

Regional flood analysis by random forest method in the Namak Lake basin

Saeid Khosrobeigi¹ | Arash Malekian¹ | Alireza Moghaddam Nia^{1,2} | Shahram Khalighi¹

1- Faculty of Natural Resources, University of Tehran, Karaj, Iran

2- Faculty of Renewable Energies and Environment, University of Tehran, Tehran, Iran

Email: malekian@ut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 25 Apr. 2020
Revised: 19 Nov. 2021
Accepted: 20 Nov. 2021
Published online: 18 Apr. 2024

Keywords:
Gamma test,
Lo Pearson III distribution,
Return period,
Maximum instantaneous discharge.

Abstract

Flood is one of the most devastating natural disasters, causing financial and human losses each year. At the same time, many rivers in Iran's watersheds lack complete and accurate statistics and information. On the other hand, estimating the flow of floods is one of the most important factors for the design and implementation of water structures. In such cases, one of the appropriate solutions to estimate the maximum flow rate with different return periods is flood analysis. In order to conduct the present study, 55 hydrometric stations with a common statistical period of 20 years were considered to perform the work after the statistical deficiencies were eliminated. Then, based on the distribution of the third type of Pearson logo with the lowest error rate and the highest number of first rank as the most suitable fit function, the amount of discharge in different return periods was estimated. The following information was collected on the types of physiography, land use, climate and geology variables. After collecting information about all independent variables using Gamma test, the most important variables affecting the maximum instantaneous flow, including area, drainage density, maximum 24-hour rainfall and watershed environment, were selected and modeled using methods. Random forest modeling and support vector modeling were performed in R software and the efficiency of these two methods was determined based on statistical indicators of R² coefficient, RMSE root mean and CE and clicker efficiency coefficient. With an efficiency coefficient of 74 to 83%, the error of 3.05 to 32.11 m³ and the coefficient of explanation of 76 to 91 are more accurate than the random forest model.

Cite this article: Khosrobeigi, S., Malekian, A., Moghaddam Nia, A.R., Khalighi, S. (2024). Regional flood analysis by random forest method in the Namak Lake basin. *Journal of Range & Watershed Management*, 77 (1), 1-15. DOI: <http://doi.org/10.22059/jrwm.2021.301588.1491>



© The Author(s).

Publisher: University of Tehran Press

تحلیل منطقه‌ای سیلاب با استفاده از روش جنگل تصادفی در حوزه آبخیز دریاچه نمک

سعید خسرویگی^۱ | آرشی ملکیان^۱ | علیرضا مقدم‌نیا^{۱،۲} | شهرام خلیقی سیگارودی^۱

۱- گروه احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی دانشگاه تهران، کرج، ایران

۲- دانشکده مهندسی انرژی و منابع پایدار، دانشکده‌گان علوم و فناوری‌های میان رشته‌ای دانشگاه تهران، تهران، ایران

رایانامه: malekian@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله:

مقاله پژوهشی

تاریخ دریافت: ۱۳۹۹/۰۲/۰۶

تاریخ بازنگری: ۱۴۰۰/۰۸/۲۸

تاریخ پذیرش: ۱۴۰۰/۰۸/۲۹

تاریخ انتشار: ۱۴۰۳/۰۱/۳۰

کلیدواژه‌ها:

آزمون گاما،

توزیع لوگ پیرسون نوع سوم،

دوره بازگشت،

دبی حداکثر لحظه‌ای.

سیل یکی از مخرب‌ترین بلایای طبیعی است که هر ساله باعث تلفات مالی و جانی می‌شود. در عین حال بسیاری از رودخانه‌های موجود در حوزه‌های آبخیز ایران فاقد آمار و اطلاعات کامل و دقیق در مورد دبی‌های سیلابی می‌باشند. از طرفی تخمین دبی سیلاب یکی از مهم‌ترین عوامل لازم جهت طراحی و اجرای سازه‌های آبی است. در چنین مواردی یکی از راه‌حل‌های مناسب برای برآورد دبی‌های حداکثر لحظه‌ای با دوره بازگشت‌های مختلف آنالیز منطقه‌ای سیلاب می‌باشد. به منظور انجام پژوهش حاضر، تعداد ۵۵ ایستگاه آب‌سنجی با دوره مشترک آماری ۲۰ ساله پس از رفع نواقص آماری برای انجام کار در نظر گرفته شدند. سپس بر اساس توزیع لوگ پیرسون نوع سوم با کمترین میزان خطا و بیشترین تعداد رتبه اول به عنوان مناسب‌ترین تابع برازش، مقدار دبی در دوره بازگشت‌های مختلف برآورد گردید. در ادامه اطلاعات مربوط به انواع متغیرهای فیزیوگرافی، کاربری اراضی، اقلیمی و زمین‌شناسی جمع‌آوری شد. پس از جمع‌آوری اطلاعات مربوط به کلیه متغیرهای مستقل با استفاده از آزمون گاما مهم‌ترین متغیرهای موثر بر دبی‌های حداکثر لحظه‌ای شامل مساحت، تراکم زهکشی، حداکثر بارندگی ۲۴ ساعته و محیط حوزه آبخیز انتخاب و مدل‌سازی با استفاده از روش‌های مدل‌سازی جنگل تصادفی و ماشین بردار پشتیبان در نرم‌افزار R انجام پذیرفت و میزان کارایی این دو روش بر اساس نمایه‌های آماری ضریب تبیین (R^2)، ریشه میانگین مربعات خطا ($RMSE$) و ضریب کارایی ناش و ساتکلیف (CE) مشخص شد. با ضریب کارایی ۷۴ تا ۸۳ درصد، خطای ۳/۰۵ تا ۳۲/۱۱ متر مکعب و ضریب تبیین ۷۶ تا ۹۱ نسبت به نسبت به مدل جنگل تصادفی از دقت بالاتری برخوردار می‌باشند.

استناد: خسرویگی، سعید؛ ملکیان، آرشی؛ مقدم‌نیا، علیرضا؛ خلیقی سیگارودی، شهرام (۱۴۰۳). تحلیل منطقه‌ای سیلاب با استفاده از روش جنگل تصادفی در حوزه آبخیز دریاچه نمک. نشریه مرتع و آبخیزداری، ۷۷(۱)، ۱-۱۵.

DOI: <http://doi.org/10.22059/jrwm.2021.301588.1491>

© نویسندگان.

ناشر: انتشارات دانشگاه تهران.

