Archive of SID

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و سوم، شماره اول، ۱۳۹۵
http://jwsc.gau.ac.ir

ارزیابی روش‌های هیدرولوژیکی و داده‌کاوی در شبیه‌سازی و پیش‌بینی دبی جریان ماهانه (مطالعه موردی: ایستگاه هیدرومتری ارازکوسه)

*خلیل قربانی[۱]، الهه سهرابیان[۲] و میثم سالاری‌جزی[۳]

[۱]دانشیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، [۲]دانش‌آموخته کارشناسی‌ارشد گروه مهندسی منابع آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، [۳]استادیار گروه مهندسی آب، دانشگاه علوم کشاورزی و منابع طبیعی گرگان

تاریخ دریافت: ۹۴/۱۱/۲۰ ؛ تاریخ پذیرش: ۹۵/۳/۳۰

چکیده

سابقه و هدف: پیش‌بینی کمی جریان در رودخانه‌ها یکی از مهم‌ترین ارکان در مدیریت منابع آب‌های سطحی به‌ویژه اتخاذ تدابیر مناسب در مواقع سیلاب و بروز خشکسالی‌ها می‌باشد. برای پیش‌بینی میزان جریان رودخانه‌ها رویکردهای متنوعی در هیدرولوژی معرفی شده است که مدل‌های مفهومی و نیز مدل‌های داده محور از مهم‌ترین آن‌ها می‌باشند. در این مطالعه برای بررسی دقت مدل‌های پیش بینی جریان رودخانه از داده‌های بلندمدت ثبت شده در حوضه آبریز ارازکوسه با مساحت ۱۶۷۸ کیلومترمربع واقع شده در شمال ایران (استان گلستان) استفاده شده است. مدل IHACRES به‌عنوان مدل مفهومی هیدرولوژیکی و مدل‌های M5 و KNN به‌عنوان مدل‌های داده‌کاوی برای مدل‌سازی جریان ماهانه رودخانه انتخاب شدند و نتایج برای بررسی صحت مدل‌های مورد مطالعه با یکدیگر مقایسه شدند. در مطالعات معدودی هر یک از مدل‌های بیان شده در پیش‌بینی دبی جریان روزانه مورد بررسی قرار گرفته است اما هدف این مطالعه بررسی هم‌زمان این مدل‌ها در یک حوضه برای پیش‌بینی جریان ماهانه رودخانه می‌باشد.

مواد و روش‌ها: داده‌های روزانه ۲۹ ساله (۲۰۱۳-۱۹۸۵) بارش و دبی ایستگاه‌های هیدرومتری و هواشناسی ارازکوسه برای استخراج سری‌های ماهانه مورد نیاز برای مدل‌سازی استفاده شد. کمیت و کیفیت داده‌های مورد نیاز برای مدل‌سازی با استفاده از آزمون‌های مختلف آماری بررسی و تأیید شد. داده‌های ثبت شده به دو جز سری تقسیم‌بندی شد که جز سری اول برای واسنجی مدل‌ها به‌کار گرفته شد و از جز سری دوم برای ارزیابی صحت‌سنجی مدل‌ها استفاده گردید. با در نظر گرفتن نتایج هر یک از مدل‌ها در دوره‌های صحت‌سنجی و واسنجی بر اساس معیارهای نکویی برازش کارایی مدل‌ها بررسی و تحلیل شد.

یافته‌ها: نتایج مدل هیدرولوژیکی مفهومی IHACRES در هر دو مرحله صحت‌سنجی و واسنجی (ضریب همبستگی برابر با ۰/۸۱ و ۰/۷۹ به‌ترتیب برای دوره واسنجی و صحت‌سنجی) نشان می‌دهد که این مدل دارای توانایی مناسب برای شبیه‌سازی جریان ماهانه رودخانه می‌باشد. همچنین بررسی نتایج دو مدل داده‌کاوی M5 و KNN (ضریب همبستگی برابر با ۰/۹۴ و ۰/۸۹ به‌ترتیب برای دوره واسنجی و صحت‌سنجی برای مدل KNN و ضریب همبستگی

* مسئول مکاتبه: ghorbani.khalil@yahoo.com

www.SID.ir

Archive of SID

برابر با ۰/۹۲ و ۰/۸۸ به‌ترتیب برای دوره واسنجی و صحت‌سنجی برای مدل M5) نشان می‌دهد که استفاده از این مدل‌ها منجر به افزایش قابل‌ملاحظه‌ای در دقت نتایج پیش‌بینی جریان ماهانه رودخانه نسبت به استفاده از مدل هیدرولوژیکی مفهومی IHACRES شده است.

نتیجه‌گیری: کاربرد مدل‌های داده کاوی یعنی M5 و KNN منجر به بهبود نتایج نسبت به کاربرد مدل هیدرولوژیکی مفهومی IHACRES شده است. این موضوع مشخص است که دقت نتایج مدل‌های داده کاوی بسیار به یکدیگر نزدیک است اما به‌دلیل آن‌که مدل M5 معادلات صریح برای پیش‌بینی ارایه می‌کند به‌عنوان مدل منتخب در این پژوهش انتخاب می‌شود. همچنین بررسی سری زمانی نتایج نشان می‌دهد که دقت مدل‌های داده‌کاوی در تخمین جریان‌های کم بهتر از تخمین جریان‌های زیاد می‌باشد.

واژه‌های کلیدی: جریان ماهانه، حوضه ارازکوسه، IHACRES، KNN، M5

مقدمه

نیاز روزافزون آب سبب گردیده است، برنامه‌ریزی‌های مدیریتی به‌منظور کنترل مصرف آب در آینده از اهمیت بیش‌تری برخوردار گردد. آبدهی رودخانه‌ها در ماه‌های آینده که پدیده‌ای تصادفی و تأثیرپذیر از برخی عوامل هواشناسی و هیدرولوژیکی نظیر دما، بارش و آبدهی می‌باشد، نقش مهمی را در مدیریت منابع آب ایفا می‌کند. با پیش‌بینی نمودن آبدهی رودخانه علاوه بر مدیریت بهره‌برداری منابع آب به‌منظور تامین نیاز، می‌توان حوادث طبیعی نظیر سیل و خشکسالی را نیز پیش‌بینی و مهار نمود. همچنین می‌توان از نتایج اجرای یک مدل شبیه‌سازی به‌منظور بررسی صحت داده‌ها، یا اصلاح و یا تکمیل داده‌ها استفاده نمود. پیش‌بینی جریان در رودخانه‌ها با دقت مناسب یکی از مهم‌ترین ارکان در مدیریت منابع آب‌های سطحی است. تاکنون روابط و الگوهای گوناگون و پیچیده‌ای برای پیش‌بینی میزان آبدهی رودخانه‌ها نظیر انواع الگوهای مفهومی بارش- رواناب، الگوهای سری زمانی و الگوهای ترکیبی (هیبرید) ارائه شده است، اما روابط ارائه شده به‌علت عدم شناخت دقیق و پیچیدگی عوامل مؤثر در آبدهی رودخانه‌ها، در بسیاری از موارد با مقادیر مشاهده شده تطابق نداشت و در مواقعی نیز میزان مقادیر محاسبه شده از روابط گوناگون تفاوت معنی‌داری با یکدیگر داشته‌اند. مطابق تقسیم‌بندی گووینداراجو (۲۰۰۰) مدل‌هایی که امروزه در هیدرولوژی مورد استفاده قرار می‌گیرند، شامل مدل‌های ریاضی- فیزیکی، مدل‌های ژئومورفولوژیکی و مدل‌های تجربی می‌باشند. دسته اول مدل‌ها، بر اساس خصوصیات فیزیکی سیستم هستند که به‌صورت معادلات دیفرانسیل بیان می‌شوند. اما دسته دوم برمبنای خصوصیات ژئومورفولوژیکی سیستم هیدرولوژیکی مورد نظر بیان می‌شوند. این نوع از مدل‌ها توانایی زیادی در شبیه‌سازی رفتار حوضه آبریز و شبکه‌های آبیاری و زهکشی دارند (۱۲). در این زمینه می‌توان به پژوهش‌های ادیب و همکاران (۲۰۱۰a، ۲۰۱۰b، ۲۰۱۱) اشاره کرد (۱، ۲، ۳). مدل‌های تجربی بدون توجه به پارامترها، سعی در ایجاد رابطه‌ای بین داده‌های ورودی و خروجی دارند که این مدل‌ها به مدل‌های جعبه سیاه یا میانگینی معروف هستند.

