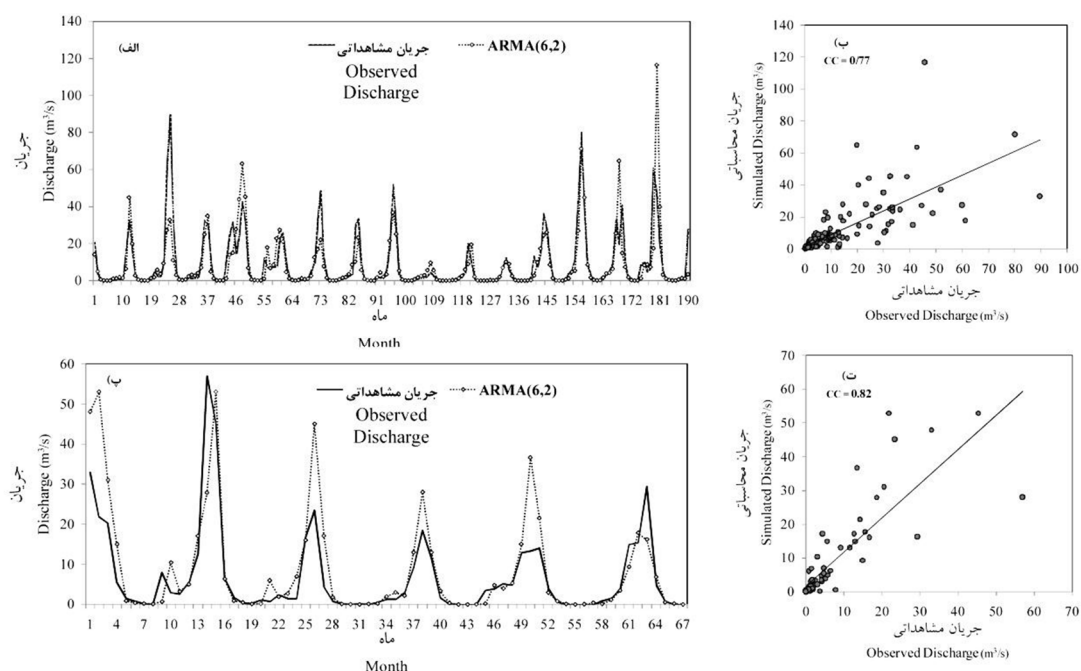
در این روابط  $Q_t$  جریان تخمین زده شده،  $Q_{t-1}$ ،  $Q_{t-2}$ ، ... و  $Q_{t-6}$  به ترتیب جریان ماهانه با تأخیرهای ۱ تا ۶ ماه و  $\varepsilon_t$ ،  $\varepsilon_{t-1}$  و  $\varepsilon_{t-2}$  به ترتیب نوفه سفید با تأخیرهای ۱ تا ۳ ماه می‌باشند.

$$Q_t = 0.02 + 0.73Q_{t-1} + \varepsilon_t \quad (23)$$

$$Q_t = 0.03 - 0.07Q_{t-1} - 0.43Q_{t-2} + 0.64Q_{t-3} + 0.05Q_{t-4} + 0.15Q_{t-5} + 0.13Q_{t-6} + 0.75\varepsilon_{t-1} + 0.95\varepsilon_{t-2} + \varepsilon_t \quad (24)$$

رودخانه در ماه‌های با مقادیر جریان متوسط و حداقل دارد ولی جهت برآورد جریان رودخانه در ماه‌های با جریان بیش‌تر عملکرد ضعیفی از خود نشان داده است.

مقادیر مشاهداتی و محاسباتی دوره‌های واسنجی و صحت‌سنجی مدل  $ARMA(6,2)$  به‌ترتیب در شکل ۱۱ آورده شده است. با توجه به این شکل مدل  $ARMA(6,2)$  در هر دو دوره واسنجی و صحت‌سنجی توانایی قابل‌قبولی در تخمین جریان



شکل ۱۱- جریان مشاهداتی و پیش‌بینی شده مدل  $ARMA(6,2)$  در دوره واسنجی (الف و ب) و صحت‌سنجی (پ و ت).

Figure 11. Observed and predicted river flow by the  $ARMA(6,2)$  model in the training (a and b) and testing periods (c and d).

سری زمانی و ماشین بردار پشتیبان به‌عنوان بهترین مدل‌ها انتخاب شدند که در جدول ۶ خلاصه‌ای از عملکرد این مدل‌ها در پیش‌بینی جریان رودخانه آورده شده است. براساس نتایج آمده در جدول ۶ مدل‌های SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی و  $ARMA(6,2)$  به لحاظ دقت در پیش‌بینی جریان به‌ترتیب در رتبه‌های اول و دوم قرار دارند.

