دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشنامه مدیریت حوزه آبخیز

تحلیل فراوانی منطقه‌ای سیلاب با استفاده از تحلیل خوشه‌ای ترکیبی و گشتاورهای خطی

علی آهنی[۱]، صمد امامقلی زاده[۲]، سید سعید موسوی ندوشنی[۳] و خلیل اژدری[۴]

۱- دانشجوی کارشناسی ارشد، دانشگاه شاهرود، (نویسنده مسوول: ali.ahani66@yahoo.com)
۲ و ۴- دانشیار، دانشگاه شاهرود
۳- استادیار، دانشگاه شهید بهشتی
تاریخ دریافت: ۹۲/۲/۲۱ تاریخ پذیرش: ۹۲/۱۱/۶

چکیده

مطالعات انجام گرفته نشان می‌دهد که روش‌های تحلیل خوشه‌ای، از کارآمدترین روش‌های منطقه‌بندی حوزه‌های آبخیز به‌منظور انجام تحلیل فراوانی منطقه‌ای سیلاب هستند. در این تحقیق به‌منظور استفاده از نقاط قوت و کم‌رنگ کردن نقاط ضعف دو گروه اساسی روش‌های تحلیل خوشه‌ای، یعنی خوشه‌بندی سلسله مراتبی و خوشه‌بندی افرازی، از ترکیب الگوریتم‌های خوشه‌بندی Ward و K-means برای منطقه‌بندی حوزه‌های آبخیز سفیدرود بزرگ و ارس استفاده شده است. همچنین تأثیر انتخاب و به‌کارگیری برخی ویژگی‌های جغرافیایی، فیزیوگرافیک و هواشناسی و ترکیبات آنها در همگنی مناطق حاصل از تحلیل خوشه‌ای مورد بررسی قرار گرفته است. ارزیابی وضعیت همگنی مناطق حاصل و اجرای تحلیل فراوانی منطقه‌ای که با استفاده از الگوریتم گشتاورهای خطی انجام گرفته است، نشان می‌دهد که از میان ویژگی‌های مورد مطالعه، ترکیب طول و عرض جغرافیایی و مساحت سطح زهکشی به‌عنوان ویژگی‌های مورد استفاده در منطقه‌بندی حوزه‌های آبخیز سفیدرود و ارس می‌تواند بهترین گزینه برای حصول بیش‌ترین تعداد مناطق همگن باشد. به‌علاوه، برای مناطقی که تا حدودی معیارهای همگنی را ارضا نمی‌کنند، استفاده از توزیع ویکبی به‌عنوان توزیع منطقه‌ای، می‌تواند مفید باشد.

واژه‌های کلیدی: تحلیل فراوانی منطقه‌ای، گشتاورهای خطی، تحلیل خوشه‌ای، خوشه‌بندی ترکیبی

مقدمه

استفاده از روش شاخص سیلاب دالریمپل (۶) برای تحلیل فراوانی منطقه‌ای به شکل نوین توسط والیس (۱۴) زمانی آغاز شد که او این روش را در ترکیب با گشتاورهای وزنی احتمال[۱] (۷) و توزیع ویکبی[۲] به‌عنوان روشی برای برآورد چندک‌ها در امتداد کران بالایی توزیع فراوانی به کار گرفت. در ادامه هاسکینگ و والیس (۹) روشی را برای تحلیل فراوانی منطقه‌ای بر اساس گشتاورهای خطی[۳] که پیش‌تر توسط هاسکینگ (۸) معرفی شده بودند، ارائه کردند. آنها در این روش شاخص‌هایی را برای آزمودن همگنی مناطق و نکویی برازش توزیع منطقه‌ای معرفی کردند. کاربرد گشتاورهای خطی در زمینه‌ی مطالعات هیدرولوژی به شکل چشمگیری افزایش یافته است. پژوهش‌های کومار و همکاران (۱۲،۱۰)، کومار و چاترجی (۱۱)، چاوشی و همکاران (۵) و سرکار و همکاران (۱۸) از جمله تحقیقات ارزشمندی هستند که در این مورد به مطالعه پرداخته‌اند. یکی از روش‌های منطقه‌بندی به‌منظور اجرای تحلیل فراوانی منطقه‌ای، استفاده از تحلیل خوشه‌ای[۴] است. تحلیل خوشه‌ای گونه‌ای از روش‌های آماری چند متغیری است که برای خوشه‌بندی داده‌ها، در گروه‌های مشابه مورد استفاده قرار می‌گیرند. در زمینه تحلیل فراوانی منطقه‌ای سیلاب، موزلی (۱۳) خوشه‌بندی سلسله مراتبی[۵] موجود در برنامه کامپیوتری BMDP2M را برای منطقه‌بندی حوزه‌های آبخیز در نیوزلند مورد استفاده قرار داد. تاسکر (۲۰) الگوریتم تمام پیوند را برای منطقه‌بندی حوزه‌های آریزونا در ایالات متحده به کار گرفت. ناتان و مک‌ماهون (۱۴) عملکرد الگوریتم‌های تک پیوند[۶]، تمام پیوند[۷]، پیوند متوسط[۸] و Ward از مجموعه‌ی روش‌های خوشه‌بندی سلسله مراتبی موجود در بسته‌ی آماری مطالعات اجتماعی (SPSS)[۹] را مقایسه کردند. برن (۲) الگوریتم K-means را برای تشخیص گروه‌بندی مناسب یک شبکه از ایستگاه‌های هیدرومتری در مانیتوبای جنوبی واقع در کانادا به‌کار گرفت. در این مطالعه آماره‌های سیلاب و طول و عرض جغرافیایی ایستگاه‌ها به‌عنوان ویژگی‌های بردارهای مشخصه انتخاب شدند. بنا بر نتایج مطالعه برن (۳) استفاده از متغیرهای یکسان مربوط به سیلاب برای تشکیل مناطق و متعاقب آن برای ارزیابی همگنی مناطق حاصل، منجر به تشکیل مناطقی می‌شود که اگرچه همگن هستند، اما ممکن است برای تحلیل فراوانی منطقه‌ای سیلاب مفید نباشند. برن و گول (۴) خوشه‌بندی سلسله مراتبی را برای منطقه‌بندی حوزه‌های آبخیز کانادا به کار گرفتند.

1- Probability Weighted Moments (PWM) 2- Wakeby Distribution 3- L-moments 4- Cluster analysis
5- Hierarchical clustering 6- Single linkage 7- Complete linkage
8- Average Linkage 9- Statistical Package for the social sciences

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

هنگامی‌که تحلیل خوشه‌ای مبتنی بر ویژگی‌های سایت‌ها باشد، آماره‌های درون‌سایتی به‌عنوان مبنای یک آزمون مستقل همگنی مناطق نهایی قابل استفاده هستند (۹). همچنین برن و گول (۴) الگوریتم K-means را بر مجموعه‌ای از سایت‌های هیدرومتری در هند به‌منظور حصول مناطقی برای اجرای تحلیل فراوانی منطقه‌ای سیلاب، اعمال کردند. ویژگی‌های مورد استفاده در این مطالعه شامل مساحت حوزه، طول و شیب آبراهه‌ی اصلی رودخانه بودند.