در سال‌های اخیر کاربرد روش‌های هوش مصنوعی از جمله منطق فازی و شبکه عصبی، در زمینه‌هایی که در آن‌ها ارتباط بین ورودی و خروجی غیرخطی بوده، کاربرد گسترده‌ای پیدا کرده‌اند. این

www.SID.ir

Archive of SID

روش‌ها به‌عنوان یک جعبه‌سیاه مناسب که کم‌تر در قید و بند مسایل فیزیکی بوده و قادرند فرآیند غیرخطی و غیرایستای جریان رودخانه را بدون نیاز به مدل‌سازی عامل‌های محیطی و ژئومتری مؤثر بر جریان رودخانه مدل‌سازی کنند، می‌باشند (۹). سلاجقه و همکاران (۲۰۱۰) 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 مورد بررسی قرار گرفت. در نهایت نتایج پژوهش نشان داد که مدل‌های سری زمانی عملکرد بهتری نسبت به مدل‌های نزدیک‌ترین همسایه داشته‌اند (۲۲). لوهانی و همکاران (۲۰۱۲) به مقایسه عملکرد مدل‌های ANFIS، ANN و خودهمبسته در پیش‌بینی جریان ماهانه ورودی به مخزن سدی در هندوستان پرداختند. یافته‌های آن‌ها بیانگر از برتری مدل ANFIS در مقایسه با مدل‌های ANN و خودهمبسته به‌ازای همه الگوهای مورد استفاده به‌عنوان ورودی مدل بود (۱۸). جین و کومار (۲۰۰۷) عملکرد شبکه‌های عصبی مصنوعی و مدل‌های سری زمانی را در پیش‌بینی مقادیر دبی ماهانه رودخانه کلرادو آمریکا برایسه سناریو مختلف مورد بررسی قرار دارند. نتایج این پژوهش نشان داد که در هر سه سناریو با افزایش تعداد ورودی‌ها برای شبکه عصبی و افزایش مرتبه مدل در مدل‌های اتورگرسیو (AR) کارایی مدل‌ها بهتر شده است. همچنین در هر سه سناریو کارایی شبکه عصبی بهتر از مدل‌های AR بوده است (۱۳). کیسی (۲۰۰۸) با بهره‌گیری از الگوهای ضمنی شبکه عصبی مصنوعی اقدام به پیش‌بینی جریان‌های روزانه و ماهانه رودخانه‌های گوسو در ژاپن، بلک‌واتر و گیلا در آمریکا و رودخانه فلیوس در ترکیه و مقایسه نتایج با روش‌های سری زمانی نمود و نشان داد که شبکه عصبی مصنوعی در مقایسه با این روش‌ها از دقت بالاتری برخوردار است (۱۶). نبی‌زاده و همکاران (۲۰۱۲) در پژوهشی از مدل‌های مبتنی بر منطق فازی شامل سیستم استنتاج فازی و سیستم استنتاج تطبیقی عصبی- فازی به‌منظور پیش‌بینی جریان روزانه رودخانه استفاده کردند. در این پژوهش از سه پارامتر بارندگی، دما و دبی روزانه حوضه آبریز لیقوان‌چای برای پیش‌بینی جریان روزانه رودخانه استفاده شد. ارزیابی نتایج پیش‌بینی‌ها نشان داد که مدل ANFIS با دقت بالاتر و پراکندگی کم‌تر نسبت به مدل FIS دبی این رودخانه را پیش‌بینی کرده است (۲۰).

توانایی مدل‌های حوضه آبریز برای شبیه‌سازی دقیق فرآیندهای هیدرولوژی با آنالیز حساسیت پارامترها، واسنجی مدل و صحت‌سنجی مدل ارزیابی می‌گردد. مدل بارش- رواناب IHACRES تاکنون به‌صورت وسیعی مورد بررسی قرار گرفته است (۷). پژوهش‌ها نشان می‌دهد که پیش‌بینی جریان توسط IHACRES در بسیاری از موارد بهتر از مدل‌های توزیعی بوده است (۱۷، ۲۸). کاربرد این مدل در هر دو دوره واسنجی و ارزیابی آسان می‌باشد. همچنین داده‌های ورودی به آسانی قابل دسترس بوده و محاسبات کوتاه می‌باشند.

استفاده از دبی ماهانه در محاسبه و تحلیل خشکسالی‌های هیدرولوژیک و مدیریت منابع آب کاربرد فراوانی دارد. فتحی (۲۰۱۶) بر اساس دبی ماهانه ایستگاه ارازکوسه به تحلیل خشکسالی هیدرولوژیک در این منطقه پرداخت و نتیجه گرفت دبی رودخانه ارازکوسه رو به کاهش است و منطقه به‌سمت خشک شدن پیش می‌رود (۱۱). تیموری (۲۰۱۴) نیز از داده‌های دبی ماهانه و بارش ماهانه جهت ارتباط دادن خشکسالی‌های هیدرولوژیک و

