نشریه آبیاری و زهکشی ایران
شماره 2، جلد 10، خرداد - تیر 1395، ص. 207-199

Iranian Journal of Irrigation and Drainage
No. 2, Vol. 10, May.-Jun. 2016, p. 199-207

بررسی عملکرد مدل‌های SVR و GEP در پیش‌بینی نوسانات ماهانه تراز آب دریاچه ارومیه

مصطفی کدخداحسینی[1*]، شایان شامحمدی[2]، حامد نوذری[3]، رسول میرعباسی نجف آبادی[4]

تاریخ دریافت: 1394/12/22 تاریخ پذیرش: 1395/3/19

چکیده

پیش‌بینی نوسانات سطح آب دریاچه‌ها از جمله موارد مهم در مدیریت و برنامه‌ریزی منابع آب است. در سال‌های اخیر کاهش شدید تراز آب دریاچه ارومیه اثرات زیست‌محیطی مخربی را در این منطقه داشته است. در این تحقیق عملکرد مدل‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن با استفاده از شش الگوی متفاوت در خلال سال‌های 1388- 1355 برای پیش‌بینی تراز آب دریاچه ارومیه بررسی شد، تا بهترین الگو برای پیش‌بینی تراز آب دریاچه معرفی گردد. در چهار الگو از داده‌های تاریخی تراز و در دوالگوی دیگر از بارش، تبخیر و نشت و تراز آب استفاده شد. نتایج نشان داد که مدل برنامه‌ریزی بیان ژن عملکرد بهتری نسبت به مدل رگرسیون بردار پشتیبان دارد و با افزایش ورودی برای آموزش مدل عملکرد آن افزایش می‌یابد. هم-چنین در بهترین حالت مقدار میانگین مربعات خطا و ضریب تبیین به ترتیب برابر 0/08 متر و 0/99 برای مدل GEP، 0/60 متر و 0/92 برای مدل SVRبه دست آمد.

واژه‌های کلیدی: برنامه‌ریزی بیان ژن، تراز دریاچه، رگرسیون بردار پشتیبان، نشت

مقدمه

پیش‌بینی تراز سطح آب در فواصل زمانی مختلف با استفاده از سری‌های زمانی داده‌های تاریخی یکی از مسائل مهم در برنامه‌ریزی منابع آب است. تغییرات تراز سطح دریاچه‌ها تحت تأثیر عواملی هم-چون بارش، رواناب‌های مستقیم و غیر مستقیم از حوضه‌های مجاور، تبخیر از سطح آزاد دریاچه، دمای آب و هوا و تعاملات بین دریاچه و سفره آب زیرزمینی است. بررسی‌های تراز سطح دریاچه‌ها همواره مورد توجه هیدرولوژیست‌ها و کارشناسان محیطزیست بوده است. مدیریت سفره‌های آب زیرزمینی منطقه، تغییرات اکولوژی منطقه، تغییرات زیست‌محیطی منطقه و تغییرات حجم آب موجود در دریاچه به شکلی تحت تأثیر سطح تراز آب دریاچه‌ها است. اگرچه برای شناسایی مدل‌های پیچیده توجه به پارامترهای فوق لازم است، اما بهتر است در شبیه‌سازی تغییرات تراز سطح دریاچه از داده‌های تاریخی ثبت شده تراز دریاچه استفاده شود (Kisi et al, 2012). در دهه‌های اخیر تکنیک‌های هوش مصنوعی که از ابزارهای تایید شده در مدل‌سازی پارامترهای غیرخطی هیدرولوژیکی می‌باشد، بسیار مورد توجه قرار گرفته است (Kisi, 2009; Wu and Chau, 2010). ماشین بردار پشتیبان[5] یکی از مدل‌های پیشرفته آموزش ماشین است، که بر پایه به حداقل رساندن خطا عمل می‌کند (Yu et al, 2006). این روش بر پایه نظریه یادگیری آماری[6] بنا نهاده شد (Vapnik, 1998) و در گستره وسیعی از مسایل هیدرولوژی (Lin et al, 2010; Lin et al, 2013; Maity et al, 2010; Wu et al, 2010b) استفاده شده است. مدل دیگر برنامه‌ریزی بیان ژن[7] می‌باشد که در ده‌های اخیر به‌عنوان یک روش کارآمد برای حل طیف گسترده‌ای از مشکلات مدل‌سازی بارش-رواناب (Nourani et al, 2012)، تعیین هیدروگراف واحد (Sajjad khan and colibaly, 2006)، روندیابی سیلاب (Sivapragasam et al, 2008) و پیش‌بینی تراز دریاچه‌ها (Ghorbani et al, 2010) مورد استفاده قرار گرفته است. عفیق و همکاران تراز روزانه سطح آب سد کلانگ مالزی را با ماشین بردار پشتیبان و شبکه نرو فازی (ANFIS) پیش‌بینی کردند. نتایج نشان از عملکرد بهتر مدل SVM نسبت به ANFIS در پیش‌بینی تراز روزانه سطح آب این سد دارد (Afiq et al, 2013). مدل SVM و الگوی MLP شبکه عصبی مصنوعی برای پیش‌بینی تراز ماهانه دریاچه

1- دانش آموخته کارشناسی ارشد مهندسی منابع آب، دانشگاه شهرکرد
2- استاد گروه علوم و مهندسی آب، دانشگاه شهرکرد
3- استادیار گروه علوم و مهندسی آب، دانشگاه بوعلی سینا
4- استادیار گروه علوم و مهندسی آب، دانشگاه شهرکرد
(*- نویسنده مسئول: Email: Hosseini_Mostafa69@yahoo.com)