رائو و سرینیواس (۱۵) عملکرد الگوریتم‌های خوشه‌بندی ترکیبی را روی ۲۶۵ ایستگاه هیدرومتری در ایالت ایندیانا در تحلیل فراوانی منطقه‌ای سیلاب مورد مطالعه قرار دادند. در این تحقیق عملکرد شاخص‌های مختلف صحت خوشه‌بندی مورد بررسی قرار گرفت که در میان آنها شاخص silhouette width بهترین عملکرد را از خود نشان داد. همچنین بر اساس توانایی کمینه‌سازی تابع هدف و مقادیر شاخص‌های صحت خوشه‌بندی ترکیب الگوریتم‌های Ward و K-means به‌عنوان روش بهینه برای منطقه‌بندی سایت‌های ایالت ایندیانا معرفی شد.

سرینیواس و همکاران (۱۹) همچنین عملکرد الگوریتم فازی c-means و نگاشت‌های خودسازمانده[1] را در زمینه‌ی منطقه‌بندی حوزه‌های آبخیز مورد بررسی قرار دادند. چاوشی و سلیمان (۵) نیز روش‌های معمول تحلیل خوشه‌ای و روش فازی را در مورد ۷۰ سایت در شمال ایران مورد استفاده قرار دادند.

در مطالعه‌ی حاضر، هدف بررسی عملکرد روش خوشه‌بندی ترکیبی حاصل از ترکیب الگوریتم‌های Ward و K-means برای منطقه‌بندی حوزه‌های آبخیز سفیدرود و ارس به‌منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب است. همچنین تأثیر استفاده از چند ویژگی جغرافیایی، فیزیوگرافیک و هواشناسی در تشکیل مناطق همگن مورد بررسی قرار گرفته است.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

حوزه‌ی آبخیز سفیدرود بزرگ با مساحت ۶۳۹۴۵ کیلومتر مربع، بزرگ‌ترین زیرحوزه‌ی حوزه‌ی آبخیز اصلی دریای خزر است و به خاطر وجود اقلیم‌های متفاوت و منابع غنی آب و خاک از اهمیت خاصی برخوردار است. این حوزه در محل تلاقی رشته کوه‌های البرز، زاگرس و مرکزی واقع شده است. این حوزه‌ی آبخیز از دو شاخه‌ی رودخانه‌ای اصلی به نام قزل اوزن و شاهرود تشکیل شده است که در محل سد سفیدرود به هم می‌پیوندند و رودخانه سفیدرود را تشکیل می‌دهند. ذینفعان این حوزه استان‌های آذربایجان شرقی، اردبیل، تهران، زنجان، قزوین، کردستان، گیلان و همدان می‌باشند (شکل ۱).

حوزه‌ی آبخیز ارس دومین زیرحوزه‌ی بزرگ حوزه‌ی آبخیز دریای خزر و دارای مساحت ۳۹۵۳۴ کیلومتر مربع (در ایران) است. رود ارس از دو شاخه‌ی مجزا در ارمنستان و ترکیه تشکیل می‌شود که این دو شاخه در منتهی‌الیه شمال غربی ایران به هم می‌پیوندند. در ایران، ارس ابتدا وارد استان آذربایجان غربی و پس از آن آذربایجان شرقی شده و سپس به سمت شمال شرقی تغییر مسیر داده، وارد دشت مغان در استان اردبیل می‌شود و در تازه کند از ایران خارج و وارد جمهوری آذربایجان می‌شود (شکل ۲).

در این پژوهش تعداد ۴۳ ایستگاه هیدرومتری حوزه‌ی سفیدرود و ۳۷ ایستگاه هیدرومتری حوزه‌ی ارس که دارای اطلاعات مورد نیاز جهت اجرای خوشه‌بندی ترکیبی بودند، برای انجام تحلیل فراوانی منطقه‌ای مورد استفاده قرار گرفته‌اند.

