

بررسی تأثیر پارامترهای فیزیوگرافی و اقلیمی حوضه در شبیه‌سازی جریان فصلی رودخانه

زهرا نعیمی کلورزی^۱، خلیل قربانی^{۲*}، میثم سالاری جزی^۳ و امیراحمد دهقانی^۲

۱. دانش‌آموخته کارشناسی ارشد مهندسی منابع آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان

۲. دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان

۳. استادیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان

(تاریخ دریافت ۱۳۹۵/۰۹/۰۱؛ تاریخ تصویب ۱۳۹۵/۱۰/۰۱)

چکیده

خصوصیات فیزیوگرافی و شرایط اقلیمی در حوضه‌های آبریز از عوامل مهم دخیل در رژیم جریان رودخانه هستند که درک روابط بین این عوامل با جریان رودخانه در یک حوضه موجب می‌شود بتوان از این روابط در زیرحوضه‌های فاقد آمار برای پیش‌بینی جریان رودخانه استفاده کرد. در این مطالعه، روابط بین پارامترهای فیزیوگرافی و اقلیمی زیرحوضه‌های آبریز استان گلستان با جریان رودخانه با کاربرد مدل درختی M5، مدل نزدیک‌ترین K-همسایگی (KNN) و رگرسیون چندمتغیره خطی (MLR) بررسی شد. داده‌های روزانه ۲۸ ساله (۱۳۶۰-۱۳۹۰) بارش، دما و دبی ایستگاه‌های هیدرومتری و هواشناسی ۳۹ زیرحوضه آبریز برای استخراج سری‌های فصلی به‌منظور مدل‌سازی استفاده شد. متوسط مقادیر R و RMSE در فصول مختلف برای مدل M5 به ترتیب برابر ۰/۷۶۸ و ۰/۸۰۰، برای مدل KNN به ترتیب برابر ۰/۸۸۵ و ۰/۵۰۱ و برای مدل MLR به ترتیب برابر ۰/۶۹۳ و ۱/۲۰۵ است که نشان‌دهنده برتری مدل KNN است. همچنین بر اساس مقادیر R و RMSE دقت نتایج مدل‌سازی در فصل‌های مختلف به ترتیب به صورت زمستان، پاییز، بهار و تابستان بوده است. به بیان دیگر نتایج پیش‌بینی جریان رودخانه در فصول تر از فصول خشک دقت بیشتری داشته است. همچنین بررسی مقادیر MBE نشان داد مدل KNN در فصل‌های بهار و زمستان به کم‌برآوردی و در تابستان و پاییز به بیش‌برآوردی منجر می‌شود. مدل M5 صرفاً در فصل بهار به کم‌برآوردی و در سایر فصول به بیش‌برآوردی و مدل MLR نیز در زمستان به کم‌برآوردی و در سایر فصول به بیش‌برآوردی از مقدار مشاهداتی منجر می‌شود.

کلیدواژگان: جریان رودخانه، حوضه فاقد آمار، مدل درخت تصمیم M5، مدل نزدیک‌ترین K-همسایگی، مدل رگرسیون چندمتغیره خطی.

مقدمه

برآورد آبدی رودخانه‌ها که پدیده‌ای تصادفی و تأثیرپذیر از برخی عوامل هواشناسی و هیدرولوژیکی نظیر دما، بارش و دبی رودخانه است، تأثیر مهمی در مدیریت منابع آب ایفا می‌کند. با پیش‌بینی آبدی رودخانه علاوه بر مدیریت بهره‌برداری منابع آب به‌منظور تأمین نیاز، می‌توان حوادث طبیعی نظیر سیل و خشکسالی را نیز پیش‌بینی و کنترل کرد. همچنین می‌توان از نتایج اجرای یک مدل شبیه‌سازی به‌منظور بررسی صحت داده‌ها، اصلاح و یا تکمیل داده‌ها بهره گرفت. تا کنون روابط و الگوهای گوناگون و پیچیده‌ای برای شبیه‌سازی آبدی رودخانه‌ها نظیر انواع الگوهای مفهومی بارش-رواناب، الگوهای سری زمانی و الگوهای ترکیبی (هیبرید) ارائه شده است، اما روابط ارائه‌شده به‌علت نشناختن دقیق و پیچیدگی عوامل مؤثر در آبدی رودخانه‌ها، در بسیاری از موارد با مقادیر مشاهده‌شده تطابق نداشت و در مواقعی نیز میزان مقادیر محاسبه‌شده از روابط گوناگون تفاوت معناداری با یکدیگر داشتند. مطابق تقسیم‌بندی گویندراجو [۱] مدل‌هایی که امروزه در هیدرولوژی استفاده می‌شوند، شامل مدل‌های ریاضی-فیزیکی، مدل‌های ژئومورفولوژیکی و مدل‌های تجربی هستند. دسته اول مدل‌ها براساس خصوصیات فیزیکی سیستم هستند که به‌صورت معادلات دیفرانسیل بیان می‌شوند، اما دسته دوم بر مبنای خصوصیات ژئومورفولوژیکی سیستم هیدرولوژیکی مد نظر بیان می‌شوند. این نوع مدل‌ها توانایی زیادی در شبیه‌سازی رفتار حوضه آبریز و شبکه‌های آبیاری و زهکشی دارند. در سال‌های اخیر کاربرد روش‌های داده‌کاوی از جمله منطق فازی و شبکه عصبی، الگوریتم KNN، ماشین بردار پشتیبان و... در زمینه‌هایی که در آنها ارتباط بین ورودی و خروجی غیرخطی است، کاربرد گسترده‌ای پیدا کرده‌اند. سلاجقه و همکارانش [۲] در پژوهشی با استفاده از مدل‌های نزدیک‌ترین همسایه و سری‌های زمانی مقادیر دبی در یک گام زمانی آینده در رودخانه کرج را پیش‌بینی کردند. در مدل‌های نزدیک‌ترین همسایه تأثیر تعداد بردارهای وضعیت و تعداد همسایه مختلف بررسی شد. در نهایت، نتایج پژوهش نشان داد مدل‌های سری زمانی عملکرد بهتری نسبت به مدل‌های نزدیک‌ترین همسایه داشته‌اند. لوهانی و همکارانش [۳] عملکرد مدل‌های ANFIS، ANN و خودهمبسته در پیش‌بینی جریان ماهانه ورودی به مخزن سدی در هندوستان را بررسی کردند. یافته‌های آنها

