

بررسی تاثیر روش‌های خوشه‌بندی سلسله مراتبی در دقت مدل‌سازی ضریب جریان در حوضه کرخه

رحیم کاظمی*^۱ و جهانگیر پرهمت^۲

^۱ مربی، پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران و ^۲ دانشیار، پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

تاریخ پذیرش: ۱۳۹۶/۰۵/۳۱

تاریخ دریافت: ۱۳۹۵/۱۲/۰۴

چکیده

برآورد میزان ضریب جریان، که متأثر از عوامل مورفومتری، زمین‌شناسی و هیدرواقليمی است، همواره یکی از موضوعات مهم در هیدرولوژی بوده، اطلاع از میزان آن، نقش به‌سزایی در برنامه‌ریزی و مدیریت بهینه منابع آب دارد. همگن‌بندی حوزه‌های آبخیز مناسب‌ترین روش برای تحلیل پارامترهای هیدرولوژیکی در غیاب پوشش کامل داده‌های هیدرولوژی است. در این پژوهش، ابتدا با بررسی داده‌های دبی روزانه ایستگاه‌های آب‌سنجی و باران‌سنجی تعداد ۲۲ ایستگاه با آمار مناسب و دوره مشترک آماری، سال‌های آبی ۱۳۷۸-۱۳۵۳ انتخاب شد. همچنین، با استفاده از نقشه توپوگرافی با مقیاس ۱:۵۰۰۰۰ و تعیین موقعیت ایستگاه‌ها، محدوده مورد پژوهش مشخص و پارامترهای اولیه حوضه‌ها با استفاده از سامانه اطلاعات جغرافیایی استخراج شد. ضریب جریان با استفاده از روش جاستین و شاخص جریان پایه به روش فیلتر رقومی برگشتی تک پارامتره، محاسبه شد. واحدهای سنگ‌شناسی با استفاده از نقشه زمین‌شناسی با مقیاس ۱:۲۵۰۰۰۰ رقومی و بر اساس نظر کارشناسی به دو طبقه عمده نفوذپذیر و نفوذناپذیر تقسیم و مساحت تحت پوشش هر واحد در هر حوضه تعیین شد. تحلیل عاملی از بین ۱۵ پارامتر مؤثر در ضریب جریان انجام و همگن‌بندی حوضه‌ها با استفاده از عوامل مستقل انتخابی به روش‌های مختلف سلسله‌مراتبی شامل: نزدیک‌ترین همسایه، دورترین همسایه، فاصله از میانه، فاصله از مرکز ثقل و روش وارد انجام شد. سپس روابط منطقه‌ای به روش رگرسیون خطی، در سطوح معنی‌داری کمتر از یک درصد برای گروه‌های همگن تمام روش‌ها به‌دست آمد. ارزیابی صحت و کارایی مدل‌های برآوردی، به روش‌های آزمون استقلال خطاها، نرمال بودن توزیع خطاها و هم‌خطی انجام شد. سپس دقت مدل‌های استخراج شده با استفاده از ایستگاه‌های شاهد محاسبه و با یکدیگر مقایسه شد. نتایج آزمون تحلیل عاملی نشان داد که کلیه متغیرها در قالب پنج عامل طبقه‌بندی می‌شوند که در مجموع در برگیرنده ۸۵/۴۵۲ درصد از واریانس داده‌ها می‌باشند. نتایج همگن‌بندی نشان داد که دسته‌بندی حوضه‌ها در روش‌های همگن‌بندی نزدیک‌ترین همسایه، دورترین همسایه، فاصله از مرکز ثقل و فاصله از میانه، تماماً مشابه هم بوده و در دو گروه با اجزاء مشابه دسته‌بندی می‌شوند. ولی در روش وارد، حوضه‌ها به سه منطقه همگن تفکیک شد، نتایج بررسی دقت مدل‌ها نشان داد که دقت مدل‌های استخراج شده در گروه‌های همگن به روش نزدیک‌ترین همسایه به‌دلیل حداقل درصد خطای نسبی به میزان ۲۵/۴ درصد، میانگین مطلق خطا به میزان ۷/۸۵ و مجذور میانگین مربعات خطا به میزان ۹/۶۲ از سایر روش‌ها مناسب‌تر است و به‌عنوان بهترین روش همگن‌بندی برای تحلیل منطقه‌ای ضریب جریان در منطقه پژوهش تشخیص داده شد.

واژه‌های کلیدی: تحلیل منطقه‌ای، خطای برآورد، سلسله مراتبی، مدل‌سازی، همگن‌بندی

مقدمه

مدیریت پایدار منابع آب، نیازمند داده‌های مناسب است که محققین را در توسعه استراتژی مدیریت کمی و کیفی منابع آب یاری کند. یکی از مشکلاتی که در اغلب حوزه‌های آبخیز ایران وجود دارد، عدم وجود آمار و اطلاعات کامل و دقیق به دلیل نبود ایستگاه آب‌سنجی است. کمبود داده‌های مناسب در حوضه مربوط به مطالعات منابع آب، منجر به عدم قطعیت در بخش‌های مربوط به مدیریت این کمبود می‌شود. توجه به این مشکل منجر به نامگذاری دهه ۲۰۰۳-۲۰۱۲ به پیش‌بینی در حوزه‌های آبخیز فاقد آمار به‌وسیله انجمن بین‌المللی هیدرولوژی شد (Chahoki و همکاران، ۲۰۱۳).

با توجه به محدودیت داده‌های اندازه‌گیری شده، تعمیم داده‌ها به حوضه‌های فاقد آمار نیازمند پژوهش و بررسی جامع در خصوص ارتباط ویژگی‌های ثابت و پایدار حوضه‌ها با داده‌های جریان در حوضه‌های همگن، برای تعمیم روابط به حوضه‌های فاقد آمار می‌باشد (Smakhtin, ۲۰۰۱). خوشه‌بندی یکی از تکنیک‌های بسیار قدرتمند برای کشف گروه‌ها و وابستگی‌های طبیعی در یک مجموعه داده و همچنین، شناخت الگوهای ساختاری و موضوعی موجود در آن، بدون داشتن هر گونه پیش‌زمینه شناختی در مورد مشخصات و ویژگی‌های داده است (Michael و همکاران، ۲۰۰۷؛ Parajka و همکاران، ۲۰۰۵).

تاکنون پژوهش‌های مختلفی برای تجزیه و تحلیل منطقه‌ای مولفه‌های هیدرولوژیکی به انجام رسیده است که در هر یک از این پژوهش‌ها، روش‌های مختلف همگن‌بندی بر اساس توصیه محققین به‌کار برده شده است. از جمله این تحقیقات می‌توان به موارد ذیل اشاره کرد. Malekiyan و همکاران (۲۰۱۳) برای تعیین عرصه‌های مناسب پخش سیلاب در حوزه آبخیز گریباگان و Smithers و Kjeldsen (۲۰۰۲) در آفریقای جنوبی برای تجزیه و تحلیل منطقه‌ای سیلاب از روش همگن‌بندی سلسله مراتبی استفاده کرده‌اند. برای برآورد دبی بیشینه روزانه استان فارس روش آزمون همگنی لانگبین به‌وسیله Mosavi و Sepaskhah (۱۹۸۹) به‌کار گرفته شد. چهار روش

نزدیک‌ترین همسایه، دورترین همسایه، وارد و متوسط گروه به‌وسیله Rohani (۲۰۰۱) و Abbassizade (۲۰۰۳) مورد بررسی و مقایسه قرار گرفت و روش وارد را برای منطقه مورد پژوهش بهینه تشخیص داده‌اند. در پژوهشی در حوزه‌های آبخیز غرب کشور روش‌های سلسله مراتبی و آزمون همگنی لانگبین به‌وسیله Khosravi و همکاران (۲۰۱۱) مورد بررسی و مقایسه قرار گرفت. نتایج منتشر شده، بهترین روش همگن‌بندی برای تحلیل منطقه‌ای سیلاب را روش سلسله مراتبی گزارش کرده‌اند. بسیاری از محققین از جمله Biyanaki و Eslamiyan (۲۰۰۵) و Ataei و Shiran (۲۰۰۱) از روش تحلیل خوشه‌ای برای تعیین مناطق همگن بر اساس ویژگی‌های هیدرولوژیکی استفاده کردند. تحلیل منطقه‌ای شاخص جریان کم Q_{95} با استفاده از چهار روش مختلف همگن‌بندی در شمال غرب ایتالیا انجام و روش موسوم به CRT^1 به‌عنوان بهترین روش گزارش شد (Veza و همکاران، ۲۰۱۰).

