

## مقایسه و ارزیابی مدل‌های شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی در تخمین بده جریان؛ مطالعه موردی حوضه صوفی چای

تاریخ دریافت: ۱۳۹۳/۰۸/۲۳

تاریخ پذیرش: ۱۳۹۴/۰۶/۰۹

رضا دهقانی<sup>۱</sup>، محمد علی قربانی<sup>۲</sup>، محمد تشنه لب<sup>۳</sup>، امیر ریخته گر غیائی<sup>۴</sup>، اسماعیل اسدی<sup>۵</sup>

### چکیده

پیش‌بینی جریان رودخانه برای برنامه‌ریزی طراحی و مدیریت مطمئن پروژه‌های منابع آب مهم است. در این پژوهش قابلیت کاربرد شبکه عصبی بیزین، برنامه‌ریزی ژن، ماشین بردار پشتیبان و رگرسیون خطی چندمتغیره برای پیش‌بینی سری زمانی جریان رودخانه صوفی چای بررسی شد. سری زمانی جریان روزانه رودخانه برای دوره ۱۳۷۶ تا ۱۳۸۹ برای ایستگاه هیدرومتری تازه کند رودخانه صوفی چای مورد استفاده واقع شد. جهت بدست آوردن بهترین نگاشت ورودی و خروجی؛ ترکیبات مختلف ورودی با استفاده از داده‌های پیشین جریان روزانه رودخانه محاسبه شد. کارایی مدل‌ها با استفاده از چهار معیار خطای: ضریب همبستگی (CC)، ریشه جذر میانگین خطا (RMSE) و ضریب نش ساتکلیف و بایاس محاسبه شد. مقایسه مدل‌ها نشان داد که شبکه عصبی بیزین با ضریب همبستگی (۰/۹۹۱)، جذر میانگین مربعات خطا (۰/۰۳۱ m<sup>3</sup>/s)، نش ساتکلیف (۰/۹۸۱) و مقدار بایاس (-۰/۰۰۶) بهترین تخمین را از داده‌های روزانه جریان رودخانه نسبت به مدل‌های برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی چندمتغیره داشته است.

واژه‌های کلیدی: برنامه‌ریزی بیان ژن، دبی جریان، شبکه عصبی بیزین، ماشین بردار پشتیبان.

<sup>۱</sup> کارشناسی ارشد کشاورزی-منابع آب، گروه مهندسی آب، دانشگاه تبریز، لرستان-خرم‌آباد-خیابان جلال آل احمد، ایران، تلفن: ۰۹۳۷۸۲۳۱۶۴۲

پست الکترونیکی: [reza.dehghani67@yahoo.com](mailto:reza.dehghani67@yahoo.com) (مسئول مکاتبه)

<sup>۲</sup> دانشیار، گروه مهندسی آب، دانشگاه تبریز، ایران، تلفن: ۰۹۱۴۴۱۴۰۳۸۱ پست الکترونیکی: [Ghorbani@tabrizu.ac.ir](mailto:Ghorbani@tabrizu.ac.ir)

<sup>۳</sup> استاد گروه مهندسی برق - کنترل، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران، تلفن: ۰۹۱۲۱۰۵۴۴۶۶ پست الکترونیکی [teshnehlabb@eetd.kntu.ac.ir](mailto:teshnehlabb@eetd.kntu.ac.ir)

<sup>۴</sup> استادیار، گروه مهندسی برق - کنترل، دانشگاه تبریز، ایران، تلفن: ۰۹۱۴۴۱۰۲۲۶۱ پست الکترونیکی: [agiasi@tabrizu.ac.ir](mailto:agiasi@tabrizu.ac.ir)

<sup>۵</sup> استادیار، گروه مهندسی آب، دانشگاه تبریز، ایران، تلفن: ۰۹۱۴۴۰۶۲۴۸۶ پست الکترونیکی: [esasadi@gmail.com](mailto:esasadi@gmail.com)

## مقدمه

امروزه با توجه به اهمیت و حساسیت امر مهار آب‌های سطحی خصوصاً در کشور ما که اکثر رودخانه‌های مناطق مختلف فصلی بوده و کمبود آبی که در پهنه وسیعی از کشور وجود دارد، نیاز به شناسایی و به مدل درآوردن رفتار رودها و شریان‌های آبی جهت برنامه‌ریزی‌های بلندمدت و استفاده بیشتر و بهتر از پتانسیل‌های آن‌ها ضروری است. لذا انتخاب مدلی که بتواند با استفاده از عوامل تأثیرگذار، جریان ورودی را به‌طور قابل قبولی برآورد نماید امری ضروری به نظر می‌رسد. امروزه سیستم‌های هوشمند به‌طور گسترده برای پیش‌بینی پدیده‌های غیرخطی مورد استفاده قرار می‌گیرد، که روش شبکه عصبی بیزین ( $BNN^1$ )، برنامه‌ریزی بیان ژن ( $GEP^2$ ) و ماشین بردار پشتیبان ( $S.V.M^3$ ) از جمله‌ی این روش‌ها است. شبکه‌های عصبی بیزین به‌عنوان یکی از روش‌های مبتنی بر شبکه‌های عصبی در مدل کردن مسایل غیرخطی و پیچیده از طریق الگوریتم‌های خاص و روش‌های آماری است. با استفاده از این روش می‌توان به مدل کردن روابط علت و معلولی یک فرآیند، آنالیز وضعیت موجود و تخمین وضعیت آینده یک سیستم پرداخت (طباطبایی و دشتی زاده، ۱۳۸۷). برنامه‌ریزی ژن یک تکنیک برنامه‌ریزی خودکار است که راه‌حل مساله را با استفاده از برنامه‌ریزی کامپیوتر ارایه کرده و عضوی از خانواده الگوریتم تکاملی می‌باشد (قربانی و صالحی، ۱۳۹۰). ماشین بردار پشتیبان نیز یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است. در سال‌های اخیر استفاده از روش‌های هوشمند مورد بررسی در مطالعات پیش‌بینی جریان ورودی به مخازن سدها مورد توجه محققین قرار گرفته است، که از جمله می‌توان به موارد ذیل اشاره نمود: در پژوهشی به‌منظور پیش‌بینی جریان روزانه رودخانه ليقوان از مدل برنامه‌ریزی بیان ژن استفاده نمودند نتایج این تحقیق

نشان داد که مدل برنامه‌ریزی بیان ژن از دقت بسیار بالایی نسبت به روش شبکه‌های عصبی مصنوعی و مدل‌های سری زمانی برخوردار است (فربودنام و همکاران، ۱۳۸۸). در پژوهشی دیگر جهت بررسی همبستگی روزانه بین ایستگاه‌های هیدرومتری در پیش‌بینی جریان ورودی به مخزن سد دز از مدل شبکه عصبی مصنوعی و رگرسیون خطی چند متغیره استفاده نمودند نتایج نشان داد با در نظر گرفتن ضریب همبستگی و معیار میانگین مربعات خطاها، مدل شبکه عصبی مصنوعی عملکرد بهتری نسبت به مدل رگرسیون خطی دارد (بنی حبیب و همکاران، ۱۳۸۹). نتایج حاصل از پژوهش استفاده از مدل ماشین بردار پشتیبان جهت پیش‌بینی دبی روزانه رودخانه قره‌سو حاکی از دقت و کارایی مناسب این مدل است (محرّم‌پور و همکاران، ۱۳۹۰). از سوی دیگر در پژوهشی کاربرد شبکه عصبی بیزین را جهت شبیه‌سازی فرآیند بارش - رواناب با استفاده از داده‌های روزانه حوضه ساگونی<sup>۴</sup> در کانادا مورد بررسی قرار دادند که نتایج حاصله نشان داد مدل شبکه عصبی بیزین خطایی ناچیز در تخمین رواناب داشته و نیز نسبت به روش شبکه عصبی مصنوعی در تخمین اکثر مقادیر دقت بالایی دارد (Khan and Coulibaly., 2006). جهت پیش‌بینی دبی جریان روزانه رودخانه شوپیل<sup>۵</sup> در آمریکا از روش برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی استفاده نمود و نشان داد که هر دو روش نتایج قابل قبولی داشته، ولی برنامه‌ریزی بیان ژن از دقت بالاتری نسبت به شبکه عصبی مصنوعی برخوردار است (Güven., 2009). در پژوهشی کاربرد شبکه عصبی بیزین را در ایستگاه جاتیسرنو واقع در اندونزی جهت تخمین بارش ساعتی مورد بررسی قرار دادند و نتایج نشان داد که شبکه عصبی بیزین دقت بالا و خطایی ناچیز نسبت به شبکه‌های عصبی مصنوعی دارد (Fithriasari et al., 2013). در پژوهشی دیگر جهت پیش‌بینی جریان رودخانه در منطقه‌ای کوهستانی و نیمه‌خشک واقع در شمال غربی