مقایسه عملکرد مدل‌های سری زمانی و SVM: پس از پیش‌بینی جریان رودخانه خرخره‌چای در ایستگاه سنته با استفاده از دو مدل سری زمانی و SVM بهترین مدل به‌ازای تأخیرهای مختلف جریان ماهانه برای هر یک از آن‌ها با توجه به شاخص‌های آماری مشخص گردید. براساس نتایج بخش‌های قبل، مدل‌های  $ARMA(6,2)$  و SVM با تابع کرنل چندجمله‌ای درجه ۴ به‌ازای ۱۴ ورودی به‌ترتیب از میان مدل‌های



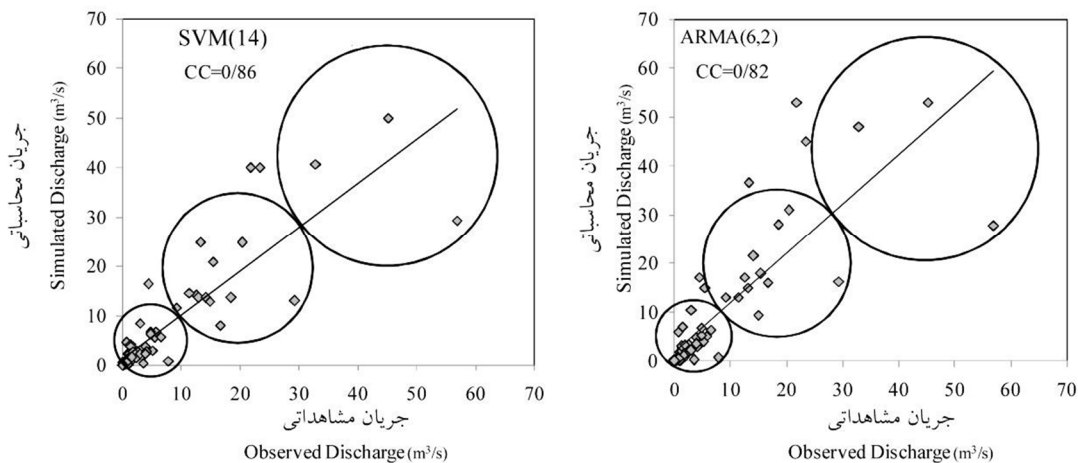
جدول ۶- ارزیابی کارایی مدل‌های ARMA و SVM در پیش‌بینی جریان ماهانه حوضه خرخره‌چای در ایستگاه سنته.

Table 6. Assessment of SVM and ARMA models performance in prediction of Kherkherehchiy monthly river flow in the Santeh station.

دوره صحت‌سنجی Testing Period			دوره واسنجی Training Period			مدل Model
NS	RMSE(m <sup>3</sup> /s)	CC	NS	RMSE(m <sup>3</sup> /s)	CC	
0.62	6.47	0.82	0.51	9.82	0.77	ARMA
0.73	5.88	0.86	0.51	8.82	0.81	SVM

انتخاب گردید. نمودار پراکنش مربوط به جریان پیش‌بینی‌شده توسط مدل‌های SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی و ARMA(۶،۲) در دوره صحت‌سنجی در شکل ۱۲ آورده شده است.

در نهایت مدل SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی به‌عنوان بهترین مدل از بین ۱۲۶ مدل مورد ارزیابی قرار گرفته در این پژوهش جهت پیش‌بینی جریان حوضه خرخره‌چای در ایستگاه سنته



شکل ۱۲- پراکنش مقادیر جریان مشاهداتی و تخمینی دوره صحت‌سنجی در مدل‌های منتخب ARMA و SVM در ۳ بازه مختلف.

Figure 12. Scatter of predicted river flow using SVM and ARMA models and observed values in the testing period for three ranges of discharge.

بیش‌تری نسبت به مدل ARMA(۶،۲) در پیش‌بینی جریان ماهانه در هر یک از سه محدوده مذکور برخوردار است. به‌طوری‌که خطای مدل SVM در هر یک از ۳ محدوده جریان کوچک‌تر از ۱۰، ۱۰ تا ۳۰ و بزرگ‌تر از ۳۰ مترمکعب بر ثانیه به‌ترتیب ۰/۶۴، ۴/۲۹ و ۲/۷ مترمکعب بر ثانیه کم‌تر از مدل ARMA(۶،۲) می‌باشد.