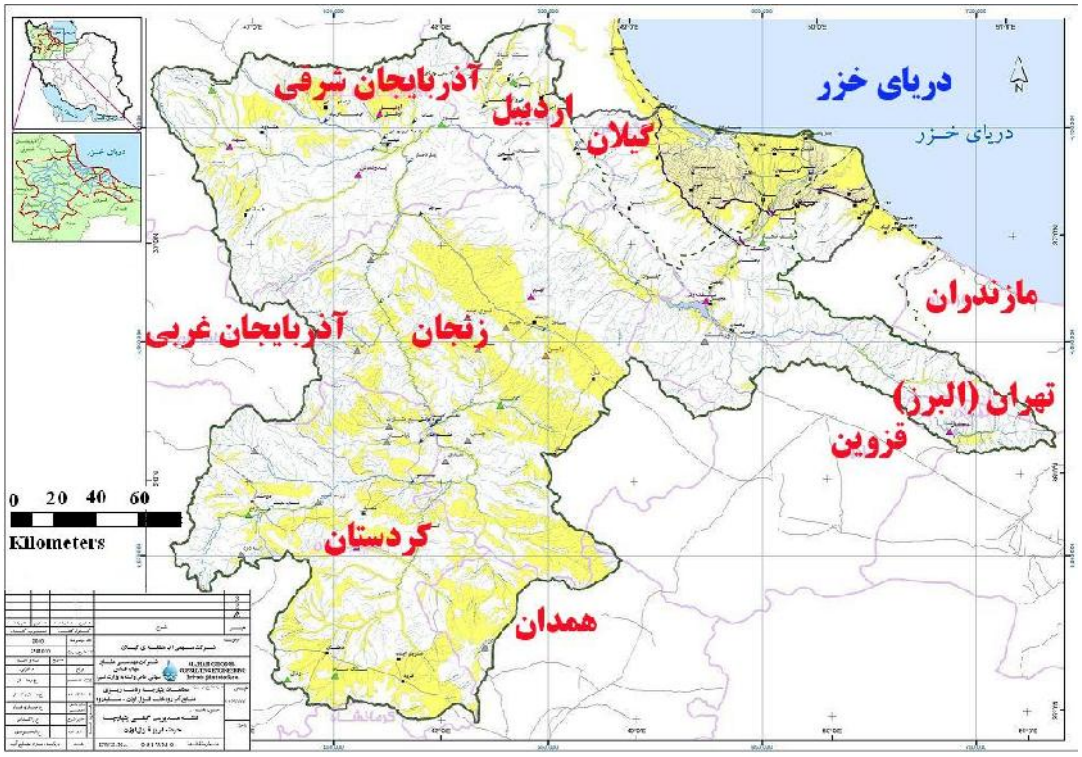
مبانی نظری و فرآیند مطالعه

نخستین گام پس از تعیین منطقه‌ی مورد مطالعه و تهیه‌ی داده‌های آماری مربوطه، انتخاب ویژگی‌ها و آماده‌سازی بردارهای مشخصه است. در این مرحله هدف انتخاب ویژگی‌های مؤثر بر واکنش سیلابی حوزه در منطقه‌ی مورد مطالعه است. یک خوشه از یک یا تعداد بیش‌تری بردار مشخصه[2] تشکیل می‌شود. یک بردار مشخصه شامل چندین ویژگی اثرگذار یک ایستگاه هیدرومتری بر فرآیند تولید سیلاب حوزه است. برخی از ویژگی‌هایی که می‌توانند برای منطقه‌بندی حوزه‌های آبخیز استفاده شوند، عبارتند از: ویژگی‌های فیزیوگرافیک، ویژگی‌های پوشش خاک، ویژگی‌های مرتبط با کاربری زمین، ویژگی‌های زهکشی حوزه، ویژگی‌های موقعیت جغرافیایی، ویژگی‌های هواشناسی، ویژگی‌های زمین‌شناسی حوزه، شاخص‌های زمان واکنش حوزه و شاخص‌های شکل حوزه. از میان ویژگی‌های مؤثر بر واکنش سیلابی سایت‌های مورد نظر در حوزه، با توجه به اطلاعات در دسترس، طول و عرض جغرافیایی به‌عنوان ویژگی‌های جغرافیایی، مساحت سطح زهکشی به‌عنوان ویژگی فیزیوگرافیک، ضریب رواناب به‌عنوان تابعی از پوشش خاک و متوسط بارندگی سالانه به‌عنوان یک ویژگی هواشناسی انتخاب شدند. خوشه‌بندی ترکیبی بر اساس هر یک از این ویژگی‌ها و یا با استفاده از ترکیبات مختلف آنها به‌عنوان ویژگی‌های بردارهای مشخصه انجام گرفت.

1- Self-organizing feature maps

2- Feature vector

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]



شکل ۱- نمایش موقعیت جغرافیایی حوزه‌ی آبخیز سفیدرود بزرگ (قزل اوزن) (شرکت مهندسی مشاور مهاب قدس)

شکل ۲- نمایش موقعیت جغرافیایی حوزه‌ی آبخیز ارس (شرکت مهندسی مشاور مهاب قدس)

پژوهش‌ها نشان می‌دهد قضاوت‌های ذهنی برای تعیین گروه‌های سایت‌های هیدرومتری با استفاده از روش‌های معمول، با به‌کارگیری شیوه‌های خوشه‌بندی کارآمد به شکل چشم‌گیری کاهش می‌یابد. بیش‌تر الگوریتم‌های خوشه‌بندی موجود می‌توانند در دو گروه طبقه‌بندی شوند: خوشه‌بندی سلسله مراتبی و

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

خوشه‌بندی‌افرازی[1] برخی از الگوریتم‌هایی که نماینده‌ی روش خوشه‌بندی سلسله مراتبی هستند، عبارتند از: تک پیوند، تمام پیوند، پیوند متوسط، و الگوریتم Ward تفاوت این الگوریتم‌ها در راهبردی است که برای تعریف نزدیک‌ترین همسایه نسبت به یک خوشه استفاده می‌کنند. از جمله انواع الگوریتم‌های خوشه‌بندی افرازی نیز می‌توان به الگوریتم‌های K-means و K-medoids اشاره کرد. در این روش‌ها، تلاش می‌شود تا گروه‌بندی طبیعی موجود در داده‌ها طی یک افراز واحد بازیابی شود.

الگوریتم‌های خوشه‌بندی‌افرازی نیازمند یک حدس اولیه در مورد تعداد خوشه‌ها و مراکز خوشه‌ها و متأثر از این حدس‌های اولیه هستند، در حالی که روش‌های خوشه‌بندی سلسله مراتبی تحت‌تأثیر تعیین شرایط اولیه قرار نمی‌گیرند. روش‌های خوشه‌بندی افرازی در حدی که بردارهای مشخصه بتوانند از یک خوشه به خوشه‌ی دیگر برای کمینه‌سازی تابع هدف حرکت کنند، پویا هستند. اما در روش‌های سلسله مراتبی، بردارهای مشخصه‌ی تعلق گرفته به یک خوشه در مراحل ابتدایی، نمی‌توانند به خوشه‌ی دیگری انتقال یابند.

به‌منظور کم‌رنگ کردن نقاط ضعف هر یک از این دو گروه از الگوریتم‌ها، می‌توان از الگوریتم‌های خوشه‌بندی ترکیبی استفاده کرد. در الگوریتم‌های خوشه‌بندی ترکیبی، مراکز خوشه‌های نهایی حاصل از یک الگوریتم سلسله مراتبی به‌عنوان مراکز اولیه‌ی خوشه‌ها در یک الگوریتم افرازی به کار می‌روند. در این مطالعه مراکز خوشه‌های به‌دست آمده به وسیله‌ی الگوریتم سلسله مراتبی Ward به‌عنوان مراکز اولیه‌ی خوشه‌ها در الگوریتم افرازی K-means مورد استفاده قرار گرفته‌اند.

الگوریتم Ward یک روش پرکاربرد برای مطالعات منطقه‌بندی در هیدرولوژی و اقلیم‌شناسی است. این الگوریتم مبتنی بر این فرض است که اگر دو خوشه به هم بپیوندند، تغییر در مقدار تابع هدف، تنها به رابطه‌ی بین این دو خوشه وابسته است و به روابط با دیگر خوشه‌ها بستگی ندارد (۱۷). تعاریف و روابط مختلفی برای سنجش فاصله‌ی میان هر دو بردار مشخصه وجود دارد که در این مطالعه از تعریف فاصله‌ی اقلیدسی استفاده شده است.

در انجام عملیات تحلیل خوشه‌ای، تابع هدف به‌صورت رابطه (۱) تعریف می‌شود،

$$F = \sum_{k=1}^{K} \sum_{j=1}^{n} \sum_{i=1}^{N_k} d^2\left(x_{ij}^k - x_{*j}^k\right) \qquad (۱)$$

که در آن، K نماد تعداد خوشه‌ها، N_k معرف تعداد بردارهای مشخصه در خوشه‌ی kام، x_{ij}^k نماینده‌ی مقدار باز تجدید مقیاس شده‌ی ویژگی jام در بردار مشخصه‌ی iام اختصاص داده شده به خوشه‌ی kام و x_{*j}^k مقدار میانگین ویژگی jام برای خوشه‌ی kام است. با کمینه‌سازی تابع هدف، فاصله‌ی هر بردار مشخصه از مرکز خوشه‌ای که به آن تعلق دارد به حداقل می‌رسد.