پژوهش مشابهی نیز در خصوص مقایسه چهار روش همگن‌بندی به‌وسیله Laaha و Blöschl (۲۰۱۴) انجام شده که از نظر دقت برآورد، روش منطقه‌بندی فصلی را در درجه اول و روش CRT را در درجه دوم اهمیت تشخیص دادند.

طی پژوهشی Salinas و همکاران (۲۰۱۳) روش‌های مختلف تحلیل منطقه‌ای شاخص‌های کم و سیل را بررسی کردند و نتیجه گرفتند که دقت روش‌های همگن‌بندی شاخص‌های کم در مناطق خشک، کمتر از مناطق مرطوب و در حوضه‌های بزرگ بیشتر از حوضه‌های کوچک است. همچنین، نتیجه گرفتند که روش‌های رگرسیون در مقیاس منطقه‌ای بهتر از مقیاس جهانی جواب می‌دهد. مشخصه‌های هیدرولوژیکی حوضه‌ها مانند ضریب جریان و سایر مولفه‌های جریان به‌وسیله پارامترهای فیزیکی مختلف، قابل برآورد است. مشخصه‌های فیزیکی عمدتاً، شامل پارامترهای فیزیوگرافی بوده، ممکن است، پارامترهای زمین‌شناسی از دیدگاه هیدروژئولوژی نیز مورد استفاده قرار گیرد. در این میان یکی از مهم‌ترین

¹ Classification and regression tree

رواناب با مساحت زمین‌های اشباع و همچنین، ارتباط آن با آب موجود در خاک و جریانات زیرسطحی است (Boughton, ۱۹۸۷).

روابط بین ضریب جریان و پوشش گیاهی با استفاده از شاخص اختلاف پوشش گیاهی نرمال شده (NDVI) در دوره زمانی ۱۹۸۲-۱۹۹۹ و به‌منظور بررسی رفتار پوشش‌های مختلف در ضریب جریان به-وسیله Carrillo و همکاران (۲۰۰۶) در برزیل مورد بررسی قرار گرفته و مشخصاً به روابط مستقیم بین شاخص اختلاف پوشش گیاهی نرمال شده و ضریب جریان تأکید شده است. در پژوهش Cerdan و همکاران (۲۰۰۴) به‌منظور بررسی اثر مقیاس در تولید رواناب، ۳۴۵ واقعه بارش-رواناب در سه حوضه با اندازه‌های مختلف در فرانسه، مورد بررسی قرار گرفت. آن‌ها نتیجه گرفتند که ضریب جریان با افزایش مساحت کاهش محسوسی را نشان می‌دهد. در پژوهشی که توسط Visessri و McIntyre (۲۰۱۶) به-منظور تحلیل منطقه‌ای ضریب جریان با تمرکز بر بررسی تاثیر تغییر کاربری اراضی در تعداد ۴۴ حوضه دارای آمار در شمال تایلند انجام شد، مشخصاً به ارتباط کاربری اراضی و رطوبت پیشین خاک و ضریب جریان اشاره کردند.

اثر مشخصه‌های مختلف هیدرولوژیکی و کلیماتولوژیکی حوضه بر ضریب جریان در ۱۵ حوضه به‌وسیله Quang و Laituri (۲۰۱۳) مورد بررسی قرار گرفت. آن‌ها بر خلاف نتایج منتشر شده پژوهش سردان و همکاران در سال ۲۰۰۴ نتیجه گرفتند که اندازه حوضه تأثیر مهمی در ضریب جریان ندارد، ولی نرخ جریان با ضریب شکل، مستقماً ارتباط دارد و ۲۷ درصد نوسانات ضریب جریان در منطقه مورد پژوهش را به شکل حوضه ربط داده، اعلام کردند که شبیه تأثیری بر روی مولفه‌های جریان در حوضه مورد پژوهش ندارد. عوامل موثر بر ضریب جریان به‌وسیله Giudice و همکاران (۲۰۱۳) در جنوب ایتالیا مورد بررسی قرار گرفته و مشخصاً روابط بین ضریب جریان و بیشینه پتانسیل نگهداشت خاک، تشریح و تاثیر شرایط رطوبت خاک قبل از بارش با استفاده از عوامل

مشخصه‌های هیدروژئولوژیکی که در پژوهش‌های مختلف بر آن تأکید شده، ضریب جریان^۱ است (Zhang و همکاران، ۲۰۱۴؛ Gottschalk و Wringartner، ۱۹۹۸؛ Dos Reis Castro و همکاران، ۱۹۹۹؛ Kirkby و Beven، ۱۹۷۹).

ضریب جریان، نسبتی است که نشان‌دهنده بخشی از بارش است که به‌طور مستقیم به رواناب تبدیل می‌شود. مفهوم ضریب جریان برای اولین بار در ابتدای قرن بیستم مطرح شده است و درحال حاضر نیز به‌منظور استفاده در طراحی و محاسبات مهندسی و همچنین، به‌عنوان یک متغیر برای شناسایی و نمایش تولید رواناب درحوضه‌ها به‌کار می‌رود (Sivapalan و همکاران، ۲۰۰۵؛ Merz و Blöschl، ۲۰۰۹). کاربرد دیگر ضریب جریان در دستیابی به مدل‌های تحلیل فراوانی سیل است که در آن‌ها با استفاده از میزان بارندگی، فراوانی سیل برآورد می‌شود و در فهم و کنترل سیل در یک رژیم هیدرولوژیکی یا اقلیمی کاربرد دارد (Parajka و همکاران، ۲۰۰۵؛ Porhemat، ۲۰۰۲).

در دهه‌های اخیر، پژوهش‌های مختلفی در زمینه بررسی نقش عوامل فیزیوگرافی، هیدرولوژیکی، اقلیمی و زمین‌شناسی بر ضریب جریان، انجام شده است. اهمیت این ضریب، همچنین در پژوهش‌های محققین مختلفی از جمله نتایج منتشر شده به‌وسیله Blöschl و همکاران (۲۰۱۳)، Giudice و همکاران (۲۰۱۳)، Gupta (۱۹۹۲)، Merz و Blöschl (۲۰۰۹) مورد تایید قرار گرفته است. بعضی از پژوهشگران مانند Carrillo و همکاران (۲۰۰۶) روابط بین ضریب جریان و درصد پوشش سطوح نفوذناپذیر حوضه را مورد تأکید قرار داده‌اند. روابط بین ضریب جریان و شرایط اولیه رطوبت خاک حوضه به‌وسیله Longobardi و همکاران (۲۰۰۳) مورد بررسی قرار گرفت. ایشان به‌دلیل عدم امکان دسترسی به شرایط رطوبت خاک، از شاخص جریان پایه به‌عنوان یک پیش‌بینی کننده اولیه برای برآورد ضریب جریان، استفاده نموده‌اند. استناد ایشان برای استفاده از شاخص جریان پایه به‌عنوان عامل پیش‌بینی کننده ضریب جریان، وجود ارتباط بین

^۲ Normalized Difference Vegetation Index

^۱ Runoff Coefficient (RC)

اقلیمی بررسی شد.

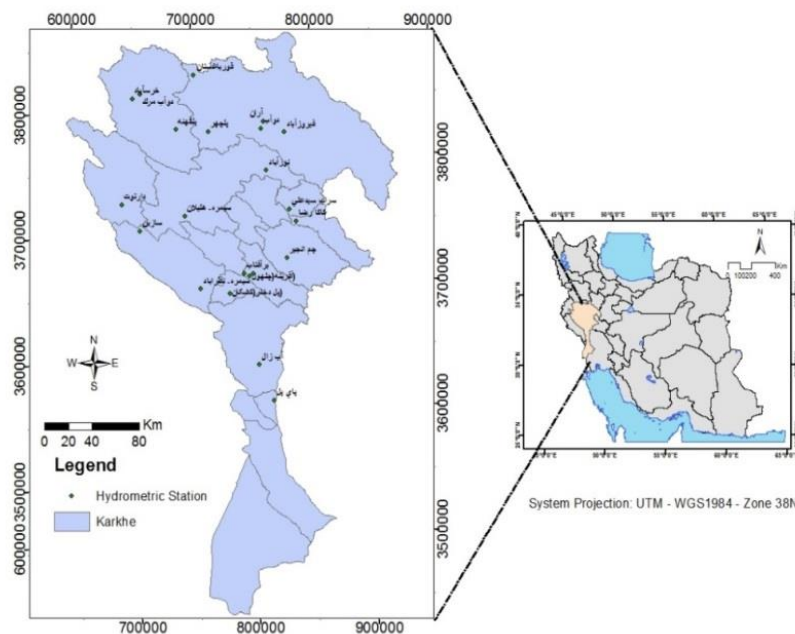
نتایج پژوهشی که به وسیله Yin و همکاران (۲۰۱۷) بر روی تاثیر تغییر اقلیم و تغییر کاربری و پوشش بر روی رواناب سطحی در منطقه شمال غرب چین با استفاده از مدل SWAT انجام شد. نشان داد که تاثیر ترکیبی تغییر اقلیم و پوشش و کاربری در مقیاس بزرگ منجر به کاهش ضریب رواناب شده است. عوامل عمده موثر بر ضریب جریان به وسیله Zhou و همکاران (۲۰۱۶) در تعدادی از حوضه‌های واقع در شمال شرق فلات تبت مورد پژوهش قرار گرفت.