<sup>1</sup>.Bayesian Neural Network

<sup>2</sup>. Gene Expression Programming

<sup>3</sup>.Support Vector Mashin

<sup>4</sup> Saguenay

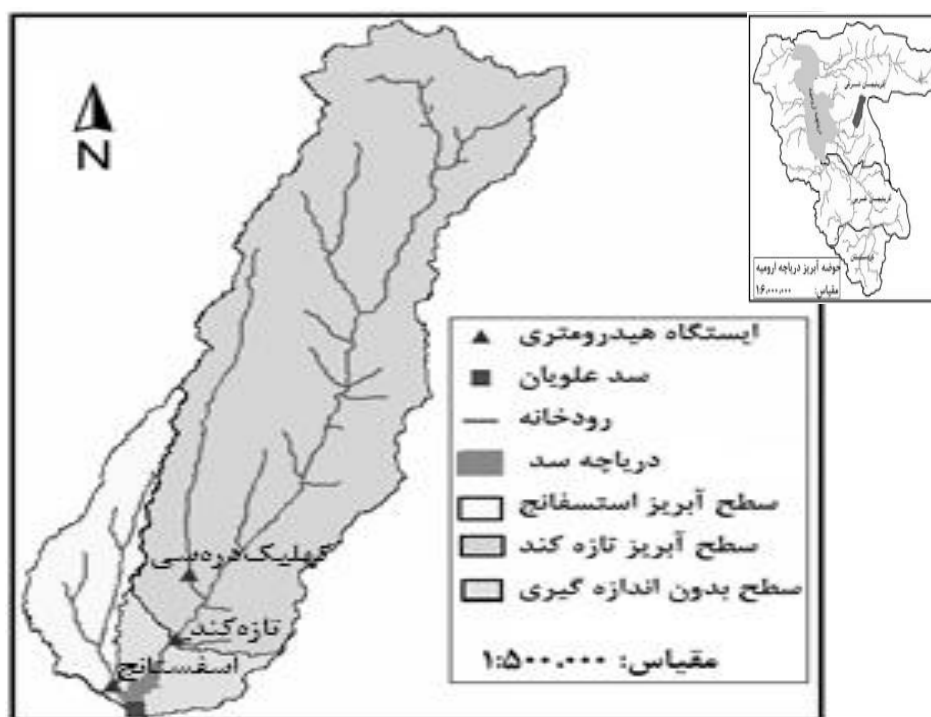
<sup>5</sup> .Shevell River

## مواد و روش‌ها

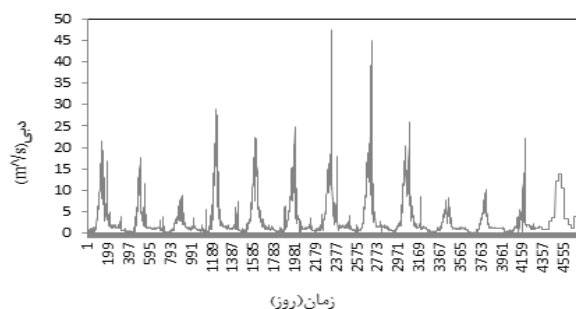
### منطقه مورد مطالعه و داده‌های مورد استفاده

رودخانه صوفی چای با طول ۱۲۰ کیلومتر از دامنه‌های جنوب غربی کوه سهند واقع در استان آذربایجان شرقی سرچشمه می‌گیرد. این رودخانه در داخل حوضه آبریز صوفی چای، یکی از زیر حوضه‌های دریاچه ارومیه، قرار دارد. در بالادست سد علویان، پنج ایستگاه هیدرومتری خرما زرد، تازه کند، چکان، اسفستانج و کهلیک دره سی احداث شده‌اند که نزدیک‌ترین ایستگاه به مخزن سد، ایستگاه تازه کند بوده و در این تحقیق از اطلاعات این ایستگاه هیدرومتری استفاده گردید. در شکل (۱) موقعیت جغرافیایی منطقه مورد مطالعه نشان داده شد. در این تحقیق، از داده‌های دبی ورودی به سد علویان در مقیاس زمانی روزانه از سال ۱۳۷۶ تا ۱۳۸۹ استفاده گردید. در شکل (۲) نمودار تغییرات دبی و در جدول (۱) خصوصیات آماری پارامتر دبی در بازه زمانی مذکور نشان داده شد.

چین از ماشین بردار پشتیبان استفاده نمودند و دریافتند که مدل ماشین بردار پشتیبان عملکردی مناسب جهت پیش‌بینی جریان رودخانه در مناطق کوهستانی نیمه‌خشک داشته و نسبت به مدل‌های شبکه عصبی مصنوعی و سیستم استنتاج فازی از دقت مطلوبی برخوردار است (He et al., 2014). در مجموع با توجه به پژوهش‌های انجام‌شده و ذکر این نکته که ایستگاه‌های هیدرومتری بالادست سد علویان به‌عنوان مهم‌ترین منبع تأمین‌کننده آب بخش‌های مختلف و نواحی مجاور خود می‌باشد، بنابراین اهمیت تخمین جریان و اقدامات مدیریتی جهت بهبود بهره‌برداری بهینه از مخازن سد ضروری است. لذا، هدف از این تحقیق تخمین دبی جریان در ایستگاه هیدرومتری بالادست سد علویان با کمک شبکه عصبی بیزین و مقایسه نتایج آن با برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی می‌باشد.



شکل (۱): موقعیت جغرافیایی ایستگاههای هیدرومتری حوضه صوفی چای



شکل (۲): مقادیر مشاهداتی دبی روزانه ایستگاه هیدرومتری (۱۳۷۶-۱۳۸۹)

جدول (۱): خصوصیات آماری مقادیر دبی روزانه مورد استفاده در بازه زمانی (۱۳۷۶-۱۳۸۹)

پارامتر	واحد	تعداد رکورد	میانگین	مینیمم	ماکزیمم	انحراف معیار	ضریب تغییرات	چولگی
Q(t)	m <sup>3</sup> /s	۴۷۴۴	۳/۱۶۳	۰	۴۷/۴۰۰	۴/۲۷۸	۱/۳۵۲	۲/۸۵۹

### شبکه‌های عصبی بیزین

این روش یکی از روش‌های سیستم پشتیبان تصمیم‌گیری می‌باشد که ابزار قدرتمندی در مدل کردن روابط علی و معلولی در قالب شبکه‌ای از احتمالات است. نکته بسیار مهم در مورد روش شبکه عصبی بیزین این است که این روش به اطلاعات دقیق و تاریخچه کامل یک واقعیت نیاز ندارد بلکه می‌تواند با استفاده از اطلاعات ناقص و غیردقیق نیز به نتایج بسیار قانع‌کننده‌ای در زمینه تخمین وضعیت فعلی یا آینده یک سیستم دست یابد. همچنین یک روش سازگار و انعطاف‌پذیر برای مدل کردن موقعیت‌های غیرقطعی و یک مدل گرافیکی مبتنی بر ادراک مستقیم از اندرکنش میان علل و معلول‌های مختلف ارائه می‌نماید. همچنین یک روش بسیار مفید در مدل کردن موقعیت‌های نامطمئن و غیرقطعی بر اساس روابط علت و معلولی است. در شبکه‌های عصبی یکی از روش‌های دستیابی مؤثر به پارامترهای بهینه، استفاده از تنظیم بیزین<sup>۱</sup> است که به‌طور اتوماتیک مقادیری مناسب

برای پارامترهای تابع قرار می‌دهد. روش تنظیم بیزین می‌تواند جهت بهبود توانایی شبکه عصبی و آموزش تابع هدف F که بصورت زیر نشان داده شده است استفاده شود (MacKay., 1992).

$$F = \beta E_D + \alpha E_W \quad (1)$$

که در آن EW مجموع مربعات وزن‌های شبکه و ED مجموع مربعات باقی مانده بین پاسخ شبکه و تابع هدف می‌باشد.  $\alpha$  و  $\beta$  نیز پارامترهای تابع هدف می‌باشند (پارامترهای تنظیم). که هر یک از این پارامترها به آموزش شبکه در کاهش باقی مانده خروجی‌ها یا حجم شبکه بستگی دارد. نکته اساسی روش تنظیم این است که پارامترهای تابع هدف را چگونه از طریق داده‌های آماری بیزین انتخاب و بهینه نماید. بطوریکه اگر  $\alpha$  و  $\beta$  بعنوان متغیرهای تصادفی در نظر گرفته شوند قانون بیزین بصورت فرمول زیر تعریف می‌شود (MacKay., 1992).

<sup>1</sup>. Bayesian

$$P(D | \alpha, \beta, M) = \frac{Z_F(\alpha, \beta)}{Z_w(\alpha) Z_D(\beta)} \quad (5)$$

که

$$Z_D(\beta) = \left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} \text{ و } Z_w(\alpha) = \left(\frac{\pi}{\alpha}\right)^{\frac{N}{2}}$$

$$\text{ و } Z_F(\alpha, \beta) = (2\pi)^{\frac{N}{2}} \det^{-\frac{1}{2}}(H) \exp(-F(w^{MP}))$$

$$H = \beta \nabla^2 E_D + \alpha \nabla^2 E_w$$

ماتریس Hessian تابع هدف F می‌باشد. در معادله (5) روش‌های لگاریتم و مشتق‌گیری به ترتیب بکار گرفته شده است با فرض اینکه معادله تفاضلی تابع برابر صفر قرار گیرد مقدار  $P(\alpha, \beta | D, M)$  به بیشینه شده و احتمال اولیه وزن‌ها مینیمم می‌شود. سپس  $\alpha$  و  $\beta$  بصورت زیر بیان می‌شوند (MacKay., 1992).

$$\alpha^{MP} = \frac{\gamma}{2E_w^{MP}} \quad (6)$$

$$\beta^{MP} = \frac{n-\gamma}{2E_D^{MP}} \quad (7)$$

$$\gamma = N - \alpha^{MP} \text{trace}^{-1} H^{MP} \quad (8)$$

که در آن n تعداد نمونه ، N مقدار کل پارامترهای شبکه،  $\gamma$  تعداد پارامترهای موثر که نسبتا اثر بیشتری در کاهش میزان خطای توابع دارند. در ابتدا  $\alpha$  و  $\beta$  بر طبق الگوریتم لونیبرگ مارکوانت فرض می‌شود و تابع هدف  $F(w)$  با آموزش شبکه منظم بیزین به کم‌ترین مقدار خود می‌رسد.  $\alpha$  و  $\beta$  بر طبق معادلات 6، 7 و 8 بروزرسانی شده و سپس مقادیر بهینه توزیع اولیه بدست می‌آید و حداقل مقدار جدید  $F(w)$  محاسبه می‌گردد و در نهایت آموزش شبکه آنقدر تکرار می‌شود تا به همگرایی برسد (Foresee., 1997). نمونه‌ایی از ساختار شبکه چهار لایه‌ای متشکل از یک لایه ورودی، یک لایه مخفی، و یک لایه خروجی در شکل (3) نشان داده شده است.