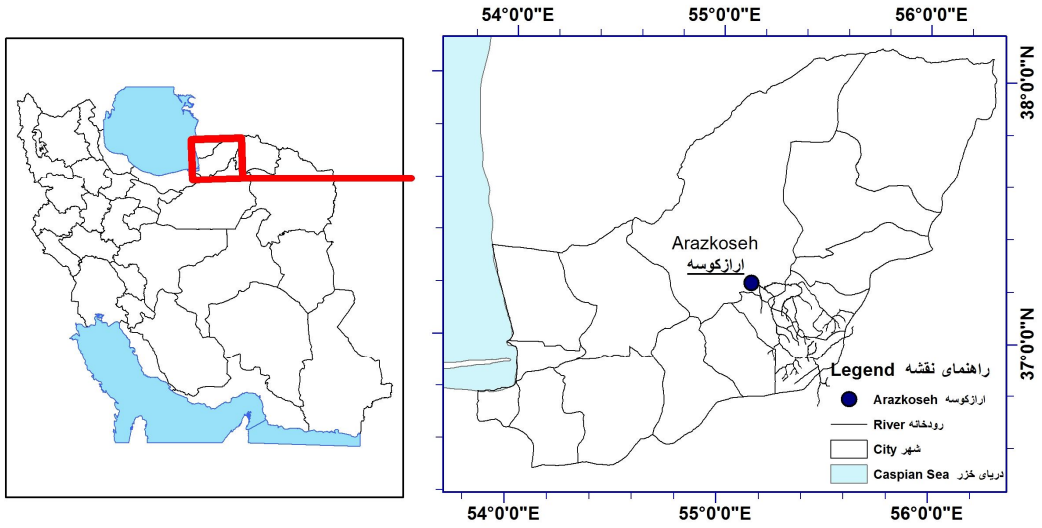
www.SID.ir

Archive of SID

هواشناسی استفاده کرد و نتیجه گرفت تطابق کاملی بین خشکسالی هواشناسی و هیدرولوژیکی وجود ندارد و با توجه به وضعیت خشکسالی در دروه‌های قبل، جریان رودخانه رفتار متفاوتی را نشان می‌دهد. همچنین کمبود بارش با زمان تاخیر کم‌تری نسبت به بیشبود بارش بر جریان رودخانه تأثیر می‌گذارد (۲۵). بر این اساس ضرورت ایجاب می‌کند تا از روش‌های مختلف دبی ماهانه رودخانه‌ها برآورد شود. استان گلستان از استان‌های شمالی کشور ایران است که هم خطر سیل و هم خطر خشکسالی آن را تهدید می‌کند. در زمینه پیش‌بینی وقایع سیلابی به کمک مدل‌های بارش- رواناب تلاش‌های زیادی انجام شده است. اما هدف از انجام این پژوهش، پیش‌بینی جریان ماهانه جهت استفاده در تحلیل خشکسالی‌های هیدرولوژیکی می‌باشد. نتایج پیش‌بینی دبی ماهانه به کمک مدل هیدرولوژیکی IHACRES با مدل‌های پایه داده‌ای بررسی و مقایسه می‌شوند.

مواد و روش‌ها

منطقه مطالعاتی: منطقه مطالعاتی این پژوهش ایستگاه ارازکوسه می‌باشد که بر روی رودخانه چهل‌چای از شاخه فرعی رودخانه گرگانرود در مجاورت شهر گنبد قرار داشته و دارای موقعیت جغرافیایی به طول ۵۵ درجه و ۸ دقیقه، عرض جغرافیایی ۳۷ درجه و ۱۳ دقیقه، ارتفاع ۳۴/۵ متر از سطح آب‌های آزاد و مساحت حوضه آبخیز ۱۶۷۸/۱ کیلومترمربع می‌باشد و دارای تجهیزات اشل، لیمنوگراف و پل تلفریک می‌باشد. برای انجام این پژوهش از آمار۲۹ ساله (۲۰۱۳-۱۹۸۵) بارندگی ایستگاه هواشناسی و دبی رودخانه ایستگاه هیدرومتری ارازکوسه در مقیاس زمانی ماهانه استفاده شد. پس از کنترل کیفی و تصحیح و تکمیل آمار با روش‌های آزمون جرم مضاعف و رابطه همبستگی، اقدام به پیش‌بینی دبی ماهانه به کمک مدل‌های هیدرولوژیکی و پایه داده‌ای شد.



شکل ۱- موقعیت منطقه مورد مطالعه.

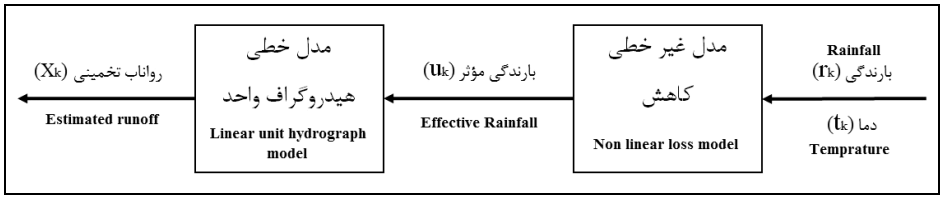
Figure 1. Location of the study area.

مدل IHACRES: IHACRES یک مدل یکپارچه مفهومی برای شبیه‌سازی بارش- رواناب می‌باشد که توسط جیکمن در سال ۱۹۹۰ توسعه یافت (۵). این مدل به پنج تا هفت متغیر برای واسنجی نیاز دارد و برای اجرا در حوضه‌های بزرگ مناسب است (۱۳). در این مطالعه از نسخه IHACRES v2.1 استفاده

www.SID.ir

Archive of SID

شده است که برای حوضه‌های دارای داده‌های پیوسته بارش، دما و رواناب کاربرد دارد. این مدل مطابق شکل ۲ شامل دو بخش به هم پیوسته غیرخطی و خطی است که به‌ترتیب برای محاسبه تلفات و تبدیل بارش مؤثر به رواناب تعریف شده است (۶، ۱۹).



شکل ۲- فلوچارت شبیه‌سازی بارش- رواناب همراه با مدول‌های خطی و غیرخطی در روش ارائه شده توسط جکمن و هورنبرگر (۱۹۹۳) (۱٤).

Figure 2. Rainfall-runoff simulations flowchart with linear and nonlinear modules in the provided method by Jakeman and Hornberger (14).

تبدیل بارش به بارش مؤثر: در این بخش، بارش مؤثر از حاصل‌ضرب بارش کل در شاخص رطوبت خاک حوزه درهر بازه زمانی محاسبه می‌شود (رابطه ۱):

$$u_k = [c(\Phi_k - l)]^p r_k \qquad (۱)$$

که در آن، c ضریب تعادل حجم بارش، l آستانه شاخص رطوبت خاک، p فاکتور واکنش غیرخطی و r_k بارش مشاهداتی می‌باشد. Φ_k شاخص رطوبت خاک است که با استفاده از رابطه ۲ محاسبه می‌گردد.

$$\Phi_k = r_k + \left(1 - \frac{1}{\tau_k}\right)\Phi_{k-1} \qquad (۲)$$

که در آن، τ_k برابر شدت خشکی خاک و به‌عنوان تابعی از دماست که با استفاده از رابطه ۳ محاسبه می‌شود:

$$\tau_k = \tau_w \exp\left(f\left(T_{ref} - T_k\right) \times 0.062\right) \qquad (۳)$$

که در آن، τ_w شدت خشکی خاک مبنا، f تابع تعدیل دما (تأثیر تغییر یک واحد دما بر میزان تلفات)، T_{ref} دمای مرجع و T_k دما در بازه زمانی مورد نظر هستند. بعد از محاسبه بارش مؤثر، هیدروگراف واحد کل با استفاده از بخش خطی در مدل محاسبه می‌گردد.

به‌عنوان نقاط ضعف مدل IHACRES هم می‌توان به این نکته اشاره کرد که بارش مؤثر در یک حوضه فقط تابع رطوبت خاک نیست بلکه به پارامترهای پوشش و به‌خصوص شیب حوضه هم بستگی دارد.