نتایج نشان داد که عمق و شدت بارش رابطه خوبی با ضریب جریان دارد. با بررسی‌های به عمل آمده تاکنون مطالعات جامعی در خصوص بررسی تاثیر روش‌های همگن‌بندی در دقت برآورد ضریب جریان، به منظور تحلیل منطقه‌ای در سطح منطقه‌ای و استانی در کشور ایران مشاهده نشد. با توجه به این که روش‌های مختلف منجر به دسته‌بندی‌های متفاوتی از حوضه‌های همگن خواهد شد، لذا پژوهش در خصوص شناخت مناسب‌ترین روش همگن‌بندی، موجب دستیابی به اطلاعاتی برای استفاده در تحلیل شرایط هیدرولوژیکی و مدیریت بهینه منابع آب حوضه در منطقه پژوهش خواهد شد. هدف از این پژوهش،

تعیین تاثیر روش‌های خوشه‌بندی سلسله مراتبی در دقت مدل‌سازی ضریب جریان در حوضه کرخه و انتخاب مناسب‌ترین روش است.

مواد و روش‌ها

ویژگی‌ها و موقعیت منطقه: محدوده این پژوهش شامل تعداد ۲۲ زیرحوضه واقع در حوضه کرخه است که بین عرض $30^{\circ} 49'$ تا $34^{\circ} 04'$ شمالی و طول جغرافیایی $46^{\circ} 06'$ تا $46^{\circ} 10'$ شرقی واقع شده است (شکل ۱). حوضه رودخانه کرخه یکی از رودخانه‌های رتبه دوم کشور است و مساحت آن برابر با 51806 کیلومتر مربع است که $3/2$ درصد از مساحت ایران را پوشش داده است. زیرحوضه‌های اصلی کرخه شامل سیمره، کشکان و زیرحوضه میانی منتهی به سد کرخه می‌باشد. منطقه مورد پژوهش متأثر از آب و هوای خشک، نیمه‌خشک، مرطوب و نیمه‌مرطوب است. در مجموع $45/1$ درصد از منطقه دارای اقلیم خشک، 28 درصد نیمه‌خشک و $26/9$ درصد مرطوب و نیمه‌مرطوب می‌باشد. بیشینه بارش سالانه در دوره آماری مورد پژوهش به میزان 556 میلی‌متر متعلق به حوضه کشکان و کمینه آن مربوط به حوضه کرخه در پای پل با 453 میلی‌متر است.



شکل ۱- موقعیت منطقه مورد مطالعه

بنگستان، گورپی و امیران، بختیاری، پابده، شهبازان، تله زنگ، کشکان و گروه فارس است و به‌طور متوسط ۳۲ درصد پوشش سطحی منطقه پژوهش را تشکیل داده‌اند. در این خصوص بیشینه پوشش سطحی آن در حوضه فیروزآباد با ۵۷ درصد پوشش سطحی است. سازندهای دوره‌های چهارم عمدتاً شامل نهشته‌های آبرفتی، پادگانه‌های آبرفتی، مخروط‌افکنه‌ها و رسوبات رودخانه‌ای است. ویژگی‌های پایه حوضه‌های مورد پژوهش در جدول ۱ ارائه شده است.

سازندهای سخت کربناته دارای گسترش مناسب هستند و حدود ۴۸ درصد سطح منطقه را پوشش می‌دهند. بیشینه پوشش سطحی سازندهای کربناته کارستی، متعلق به حوضه‌های کاکا رضا و آب‌زال با پوشش سطحی ۴۵ درصد و کمینه آن مربوط به حوضه فیروزآباد با پوشش دو درصد می‌باشد. سازندهای سخت غیرکربناته عمدتاً شامل سنگ‌های دگرگونی زون سنندج-سیرجان، ولکانیک‌ها و توف‌های همراه، شیست‌های همدان، سازندهای متعلق به گروه

جدول ۱- مشخصه‌های حوضه‌های مورد پژوهش

ایستگاه هیدرومتری	کد ایستگاه	مساحت (km ²)	ارتفاع متوسط (m)	شیب متوسط حوضه (/.)	طول رودخانه اصلی (km)	طول جغرافیایی	عرض جغرافیایی
فیروزآباد	۲۱-۱۰۹	۸۶۹	۱۹۵۲	۱۷/۳	۴۸/۹	۲۱° ۳۴'	۰۷° ۴۸'
آران	۲۱-۱۱۳	۲۲۹۸	۱۷۸۰	۱۵/۵	۶۱/۴	۲۵° ۳۴'	۵۵° ۴۷'
دوآب	۲۱-۱۱۵	۸۰۲۶	۱۸۹۵	۱۵/۱	۱۲۱/۸	۲۲° ۳۴'	۵۴° ۴۷'
پل‌چهر	۲۱-۱۲۷	۱۰۲۰۸	۱۸۹۱	۱۷/۳	۱۶۱/۷	۱۹° ۳۴'	۲۵° ۴۷'
خرس‌آباد	۲۱-۱۳۱	۱۴۳۴	۱۵۲۷	۸/۷	۷۷/۸	۳۱° ۳۴'	۴۳° ۴۶'
دوآب مرک	۲۱-۱۳۳	۱۲۹۴	۱۵۴۴	۱۳/۷	۴۶	۳۳° ۳۴'	۴۶° ۴۶'
پل‌کهنه	۲۱-۱۴۱	۵۰۴۱	۱۵۶۷	۱۴/۲	۸۱/۵	۱۹° ۳۴'	۰۷° ۴۷'
قورباغستان	۲۱-۱۴۳	۵۳۰۹	۱۵۶۲	۱۴/۱	۹۵	۴۳° ۳۴'	۱۵° ۴۷'
نورآباد	۲۱-۱۴۵	۶۲۱	۲۰۴۳	۱۶/۸	۳۴/۱	۰۴° ۳۴'	۵۸° ۴۷'
سیمره-هلیلان	۲۱-۱۴۷	۱۹۹۷۷	۱۷۵۲	۱۷/۵	۱۸۱/۷	۴۲° ۳۳'	۱۵° ۴۷'
دارتوت	۲۱-۱۵۷	۲۶۵۷	۱۵۵۱	۱۷/۶۱	۸۳	۴۵° ۳۳'	۴۰° ۴۶'
سازبن	۲۱-۱۵۹	۲۶۱۲۸	۱۳۶۵	۲۰	۹۲۲۳/۶	۳۴° ۳۳'	۵۱° ۴۶'
کاکا رضا	۲۱-۱۶۹	۱۱۳۰	۲۰۲۷	۲۴/۱	۷۱/۸	۴۳° ۳۳'	۱۶° ۴۸'
سراب سیدعلی	۲۱-۱۷۱	۷۸۶	۲۱۰۴	۲۷/۴	۲۶/۸	۴۸° ۳۳'	۱۳° ۴۸'
چم انجیر	۲۱-۱۷۵	۱۶۳۰	۱۶۵۰	۲۰/۵	۵۱/۶	۲۷° ۳۳'	۱۳° ۴۸'
آفرینه (کشکان)	۲۱-۱۷۷	۶۸۴۲	۱۷۱۸	۲۳/۴	۸۶/۳	۱۹° ۳۳'	۵۴° ۴۷'
آفرینه (چلهول)	۲۱-۱۷۹	۸۰۸	۱۶۴۷	۲۳/۷	۵۴/۴	۱۸° ۳۳'	۵۲° ۴۷'
برآفتاب	۲۱-۱۸۱	۱۱۳۶	۱۳۷۳	۹/۳	۳۱۸/۳۵	۱۹° ۳۳'	۴۹° ۴۷'
پل	۲۱-۱۸۳	۹۲۶۷	۱۶۳۲	۲۲/۴	۱۱۱/۵	۱۰° ۳۳'	۴۳° ۴۷'
آب زال	۲۱-۱۸۹	۶۰۰	۱۴۰۵	۳۴/۳	۴۱/۴	۴۰° ۳۲'	۰۴° ۴۸'
پای پل	۲۱-۱۹۱	۴۲۱۹۱	۱۵۴۴	۱۹/۲۷	۳۱۲	۲۵° ۳۲'	۰۹° ۴۸'
سیمره-نظرآباد	۲۱-۴۱۱	۲۸۲۸۱	۱۶۳۰	۱۳/۱	۲۰۶/۸	۱۱° ۳۳'	۲۶° ۴۷'