با توجه به این شکل مدل‌های SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی و ARMA(۶،۲) در تخمین جریان در ۳ محدوده جریان کوچک‌تر از ۱۰، ۱۰ تا ۳۰ و بزرگ‌تر از ۳۰ مترمکعب بر ثانیه عملکرد متفاوتی از خود نشان دادند که نتایج مربوط به آن‌ها در جدول ۷ آورده شده است. بر اساس مقادیر شاخص‌های آماری CC و RMSE آمده در این جدول، مدل SVM با تابع کرنل چندجمله‌ای از دقت

جدول ۷- ارزیابی دقت مدل‌های منتخب ARMA و SVM در تخمین مقادیر جریان (Q) در ۳ بازه مختلف.

Table 7. Evaluation of SVM and ARMA models accuracy in prediction of monthly river flow in three ranges of discharge.

30 ≤ Q ≤ 60		10 ≤ Q ≤ 30		0 ≤ Q ≤ 10		
ARMA	SVM	ARMA	SVM	ARMA	SVM	
18.90	16.83	13.81	9.52	3.02	2.38	RMSE(m <sup>3</sup> /s)
-0.92	-0.54	0.41	0.38	0.66	0.68	CC

میانگین متحرک و تفاضل‌های مختلف در پیش‌بینی جریان ماهانه حوضه خرخره‌چای تشخیص داده شد. نتایج نشان داد که مدل SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی با مقادیر شاخص‌های CC، NS و RMSE به‌ترتیب برابر با ۰/۷۳، ۰/۸۶ و ۵/۸۸ (m<sup>3</sup>/sec) در دوره تست (جدول ۴) نسبت به مدل‌های SVM با توابع کرنل خطی و RBF بهترین کارایی را دارا می‌باشد.

مقایسه مقادیر جریان پیش‌بینی شده در ۳ محدوده کوچک‌تر از ۱۰، ۱۰ تا ۳۰ و بزرگ‌تر از ۳۰ مترمکعب بر ثانیه توسط مدل‌های SVM و سری زمانی با مقادیر مشاهداتی آن‌ها در بازه‌های مذکور نشان داد که مدل SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی در هر یک ۳ محدوده جریان مذکور دارای خطای برآورد کم‌تری نسبت به مدل ARMA(۶،۲) می‌باشد. اگرچه در اکثر پژوهش‌های صورت گرفته در زمینه پیش‌بینی‌های هیدرولوژی عمدتاً از ماشین بردار پشتیبان با تابع پایه شعاعی استفاده می‌شود (۱۱)، اما برتری تابع کرنل چندجمله‌ای در پژوهش حاضر و مطالعاتی از قبیل ازانی و همکاران (۲۰۱۴)، لزوم بررسی کارایی توابع کرنل مختلف را نشان می‌دهد (۴).

مقایسه عملکرد مدل‌های مورد بررسی نشان داد که مدل SVM با تابع کرنل چندجمله‌ای به‌ازای ۱۴ ورودی از دقت و کارایی بیش‌تری در پیش‌بینی جریان ماهانه حوضه خرخره‌چای در ایستگاه سنته نسبت به مدل‌های سری زمانی AR، ARMA و ARIMA برخوردار است.

بررسی مطالعات صورت گرفته نیز نشان داد در پژوهش‌هایی از قبیل جیان و همکاران (۲۰۰۶) و محرم‌پور و همکاران (۲۰۱۳) مدل SVM با توابع کرنل مختلف نسبت به مدل‌های سری زمانی ARMA و ARIMA عملکرد بهتری در پیش‌بینی جریان رودخانه‌ها از خود نشان می‌دهد (۱۴، ۱۹) که این موضوع با نتایج پژوهش حاضر مطابقت دارد.

### نتیجه‌گیری

در پژوهش حاضر سعی گردید عملکرد مدل‌های سری زمانی و ماشین بردار پشتیبان (SVM) در پیش‌بینی جریان ماهانه حوضه خرخره‌چای در ایستگاه سنته مورد بررسی قرار گرفته و دقت این مدل‌ها در برآورد جریان ماهانه این ایستگاه مقایسه شود. در این راستا ابتدا توزیع احتمالاتی داده‌های جریان مشاهداتی به نرمال تبدیل گردیده و سپس آزمون‌های کلموگروف-اسمیرنوف و شاپیرو-ویلک در سطح اطمینان ۰/۹۵ برای آن‌ها انجام گرفت. در ادامه جهت لحاظ نمودن تأثیر جریان با تأخیرهای ماهانه مختلف و همچنین ساختار به‌کار رفته در مدل‌ها بر عملکرد آن‌ها در پیش‌بینی جریان ماهانه، کارایی مدل‌های SVM با توابع کرنل مختلف و سری زمانی به‌ازای تأخیرهای ۱ تا ۱۵ جریان ماهانه مورد ارزیابی قرار گرفت و نتایج زیر به‌دست آمد.