بیانگر برتری مدل ANFIS در مقایسه با مدل‌های ANN و خودهمبسته به‌آرامی همه الگوهای استفاده‌شده به‌عنوان ورودی مدل بود. در مطالعه دیگری با بهره‌گیری از الگوهای ضمنی شبکه عصبی مصنوعی اقدام به پیش‌بینی جریان‌های روزانه و ماهانه رودخانه‌های گوس و در ژاپن، بلکواتر و گیلا در آمریکا و رودخانه فلیوس در ترکیه پرداخته شد و مقایسه نتایج با روش‌های سری زمانی نشان داد شبکه عصبی مصنوعی در مقایسه با این روش‌ها دقت بیشتری دارد [۴]. همچنین در تحقیقی با استفاده از مدل‌های مبتنی بر منطق فازی شامل سیستم استنتاج فازی و سیستم استنتاج تطبیقی عصبی-فازی به‌منظور پیش‌بینی جریان روزانه رودخانه پرداخته شد. در این پژوهش از سه پارامتر بارندگی، دما و دبی روزانه حوضه آبریز ليقوان چای برای پیش‌بینی جریان روزانه رودخانه استفاده شد. ارزیابی نتایج پیش‌بینی‌ها نشان داد مدل ANFIS با دقت بیشتر و پراکندگی کمتر نسبت به مدل FIS دبی این رودخانه را پیش‌بینی کرده است [۵]. پرویز و همکارانش [۶] برای پیش‌بینی جریان فصلی رودخانه از روش‌های تفکیک زمانی و شبکه‌های عصبی مصنوعی^۱ استفاده کردند. آماره‌های نکویی برازش مدل‌های بررسی‌شده RMSE، MR و SE مبین عملکرد بهتر مدل مبسوط و شبکه عصبی در پیش‌بینی جریان فصلی در مقایسه با مدل پایه بودند، ولی در تفکیک مقیاس‌های زمانی کوچک‌تر مدل مبسوط توصیه می‌شود. زارع‌زاده مهریزی و بزرگ حداد [۷] در تحقیقی برای شبیه‌سازی و پیش‌بینی آبدی با استفاده از الگوریتم ترکیبی ANN-GA پرداختند. در این تحقیق مؤثر بودن یا نبودن هر یک از عوامل دما، بارش، آبدی و تأخیرهای زمانی این پارامترها در شبیه‌سازی و پیش‌بینی آبدی رودخانه بررسی شد. مقادیر ضریب همبستگی بین آبدی‌های محاسباتی و موجود برای دو مدل شبیه‌سازی و پیش‌بینی به ترتیب برابر ۰/۸۶ و ۰/۷۹ به دست آمد که بیانگر کارایی الگوریتم ANN-GA در شبیه‌سازی و پیش‌بینی آبدی با دقت مطلوب است. سیدیان و همکارانش [۸] در مطالعه‌ای دبی جریان رودخانه با استفاده از داده‌کاوی و سری زمانی حوضه آبریز گرگان‌رود برای سال‌های آبی ۱۳۸۱-۱۳۹۰ را پیش‌بینی کردند. نتایج این مطالعه نشان داد روش SVM^۲ در هر دو مقیاس ماهانه و هفتگی دقت بیشتری نسبت

1. Artificial Neural Network
2. Support Vector Machine

از طریق رگرسیون چندمتغیره به مدل سازی تولید رواناب پرداخته شد. نتایج رگرسیون چندمتغیره برای مدل سازی تولید رواناب نشان می دهد عوامل تعیین کننده بر میزان دبی در دوره های بازگشت مختلف در حوضه آبریز رودخانه کشکان، فاکتورهای بیشترین بارش ماهانه، مساحت، زمان تمرکز، ضریب فشردگی و بیشترین ارتفاع حوضه هستند. زارع ایبانه و همکارانش [۱۳] مدل های هوشمند عصبی و تجربی در تخمین رواناب سالانه با استفاده از اطلاعات فیزیوگرافی و اقلیمی سه ایستگاه آب سنجی پل زمانخان، قلعه شاهرخ و سد زاینده رود واقع در حوضه زاینده رود را ارزیابی کردند. پارامترهای ورودی شامل متوسط بارش سالانه، متوسط دمای هوا، کمترین و بیشترین دمای هوا بود. نتایج نشان داد مدل های هوشمند عصبی دقت مناسبی در برآورد رواناب داشتند.

اسکندری نیا و همکارانش [۱۴] در تحقیقی تأثیر عامل بارش پیشین در برآورد جریان رودخانه توسط مدل سازی هوشمند بارش- رواناب را بررسی کردند. در این تحقیق به منظور پیش بینی روزانه جریان رودخانه بختیاری از شبکه پرسپترون چندلایه (MLP) استفاده و برای نیل به اهداف یادشده به بررسی تأثیر بارش پیشین (API) در فرایند بارش- رواناب پرداخته شد و این پارامتر نیز به ورودی های مدل اضافه و بهبود شایان توجهی در نتایج پیش بینی به دست آمد. بررسی آنالیز حساسیت مدل نسبت به پارامترهای ورودی نشان دهنده این مطلب بود که دبی یک روز قبل، میزان بارش هفت روز قبل و بارش در ایستگاه باران سنجی تنگ پنج پارامترهای اصلی این حوضه در پیش گویی جریان روزانه رودخانه هستند. خدمتی و همکارانش [۱۵] برای منطقه بندی و برآورد دبی سیلابی در حوضه های آبریز فاقد آمار جنوب شرق ایران با ترکیب روش شاخص سیلاب^۲ و رگرسیون چندمتغیره^۳ استان های کرمان، یزد، سیستان و بلوچستان و هرمزگان به کمک روش هایی نظیر شاخص سیلاب، تحلیل خوشه ای و رگرسیون چندمتغیره پرداختند. در این تحقیق با استفاده از ویژگی های فیزیوگرافی و گشتاورهای متعارف و خطی، گروه زیرحوضه های همگن تعیین شد. توزیع های احتمالاتی مرسوم آزمون شد و در نهایت توزیع احتمالاتی غالب منطقه توزیع لوگ پیرسون تیپ ۳ تعیین شد و براساس آن دبی اوج با دوره بازگشت های مختلف برآورد و مدل های ریاضی

به سری زمانی دارد، همچنین دقت هر دو روش در مقیاس ماهانه بیشتر از مقیاس هفتگی است. احمدی و همکارانش [۹] در تحقیقی روش های برنامه ریزی ژنتیک و ماشین بردار پشتیبان در پیش بینی جریان رودخانه باراندوزچای را مقایسه کردند. مقایسه نتایج دو مدل نشان داد گرچه دقت روش برنامه ریزی ژنتیک با ضریب همبستگی برابر ۰/۹۷۸ و ریشه میانگین مربعات خطا برابر ۱/۶۶ نسبت به روش ماشین بردار پشتیبان با ضریب همبستگی برابر ۰/۹۷۶ و ریشه میانگین مربعات خطا ۱/۸ اندکی بیشتر بود، روش SVM به مراتب ساده تر از روش GP است و این روش می تواند به عنوان روشی کاربردی برای پیش بینی جریان روزانه به کار رود. ثانی خانی و همکارانش [۱۰] در مطالعه ای برای مدل سازی جریان رودخانه باراندوزچای از روش های k- نزدیک ترین همسایه^۱ و روش های هوشمند استفاده کردند. نتایج به دست آمده بیانگر عملکرد قابل قبول روش های استفاده شده در پیش بینی مقدار جریان ماهانه بودند. با اضافه کردن ضریب فصلی جریان به الگوی ورودی مدل، عملکرد مدل های هوشمند در پیش بینی به صورت شایان توجهی افزایش یافت. فیرات و گانگر [۱۱] در تحقیقی مدل عصبی فازی را برای پیش بینی جریان رودخانه بزرگ مئاندر در ترکیه ارزیابی کردند و به این نتیجه رسیدند که مدل عصبی- فازی می تواند با موفقیت، صحت و قابلیت اطمینان زیاد برای پیش بینی جریان رودخانه به کار رود.

پیش بینی دبی جریان روزانه رودخانه شویلیکیل در ایالات متحده با استفاده از روش برنامه ریزی ژنتیک خطی توسط گوون انجام گرفت. نتایج به دست آمده با دو الگو از روش های شبکه عصبی مقایسه شد. نتایج نشان داد هر دو روش دقت قابل قبولی داشته اند، اما روش برنامه ریزی ژنتیک خطی دقت بیشتری نسبت به روش های شبکه عصبی دارد.