دارای آمار کامل و بدون نیاز به بازسازی انتخاب شدند. همچنین، با استفاده از نقشه توپوگرافی با مقیاس ۱:۵۰۰۰۰ و تعیین موقعیت ایستگاه‌ها، محدوده مورد پژوهش مشخص و پارامترهای اولیه حوضه با استفاده از سامانه اطلاعات جغرافیایی، استخراج شدند. ضریب جریان با استفاده از روش جاستین محاسبه و سپس

روش پژوهش: برای انجام این پژوهش، ابتدا با بررسی داده‌های دبی روزانه ایستگاه‌های آب‌سنجی و باران‌سنجی تعداد ۲۲ ایستگاه با آمار مناسب و دوره مشترک آماری سال‌های آبی ۱۳۷۸-۱۳۵۳ انتخاب شد. با توجه به اهمیت صحت و دقت آمار ایستگاه‌های

$$Y_L = Y + K_n S_y \quad (2)$$

که در آن، Y_H آستانه داده‌های پرت بالا به صورت لگاریتمی و Y_L آستانه داده‌های پرت پایین به صورت لگاریتمی و K_n ضریبی است که از جداول مربوطه به تعداد داده‌ها انتخاب می‌شود. S_y انحراف از معیار داده‌ها می‌باشد. در سال‌هایی که داده‌های دبی ثبت نشده بود، با استفاده از ماتریس همبستگی و رابطه همبستگی با ایستگاه‌های دیگر نواقص دبی کلیه ایستگاه‌ها تا ۲۵ سال آماری مشترک تکمیل شد.

روش محاسبه رواناب: برای محاسبه رواناب سالانه حوضه‌های فاقد آمار، روش‌های مختلفی وجود دارد که مناسب‌ترین آن‌ها روش جاستین است. جاستین برای برآورد رواناب سالانه با استفاده از سه پارامتر بارندگی، درجه حرارت و شیب حوضه رابطه‌ای به شکل زیر را ارائه کرد.

$$R = K \times S^{0.155} \frac{P^2}{(1.8T + 32)} \quad (3)$$

که در آن، R رواناب سالانه (سانتی‌متر)، K ضریب منطقه‌ای، S پارامتر فیزیوگرافی است که از رابطه زیر به دست می‌آید.

$$S = \frac{H_{\max} - H_{\min}}{\sqrt{A}} \quad (4)$$

که در آن، H_{\max} ارتفاع بیشینه حوضه (کیلومتر)، H_{\min} ارتفاع کمینه حوضه (کیلومتر)، A سطح حوضه (کیلومتر مربع)، P بارندگی سالانه (سانتی‌متر)، T متوسط درجه حرارت سالانه (درجه سانتی‌گراد) است. **بررسی عوامل تأثیرگذار بر ضریب جریان:** تجزیه و تحلیل عاملی به روش PCA^۶ با استفاده از نرم‌افزار SPSS برای ۱۵ متغیر هیدرولوژیکی، فیزیوگرافی، اقلیمی و زمین‌شناسی در حوضه‌های منتخب به کار برده شد. متغیرهای مورد استفاده شامل ویژگی‌های مختلف حوضه، از جمله مساحت حوضه، طول رودخانه اصلی، ارتفاع متوسط، شیب حوضه، ضریب جریان، درصد پوشش سازندهای نفوذپذیر و کارستی، درصد پوشش سازندهای نفوذناپذیر، بارش متوسط سالانه، طول حوضه، چگالی زهکشی، شاخص پوشش گیاهی (NDVI)، شیب رودخانه اصلی، دبی کل، جریان پایه و درصد پوشش برف در اسفند ماه هستند.

جریان پایه با استفاده از داده‌های روزانه جریان و پس از کنترل سری زمانی، به وسیله نرم‌افزار Hydro Office (۲۰۱۰) به روش فیلتر رقومی برگشتی تک-پارامتره، استخراج شد. واحدهای سنگ‌شناسی با استفاده از نقشه زمین‌شناسی رقومی شده، با مقیاس ۱:۲۵۰۰۰۰ و بر اساس نوع لیتولوژی به دو طبقه عمده نفوذپذیر و نفوذناپذیر تقسیم و مساحت تحت پوشش هر واحد در هر حوضه به تفکیک محاسبه شد. تجزیه و تحلیل عاملی برای شناخت عوامل مستقل موثر بر ضریب جریان انجام شد. همگن‌بندی حوضه‌ها با استفاده از عوامل مستقل انتخابی به روش‌های مختلف سلسله مراتبی شامل: نزدیک‌ترین همسایه^۱، دورترین همسایه^۲، فاصله از میانه^۳، فاصله از مرکز ثقل^۴ و روش وارد^۵، انجام شد. سپس روابط منطقه‌ای به روش رگرسیون خطی، در سطوح معنی‌داری کمتر از یک درصد برای گروه‌های همگن تمام روش‌ها به دست آمد. تامین اعتبار روابط به دست آمده به روش‌های آزمون استقلال خطاها، بررسی توزیع نرمال خطاها و آزمون هم‌خطی انجام شد. سپس دقت مدل-های استخراج شده با استفاده از ایستگاه‌های شاهد محاسبه و با استفاده از آماره‌های میانگین مطلق خطا، مجذور میانگین مربعات خطا و خطای نسبی با یکدیگر مقایسه شد.

بررسی و آماده‌سازی اطلاعات: در بسیاری از مطالعات هیدرولوژیکی، با فقدان و یا کمبود ایستگاه-های آب‌سنجی مواجه بوده و نیز عوامل بسیاری وجود دارند که باعث از دست رفتن آمار برخی ایستگاه‌ها می‌شوند، از این‌رو در ابتدا اقدام به جمع‌آوری، اصلاح و بازسازی آمار دبی در منطقه پژوهش شد. به منظور شناسایی داده‌های پرت، از آزمون داده‌های پرت استفاده شد. آزمون مربوطه برای هر دو وضعیت داده-های بالا و پایین، انجام شد که برای تعیین آستانه داده‌های پرت بالا از رابطه (۱) و برای تعیین آستانه داده‌های پرت پایین از رابطه (۲) استفاده شد.

$$Y_H = Y + K_n S_y \quad (1)$$

¹ Nearest neighbor

² Furthest neighbor

³ Median clustering

⁴ Centroid clustering

⁵ Ward

⁶ Principal Component Analysis

خطاها با P نشان داده شود، در این صورت آماره دوربین-واتسون به کمک رابطه $DW = 2(1-P)$ محاسبه می‌شود.

اگر $P=0$ آن‌گاه مقدار آماره $DW=2$ خواهد بود که نشانه عدم همبستگی خطاها است.

اگر $P=1$ باشد، آن‌گاه آماره $DW=0$ خواهد بود که نشان می‌دهد، خطاها دارای خود همبستگی مثبت هستند.

همچنین، اگر $P=-1$ باشد، آن‌گاه مقدار آماره $DW=4$ خواهد بود که نشانه همبستگی منفی بین خطاها است، (Wade, 2005).

از آنجایی که فرض H_0 بر عدم همبستگی بین خطاها تأکید دارد، چنانچه این آماره در بازه $1/5$ تا $2/5$ قرار گیرد، فرض H_0 پذیرفته می‌شود و در غیر این صورت فرض H_0 رد می‌شود، یعنی بین خطاها همبستگی وجود دارد.

بررسی نرمال بودن خطاها: به منظور بررسی نرمال بودن خطاها، مقادیر استاندارد خطاها محاسبه و نمودار توزیع خطاها و نمودار نرمال داده‌ها رسم می‌شود. توجه به میزان میانگین و انحراف معیار خطاها برای بررسی اعتبار روابط لازم است.