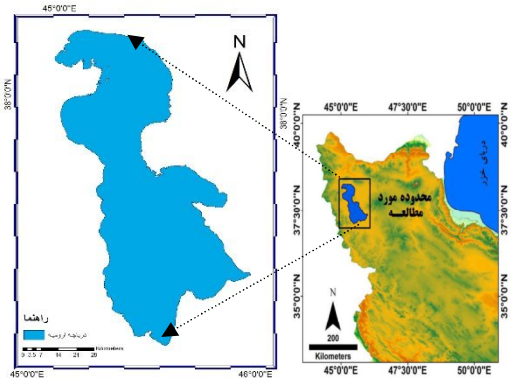
5- Support Vector Machine
6- Statistical Learning Theory
7- Genetic Expression Programming

اریس در شمال آمریکا به کار برده شد (Sajjad khan and colibaly, 2006). نتایج نشان داد که مدل SVM عملکرد بالایی در پیش‌بینی تراز این دریاچه داشته است. برای پیش‌بینی تغییرات دریای خزر ایمانی و همکاران از مدل‌های SVM و GEP به کمک اطلاعات ماهواره‌ای استفاده کردند. نتایج نشان داد که مدل SVM با میانگین مربعات خطا 0/035 متر و ضریب تبیین 0/96 عملکرد بهتری نسبت به مدل GEP دارد (Imani et al, 2014). پیش‌بینی روزانه تراز سطح دریاچه ایزنیک در شرق ترکیه با استفاده از مدل‌های ANN، GEP و ANFIS مورد بررسی قرار گرفت (Kisi et al, 2010). نتایج این تحقیق نشان داد که مدل GEP برآورد بهتری نسبت به مدل‌های شبکه عصبی مصنوعی و نروفازی دارد. در سال‌های اخیر مطالعات زیادی برای پیش‌بینی تراز سطح آب دریاچه ارومیه انجام شده است. کاوه کار و همکاران (1392) در مطالعه‌ای برای پیش‌بینی تراز روزانه سطح آب دریاچه ارومیه از مدل برنامه‌ریزی بیان ژن استفاده کردند. در این مطالعه از نوسانات تراز دریاچه از یک تا نه روز قبل برای ورودی استفاده شد و ترکیب‌های مختلف مورد بررسی قرار گرفت. نتایج نشان داد این مدل دقت مطلوبی در شبیه‌سازی نوسانات سطح آب دارد. پیش‌بینی نوسانات ماهانه آب سطح دریاچه ارومیه با برنامه‌ریزی ژنتیک و مدل شبکه درختی در مقیاس‌های 6، 12، 18 و 24 ماه در مطالعه‌ای مورد بررسی قرار گرفت (صمدیان فرد و همکاران، 1392). نتایج نشان داد که مدل برنامه‌ریزی ژنتیک نسبت به شبکه درختی عملکردبالاتری را دارد. مدل‌های ماشین بردار پشتیبان و شبکه عصبی موجک با استفاده از مقادیر سالانه بارش، دما، رواناب ورودی و تراز آب، برای پیش‌بینی تراز دریاچه ارومیه در مطالعه‌ای به کار گرفته شد (Nouri et al, 2014). نتایج این مطالعه نشان داد، مدل ماشین بردار پشتیبان عملکرد بالاتری نسبت به مدل عصبی موجک در پیش‌بینی نوسانات سالانه تراز این دریاچه دارد. در این مطالعه عملکرد مدل‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن در پیش‌بینی تراز ماهانه دریاچه ارومیه تحت سناریوهای مختلف مورد بررسی قرار گرفته است.

مواد و روش‌ها

منطقه مورد مطالعه

دریاچه ارومیه از لحاظ موقعیت جغرافیایی در عرض 37 درجه و 5 دقیقه تا 38 درجه و 16 دقیقه و طول 45 درجه و 1 دقیقه تا 46 درجه و در ارتفاع 1275 متری از سطح دریا قرار گرفته است (شکل 1). در این مطالعه از دوره آماری 33 ساله از سال آبی 56-55 تا 88-87 به‌صورت ماهانه استفاده شده است. این داده‌ها شامل بارش ثبت شده در سه ایستگاه باران‌سنجی شبستر، ارومیه و میاندوآب، (میانگین بارش این ایستگاه‌ها به‌عنوان بارش سطح دریاچه در نظر گرفته شده است)، تبخیر از سطح دریاچه، نشت (میرعباسی و همکاران، 1393) و سطح تراز دریاچه می‌باشد.



شکل 1- موقعیت جغرافیایی دریاچه ارومیه

برنامه‌ریزی بیان ژن

این روش جز روش‌های الگوریتم گردشی محسوب می‌شود که مبانی تمامی آن‌ها بر اساس نظریه تکامل داروین استوار است. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیار-های کیفی و سپس تابع یاد شده را برای اندازه‌گیری و مقایسه روش-های مختلف حل، در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، روش حل مناسب را ارایه می‌نمایند. برنامه ریزی بیان ژن جدیدترین شیوه از بین روش‌های الگوریتم گردشی می‌باشد که به دلیل دارا بودن دقت کافی، به عنوان مرسوم‌ترین شیوه بوده و از کاربرد بیش‌تری برخوردار است. زمینه اصلی برنامه ریزی بیان ژن، همان الگوریتم ژنتیک است، با این تفاوت که در این روش از شاخه‌های مجزا به جای نوارهای بیت استفاده می‌شود. هر شاخه نیز از یک مجموعه پایانه‌ها (متغیرهای مسئله) و مجموع توابع (عملگر-های اصلی) تشکیل می‌شود (Borelli et al, 2006). فرآیند گام به

گام برنامه‌ریزی بیان ژن به صورت مراحل زیر است:

1) یک جمعیت اولیه از توابع مرکب نشان دهنده مدل‌های پیش بینی، به صورت تصادفی در نظر گرفته می‌شود

2) هر یک از افراد جمعیت مذکور با استفاده از توابع برازش، مورد ارزیابی قرار می‌گیرند

3) در هر تولید، مراحل زیر برای انتخاب یک جمعیت جدید دنبال می‌شود: الف) یکی از عملگرهای عبور، جهش و کپی انتخاب می‌شود. ب) تعداد مناسبی از افراد جمعیت حاضر انتخاب می‌شوند. ج) از عملگر انتخابی برای تولید فرزند استفاده می‌شود.د) فرزند یاد شده در یک جمعیت جدید وارد می‌شود. ه) مدل مورد نظر با استفاده از برازش‌های مختلف مورد ارزیابی واقع می‌شود.