$$P(\alpha, \beta | D, M) = \frac{P(D | \alpha, \beta, M) P(\alpha, \beta | M)}{P(D | M)} \quad (2)$$

در فرمول بالا D داده‌های آموزش، M مدل شبکه و W نیز وزن شبکه می‌باشد. بر طبق قانون بیزین اگر  $\alpha$  و  $\beta$  از توزیع یکنواخت پیروی نمایند در این صورت احتمال  $P(D | \alpha, \beta, M)$  زمانی ماکزیمم می‌شود که احتمال توزیع اولیه  $\alpha$  و  $\beta$  در معادله (2) حداکثر به مقدار ماکزیمم خواهد رسید. با فرض اینکه باقی مانده‌ها و وزن‌ها متغیرهای تصادفی باشند و براساس قانون بیزین معادله (3) بصورت زیر می‌باشد.

$$P(w | D, \alpha, \beta, M) = \frac{P(D | w, \beta, M) P(w | \alpha, M)}{P(D | \alpha, \beta, M)} \quad (3)$$

در صورتی که فرض شود باقی مانده‌ها و وزن‌ها از توزیع گوسین پیروی کرده داریم:

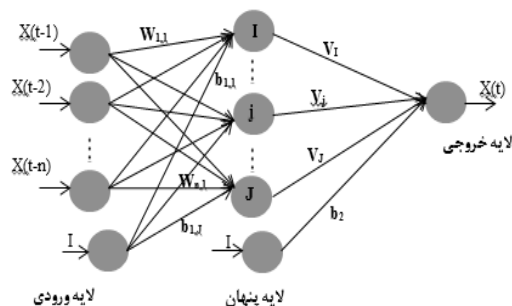
$$P(D | w, \beta, M) = \frac{\exp(-\beta E_D)}{Z_D(\beta)} \quad (4)$$

$$P(w | \alpha, M) = \frac{\exp(-\alpha E_w)}{Z_w(\alpha)} \quad (a-4)$$

برای اطمینان از اینکه  $P(D | \alpha, \beta, M)$  بعنوان فاکتور تنظیم در معادله (3) گردد معادله (b-4) بصورت زیر می‌باشد.

$$P(w | D, \alpha, \beta, M) = \frac{\exp(-F(w))}{Z_F(\alpha, \beta)} \quad (b-4)$$

اگر معادله‌های a-4 و b-4 با هم در فرمول (3) جایگذاری شوند آنگاه خواهیم داشت.



شکل (۳): نمای کلی یک شبکه عصبی بیزین سه لایه‌ای

برنامه‌ریزی بیان ژن روش برنامه‌ریزی بیان ژن، در سال ۱۹۹۹ توسط فریرا ارایه شد (Ferreira.,2001). این روش ترکیبی از روش‌های برنامه‌ریزی ژنتیک (GP) و الگوریتم ژنتیک بوده (GA) که در آن، کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک، ترکیب می‌شوند. از آنجایی‌که در این روش تمام ساختارهای شاخه‌ای با اندازه و اشکال متفاوت، در کروموزوم‌های خطی با طول ثابت کدگذاری می‌شوند، سبب گردیده که در این روش فنوتیپ و ژنوتیپ از هم جدا شوند و سیستم بتواند از تمام مزایای تکاملی به سبب وجود آنها بهره‌مند شود. اکنون باوجود اینکه فنوتیپ در GEP، همان نوع از ساختارهای شاخه‌ای مورد استفاده در GP را شامل می‌شود، اما ساختارهای شاخه‌ای که به وسیله GEP استنتاج می‌شوند (که بیان درختی نیز نامیده می‌شود) مبین تمامی ژنوم‌های مستقل هستند. به‌طور خلاصه می‌توان گفت در GEP به‌سازی‌ها در یک ساختار خطی اتفاق افتاده و سپس به‌صورت ساختار درختی بیان می‌شود و این موجب می‌شود تنها ژنوم اصلاح‌شده به نسل بعد منتقل‌شده و نیازی به ساختارهای سنگین برای تکثیر و جهش وجود نداشته باشد(فریرا، ۲۰۰۱). در این روش پدیده‌های مختلف با استفاده از مجموعه‌ای از توابع و مجموعه‌ای از ترمینال‌ها، مدل‌سازی می‌شوند. مجموعه توابع، معمولاً شامل توابع اصلی حسابی

### برنامه‌ریزی بیان ژن

برنامه‌ریزی بیان ژن، در سال ۱۹۹۹ توسط فریرا ارایه شد (Ferreira.,2001). این روش ترکیبی از روش‌های برنامه‌ریزی ژنتیک (GP) و الگوریتم ژنتیک بوده (GA) که در آن، کروموزوم‌های خطی و ساده با طول ثابت، مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت، مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک، ترکیب می‌شوند. از آنجایی‌که در این روش تمام ساختارهای شاخه‌ای با اندازه و اشکال متفاوت، در کروموزوم‌های خطی با طول ثابت کدگذاری می‌شوند، سبب گردیده که در این روش فنوتیپ و ژنوتیپ از هم جدا شوند و سیستم بتواند از تمام مزایای تکاملی به سبب وجود آنها بهره‌مند شود. اکنون باوجود اینکه فنوتیپ در GEP، همان نوع از ساختارهای شاخه‌ای مورد استفاده در GP را شامل می‌شود، اما ساختارهای شاخه‌ای که به وسیله GEP استنتاج می‌شوند (که بیان درختی نیز نامیده می‌شود) مبین تمامی ژنوم‌های مستقل هستند. به‌طور خلاصه می‌توان گفت در GEP به‌سازی‌ها در یک ساختار خطی اتفاق افتاده و سپس به‌صورت ساختار درختی بیان می‌شود و این موجب می‌شود تنها ژنوم اصلاح‌شده به نسل بعد منتقل‌شده و نیازی به ساختارهای سنگین برای تکثیر و جهش وجود نداشته باشد(فریرا، ۲۰۰۱). در این روش پدیده‌های مختلف با استفاده از مجموعه‌ای از توابع و مجموعه‌ای از ترمینال‌ها، مدل‌سازی می‌شوند. مجموعه توابع، معمولاً شامل توابع اصلی حسابی

### ماشین بردار پشتیبان

اولین کاربرد این روش در مسائل آب توسط دیباک و همکاران در سال ۲۰۰۱ با مدل‌سازی بارندگی- رواناب ارائه شد. 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. 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  $f(x)$  به علاوه مقداری اغتشاش (خطای مجاز  $\epsilon$ ) مشخص شود.

$$f(x) = W^T \cdot \phi(x) + b \quad (9)$$

$$y = f(x) + \text{noise} \quad (10)$$

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) \quad (14)$$

فرآیند محاسبات این مدل، با کدنویسی در محیط متلب، انجام شد و پارامترها  $(c, \varepsilon, \sigma)$  از طریق سعی و خطا بهینه گردیدند.

### رگرسیون خطی چند متغیره

تحلیل رگرسیون روشی آماری است که در آن از رابطه بین دو یا چند متغیر کمی (متغیرهای مستقل و یا پیش‌بینی کننده) برای پیش‌بینی متغیر وابسته (متغیر پاسخ) استفاده می‌شود. یک مدل رگرسیون خطی چند متغیره به صورت زیر بیان می‌شود:

$$y = \sum_{i=1}^N \beta_i X_i + \varepsilon \quad (15)$$

که در آن  $\beta_i$  ضرایب رگرسیون (پارامترهای رگرسیون)،  $X_i$  متغیر مستقل،  $\varepsilon$  خروج از مرکز و  $N$  تعداد متغیرهای مستقل است. به منظور تخمین ضرایب رگرسیون  $(\beta_i)$  از روش حداقل مربعات استفاده می‌شود (بنی حبیب و همکاران، ۱۳۸۹).

### معیارهای ارزیابی

به منظور ارزیابی دقت و کارایی مدل‌ها، از نمایه‌های ضریب همبستگی (CC)، ریشه میانگین مربعات خطا (RMSE)، ضریب ناش ساتکلیف (NS) و بایاس (Bias) به صورت زیر محاسبه شد. بهترین مقادیر برای این چهار معیار به ترتیب یک، صفر، یک و صفر می‌باشند.

$$CC = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad -1 \leq R \leq 1 \quad (16)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (17)$$

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 آموزش مدل SVM توسط مجموعه‌ای از نمونه‌ها (مجموعه آموزش) محقق می‌شود. برای محاسبه  $w$  و  $b$  لازم است تابع خطا (معادله ۱۱) در مدل SVM رگرسیونی با در نظر گرفتن شرایط مندرج (قیود) در معادله (۱۲) بهینه شود

$$W^T \cdot \phi(X_i) + b - y_i \leq \varepsilon + \varepsilon_i^* \quad (11)$$

$$\frac{1}{2} W^T \cdot W + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^*$$

$$\frac{1}{2} W^T \cdot W + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^* \quad (12)$$

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

در معادلات بالا  $C$  عددی صحیح و مثبت است، که عامل تعیین جرمه در هنگام رخ دادن خطای آموزش مدل می‌باشد،  $\phi$  تابع کرنل،  $N$  تعداد نمونه‌ها و دو مشخصه  $\varepsilon_i, \varepsilon_i^*$  متغیرهای کمبود هستند. در نهایت تابع SVM رگرسیونی را می‌توان به فرم زیر باز نویسی کرد:

$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \phi(x_i)^T \cdot \phi(x) + b \quad (13)$$

در معادله (۱۳)،  $\bar{\alpha}_i$  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 رگرسیون انتخاب یک تابع کرنل است. می‌توان توابع مختلف کرنل برای ساخت انواع مختلف SVM رگرسیونی را به کار برد. در مطالعات هیدرولوژی عمدتاً از تابع کرنل پایه شعاعی<sup>۱</sup> (RBF) استفاده می‌گردد (اسکندری و نوری، ۱۳۸۹)، که رابطه آن به صورت زیر می‌باشد.