آزمون هم‌خطی^۵: هم‌خطی وضعیتی است که نشان می‌دهد یک متغیر مستقل، تابعی خطی از بین سایر متغیرهای مستقل است. اگر هم‌خطی در یک معادله رگرسیون بالا باشد، بدین معنی است که بین متغیرهای مستقل همبستگی بالایی وجود دارد و ممکن است با وجود بالا بودن ضریب تعیین، مدل دارای اعتبار بالایی نباشد. به عبارت دیگر با وجود آن که مدل دارای دقت مناسب به نظر می‌رسد، ولی دارای متغیرهای مستقل معنی‌داری نیست.

همگن‌بندی حوضه‌ها: یکی از روش‌هایی که می‌توان به وسیله آن دبی جریان را به سایر حوضه‌ها تعمیم داد، تعیین همگنی حوضه‌های آبخیز می‌باشد. در این پژوهش به منظور همگن‌بندی مناطق مورد نظر، از روش‌های مختلف تحلیل خوشه‌ای سلسله مراتبی شامل: نزدیک‌ترین همسایه، دورترین همسایه، فاصله از مرکز ثقل و روش وارد، استفاده شد. تحلیل خوشه-

با توجه به پیچیدگی نتایج حاصل از تحلیل عاملی، به منظور به پیشینه رساندن واریانس بارهای هر یک از عامل‌ها، تسهیل در تفسیر ساختار عاملی و ایجاد استقلال در عامل‌ها، محورهای عاملی با روش واریماکس^۱ دوران یافت. سپس با استفاده از روش مدل رگرسیونی، ماتریس امتیازات عاملی استخراج شد. برای خارج کردن متغیرهای غیرضروری و توجه به همبستگی جزئی، از ماتریس همبستگی آنتی‌ایمیج^۲ و آماره کفایت اندازه نمونه (MSA)^۳ استفاده شد. به طوری که هر متغیر که دارای کمترین مقدار MSA بود، با در نظر گرفتن سطوح معنی‌داری ماتریس ضرایب همبستگی میان متغیرها، از مجموعه متغیرها حذف شدند. پس از انتخاب متغیرهای مهم، تجزیه عاملی بر اساس این متغیرها صورت گرفته، مهمترین پارامترها به عنوان عوامل مستقل مشخص شد. همگن‌بندی با استفاده از عوامل انتخابی و به روش‌های مختلف سلسله مراتبی انجام شد. سپس روابط بین عامل وابسته که ضریب جریان (RC) است، با سایر عوامل مستقل، با استفاده از روش رگرسیونی رابطه (۵) در مناطق همگن بررسی و مناسب‌ترین آن‌ها ارائه شد.

$$Rc = aLi + bMr + cHc + d \quad (5)$$

که در آن، Li عامل زمین‌شناسی، Mr عامل مورفومتری، Hc عامل هیدرواقلمی و پارامترهای a ، b ، c و d ضرایب ثابت رابطه هستند.

بررسی اعتبار رگرسیون‌ها: انجام هر رگرسیونی، مستلزم فرضیاتی از قبیل، مستقل بودن خطاها از یکدیگر، نرمال بودن توزیع خطاها و پایین بودن میزان هم‌خطی بین متغیرهای مستقل است و تا زمانی که این مفروضات برقرار نباشد، اعتبار رگرسیون خدشه دار خواهد بود.

بررسی استقلال خطاها: خطا عبارت است از تفاوت بین مقادیر واقعی و مقادیر پیش‌بینی شده می‌باشد. برای بررسی استقلال خطاها از همدیگر از آزمون دوربین-واتسون^۴ استفاده شد. اگر همبستگی بین

¹ Varimax Rotation

² Anti-image

³ Measure of Sampling Adequacy

⁴ Durbin-Watson

⁵ Colinerity

روش فاصله از مرکز ثقل^۵: این روش نیز جزء روش-های خوشه‌بندی سلسله مراتبی پایین به بالا محسوب می‌شود. در این روش برای محاسبه شباهت بین دو خوشه A و B از رابطه (۸) استفاده می‌شود.

$$d_{AB} = d\left(\frac{\sum_{i=A} X_i}{N_A}, \frac{\sum_{i=B} X_j}{N_B}\right) \quad (8)$$

که در آن، X_i یک نمونه داده متعلق به خوشه A، X_j یک نمونه داده متعلق به خوشه B، N_A تعداد اعضاء خوشه A و N_B تعداد اعضاء خوشه B است. در واقع در این روش، شباهت بین دو خوشه فاصله بین بردار میانگین تمام اعضاء یکی با بردار میانگین تمام اعضاء دیگری است.

روش وارد^۶: در این روش خوشه‌بندی، برای کاهش تلفات ناشی از داده‌های دور افتاده^۷ از معیار جدیدی برای محاسبه عدم شباهت بین خوشه‌ها استفاده می‌شود. بدین منظور از مجموع مربعات تفاضل هر داده از یک خوشه با بردار میانگین آن خوشه، به‌عنوان معیاری برای سنجش یک خوشه استفاده می‌شود. الگوریتم مربوطه بدین صورت است که ابتدا هر داده به‌عنوان یک خوشه در نظر گرفته می‌شود. به‌ازاء تمام جفت خوشه‌های ممکن از مجموعه خوشه‌ها آن دو خوشه‌ای که مجموع مربعات تفاضل داده‌های خوشه حاصل از اجتماع آن‌ها با بردار میانگین خوشه حاصل، کمینه باشد، انتخاب می‌شود. دو خوشه انتخاب شده با هم ترکیب می‌شوند و تا زمانی که تعداد خوشه‌ها به تعداد مورد نظر نرسیده است، مراحل تکرار می‌شود.

نتایج و بحث

در نتیجه بررسی ماتریس همبستگی و جدول اشتراک استخراجی، متغیر طول رودخانه اصلی به‌دلیل همبستگی شدید داخلی با برخی عوامل و همچنین، به‌علت اینکه مقدار اشتراک استخراجی آن کمتر از ۰/۵ بود از فرایند تحلیل عاملی حذف شد. آزمون تحلیل عاملی نشان داد که کلیه متغیرها در قالب پنج عامل طبقه‌بندی می‌شوند که در مجموع در برگیرنده ۸۵/۴۵۲ درصد از واریانس داده‌ها هستند (جدول ۲).

ای به‌عنوان نوعی تحلیل چندمتغیره، ابزاری آماری است که در آن داده‌ها و مشاهدات بر اساس متغیرهای تعیین شده به‌وسیله محقق طبقه‌بندی می‌شوند. در این روش، حوضه‌ها از نظر خصوصیات فیزیکی دو به دو با هم مقایسه می‌شوند و باید از مقیاس‌های کمی استفاده کرد تا بتوان بیشترین شباهت یا کمترین فاصله بین داده‌ها و مشاهدات را اندازه‌گیری نمود.

خوشه‌بندی سلسله مراتبی^۱: خوشه‌بندی سلسله-مراتبی، روشی است که در آن به خوشه‌های نهایی بر اساس میزان عمومیت آن‌ها ساختاری سلسله مراتبی، نسبت داده می‌شود.

روش نزدیک‌ترین همسایه^۲: این روش یکی از قدیمی‌ترین و ساده‌ترین روش‌های خوشه‌بندی است و جزء روش‌های خوشه‌بندی سلسله مراتبی پایین به بالا^۳ محسوب می‌شود. در این روش ابتدا تمام داده‌ها به‌عنوان خوشه‌ای مجزا در نظر گرفته می‌شود و در طی فرایندی تکراری در هر مرحله خوشه‌هایی که شباهت بیشتری با یکدیگر دارند، با همدیگر ترکیب می‌شوند تا در نهایت یک خوشه و یا تعداد مشخصی خوشه حاصل شود. در این روش برای محاسبه شباهت بین دو خوشه A و B از رابطه (۶) استفاده می‌شود.

$$d_{AB} = \min d_{ij} \quad i = A \quad j = B \quad (6)$$

که در آن، i یک نمونه داده متعلق به خوشه A و j یک نمونه داده متعلق به خوشه B می‌باشد. در واقع در این روش شباهت بین دو خوشه، کمترین فاصله بین یک عضو از یکی با یک عضو از دیگری است.

روش دورترین همسایه^۴: این روش نیز جزء روش-های خوشه‌بندی سلسله مراتبی پایین به بالا محسوب می‌شود. در این روش برای محاسبه شباهت بین دو خوشه A و B از رابطه (۷) استفاده می‌شود.

$$d_{AB} = \max d_{ij} \quad i = A \quad j = B \quad (7)$$

که در آن، i یک نمونه داده متعلق به خوشه A و j یک نمونه داده متعلق به خوشه B می‌باشد. در واقع در این روش شباهت بین دو خوشه، بیشترین فاصله بین یک عضو از یکی با یک عضو از دیگری است.