4) گام سوم تا نیل به حداکثر تعداد تولید، تکرار خواهد شد

در این مطالعه برای پیش‌بینی، توسعه و اجرای مدل‌های مبتنی بر برنامه‌ریزی بیان ژن از نرم‌افزار GeneXpro Tools 4 استفاده شده است. فرآیند مدل‌سازی پیش‌بینی تراز دریاچه ارومیه به‌صورت زیر انجام شده است. گام اول، انتخاب توابع برازش مناسب کهدراینمطالعهتابعجذرمیانگینمربعاتخطابه عنوانتابع برازشانتخابگردید. گام دوم، انتخاب مجموعه متغیرهای ورودی و مجموعه توابع به‌منظور تولید کروموزم‌ها می‌باشد. در مسئله حاضر ترمینال‌ها متشکل از مقادیر جریان رودخانه با تاخیرهای زمانی می‌باشد. در این مطالعه، از چهار عملگر اصلی که شامل {+، -، x، ÷} و توابع ریاضی { Log (x)، $\sqrt{x}$ ، $\sqrt[3]{x}$ ، X2 ، X3 و tan(X)} استفاده شد. گام سوم، شامل انتخاب ساختار و معماری کروموزوم‌ها می‌باشد. گام چهارم انتخاب تابع پیوندی است که در این مطالعه عمل جمع برای ایجاد پیوند بین زیر شاخه‌ها مورد استفاده قرار گرفت. درنهایت، در گام پنجم عملگرهای ژنتیک و نرخ هر یک از آن‌ها انتخاب می‌شود.

ماشین بردار پشتیبان رگرسیونی

در یک مدل رگرسیونی ماشین بردار پشتیبان[1]لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین fبه‌علاوه یک مقدار اضافی نویز[2] مشخص می‌شود که در رابطه 1 نشان داد شده است.

$$y = f(x) + Noise \qquad (1)$$

پیش‌بینی کند. این تابع به‌وسیله آموزش مدل SVR بر روی یک مجموعه داده به‌عنوان مجموعه آموزش که شامل فرآیندی به‌منظور بهینه‌سازی دائمی تابع خطا است، قابل دسترسی است. بر مبنای تعریف این تابع خطا، دو نمونه از مدل‌های SVR شناخته شده است که عبارت‌اند از الف) مدل‌های رگرسیونی SVR نوع اول که مدل‌های $SVR - \nu$ نیز نامیده می‌شوند و ب) مدل‌های رگرسیونی SVR نوع دوم که بانام$SVR - \varepsilon$ شناخته شده هستند. در این مطالعه $SVR - \varepsilon$ به دلیل کاربرد گسترده آن در مسائل رگرسیونی مورد استفاده قرار گرفت. برای این مدل، تابع خطا به‌صورت رابطه 2 تعریف می‌شود:

$$\frac{1}{2}W^T W + C\sum_{i-1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^* \qquad (2)$$

تابع خطای فوق لازم است که با توجه به محدودیت‌های رابطه 3 کمینه گردد (Hamel, 2009):

$$\begin{aligned} &W^T\phi(X_i) + b - y_i \le \varepsilon + \xi_i^* \\ &y_i - W^T\phi(X_i) - b \le \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \ge 0 \end{aligned} \qquad (3)$$

که در این روابط C ثابت گنجایش، W بردار ضرایبWT، ترانهاده بردار ضرایب، ξ_i, ξ_i^* ضرایب کمبود، ضریب ثابت، N الگوهای آموزش مدل و ϕ تابع کرنل است.

اطلاعات کمی در مورد انتخاب تابع غیرخطی مناسب ϕ در دسترس می‌باشد. ماشین‌های بردار پشتیبان برای حل مسائل غیرخطی، ابعاد مسئله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای SVR به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. در عمل چهار نوع کرنل خطی[3]،کرنل چندجمله‌ای[4]، کرنل تانژانت هیپربولیک[5]و کرنل گوسی[6](RBF) به کار گرفته می‌شوند. در در جدول 1 معادلات برخی از کرنل‌های رایج آوردهشده‌است.

جدول 1- توابع کرنل رایج در ماشین‌های بردار پشتیبان (Hamel, 2009)

نوع تابع	تابع کرنل
خطی	$K(x_i, x_j) = x_i^T . x_j$
چندجمله‌ای	$K(x_i, x_j) = (\gamma x_i^T x_j + C)^d$
تانژانت هیپربولیک	$K(x_i, x_j) = tanh\,(\gamma x_i^T x_j + C)$
	$K(x_i, x_j) = \exp\,(-\gamma\lvert x_i - x_j\rvert^2)$

درنهایت، تابع تصمیم رگرسیون بردار پشتیبان غیرخطی به‌صورت معادله زیر خواهد بود که کنترل‌کننده میزان نوسان تابع گوسی و هم-چنین کنترل‌کننده نتایج پیش‌بینی و تعمیم‌دهنده مدلSVR است

1- Support Vector Regression
2- Noise

3- Linear kernel
4- Polynomial kernel
5- Hyperbolic tangent kernel
6- Radial Base Function kernel