با بررسی و مطالعه تحقیقات گذشته نتیجه گرفته می شود که در زمینه برآورد دبی رودخانه با استفاده از پارامترهای حوضه مطالعات محدودی صورت گرفته است و یا منطقه مطالعاتی تنها یک حوضه بوده است. به طوری که نگارش و همکارانش [۱۲] برای مدل سازی تولید رواناب حوضه آبریز رودخانه کشکان براساس روش های آماری از داده های بارش و دبی در مقیاس ماهانه و ویژگی های فیزیوگرافی پنج زیرحوضه رودخانه کشکان که پراکندگی مناسبی در سطح حوضه دارند،

2. Flood Index
3. Multivariate Regression

1. K-Nearest Neighbors

هیدرومتری کل زیرحوضه‌های استان مد نظر استفاده شده است.

مواد و روش‌ها

منطقه مطالعه شده

استان گلستان از نظر موقعیت جغرافیایی بین ۳۶ درجه و ۲۴ دقیقه تا ۳۸ درجه و ۵ دقیقه عرض شمالی و ۵۱ درجه و ۵ دقیقه تا ۵۶ درجه و ۱۴ دقیقه طول شرقی از نصف‌النهار گرینویچ قرار گرفته است. مساحت استان ۲۰۴۳۷/۷۴ کیلومترمربع است که معادل ۱/۳ درصد مساحت کل کشور را دارد. این استان از شمال به کشور ترکمنستان، از شرق به استان خراسان شمالی، از جنوب به استان سمنان و از غرب به استان مازندران و دریای خزر محدود می‌شود. از نظر توپوگرافی این استان از نقاط کوهستانی و جلگه‌ای تشکیل شده است که قسمت جنوب و شرق را مناطق کوهستانی دربر گرفته و به‌طور موازی و با جهت تقریباً شرقی- غربی و ادامه رشته‌کوه‌های البرز هستند. رودخانه‌های اصلی استان از جنوب به شمال که به دشت گرگان وارد می‌شوند شامل رودهای قره‌سو، گرگان‌رود و اترک هستند.

داده‌های مورد نیاز

داده‌های استفاده شده در این پژوهش شامل اطلاعات دبی، بارش روزانه و دمای روزانه ایستگاه‌های هیدرومتری و هواشناسی است که طی یک دوره آماری بلندمدت ۱۳۶۳-۱۳۹۰ از شرکت آب منطقه‌ای استان گلستان و سازمان هواشناسی گلستان تهیه شد. همچنین نقشه مدل رقومی ارتفاعی نیز که برای استخراج زیرحوضه‌ها و پارامترهای فیزیوگرافی حوضه لازم بود از شرکت آب منطقه‌ای استان تهیه شد. در منطقه مطالعه شده تعداد ۳۹ ایستگاه هیدرومتری و ۱۸ ایستگاه هواشناسی استفاده شده است.

شکل ۱ موقعیت استان گلستان به همراه ۳۹ زیرحوضه آبریز را نشان می‌دهد. جهت استخراج زیرحوضه‌های آبریز در محیط نرم افزار ArcGIS9.3 و با استفاده از برنامه الحاقی Archydro اقدام گردید.

ترسیم زیرحوضه‌ها و محاسبه پارامترهای فیزیوگرافی ابتدا براساس نقشه توپوگرافی در محیط ArcGIS، زیرحوضه‌های منطقه مطالعاتی استخراج و پارامترهای فیزیوگرافی آنها محاسبه شدند تا به‌عنوان متغیرهای مستقل

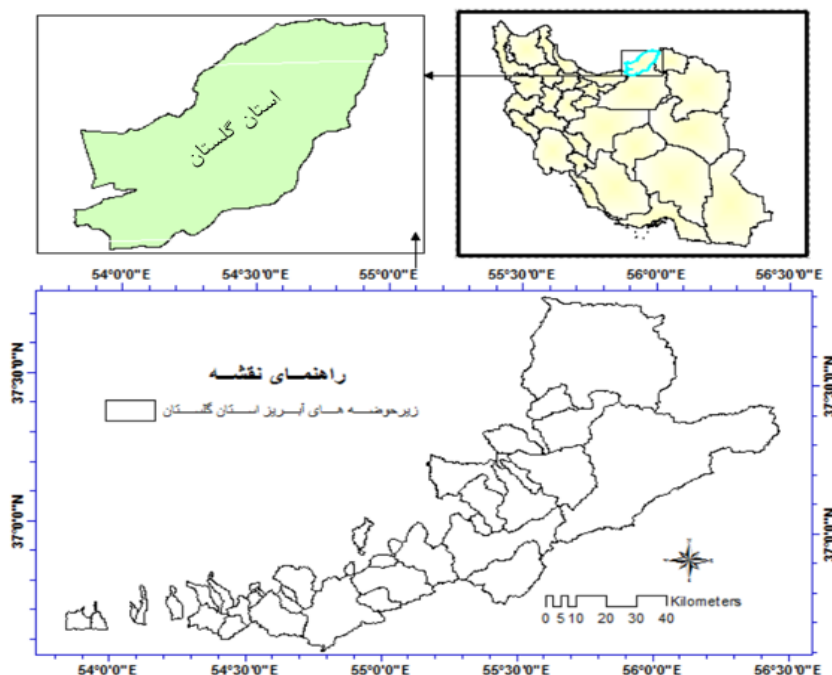
تولید شد. سپس مدل‌های به‌دست‌آمده با استفاده از سه زیرحوضه دیگر آزمون شدند که در پیدایش مدل‌های ریاضی مشارکت نداشتند. در نهایت مناسب‌ترین روابط ریاضی برای برآورد دبی سیلابی در دوره بازگشت‌های مختلف در حوضه‌های آبریز فاقد آمار در منطقه جنوب شرق ایران به‌دست آمد. در زمینه تخمین دبی رودخانه در مقیاس‌های زمانی مختلف علاوه بر روابط تجربی، از مدل‌های داده‌کاوی نیز استفاده شده است. اکبری و همکارانش [۱۶] برای پیش‌بینی جریان روزانه رودخانه با استفاده از الگوریتم نزدیک‌ترین K- همسایگی پرداختند که نتایج بیانگر عملکرد مطلوب روش KNN بوده است

قربانی و همکارانش [۱۷] برای شبیه‌سازی و پیش‌بینی دبی جریان ماهانه با ارزیابی روش‌های هیدرولوژیکی و داده‌کاوی نتیجه گرفتند که استفاده از مدل‌های داده‌کاوی به افزایش شایان توجهی در دقت نتایج پیش‌بینی جریان ماهانه رودخانه نسبت به استفاده از مدل هیدرولوژیکی مفهومی IHACRES منجر شده است. همچنین در تحقیقی دیگر مدل‌های هیدرولوژیکی و پایه داده‌ای در برآورد رواناب روزانه در حوضه گالیکش را ارزیابی کردند. نتایج نشان داد ضریب همبستگی بین مقادیر مشاهداتی و شبیه‌سازی شده برای مدل‌های ANN، KNN و ANFIS به ترتیب برابر ۰/۸۵، ۰/۸۳ و ۰/۸۵ به‌دست آمد. در حالی که برای مدل IHACRES ضریب همبستگی برابر ۰/۶۷ محاسبه شد. بنابراین، مدل‌های پایه داده‌ای عملکرد بهتری نسبت به مدل‌های هیدرولوژیکی دارند [۱۸].

نعیمی‌کلورزی و همکارانش [۱۹] با در نظر گرفتن پارامترهای فیزیوگرافی و اقلیمی زیرحوضه‌های آبریز استان گلستان اقدام به شبیه‌سازی دبی ماهانه رودخانه پرداختند. در این تحقیق با استفاده از رگرسیون چندمتغیره خطی و مدل درختی M5 به این نتیجه رسیدند که برآورد دبی رودخانه در مقیاس ماهانه با استفاده از مدل M5 دقت بیشتر و خطای کمتری دارد.