برای اجرای الگوریتم K-means باید ابتدا به تعداد خوشه‌های مورد نظر مرکز خوشه تعریف شود. هر خوشه توسط مرکز خود که میانگین بردارهای مشخصه‌ی موجود در خوشه است، معرفی می‌شود. اختصاص داده‌ها به خوشه‌های مختلف بر اساس دوری و نزدیکی به مراکز خوشه‌ها انجام می‌گیرد. پس از اتمام عملیات خوشه‌بندی، مراکز دسته‌ها با توجه به بردارهای مشخصه‌ی موجود در آنها به‌هنگام‌سازی شده و عملیات اختصاص بردارهای مشخصه تکرار می‌شود. این عملیات تا زمانی که بهترین نتیجه از نظر کمینه‌سازی تابع هدف حاصل شود، تکرار می‌گردد. بردارهای مشخصه در تکرارهای متعدد از خوشه‌ای به خوشه‌ی دیگر انتقال می‌یابند تا مقدار تابع هدف کمینه شود. روش K-means به‌دلیل کارایی خود در خوشه‌بندی مجموعه‌های بزرگ داده‌ها با ویژگی‌های عددی مشهور است (۱۷).

خوشه‌های تشکیل شده به‌صورت بصری و با استفاده از شاخص‌های صحت‌سنجی خوشه برای تعیین تعداد بهینه‌ی مناطق تفسیر می‌شوند. در مطالعه‌ی حاضر از میان شاخص‌های سنجش صحت خوشه‌بندی، شاخص silhouette width به‌دلیل عملکرد قابل قبول در مطالعات پیشین انتخاب شده است. این شاخص برای iامین بردار مشخصه در خوشه با s(i) نشان داده شده و مطابق رابطه (۲) تعریف می‌شود.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \qquad (۲)$$

که در آن a(i) فاصله‌ی متوسط بردار مشخصه‌ی iام نسبت به تمام بردارهای مشخصه‌ی دیگر موجود در خوشه و b(i) حداقل فاصله‌ی متوسط بردار مشخصه‌ی iام نسبت به تمام بردارهای مشخصه‌ی خوشه‌ی دیگر است. بر اساس رابطه (۲)، خواهیم داشت $-1 \le s(i) \le 1$. اگر مقدار s(i) نزدیک به ۱ باشد، می‌توان این‌طور نتیجه‌گیری کرد که بردار مشخصه‌ی iام در خوشه‌ای مناسب جای گرفته است. از سوی دیگر چنان‌چه مقدار s(i) به ۱- نزدیک باشد، می‌توان این گونه استنتاج کرد که iامین بردار مشخصه به خوشه‌ی مناسبی تعلق نیافته است. از میانگین تمام مقادیر s(i) که با S_{av} نشان داده می‌شود، برای قضاوت کلی در مورد خوشه‌بندی انجام گرفته استفاده می‌شود. بالا بودن کیفیت خوشه‌بندی بر اساس شاخص S_{av} بدین معنا است که بردارهای مشخصه‌ی موجود در یک خوشه حتی‌الامکان به هم نزدیک و بردارهای مشخصه‌ی موجود در خوشه‌های

1- Partitional clustering

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

مختلف حتی‌الامکان از یکدیگر دور هستند.

برآورد سیلاب در تحلیل فراوانی منطقه‌ای تا زمانی قابل اعتماد است که تعداد سال‌های آمار موجود در ایستگاه‌های یک منطقه بزرگ‌تر یا مساوی پنج برابر دوره‌ی بازگشت مورد نظر جهت برآورد بزرگی سیلاب باشد (۱۷). از این‌رو تعداد سایت‌های موجود در هر منطقه و سال‌های آماری موجود برای هر یک از آنها عاملی تعیین کننده در انتخاب تعداد خوشه‌ها است. در این تحقیق با توجه به تعداد سایت‌ها، نتایج حاصل از اختصاص سایت‌های موجود در حوزه‌های آبخیز سفیدرود و ارس به ۲ تا ۶ خوشه یا منطقه با استفاده از خوشه‌بندی ترکیبی مورد ارزیابی قرار گرفته است.

مناطق تعریف شده در گام پیشین، با استفاده از آزمون‌های همگنی آماری ارزیابی می‌شوند. ابتدا با محاسبه‌ی شاخص ناجوری D برای سایت‌های موجود در هر منطقه، سایت‌هایی که از نظر آماری با سایر سایت‌های مورد بررسی ناسازگار هستند، شناسایی شده و حذف می‌شوند. سپس همگنی مناطق حاصل از عملیات خوشه‌بندی، با استفاده از شاخص‌های ناهمگنی H مورد ارزیابی قرار می‌گیرند. سه شاخص ناهمگنی H_1، H_2 و H_3 بر اساس گشتاورهای خطی تعریف می‌شوند. در هر منطقه اگر $H<1$ باشد، منطقه همگن، اگر $1\leq H\leq 2$ باشد، منطقه نسبتاً ناهمگن و اگر $H\geq 2$ باشد، منطقه کاملاً ناهمگن است. از آنجا که در منطقه‌بندی با استفاده از روش‌های تحلیل خوشه‌ای، شاخص H_1 اهمیت و قابلیت اعتماد بیش‌تری دارد، لذا در بررسی همگنی مناطق حاصل، توجه و تأکید بیش‌تری به این شاخص معطوف می‌گردد. مناطق ناهمگن، به‌منظور بهبود وضعیت همگنی‌شان می‌توانند با حذف یک یا چند ایستگاه که تأثیر بیش‌تری در افزایش ناهمگنی دارند یا جابه‌جایی محدود برخی سایت‌ها بین خوشه‌ها و یا باز تعریف مناطق در صورت نیاز، اصلاح شوند (۹).

هدف این مرحله اجرای آزمون‌های نکویی برازش منطقه‌ای برای تشخیص و برازش یک توزیع فراوانی سیلاب مناسب بر داده‌های سیلاب سایت‌ها در یک منطقه است. توزیع برازش یافته، برای برآورد چندک‌های[1] سیلاب استفاده می‌شود. در این مطالعه به‌منظور شناسایی توزیع بهینه برای هر منطقه از شاخص نکویی برازش Z استفاده شده است. چنانچه در یک منطقه برای توزیعی خاص ($DIST$)، $|Z^{DIST}|\leq 1.64$ باشد، آن توزیع می‌تواند به‌عنوان توزیع منطقه‌ای انتخاب شود. نزدیک‌تر بودن این مقدار به صفر می‌تواند نشان‌دهنده‌ی بهینه بودن انتخاب توزیع مورد نظر باشد. در صورتی‌که هیچ یک از توزیع‌های دو یا سه پارامتری به‌عنوان توزیع منطقه‌ای برازش مناسبی را نشان ندهند، می‌توان از توزیع پنج پارامتری ویکبی استفاده کرد (۹). در این مطالعه چگونگی برازش توزیع‌های سه پارامتری لجستیک تعمیم‌یافته (GLO)، مقادیر حدی تعمیم‌یافته (GEV)، نرمال تعمیم‌یافته (GNO)، پیرسون تیپ III(PE3) و پارتوی تعمیم‌یافته (GPA) با استفاده از شاخص نکویی برازش Z مورد بررسی قرار گرفته است. این توزیع‌ها در حالت‌های خاص خود به‌صورت برخی توزیع‌های دو پارامتری شناخته شده در می‌آیند. همچنین در صورتی که اصلاح یک منطقه‌ی ناهمگن به شکلی مناسب دشوار یا ناممکن باشد، یا هیچ یک از توزیع‌های سه پارامتری معرفی شده، به‌عنوان توزیع منطقه‌ای برازش خوبی را نشان ندهند، می‌توان از توزیع پنج پارامتری ویکبی که انعطاف‌پذیری بالاتری نسبت به توزیع‌های دو یا سه پارامتری دارد، به‌عنوان توزیع منطقه‌ای استفاده کرد (۹). توزیع منتخب منطقه‌ای معمولاً با اعمال ضریبی مشخص برای هر سایت، تبدیل به توزیع ویژه‌ی آن سایت می‌گردد. این ضریب می‌تواند دبی متوسط هر سایت باشد.