(Yu et al, 2006).

$$f(x_i) = \sum_{i=1}^{1}(-\partial_i - \partial_i^*)\, K\left(x_i, x_j\right) + b \qquad (4)$$

در این تحقیق از مدل SVR برای پیش‌بینی تراز دریاچه ارومیه توسط نرم‌افزار Statistica 10 و از تابع کرنل استفاده شد. تابع کرنل به دلیل اینکه دقت بالاتری در پیش‌بینی تراز دریاچه داشت به کار گرفته شد. مقادیر بهینه مشخصه‌های مدل SVRشامل ε، C و γباید تعیین شوند. بدین منظور مشخصه‌های ε و C توسط الگوریتم بهینه‌سازی جست‌وجوی شبکه[1]و متغیر γ با آزمون و خطا محاسبه شد. ابتدا شبکه‌هایی با ابعاد بزرگ محدوده مشخصه‌های ε و C به ازای مقدار ثابت γ تعیین می‌شود. سپس این محدوده به ابعادکوچک‌تر تقسیم می‌شود تا مقادیر دقیق‌تر این متغیرها به دست آیند. این فرآیند برای γ مختلف اجرا شده و مدل‌های زیادی به ازای تغییر در مقدار γ ایجاد می‌شوند که از میان آن‌ها مدلی که دارای کم-ترین خطا باشد انتخاب و مشخصه‌های آن به‌عنوان مقادیر بهینه انتخاب می‌شوند.

معیارهای ارزیابی مدل‌ها

به‌منظور مقایسه بارش ماهانه مشاهده‌ای وتخمین زده شده توسطمدل‌هایرگرسیون ماشین بردار پشتیبان (SVR) و برنامه‌ریزی ژنتیک (GEP)، از معیارهای ضریب تبیین (R2) و ریشه میانگین مربعات خطا (RMSE) استفاده شد (احمدی و همکاران، 1393).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_{(pre)i} - R_{(obs)i})^2} \qquad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(R_{(obs)i} - R_{(pre)i})^2}{\sum_{i=1}^{n}(R_{(obs)i} - R_m)} \qquad (6)$$

در این روابط $R_{(pre)i}$ بارش تخمینی در زمان iام، $R_{(obs)i}$ بارش مشاهده شده در همان زمان i و R_m میانگین مقادیر مشاهداتی می‌باشند. هرچه ضریب تبیین بالاتر و RMSE پایین‌تر باشد نتایج دقیق‌تر خواهند بود. برای مدل‌سازی در این مطالعه از داده‌های 25 سال (80-1355) برای آموزش مدل و 8 سال (88-1381) به‌عنوان داده‌های اعتبارسنجی استفاده شد.