هدف این تحقیق پیش‌بینی دبی فصلی با استفاده از پارامترهای فیزیوگرافی و اقلیمی زیرحوضه‌ها از طریق مدل‌های نزدیک‌ترین k- همسایگی (KNN) و درخت تصمیم‌گیری^۱ M5 است. منطقه پژوهش شده استان گلستان است که از آمار و اطلاعات ایستگاه‌های سینوپتیک و

1. Decision Tree



شکل ۱. موقعیت استان گلستان و زیرحوضه های آن

خصوصیات فیزیوگرافی حوضه های آبریز خصوصیات فیزیکی حوضه به مجموعه پارامترهایی گفته می شود که مقادیر آنها برای حوضه آبریز نسبتاً ثابت است و در واقع نشان دهنده وضع ظاهری حوضه است. به طوری که قبلاً اشاره شد پارامترهای فیزیکی از این نظر اهمیت دارند که میان آنها و رواناب به دست آمده از بارش های جوی روابطی وجود دارد. بنابراین، با استفاده از این روابط می توان مقدار دبی رودخانه یا شدت سیلاب ها را برای حوضه های فاقد آمار، برآورد و یا محاسبه کرد. با استفاده از نرم افزار ArcGIS 9.3 شبکه آبراهه ها ترسیم و مرز زیرحوضه ها بسته و پارامترهای فیزیوگرافی زیرحوضه ها محاسبه شد.

محاسبه پارامترهای اقلیمی بارش و دما

داده های روزانه بارش و دما طی سال های آماری ۱۳۶۳-۱۳۹۰ از سازمان هواشناسی و آب منطقه ای استان گلستان تهیه و ایستگاه هایی انتخاب شدند که آمار کافی داشتند. سپس براساس روش های درون یابی، مقادیر ماهانه بارش و دما برای هر یک از ماه های سال طی این دوره پهنه بندی و میانگین منطقه ای آنها در محیط GIS در هر یک از زیرحوضه ها محاسبه شد و پارامترهای بارش و دما به صورت مقیاس فصلی مرتب سازی شدند تا به عنوان متغیرهای مستقل وارد مدل شوند.

وارد مدل شوند. به منظور محاسبه مشخصات فیزیوگرافی زیرحوضه های آبریز با استفاده از برنامه الحاقی HecGeoHMS در محیط نرم افزاری ArcGIS اقدام شد. این پارامترها شامل مساحت حوضه (A)، محیط حوضه (Per)، ارتفاع متوسط حوضه (Hmean)، طول آبراهه اصلی (Lab) و شیب آبراهه اصلی (Sab) است. برخی از پارامترها از قبیل عامل شکل حوضه (SF)، فرم حوضه، تراکم شبکه رودخانه (μ)، ضریب گراویلیوس (C)، زمان تمرکز (T_c)، طول مستطیل معادل (L) و عرض مستطیل معادل (B) با استفاده از روابط موجود به صورت زیر برآورد شدند:

$$t_c = 0.949 \left(\frac{L^r}{H} \right)^{0.785} \quad (1)$$

$$C = \frac{0.28P}{\sqrt{A}} \quad (2)$$

$$\mu = \frac{\sum L_i}{A} \quad (3)$$

$$SF = \frac{L^r}{A} \quad (4)$$

$$FF = \frac{A}{L^r} \quad (5)$$

$$L = \frac{C\sqrt{A} + \sqrt{C^2A - 1/2544A}}{1/12} \quad (6)$$

$$B = \frac{C\sqrt{A} - \sqrt{C^2A - 1/2544A}}{1/12} \quad (7)$$

پیش‌بینی داده‌های پیوسته ارائه کرد. این مدل، برخلاف مدل‌های درخت تصمیم معمول که کلاس یا رده‌های گسسته را به‌عنوان خروجی ارائه می‌کنند، نوعی مدل خطی چندمتغیره را برای داده‌ها در هر گره از مدل درختی می‌سازد. تشکیل ساختار مدل‌های درخت تصمیم‌گیری شامل مراحل ایجاد درخت و هرس کردن آن است. در مرحله ساختن درخت، از نوعی الگوریتم استنتاجی یا معیار تقسیم (انشعاب) برای تولید یک درخت تصمیم استفاده می‌شود.

الگوریتم نزدیک‌ترین K-همسایگی (KNN)

یکی دیگر از روش‌های مدل‌سازی در داده‌کاوی الگوریتم K-نزدیک‌ترین همسایگی است. این الگوریتم جز روش‌های یادگیری نظارت شده است که هم در طبقه‌بندی و هم در پیش‌بینی استفاده می‌شود. چگونگی عملکرد این الگوریتم براساس مشاهدات و نمونه‌هاست. براساس این الگوریتم می‌توان نمونه جدیدی را براساس بیشترین K گروه و دسته که نزدیک‌ترین همسایگی‌ها را با آن نمونه داشته باشند، تقسیم کرد. به بیان دیگر، می‌توان گفت که این روش K تعداد از الگوهای مشابه را پیدا می‌کند و براساس آنها ارزش نمونه مطالعه‌شده را پیش‌بینی می‌کند. در روش K-نزدیک‌ترین همسایگی نمونه ناشناخته S در یک کلاس از پیش تعریف‌شده C_i متعلق به C براساس نمونه‌های کلاس‌بندی‌شده قبلی (داده‌های آموزش) طبقه‌بندی می‌شود. زمانی که نمونه S کلاس‌بندی می‌شود، KNN فاصله آن را با همه نمونه‌های موجود در داده‌های آموزشی اندازه‌گیری می‌کند. فاصله اقلیدسی از متداول‌ترین معیارهای اندازه‌گیری فاصله است. سپس همه مقادیر فاصله مرتب می‌شوند به طوری که $d_i \leq d_{i+1}$, $i=1,2,3,\dots,n$ است. K نمونه با کمترین فاصله به نمونه جدید به‌عنوان نزدیک‌ترین K-همسایه شناخته می‌شوند و برای کلاس‌بندی نمونه جدید S به کلاس موجود C_i , $1 < i \leq n \in C$ استفاده می‌شود. تصمیم طبقه‌بندی به طبیعت داده‌ها بستگی دارد. این الگوریتم جزء روش‌های تبیل به حساب می‌آید. به این دلیل که مرحله آموزش را همان زمان اجرا می‌کند که نمونه جدید باید طبقه‌بندی شود برخلاف الگوریتم‌های یادگیری، نقطه مقابل آن که داده‌های آموزشی را قبل از ورود نمونه جدید طبقه‌بندی می‌کنند. بر این اساس، KNN نسبت به دیگر الگوریتم‌های یادگیری محاسبات بیشتری را لازم

پس از استخراج پارامترهای فیزیوگرافی و محاسبه پارامترهای اقلیمی حوضه آبریز، به‌عنوان متغیرهای مستقل و پیش‌بینی‌کننده دبی، به جست‌وجوی روابط بین آنها با متغیر هدف یا دبی فصلی پرداخته شد. در این پژوهش از مدل داده‌کاوی شامل M5 و KNN برای مدل‌سازی غیرخطی بین داده‌ها استفاده شد. همچنین از مدل رگرسیون چندمتغیره خطی نیز برای بررسی رابطه بین متغیرهای مستقل و وابسته استفاده شد.