مدل ARMA(۶،۲) با مقادیر شاخص‌های آماری CC، NS و RMSE به‌ترتیب برابر با ۰/۸۲، ۰/۶۲ و ۶/۴۷ (m<sup>3</sup>/sec) در دوره تست، بهترین مدل از بین مدل‌های سری زمانی به‌کار گرفته شده با تأخیر،

## منابع

1. Adamowski, J. 2013. Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India. *Versita J.* 45: 1. 71-83.
2. Adeli, A., Fathi-Moghadam, M., and Musavi Jahromi, H. 2014. Using Stochastic Models to Produce Artificial Time Series and Inflow Prediction: A Case Study of Talog Dam reservoir, Khuzestan Province, Iran. *J. Inter. Bull. Water Resour. Dev.* 2: 5. 1-13. (In Persian)
3. Asefa, T., Kemblowski, M., McKee, M., and Khalil, A. 2005. Multi-time scale stream flow predictions: The Support vector machines approach. *J. Hydrol.* 318: 1-4. 7-16.
4. Azani, A., Fazelifard, M.H., and Ghorbani, M.A. 2014. Simulation of Urmia Lake water level using support vector machines and artificial neural network. The 13<sup>th</sup> Conference of the hydraulic Iran, Tabriz University. (In Persian)
5. Baofeng, G., Gunn, S.R., Damper, R.I., and Nelson, J.D.B. 2008. Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification. *IEEE Transactions on Image Processing.* 17: 4. 622-629.
6. Bani Habib, M.A., and Valipour, M. 2008. Comparative assessment of ARMA, ARIMA and autocorrelated artificial neural network models in forecasting inflow to the Dez reservoir. First International Conference on Water Crisis, Zabol University. (In Persian)
7. Basak, D., Pal, S., and Patranabis, D.C. 2007. Support vector regression. *Neural Inf. Process.* 11: 203-225.
8. Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control.* Third edition, Prentice Hall. 598p.
9. Damle, C., and Yalcin, A. 2007. Flood prediction using time series data mining. *J. Hydrol.* 333: 2-4. 305-316.
10. Dibike, Y., Velickov, S., Solomatine, D., and Abbott, M. 2001. Model induction with of support vector machines: Introduction and applications. *J. Comp. Civil Engin.* 15: 3. 208-216.
11. Eskandari, A., Nouri, R., Meraji, H., and Kiaghadi, A. 2012. Development of appropriate model based on artificial neural network and support vector machine for forecasting 5-Days Biochemical Oxygen Demand (BOD5). *J. Ecol.* 61: 71-82. (In Persian)
12. Fletcher, R. 1987. *Practical Methods of Optimization.* Wiley, New York. 456p.
13. Jain, A., and Kumar, A.M. 2007. Hybrid neural network models for hydrologic time series forecasting. *Appl. Soft Com. J.* 7: 2. 585-592.
14. Jian, Y., Cheng, C.T., and Chau, K.W. 2006. Using support vector machines for long-term discharge prediction. *J. Hydrol. Sci. – des Sci. Hydrol.* 51: 4. 599-612.
15. Kavzoglu, T., and Colkesen, I. 2009. A kernel functions analysis for support vector machines for land cover classification. *Inter. J. Appl. Earth Obs. Inf.* 11: 5. 352-359.
16. Kakaei Lafadani, E., Moghaddam Nia, A., Ahmadi, A., Jajarmizadeh, M., and Ghafari, M. 2013. Stream flow simulation using SVM, ANFIS and NAM models (A Case study). *Caspian J. Appl. Sci. Res.* 2: 4. 86-93.
17. Liu, G.Q. 2011. Comparison of Regression and ARIMA models with Neural Network models to forecast the daily streamflow. PhD thesis, University of Delaware. 545p.
18. Misra, D., Oommen, T., Agarwal, A., and Mishra, S.K. 2009. Application and analysis of Support Vector machine based simulation for runoff and sediment yield. *J. Biosyst. Engin.* 103: 9. 527-535.
19. Moharrampour, M., Mehrabi, A., Hajikandi, H., and Sohrabi, S. 2013. Comparison of Support Vector Machines (SVM) and Autoregressive integrated moving average (ARIMA) in daily flow forecasting. *J. River Engin.* 1: 1. 34-45.
20. Salas, J.D. 1993. Analysis and modeling of hydrological time series. P 1-19, In: R, David (Ed.), *Handbook of Hydrology*, McGraw-Hill, New York.
21. Vapnik, V., and Chervonenkis, A. 1991. The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognition and Image Analysis.* 1: 3. 283-305.