بحث و نتایج

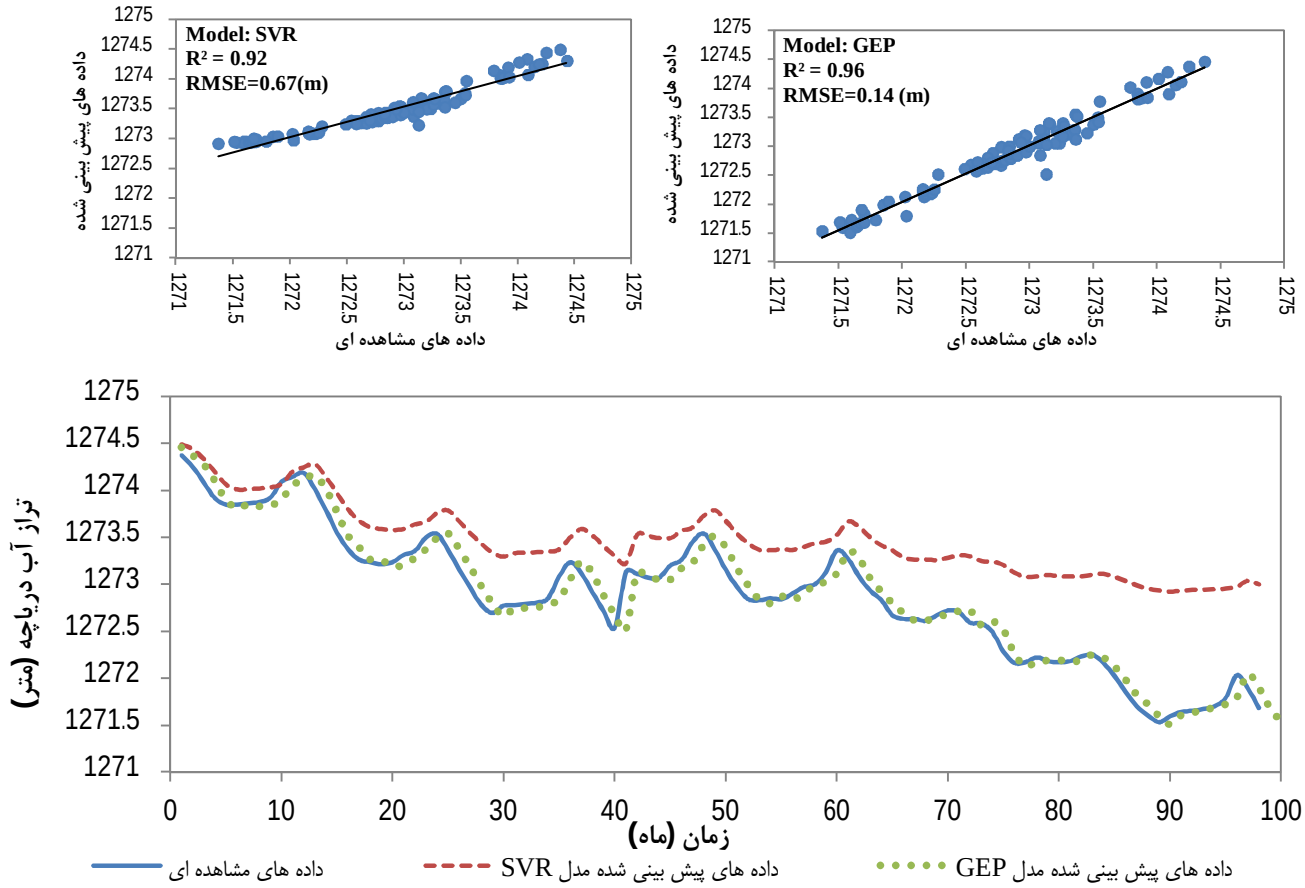
در این مطالعه برای مدل‌سازی نوسانات ماهانه تراز آب دریاچه ارومیه با استفاده از مدل‌های رگرسیون ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن، مدل‌ها به‌صورت شش الگو مورد آموزش و ارزیابی قرار گرفتند که نتایج آن در جدول 2 نشان داده شده است. مقادیر بهینه مشخصه‌های مدل SVR شامل ε، C و γ در بهترین الگو (6) به‌ترتیب 0/2، 10 و 0/325 در نظر گرفته شده است. در مدل‌های SVR و GEP انتخاب جمعیت اولیه و تأثیر گذار در پیش‌بینی به‌منظور آموزش، از عوامل مهم در بهبود عملکرد مدل‌ها می‌باشد. لذا در الگوسازی تراز دریاچه بایستی سعی در انتخاب موثرترین داده‌های مشاهداتی برای آموزش مدل‌ها کرد. در جدول 2 نمایه‌های H، R، E و L به ترتیب نشان دهنده تراز آب (متر)، بارش (میلی‌متر)، تبخیر (میلی‌متر) و نشت (میلی‌متر) می‌باشد. در چهار الگوی اولیه از داده‌های خود تراز آب برای شبیه‌سازی تراز در ماه آینده استفاده شد، به‌طوری که برای پیش‌بینی تراز در الگوی اول از داده‌های ماه قبل، در الگوی دوم از داده‌های سه ماه قبل، در الگوی سوم از داده‌های شش ماه قبل و در الگوی چهارم از داده‌های 12 ماه قبل از ماه مورد نظر استفاده شد. همان‌طور که نتایج نشان می‌دهد با افزایش تاخیرهای زمانی تراز آب عملکرد مدل SVR روند افزایش خطای پیش‌بینی را داشته، و در مقابل عملکرد مدل GEP با توجه به اینکه تغییر چندانی نداشته اما در کل روند افزایش دقت و کاهش خطا را داشته است که در چهار الگو بهترین عملکرد برای مدل SVR (الگو 1) و برای مدل GEP (الگو 4) مقدار ضریب تبیین و RMSE به‌ترتیب 0/92 و 0/67 برای SVR و 0/97 و 0/12 برای مدل GEP محاسبه شد. در الگوهای پنجم برای پیش‌بینی تراز آب در ماه مورد نظر از داده‌های تراز ماه قبل، تبخیر و بارش همان ماه و در الگوی ششم علاوه بر این‌ها از داده‌های نشت به سفره‌های آب زیرزمینی نیز استفاده شد. در الگوی پنجم مدل GEP دقت بالا و خطای پیش‌بینی کم‌تری نسبت به مدل SVR دارد و همان‌طور که نتایج نشان می‌دهد با اضافه شدن داده‌های نشت به آموزش مدل‌ها دقت هر دو مدل افزایش یافته که در این حالت مقدار ضریب تبیین و RMSE به‌ترتیب 0/92 و 0/6 (متر) برای مدل SVR و 0/99 و 0/08 (متر) برای مدل GEP به‌دست آمد و نشان از دقت بالای مدل GEP در پیش‌بینی تراز دریاچه دارد.

به‌منظور ارزیابی و بررسی عملکرد مدل‌های SVR و GEP مقادیر مشاهده شده در مقابل پیش‌بینی شده در شکل‌های 2، 3 و 4 ترسیم گردید. شکل 2 و 3 شبیه‌سازی الگوهای اول و چهار در مرحله اعتبارسنجی را نشان می‌دهد. همان‌طور که این شکل‌ها نشان می‌دهند مدل GEP با وجود اینکه تغییرات کمی داشته، اما دقت آن افزایش یافته است. در حالی که در مدل SVR با افزایش تعداد تاخیرهای زمانی تراز به‌عنوان داده‌های آموزش به مدل، میزان خطا افزایش و عملکرد مدل کاهش یافته است که این نتایج در جدول 2 نشان داده شده است. شکل 4 عملکرد مدل‌ها زمانی که از الگوی ششم برای آموزش مدل‌ها استفاده شده است را نشان می‌دهد. در این حالت مدل‌ها بالاترین دقت و عملکرد را داشته و مدل GEP با میزان خطای 0/08 (متر) بیش‌ترین دقت را در بین سایر الگوها از خود نشان داده است.