¹ Hierarchical Clustering

² Nearest Neighbor

³ Bottom-Up

⁴ Furthest neighbor

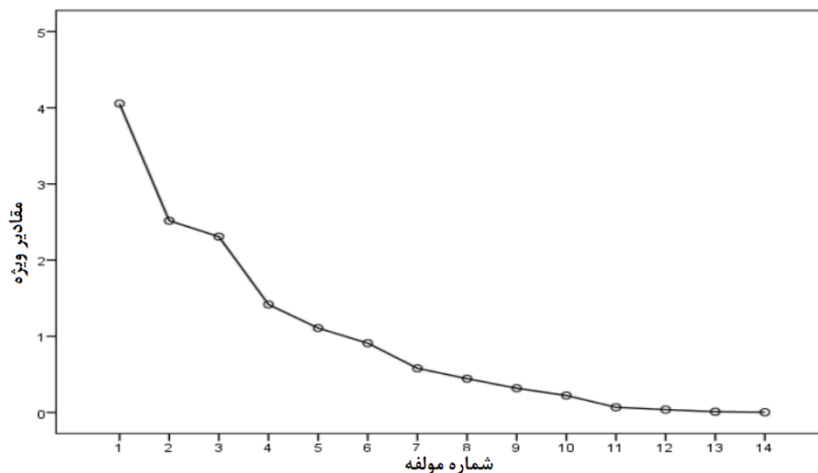
⁵ Centroid clustering

⁶ Ward

⁷ Outlier

حوضه و شاخص NDVI به‌عنوان موثرترین پارامترها شناخته شدند. تصویر گرافیکی مقادیر ویژه در هر یک از عامل‌های استخراج شده در شکل ۲ نمایش داده شده است. همان‌طور که مشخص است، مقدار واریانس توجیه شده (مقدار ویژه) با استخراج عامل‌های پنجم به بعد سریعاً افت می‌کند.

عوامل طول حوضه، چگالی زهکشی، درصد پوشش برف در اسفند ماه و شیب رودخانه اصلی در قالب عامل اول، سازندهای نفوذپذیر و نفوذناپذیر از میان عوامل زمین‌شناسی در عامل دوم، شاخص جریان پایه، ارتفاع متوسط حوضه و بارش متوسط سالانه در عامل سوم، عامل مساحت حوضه در عامل چهارم و شیب



شکل ۲- مقادیر ویژه عامل‌های استخراج شده

جدول ۲- مقادیر کل درصد از واریانس توضیح داده شده بر اساس عوامل مستقل انتخابی در آزمون تحلیل عاملی

عامل	درصد از واریانس	درصد واریانس جمع‌ی
۱	۳۰/۸۹۳	۳۰/۸۹۳
۲	۱۸/۳۴۰	۴۹/۲۳۳
۳	۱۷/۱۴۰	۶۶/۳۷۳
۴	۱۰/۶۹۶	۷۷/۰۷۰
۵	۸/۳۸۳	۸۵/۴۵۲

مشابه هم بوده و در دو گروه با اجزاء مشابه دسته‌بندی می‌شوند.

ولی در روش وارد، حوضه‌ها به سه منطقه همگن تفکیک شد، با این استثناء که در این روش، حوضه‌های واقع در منطقه همگن سه، دقیقاً مشابه منطقه دو در سایر روش‌ها است، لذا، به‌دلیل تشابه منطقه سه این روش با منطقه دو سایر روش‌ها، فقط برای منطقه یک و دو این روش، مراحل بعدی روش پژوهش اجرا شد. همچنین، در تمام روش‌های مورد استفاده، ایستگاه پای‌پل (کد ۱۹۱-۲۱) در خارج از منطقه همگن‌بندی قرار گرفت.

همگن‌بندی مناطق مورد نظر، به پنج روش سلسله مراتبی بر اساس مهمترین عوامل موثر انتخابی، انجام و سپس برای اعتباریابی گروه‌های همگن، روش تابع تشخیص مورد استفاده قرار گرفت و تعویض گروه‌ها بر مبنای ارزیابی مدل تا زمانی که میزان دقت مدل برابر ۱۰۰ درصد شود، ادامه داده شد. گروه‌های همگن بر اساس تابع فاصله اقلیدوسی و در فاصله اقلیدوسی "پنج" انتخاب شدند که نتایج در جدول ۳ ارائه شده است. نتایج همگن‌بندی نشان داد که حوضه‌ها در روش‌های همگن‌بندی نزدیک‌ترین همسایه، دورترین همسایه، فاصله از مرکز ثقل و فاصله از میانه، تماماً

است که در آن‌ها، Rc ضریب جریان، A مساحت حوضه، H ارتفاع متوسط حوضه، IPE درصد پوشش سازندهای نفوذناپذیر، P بارش متوسط سالانه، CL طول حوضه، Sb شیب متوسط حوضه، Q دبی متوسط سالانه، SN سطح پوشش برف در اسفند ماه است.

روابط رگرسیونی برآورد ضریب جریان: روابط استخراجی با دخالت زیرحوضه‌های هر منطقه، نشان داد که مقادیر ضریب جریان در برابر پارامترهای هیدرولوژیکی، زمین‌شناسی و هندسی حوضه، همبستگی معنی‌داری را در سطح کمتر از یک درصد دارد که به صورت مدل‌های رگرسیونی چندمتغیره با استفاده از روش گام به گام در جدول ۴ ارائه شده

جدول ۳- زیرحوضه‌های انتخابی واقع در هر یک از زیرگروه‌ها بر اساس روش‌های پنج گانه خوشه‌بندی

روش	شماره خوشه	کد ایستگاه
نزدیک‌ترین همسایه	۱	پل چهر، پل کهنه، قورباغستان، دوآب، آفرینه (کشکان)، پل دختر (کشکان)، خرس‌آباد، دوآب مرک، آران، چم انجیر، برآفتاب، آفرینه (چلهول)، آب زال، نورآباد، سراب سیدعلی، کاکارضا، فیروزآباد، دارتوت
نزدیک‌ترین همسایه	۲	سیمره (هلیلان)، سازین، سیمره (نظرآباد)
وارد	۱	خرس‌آباد، دوآب مرک، آران، چم انجیر، برآفتاب، آفرینه (چلهول)، آب زال، نورآباد، سراب سیدعلی، کاکارضا، فیروزآباد، دارتوت
وارد	۲	پل چهر، پل کهنه، قورباغستان، دوآب، آفرینه (کشکان)، پل دختر (کشکان)
وارد	۳	سیمره (هلیلان)، سازین، سیمره (نظرآباد)

جدول ۴- روابط رگرسیونی در مناطق همگن

روش	شماره خوشه	میانگین مطلق خطا	مجذور میانگین مربعات خطا	خطای نسبی (درصد)	رابطه
وارد	۱	۲/۰۶	۴/۶۴	۱۰/۳	$Rc = 1.230 \times Sb + 0.030 \times P - 11.173$
وارد	۲	۳۲/۷۶	۵۴/۸۴	۲۳	$Rc = 0.002 \times A + 0.005 \times H + 1.876 \times Sb - 0.443 \times IMPE - 67.549$
نزدیک‌ترین همسایه	۱	۷/۸۵	۹/۶۲	۲۵/۴	$Rc = 3.018 \times Sb + (-1.146 \times SN) - 18.530$
نزدیک‌ترین همسایه	۲	۱۰/۴	۱۸/۴۲	۲۹/۹	$Rc = 0.683 \times Sb + (-0.047 \times CL) + 0.003 \times Q + 15.982$

برآوردی استاندارد شده و خطای (باقی‌مانده) استاندارد شده را شامل می‌شود. مقدار آماره دوربین-واتسون، برای تمام مدل‌ها در فاصله ۱/۵ تا ۲/۵ قرار دارد و

خلاصه نتایج اجرای آزمون دوربین-واتسون در جدول ۵ ارائه شده است. این خروجی‌ها آماره برآوردی به‌وسیله رگرسیون، خطاها (باقی‌مانده)، مقادیر

یک، مقادیر مربوط به پارامترهای حوضه‌های واقع در منطقه دو به‌عنوان شاهد در نظر گرفته شد. سپس با اجرای رابطه مربوطه برای پارامترهای شاهد، در مجموع میزان خطای نسبی برآوردی برای رابطه منطقه یک به روش "وارد" به میزان ۱۰/۳ درصد به-دست آمد. با همین روش با در نظر گرفتن ایستگاه شاهد برای منطقه دو در روش "وارد"، میزان خطای نسبی، رابطه منطقه دو به میزان ۲۳ درصد به‌دست آمد.