<sup>۱</sup> Radial Basis Function

مدل در نظر گرفته شد. با توجه به همبستگی متقابل و معنی‌دار بین داده‌های ورودی و خروجی، ترکیب‌های مختلفی از پارامترهای ورودی به منظور دستیابی به مدل بهینه جهت تخمین دبی ورودی به سد علویان استفاده شد که در جدول ۳ آمده است. بدین منظور از داده‌های ایستگاه هیدرومتری تازه کند که دارای ۴۷۴۴ رکورد ثبت شده طی بازه زمانی (۱۳۷۶-۱۳۸۹)، در مقیاس زمانی روزانه می‌باشد، استفاده شد. در نهایت تعداد ۳۷۹۴ رکورد برای آموزش و تعداد ۹۵۰ رکورد باقی‌مانده برای صحت‌سنجی روش‌های موردبررسی انتخاب شد.

$$NS=1-\frac{\sum_{i=1}^N (x_i-y_i)^2}{\sum_{i=1}^N (x_i-\bar{x})^2} \quad -\infty \leq NS \leq 1 \quad (18)$$

$$Bias=\frac{1}{N} \sum_{i=1}^N (x_i-y_i)^2 \quad (19)$$

در روابط بالا،  $x_i$  و  $y_i$  به ترتیب مقادیر مشاهداتی و محاسباتی در گام زمانی  $i$ ام،  $N$  تعداد گام‌های زمانی،  $\bar{x}$  و  $\bar{y}$  نیز به ترتیب میانگین مقادیر مشاهداتی و محاسباتی می‌باشد. یک نکته مهم در آموزش شبکه‌های عصبی نرمال‌سازی داده‌ها قبل از استفاده در مدل می‌باشد این عمل خصوصاً وقتی دامنه تغییرات ورودی‌ها زیاد باشد کمک شایانی به آموزش بهتر و سریع‌تر مدل می‌کند. اصولاً وارد کردن داده‌ها بصورت خام باعث کاهش سرعت و دقت شبکه می‌شود (Xu et al., 2007). برای نرمال‌سازی داده‌های تحقیق از رابطه زیر استفاده شده است:

$$X_n=0.1+0.8 \frac{X_i-X_{min}}{X_{max}-X_{min}} \quad (20)$$

## نتایج و بحث

یکی از مهمترین مراحل در مدل‌سازی، انتخاب ترکیب مناسبی از متغیرهای ورودی است. از این رو ابتدا همبستگی متقابل<sup>۱</sup> بین متغیرهای ورودی و خروجی محاسبه گردید و پارامترهای ورودی به منظور دستیابی به مدل بهینه جهت تخمین میزان دبی روزانه به سد علویان انتخاب و در جدول ۲ آرایه شد.

جدول (۲): همبستگی متقابل بین پارامترهای ورودی و

خروجی

	Q(t-1)	Q(t-2)	Q(t-3)	Q(t-4)
Q(t)	۰/۹۳۳	۰/۸۹۰	۰/۸۶۹	۰/۸۵۰

در این جدول  $Q(t-1)$ ،  $Q(t-2)$ ،  $Q(t-3)$ ،  $Q(t-4)$  (به ترتیب دبی در زمان  $t-4$ ،  $t-3$ ،  $t-2$  و  $t-1$  می‌باشند) به‌عنوان ورودی و  $Q(t)$  (دبی در زمان  $t$ ) به‌عنوان خروجی

<sup>1</sup> Cross-correlation



جدول (۳): ترکیب‌های منتخب پارامترهای ورودی مدل‌های شبکه عصبی بیزین، برنامه‌ریزی بیان ژن و ماشین بردار پشتیبان

شماره	ساختار ورودی	خروجی
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)

### نتایج شبکه عصبی بیزین

شبیه‌سازی انتخاب شد. همانطور که در جدول (۵) مشاهده می‌شود شبکه عصبی بیزین بخوبی تعمیم داده‌شده است که توانسته داده‌های بخش صحت سنجی را بخوبی داده‌های بخش آموزش اجرا نماید (MacKay., 1992). همان‌طور که در شکل (۴-ب) مشخص است تطابق مقادیر محاسباتی دبی مربوط به مرحله صحت‌سنجی مدل شبکه عصبی بیزین با مقادیر مشاهداتی وجود دارد. همچنین با توجه به نمودار (۴-الف) قابلیت بالای این مدل در تخمین اکثر مقادیر مشهود است. این نتایج با پژوهش‌های (Fithriasari et al., 2013) و (Khan and Coulibaly., 2006) مطابقت دارد. در تبیین این نتایج می‌توان گفت در شبکه‌های عصبی بیزین با اختصاص وزن مناسب به هر یک از پارامترهای ورودی طبق قانون بیز، باعث کاهش میزان خطای شبکه می‌شود. همچنین استفاده از تقریب گوس نیوتن جهت محاسبه پارامترهای تنظیم، سبب تسریع در روند همگرایی شبکه، بهینه نمودن وزن‌ها و کاهش خطای شبکه می‌شود.

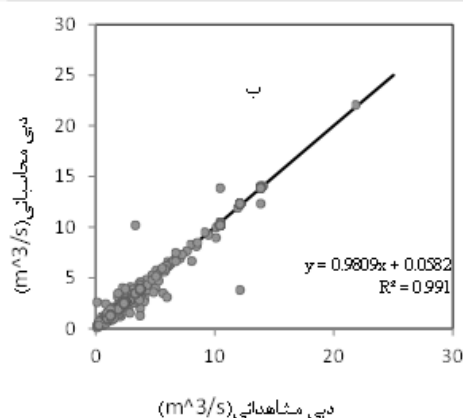
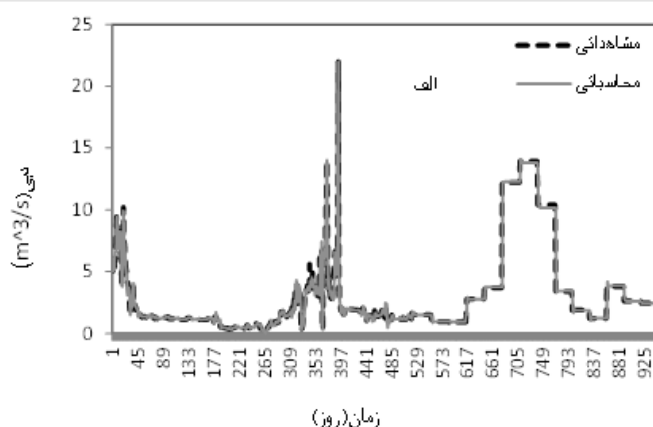
به‌منظور تخمین جریان ورودی به سد علویان از روش شبکه عصبی بیزین از نوع شبکه پرسپترون چندلایه با تعداد نرون‌های متفاوت استفاده شد. در این شبکه پارامترهای تابع هدف (پارامترهای تنظیم) با استفاده از تقریب گوس نیوتن محاسبه و نتایج آن در جدول (۴) ارائه شد با توجه به جدول (۴) خطای شبکه کاهش یافته و به پارامترهای ورودی به شبکه وزن مناسب اختصاص یافته است با بهینه نمودن پارامترهای تنظیم و تأثیر آن در تابع هدف از بیش برآزشی آموزش شبکه جلوگیری شد (MacKay., 1992). روند کار آموزش شبکه با تعداد نرون‌های کم آغاز و افزودن نرون‌های اضافی تا زمانی ادامه می‌یابد که افزایش نرون‌های بیشتر تأثیری در بهبود خطا نداشته باشد، در این روش با توجه به جدول (۵) مدل ترکیبی شماره ۴ بصورت شبکه‌ای با ۱۶ نرون در لایه پنهان اول با بیشترین ضریب همبستگی  $CC=0/991$  و کمترین ریشه میانگین مربعات خطا  $RMSE=0/031$  m3/s و  $NS=0/981$  و بایاس  $-0/006$  در مرحله صحت‌سنجی به‌عنوان بهترین مدل در فرآیند

جدول (۴): مقادیر پارامترهای مورد استفاده در شبکه عصبی بیزین

پارامتر	$E_D$	$E_W$	$\beta$	$\alpha$	Gradient	Epoch
مجموع مربعات خطای شبکه	مجموع مربعات وزن‌ها	مجموع مربعات وزن‌ها	پارامتر تنظیم	پارامتر تنظیم	شیب	تعداد دور
۰/۰۰۰۲	۱۶/۴۶۱	۱۶/۴۶۱	$1/79 \times 10^{-5}$	۰/۹۹۹۹	$10^{-1}$	۱۰۰۰

جدول (۵): نتایج شبکه عصبی بیزین برای مدل‌های مختلف در مراحل آموزش و صحت‌سنجی (تست)

		صحت سنجی (تست)				آموزش				
		RMSE				RMSE				
مدل	ساختار	CC	NS	$(m^3/s)$	Bias	CC	NS	$(m^3/s)$	Bias	
۱	۱-۸-۱	۰/۹۴۷	۰/۰۸۸	۰/۸۹۷	۰/۰۰۶	۰/۹۷۲	۰/۰۵۳	۰/۹۴۶	۰/۰۰۸	
۲	۲-۱۲-۱	۰/۹۵۷	۰/۰۸۰	۰/۹۱۶	-۰/۰۰۲	۰/۹۸۴	۰/۰۴۰	۰/۹۶۹	۰/۰۰۴	
۳	۳-۱۵-۱	۰/۹۶۰	۰/۰۷۶	۰/۹۲۳	۰/۰۰۶	۰/۹۸۹	۰/۰۳۴	۰/۹۷۸	۰/۰۰۵	
۴	۴-۱۶-۱	۰/۹۶۶	۰/۰۷۰	۰/۹۳۴	۰/۰۰۸	۰/۹۹۱	۰/۰۳۱	۰/۹۸۱	۰/۰۰۶	



شکل (۴): نمودار حاصل از مقادیر بهینه مدل شبکه عصبی بیزین برای داده‌های مرحله صحت‌سنجی (الف) مقادیر محاسباتی و مشاهداتی نسبت به زمان ب) نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