در این مطالعه به‌منظور اجرای عملیات تحلیل خوشه‌ای، منطقه‌بندی و تحلیل فراوانی منطقه‌ای سیلاب از نسخه ۲.۱۵.۳ زبان برنامه‌نویسی R و محیط نرم افزاری مربوط به آن استفاده شده است. در این راه از بسته‌های[2] cluster نسخه ۱.۱۴.۴ و lmomRFA نسخه ۲.۵ که برای استفاده در محیط نرم‌افزاری R به نگارش در آمده‌اند، بهره گرفته شده است.

نتایج و بحث

در این تحقیق ابتدا با استفاده از الگوریتم خوشه‌بندی ترکیبی منطقه‌بندی بر اساس ویژگی‌های مختلف انجام گرفت. بدین صورت که ابتدا ایستگاه‌های هیدرومتری موجود در هر یک از دو حوزه‌ی سفیدرود و ارس با استفاده از الگوریتم خوشه‌بندی سلسله مراتبی Ward در تعداد ۲ تا ۶ خوشه جای گرفتند. سپس مراکز هر یک از این خوشه‌ها به‌عنوان مراکز اولیه‌ی خوشه‌ها در الگوریتم K-means برای حالت‌های ۲ تا ۶ خوشه‌ای مورد استفاده قرار گرفتند. تعداد دفعات تکرار الگوریتم K-means برابر ۱۰۰۰ تنظیم شد. در ادامه شاخص ناجوری D برای سایت‌های موجود در هر منطقه و شاخص‌های H برای مناطق تشکیل شده محاسبه شدند. از میان ۴۳ سایت مورد مطالعه در حوزه‌ی سفیدرود، تعداد ۲ سایت و از بین ۳۷ سایت مورد بررسی در حوزه‌ی ارس، تعداد ۶ سایت که در اجرای خوشه‌بندی با استفاده از ویژگی‌های مختلف بر اساس شاخص D ناجور تشخیص داده شدند، از مجموعه‌ی سایت‌های مورد بررسی حذف شدند. سپس خوشه‌بندی دوباره

1- Quantiles

2- Packages

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

انجام گرفت و مقدار شاخص صحت خوشه‌بندی S_{av} و شاخص‌های ناهمگنی H برای مناطق تشکیل شده مجدداً محاسبه شدند. بر اساس نتایج حاصل از محاسبه‌ی شاخص S_{av} در خوشه‌بندی بر اساس ویژگی‌های مختلف که در جدول (۱) ارائه شده است، بالاترین مقدار S_{av} برای منطقه‌بندی حوزه‌ی سفیدرود مربوط به حالت دو منطقه‌ای و برای حوزه‌ی ارس مربوط به حالت ۶ منطقه‌ای و هر دو بر اساس متوسط بارندگی سالانه است.

پس از محاسبه‌ی مجدد شاخص‌های ناهمگنی H، خوشه‌های دارای مقدار $H_1<1$ به‌عنوان مناطق همگن شناخته شده و تعداد آنها در هر مرحله از خوشه بندی ثبت شد. در جدول (۲) تعداد مناطق همگن، در مراحل مختلف خوشه‌بندی با استفاده از ویژگی‌های مختلف، ارائه شده است.

جدول ۱- مقدار شاخص صحت خوشه‌بندی S_{av} در خوشه‌بندی با استفاده از ویژگی‌های مختلف، L نماد طول و عرض جغرافیایی، A نماد مساحت سطح زهکشی، C نماد ضریب رواناب و P نماد متوسط بارندگی سالانه هستند.

حوزه	تعداد مناطق	L	A	C	P	LA	LC	LP	AC	AP	PC	APC	LAC	LAP	LPC	LAPC
									$\times 10^{-2}$							
سفیدرود	۲	۴۴	۵۷	۶۲	۸۱	۳۷	۴۱	۴۵	۴۳	۵۲	۵۳	۴۶	۳۸	۳۸	۴۳	۳۹
	۳	۵۹	۶۷	۵۹	۶۸	۴۵	۴۸	۵۲	۴۴	۵۰	۶۱	۴۷	۴۱	۴۰	۴۶	۳۸
	۴	۶۶	۶۳	۶۵	۶۷	۴۶	۴۹	۶۲	۴۶	۵۳	۴۵	۴۳	۴۰	۴۶	۵۴	۴۲
	۵	۶۵	۵۹	۶۶	۶۱	۵۰	۴۶	۶۰	۴۲	۵۵	۴۸	۴۱	۴۱	۴۸	۴۸	۴۴
	۶	۵۸	۵۹	۶۴	۶۰	۴۹	۴۵	۵۹	۴۵	۴۷	۴۹	۴۲	۳۸	۴۶	۴۳	۴۱
ارس	۲	۵۷	۵۹	۸۰	۶۸	۳۶	۴۴	۴۱	۴۲	۳۸	۴۰	۳۱	۳۶	۳۱	۳۴	۲۹
	۳	۵۳	۵۸	۵۷	۷۷	۴۲	۵۰	۳۳	۴۷	۴۳	۴۸	۳۱	۳۷	۳۰	۳۷	۲۷
	۴	۵۷	۵۸	۵۳	۹۰	۴۳	۴۴	۳۱	۴۰	۴۴	۵۰	۳۳	۳۳	۲۸	۳۱	۲۶
	۵	۵۸	۵۶	۵۴	۹۲	۴۴	۴۵	۳۲	۴۱	۳۹	۵۳	۳۵	۳۷	۲۷	۳۳	۲۸
	۶	۵۵	۵۴	۵۱	۹۴	۴۳	۴۴	۳۶	۳۹	۴۳	۵۱	۳۴	۳۶	۲۹	۳۶	۲۹