۱. مقدمه

سیل از جمله پدیده‌هایی است که هر ساله خسارت‌های جبران ناپذیری به اقتصاد کشورها وارد می‌کند. لذا پژوهش در خصوص ویژگی‌های این پدیده طبیعی از اهمیت بالایی برخوردار است. از پارامترهای مهم در معرفی سیلاب‌ها حداکثر دبی با دوره برگشت‌های مختلف می‌باشد که همیشه در طراحی سازه‌های آبی تخمین مناسبی از آن مورد نیاز بوده و از نظر اقتصادی نقش تعیین کننده ای در محاسبه حجم عملیات سازه ای ایفا می‌کند. در بسیاری از حوضه‌ها خصوصیات فیزیوگرافی و زمین شناسی، نقش موثری در تولید دبی سیلابی ایفا می‌کند، این ویژگی‌ها با گذشت زمان‌های نسبتاً طولانی ثابت بوده و لذا می‌توان از آنها به عنوان پارامتر مستقل برای تخمین دبی حداکثر به خصوص در مناطقی که فاقد ایستگاه هیدرومتری هستند، استفاده نمود و این ویژگی‌ها در حوضه‌های فاقد آمار مهمترین نقش را در برآوردهای هیدرولوژی دارند (عزیز^۱ و همکاران، ۲۰۱۷). تحلیل منطقه‌ای سیلاب روشی است که در حوضه‌های فاقد آمار جریان با استفاده از آمار جریان سیلاب در حوضه‌های مجهز به ایستگاه‌های هیدرومتری روابطی بین مقادیر جریان سیل و برخی از ویژگی‌های مرفولوژی، فیزیوگرافی، اقلیمی، زمین شناسی، خاکشناسی، پوشش گیاهی و نحوه کاربری اراضی حوضه‌ها برآورد می‌کند (عزیز و همکاران، ۲۰۱۴). در تحلیل فراوانی سیل منطقه‌ای اغلب یک روش‌های خطی مبتنی بر رگرسیون برای تعیین دبی حداکثر با دوره بازگشت معین استفاده می‌شود که از آنجایی که بیشتر فرآیندهای هیدرولوژیک غیر خطی هستند، دارای تغییرپذیری زمانی و مکانی زیادی هستند، مدل‌سازی خطی نمی‌تواند این فرآیندها را به خوبی نمایش دهند بنابراین مدل‌های غیر خطی می‌تواند فرآیندهای هیدرولوژیک، که طبیعتاً غیر خطی هستند، را بهتر توجیه کنند (عزیز و همکاران، ۲۰۱۴). یکی از رویکردهای غیر خطی برای مدل‌سازی فرآیندهای هیدرولوژیک روش‌های مبتنی بر ماشین یادگیری است. از روش‌های ماشین یادگیری، که می‌تواند به صورت بالقوه برای تحلیل فراوانی سیل منطقه‌ای استفاده شود، تکنیک‌های رگرسیون پشتیبان بردار^۲ می‌باشد و بر اساس الگوریتم طبقه بندی پایه کرنال که ماشین بردار پشتیبان نیز نامیده می‌شود، توسعه یافته است. چندین سال است که ماشین بردار پشتیبان به عنوان یک ابزار رگرسیون تحت نام ماشین رگرسیون بردار پشتیبان توسعه یافته است (زکریا و شبیری^۳، ۲۰۱۲). جنگل‌های تصادفی، روش‌های یادگیری جدیدی است، که از الگوریتم‌های پایه برای پیش بینی تکراری چندگانه استفاده می‌کند. جنگل‌های تصادفی می‌تواند تعداد بی شماری متغیر ورودی را بدون حذف یکی از آنها تجزیه و تحلیل کند. همچنین نسبت به شبکه عصبی مصنوعی در شرایط غیر افتادن در کمینه محلی و داده‌های نویزدار، حساسیت کمتری دارد و می‌تواند تخمین بهتری از پارامترها داشته باشد (پورقاسمی و کلر^۴، ۲۰۱۶).

زکریا و شبیری (۲۰۱۲) روش مدل‌سازی ماشین بردار پشتیبان را به منظور پیش‌بینی دبی حداکثر در دوره بازگشت‌های ۱۰ و ۱۰۰ ساله با استفاده از آمار ۸۸ ایستگاه آب‌سنجی در مالزی به کار گرفتند و نتایج آن را با روش رگرسیون چند متغیره مقایسه کردند. بر این اساس نتایج نشان داد که ماشین بردار پشتیبان از عملکرد بهتری نسبت به رگرسیون چند متغیره برخوردار است. عزیز و همکاران (۲۰۱۴) شبکه عصبی مصنوعی را برای انجام تحلیل منطقه‌ای سیلاب در شرق استرالیا استفاده کردند. برای این منظور تعداد ۴۵۲ حوزه آبخیز و دو متغیر ورودی شامل مساحت و شدت بارندگی به کار گرفته شد. نتایج حاصل از این پژوهش نشان داد که شبکه عصبی مصنوعی از دقت بالاتری نسبت به روش‌های سنتی رگرسیونی برخوردار است. دروچر^۵ و همکاران (۲۰۱۷) از کاپولای فضایی برای پیش بینی دبی‌های حداکثر در دوره‌های بازگشت مختلف حوضه‌های فاقد آمار در کبک کشور کانادا استفاده کردند. نتایج نشان داد، که کاپولای فضایی قادر به حل مشکل مربوط به ارب شدن مدل را دارد و در حضور ایستگاه‌های مشکل دار مناسب است. به علاوه، کیفیت پیش‌بینی دبی‌های حداکثر را بهبود می‌بخشد.

¹ Aziz

² Support Vector Machine - Regression (SVR)

³ Zakaria & Shabri

⁴ Pourghasemi & Kerle

⁵ Durocher

ژیانک^۱ و همکاران (۲۰۱۷) به تحلیل فراوانی سیل منطقه‌ای با استفاده از روش توزیع هالفن همراه با اصل حداکثر آنتروپی برای ۱۲ ایستگاه هیدرومتری در کشور آمریکا پرداختند. نتایج نشان داد، که کارایی این مدل در مقایسه با روش دیگر بهتر است. رحمان^۲ و همکاران (۲۰۱۸) به تحلیل فراوانی سیل منطقه‌ای با استفاده از مدل‌های افزایشی تعمیم یافته (GAM) برای ۸۵ ایستگاه هیدرومتری در جنوب کشور استرالیا پرداختند. نتایج نشان داد، که کارایی GAM نسبت به سایر مدل‌ها مانند رگرسیون غیر خطی بهتر می‌باشد. نتایج این پژوهش نشان داد که بر اساس آزمون نکوتی برازش Z^{DIST} برای نواحی همگن ۱، ۲ و ۳ توزیع لجستیک تعمیم یافته به عنوان بهترین توزیع منطقه‌ای است. پناهی و علیجانی (۲۰۱۳) پژوهشی را برای پیش‌بینی دبی اوج با استفاده از روش مدل‌سازی شبکه عصبی مصنوعی و رگرسیون چندمتغیره در حوزه آبخیز مدارس انجام دادند و بیان داشتند که دقت روش مدل‌سازی شبکه عصبی مصنوعی بیشتر از مدل رگرسیون چند متغیره است. نگارش و همکارا (۲۰۱۳) پژوهشی را به منظور شبیه‌سازی و پیش‌بینی سیلاب در رودخانه سرباز در استان سیستان و بلوچستان با استفاده از روش مدل‌سازی شبکه عصبی مصنوعی به انجام رساندند. مقایسه نتایج روش‌های مدل‌سازی شبکه عصبی مصنوعی و رگرسیونی نشان داد که روش مدل‌سازی شبکه عصبی از عملکرد مناسب‌تری نسبت به روش رگرسیونی به‌منظور پیش‌بینی جریان در این رودخانه برخوردار است. با توجه به مطالعات قبلی تعداد محدودی از الگوریتم‌های یادگیری همچون ماشین بردار پشتیبان با جنگل تصادفی برای مدل‌سازی دبی‌های حداکثر لحظه‌ای استفاده شده است. لذا هدف از پژوهش حاضر، استفاده و مقایسه عملکرد ماشین بردار پشتیبان با جنگل تصادفی به منظور مدلسازی دبی‌های حداکثر لحظه‌ای در حوزه آبخیز دریاچه نمک می‌باشد.

۲. مواد و روش‌ها

۲-۱. منطقه مورد مطالعه

حوزه آبخیز دریاچه نمک با کد ۴۱ در طول جغرافیایی ۴۸ تا ۵۳ شرقی و عرض جغرافیایی ۳۳ تا ۳۶ شمالی با مساحت ۹۲۵۵۰ کیلومتر مربع واقع شده است (شکل ۱). بخش شمالی این حوزه آبخیز به وسیله رشته کوه البرز و کوه‌های رودبار و بخش شمال غرب و غرب آن با انشعابات زاگرس و زاگرس اصلی، بخش جنوبی با کوه‌های کرکس و شرق آبخیز به مناطق کویری محدود می‌گردد. از این میان حدود ۴۲۰۲۶ کیلومتر مربع آن را مناطق کوهستانی و ۵۰۵۲۴ کیلومتر مربع باقیمانده را دشت‌ها، کوهپایه‌ها و دریاچه تشکیل می‌دهند. طیف ارتفاعی از ۸۰۰ متر در اطراف دریاچه نمک تا ۴۳۷۵ متر در ارتفاعات جاجرود متغیر است (وفاخواه، ۱۹۹۸).