22. Vapnik, V., and Cortes, C. 1995. Support vector networks. *Machine Learning*. 20: 273-297.
23. Wei, W.W.S. 2006. *Time Series analysis: univariate and multivariate methods* (second edition). Greg Tobin Publisher, ISBN 0-321-32216-9.
24. Yang, K., Shan, G., and Zhao, L. 2006. Correlation Coefficient Method for Support Vector Machine Input Samples. *International Conference on Machine Learning and Cybernetics*. Pp: 2857-2861.
25. Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., and Lee, K.K. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* 396: 1-2. 128-138.
26. Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neuro computing*. 50: 159-175.



## **Assessment of kernel functions performance in river flow estimation using support vector machine**

**\*M. Isazadeh<sup>1</sup>, H. Ahmadzadeh<sup>2</sup> and M.A. Ghorbani<sup>3</sup>**

<sup>1</sup>M.Sc. Student, Dept. of Water Engineering, Tabriz University, <sup>2</sup>Ph.D Student, Dept. of Water Engineering, Tabriz University, <sup>3</sup>Associate Prof., Dept. of Water Engineering, Tabriz University

Received: 03/13/2015; Accepted: 01/10/2016

### **Abstract**

**Background and Objectives:** Accurate prediction of river flow has an important role in the optimum management of available water resources. In recent years, support vector machine (SVM) that is one of the most important data-driven models, has been considered in this regards. This model is a useful learning system based on constrained optimization theory that uses induction of structural error minimization principle and results a general optimized answer. Such as other data mining models, the SVM model can also be used for runoff simulation when the only available data is runoff (autoregressive simulation). Typically, three kernel functions, namely, radial basis, polynomial of degree  $d$  and linear are applied in SVM that use of each function with various parameters for river flow estimation may have different results. Therefore, it is necessary to evaluate the accuracy of each of these functions and select the appropriate kernel function for runoff simulation. Since time series models, namely, AR, ARMA and ARIMA are the main models for autoregressive simulation of runoff, relative accuracy of kernel functions can be investigated by comparing their performance with these models. Therefore, assessment of the accuracy of kernel functions for monthly river flow simulation and comparison of their performance with time series models is main aim of this study.

**Materials and Methods:** In this study Kherkherehchiy river basin was selected as the study area and observed monthly river flow of this basin in the Santeh gauging station were applied for calibration and validation of models. For this purpose, first 75 percent of monthly river flow data (1367-1384) were selected to calibrate models and 25 percent of data (1385-1390) were used to validate models. Next, probability distribution of monthly river flow data in Santeh station were studied based on Kolmogorov-Smirnov and Shapiro-Wilk test and then normalization of data distribution were done. After optimization of parameters for each kernel functions the monthly flow values were predicted in Santeh station and the performance of these functions were evaluated using root mean square errors (RMSE) and the correlation coefficient (CC).

**Results:** The investigations of this study indicated that although there is no significant difference in the results of three kernel functions, but the polynomial kernel function of degree 4 with CC and RMSE values of 0.86 and 5.88 ( $m^3/sec$ ) respectively in the testing period, has high accuracy and better performance in prediction of monthly flow in comparison to other kernel functions. Also the results showed that ARMA(6,2) with CC and RMSE values of 0.82 and 6.47 ( $m^3/sec$ ) respectively in the testing period, has good performance in prediction of Kherkherehchiy monthly flow compared to the other time series models.

**Conclusion:** Finally, the predicted monthly river flow using polynomial kernel function of degree 4 (as a representative of SVM model) was compared with the results of ARMA(6,2) (as a representative of time series model) and this conclusion was obtained that the SVM model has a better performance than time series models in the monthly river flow prediction of the Kherkherehchiy basin.

**Keywords:** Kernel functions, Kherkherehchiy river basin, River flow prediction, Support vector machine, Time series models

---

\* Corresponding Author; Email: mohammadisazade@gmail.com