رگرسیون چندمتغیره خطی (MLR)

در بیشتر مطالعات آماری بیشتر همبستگی یک عامل را فقط با عاملی در نظر می‌گیرند و اثر دیگر عوامل و ترکیب این عوامل بر عامل مد نظر نادیده گرفته می‌شود. در صورتی که بخواهیم اثر چند عامل را بر یک عامل در نظر بگیریم، از ساده‌ترین راه‌ها استفاده از رابطه خطی چندمتغیره است که شکل کلی آن به صورت زیر است:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (8)$$

مدل درخت تصمیم M5

در مدل‌سازی پدیده‌ها، وجود شرایط محلی ممکن است موجب شود تا استفاده از یک رابطه کلی نتایج خوبی به همراه نداشته باشد و تغییرات محلی به خوبی دیده نشوند. در صورت امکان، شناسایی محدوده‌های همگن و ارائه روابط ساده خطی برای هر یک از این محدوده‌ها می‌تواند موجب افزایش دقت مدل شود. بر این اساس، معمولاً برای حل مسائل پیچیده، آن را به چند مسئله کوچک‌تر و ساده‌تر تقسیم و سپس جواب‌های به‌دست‌آمده را با هم ترکیب می‌کنند. همین ایده ساده در مدل‌های درخت تصمیم استفاده می‌شود. به این منظور، فضا یا محدوده مقادیر داده‌های ورودی به چند زیربازه یا ناحیه تقسیم و برای هر ناحیه یک معادله یا مدل مناسب استخراج می‌شود درخت‌های تصمیم روشی برای نمایش یک سری از قوانین هستند که به یک رده یا مقدار منتهی می‌شوند. درخت‌های تصمیم به کمک جداسازی متوالی داده‌ها به یک سری گروه مجزا تشکیل شده و سعی می‌شود در فرایند جداسازی، فاصله بین گروه‌ها افزایش یابد. ساختار یک مدل درختی شامل ریشه، گره‌های داخلی و برگ است. از مدل‌های درخت تصمیم در حل بسیاری از مسائل طبقه‌بندی و رگرسیون استفاده شده است. برای نخستین بار Quinlan (۱۹۹۲) مدل درخت تصمیم موسوم به M5 را برای

دارد. KNN برای داده‌های پویا و داده‌هایی مناسب است که سریع تغییر می‌کنند و به‌روز می‌شوند. در روش آماری ناپارامتری KNN دو عامل مهم دخیل‌اند. اولین عامل، انتخاب تابع فاصله‌سنجی و وزن‌های مربوط به آن و دومی انتخاب بهترین تعداد همسایگی است. در انجام پیش‌بینی‌ها توسط KNN استفاده از تابع فاصله‌سنجی اقلیدسی وزن‌دار شده بسیار معمول است. روش KNN با اختصاص وزن‌های بیشتر به همسایگی‌های منتخبی که از نظر فاصله زمانی نسبت به زمان حاضر نزدیک‌تر هستند توانسته است نتایج پیش‌بینی را نسبت به اتفاقات همسایگی‌های نزدیک‌تر به زمان فعلی وابسته‌تر کند (Yates et al, 2003). تحقیقات گسترده‌ای برای انتخاب بهترین همسایگی و وزن‌ها انجام شده است (Piechota et al, 2001). همچنین تصحیح و یا حذف اطلاعات پرت و یا مشکوک می‌تواند توانایی و قدرت کارایی این روش را افزایش دهد. بنابراین، تعداد متغیرهای پیش‌بینی‌کننده و طول آماری مناسب می‌تواند تأثیر شایانی را روی دقت مدل داشته باشد (Tarboton et al, 1993).

از مزایای استفاده از الگوریتم KNN می‌توان به موارد زیر اشاره کرد:

- اجرای ساده؛
- بی‌نیازی به مرحله تخمین پارامترها؛
- قابلیت مدل‌سازی غیرخطی؛
- مؤثر بودن و عملکرد با بازدهی فراوان در برخورد با تعداد دسته‌های زیاد از داده‌ها.

همان‌طور که بیان شد در روش KNN نیاز به یک تابع فاصله‌سنجی به‌منظور یافتن بهترین همسایگی‌ها است که دو تابع فاصله‌سنجی اقلیدسی و ماهانالوبیس از معمول‌ترین این توابع‌اند. تابع فاصله‌سنجی اقلیدسی براساس فاصله متریک دو نقطه در فضای اقلیدسی تعریف می‌شود که در این تحقیق نیز از فاصله اقلیدسی استفاده شده است. چنانچه m متغیر پیش‌بینی‌کننده با طول سری زمانی تاریخی n در نظر گرفته شود این تابع فاصله‌سنجی را می‌توان به‌صورت زیر تعریف کرد:

$$Dist(t-i) = \sqrt{\sum_{j=1}^m W_j (x_{js(t-i)} - x_{j,t})^2} \quad (9)$$

که در این رابطه:

مؤلفه‌های استفاده‌شده در شبیه‌سازی

در شبیه‌سازی با استفاده از مدل درخت تصمیم‌گیری و نزدیک‌ترین K همسایگی، از عناصر مختلفی استفاده شده است. این عناصر به‌عنوان متغیر مستقل به مدل معرفی و شبیه‌سازی‌ها برای پیش‌بینی متغیر هدف صورت گرفته است. متغیرهای مستقل استفاده‌شده در این پژوهش شامل متوسط بارش فصلی (P) برحسب میلی‌متر و متوسط دمای فصلی (T) برحسب درجه سانتی‌گراد با طول دوره آماری ۱۳۶۳-۱۳۹۰ و پارامترهای فیزیوگرافی مانند مساحت حوضه (A) برحسب کیلومتر مربع، محیط حوضه (Per) برحسب کیلومتر، ارتفاع متوسط حوضه (Hmean) برحسب متر، عامل شکل حوضه (SF) بی‌بعد، فرم حوضه (FF) بی‌بعد، طول آبراهه اصلی (Lab) برحسب کیلومتر، شیب آبراهه اصلی (Sab) برحسب درصد، تراکم شبکه رودخانه (μ) برحسب کیلومتر بر کیلومتر مربع، ضریب گراویلیوس (C)، زمان تمرکز (Tc) برحسب ساعت، طول مستطیل معادل (L) و عرض مستطیل معادل (B) که برحسب کیلومتر است.

ارزیابی نتایج

برای ارزیابی نتایج مدل‌سازی از معیارهای ارزیابی ریشه میانگین مربعات خطا (RMSE)، میانگین خطای اریب (MBE) و ضریب همبستگی (R) استفاده شد. RMSE بزرگی خطا را نشان می‌دهد و هرچه مقدار آن کوچک‌تر باشد، دقت مدل بیشتر است. MBE انحراف از خطای میانگین یا بیش‌برآورد و کم‌برآورد مدل را نشان می‌دهد باید متمایل به صفر باشد. R نیز ضریب همبستگی بین مقادیر دبی مشاهداتی (Q_o) و دبی پیش‌بینی‌شده (Q_p) است که زیادبودن آن دقت مدل را نشان می‌دهد که روابط به‌صورت زیر ارائه داده شد:

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (Q_p - Q_o)^2}{\sum_{i=1}^n (Q_o - \bar{Q})^2}} \quad (10)$$

زیرحوضه‌ها به‌عنوان متغیرهای مستقل وارد مدل شدند و به‌کمک مدل‌های داده‌کاوی M5، KNN و رگرسیون چندمتغیره خطی مقدار متوسط آبدهی فصلی رودخانه مدل‌سازی شد. سپس برای مشخص کردن تأثیر بارش فصل قبل در پیش‌بینی دبی، بارش یک فصل قبل نیز به‌عنوان متغیر مستقل وارد مدل شدند (الگوی دوم). نتایج ارزیابی مدل‌های به‌کار گرفته‌شده در جدول‌های ۱ و ۲ ارائه شده است.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_p - Q_o)^2}{n}} \quad (11)$$

$$MBE = \frac{\sum_{i=1}^n (Q_p - Q_o)}{n} \quad (12)$$

نتایج و بحث

برای پیش‌بینی جریان فصلی رودخانه از دو الگوی ورودی متغیرها به مدل استفاده شد. در الگوی نخست ابتدا داده‌های اقلیمی بارش و دما به همراه مشخصات فیزیوگرافی