۲-۲. روش تحقیق

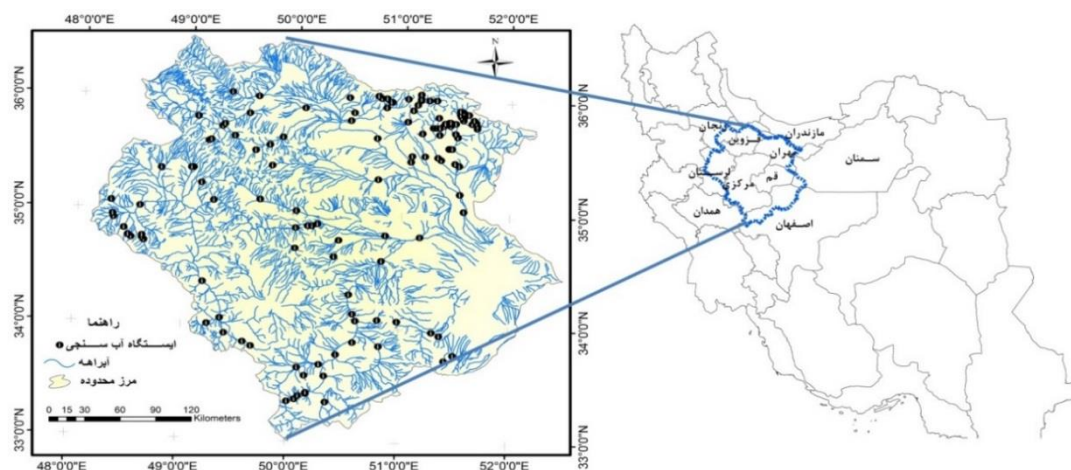
۲-۲-۱. جمع آوری اطلاعات

برای انجام تحقیق حاضر، ابتدا داده‌های حداکثر دبی لحظه‌ای سیلاب سالانه کلیه ایستگاه‌های موجود در منطقه از سازمان تحقیقات منابع آب ایران و سازمان آب و منطقه‌ای استان‌های موجود در حوزه آبخیز دریاچه نمک اخذ و مورد بازبینی قرار گرفت. جمعا ۱۹۸ ایستگاه هیدرومتری در حوزه آبخیز مطالعاتی موجود بود. تعدادی از ایستگاه‌ها آماری لازم نبود و تعدادی از ایستگاه‌ها نیز در آزمون ران تست برای همگن نبودن داده‌ها کنار گذاشته شدند. لذا به عنوان مطلوب‌ترین حالت ممکن تعداد ۵۵ ایستگاه با دوره آماری از سال آبی ۲۰ ساله وارد محاسبات گردیدند، برای تعیین دبی حداکثر لحظه‌ای در دوره بازگشت‌های مختلف، از نرم افزار EasyFit استفاده شد. برای نیل به این هدف ابتدا دبی‌های حداکثر لحظه‌ای با توزیع‌های مختلف آماری برازش داده شدند. سپس رتبه‌بندی، هریک از توزیع‌های فوق بر اساس آزمون برازش کلموگروف-اسمیرنوف از نمره ۱ تا ۷ مشخص شد. در ادامه مجموع نمرات برای هر کدام از توزیع‌ها در ۵۵ ایستگاه منتخب

¹ Xiong

² Rahman

محاسبه و بهترین توزیع آماری با کمترین نمره به دست آمد. در نهایت دبی با دوره بازگشت‌های ۲، ۵، ۱۰، ۲۵، ۵۰ و ۱۰۰ ساله با استفاده از بهترین توزیع آماری تعیین شد. به منظور استخراج خصوصیات فیزیوگرافی هر یک از ایستگاه‌ها از مدل رقومی ارتفاع (DEM) به پیکسل سایز ۵۰*۵۰ متر، برای تعیین پارامتر زمین‌شناسی ابتدا کلیه سازندهای زمین‌شناسی زیر حوزه‌های بالادست ایستگاه‌های منتخب مساحی و سپس با توجه به نفوذپذیر (سازندهای کواترنری) یا غیر قابل نفوذپذیر بودن هر یک از سازندها درصد سازندهای نفوذپذیر در هر یک از زیر حوزه‌های منطقه با استفاده از نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ تعیین شد. در ادامه برای استخراج متغیر کاربری اراضی از نقشه‌های کاربری اراضی ۱:۲۵۰۰۰۰ و با کمک عکس هوایی ۱:۵۰۰۰۰ در نرم‌افزار ArcGIS و برای تعیین پارامترهای اقلیمی از جمله حداکثر بارندگی ۲۴ ساعته، متوسط بارندگی سالانه، متوسط درجه حرارت سالانه و متوسط تبخیر سالانه از آمار ایستگاه‌های هواشناسی منطقه مورد مطالعه به روش گردایان استفاده شد.



شکل ۱. موقعیت منطقه مورد مطالعه

۲-۲-۲. تعیین مهم‌ترین متغیرها در تحلیل منطقه‌ای سیلاب

در این مرحله می‌بایست مهم‌ترین متغیرهای موثر بر دبی سیلاب را تعیین کرد. برای این منظور از آزمون گاما به کمک نرم‌افزار WinGamma بهره گرفته شد. در واقع آزمون گاما ابزاری مناسب برای تعیین ترکیب بهینه ورودی‌ها و تعداد داده‌های مناسب برای رسیدن به کم‌ترین میانگین مربعات خطا در هر گونه مدل‌سازی غیر خطی پیوسته می‌باشد (بسالت‌پور و همکاران، ۱۳۹۱). قابلیت دیگر نرم‌افزار WinGamma تقسیم بندی داده‌های استاندارد شده به دو دسته آموزش و آزمون مدل با استفاده از آزمون M، می‌باشد.

۲-۲-۳. استانداردسازی داده‌ها

اصولاً وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت مدل می‌شود. برای احراز از چنین شرایطی و همچنین به منظور یکسان کردن ارزش داده‌ها، برای اجرای جنگل تصادفی و ماشین بردار پشتیبان استانداردسازی داده‌ها انجام می‌پذیرد (کیشی و همکاران، ۲۰۱۳). رابطه ۱ رابطه استاندارد سازی داده‌ها به منظور استفاده در مدل‌های مورد نظر را نشان می‌دهد. این رابطه، استانداردسازی داده‌ها بین دو مقدار ۰/۱ تا ۰/۹ را نشان می‌دهد.

$$X_{Normal} = 0.8 \times \frac{(X_i - X_{min})}{(X_{max} - X_{min})} + 0.1 \quad (\text{رابطه ۱})$$

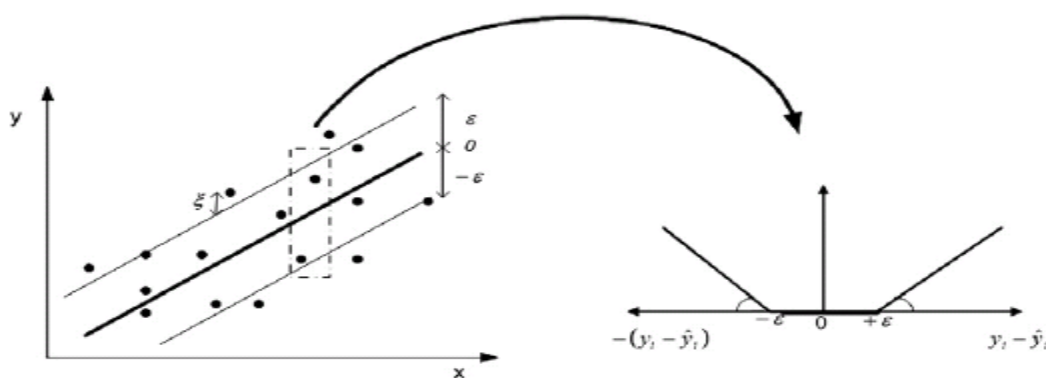
که در آن: X_{normal} : داده نرمال شده X_i : داده مشاهده شده X_{min} : داده حداقل X_{max} : داده حداکثر

۲-۲-۴. ماشین بردار پشتیبان (SVM)

ماشین پشتیبان بردار یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. برای ساختن شکل رگرسیونی ماشین پشتیبان بردار از تابع خطای جدیدی استفاده نمود که به نام تابع خطای ϵ -insensitive خوانده می‌شود و به صورت زیر تعریف می‌گردد (شکل ۲).

$$L(y, f(x, a)) = |y - f(x, a)|_{\epsilon} \quad (\text{رابطه ۲})$$

با توجه به رابطه ۲، می‌توان دریافت که خطاهایی با مقادیر کمتر از ϵ در نظر گرفته نشده‌اند.