جدول ۲- تعداد مناطق همگن در خوشه‌بندی با استفاده از ویژگی‌های مختلف

حوزه	تعداد مناطق	L	A	C	P	LA	LC	LP	AC	AP	PC	APC	LAC	LAP	LPC	LAPC
سفیدرود	۲	۱	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
	۳	۳	۱	۱	۱	۲	۲	۲	۱	۱	۱	۱	۲	۲	۲	۲
	۴	۳	۱	۱	۲	۳	۳	۳	۰	۲	۱	۱	۳	۳	۳	۳
	۵	۴	۲	۲	۲	۴	۴	۳	۲	۳	۲	۲	۴	۴	۴	۴
	۶	۵	۳	۳	۲	۶	۵	۵	۴	۳	۳	۳	۵	۵	۵	۵
ارس	۲	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰
	۳	۱	۱	۰	۱	۰	۰	۱	۰	۰	۰	۱	۲	۱	۰	۲
	۴	۲	۱	۲	۲	۱	۲	۲	۱	۱	۱	۰	۱	۲	۰	۱
	۵	۳	۲	۲	۳	۳	۲	۲	۲	۱	۱	۱	۱	۱	۱	۱
	۶	۴	۳	۴	۴	۴	۳	۲	۲	۲	۲	۲	۳	۲	۲	۲

L: نماد طول و عرض جغرافیایی، A: نماد مساحت سطح زهکشی، C: نماد ضریب رواناب و P: نماد متوسط بارندگی سالانه هستند.

بر اساس جدول (۲)، در منطقه‌بندی حوزه‌ی آبخیز سفیدرود تنها در دو حالت تمامی مناطق تشکیل شده همگن هستند.حالت نخست مربوط به تقسیم حوزه به ۳ منطقه با استفاده از طول و عرض جغرافیایی و حالت دوم مربوط به تقسیم حوزه به ۶ منطقه با استفاده از طول و عرض جغرافیایی و مساحت سطح زهکشی به‌عنوان ویژگی‌های بردارهای مشخصه است. لذا مناطق حاصل از این دو حالت برای ادامه‌ی تحلیل فراوانی منطقه‌ای سیلاب در حوزه‌ی سفیدرود مورد استفاده قرار گرفتند. در مورد حوزه‌ی آبخیز ارس، همان‌گونه که در جدول (۲) دیده می‌شود، در هیچ یک از حالات منطقه‌بندی، تمام مناطق تشکیل شده همگن نیستند. لذا از بین حالاتی که بیش‌ترین نسبت مناطق همگن به کل مناطق را دارا هستند، باید بهترین حالت برگزیده شود. بدین منظور حالت‌های ۳ منطقه‌ای با ۲ منطقه‌ی همگن، حالت‌های ۵ منطقه‌ای با ۳ منطقه‌ی همگن و حالت‌های ۶ منطقه‌ای با ۴ منطقه‌ی همگن مورد بررسی قرار گرفتند. در جدول (۳) تعداد ایستگاه‌های جای گرفته در مناطق همگن در هر یک از این حالت‌ها ارائه شده است که در میان ۹ حالت مورد بررسی، حالت ۵ منطقه‌ای که مناطق آن بر اساس طول و عرض جغرافیایی و مساحت سطح زهکشی شکل گرفته‌اند، بیش‌ترین تعداد ایستگاه‌های جای گرفته در مناطق همگن را داراست و بنابراین مناطق حاصل از این حالت برای اجرای تحلیل فراوانی منطقه‌ای در حوزه‌ی آبخیز ارس انتخاب شدند.

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

در ادامه با استفاده از شاخص نکویی برازش Z، توزیع منطقه‌ای برگزیده برای هر یک از مناطق تشکیل شده در دو حالت منتخب برای حوزه‌ی سفیدرود و یک حالت منتخب برای حوزه‌ی ارس، مطابق نتایج مندرج در جدول ۴ مشخص شدند.

جدول ۳- تعداد ایستگاه‌های مناطق همگن برای حالاتی از منطقه‌بندی حوزه‌ی ارس که بیش‌ترین نسبت مناطق همگن به کل مناطق را دارند

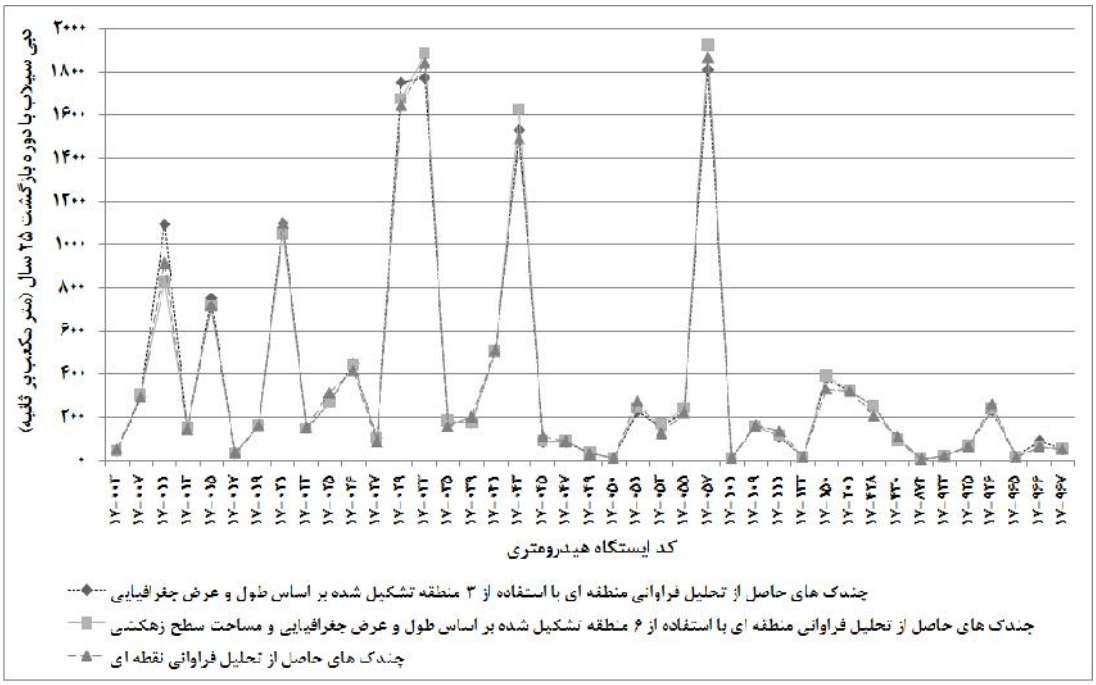
تعداد کل مناطق	۳		۵			۶				
ویژگی‌ها	LAC	LAPC	L	P	LA	L	C	P	LA
ایستگاه‌های موجود در مناطق همگن	۱۷	۱۷	۱۷	۱۵	۲۱	۱۷	۱۹	۱۵	۱۹