میزان میانگین مطلق خطا و مجذور میانگین مربعات خطا برای منطقه یک به ترتیب، به میزان ۲/۰۶ و ۴/۶۴ و برای منطقه دو برابر ۳۲/۷۶ و ۵۳/۸۴ به‌دست آمد. میزان خطای نسبی برآورد، میانگین مطلق خطا و مجذور میانگین مربعات خطا برای منطقه یک در روش نزدیکترین همسایه به ترتیب برابر ۲۵/۴، ۷/۸۵ و ۹/۶۲ به‌دست آمد و برای منطقه دو به ترتیب برابر با ۲۹/۹، ۱۰/۴ و ۱۸/۴ بود که به‌طور مقایسه‌ای روش نزدیکترین همسایه به‌دلیل پایین-ترین میزان میانگین مطلق خطا، مجذور میانگین مربعات خطا و خطای نسبی کم نسبت به روش دیگر از دقت بیشتری در برآورد منطقه‌ای ضریب جریان برخوردار است.

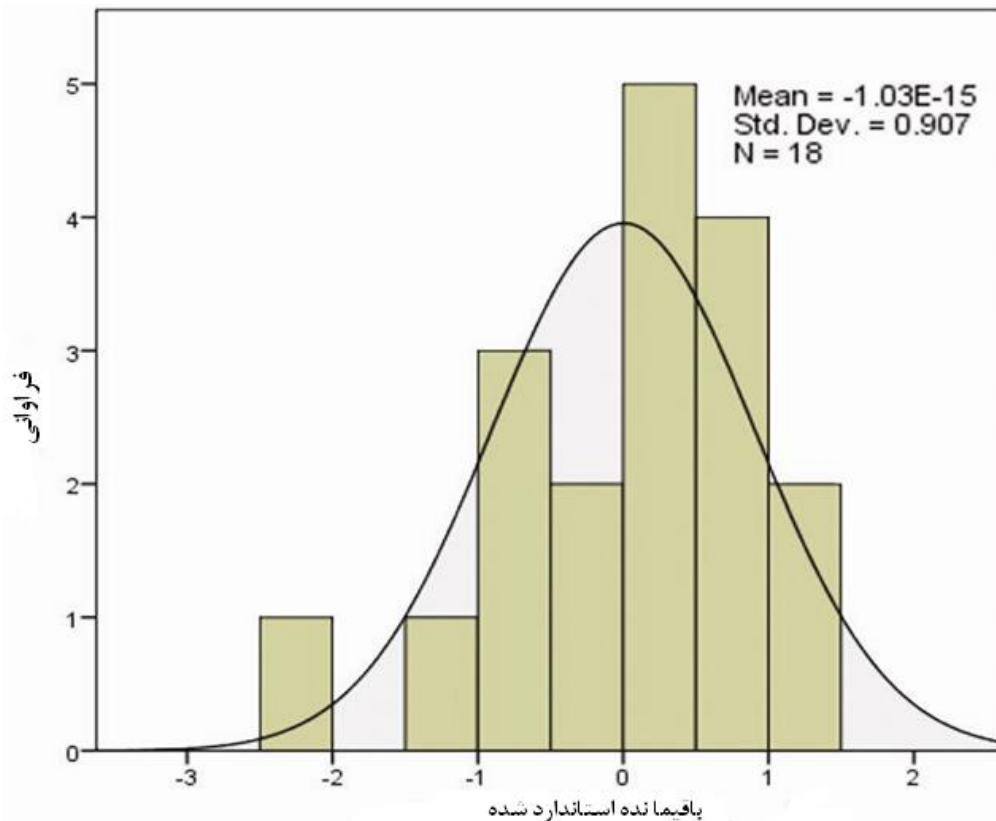
اعتبار استفاده از مدل‌ها را تایید می‌کند. به‌منظور بررسی نرمال بودن خطاها، نمودار توزیع خطاها ترسیم و با نمودار توزیع نرمال مقایسه شد. مشاهده شد که توزیع خطاها برای تمام روابط استخراجی در مناطق همگن، تقریباً نرمال است. همچنین، مقدار میانگین مربوطه، بسیار کوچک (نزدیک به صفر) و انحراف معیار آن نیز تقریباً یک است. شکل ۳ نمودار توزیع خطاها برای منطقه همگن یک، به روش نزدیکترین فاصله را به‌عنوان نمونه نشان می‌دهد. آزمون هم‌خطی برای روش وارد، نشان داد که فقط در یکی از متغیر-های مستقل (به‌دلیل پایین بودن بسیار زیاد مقادیر ویژه و بالا بودن شاخص وضعیت) بیانگر وجود مشکل در استفاده از رگرسیون مربوط به آن است.

برای رفع این مشکل با بررسی فاصله انحراف معیار متغیرها از خط رگرسیون، مشخص شد که داده مربوطه نیز در فاصله کمتر از یک انحراف معیار از خط رگرسیون است و استفاده از آن خطای جدی در پیش‌بینی‌های مدل ایجاد نمی‌کند. بر اساس نتایج آزمون هم‌خطی برای روش نزدیکترین همسایه نیز پارامترهایی که دارای شاخص وضعیت بیش از ۱۵ بودند، در مرحله انتخاب پارامترها برای مدل، حذف شد.

برای بررسی میزان خطای برآوردی رابطه منطقه

جدول ۵- خلاصه اجرای آزمون دوربین-واتسون

مدل	روش	ضریب همبستگی	ضریب تعیین	ضریب تعیین تعدیل شده	خطای برآورد استاندارد مدل	ضریب دوربین-واتسون
منطقه ۱	وارد	۰/۹۸۶	۰/۹۷۲	۰/۸۸۳	۷/۱۷۸۹	۱/۹۳۲
منطقه ۲	وارد	۰/۹۸۰	۰/۹۶۱	۰/۹۳۶	۵/۱۶	۲/۵
منطقه ۱	نزدیکترین همسایه	۰/۸۹۴	۰/۷۹۹	۰/۷۷۲	۱۰/۰۱	۲/۵
منطقه ۲	نزدیکترین همسایه	۰/۹۹۸	۰/۹۹۶	.	.	۲



شکل ۳- نمودار توزیع خطاها برای ضریب جریان در منطقه همگن یک به روش نزدیکترین فاصله

نتیجه‌گیری

در این پژوهش با استفاده از تحلیل عاملی از بین ۱۵ پارامتر مؤثر در ضریب جریان و همگن‌بندی حوضه‌ها با استفاده از عوامل مستقل انتخابی به روش-های مختلف سلسله مراتبی شامل: نزدیک‌ترین همسایه، دورترین همسایه، فاصله از میانه، فاصله از مرکز ثقل و روش وارد انجام شد. سپس روابط منطقه‌ای به روش رگرسیون خطی، در سطوح معنی داری کمتر از یک درصد برای گروه‌های همگن تمام روش‌ها به دست آمد.

نتایج ارزیابی صحت و کارایی مدل‌های برآوردی، به روش‌های آزمون استقلال خطاها، نرمال بودن توزیع خطاها و هم‌خطی و بررسی دقت مدل‌های استخراج شده با استفاده از ایستگاه‌های شاهد نشان داد که

دقت مدل‌های استخراج شده در گروه‌های همگن به روش نزدیک‌ترین همسایه به دلیل کمینه درصد خطای نسبی به میزان ۲۵/۴ درصد، میانگین مطلق خطا به میزان ۷/۸۵ و مجذور میانگین مربعات خطا به میزان ۹/۶۲ از سایر روش‌ها مناسب‌تر است و به‌عنوان بهترین روش همگن‌بندی برای تحلیل منطقه‌ای ضریب جریان در منطقه پژوهش تشخیص داده شد.

تقدیر و تشکر

این پژوهش با استفاده از امکانات پژوهشکده حفاظت خاک و آبخیزداری به انجام رسیده که بدین وسیله از همراهی و مساعدت مسئولین محترم پژوهشکده تشکر به عمل می‌آید.