پژوهش جهت ساخت درخت تجزیه و تولید یک جمعیت اولیه از راه‌حلهایی استفاده شد که این راه‌حل‌ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند نحوه و دلیل انتخاب این نوع عملگرها با توجه به مطالعات (Ghorbani et al., 2012) و (Khatibi et al., 2012) صورت گرفت.

$$F_1: \{+, -, *, /\}$$

(۲۱)

### نتایج برنامه‌ریزی بیان ژن

استفاده از برنامه‌ریزی بیان ژن به دلیل توان انتخاب متغیرهای مؤثر در مدل و حذف متغیرهایی با تأثیر کمتر و همچنین توانایی ارایه رابطه صریح جهت تخمین جریان ورودی به سد علویان مد نظر قرار گرفت. در این تحقیق از ساختارهای ورودی برای تعیین متغیرهای معنی‌دار استفاده و جهت بررسی بیشتر علاوه بر مجموعه چهار عملگر اصلی ( $F_1$ )، حالت‌هایی بر اساس عملگرهای ریاضی پیش فرض برنامه ( $F_2$ ) لحاظ شد. همچنین در این

برنامه‌ریزی بیان ژن به طور خلاصه در جدول (۶) ارائه شد.

$$F_2: \{+, -, *, /, \sqrt{\quad}, \text{Exp}, \text{Ln}, ^2, ^3, \sqrt[3]{\quad}, \text{Sin}, \text{Cos}, \text{Atan}\} \quad (22)$$

پارامترهای مورد استفاده و نرخ آن‌ها در استخراج مدل‌هایی جهت تخمین جریان ورودی با استفاده از روش

جدول ۶- مقادیر پارامترهای مورد استفاده با استفاده از روش برنامه‌ریزی بیان ژن

تنظیمات کلی	عملگرهای ژنتیکی		
تعداد کروموزوم‌ها	نرخ جهش	۳۰	۰/۰۴۴
اندازه راس	نرخ وارون‌سازی	۸	۰/۱
تعداد ژن‌ها در هر کروموزوم	نرخ ترانهش درج متوالی	۳	۰/۱
تعداد جمعیت تولیدی	نرخ ترانهش ریشه درج متوالی	۱۰۰۰	۰/۱
تابع پیوند	نرخ ترکیب تک نقطه ایی	جمع (+)	۰/۳
	نرخ ترکیب دو نقطه ایی		۰/۳

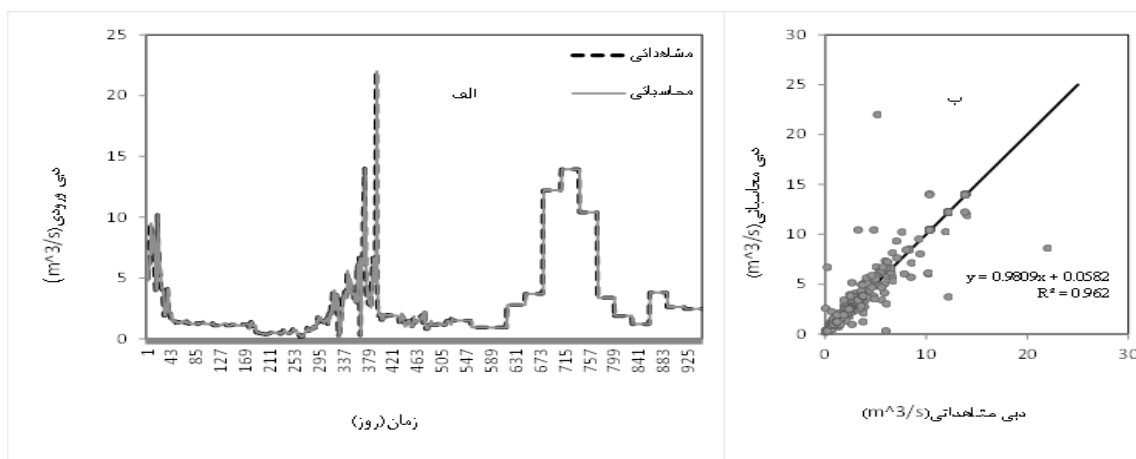
اساس خط  $(y=x)$  است. همچنین لازم به ذکر است با توجه به شکل (۵-الف) این مدل در تخمین مقادیر جریان ورودی خوب عمل کرده، به طوری که این مقادیر محاسباتی را برابر مقدار واقعی آن‌ها (مشاهداتی) تخمین زد. که این نتایج با نتایج (Aytek et al., 2008)، (سلطانی و همکاران، ۱۳۸۹) و (فربودنام و همکاران، ۱۳۸۸) مطابقت دارد. در تبیین این نتایج می‌توان بیان نمود معادله حاصل از برنامه‌ریزی بیان ژن، از ترکیب تصادفی مجموع ترمینال‌ها و توابع بدست می‌آید بنابراین اگر رابطه بین ورودی‌ها و خروجی‌ها خطی باشد ولی عملگرهای  $\sin$ ،  $\cos$  و ... در مجموعه توابع انتخاب شود برنامه‌ریزی بیان ژن در استخراج رابطه از آن عملگرهای انتخابی استفاده می‌کند که این امر باعث کاهش دقت مدل می‌شود که در این پژوهش جهت افزایش دقت مدل عملگرهای  $\sin$ ،  $\cos$  و ... کاربرد ندارد و نیز با توجه به دقت و سادگی، مدل حاصل از چهار عمل اصلی ریاضی

نتایج حاصل از مدل برنامه‌ریزی بیان ژن با هر دو عملگر در جدول (۷) بیانگر این است که عملگر  $F1$  در هر دو مرحله آموزش و صحت‌سنجی با بیشترین ضریب همبستگی  $CC=0/962$ ،  $RMSE=1/760$ ،  $NS=0/923$  و بایاس  $-0/005$  نسبت به سایر عملگرها از دقت بیشتری برخوردار است. بنابراین برنامه‌ریزی بیان ژن با عملگر  $F1$  شامل چهار عملگر اصلی ریاضی، با وجود رابطه ساده ریاضی بیشترین دقت را در تخمین جریان ورودی به سد علویان دارد. نمودارهای پراکنش برنامه‌ریزی بیان ژن مربوط به مرحله صحت‌سنجی در شکل (۵-ب) نشان‌دهنده خط برازش مقادیر محاسباتی با بهترین خط برازش  $y=x$  است. همان‌طور که از این شکل مشخص است مقادیر تخمینی و مشاهداتی به‌غیر از چند نقطه که روی خط نیمساز قرار نگرفته در سایر نقاط بر روی خط برازش قرار دارد که این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی بر

{+، -، \*} جهت تخمین جریان ورودی به سد علویان پیشنهاد شد.

جدول (۷): نتایج مدل برنامه‌ریزی بیان ژن با استفاده از دو مجموعه عملگر ریاضی منتخب

مدل	عملگر	آموزش				صحت سنجی (تست)			
		Bias	NS	RMSE	CC	Bias	NS	RMSE	CC
۱	F1	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		-۰/۲۰۴	۰/۹۱۰	۱/۹۱۴	۰/۹۶۱	-۰/۱۸۳	۰/۸۵۰	۶/۳۵۴	۰/۹۳۱
۲	F2	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		-۰/۰۱۴	۰/۹۲۳	۱/۷۶۰	۰/۹۶۱	-۰/۰۰۶	۰/۸۶۵	۶/۰۳۴	۰/۹۳۰
۳	F1	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		۰/۰۰۳	۰/۹۲۱	۱/۷۸۷	۰/۹۶۰	۰/۰۱۷	۰/۸۶۴	۶/۰۵۶	۰/۹۳۱
۴	F2	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		۰/۰۳۹	۰/۹۲۲	۱/۷۷۰	۰/۹۶۰	۰/۰۶۵	۰/۸۶۳	۶/۸۱۰	۰/۹۳۰
۵	F1	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		-۰/۰۰۱	۰/۹۲۲	۱/۷۷۳	۰/۹۶۰	۰/۰۰۹	۰/۸۷۰	۵/۹۱۴	۰/۹۳۳
۶	F2	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		۰/۰۶۴	۰/۹۲۲	۱/۷۷۵	۰/۹۶۰	۰/۰۷۹	۰/۸۷۱	۵/۹۰۳	۰/۹۳۳
۷	F1	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		-۰/۰۰۵	۰/۹۲۳	۱/۷۶۰	۰/۹۶۲	۰/۰۰۱	۰/۸۵۹	۶/۱۶۳	۰/۹۲۹
۸	F2	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
		-۰/۰۶۶	۰/۹۲۱	۱/۷۹۰	۰/۹۶۰	۰/۰۳۳	۰/۸۷۵	۵/۸۱۲	۰/۹۳۵



شکل (۵): نمودار حاصل از مقادیر بهینه مدل برنامه‌ریزی بیان ژن برای داده‌های مرحله صحت‌سنجی

الف (مقادیر محاسباتی و مشاهداتی نسبت به زمان ب) نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