جدول ۱. نتایج ارزیابی مدل‌های درخت تصمیم و نزدیک‌ترین k همسایگی برای شبیه‌سازی آبدهی فصلی در الگوی اول

فصل	مدل M5			الگوریتم KNN			مدل MLR		
	MBE	RMSE	R	MBE	RMSE	R	MBE	RMSE	R
بهار	-۰/۰۰۹	۱/۵۴	۰/۸۰۱	-۰/۰۲۱	۱/۰۸۶	۰/۹۱۸	۰/۰۲۸	۱/۹۷۹	۰/۶۹۶
تابستان	-۰/۰۰۰۲	۰/۷۴۳	۰/۳۹۷	۰/۰۰۰۲	۰/۴۲	۰/۷۲۵	۰/۰۱۶	۰/۵۱۵	۰/۵۶۷
پاییز	-۰/۰۰۱۷	۰/۳۳۸	۰/۸۸۲	۰/۰۲۳	۰/۲۹۳	۰/۹۱۵	۰/۰۰۴	۰/۴۹۷	۰/۷۳۱
زمستان	-۰/۰۰۴	۰/۷۱۷	۰/۹۰۱	-۰/۰۱۶	۰/۵۵۴	۰/۹۳۹	-۰/۰۱۰	۱/۰۳۲	۰/۷۶۸

جدول ۲. نتایج ارزیابی مدل‌های درخت تصمیم و نزدیک‌ترین k همسایگی برای شبیه‌سازی آبدهی فصلی در الگوی دوم

فصل	مدل M5			الگوریتم KNN			مدل MLR		
	MBE	RMSE	R	MBE	RMSE	R	MBE	RMSE	R
بهار	-۰/۰۱۶	۱/۴۹۹	۰/۸۱۶	-۰/۰۹۶	۰/۹۱	۰/۹۳۶	۰/۰۱۸	۱/۹۶۷	۰/۵۹۶
تابستان	۰/۰۰۰۳	۰/۷۶۷	۰/۴۱۲	-۰/۰۰۰۱	۰/۴۲۸	۰/۷۲۱	۰/۰۱۵	۱/۲۳۱	۰/۵۳۴
پاییز	۰/۰۰۰۶	۰/۳۸۹	۰/۹۰۱	۰/۰۲۳	۰/۲۹۴	۰/۹۲۶	۰/۰۰۶	۰/۹۸۹	۰/۷۳۱
زمستان	۰/۰۰۷	۰/۵۴۶	۰/۹۴۶	-۰/۰۰۱	۰/۳۷۵	۰/۹۵۷	-۰/۰۰۸	۰/۶۲۱	۰/۷۶۷

را از مقدار مشاهداتی کمتر برآورد کرده است و در فصول تابستان و پاییز شبیه‌سازی با بیش‌برآوردی همراه بوده است. نتایج نشان داد ورود پارامتر بارش یک فصل قبل به مدل‌سازی سبب افزایش دقت پیش‌بینی و کاهش خطا شده است. با توجه به جدول ۲ در فصل زمستان با استفاده از مدل درخت تصمیم میزان ضریب همبستگی بین مقادیر دبی مشاهداتی و پیش‌بینی‌شده برابر ۰/۹۴۶ و ریشه میانگین مربعات خطای آن برابر ۰/۵۴۶ است و در مدل KNN میزان همبستگی بین مقادیر دبی مشاهداتی و پیش‌بینی‌شده برابر ۰/۹۵۷ و ریشه میانگین مربعات خطای برابر ۰/۳۷۵ بوده است که برتری مدل KNN نسبت به M5 را نشان می‌دهد. در سایر فصول نیز بعد از فصل زمستان، فصل پاییز بیشترین دقت از شبیه‌سازی را به خود اختصاص داده است و در نهایت فصل بهار و تابستان از نظر دقت و خطا در رتبه‌های پایین‌تر قرار می‌گیرند. همچنین با توجه به نتایج روش رگرسیون

با مقایسه سه معیار ارزیابی شامل ضریب همبستگی، ریشه میانگین مربعات خطا و میانگین خطای اریب در روش درخت تصمیم، الگوریتم نزدیک‌ترین K همسایگی و رگرسیون چندمتغیره خطی مشخص شد که فصل زمستان در بین سایر فصول نتایج بهتری دارد. براساس نمایه‌های ارزیابی محاسبه‌شده از بیشترین دقت تا پایین‌ترین دقت به ترتیب فصل زمستان، پاییز، بهار و تابستان است. با توجه به نزدیک‌بودن نتایج دو مدل M5 و KNN، این دو روش نتایج بهتری را نسبت به رگرسیون چندمتغیره ارائه داده‌اند. مدل M5 در فصول بهار، پاییز و زمستان دبی فصلی رودخانه را از مقدار واقعی آن کمتر شبیه‌سازی کرده است (کم‌برآوردی) و در فصل تابستان اختلاف بین مقادیر دبی شبیه‌سازی‌شده و مشاهداتی میزان ناچیزی بود و بیش‌برآوردی اتفاق افتاده است. مدل KNN در فصول بهار و زمستان میزان دبی فصلی

آبراهه اصلی، طول آبراهه اصلی، مساحت حوضه و فرم حوضه از عوامل تأثیرگذار بر دبی فصل زمستان هستند.

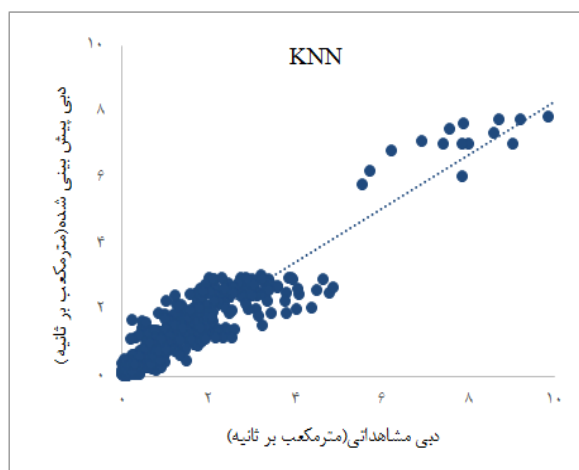
$$Q = -0.558 + 1/88S_{ab} + 0.1722L_{ab} + 0.08759A - 0.2102SF \quad (13)$$

شکل‌های ۲ و ۳ میزان پراکنش بین داده‌های دبی مشاهداتی و دبی پیش‌بینی شده را با استفاده از دو مدل داده‌کاوی رگرسیون درختی M5 و الگوریتم نزدیک‌ترین K-همسایگی نشان می‌دهد.

نتایج خروجی از مدل درختی M5 برای فصل زمستان در زیر آورده شده است. از پارامترهای مؤثر بر تخمین دبی با استفاده از مدل یادشده می‌توان به محیط حوضه، مساحت حوضه، شیب آبراهه اصلی، ارتفاع متوسط حوضه، تراکم زهکشی، شیب حوضه و بارش یک فصل قبل اشاره کرد که در این بین محیط و مساحت حوضه، بارش یک فصل قبل، شیب حوضه و عرض مستطیل معادل تأثیر بیشتری در برآورد متوسط آبدهی فصل زمستان بوده است.