شکل ۲. تابع خطای ϵ -insensitive

به عبارت دیگر در این تابع، خطاهایی در محدوده کمتر از ϵ محتمل جریمه نمی‌شوند. از این محدوده به نام محدوده ϵ یاد می‌شود و در مسائل چند بعدی ناحیه ϵ -insensitive شکلی نظیر یک دال خواهد داشت و یا به طور کلی این محدوده بین دو ابرصفحه موازی قرار گرفته است. برای توسعه الگوریتم در ابتدا باید تخمین یک تابع خطی بررسی شود. کلیه توابع خطی به شکل کلی زیر می‌باشد

$$f(x) = \langle w, x \rangle + b, \quad w, x \in X, b \in R. \quad (\text{رابطه ۳})$$

در رابطه ۳ (w, x) مشخص کننده ضرب داخلی دو بردار w بردار وزن و x فضای ورودی در فضای هیلبرت است. هدف روند یادگیری، یافتن تابع f با کمترین خطا براساس داده‌های مستقل و با توزیع یکنواخت $(x_1, y_1), \dots, (x_m, y_m) \subseteq X \times Y$ است که الگوریتم ϵ -SVR نامیده می‌شود. برای کمینه سازی تابع خطای عمومی شده Rreg و براساس تابع خطای ϵ -insensitive پرداخته شد. Rreg را می‌توان براساس شکل باز شده Remp باز نویسی کرد.

$$R_{emp}^{\epsilon}[f] = \frac{1}{m} \sum_{i=1}^m |y_i - f(x_i)|_{\epsilon} \quad (\text{رابطه ۴})$$

$$R_{reg}[f] = \frac{1}{2} |w_i|^2 + CR_{emp}^{\epsilon}[f] \quad (\text{رابطه ۵})$$

که در این رابطه، R_{emp}^{ϵ} خطای آموزش را در قالب تابع خطای ϵ -insensitive محاسبه می‌نماید و ضریب c ثابتی است که به نوعی مقدار $|w_i|^2$ را با توجه به پیچیدگی تابع تعیین می‌کند. کمینه سازی رابطه ۴ نشان دهنده ایده اصلی تئوری یادگیری آماری است که

بیان می‌کند برای دستیابی به کمترین خطا واقعی، کنترل پیچیدگی مدل و همچنین خطای مربوط به داده‌های آموزش ضروری است. عمومیت پذیری با استفاده از داده‌های محدود امکان پذیر است اگر و تنها اگر تابع تخمین ظرفیت محدودی داشته باشد. مسئله کمینه سازی خطای هموار شده پس تبدیل به یک تابع کمینه ساز و ساخت عبارت لاگرانژی مورد نظر به یک مسئله بهینه‌سازی مقید محذب تبدیل می‌گردد که با میل دادن مشتقات جزئی آن نسبت به متغیرهای تصمیم قابل حل است. پس از حل مسئله بهینه‌سازی بالا مقادیر W و f به صورت زیر بدست می‌آیند.

$$w = \sum_{i=1}^m (\alpha_i^* - \alpha_i) x_i \quad (\text{رابطه ۶})$$

$$f(x) = \sum_{i=1}^m (\alpha_i^* - \alpha_i) \langle x_i, x \rangle + b \quad (\text{رابطه ۷})$$

قبل از تقسیم خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا پیش‌بینی کند داده‌ها به وسیله تابع فی به فضای با ابعاد خیلی بالاتر منتقل می‌شوند. برای حل مسئله با ابعاد خیلی بالا از قضیه دوگانگی لاگرانژ برای تبدیل مسئله کمینه‌سازی مورد نظر به فرم دوگانگی آن استفاده می‌گردد که در آن به جای تابع پیچیده فی که به فضایی با ابعاد بالا می‌رود. تابع سادتری به نام هسته (کرنل) که ضرب برداری تابع فی است ظاهر می‌شود. از توابع هسته‌ای مختلفی از جمله هسته‌های خطی، شعاعی، چند جمله‌ای و سیگموئید می‌توان استفاده نمود. بنابراین کافی است در مسائل غیر خطی از کرنل مقادیر ورودی به جای خود تابع استفاده شود. با توجه به تئوری توضیح داده شده دقت در تعیین پارامترهای هموارساز و مقدار پارامترهای موجود در تابع کرنل ست، در کاهش خطای مسئله تاثیر بسیار زیادی دارد (زکریا و شبیری، ۲۰۱۲).

۲-۲-۵. مدل جنگل تصادفی (RF)

جنگل تصادفی یکی از روش‌های ناپارامتریک مناسب برای مدل سازی داده‌های پیوسته و گسسته براساس روش درخت تصمیم می‌باشد. از مشکلات این روش نوسانات بالای نتایج هر درخت است. به منظور کاهش این نوسانات و کاهش واریانس برآورد، روش جنگل تصادفی پیشنهاد شده است. جنگل تصادفی با استفاده از مجموعه‌ای از درخت‌ها با در نظر گرفتن n داده مشاهده‌ای مستقل ساخته می‌شود (پورقاسمی و کلر، ۲۰۱۶).

$$(Y, X), i=1, \dots, n$$

این روش ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه بوت استرپ از داده‌ها شرکت دارند و در ساخت هر درخت به طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند. با استفاده از روش بوت استرپ به تعداد زیاد (به عنوان مثال ۲۰۰۰ مرتبه) نمونه‌هایی n تایی از مجموعه داده‌های مشاهداتی اولیه، نمونه برداری همراه با جای گذاری می‌شوند. در طی فرآیند نمونه گیری حدود یک سوم از داده‌ها نمونه‌گیری نمی‌شوند و به عنوان نمونه خارج از کیسه ۱ (از این داده‌ها برای تعیین متغیرهای مهم و همچنین برآورد نااریب خطا استفاده می‌شود) در نظر گرفته می‌شوند. سپس بر روی هر نمونه بوت استرپ یک درخت گسترش داده می‌شود. در طی فرآیند ساخت درخت در هر شاخه، از بین تمام M متغیر مستقل به صورت تصادفی m متغیر برای تقسیم شدن انتخاب می‌شود. برای حالت رگرسیونی نسبت m/M برابر با یک سوم است و برای کلاسه بندی برابر با $m = \sqrt{M}$ پیشنهاد شده است. پس از ساخت تمام درخت‌ها، داده‌های آزمون به درخت معرفی شده و به تعداد درخت‌ها برای بردار ورودی یک خروجی به دست می‌آید. با میانگین گیری این خروجی‌ها،

¹ Sample out of the bag

خروجی نهایی مدل و با در نظر گرفتن توزیع تجربی خروجی‌ها مقادیر صدک‌ها و دامنه عدم قطعیت محاسبه می‌شود. روش درخت رگرسیون جنگل تصادفی به ویژه هنگامی که تعداد مشاهدات در مقایسه با تعداد پیش‌بینی‌کننده‌ها نسبتاً کم باشد یک روش پیش‌بینی کارآمد است (پورقاسمی و کلر، ۲۰۱۶). در این پژوهش محاسبات مدل جنگل تصادفی در محیط نرم افزار R انجام شد. متغیر اندازه‌گیر (که نشان دهنده تعداد برگ‌ها در هر شاخه است) با آزمون و خطا تعیین شد.

۲-۲-۶. ارزیابی مدل‌ها

به منظور ارزیابی نتایج جنگل تصادفی و ماشین بردار پشتیبان از نمایه‌های آماری ضریب تبیین^۱ (R^2) رابطه ۸، ریشه میانگین مربعات خطا^۲ (RMSE) رابطه ۹ و ضریب کارایی ناش و ساتکلیف^۳ (CE) رابطه ۱۰ استفاده گردید.

$$R^2 = \frac{\sum_{i=1}^n (q_{O(t)} - \overline{q_{O(t)}})(q_{P(t)} - \overline{q_{P(t)}})}{\left(\sum_{i=1}^n (q_{O(t)} - \overline{q_{O(t)}})^2\right)^{0.5} \left(\sum_{i=1}^n (q_{P(t)} - \overline{q_{P(t)}})^2\right)^{0.5}} \quad (\text{رابطه ۸})$$

$$RMSE = \frac{1}{n} \sum_{i=1}^n (q_{P(t)} - q_{O(t)})^2 \quad (\text{رابطه ۹})$$

$$CE = 1 - \frac{\sum_{i=1}^n (q_{O(t)} - q_{P(t)})}{\sum_{i=1}^n (q_{O(t)} - \overline{q_{O(t)}})} \quad (\text{رابطه ۱۰})$$

که در آن $q_{O(t)}$ دبی مشاهده شده با دوره بازگشت معین، $q_{P(t)}$ دبی پیش‌بینی شده با دوره بازگشت معین، n تعداد داده‌های مشاهداتی، $\overline{q_{O(t)}}$ میانگین دبی مشاهده شده با دوره بازگشت معین و $\overline{q_{P(t)}}$ میانگین دبی پیش‌بینی شده با دوره بازگشت معین است.