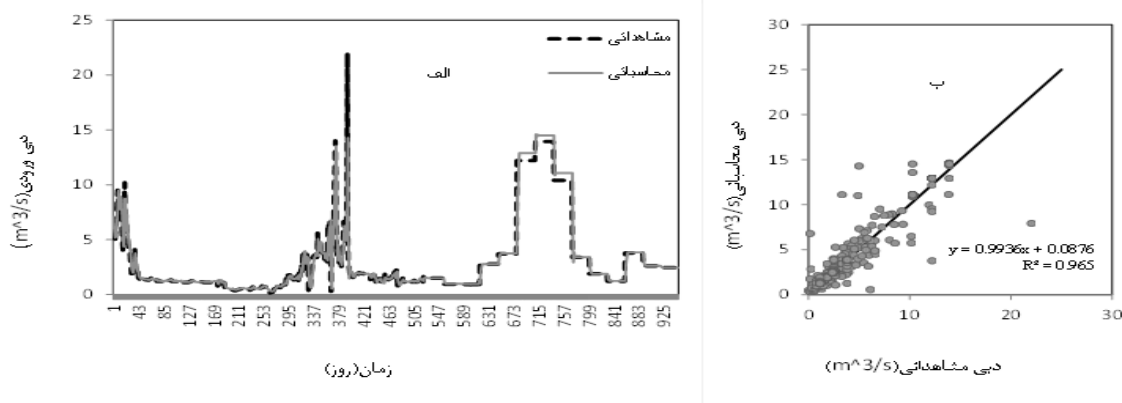
مشاهداتی به‌غیراز چند نقطه روی خط نیمساز قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی بر اساس خط  $(y=x)$  است. در شکل (۶-الف) نیز تغییرات مقادیر محاسباتی و مشاهداتی نسبت به زمان نشان داده شد، به‌طوری‌که این مدل در تخمین مقادیر بیشینه در شرایط مختلف زمانی دقت کافی نداشت. که نتایج جدول (۸) با نتایج (اسکندری و نوری، ۱۳۸۹) همخوانی دارد که می‌توان بیان نمود با توجه به اینکه ماشین بردار پشتیبان بر مبنای استفاده از اصل استقرای کمینه‌سازی خطای ساختاری است بنابراین در شبیه‌سازی، با بکارگیری روش یادگیری با نظارت در توابع پایه شعاعی باعث می‌شود تخمین پارامتر از سرعت بالا و خطای کمتری نسبت به سایر کرنل‌ها داشته باشد و این یکی از ویژگی‌های ممتاز توابع پایه شعاعی محسوب می‌شود.

### نتایج ماشین بردار پشتیبان

به‌منظور تخمین جریان ورودی به سد علویان توسط مدل SVM توابع مختلف کرنل شامل کرنل خطی، چندجمله‌ای و توابع پایه شعاعی که از انواع رایج مورد استفاده در هیدرولوژی است را می‌توان مورد بررسی قرار داد و بهینه‌ترین را انتخاب نمود. نتایج حاصل بررسی کرنل‌ها در جدول (۸) آورده شد. با توجه به جدول مذکور مدل ترکیبی شماره ۴ با کرنل توابع پایه شعاعی دارای بیشترین میزان ضریب همبستگی  $CC=0/965$  و کمترین میزان ریشه مربعات خطا  $RMSE = 1/680 \text{ m}^3/\text{s}$  و  $NS = 0/928$  و بایاس  $-0/067$  در مرحله صحت‌سنجی جواب بهینه‌تری نسبت به سایر مدل‌ها به همراه داشت. در شکل (۶) نمودار بهترین مدل حاصل شده برای داده‌های بخش صحت‌سنجی نشان داده شد. همان‌طور که از شکل (۶-ب) مشخص است بیشتر مقادیر تخمینی و

جدول (۸): نتایج سه کرنل استفاده‌شده در روش ماشین بردار پشتیبان برای داده‌های بخش آموزش و صحت‌سنجی (تست)

	صحت‌سنجی (تست)				آموزش				کرنل
	Bias	NS	RMSE ( $\text{m}^3/\text{s}$ )	CC	Bias	NS	RMSE ( $\text{m}^3/\text{s}$ )	CC	
۱	-۰/۰۵۸	۰/۹۲۶	۱/۷۳۱	۰/۹۶۳	۰/۰۰۰۶	۰/۸۸۶	۵/۵۴۲	۰/۹۴۱	توابع پایه شعاعی
	-۰/۰۶۲	۰/۹۲۵	۱/۷۳۷	۰/۹۶۲	-۰/۰۰۰۳	۰/۸۸۶	۵/۵۴۷	۰/۹۴۱	چندجمله‌ای
	-۰/۰۱۳	۰/۹۲۱	۱/۷۸۴	۰/۹۶۰	۰/۰۰۰۵	۰/۸۶۴	۶/۰۵۴	۰/۹۲۹	خطی
۲	-۰/۰۵۶	۰/۹۲۲	۱/۷۷۰	۰/۹۶۲	-۰/۰۰۰۱	۰/۸۸۸	۵/۴۸۲	۰/۹۴۲	توابع پایه شعاعی
	-۰/۰۷۷	۰/۹۲۶	۱/۷۲۸	۰/۹۶۳	۰/۰۰۰۴	۰/۸۸۸	۵/۴۹۵	۰/۹۴۲	چندجمله‌ای
	-۰/۰۱۳	۰/۹۲۲	۱/۷۶۹	۰/۹۶۱	۰/۰۰۰۱	۰/۸۶۷	۵/۹۹۱	۰/۹۳۱	خطی
۳	-۰/۰۷۴	۰/۹۲۷	۱/۷۱۸	۰/۹۶۵	۰/۰۰۰۹	۰/۹۰۲	۵/۱۴۳	۰/۹۴۹	توابع پایه شعاعی
	-۰/۰۷۵	۰/۹۳۰	۱/۶۷۷	۰/۹۶۷	۰/۰۰۰۳	۰/۸۹۸	۵/۲۲۹	۰/۹۴۷	چندجمله‌ای
	-۰/۰۱۳	۰/۹۲۳	۱/۷۶۳	۰/۹۶۱	-۰/۰۰۰۴	۰/۸۷۱	۵/۸۸۷	۰/۹۳۳	خطی
۴	-۰/۰۶۷	۰/۹۲۸	۱/۶۸۰	۰/۹۶۵	۰/۰۰۰۲	۰/۹۰۱	۵/۱۴۸	۰/۹۴۹	توابع پایه شعاعی
	-۰/۰۸۲	۰/۹۳۰	۱/۷۰۲	۰/۹۶۵	۰/۰۰۰۱	۰/۸۹۳	۵/۳۷۰	۰/۹۴۴	چندجمله‌ای
	-۰/۰۱۳	۰/۹۲۴	۱/۷۵۰	۰/۹۶۱	-۰/۰۰۰۵	۰/۸۷۲	۵/۸۶۷	۰/۹۳۳	خطی



شکل (۶): نمودار حاصل از مقادیر بهینه مدل ماشین بردار پشتیبان برای داده‌های مرحله صحت‌سنجی

الف (مقادیر محاسباتی و مشاهداتی نسبت به زمان ب) نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

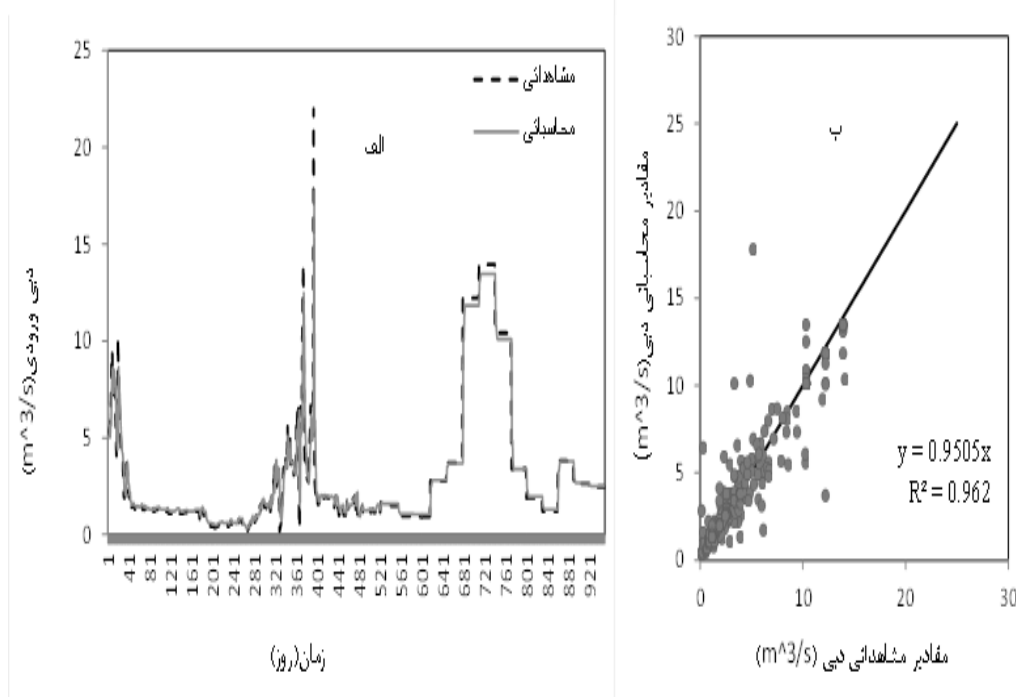
است. در شکل ۷ نمودار بهترین مدل حاصل شده برای داده‌های بخش صحت‌سنجی نشان داده شد. همان‌طور که از شکل (۷-ب) مشخص است بیشتر مقادیر تخمینی و مشاهداتی به‌غیر از چند نقطه روی خط نیمساز قرار گرفته و این امر دلالت بر برابری مقادیر مشاهداتی و محاسباتی بر اساس خط  $(y=x)$  است. در شکل (۷-الف) نیز تغییرات مقادیر محاسباتی و مشاهداتی نسبت به زمان نشان داده شد، به‌طوری‌که این روش عملکرد کافی را در تخمین مقادیر بیشینه و کمینه نداشته است بگونه‌ای که این مقادیر را نزدیک به مقدار واقعیشان پیش بینی نکرده است.