چندمتغیره خطی میزان ضریب همبستگی در فصل زمستان بیشتر از سایر فصول سال بوده است و میزان خطای بین دبی مشاهداتی و پیش‌بینی شده در فصل زمستان به کمترین مقدار خود رسیده است. به‌طور کلی، نتایج نشان می‌دهد مدل‌های M5 و KNN توانایی بهتری در شبیه‌سازی دبی فصلی رودخانه نسبت به رگرسیون چندمتغیره خطی دارند. همچنین با توجه به معیار MBE مقادیر مثبت آن بیش‌برآوردی از دبی و مقادیر منفی کم‌برآوردی از دبی فصلی رودخانه را نشان می‌دهد. در جدول ۲ همان‌طور که مشاهده می‌شود مدل M5 میزان دبی فصل زمستان را نسبت به مقدار مشاهداتی آن بیشتر برآورد کرده است و با استفاده از مدل KNN شاهد کم‌برآوردی از دبی فصل زمستان هستیم. هرچه مقادیر MBE به صفر نزدیک‌تر باشند، نشان‌دهنده برآورد مطلوب از دبی فصلی است.

رابطه زیر معادله رگرسیونی بین پارامترهای فیزیوگرافی و اقلیمی زیرحوضه‌های آبریز را نشان می‌دهد. پارامتر شیب



شکل ۳. نمودار پراکنش بین دبی مشاهداتی و پیش‌بینی شده فصل زمستان با روش KNN

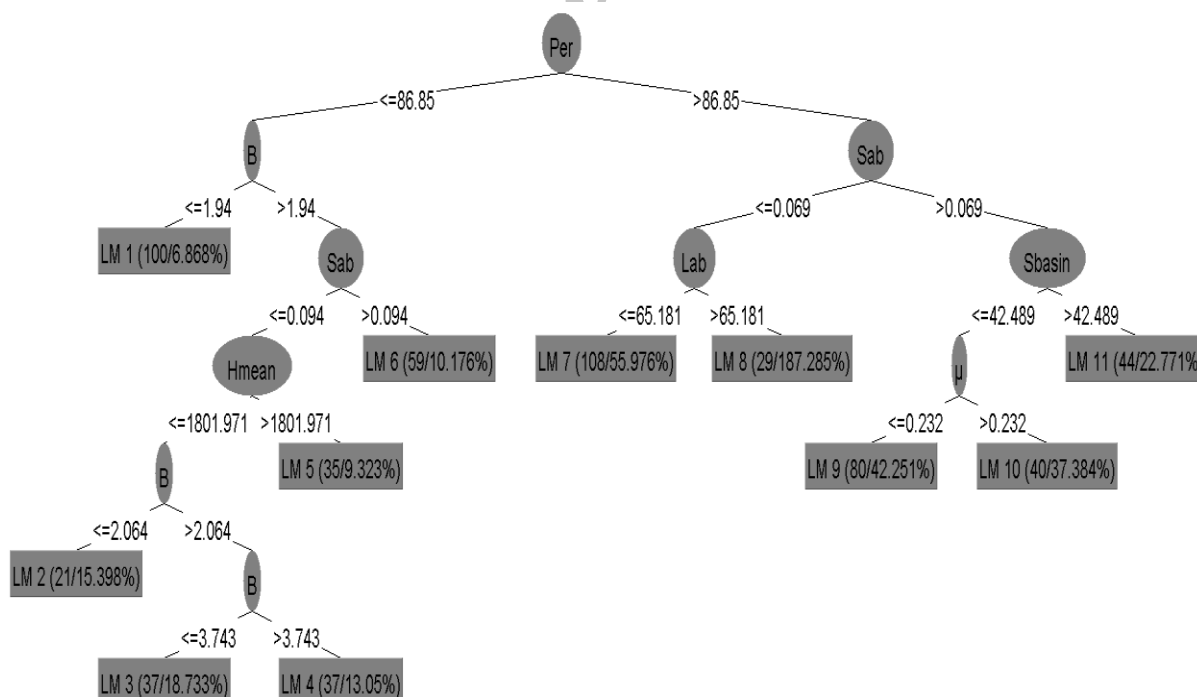


شکل ۲. نمودار پراکنش بین دبی مشاهداتی و پیش‌بینی شده فصل زمستان با روش M5

Per <= 86.85 :
 | B <= 1.94 : LM1 (100/7.116%)
 | B > 1.94 :
 || Sab <= 0.094 :
 ||| Hmean <= 1801.971 :
 |||| B <= 2.064 : LM2 (21/10.317%)
 |||| B > 2.064 :
 ||||| B <= 3.743 : LM3 (37/18.506%)
 ||||| B > 3.743 : LM4 (37/13.922%)
 ||| Hmean > 1801.971 : LM5 (35/9.453%)
 || Sab > 0.094 : LM6 (59/9.237%)
 Per > 86.85 :
 | Sab <= 0.069 :

|| $\mu \leq 0.249$: LM7 (103/55.142%)
 || $\mu > 0.249$: LM8 (29/69.247%)
 | Sab > 0.069 :
 || Sbasin ≤ 42.489 :
 ||| $\mu \leq 0.232$: LM9 (80/45.073%)
 ||| $\mu > 0.232$: LM10 (40/38.102%)
 || Sbasin > 42.489 : LM11 (44/20.296%)

LM num: 1
 $Q = -0.0004 * A - 0.2838 * \mu + 0.0036 * Sab - 0.0001 * Hmean + 0.1546 * C + 0.0539 * B + 0.0019 * Sbasin + 15.9227 * V - 0.1778$
 LM num: 2
 $Q = 0.0073 * A - 0.2838 * \mu - 0.7327 * Sab - 0.0001 * Hmean + 0.1546 * C + 0.0332 * B + 0.0019 * Sbasin - 0.0949$
 LM num: 3
 $Q = -0.0002 * A - 0.2838 * \mu - 0.7327 * Sab - 0.0002 * Hmean + 0.1546 * C - 0.0018 * B + 0.0019 * Sbasin + 3.5132 * V1 + 0.5934$
 LM num: 4
 $Q = -0.0002 * A - 0.2838 * \mu - 0.7327 * Sab - 0.0001 * Hmean + 0.1546 * C - 0.0018 * B + 0.0019 * Sbasin + 3.5132 * V1 + 0.3809$
 LM num: 5
 $Q = 0 * A - 0.2838 * \mu - 0.7327 * Sab - 0.0001 * Hmean + 0.1546 * C + 0.0228 * B + 0.0019 * Sbasin - 5.9083 * V + 0.2475$
 LM num: 6
 $Q = 0.0022 * A - 0.2838 * \mu - 1.06 * Sab - 0.0001 * Hmean + 0.1546 * C + 0.0393 * B + 0.0019 * Sbasin - 12.933 * V - 0.0494$
 LM num: 7
 $Q = -0.0002 * A + 10.0893 * \mu - 4.4818 * Sab - 0.0001 * Hmean - 0.8935 * C + 0.032 * B + 0.0018 * Sbasin - 1.8887 * V - 0.0328 * avgT + 2.2619$
 LM num: 8
 $Q = 0.0182 * A + 24.4182 * \mu - 4.4818 * Sab - 0.0001 * Hmean + 1.0547 * C + 0.032 * B + 0.0018 * Sbasin + 33.3696 * V - 15.1396$
 LM num: 9
 $Q = -0.0002 * A - 1.2637 * \mu - 3.7331 * Sab - 0.0001 * Hmean + 0.1511 * C + 0.032 * B + 0.0003 * Sbasin + 0.6024 * V1 + 1.7585$
 LM num: 10
 $Q = -0.0002 * A - 2.2727 * \mu - 3.7331 * Sab - 0.0001 * Hmean + 0.1511 * C + 0.032 * B + 0.0003 * Sbasin + 5.0732 * V1 + 1.5085$
 LM num: 11
 $Q = 0.0081 * A - 1.3132 * \mu - 3.7331 * Sab - 0.0001 * Hmean + 0.1511 * C + 0.032 * B - 0.0017 * Sbasin - 1.905$



شکل ۴. نمودار درختی فصل زمستان

Agricultural Sciences and Natural Resources. 2010. [Persian].