منابع مورد استفاده

1. Abbasszade, G.H. 2003. Regional flood frequency analysis using a hybrid method in South Alborz. MSc Thesis, Department of Natural Resources, Tehran University, Iran. 107 pages (in Persian).
2. Ataei, H. and M. Shiran. 2011. Identifying homogeneous hydrological basins based on effective geomorphologic variants on flood by cluster analysis. Journal of Geography and Environmental Planning, 42(2): 79-98 (in Persian).
3. Beven, K.J. and M.J. Kirkby. 1979. A physically-based variable contributing-area model of catchment

- hydrology. *Hydrological Science Bulletin*, 24(1): 43-69.
4. Biyabanaki, M. and S.S. Eslamiyan. 2005. Using cluster method in the determination of hydrologic homogeneous and evaluation by discriminate analysis and Andrew curves in Karkheh catchment. *Journal of Agriculture*, 2: 13-26 (in Persian).
 5. Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione and H. Savenije. 2013. *Runoff prediction in ungauged basins: Synthesis across Processes, Places and Scales*. 1st edition, Cambridge University Press, New York, 484 pages.
 6. Boughton, W.C. 1987. Evaluating partial areas of watershed runoff. *Journal of Irrigation and Drainage Engineering*, 113(3): 356-366.
 7. Carrillo, F., J.V. Soares, N.J. Ferreira and D.A. Rodriguez. 2006. Runoff coefficient and its relation to phytogeography in BRAZIL. *Proceedings of 8th ICSHMO*, Foz do Iguacu, Brazil, 229-233.
 8. Cerdan, O., Y. Le Bissonnais, G. Govers, V. Leconte, K. Van Oost, A. Couturier, C. King and N. Dubreuil. 2004. Scale effects on runoff from experimental plots to catchments in agricultural areas in Normandy. *Journal of Hydrology*, (299): 4-14.
 9. Chahoki, A., A. Salajeghe, M. Mahdavi, SH. Khaligi Sigarodi and A. Sadi. 2013. Regional modeling of flow duration curve at ungauged catchments of arid region, case study: central of Iran. *Journal of Range and Watershed Management*, 66(2): 251-265 (in Persian).
 10. Dos Reis Castro, N.M., A.V. Auzet and P. Chevallier. 1999. Land use change effects on runoff and erosion from plot to catchment scale on the basaltic plateau of Southern Brazil. *Hydrological Processes*, 13: 1621-1628.
 11. Wade, A. 2005. *A handbook of statistical analyses using SPSS*. Chapman and Hall/CRC, Boca Raton, 345 pages.
 12. Giudice, G.D., R. Padulano and G. Rasulo. 2013. Spatial prediction of the runoff coefficient in Southern Peninsular Italy for the index flood estimation. *Hydrology and Earth System Sciences*, 9: 4919-4941.
 13. Gottschalk, L. and R. Weingartner. 1998. Distribution of peak flow derived from a distribution of rainfall volume and runoff coefficient and a unit hydrograph. *Journal of Hydrology*, 208: 148-162.
 14. Gupta, B.L. 1992. *Engineering hydrology. Estimation of Runoff*, 4: 63-64.
 15. Kjeldsen, T. and J.C. Smithers. 2002. Regional flood frequency analysis in the KwaZulu-Natal provinces, South Africa, using the index-flood method. *Journal of Hydrology*, 255: 194-211.
 16. Khosravi, M., M. Mohseni Saravi, M. Vafakhah and G. Fathi. 2011. Determination of the best hydrometric clustering method for regional flood analysis in west catchments of country. *Journal of Watershed Management Science and Engineering*, 4(13): 15-20 (in Persian).
 17. Laaha, G. and G. Blöschl. 2014. A comparison of low flow regionalization methods-catchment grouping. *Journal of Hydrology*, 323(1-4): 193-214.
 18. Longobardi, A., P. Villani, R.B. Grayson and A.W. Western. 2003. On the relationship between runoff coefficient and catchment initial conditions. *Proceedings of the International Congress on Modeling and Simulation of Australia and New Zealand*, Townsville, Australia, 2: 867-872.
 19. Malekiyan, A., S. Hasanpor and H. Faraji-Sabokbar. 2013. Application and comparing of gray cluster analysis and hierarchical analysis in locating suitable areas for flood spreading, case study: Garebayegan catchment. *Journal of Watershed Management Science and Engineering*, 7(20): 35-46 (in Persian).
 20. Merz, R. and G. Blöschl. 2009. A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resources Research*, 45(1): 14-25.
 21. Michael, W. and M.C. Berry. 2007. *Survey of text mining: clustering, classification, and retrieval*. Second edition, ISBN: 1848000456 | 239 pages.
 22. Parajka, J., R. Merz and G. Blöschl. 2005. A comparison of regionalization methods for catchment model parameters, hydrology and earth system sciences discussions. *Journal of European Geosciences Union*, 9(3): 157-171.
 23. Porhemat, J. 2002. *Snow melt runoff distribution model using remote sensing data*. PhD Thesis, Islamic Azad University, Tehran, 261 pages (in Persian).
 24. Quang, B.T. and M.J. Laituri. 2013. The effects of watershed characteristics on storm runoff relationships in Vietnam. *Journal of Environmental Science and Water Resources*, 2(2): 40-52.
 25. Rohani, H. 2001. *Regional flood frequency analysis using a hybrid method in arid and semi-arid area, case study: Khorasan*. MSc Thesis, Department of Natural Resources, Tehran University, 132 pages (in Persian).
 26. Salinas, J.L., G. Laaha, M. Rogger, J. Parajka, A. Viglione, M. Sivapalan and G. Blöschl. 2013. Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies. *Hydrology and Earth System Sciences*, 17: 2637-2652.
 27. Smakhtin, V.Y. 2001. Low flow hydrology: a review. *Journal of Hydrology*, 240: 147-186.

28. Sivapalan, M., G. Bloschl, R. Merz and D. Gutknecht. 2005. Linking flood frequency to long-term water balance: Incorporating effects of seasonality. *Water Resources Research*, 41: W06012.
29. Vezza, P., C. Comoglio, M. Rosso and A. Viglione. 2010. Low flows regionalization in north-western Italy. *Water Resources Management*, 24(14): 4049-4074.
30. Visessri, S. and N. McIntyre. 2016. Regionalization of hydrological responses under land-use change and variable data quality. *Hydrological Sciences Journal*, 61(2): 302-320.
31. Yin, J., F. He, Y.J. Xiong and G.Y. Qiu. 2017. Effects of land use/land cover and climate changes on surface runoff in a semi-humid and semi-arid transition zone in northwest China. *Hydrology and Earth System Sciences*, 21: 183-196.
32. Zhang, Z., X. Chen, Y. Huang and Y. Zhang. 2014. Effect of catchment properties on runoff coefficient in a karst area of southwest China. *Hydrological Processes*, 28: 3691-3702.
33. Zhou, L., L. Dong, Y. L. Xiao and L. Guang-Yong. 2016. Runoff coefficient characteristics and its dominant influencing factors in a riparian grassland in the Qinghai Lake watershed, NE Qinghai-Tibet Plateau. *Arabian Journal of Geosciences*, 9: 397-410.

Investigating the effect of hierarchical clustering methods on accurately modeling of runoff coefficient in Karkheh Basin

Rahim Kazemi^{*1} and Jahangir Porhemmat²

¹ Scientific Board, Soil Conservation and Watershed Management Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran and ² Associate Professor, Soil Conservation and Watershed Management Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran

Received: 22 February 2017

Accepted: 22 August 2017

Abstract

Estimating the runoff coefficient that is influenced by morphometric, geologic and hydro-climatologically factors are the most important issues in hydrology and information of its role in the planning and management of water resources is more important. Clustering catchments is the best method for the analysis of hydrological parameters in the absence of full coverage of hydrological data. In this research, twenty two hydrometric stations with common period from 1974 to 1999 were selected. Physiographic parameters of the catchments were extracted. Runoff coefficient was calculated and then base flow was extracted from using one parameter recursive digital filters. Lithological units using digital geological map, with the scale of 1: 250,000, based on expert opinion divided on two classes and area covered by each unit in each catchment were calculated. Factor analysis using 15 parameters were conducted. Catchments using independent factors in different hierarchy methods includes: nearest neighbor, furthest neighbor, median clustering, centroid clustering and Ward method were classified. Then, the regional equations using linear regression at 1% significant level were determined. To compare and evaluate the accuracy and efficiency of the models, independence errors, colinerity and normal distribution of error were tested. The results of factor analysis showed that all variables are to be classified in terms of five factors which 85.9% of the variance was included. Results of homogeneity showed that the basins in homogeneous methods of nearest neighbor, furthest neighbor, centroid clustering and median clustering, were all the same and classified in two groups with the similar components. The results of accuracy assessment showed that the accuracy of nearest neighbor methods was more accurate, and because of low relative error (25.4%) and MAE of 7.85 and RMSE of 9.62 was diagnosed as the best method for regional analyzing of runoff coefficient in the study area.

Key words: Clustering, Hierarchical, Modeling, Regional analysis, Estimation error

* Correspondent author: ra_hkazemi@yahoo.com