جدول ۴- توزیع‌های منطقه‌ای برگزیده برای مناطق تشکیل شده در حالات منتخب

حوزه‌ی آبخیز	سفیدرود									ارس								
ویژگی‌ها	L			LA							LA							
شماره منطقه	۱	۲	۳	۱	۲	۳	۴	۵	۶	۱	۲	۳	۴	۵
توزیع منطقه‌ای	GNO	GNO	WAK	GNO	PE3	GLO	GNO	WAK	GLO	GEV	WAK	GLO	GEV	WAK

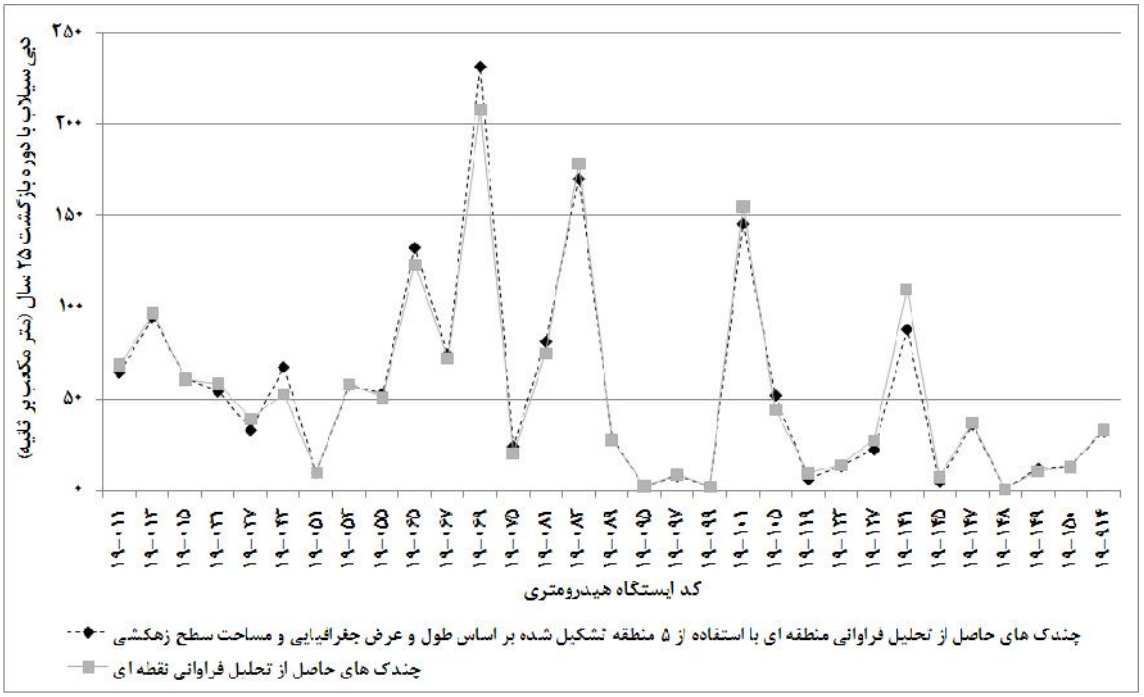
GNO معرف توزیع نرمال تعمیم یافته، GLO نماد توزیع لجستیک تعمیم یافته، GEV معرف توزیع مقادیر حدی تعمیم یافته، PE3 نماینده‌ی توزیع پیرسون تیپ III و WAK نشانه‌ی توزیع ویکبی است.

سپس برآورد چندک‌های سیلاب با دوره بازگشت ۲۵ سال برای هر یک از این حالات با استفاده از تئوری گشتاورهای خطی انجام گرفت. به‌علاوه تحلیل فراوانی درون سایتی یا نقطه‌ای سیلاب نیز برای هر یک از ایستگاه‌ها به‌صورت جداگانه صورت گرفت. چندک‌های برآورد شده برای سیلاب با دوره بازگشت ۲۵ سال به دو روش منطقه‌ای و نقطه‌ای در دو حوزه‌ی سفیدرود و ارس به‌ترتیب در شکل‌های (۳ و ۴) مشاهده می‌شود.



شکل ۳- برآورد سیلاب با دوره بازگشت ۲۵ سال با استفاده از تحلیل فراوانی منطقه‌ای و نقطه‌ای در حوزه‌ی سفیدرود

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]



شکل ۴- برآورد سیلاب با دوره بازگشت ۲۵ سال با استفاده از تحلیل فراوانی منطقه‌ای و نقطه‌ای در حوزه‌ی ارس

بررسی نتایج به‌دست آمده در این مطالعه نشان می‌دهد که در مواردی که ویژگی‌های مورد استفاده در خوشه‌بندی با ویژگی‌های به‌کار رفته در آزمون همگنی متفاوت باشند، کیفیت خوشه‌بندی که بر اساس شاخص صحت خوشه‌بندی تعیین می‌شود، لزوماً ارتباط مشخصی با همگنی مناطق تشکیل شده ندارد. افزون بر این، با توجه به نتایج مندرج در جدول (۱) به‌نظر می‌رسد به‌طور معمول با افزایش تعداد ویژگی‌های مورد استفاده در خوشه‌بندی دو حوزه‌ی مورد مطالعه، مقدار شاخص S_{av} کاهش می‌یابد.

همچنین با بررسی نتایج ارائه شده در جداول (۲ و ۳) می‌توان دریافت که بر اساس تعداد مناطق همگن تشکیل شده و ایستگاه‌های موجود در آنها، حالت‌های مربوط به تقسیم حوزه‌ی سفیدرود به ۳ منطقه با استفاده از طول و عرض جغرافیایی و ۶ منطقه بر اساس طول و عرض جغرافیایی و مساحت زهکشی، حالت‌های بهینه هستند. شایان ذکر است که حالت ۳ منطقه‌ای به دلیل تشکیل مناطقی با تعداد ایستگاه‌ها و داده‌های آماری بیش‌تری، امکان برآورد سیلاب با دوره‌های بازگشت بزرگ‌تر را فراهم می‌کند که از این نظر، حالت ۳ منطقه‌ای نسبت به حالت ۶ منطقه‌ای برتری دارد. به علاوه مطابق نتایج ثبت شده در جداول (۲ و ۳) می‌توان تقسیم حوزه‌ی آبخیز ارس به ۵ منطقه بر اساس طول و عرض جغرافیایی و مساحت سطح زهکشی را حالت بهینه‌ی منطقه‌بندی این حوزه از جهت تعداد مناطق همگن و فراوانی ایستگاه‌های جای گرفته در مناطق همگن دانست.