#### نتایج رگرسیون خطی چند متغیره

در پژوهش حاضر جهت تخمین جریان از روش رگرسیون خطی چند متغیره استفاده گردید. در این روش از پارامترهای دبی جریان در تأخیرهای زمانی در مقیاس روزانه استفاده شد. نتایج حاصل از عملکرد روش رگرسیون خطی در جدول ۹ قابل‌مشاهده است همان‌طور که در جدول ۹ مشخص است مدل شماره ۴ با بیشترین دقت  $CC=0/961$  و کمترین ریشه میانگین مربعات خطا  $RMSE = 1/747 \text{ m}^3/\text{s}$  و بیشترین ضریب ناش-ساتکلیف  $NS=0/925$  و کمترین مقدار بایاس  $0/004$  در مرحله صحت‌سنجی جواب بهتری نسبت به سایر مدل‌ها از خود نشان داده

جدول (۹): نتایج مدل رگرسیون خطی برای داده‌های بخش آموزش و صحت‌سنجی (تست)

مدل	آموزش				صحت‌سنجی (تست)			
	Bias	NS	RMSE	CC	Bias	NS	RMSE	CC
۱	۰/۸۶۴	۰/۸۶۴	۶/۰۴۵	۰/۹۲۹	-۰/۰۰۵	۰/۹۲۱	۱/۷۸۲	۰/۹۶۰
۲	۰/۸۶۷	۰/۸۶۷	۵/۹۹۰	۰/۹۳۱	-۰/۰۰۲	۰/۹۲۳	۱/۷۶۷	۰/۹۶۱
۳	۰/۸۷۲	۰/۸۷۲	۵/۸۷۸	۰/۹۳۳	-۰/۰۰۱	۰/۹۲۴	۱/۷۶۵	۰/۹۶۱
۴	۰/۸۷۱	۰/۸۷۱	۵/۸۶۵	۰/۹۳۴	-۰/۰۰۴	۰/۹۲۵	۱/۷۴۷	۰/۹۶۱



شکل (۷): نمودار حاصل از مقادیر بهینه مدل رگرسیون خطی برای داده‌های مرحله صحت‌سنجی (الف) مقادیر محاسباتی و مشاهداتی نسبت به زمان (ب) نمودار پراکنش میان مقادیر مشاهداتی و محاسباتی

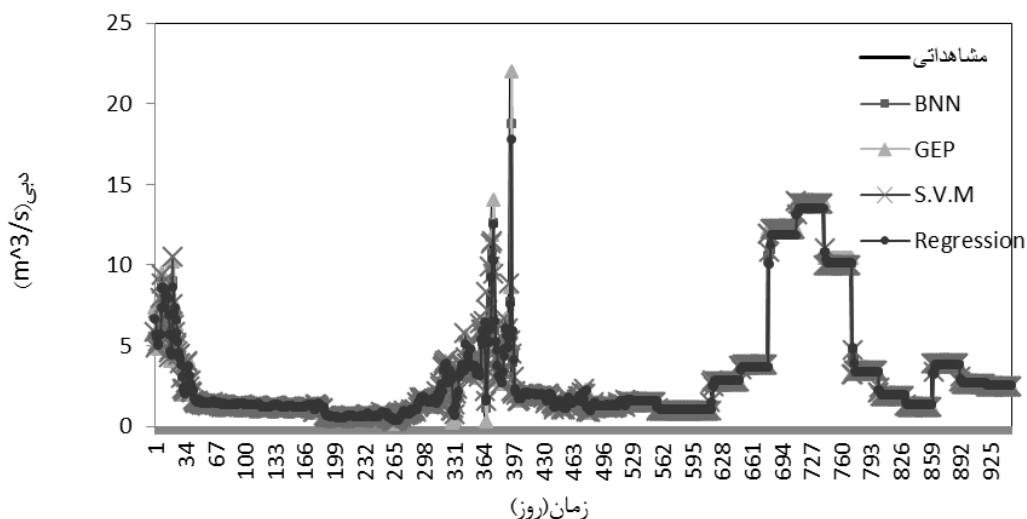
بیشتر مقادیر عملکرد قابل قبولی دارند، در حالی که مدل شبکه عصبی بیزین مقادیر کمینه، بیشینه و میانه دبی جریان را به خوبی پوشش داده است. همچنین در جدول (۱۱) نتایج حاصل از مقادیر پیک‌ها و خطاها در هر چهار مدل نشان داده شد. همانطور که در این جدول مشاهده می‌شود از میان روش‌های موردبررسی روش شبکه عصبی بیزین با اختصاص وزن بهینه به هریک از پارامترهای ورودی که موجب کاهش اختلاف خروجی با پاسخ شبکه می‌شود در تخمین مقادیر پیک عملکرد قابل قبولی دارد به‌گونه‌ای که در تخمین مقادیر پیک دارای خطای ناچیزی است.

#### مقایسه عملکرد مدل‌ها

همانطور که در جدول (۱۰) مشاهده می‌شود با انتخاب جواب بهینه هر کدام از مدل‌ها و مقایسه آن‌ها با یکدیگر مشخص شد که هر چهار مدل با دقت خوبی می‌توانند دبی جریان را شبیه‌سازی کنند. که از بین مدل‌های به‌کاررفته مدل شبکه عصبی بیزین، بیشترین دقت  $CC=0/991$  و کمترین ریشه میانگین مربعات خطا  $RMSE = 0/031 \text{ m}^3/\text{s}$  و بیشترین ضریب ناش-ساتکلیف  $NS=0/981$  و کمترین مقدار بایاس  $-0/006$  در مرحله صحت‌سنجی را دارد. در شکل (۸) نتایج هر چهار مدل نسبت به مقادیر مشاهداتی در طی زمان نشان داده شد همانطور که در شکل (۸) قابل مشاهده است هر چهار مدل در تخمین

جدول (۱۰): نتایج نهایی حاصل از آموزش و صحت‌سنجی مدل‌های شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی

مدل	آموزش				صت سنجی (تست)			
	Bias	NS	RMSE	CC	Bias	NS	RMSE	CC
	(m <sup>3</sup> /s)				(m <sup>3</sup> /s)			
شبکه عصبی بیزین	-۰/۰۰۸	۰/۹۳۴	۰/۰۷۰	۰/۹۶۶	-۰/۰۰۶	۰/۹۸۱	۰/۰۳۱	۰/۹۹۱
برنامه‌ریزی بیان ژن	-۰/۰۰۱	۰/۸۵۹	۶/۱۶۳	۰/۹۲۹	-۰/۰۰۵	۰/۹۲۳	۱/۷۶۰	۰/۹۶۲
ماشین بردار پشتیبان	-۰/۰۰۲	۰/۹۰۱	۵/۱۴۸	۰/۹۴۹	-۰/۰۰۶۸	۰/۹۲۸	۱/۶۸۰	۰/۹۶۵
رگرسیون خطی	-۰/۰۰۳	۰/۸۷۱	۵/۸۶۵	۰/۹۳۴	-۰/۰۰۴	۰/۹۲۵	۱/۷۴۷	۰/۹۶۱

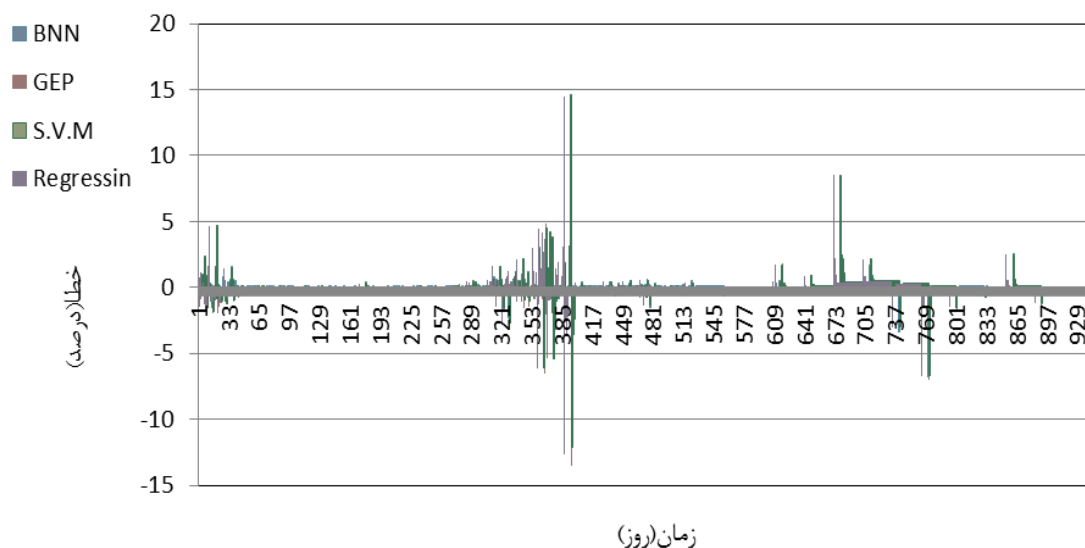


شکل (۸): نمودار پراکنش و مشاهداتی - محاسباتی مقادیر بهینه مدل شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی برای داده‌های ثبت‌شده مرحله صحت‌سنجی



جدول (۱۱): مقادیر دبی پیک و خطا مدل شبکه عصبی بیزین، برنامه ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی

خطا (درصد)				مقدار دبی پیک				مشاهداتی	پیکها
Reg	S.V.M	GEP	BNN	Reg	S.V.M	GEP	BNN		
۲/۱۸۳	۱/۷۱۴	۱/۵۰۰	۰/۲۰۹	۷/۳۱۷	۷/۷۸۵	۸	۹/۲۹۰	۹/۵۰۰	۱
۴/۶۵۳	۴/۴۷۸	۴/۲۰۰	۰/۱۴۶	۵/۵۴۷	۵/۷۲۱	۶	۱۰/۰۵۳	۱۰/۲۰۰	۲
۳/۶۹۱	۲/۹۲۹	۲/۱۵۰	۰/۰۹۱	۱۰/۳۱۹	۱۱/۰۸۰	۱۱/۸۶۰	۱۳/۹۱۸	۱۴/۰۱۰	۳
۱۴/۴۴۷	۱۴/۱۰۲	۱۳/۴۱۰	۰/۱۸۹	۷/۵۳۳	۷/۸۷۷	۸/۵۷۰	۲۱/۷۹۰	۲۱/۹۸۰	۴
۰/۴۰۳	-۰/۶۹۵	۰	-۰/۰۱۲	۱۱/۸۰۷	۱۲/۹۰۵	۱۲/۲۱۰	۱۲/۲۲۲	۱۲/۲۱۰	۵
۰/۴۸۰	-۰/۵۶۲	۰	۰/۱۸۷	۱۳/۴۶۰	۱۴/۵۰۲	۱۳/۹۴۰	۱۳/۷۵۲	۱۳/۹۴۰	۶
۰/۳۲۱	-۰/۶۸۳	۰	۰/۳۰۴	۱۰/۰۷۹	۱۱/۰۸۳	۱۰/۴۰۰	۱۰/۰۹۵	۱۰/۴۰۰	۷