۳. یافته‌های پژوهش

بعد از مشخص شدن ایستگاه‌های منتخب، پارامترهای فیزیوگرافی، اقلیمی، زمین‌شناسی و کاربری اراضی حوزه محاسبه گردید (جدول ۱). به منظور تعیین دبی‌های حداکثر لحظه‌ای با دوره بازگشت‌های مختلف، به دبی‌های حداکثر لحظه در سال‌های مختلف انواع توزیع‌های آماری برازش داده شد. جدول ۲ نتایج انتخاب بهترین تابع توزیع را نشان می‌دهد. بر این اساس توزیع لوگ پیرسون نوع سوم با کمترین میزان خطا و بیشترین تعداد رتبه اول به عنوان مناسب‌ترین تابع برازش برای دبی‌های حداکثر لحظه‌ای انتخاب شد. همچنین نتایج جدول ۲ نشان می‌دهد که توزیع‌های آماری لوگ نرمال سه پارامتری، لوگ نرمال دو پارامتری، پیرسون نوع سوم، ویبول، گامبل و نرمال به ترتیب با مجموع نمرات ۱۷۲، ۲۴۴، ۲۵۸، ۲۶۵، ۲۷۴ و ۲۹۵ در رتبه‌های بعدی انتخاب بهترین توزیع آماری قرار گرفتند. با توجه به نتایج حاصل از

¹Determination Coefficient

²Root Mean Square Error

³Nash-Sutcliffe Coefficient

این تحقیق و انتخاب توزیع لوگ پیرسون نوع سوم به‌عنوان بهترین توزیع آماری می‌توان بیان داشت دبی‌های حداکثر لحظه‌ای پدیده‌ای حدی می‌باشند، به همین دلیل توزیع لوگ پیرسون تیپ سوم به عنوان توزیع مناسب در این پژوهش انتخاب گردیده است.

جدول ۱. خصوصیات آماری متغیرهای اندازه‌گیری شده حوزه آبخیز دریاچه نمک

متغیر	حداکثر	حداقل	میانگین	ضریب تغییرات
مساحت (کیلومتر مربع)	۱۷۴۱۳/۱۴	۲۵	۱۳۳۶/۴۶۶	۲/۳۴
محیط (کیلومتر)	۸۳۳/۸۴	۲۴/۸۲	۱۵۵/۱۰	۱/۰۷
شیب متوسط وزنی (درصد)	۶۰	۱/۶	۱۸/۱۵	۰/۷۷
حداکثر ارتفاع از سطح دریا (متر)	۴۳۷۵	۱۷۰۰	۳۱۰۸/۱۱	۰/۰۲
حداقل ارتفاع از سطح دریا (متر)	۳۳۹۷	۹۷۱/۱۵	۲۵۴۲/۹۰	۲/۲۲
ارتفاع متوسط از سطح دریا (متر)	۲۸۵۹	۱۱۳۳/۱۷	۲۱۱۹/۶۲	۰/۱۸
شیب آبراهه اصلی (درصد)	۳۶۲/۰۸	۷/۵	۵۱/۹۱	۱/۲۷
تراکم زهکشی (کیلومتر بر کیلومتر مربع)	۱۴/۲۲	۰/۰۱	۱/۱۲	۲/۱۴
طول آبراهه اصلی (کیلومتر)	۳۷	۰/۲۱	۳/۷۴	۱/۶۹
حداکثر بارندگی ۲۴ ساعته (میلی متر)	۷۱	۰/۵	۴۱۴/۱۴	۰/۳۰
متوسط درجه حرارت سالیانه (میلی متر)	۱۱/۷۲	۴/۸۶	۷/۷۲	۰/۲۲
متوسط تبخیر پتانسیل سالانه (میلی متر)	۱۸۸۱/۰۵	۷۴۱/۱۳	۱۳۶۸/۱۵	۰/۲۳
درصد سازندهای نفوذ پذیر (درصد)	۷۹/۷۲	۰/۰۵	۲۳/۱۸	۱/۰۲
درصد اراضی مرتعی (درصد)	۹۶/۳۹	۲۲	۶۸/۰۵	۰/۳۱
درصد اراضی کشاورزی آبی (درصد)	۵۳/۴۱	۰	۸/۵۶	۱/۳۲
درصد اراضی کشت دیم (درصد)	۴۴	۰	۱۲/۲۵	۱/۰۱
درصد اراضی باغی (درصد)	۲۱/۶۵	۰	۳/۹۳	۱/۴۶
درصد اراضی سنگی (درصد)	۵۴/۸۶	۰/۲	۵/۶۰	۱/۷۳

جدول ۲. انتخاب مناسب‌ترین توزیع آماری منطقه‌ای در حوزه آبخیز دریاچه نمک

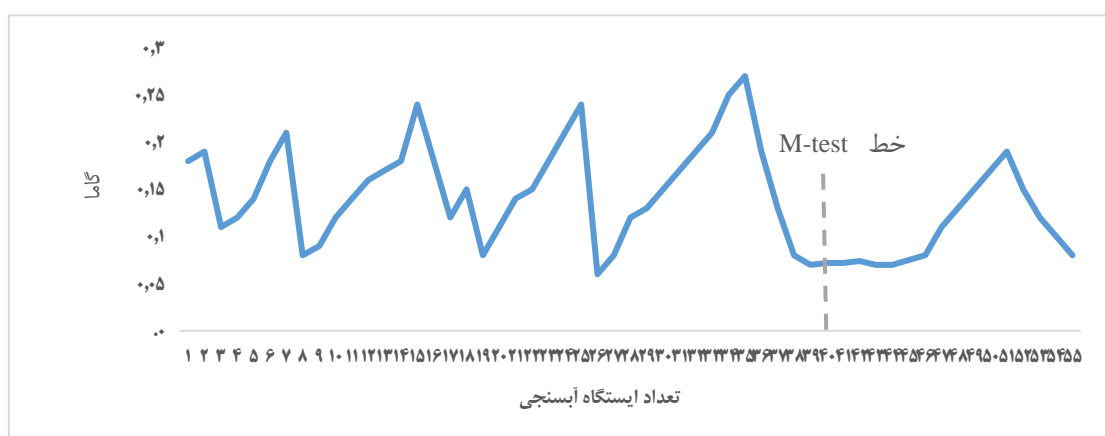
توزیع آماری	لوگ پیرسون نوع سوم	لوگ نرمال سه پارامتری	لوگ نرمال دو پارامتری	پیرسون نوع سوم	ویبول	گامبل	نرمال
جمع نمرات	۱۴۵	۱۷۲	۲۴۴	۲۵۸	۲۶۵	۲۷۴	۲۹۵
تعداد رتبه اول	۱۹	۱۰	۸	۷	۴	۳	۳

به منظور انتخاب بهترین ترکیب ورودی در مدل‌های جنگل تصادفی و ماشین بردار پشتیبان از آزمون گاما استفاده شد که نتایج آن در جدول ۳ آمده است. بر این اساس متغیرهای مساحت، تراکم زهکشی، حداکثر بارش ۲۴ ساعته و محیط به‌عنوان متغیرهای تاثیرگذار در مدل‌های مورد نظر به کار گرفته می‌شوند. همچنین براساس نتایج آزمون M تعداد ایستگاه‌های آب‌سنجی مورد نیاز برای آموزش و آزمون مدل تعیین شد. برای این منظور با استفاده از آماره گاما (Γ) می‌توان داده‌های مورد نیاز را برای ایجاد خط مجانب پایدار (M) را مشخص کرد. همان‌گونه که از شکل ۲ مشاهده می‌گردد، هم‌سانی خط مجانب آماره گاما با مقدار ۰/۰۷ در نزدیکی نقاط ۳۹ تا ۴۴ مشخص شده است. به همین دلیل می‌توان داده ۴۰ را به‌عنوان نقطه مرزی برای تعیین تعداد داده‌های آموزش و آزمون مدل‌ها در نظر گرفت. بر این اساس تعداد ۴۰ ایستگاه به‌عنوان داده‌های آموزش و تعداد ۱۵ ایستگاه به‌عنوان داده‌های آزمون مدل‌ها در نظر گرفته شدند. در تحقیق حاضر به منظور انجام مدل‌سازی با ماشین بردار پشتیبان از روش رگرسیونی سه تابع کرنل خطی، چندجمله‌ای و شعاعی