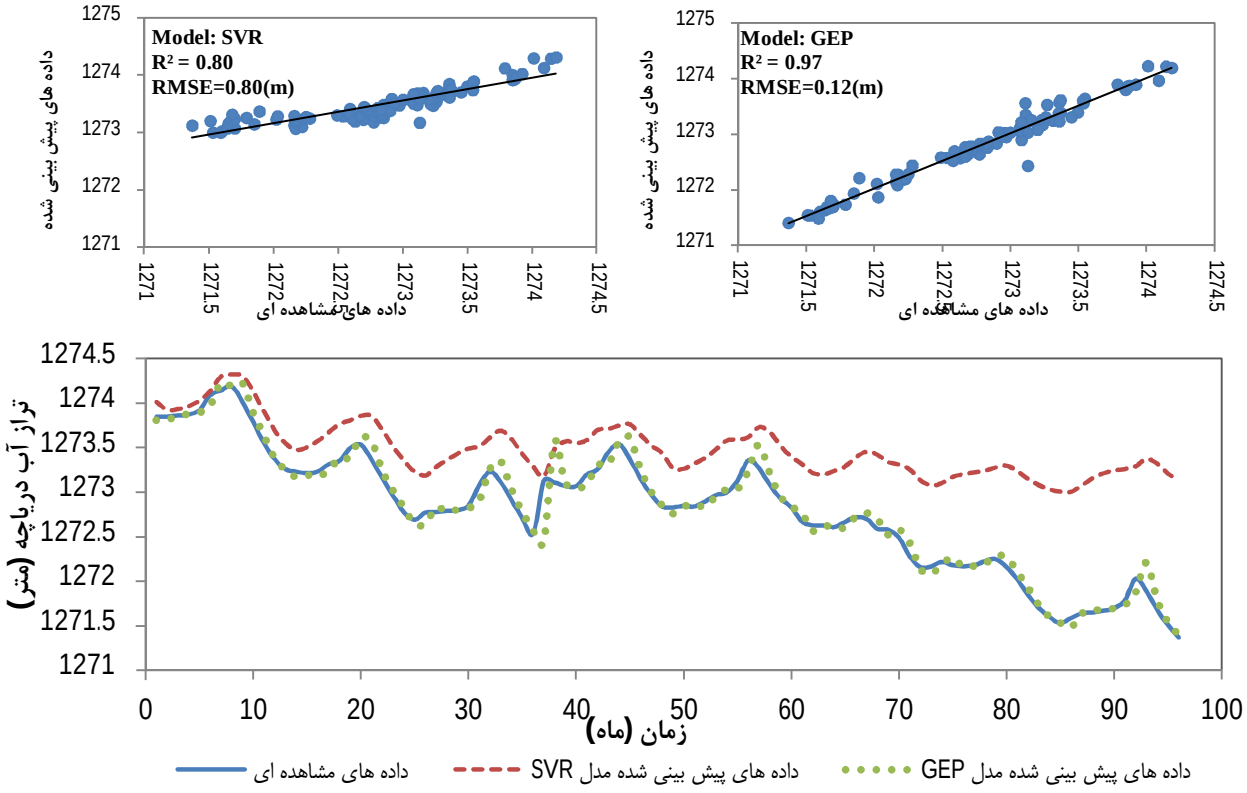
1- Grid Search

جدول 2- عملکرد ترکیب‌های مختلف مدل‌های SVR و GEP در پیش‌بینی نوسانات ماهانه تراز آب دریاچه ارومیه

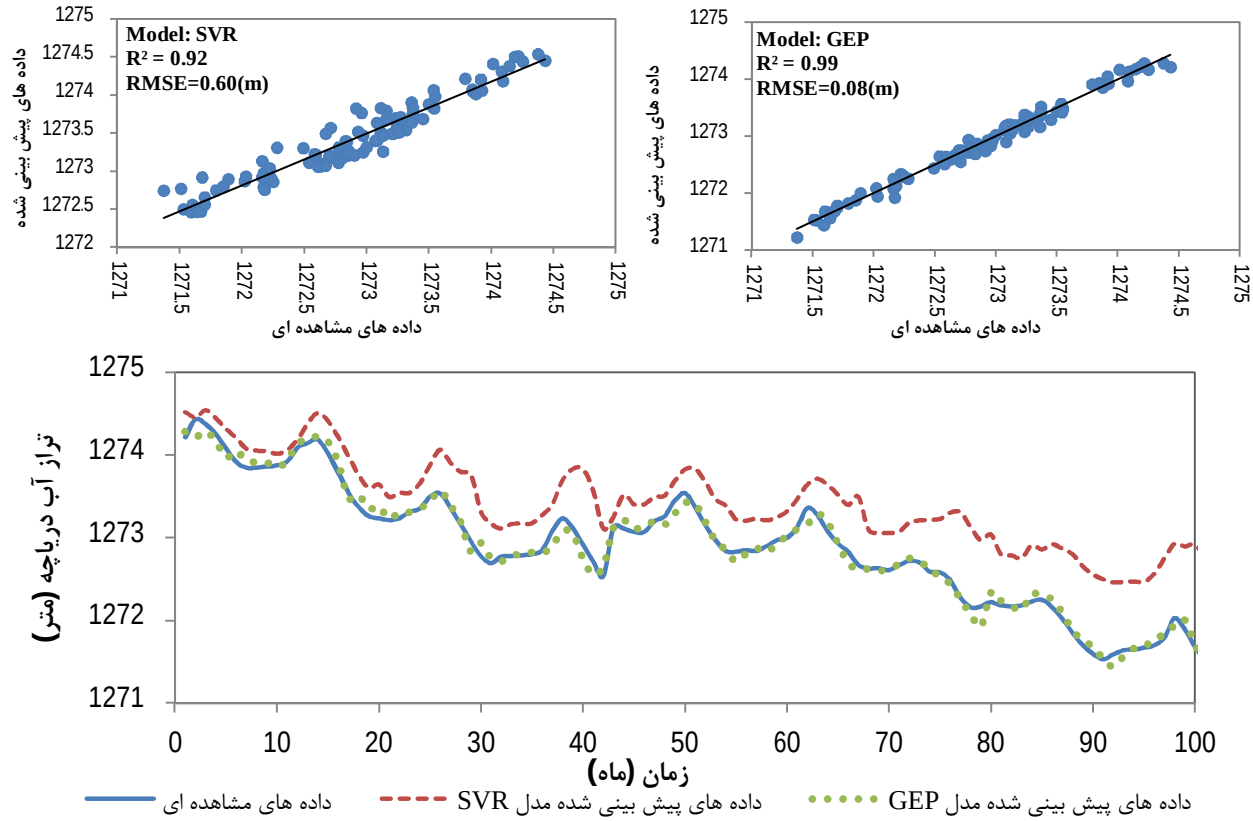
GEP		SVR			
R^2	RMSE (m)	R^2	RMSE (m)	الگو	شماره
0/96	0/14	0/92	0/67	$H_t = f(H_{t-1})$	1
0/97	0/13	0/87	0/80	$H_t = f(H_{t-1}, H_{t-2}, H_{t-3})$	2
0/97	0/12	0/86	0/81	$H_t = f(H_{t-1}, H_{t-2}, ..., H_{t-6})$	3
0/97	0/12	0/80	0/80	$H_t = f(H_{t-1}, H_{t-2}, ..., H_{t-12})$	4
0/97	0/10	0/87	0/65	$H_t = f(H_{t-1}, R_t, E_t)$	5
0/99	0/08	0/92	0/60	$H_t = f(H_{t-1}, R_t, E_t, L_t)$	6