در نهایت با بررسی نتایج مربوط به چندک‌های برآورد شده‌ی سیلاب با استفاده از دو روش تحلیل فراوانی منطقه‌ای و نقطه‌ای در دو حوزه‌ی سفیدرود و ارس که در شکل‌های (۳ و ۴) نشان داده شده‌اند، مشخص می‌شود که مقادیر برآوردهای انجام شده با استفاده از تحلیل‌های منطقه‌ای و نقطه‌ای تا حد زیادی به هم نزدیک بوده و در مواردی که ایستگاهی فاقد آمار یا دارای آمار کوتاه مدت باشد، می‌توان با قابلیت اعتماد قابل قبولی از تحلیل فراوانی منطقه‌ای و منطقه‌بندی با استفاده از خوشه‌بندی ترکیبی برای برآورد چندک‌های سیلاب در آن ایستگاه استفاده کرد که این نتیجه با نتایج اعلام شده در مطالعه رائو و سرینیواس (۱۵) مطابقت دارد. در ضمن استفاده از توزیع پنج پارامتری ویکبی در مواردی که منطقه‌ای شرایط همگنی را ارضا نمی‌کند و یا سایر توزیع‌ها برازش مناسبی بر داده‌های یک منطقه نشان نمی‌دهند، به‌عنوان توزیع منطقه‌ای می‌تواند نتایج مناسبی را حاصل کند که این موضوع در هماهنگی با نتیجه اعلام شده در مطالعه آهنی و همکاران (۱) می‌باشد.

عملکرد الگوریتم‌های خوشه‌بندی ترکیبی در دیگر حوزه‌های آبخیز و چگونگی تأثیر استفاده از سایر ویژگی‌های مؤثر بر فرآیند تولید سیلاب در تشکیل بردارهای مشخصه و منطقه‌بندی حوزه‌ها از جمله

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

مواردی است که می‌تواند در مطالعات آتی در این زمینه مورد بررسی قرار گیرد. همچنین با توجه به اهمیت فراوان مطالعات مربوط به پدیده خشک‌سالی در ایران، مطالعه‌ی کاربرد روش‌های تحلیل خوشه‌ای و تحلیل فراوانی منطقه‌ای با استفاده از گشتاورهای خطی پیشنهاد می‌شود.

منابع

1. Ahani, A., S.S. Mousavi Nadoushani and M. Zare Karizi. 2012. Comparison of uncertainty of flood quantiles estimation in regional frequency analysis using Dalrymple flood index method and L-moments. Journal of Irrigation and Water Engineering, 10: 79-91. (In Persian)
2. Burn, D.H. 1989. Cluster analysis as applied to regional flood frequency. Journal of Water Resources Planning and Management, 115: 567-582.
3. Burn, D.H. 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology, 202: 212-230.
4. Burn, D.H and N.K. Goel. 2000. The formation of groups for regional flood frequency analysis. Hydrological Sciences Journal, 45: 97-112.
5. Chavoshi, S and W.N.A. Soleiman. 2009. Delineating pooling group for flood frequency analysis using soft computing. European Journal of Scientific Research, 35: 181-187.
6. Dalrymple, T. 1960. Flood frequency analysis. U.S. Geological Survey, Water Supply Paper, 1543-A., U.S. Department of the Interior, Washington DC, USA.
7. Greenwood, J.A., J.M. Landwehr, N.C. Matalas and J.R. Wallis. 1979. Probability weighted moments: Definition and relation to parameters of distribution expressible in inverse form. Water Resources Research, 15: 1049-1054.
8. Hosking, J.R.M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society, Series B, 52: 105-24.
9. Hosking, J.R.M. and J.R. Wallis. 1997. Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, New York, USA, 224 pp.
10. Kumar, R., R.D. Singh and S.M. Seth. 1999. Regional flood formula for seven subzones of India. Journal of Hydrologic Engineering, 4: 240-244.
11. Kumar, R and C. Chatterjee. 2005. Regional flood frequency analysis using L-moment for North Brahmaputra region of India. Journal of Hydrologic Engineering, 10: 1-7.
12. Kumar, R., C. Chatterjee, S. Kumar, A.K. Lohani and R.D. Singh. 2003. Development of regional flood frequency relationships using L-moments for Middle Ganga Plains Subzone 1(f) of India. Water Resources Management, 17: 243-257.
13. Mosley, M.P. 1981. Delimitation of New Zealand hydrological regions. Journal of Hydrology, 49: 173-192.
14. Nathan, R.J. and T.A. McMahon. 1990. Identification of homogeneous regions for the purposes of regionalisation. Journal of Hydrology, 121: 217-238.
15. Rao, A.R. and V.V. Srinivas. 2006. Regionalization of watersheds by hybrid cluster analysis. Journal of Hydrology, 318: 37-56.
16. Rao, A.R. and V.V. Srinivas. 2006. Regionalization of watersheds by fuzzy cluster analysis. Journal of Hydrology, 318: 57-79.
17. Rao, A.R. and V.V. Srinivas. 2008. Regionalization of Watersheds: An Approach Based on Cluster Analysis. Series: Water Science and Technology Library, Vol. 58, Springer Publishers, Berlin, Germany, 241 pp.
18. Sarkar, S., N.K. Goel and B.S. Mathur. 2010. Development of isopluvial map using L-moment approach for Tehri-Garhwal Himalaya. Stochastic Environmental Research and Risk Assessment, 24: 411-423.
19. Srinivas, V.V., S. Tripathi, A.R. Rao and R.S. Govindaraju. 2008. Regional flood frequency analysis by combining self-organizing feature maps and fuzzy clustering. Journal of Hydrology, 348: 148-166.
20. Tasker, G.D. 1982. Comparing methods of hydrologic regionalization. Water Resources Bulletin,18: 965-970.
21. Wallis, J.R. 1982. Hydrologic problems associated with oil shale development. In Environmental Systems and Management, edited by S. Rinaldi, 85-102 pp., Amsterdam, North-Holland.

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]

Regional Flood Frequency Analysis by Hybrid Cluster Analysis and L-moments

Ali Ahani[1], Samad Emamgholizadeh[2], Seyed Saeid Mousavi Nadoushani[3] and Khalil Azhdari[4]

1- M.Sc. Student, Shahrood University (Corresponding author: ali.ahani66@yahoo.com)
2 and 4- Associate Professor, Shahrood University
3- Assistant Professor, Shahid Beheshti University
Received: May 11, 2013 Accepted: January 26, 2014

Abstract

Recent studies show that cluster analysis methods are one of the most useful techniques of regionalization of watersheds for regional flood frequency analysis. In this study combination of Ward and K-means clustering algorithms is used for regionalization of gauging stations in Sefidrood and Aras watersheds in order to use advantages and decrease influences of disadvantages of two main categories of cluster analysis methods. Also effects of selection and usage of some geographical, physiographic and meteorological attributes and their combinations on homogeneity of regions formed by cluster analysis is studied. Assessment of homogeneity of final regions and performance of regional flood frequency analysis using L-moments show that combination of longitude, latitude and drainage area as attributes used for regionalization of Sefidrood and Aras watersheds may be the best option to form maximum number of homogeneous regions. Furthermore, Wakeby distribution may be used as regional distribution for heterogeneous regions.

Keywords: Regional frequency analysis, L-moments, Cluster analysis, Hybrid clustering

[Downloaded from jwmr.sanru.ac.ir on 2026-04-22]