استفاده شد. به همین منظور ابتدا مقادیر بهینه پارامترهای مورد استفاده در هر کدام از توابع در دوره بازگشت‌های مختلف مشخص و پس از ایجاد مدل‌ها، مناسب‌ترین تابع کرنل انتخاب شد. در انتخاب مقادیر بهینه پارامترها سعی می‌شود که مدل با استفاده از این مقادیر کم‌ترین خطا در مرحله آزمون را داشته باشد. جدول ۴ نتایج حاصل از انتخاب مقادیر بهینه پارامترهای مورد نیاز برای توابع خطی، چندجمله‌ای و شعاعی را نشان می‌دهد. با توجه به نتایج جدول ۵ مقدار بهینه c ۵۲ الی ۷۲، ε ۰/۳۵ الی ۰/۵۵ و مقدار Γ ۱/۲ الی ۳/۲ می‌باشد. جدول ۵ نمایه‌های آماری حاصل از اجرای مدل‌سازی ماشین بردار پشتیبان را نشان می‌دهد.

جدول ۳. انتخاب متغیرهای تاثیرگذار بر دبی حداکثر سیلاب با استفاده از آزمون گاما

متغیر	گاما (Γ)	متغیر	گاما (Γ)
مساحت	۰/۲۲۲۵	شیب متوسط وزنی	۰/۱۷۸۲
تراکم زهکشی	۰/۲۱۸۵	ارتفاع میانگین	۰/۱۷۷۳
حداکثر بارندگی ۲۴ ساعته	۰/۲۰۷۸	حداکثر بارندگی ۲۴ ساعته	۰/۱۷۷۰
محیط	۰/۱۹۴۵	طول آبراهه اصلی	۰/۱۷۶۷
درصد سازند نفوذپذیر	۰/۱۹۲۴	شیب آبراهه اصلی	۰/۱۷۵۲
درصد اراضی کشاورزی	۰/۱۹۱۸	متوسط درجه حرارت سالانه	۰/۱۷۰۶
درصد اراضی مرتعی	۰/۱۸۸۸	درصد اراضی سنگی	۰/۱۶۹۷
درصد اراضی باغی	۰/۱۸۶۵	درصد اراضی دیم	۰/۱۶۴۵
متوسط بارندگی سالانه	۰/۱۸۴۲	ارتفاع حداکثر	۰/۱۶۳۵
متوسط تبخیر سالانه	۰/۱۸۱۶	ارتفاع حداقل	۰/۱۵۹۹



شکل ۲. نمودار M-test مقادیر گاما

با توجه به نتایج جدول ۵ حاصل از اجرای مدل‌ها در دوره بازگشت‌های ۲، ۵، ۱۰، ۲۵، ۵۰ و ۱۰۰ سال تابع پایه شعاعی با بیش‌ترین مقدار کارایی و کم‌ترین میزان خطا به‌عنوان بهترین تابع انتخاب شدند که با نتایج زکریا و شبیری (۲۰۱۲) هم‌خوانی دارد. ارزیابی عملکرد مدل ماشین بردار پشتیبان بر مبنای آماره ضریب کارایی ناش-ساتکلیف نشان می‌دهد که دقت مدل‌سازی روند سینوسی دارد به طوری که تا دوره بازگشت ده سال رو به افزایش می‌یابد و درباره دقت مدل‌سازی در دبی با دوره بازگشت ۵۰ و ۱۰۰ ساله افزایش می‌یابد است. همچنین نتایج جدول ۵ و شکل نشان ۳ داد که دقیق‌ترین مدل ماشین بردار پشتیبان در دوره بازگشت ۵ سال با میزان ضریب

کارایی ۰/۹۱ میزان خطای ۱۱/۱۵ متر مکعب بر ثانیه و میزان ضریب تبیین ۰/۹۱ و کمترین کارایی مدل مربوط به دوره بازگشت ۲۵ سال با میزان ضریب کارایی ۰/۷۴ میزان خطای ۲۸/۰۳ متر مکعب بر ثانیه و میزان ضریب تبیین ۷۶ در مرحله آزمون در مرحله آزمون می‌باشد. در مدل سیستم پشتیبان بردار به جز در دبی با دوره بازگشت ۲۵ ساله، با افزایش دوره بازگشت کارایی مدل افزایش پیدا می‌کند.

جدول ۴. مقادیر بهینه پارامترهای مورد استفاده در مدل ماشین بردار پشتیبان

دبی	مقادیر بهینه پارامترها		
	Γ	ϵ	c
Q۲	۳/۲	۰/۴۵	۷۲
Q۵	۳	۰/۵۰	۷۲
Q۱۰	۳/۱	۰/۵۰	۶۸
Q۲۵	۱/۲	۰/۳۵	۶۲
Q۵۰	۱/۲	۰/۳۵	۵۵
Q۱۰۰	۱/۲	۰/۵۵	۵۲

جدول ۵. نمایه‌های آماری حاصل از مدل‌سازی با استفاده از ماشین پشتیبان بردار

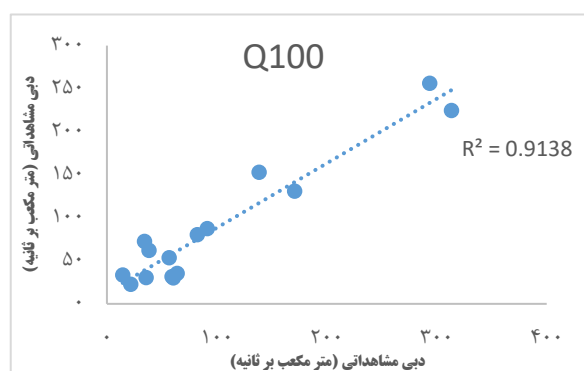
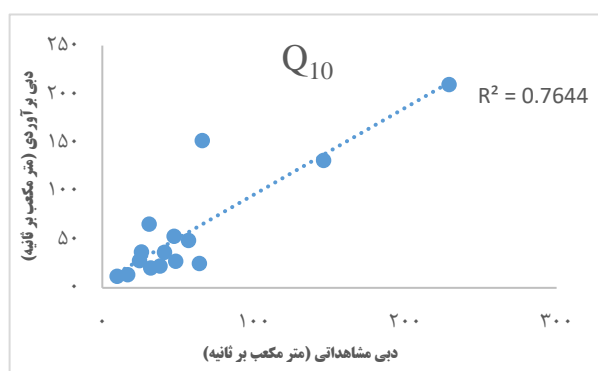
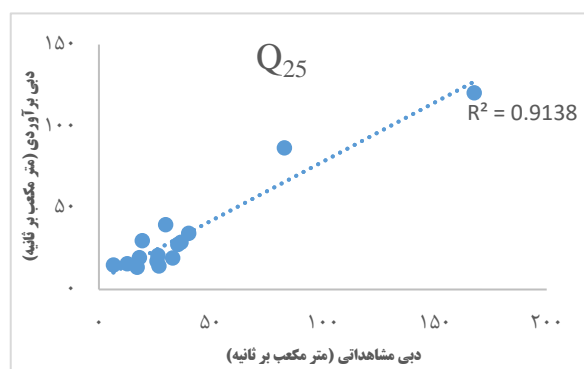
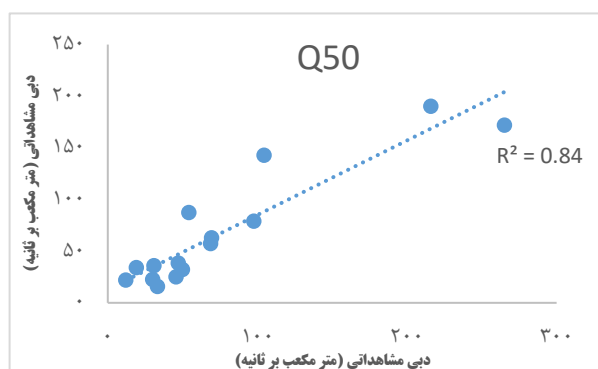
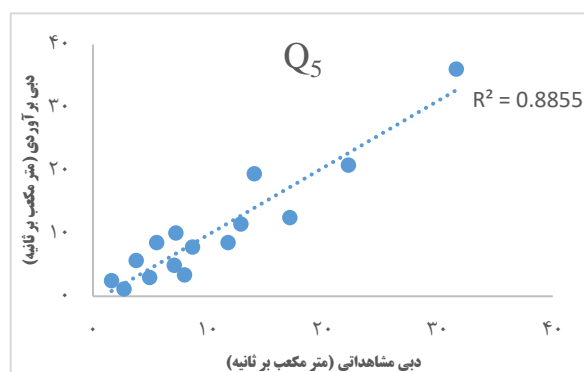
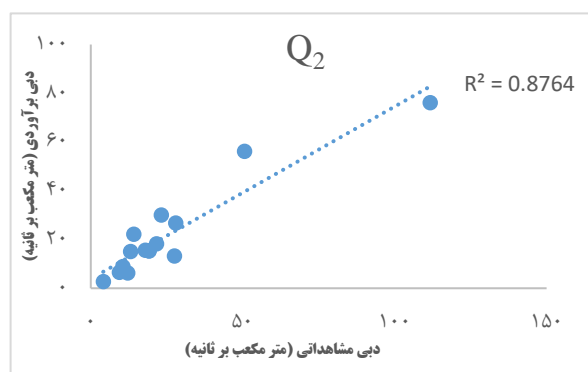
دبی	نوع تابع	شاخص آماری			
		RMSE		R ²	
		آموزش	آزمون	آموزش	آزمون
Q۲	شعاعی	۰/۹۴	۰/۸۸	۱/۴۶	۳/۰۵
Q۵	شعاعی	۰/۹۳	۰/۸۷	۴/۴۲	۹/۴۷
Q۱۰	شعاعی	۰/۹۳	۰/۹۱	۸/۹۱	۱۱/۱۵
Q۲۵	شعاعی	۰/۸۷	۰/۷۶	۱۴/۷۶	۲۸/۰۳
Q۵۰	شعاعی	۰/۸۶	۰/۹۱	۱۶/۸۴	۳۰/۴۸
Q۱۰۰	شعاعی	۰/۹۴	۰/۷۶	۱۶/۹۱	۳۲/۱۱