شکل (۹): نمودار خطای بهینه هر چهار مدل به صورت درصدی از میانگین مقادیر مشاهداتی

آماري ترسيم شد (شکل ۹). همان طور که در این شکل مشخص است برای هر چهار مدل بیشتر خطاها در باند  $\pm 5$  درصد قرار گرفته است و بیشترین میزان خطای مدل شبکه عصبی بیزین، برنامه ریزی بیان ژن،

در نهایت اختلاف مقادیر مشاهداتی جریان ورودی و محاسباتی بهینه مدلها به صورت درصدی از میانگین مقادیر مشاهداتی (مقدار خطا)، محاسبه و نمودار آن نسبت به نمونه داده های ثبت شده دوره

بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی)، با استفاده از معیارهای ارزیابی مورد مقایسه قرار گرفت. نتایج تحقیق را می‌توان بصورت زیر خلاصه نمود: نتایج نشان داد که در هر چهار روش شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی، ساختاری که در آن ورودی مدل از ترکیب پارامترهای جریان در ۴ تأخیر زمانی تشکیل شده، از دقت قابل قبولی برخوردار است. همچنین با توجه به معیارهای ارزیابی نتیجه شد که هر چهار روش مورد بررسی، می‌توانند با دقت نسبتاً بالایی میزان جریان روزانه را پیش‌بینی نمایند. در این میان، مدل شبکه عصبی بیزین با توجه به خطای کمتر نسبت به سایر مدل‌ها و نیز استفاده از قانون بیز جهت بهینه‌سازی پارامترهای تنظیم و اضافه شدن مجموع مربعات وزن‌ها در تابع هدف عملکرد بهتری نسبت به مدل‌های برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی از خود نشان داده است.

ماشین بردار پشتیبان و رگرسیون خطی به ترتیب ۰/۱۸۹، ۱۳/۴۱۰، ۱۴/۱۰۲ و ۱۴/۴۴۷ درصد از میانگین مقادیر مشاهداتی می‌باشد که در میان مدل‌های مذکور (شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی)، مدل شبکه عصبی بیزین کم‌ترین میزان خطا را به علت اضافه شدن پارامتر مجموع مربعات وزن‌ها به تابع هدف داشته است (MacKay, 1992).

### نتیجه‌گیری

در این تحقیق سعی بر آن شد که عملکرد مدل‌هایی جهت شبیه‌سازی جریان ورودی به ایستگاه بالادست مخزن سد علویان با استفاده از داده‌های روزانه مورد ارزیابی قرار گیرد. مدل‌های به کار گرفته شده شامل شبکه عصبی بیزین، برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان و رگرسیون خطی است. مقادیر جریان ورودی مشاهداتی با جریان ورودی تخمین زده شده در مدل‌های مذکور (شبکه عصبی

### منابع

- اسکندری، ع. و ر. نوری. ۱۳۸۹. توسعه مدلی مناسب بر مبنای شبکه عصبی مصنوعی و ماشین بردار پشتیبان برای پیش‌بینی به هنگام اکسیژن خواهی بیوشیمیایی ۵ روزه. محیط شناسی سال سی و هشتم، شماره ۶۱. ص ۷۶-۷۴.
- بنی‌حبیب، م.، ا. موسوی، م. سادات جمالی. ۱۳۸۹. مدل شبکه عصبی مصنوعی برای بررسی همبستگی روزانه بین ایستگاه‌ها در پیش‌بینی جریان ورودی به مخزن سد دز. مجله پژوهش آب ایران. سال چهارم. شماره ۷. ص ۳۲-۲۵.
- سلطانی، ع.، م. ع. قربانی، ا. فاخری فرد، ص. دربندی و د. فرسادی زاده. ۱۳۸۹. برنامه‌ریزی ژنتیک و کاربرد آن در مدل‌سازی فرآیند بارش- رواناب. مجله دانش آب و خاک. شماره ۴. ص ۶۲-۷۱.
- طباطبایی، س. ع. و پ. دشتی‌زاده. ۱۳۸۷. معرفی شبکه‌های عصبی Bayesian و کاربرد آن‌ها در نگهداری پل‌ها. چهاردهمین کنفرانس دانشجویان مهندسی عمران سراسر کشور.
- فربودنام، ن.، م. ع. قربانی و م. ت. اعلمی. ۱۳۸۸. پیش‌بینی جریان رودخانه با استفاده از برنامه‌ریزی ژنتیک (مطالعه موردی: حوضه آبریز رودخانه ليقوان). نشریه دانش آب و خاک. شماره ۱۹. ص ۱۲۳-۱۰۷.
- قربانی، م. ع. و ا. صالحی. ۱۳۹۰. استفاده از برنامه‌ریزی بیان ژن در بررسی تغییرات داده‌های کیفی آب زیرزمینی با نوسانات سطح آب در دشت برخوردار اصفهان. ششمین کنگره ملی مهندسی عمران.

- محرم‌پور، م.، ع. محرابی و م. کاتوزی. ۱۳۹۰. به کارگیری ماشین بردار پشتیبان SVM برای پیش‌بینی دبی روزانه. چهارمین کنفرانس مدیریت منابع آب ایران.
- Aytek A., M. Asce and M. Alp. 2008. An application of artificial intelligence for rainfall-runoff modeling. *J Earth System Science* 117: 145-155.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*, Vol.13(2): 87-129.
- Fithriasari, K., N. Brodjol, U. Sutikno, H. Kuswanto. 2013. Prediction of Hourly Rainfall using Bayesian Neural Network with Adjusting Procedure. *The Third Basic Science International Conference*.
- Foresee, F. D. 1997. Gauss-newton approximation to Bayesian regularization. *International Joint Conference on Neural Network*, Houston:1930-1935.
- Ghorbani, M. A., R. Khatibi, H. Asadi and P. Yousefi. 2012. Inter- Comparison of an Evolutionary Programming Model of Suspended Sediment Time-series with other Local Model. *INTECH*. doi. org/10.5772/47801.; 255-282.
- Guyen, A. 2009. Linear genetic programming for time-series modelling of daily flow rate. *Journal Earth System Science*, vol 118: 157-173.
- He. Z., X. Wen. H. Liu. and J. Du. 2014. "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region". *Journal of Hydrology* 509 : 379-386
- Khan, M. S. and P. Coulibaly .2006. Bayesian neural network for rainfall-runoff modeling, *Water Resour. Res.*, 42,W07409, doi:10.1029/2005WR003971.
- Khatibi, R., L. Naghipour, M. A. Ghorbani and M. T. Aalami. 2012. Predictability of relative humidity by two artificial intelligence techniques using noisy data from two Californian gauging stations. *Neural computing and application*: 643-941.
- MacKay, D. J. C.1992. Bayesian Interpolation, *Neural Computation*, vol. 4, pp. 415-447.
- Othman, F. and M.Naseri. 2011. Reservoir inflow forecasting using artificial neural Network. *International Journal of the Physical Sciences* Vol. 6(3): 434-440.
- Xu, L., J. Wang, J. Guan, and F. Huang. 2007. A Support Vector Machine Model for Mapping of Lake Water Quality from Remote-Sensed Images. *IC-MED*. Vol. 1(1): 57-66.

## Comparison and Evaluation of Bayesian Neural Network, Gene expression programming, Support Vector Machine and Multiple Linear Regression in River Discharge Estimation (Case Study: Sufi Chay Basin)

Reza Dehghani<sup>۱</sup>, Mohammad Ali Ghorbani<sup>۲</sup>, Mohammad teshnehlab<sup>۳</sup>, Amir Rikhtehgar Gheasi<sup>۴</sup>,  
Esmail Asadi<sup>۵</sup>

### Abstract

Prediction of river discharge is importance for reliable planning, design and management of water resources projects. This study investigates the applicability of Bayesian Neural Network (BNN), Gene-Expression Programming (GEP), Support Vector Machine (SVM), and Multi linear Regression (MLR) for prediction of river discharge time series in the Soufichay river, Iran. Daily river discharge time series for period of 1997 to 2010 of Tazehkand hydrometric station from Soufichay river was used. To obtain the best input–output mapping, different input combinations of antecedent daily river discharge were evaluated. The performance of the models were evaluated through the four performance criteria: Correlation Coefficient (CC), Root Mean Square Error (RMSE), the Nash–Sutcliffe efficiency coefficient (N-S) and Bias criteria. A comparison of models indicates that BNN with CC (0.991), RMSE (0.031m<sup>3</sup>/s), N-S (0.981) and Bias(-0.006) predicted better than GEP, SVM, and MLR models for daily river discharge time series.

**Key words :** Bayesian neural networks, Discharge, Genetic programming, Support vector machine.

<sup>1</sup>. MSC, Agricultural Engineering -Water Resources, University of Tabriz, Iran  
(Corresponding author.; Email: reza.dehghani67@yahoo.com)

<sup>2</sup> .Associate Professor of Water Engineering ,University of Tabriz, Tabriz, Iran, Ghorbani@tabrizu.ac.ir

<sup>3</sup> .Professor of Electrical Engineering - Control ,Electrical Engineering Department ,K.N Toosi University, Tehran, Iran, teshnehlab@eetd.kntu.ac.ir

<sup>4</sup> .Assistant Professor of Electrical Engineering - Control, University of Tabriz, Tabriz, Iran, agiasi@tabrizu.ac.ir

<sup>5</sup> .Assistant Professor of Water Engineering ,University of Tabriz, Tabriz, Iran, esasadi@gmail.com