تبدیل بارش مؤثر به رواناب (بخش خطی): بخش خطی دارای سه پارامتر τ_q، τ_s و V_s می‌باشد. ترکیب دو مؤلفه جریان سریع $X_k^{(q)}$ و جریان آهسته $X_k^{(s)}$ منجر به تولید رواناب X_k می‌شود که با استفاده از رابطه‌های (۴ تا ۶) محاسبه می‌گردد (۲۴).

$$X_k = X_k^{(q)} + X_k^{(s)} \qquad (۴)$$

$$X_k^{(q)} = -\alpha_q X_{k-1}^{(q)} + \beta_q U_k \qquad (۵)$$

$$X_k^{(s)} = -\alpha_s X_{k-1}^{(s)} + \beta_s U_k \qquad (۶)$$

که در آن، α_q و β_q به‌ترتیب، ثابت زمانی جریان سریع و α_s و β_s به‌ترتیب ثابت زمانی جریان آهسته می‌باشند که با استفاده از رابطه‌های ۷ و ۸ به‌دست می‌آیند.

www.SID.ir

Archive of SID

$$\tau_q = \frac{-\Delta}{\ln(-\alpha_q)} \quad (۷)$$

$$\tau_s = \frac{-\Delta}{\ln(-\alpha_s)} \quad (۸)$$

که در آن، Δ بازه زمانی و τ_q و τ_s ثابت زمانی فروکش برای جریان سریع و آهسته در مخازن متوالی (بر حسب روز) می‌باشند. نسبت حجمی جریان سریع و آهسته به‌صورت رابطه ۹ در نظر گرفته می‌شوند:

$$V_q = 1 - V_s = \left(\frac{\beta_q}{1+\alpha_q}\right) = 1 - \frac{\beta_s}{1+\alpha_s} \quad (۹)$$

داده کاوی: داده‌کاوی فرآیندی است که ابزارهای مختلف تحلیل داده را به‌کار می‌گیرد تا الگوها و روابط فیزیکی متغیرها را در مجموعه داده‌های مختلف کشف کند (۲۶). در داده‌کاوی الگوریتم‌های متعددی به‌منظورهای مختلف مانند طبقه‌بندی و پیش‌بینی مورد استفاده قرار می‌گیرند. در این پژوهش دو روش مدل درخت تصمیم M5 و نزدیک‌ترین K- همسایگی مورد استفاده قرار گرفت.

مدل درختی M5: مدل درختی M5 که توسط کوئینلن (۱۹۹۲) (۲۱) ارایه شد زیرمجموعه‌ای از روش‌های یادگیری ماشینی و داده‌کاوی است. داده‌کاوی به فرآیند جست و جو و کشف مدل‌های گوناگون، مختصرسازی‌ها و اخذ مقادیر از مجموعه‌ای از مقادیر معلوم اطلاق می‌گردد (۴). روش‌های داده‌کاوی برای مجموعه داده‌های بزرگ با متغیرهای زیاد ساخته شده‌اند، بنابراین متفاوت از روش‌های آماری قدیمی هستند که برای مجموعه داده‌های کوچک با متغیرهای اندک طراحی شده‌اند. روش‌های بر مبنای درخت[1] یکی از تکنیک‌های داده‌کاوی است که در این روش‌ها خروجی به‌صورت یک مدل با سازه درختی با استفاده از داده‌های ورودی و خروجی می‌باشد.

1- Tree-Based Methods

الگوریتم M5 رایج‌ترین طبقه‌بندی استفاده شده در خانواده مدل تصمیم‌گیری درختی است. مدل درختی M5 یک الگوریتم پیش‌بینی عددی است و گره‌های درخت با خواص بیشینه خطاهای مورد انتظاری که به‌عنوان تابعی از انحراف استاندارد پارامترهای خروجی می‌باشد انتخاب می‌شوند. یک درخت تصمیم معمولاً از چهار بخش ریشه[2]، شاخه[3]، گره‌ها[4] و برگ‌ها[5] تشکیل شده است که گره‌ها با دایره نشان داده می‌شوند و شاخه‌ها نشان‌دهنده اتصال بین گره‌ها می‌باشند (۱۰). اولین مرحله برای ایجاد یک مدل درختی، استفاده از یک معیار انشعاب است. معیار انشعاب برای الگوریتم M5 بر اساس عملکرد انحراف استاندارد مقادیر هر کلاس و یا طبقه است که در هر گره به‌دست آمده است. این روش اساس روش‌های طبقه‌بندی است که آنتروپی[6] نامیده می‌شود. آنتروپی می‌تواند به‌عنوان معیار میزان آشفتگی و بی‌نظمی یک سیستم تفسیر شود. شکل ۳ نشان می‌دهد که یک ساختار درختی از روش‌های آموزشی مربوط به M5 دامنه پارامتر ورودی ۱ و ۲ می‌باشد. معیار انشعاب بیانگر میزان خطا در آن گره می‌باشد و مدل حداقل خطای مورد انتظار را به‌عنوان نتیجه آزمایش هر صفت در آن گره محاسبه می‌کند. خطای مدل عموماً با اندازه‌گیری دقت پیش‌بینی مقادیر هدف موارد دیده‌نشده سنجش می‌شود. فرمول محاسبه کاهش انحراف استاندارد (SDR) به‌صورت رابطه‌های ۱۰ و ۱۱ می‌باشد.

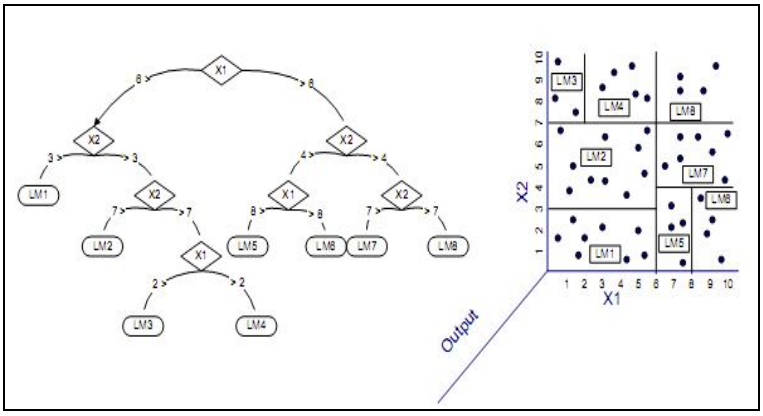
$$SDR = sd(T) - \sum_{i=1}^{N} \frac{|T_i|}{|T|} sd(T_i) \quad (۱۰)$$

$$sd(T) = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right)} \quad (۱۱)$$

2- Root
3- Beach
4- Nodes
5- Leafs
6- Entropy

www.SID.ir

Archive of SID



شکل ۳- نمونه‌ای از مدل درختی M5.

Figure 3. An example of M5 tree model.