به منظور انجام مدل‌سازی جنگل تصادفی برای تعیین گره‌ها از روش سعی و خطا استفاده شد. جدول ۶ نمایه‌های آماری حاصل از اجرای جنگل تصادفی را نشان می‌دهد. با توجه به نتایج حاصل از اجرای مدل‌ها در دوره بازگشت‌های ۲، ۵، ۱۰، ۲۵، ۵۰ هر چه دوره بازگشت بیشتر می‌شود تعداد گره‌ها در مدل‌سازی افزایش می‌یابد.

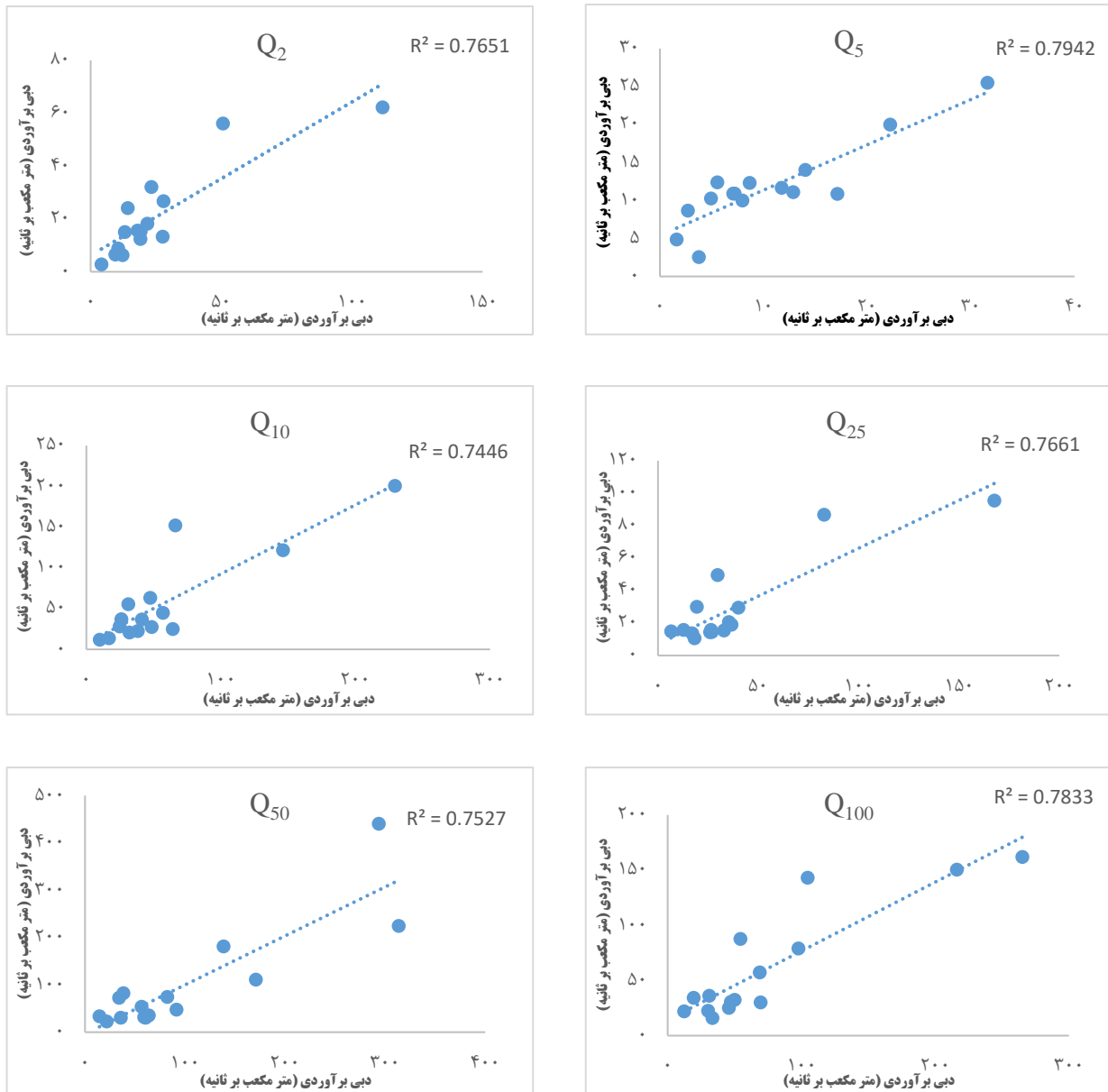
جدول ۶. نمایه‌های آماری حاصل از مدل‌سازی با جنگل تصادفی

دبی	تعداد گره	شاخص آماری			
		RMSE		R ²	
		آموزش	آزمون	آموزش	آزمون
Q۲	۱۸	۰/۸۵	۰/۷۹	۲/۳۲	۴/۱۱
Q۵	۲۳	۰/۸۳	۰/۷۶	۷/۲۱	۱۴/۱۸
Q۱۰	۳۲	۰/۸۷	۰/۷۶	۱۲/۷۸	۲۱/۹۱
Q۲۵	۳۸	۰/۸۱	۰/۷۴	۱۸/۱۷	۲۸/۶۵
Q۵۰	۴۶	۰/۸۳	۰/۷۸	۲۲/۸۷	۳۷/۶۴
Q۱۰۰	۵۲	۰/۸۷	۰/۷۵	۲۷/۴۶	۵۳/۵۷

در فرآیند مدل سازی طبق نظر (کیشی و همکاران، ۲۰۱۳) مدلی قابل پذیرش است که ضریب کارایی بالاتری در مرحله آزمون داشته باشد. پس از پایان مراحل مدل سازی با استفاده از جنگل تصادفی و ماشین بردار پشتیبان، نتایج حاصله نشان داد که در دوره بازگشت های ۲، ۵، ۱۰، ۲۵، ۵۰ و ۱۰۰ سال ماشین بردار پشتیبان نسبت به مدل جنگل تصادفی از عملکرد دقیق تری برخوردار است که با نتایج زکریا و شبیری (۲۰۱۲) در برتری مدل شده ماشین بردار پشتیبان نسبت به سایر مدل ها مطابقت دارد.