- [3]. Lohani AK, Kumar R, Singh RD. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology*. 2012; 442: 23-35.
- [4]. Kisi Ö. Daily river flow forecasting using artificial neural networks and auto-regressive models. *Turkish Journal of Engineering and Environmental Sciences*. 2005; 29(1): 9-20.
- [5]. Nabizadeh M, Mosaedi A, Hesam M, Dehghani AA. Comparing the performance of Fuzzy based models in stream flow on Lighvan River. *J. of Water and Soil Conservation*. 2012; 19(1): 117-134. [Persian].
- [6]. Parviz L, Kholghi M, Malmir M. Comparison of Methods temporal resolution and artificial neural networks in anticipation of a seasonal river flow. *Journal of Iran Water Research*. 2008; 2(2): 9-17. [Persian].
- [7]. Zarezadeh-Mehrzi M, Bozorg Haddad O. Inflow Simulation and Forecasting Optimization Using Hybrid ANN-GA Algorithm. *Journal of Water and Soil*. 2010; 24(5): 942-954. [Persian].
- [8]. Seyedian SM, Soleimani M, Kashani M. Predicting streamflow data-driven model and time series. *Iranian Journal of Eco Hydrology*. 2015; 1(3): 167-179. [Persian].
- [9]. Ahmadi F, Radmanesh F, Mirabbasi Najaf abadi R. Comparison between Genetic Programming and Support Vector Machine Methods for Daily River Flow Forecasting (Case Study: Barandoozchay River). *Journal of Water and Soil*. 2014; 28(6): 1162-1171. [Persian].
- [10]. Sanikhani H, Dinpajuh Y, Ghorbani MM. River flow modeling using K- nearest neighborhood and intelligent methods. *Journal of Water and Soil Science*. 2015; 25(1): 219-233. [Persian].
- [11]. Firat M, Gungor M. River flow estimation using adaptive neuro-fuzzy inference system. *Journal of Mathematics and Computers in Simulation*. 2006; 75(3-4): 87-96
- [12]. Negaresh H, Tavousi T, Mehdinasab M. Modeling the Production of Runoff in Kashkan River Catchment Based on the Statistical Methods. *Journal of Research in urban ecology*. 2014; 3(6): 81-92. [Persian].
- [13]. Zare Abyaneh H, Bayat Varkeshi M. Evaluation of Artificial Intelligent and Empirical Models in Estimation of Annual Runoff. *Journal of Water and Soil*. 2010; 25(2): 365-379. [Persian].

نتیجه‌گیری

برآورد متوسط آبدهی فصلی رودخانه ابزاری مهم در مدیریت و برنامه‌ریزی منابع آب و محیط زیست است. در حوضه‌هایی که آمار محدود ثبت شده هستند برآورد آبدهی رودخانه به مراتب پیچیدگی بیشتری دارد.

در این پژوهش از سه روش مدل‌سازی به نام رگرسیون درختی M5، الگوریتم نزدیک‌ترین k- همسایگی و رگرسیون چندمتغیره خطی برای تخمین دبی فصلی رودخانه استفاده شد. برای شبیه‌سازی دبی در مقیاس فصلی از پارامترهای اقلیمی و فیزیوگرافی ۳۹ زیرحوضه آبریز استان گلستان به عنوان متغیرهای ورودی به مدل‌های یادشده استفاده شد. مدل درخت تصمیم M5 ضمن سادگی محاسبه‌ها و معادله‌های ارائه شده، قابلیت خوبی در تخمین آبدهی رودخانه داشته است. این مدل به ازای پارامترهای اقلیمی شامل متوسط بارش و دما و پارامترهای فیزیوگرافی مقدار دبی رودخانه را برای فصل‌های مختلف به خوبی تخمین می‌زند. در پژوهش نیمیه کلورزی و همکارانش [۱۹] مدل M5 در برآورد دبی ماهانه رودخانه عملکرد مطلوبی داشت. از طرفی الگوریتم KNN نیز به دلیل تعداد زیاد داده‌های ورودی به مدل، نتایج مطلوب‌تری از مدل‌سازی دبی فصلی رودخانه نسبت به M5 ارائه می‌دهد که با نتایج پژوهش قربانی و همکارانش [۱۸] مبنی بر عملکرد خوب مدل‌های داده‌کاوی در شبیه‌سازی دبی رودخانه مطابقت دارد. ارزیابی نتایج به دست آمده از مدل‌های یادشده نشان داد ضریب همبستگی داده‌ها برای فصل‌های پرباران بیشتر از دیگر فصل‌های سال است و دلیل آن شرایط آب و هوایی در بهار و تابستان و همچنین استفاده از آب برای مصارف کشاورزی توسط کشاورزان است که سبب کاهش دبی رودخانه می‌شود. نتایج نشان داد مدل‌های M5 و KNN دقت زیاد و خطای کمتری نسبت به روش رگرسیون چندمتغیره خطی داشتند و مدل KNN نسبت به سایر مدل‌ها توانایی بیشتری در شبیه‌سازی دبی فصلی رودخانه داشته است.

منابع

- [1]. Govindaraju RS. Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*. 2000; 5(2): 115-123.
- [2]. Salajegheh A, Fathabadi A, Gholami H. Predict river discharge using the nearest neighbor. 5th national conference on science and management engineering Iran. Gorgan University of

- [14].Eskandarinia AR, Nazarpour H, Ahmadi MZ, Teimouri M, Moshfegh MZ. Examine the effect of antecedent precipitation in the river flow estimates by artificial neural network (case study: Bakhtiari River). *Journal of watershed management*. 2011; 2(3): 51-62. [Persian].
- [15].Khedmati H, Manshoury M, Heydarizade M, Sedghi H. Zonation and Estimation of Flood Discharge in Ungauged Sites Located in South-East Basins of Iran Using a Combination of Flood Index and Multi-Variable Regression Methods (Sistan and Baluchistan, Kerman, Yazd and Hormozgan Provinces). *J. Water Soil*. 2010; 24: 3: 593-609. [Persian].
- [16].Akbari M, Van Overloop PJ, Afshar A. Clustered K nearest neighbor algorithm for daily inflow forecasting. *Water resources management*. 2011; 25(5):1341-57.
- [17].Ghorbani Kh, Sohrabian E, Salarijazi M. Evaluation of hydrological and data mining models in monthly river discharge simulation and prediction (Case study: Araz-Kouseh watershed). *Journal of Water and Soil Conservation*. 2016; 23(1): 203-217. [Persian].
- [18]. Ghorbani Kh, Meftah Halaghi M, Sohrabian E. Evaluation of hydrological and data-based models in estimation of daily runoff in Galikesh watershed. *Int. J. Hydrology Science and Technology*. 2016; 6(1): 27-44.
- [19].Naeimi Kalourazi Z, Ghorbani Kh, Salarijazi M, Deghani A. A. Estimation of monthly discharge using climatic and physiographic parameters of ungauged basins. *Journal of Water and Soil Conservation*. 2016; 23(3): 207-224. [Persian].

Archive of SID