الگوریتم نزدیک‌ترین k- همسایگی: 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 مورد استفاده قرار می‌گیرد. نحوه عملکرد این الگوریتم بر اساس مشاهدات و نمونه‌ها می‌باشد. بر اساس این الگوریتم می‌توان یک نمونه جدید را بر اساس اکثریت k گروه و دسته که نزدیک‌ترین همسایگی‌ها را با آن نمونه داشته باشند، تقسیم‌بندی نمود. به‌عبارت دیگر می‌توان گفت این روش k تعداد از الگوهای مشابه را پیدا کرده و بر اساس آن‌ها ارزش نمونه مورد مطالعه را پیش‌بینی می‌کند.

منطق مورد استفاده در این روش به این صورت است که با اندازه‌گیری و مشاهده متغیرهای واقعی مدل به جستجوی حالت‌های مشابه به حالت فعلی در سری تاریخی می‌پردازد. وقایعی که در سری تاریخی بعد از این حالت‌ها پیش آمده‌اند می‌توان به‌عنوان گزینه‌های محتمل در شرایط فعلی نیز در نظر گرفته شوند. احتمال وقوع هر یک از این حالت‌ها در شرایط حاضر بستگی به شباهت بردار مشاهداتی متغیرهای فعلی با بردار متغیرهای مستقل مشاهداتی در سری تاریخی دارد (۱۵).

در روش آماری ناپارامتری K-NN دو عامل مهم دخیل است. اولین عامل، انتخاب تابع فاصله‌سنجی و وزن‌های مربوطه و دومی انتخاب بهترین تعداد همسایگی است. در انجام پیش‌بینی‌ها توسط K-NN استفاده از تابع فاصله‌سنجی اقلیدسی وزن‌دار شده[1] بسیار معمول است. روش K-NN با اختصاص وزن‌های بیش‌تر به همسایگی‌های منتخبی که از نظر فاصله زمانی نسبت به زمان حاضر نزدیک‌تر می‌باشند توانسته است نتایج پیش‌بینی را نسبت به اتفاقات همسایگی‌های نزدیک‌تر به زمان فعلی وابسته‌تر گردند (۲۳، ۲۷).

معیارهای ارزیابی مدل‌ها: در این پژوهش در هر مرحله برای تعیین بهترین ساختار مدل و مقایسه مدل‌های مختلف از معیارهای ارزیابی ضریب همبستگی (r)، ریشه میانگین مربعات خطا (RMSE)، ناش ساتکلیف (NSE) و MBE استفاده شد. مقدار RMSE، نشان‌دهنده بزرگی خطا، MBE انحراف خطا از خط نیمساز می‌باشد.

$$r = \frac{\sum_{i=1}^{N}(Qobs - \overline{Q}_{obs})(Qsim - \overline{Q}_{sim})}{\sqrt{\sum_{i=1}^{N}(Qobs - \overline{Q}_{obs})^2 \sum_{i=1}^{N}(Qsim - \overline{Q}_{sim})^2}} \qquad (۱۲)$$

1- Weighted Euclidean

www.SID.ir

Archive of SID

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_{sim} - Q_{obs})^2}{N}} \quad (۱۳)$$

$$MBE = \frac{\sum_{i=1}^{N}(Q_{sim} - Q_{obs})}{N} \quad (۱۴)$$

$$NSE = \left[1 - \frac{\sum_{i=1}^{N}(Q_{obs} - Q_{sim})^2}{\sum_{i=1}^{N}(Q_{obs} - \overline{Q}_{obs})^2}\right] \quad (۱۵)$$

که در آن‌ها، $\overline{Q}_{obs}$ و $\overline{Q}_{sim}$ به‌ترتیب متوسط دبی مشاهداتی و شبیه‌سازی شده بر حسب مترمکعب بر ثانیه می‌باشد.

نتایج

در این مطالعه، از داده‌های ۱۶ سال اول (۲۰۰۰-۱۹۸۵) به‌منظور آموزش و از ۱۳ سال بقیه (۲۰۱۳-۲۰۰۱) به‌منظور آزمون استفاده شد. پارامترهای آماری ماهانه داده‌های جریان در جدول ۱ درج شده است.

جدول ۱- پارامترهای آماری ماهانه جریان.

Table 1. Statistical parameters of monthly streamflow.

	دوره آماری Statistical period	میانگین Average (m^3/s)	حداکثر Maximum (m^3/s)	حداقل Minimum (m^3/s)	انحراف معیار SD (m^3/s)	ضریب تغییرات Coefficient Variation
آموزش Train	1985-2000	5.23	25.83	0.004	5.13	0.97
آزمون Test	2001-2013	4.63	22.62	0.039	4.97	1.08

جهت پیش‌بینی جریان در ماه جاری از داده‌های جریان در ماه‌های قبل به‌عنوان ورودی در مدل‌های مورد بررسی استفاده گردید. ترکیبات ورودی‌های مختلف در مدل‌ها عبارتند از:

1) $Q_t=f(P,T,Q)$
2) $Q_t=f(P,T,Q,Q_{t-1})$
3) $Q_t=f(P,T,Q,P_{t-1},Q_{t-1})$
4) $Q_t=f(P,T,Q,P_{t-1},T_{t-1},Q_{t-1})$
5) $Q_t=f(P,T,Q,P_{t-2},T_{t-2},Q_{t-2})$
6) $Q_t=f(P,T,Q,P_{t-3},T_{t-3},Q_{t-3})$

که در این مدل‌ها، P بارش ماه، T میانگین دمای ماه، Q میانگین دبی ماه و اندیس‌های t-L نیز مقادیر این پارامترها برای L ماه قبل می‌باشد.

نتایج مدل هیدرولوژیکی IHACRES: در این پژوهش برای پیش‌بینی دبی جریان ماهانه با استفاده از مدل‌های هیدرولوژیکی از مدل IHACRES استفاده شده است. شکل ۴ و جدول ۲ وضعیت عملکرد مدل IHACRES را در دو دوره واسنجی و صحت‌سنجی نشان می‌دهد. بر اساس نتایج به‌دست آمده مدل IHACRES با ضریب همبستگی ۰/۸۱ در دوره واسنجی و ۰/۷۹ در دوره صحت‌سنجی دارای توانایی لازم برای شبیه‌سازی‌های مقادیر روزانه دبی حوضه ارازکوسه است. باید توجه داشت که بر اساس نتایج شکل ۲ مدل توانایی کم‌تری در شبیه‌سازی دبی‌های حداکثر دارد، به‌طوری‌که مدل دبی‌های کم‌تری را نسبت به دبی‌های حداکثر مشاهداتی شبیه‌سازی کرده است.