شکل 2- مقایسه داده‌های مشاهده‌ای و شبیه‌سازی شده تراز آب دریاچه ارومیه با اجرای الگو 1



شکل 3- مقایسه داده‌های مشاهده‌ای و شبیه‌سازی شده تراز آب دریاچه ارومیه با اجرای الگوی 4



شکل 4- مقایسه داده‌های مشاهده‌ای و شبیه‌سازی شده تراز آب دریاچه ارومیه با اجرای الگو 6

نتیجه‌گیری

در این تحقیق عملکرد مدل‌های ماشین بردار پشتیبان و برنامه‌ریزی بیان ژن در پیش‌بینی ماهانه تراز آب دریاچه ارومیه در خلال سال‌های 1355 تا 1388 مورد بررسی قرار گرفت. در بخش آموزش مدل‌ها از شش الگو استفاده شد. در الگوهای اول تا چهارم برای پیش‌بینی در ماه مورد نظر از خود داده‌های تاریخی اندازه‌گیری شده تراز آب، به‌صورت 1، 3، 6 و 12 ماه پیش از ماه مورد نظر استفاده شد. در الگوهای پنج و شش ترکیبی از داده‌های بارش، تبخیر، تراز آب و نشت برای آموزش مدل‌ها مورد بررسی قرار گرفت. نتایج این مطالعه به‌صورت زیر ارائه شده است :

با افزایش تاخیرهای زمانی تراز آب (الگوهای 1 تا 4) به ترتیب 1، 3، 6 و 12 ماه قبل از ماه مورد نظر برای آموزش مدل‌ها میزان دقت و عملکرد مدل SVR از مقدار 0/92 به 0/80 برای R^2 و 0/67 به 0/80 متر برای RMSE کاهش، و دقت و عملکرد مدل GEP از مقدار 0/96 به 0/97 برای R^2 و 0/14 به 0/12 متر برای RMSE افزایش یافته است.

با توجه به اینکه مدل GEP زمانی که از داده‌های تراز برای پیش‌بینی استفاده می‌کند از الگوی اول تا الگوی چهارم با وجود افزایش اندک دقت، عملکرد تغییر چندانی نداشته است بنابراین هرگاه با کمبود داده مواجه هستیم می‌توان از این مدل و الگو استفاده کرد.

برای بهترین عملکرد مدل SVR (الگوی 6) مقادیر بهینه پارامترهای این مدل یعنی ε و C با استفاده از الگوریتم بهینه‌سازی جستجوی شبکه به ترتیب معادل 0/100 و 10 و پارامتر γ با سعی و خطا برابر 0/480 به دست آمد.

در الگوی ششم با اضافه کردن داده‌های نشت برای آموزش مدل‌ها عملکرد افزایش، و خطا به میزان قابل توجهی کاهش یافت. در این حالت میزان ضریب تبیین و میانگین مربعات خطا برای مدل SVR و GEP به ترتیب 0/92 و 0/60 متر و 0/99 و 0/08 متر به دست آمد.

به‌طور کلی در تمامی الگوها مدل GEP عملکرد بهتری نسبت به مدل SVR از خود نشان داد، که این نتایج با تحقیقات صمدیان فرد و همکاران (1392)، کاوه کار و همکاران (1392) که نشان دادند برنامه‌ریزی ژنتیک عملکرد بالایی در پیش‌بینی روزانه تراز دریاچه ارومیه دارد مطابقت دارد. هم‌چنین نوری و همکاران (2014) نشان دادند SVR عملکرد مناسبی در پیش‌بینی تراز دریاچه ارومیه دارد. بنابراین این دو مدل توانایی بالایی در پیش‌بینی تراز دریاچه ارومیه در مطالعات گذشته از خود نشان داده‌اند.

منابع

احمدی،ف.، رادمنش،ف و میرعباسی،ر. 1393. مقایسه روش‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان در پیش‌بینی جریان روزانه رودخانه (مطالعه موردی: رودخانه باراندوز چای). مجله دانش آب و خاک مشهد. 26: 1025-1010.

کاوه کار،ش.، قربانی،م.ا.، افشار زاده،ا و دربندی،س. 1392. شبیه‌سازی نوسانات سطح آب با استفاده از برنامه‌ریزی بیان ژن. مجله عمران و محیطزیست. 63:43-53.

میرعباسی،ر.، کدخداحسینی،م.، شامحمدی،ش و نوذری،ح. 1393. بررسی میزان تبادل آب بین دریاچه ارومیه و سفره‌های آب زیر زمینی، دومین همایش بحران آب، 19-18 شهریور، دانشگاه شهرکرد.