شکل ۳. نمودار مقایسه دبی مشاهداتی با دبی برآوردی با دوره بازگشت مختلف در مدل سیستم پشتیبان بردار



شکل ۴. نمودار مقایسه دبی مشاهداتی با دبی برآوردی با دوره بازگشت مختلف در مدل جنگل تصادفی

۴. بحث و نتیجه گیری

این مطالعه در حوزه‌های بالادست ۵۵ ایستگاه آسنجی در حوزه آبخیز دریاچه نمک با طول دوره آماری مشترک ۲۰ ساله بین سال‌های انجام پذیرفت. متغیر وابسته در این پژوهش دبی‌های حداکثر سیلاب در دوره بازگشت‌های مختلف و متغیرهای مستقل شامل متغیرهای فیزیوگرافی، زمین‌شناسی، کاربری اراضی و اقلیمی می‌باشد. توزیع لوگ پیرسون نوع سوم با کم‌ترین میزان خطا به عنوان بهترین توزیع آماری به منظور برازش دبی‌های حداکثر لحظه‌ای انتخاب شد. توزیع لوگ پیرسون نوع سوم کاربرد زیادی در هیدرولوژی و هواشناسی

داشته و به صورت استاندارد برای محاسبه دبی سیلاب‌ها به کار می‌رود. با توجه به نتایج حاصل از این تحقیق و انتخاب توزیع لوگ پیرسون نوع سوم به عنوان بهترین توزیع آماری می‌توان بیان داشت که دبی‌های حداکثر لحظه‌ای پدیده‌ای حدی می‌باشند، به همین دلیل توزیع لوگ پیرسون تیپ سوم به عنوان توزیع مناسب در این پژوهش انتخاب گردیده است. نتایج اجرای آزمون گاما نشان داد متغیر مساحت با بیش‌ترین مقدار آماره گاما (T) به عنوان تاثیرگذارترین متغیر ورودی انتخاب و چهار متغیرهای مساحت، تراکم زهکشی، حداکثر بارندگی ۲۴ ساعته و محیط به عنوان چهار عوامل مهم و تاثیرگذار بر سیل مشخص و در دو مدل جنگل تصادفی و ماشین بردار پشتیبان به کار گرفته شدند. بررسی نتایج پژوهش اچکدالی^۱ و همکاران (۲۰۱۸) و هالیجریجیس^۲ و همکاران (۲۰۱۷) نشان می‌دهد عمدتاً چهار متغیر انتخابی در مطالعات مرتبط با آنالیز سیل به کار گرفته شده‌اند. پس از اجرای مدل‌ها، نتایج نشان داد به طور کلی در دوره بازگشت‌های محاسبه شده ماشین بردار پشتیبان از عملکرد دقیق‌تری برخوردار است. بنابراین می‌توان اینگونه نتیجه گرفت که روش مدل سازی ماشین بردار پشتیبان به عنوان مناسب‌ترین روش مدل سازی داده‌های دبی حداکثر لحظه‌ای نسبت به روش جنگل تصادفی در حوزه آبخیز دریاچه نمک می‌باشد.

References

- Aziz, K., Haque, M. M., Rahman, A., Shamseldin, A. Y., & Shoaib, M. (2017). Flood estimation in ungauged catchments: application of artificial intelligence-based methods for Eastern Australia. *Stochastic Environmental Research and Risk Assessment*, 31(6), 1499-1514.
- Aziz, K., Rahman, A., Shamseldin, A.Y., & Shoaib, M. (2014). Co-Active Neuro Fuzzy Inference System for Regional Flood Estimation in Australia. *Journal of Hydrology and Environment Research*, 1(1): 11-20.
- Besalatpour, A., Haj Abbasi, M. A., & Ayoubi S. A. (2013). Using gamma test to select the optimal inputs in soil shear strength modeling using artificial neural networks. *Journal of Research Water and Soil Protection*, 20 (1): 97-114. (In Persian).
- Chua, K.W., Wu, C.L., & Li, Y.S. (2005). Comparison of several flood forecasting models in Yangtze River. *Journal of Hydrologic Engineering*, 10, 485-491.
- Du, J., Fang, J., Xu, W., & Shi, P. (2013). Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province, China. *Stochastic Environmental Research and Risk Assessment*, 27, 377-387.
- Durocher, M., Chebana, F., & Ouarda, T. B. M. J. (2015). A Nonlinear Approach to Regional Flood Frequency Analysis Using Projection Pursuit Regression. *Journal of Hydrometeorology*, 16(4), 1561-1574.
- Echogdali, F. Z., Boutaleb, S., Elmouden, A., & Ouchchen, M. (2018). Assessing Flood Hazard at River Basin Scale: Comparison between HECRAS-WMS and Flood Hazard Index (FHI) Methods Applied to El Maleh Basin, Morocco. *Journal of Water Resource and Protection*, 10(9), 957-977.
- Hailegeorgis, T. T., & Alfredsen, K. (2017). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology*, 27, 377-387.
- Kisi, O., Shiri, J., & Tombul M. (2013). Modeling Rainfall-Runoff Process Using Soft Computing Techniques. *Computers & Geosciences*, 51, 108-117.
- Kornejady, A., Heidary, K., Sarparast, M., Khosravi, G., & Mombeini, M. (2014). Performance Assessment of Two "LNRF" and "AHP-Area Density" Models in landslide Susceptibility Zonation. *American Journal of Biomedical and Life Sciences*, 4, 169-176.
- Negaresh, H., Ajdari Moghaddam, M., & Armesh M. (2013). Application of artificial neural network in simulation and flood prediction in Sarbaz Watershed. *Journal of Geography and Development*, 31: 15-28.
- Nikoo, M., Ramezani, F., Hadzima-Nyarko, M., Nyarko, E.K., & Nikoo, M. (2016). Flood-routing modeling with neural network optimized by social-based algorithm. *Natural Hazards*, 128-152

¹ Echogdali

² Hailegeorgis

- Panahi, A., & Alijani, B. (2013). Predicting Dubai Peak Flood Using Artificial Neural Network Modeling and Multivariate Regression (Madarsoo Watershed of Golestan Province). *Geography (International Quarterly of the Geographical Society of Iran)*, 11 (38): 113-132. (In Persian).
- Pierdicca, N., Pulvirenti, L., Chini, M., Guerriero, L., & Ferrazzoli, P. (2010). A fuzzy-logic-based approach for flood detection from cosmo- skymed data Dept. Information, Electronic and Telecommunications Engineering, Sapienza University of Rome.
- Pourghasemi, H.R., & Kerle, N., (2016). Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environmental Earth Sciences*, 75, 1–17.
- Rahman, A., Charron, C., Ouarda, T. B., & Chebana, F. (2018). Development of regional flood frequency analysis techniques using generalized additive models for Australia. *Stochastic Environmental Research and Risk Assessment*, 32(1), 123-139.
- Sahoo, G.B., Schladow, S.G., & Reuter, J.E. (2009). Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. *Journal of Hydrology*, 378, 325–342.
- Sivapalan, M., & Blöschl G. (1997). Transformation of point rainfall to areal rainfall: intensity-duration-frequency curves. *Journal of Hydrology*, 98(240): 150-167.
- Vafkhah, M. (1998). Estimation of the abundance of regional currents of the minimum seasonal rivers (study in the arid regions of central Iran), Bachelor's degree in watershed management. *Tarbiat Modares University*, 145 p. (In Persian).
- Xiong, F., Guo, S., Chen, L., Yin, J., & Liu, P. (2018). Flood Frequency Analysis Using Halphen Distribution and Maximum Entropy. *Journal of Hydrologic Engineering*, 23(5),
- Youssef, A.M., Pradhan, B., & Hassan, A.M. (2011). Flash flood risk estimation along the St. Katherine Road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. *Environmental Earth Sciences*, 62, 611–623.
- Zakaria, A.Z., & Shabri A. (2012). Streamflow Forecasting at Ungaged Sites using Support Vector Machines. *Applied Mathematical Sciences*, 60(6), 3003-3014.