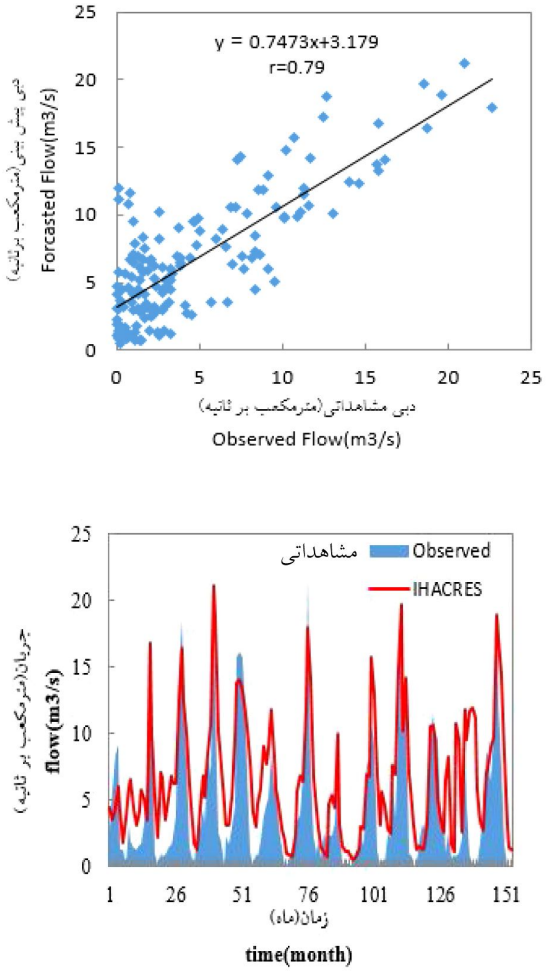
www.SID.ir

Archive of SID

جدول ۲- مشخصات آماری مدل IHACRES.

Table 2. Statistical properties of IHACRES.

NS	MBE	RMSE	r	
0.61	0.52	3.2	0.81	واسنجی (calibration)
0.47	2	3.6	0.79	صحت‌سنجی (validation)



شکل ٤- مقادیر مشاهداتی و پیش‌بینی‌شده دبی ماهانه رودخانه با مدل IHACRES.

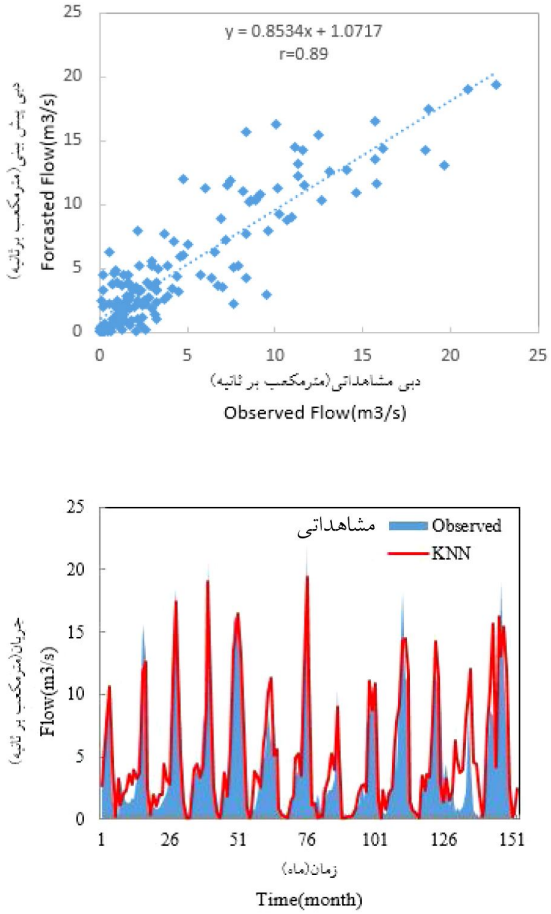
Figure 4. Observed and forecasted monthly streamflow using IHACRES model.

نتایج مدلسازی دبی ماهانه با استفاده از الگوریتم نزدیک‌ترین k- همسایگی: در گام اول عملکرد روش K-NN در پیش‌بینی جریان مورد بررسی قرار گرفت. به‌طوری‌که قبلاً اشاره شد، در این روش پیش‌بینی با توجه به تعداد تاخیرها و همچنین شعاع همسایگی انجام می‌شود. البته انتخاب منطقی تعداد همسایگی نیز از پارامترهای مهم در پیش‌بینی، توسط این روش می‌باشد. در شکل ۵، تغییرات و پراکندگی داده‌های مشاهداتی و پیش‌بینی شده جریان توسط مدل نزدیک‌ترین همسایگی به‌ازای بهینه‌ترین ساختار

www.SID.ir

Archive of SID

مدل (سناریوی سوم) نشان داده شده است. در این حالت مقادیر شاخص‌های آماری شامل r، RMSE، MBE و NSE به‌ترتیب برابر ۰/۸۸۹، ۲/۳۵، ۳/۹۰ و ۰/۷۸۰ می‌باشد. نتایج الگوریتم نزدیک‌ترین همسایگی برای ترکیب‌های مختلف ورودی در پیش‌بینی جریان در جدول ۳ آورده شده است.



شکل ۵- مقادیر مشاهداتی و پیش‌بینی‌شده دبی ماهانه رودخانه با مدل K-NN.

Figure 5. Observed and forecasted monthly streamflow using K-NN model.

جدول ۳- مشخصات آماری الگوریتم نزدیکترین k- همسایگی.

Table 3. Statistical properties of K nearest neighborhood.

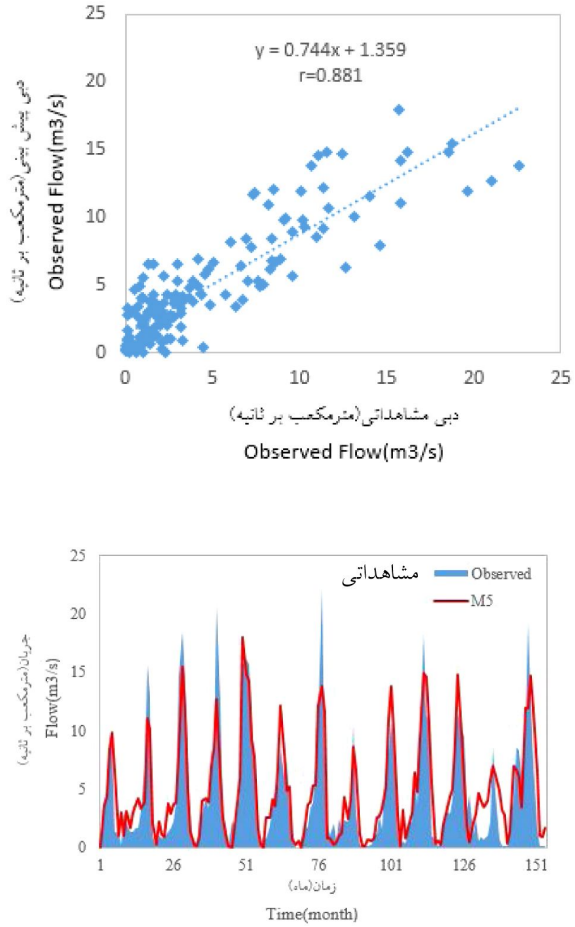
آزمون (Test)				آموزش (Train)				سناریو Scenario
NS	MBE	r	RMSE	NS	MBE	r	RMSE	
0.775	0.148	0.878	2.34	0.858	0.083	0.932	1.88	1
0.779	0.151	0.886	2.35	0.863	0.075	0.928	1.91	2
0.780	0.390	0.889	2.35	0.878	-0.019	0.937	1.80	3
0.777	0.108	0.883	2.36	0.855	0.011	0.924	1.97	4
0.747	0.328	0.869	2.52	0.826	-0.051	0.909	2.15	5
0.775	0.099	0.880	2.37	0.806	-0.029	0.898	2.28	6

www.SID.ir

Archive of SID

با توجه به شکل ۵ می‌توان دریافت کـه عملکـرد مدل در پیش‌بینی مقادیر حداقل جریان به مراتب بهتر از پیش‌بینی مقادیر حداکثر جریان بوده است کـه ایـن مسأله ناشی از وجود مقـادیر دبـی متنـاظر بـا مقـادیر حداقل جریان در همـسایگی نقطـه مـورد پـیش‌بینـی می‌تواند باشد که در عملکرد مدل تأثیر زیادی دارد.