صمدیان فرد،س.، حسنیا،د و ستاری،م.ت. 1392. پیش‌بینی نوسانات آب دریاچه ارومیه با استفاده از شبکه درختی M5 و برنامه ژنتیک، سی و دومین گردهمایی و نخستین کنگره بین المللی علوم زمین، 20-17 دی. دانشگاه ارومیه.

Afiq,H., Ahmed,E., Ali,N., Othman, A., Aini, K.,Mukhlisi,H.M. 2013. Daily forecastingofdam water levels: comparing a support vector machine (SVM) model withadaptiveneuro fuzzy inference system (ANFIS). Water Resources Management. 27:3803-3823.

Borelli,A., De Falco,I., Della,C.A., Nicodemi,M.,Trautteur,G. 2006. Performance of Genetic Programming toExtract the Trend in Noisy Data Series. Physica A. 370: 104-108.

Ghorbani,M., Khatibi,R., Aytek,A.,Makarynskyy,O. 2010. Sea water level forecasting using genetic programming and artificial neural networks. Journal of Computers and Geosciences. 36.5:620 627.

Hamel,L. 2009. Knowledge Discovery with Support Vector Machines. Hoboken. N.J. John Wiley.

Imani,M., You,R.J., Kou,C.Y. 2014. Forecasting Caspian Sea level changes using satellite altimetry data (June 1992–December 2013) based on evolutionary support vector regression algorithms and gene expression programming. Global and Planetary Change. 121:53–63.

Kisi,O. 2009. Neural networks and wavelet conjunction model for intermittent stream flow forecasting. Journal of Hydrologic Engineering. 14.8: 773–782.

Kisi,O., Shiri,J., Nikoofar.,B. 2012. Forecasting daily lake levels using artificial intelligence approaches. Computers and Geosciences Journal. 41: 169-180.

Lin,G.F., Chen,G.R., Huang,P.Y. 2010. Effective typhoon characteristics and their effects on hourly reservoir inflow forecasting. Advance Water Resources. 33.8:887–898.

Lin,G.F., Chou,Y.C., Wu,M.C. 2013. Typhoon flood forecasting using integrated two-stage support vector

Sivapragasam,C., Maheswaran,R., Veena,V. 2008. Genetic programming approach for flood routing in natural channels. Hydrological Processes. 22:623–628.

Vapnik,V.N. 1998. Statistical Learning Theory. Wiley, New York.

Wu,C.L., Chau,K.W. 2010. Data-driven models for monthly streamflow time series prediction. Engineering Applications of Artificial Intelligence. 23:1350–1367.

Wu,J., Liu,M., Jin,L. 2010. Least square support vector machine ensemble for daily rainfall forecasting based on linear and nonlinear regression. Advances in Neural Network Research and Applications. Lecture Notes in Electrical Engineering. 67.1:55–64.

Yu,P.S., Chen,S.T., Chang,I.F. 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology. 328.3–4:704–716.

machine approach. Journal of Hydrology. 486:334–342.

Maity,R., Bhagwat,P.P.,Bhatnagar,A. 2010. Potential of support vector regression for prediction of monthly streamflow using endogenous property. Hydrological Processes. 24.7:917–923.

Nourani,V., Komasi,M., Alami,M.T. 2012. Hybrid wavelet–genetic programming approach to optimize ANN modelling of rainfall–runoff process. Journal of Hydrologic Engineering. 17.6:724–741.

Noury,M., Sedghi,H., Babazadeh,H.,Fahimi,H. 2014. Urmia Lake Water Level Fluctuation Hydro Informatics Modeling Using Support Vector Machine and Conjunction of Wavelet and Neural Network. Hydropgysical Processes. 41:261-269.

Sajjad khan,M., Colibaly,P. 2006. Application of Support Vector Machine in Lake Water Level Prediction. Journal of Hydrology. 11:199-2-5.

Iranian Journal of Irrigation and Drainage
No. 2, Vol. 10, May.-Jun. 2016, p. 199-207

نشریه آبیاری و زهکشی ایران
شماره2، جلد10، خرداد - تیر 1395، ص. 207-199

Evaluate the Performance of SVR and GEP Models in Predicting the Monthly Fluctuations in Water Level of Urmia Lake

M. Kadkhodahosseini[1]*, Sh. Shamohammadi[2], H. Nozari[3], R. Mirabbasi[4]
Recived: Mar.12, 2016 Accepted: Jun.08, 2016

Abstract

Prediction of lake level fluctuations is one of the most important issues in water resources planning and management. In recent years, the significant decline in water level of Urmia Lake had detrimental environmental impacts on this region. In this study, the performance of genetic expression programming and support vector regression models for predicting Urmia Lake water level was evaluated based on six different patterns during 1976-2009 to determine the best input pattern. The historical data of water level were used in four patterns, and precipitation, evaporation, seepage and water level were used in two other patterns. The results showed that the genetic expression programming had better performance than SVR model and the model performance improves with increasing input for model training. Also, at the best pattern the mean square error and coefficient of determination were calculated 0.08 m and 0.99 for GEP model and 0.60 m and 0.92 for SVM, respectively.

Keywords: Genetic expression programming, Lake level, Regression vector machine, Urmia lake, Seepage

1- Graduated. Student of Water Resources Engineering, Department of Water Engineering, College of Agriculture, Shahrekord University, Shahrekord
2- Professor, Department of Water Engineering, College of Agriculture, Shahrekord University, Shahrekord
3- Assistant Professor, Department of Water Engineering, College of Agriculture, BualiSina University, Hamedan
4- Assistant Professor, Department of Water Engineering, College of Agriculture, Shahrekord University, Shahrekord
(*- Corresponding Author Email: Hosseini_mostafa69@yahoo.com)