نتایج مدلسازی دبی ماهانه با استفاده از مدل درختی M5: در گام بعـدی، عملکـرد مـدل درختـی M5 در پـیش‌بینـی جریـان ماهانـه بـه‌ازای الگوهـای ورودی مختلف ارزیابی گردید. با توجه به نتایج به‌دست آمده از مدل M5 کـه در جـدول ۴ آمـده اسـت، ملاحظـه می‌شود که این روش نیز عملکرد مناسبی در پیش‌بینی جریان ماهانه حوضه آبریز ارازکوسه داشته اسـت. در شکل ۶، تغییرات و پراکندگی داده‌هـای مـشاهداتی و پیش‌بینی شده جریان توسط مدل درختی M5 بـه‌ازای بهینه‌ترین ساختارمدل (الگوی چهارم) نشان داده شده است. در این حالت مقادیر شاخص‌های آماری شـامل r، RMSE، MBE و NSE بـه‌ترتیـب برابـر ۰/۸۸، ۲/۳۸، ۰/۱۷۵ و ۰/۷۷۴ می‌باشد.



شکل ٦- مقادیر مشاهداتی و پیش‌بینی‌شده دبی ماهانه رودخانه با مدل M5.

Figure 6. Observed and forecasted monthly streamflow using M5 model.

www.SID.ir

Archive of SID

جدول ٤- مشخصات آماری مدل درختی M5.

Table 4. Statistical properties of M5 tree model.

سناریو Scenario	آموزش (Train)				آزمون (Test)			
	RMSE	r	MBE	NS	RMSE	r	MBE	NS
1	2.15	0.909	-0.027	0.823	2.48	0.849	0.252	0.762
2	2.16	0.911	-0.025	0.825	2.59	0.858	0.248	0.756
3	2.13	0.914	-0.02	0.830	2.61	0.855	0.196	0.729
4	2.06	0.920	-0.017	0.840	2.38	0.881	0.175	0.774
5	2.30	0.897	0.068	0.801	2.76	0.841	0.536	0.695
6	2.32	0.894	0.011	0.799	2.73	0.847	0.479	0.703

نتیجه‌گیری

در این پژوهش از داده‌های ایستگاه ارازکوسه به‌عنوان ورودی جهت پیش‌بینی جریان ماهانه با استفاده از سه روش IHACRES، KNN و M5 استفاده گردید. با توجه به مقدار بالای ضریب همبستگی حاصل از دو روش KNN و M5 نسبت به روش هیدرولوژیکی IHACRES نتیجه می‌شود که در حوضه ارازکوسه مدل‌های پایه داده‌ای عملکرد بهتری نسبت به مدل‌های هیدرولوژیکی دارند. روش KNN به‌دلیل سادگی و عدم پیچیدگی بالا می‌تواند همواره یکی از بهترین گزینه‌ها برای انجام پیش‌بینی در علوم مختلف به‌خصوص هیدرولوژی و هواشناسی باشد. این روش برای شرایطی که دارای سری‌های تاریخی بلندمدت است دارای ارجحیت بالاتری خواهد بود زیرا تأثیر وقایع متنوع حدی و غیرحدی در پیش‌بینی‌های آتی می‌تواند بسیار مؤثر باشد. همچنین مدل درختی M5 به لحاظ سادگی، قابل فهم بودن و ارائه روابط خطی ساده در محدوده خاصی از داده‌های ورودی که به‌صورت روابط خطی ارائه می‌گردد، می‌تواند به‌عنوان یک روش جایگزین قابل رقابت با سایر روش‌ها مطرح شود.

با توجه به نتایج به‌دست آمده از سه مدل به‌وضوح روشن است که مدل‌ها توانایی خوبی در برآورد جریان حوضه ارازکوسه دارند. هر چند نتایج شاخص‌های آماری بیانگر از برتری نسبی مدل KNN نسبت به IHACRES و M5 می‌باشد.

جدول ٥- مقایسه مدل‌های هیدرولوژیکی و داده‌کاوی در پیش‌بینی و شبیه‌سازی دبی ماهانه.

Table 5. Hydrologic and data mining models comparison in monthly discharge estimation and prediction.

	r	RMSE	MBE	NS
واسنجی (۲۰۰۰-۱۹۸۵) Calibration (1985-2000)				
IHACRES	0.81	3.2	0.52	0.61
KNN	0.937	1.80	-0.019	0.878
M5	0.920	2.06	-0.017	0.840
صحت‌سنجی (۲۰۱۳-۲۰۰۱) Validation (2001-2013)				
IHACRES	0.79	3.6	2	0.47
KNN	0.889	2.35	0.390	0.780
M5	0.881	2.38	0.175	0.774

www.SID.ir

Archive of SID

منابع

1.Adib, A., Salarijazi, M., and Najafpour, K. 2010a. Evaluation of Synthetic Outlet Runoff Assessment Models. J. Appl. Sci. Environ. Manage. 14: 3. 13-18.

2.Adib, A., Salarijazi, M., Shooshtari, M.M., and Akhonodali, A.M. 2011. Comparison between characteristics of geomorphoclimatic instantaneous unit hydrograph be produced by GcIUH based Clark Model and Clark IUH model. J. Mar. Sci. Technol. 19: 2. 201-209.

3.Adib, A., Salarijazi, M., Vaghefi, M., Mahmoodian-shooshtari, M., and Akhonali, A.M. 2010b. Comparison between GcIUH-Clark, GIUH-Nash, Clark-IUH, and Nash-IUH models. Turk. J. Engin. Environ. Sci. 34: 91-103.

4.Alikhanzadeh, A. 2007. Data mining, Edition 1, publishing of computer science, Babol. 344p. (In Persian)

5.Besaw, L.E., Rizzo, D.M., Bierman, P.R., and Hackett, W.R. 2010. Advances in ungauged stream flow prediction using artificial neural networks. Hydrology. 386: 27-37.

6.Crooke, B.F.W., Andrews, F., Spate, J., and Cuddy, S.M. 2005. IHACRES user guide. Technical Report 2005/19. Second Edition. iCAM, School of Resources, Environment and Society, The Australian National University, Canberra. http://www.toolkit.net.au/ihacres.

7.Croke, B.F.W., and Jakeman, A.J. 2008. Use of the IHACRES rainfall-runoff model in arid and semi-arid regions, P 41-48. In: H.S. Wheater, S. Sorooshian and K.D. Sharma (Eds.), Hydrological Modelling in Arid and Semi-arid Areas. Cambridge University Press, Cambridge.

8.Croke, B.F.W., and Jakeman, A.J. 2004. A catchment moisture deficit module for the IHACRES rainfall-runoff model. Environmental Modelling and Software. 19: 1-5.

9.El-Shafie, A., RedaTaha, M., and Noureldin, A. 2007. A neuro-fuzzy model for inflow forecasting of the Nile River at Aswan high dam. Water Resource Manage. 21: 533-556.

10.Fallahi, M., Varvani, H., and Golian, S. 2012. Forecast precipitation using regression tree for flood control. 5th conference of watershed and water resource management and land, Kerman. (In Persian)

11.Fathi, R. 2016. Spatial analysis of the hydrological drought. M.Sc. thesis, Gorgan University of Agricultural Sciences and Natural Resources, 99p. (In Persian)

12.Govindaraju, R.S. 2000. Artificial neural network in hydrology. I: Preliminary Concepts. J. Hydrol. Engin. 5: 2. 115-123.

13.Jain, A., and Kumar, A.M. 2007. Hybrid neural network models for hydrologic time series forecasting. Appl. Soft Comp. J. 7: 2. 585-592.

14.Jakeman, A.J., and Hornberger, G.M. 1993. How Much Complexity Is Warranted in a Rainfall-Runoff Model?. Water Resources Research. 29: 2637-2649.

15.Karamooz, M., and Araghinejad, Sh. 2005. Advanced Hydrology. Amirkabir University Press, 464p. (In Persian)

16.Kisi, O. 2005. Daily river flow forecasting using artificial neural networks and auto regressive models. Turk. J. Engin. Environ. Sci. 29: 9-20.

17.Littlewood, I.G., and Jakeman, A.J. 1994. A New Method of Rainfall-Runoff Modelling and its Applications in Catchment Hydrology. Environmental Modelling. 2: 142-171.

18.Lohani, A.K., Kumar, R., and Singh, R.D. 2012. Hydrological time series modeling: A comparison between adaptive neurofuzzy, neural network and autoregressive techniques. J. Hydrol. 442-443: 23-35.

19.McIntyre, N., and Al-Qurashi, A. 2009. Performance of ten rainfall-runoff models applied to an arid catchment in Oman. Environmental Modelling and Software. 24: 726-738.

20.Nabizadeh, M., Mosaedi, A., Hesam, M., and Dehghani, A.A. 2012. Comparing the performance of Fuzzy based models in stream flow on Lighvan River. J. Water Soil Cons. 19: 1. 117-134. (In Persian)

21.Quinlan, J.R. 1992. Learning with continuous classes. In: proceedings AI, 92 (Adams & Sterling, Eds), P 343-348. Singapore: World Scientific.

www.SID.ir

Archive of SID

22.Salajegheh, A., Fathabadi, A., and Gholami, H. 2010. Predict river discharge using the nearest neighbor. 5th national conference on science and management engineering Iran. Gorgan University of Agricultural Sciences and Natural Resources. (In Persian)
23.Sorjamaa, A., Reyhani, N., and Lendasse, A. 2005. Input and structure selection for K-NN approximator. 8th International Conference on Artificial Neural Networks, Lecture Notes in Computer Science Springer, IWANN, Berlin, Pp: 958-992.
24.Taesombat, W., and Sriwongsitanon, N. 2010. Flood Investigation in the Upper Ping river basin using mathematical models. Kasetsart Natural Science. 44: 152-166.
25.Teimoori, F. 2014. Comparative study if meteorological indices with hydrological indices for drought monitoring using data mining method. M.Sc. thesis, Gorgan University of Agricultural Sciences and Natural Resources, 95p. (In Persian)
26.Two Crows Corporation. 1999. Introduction to data mining and knowledge discovery, third ed., Postmac, MD. Available at: www.twocrows.com, (April 29, 2000).
27.Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K. 2003. A technique for generating regional climate scenarios using a nearest-neighbor algorithm. Water Resources Research. 39: 7. 1114-1121.
28.Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M., and Jakeman, A.J. 1997. Performance of Conceptual Rainfall-Runoff Models in Low-Yielding Ephemeral Catchments. Water Resources Research. 33: 1. 153-166.

www.SID.ir

Archive of SID

Gorgan University of Agricultural Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 23(1), 2016
http://jwsc.gau.ac.ir

Evaluation of hydrological and data mining models in monthly river discharge simulation and prediction (Case study: Araz-Kouseh watershed)

***Kh. Ghorbani[1], E. Sohrabian[2] and M. Salarijazi[3]**

[1]Associate Prof., Dept. of Water Engineering, Gorgan University of Agriculture Science and Natural Resources, [2]M.Sc. Graduate, Dept. of Water Resource Engineering, Gorgan University of Agriculture Science and Natural Resources, [3]Assistant Prof., Dept. of Water Engineering, Gorgan University of Agriculture Science and Natural Resources

Received: 02/09/2016; Accepted: 06/19/2016

Abstract

Background and Objectives: Quantitative prediction of river discharge is one of the most important elements in the management of surface water resources, especially to take suitable decisions in occurrence of floods and drought events. Various approaches introduced in hydrology to predict river discharge among them conceptual models as well as data-driven models are the most important ones. In this study, long term recorded data sets in Araz-Kouseh watershed with 1678 km^2 area located in northern Iran (Golestan province) were used to investigate the precision of different river discharge prediction models. The IHACRES model as a conceptual hydrological model and KNN and M5 as data mining models were selected for modeling of monthly river discharge and the results were compared to examine the accuracy of studied models. In some studies, the expressed models were used for daily river discharge prediction but the main objectives of this study are application of these models to predict monthly discharge for a watershed.

Materials and Methods: The 29 years (1985-2013) daily rainfall and discharge data belonging to Araz-Kouseh hydrometry and meteorological stations were used to extract monthly time series for modeling. The required quantity and quality conditions of datasets for modeling were confirmed using different statistical tests. Recorded datasets were divided in two subseries, first one was used for calibration period and second one used for validation of investigated models. The results of models in calibration and validation period were analyzed considering model efficiency goodness of fit criteria.

Results: The results of IHACRES conceptual hydrological model for both calibration and validation periods (correlation coefficients equal to 0.81 and 0.79 for calibration and validation periods respectively) showed suitable ability of this model to predict monthly river discharge. Moreover investigation of results of both M5 and KNN data mining models (correlation coefficient equal to 0.94 and 0.89 for calibration and validation periods respectively for KNN model and equal to 0.92 and 0.88 for calibration and validation periods respectively for M5 model) reveals that application of these models led to significant increase in prediction precision in comparison with IHACRES model.

Conclusions: The results of this study indicate the data mining models, i.e. M5 and KNN, outperform conceptual hydrological model, i.e. IHACRES, for prediction of monthly river discharge considering different goodness of fit criteria. It is clear that the accuracy of the prediction of data mining models are very close to each other but the M5 model is selected as best model in this study because of its explicit equations for prediction. Furthermore, investigation of time series of predicted river discharge show data mining models had better prediction for low discharges in comparison with high discharges.

Keywords: Monthly river discharge, Araz-Kouseh watershed, IHACRES, KNN, M5

* Corresponding Author; Email: ghorbani.khalil@yahoo.com

www.SID.ir