دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشنامه اصلاح گیاهان زراعی

منطقه‌بندی حوزه‌های آبخیز با به‌کارگیری نوعی از شبکه‌های عصبی مصنوعی به منظور تحلیل فراوانی منطقه‌ای سیلاب

علی آهنی[۱]، صمد امامقلی‌زاده[۲]، سیدسعید موسوی ندوشنی[۳] و خلیل اژدری[۲]

۱- دانشجوی کارشناسی ارشد، دانشگاه صنعتی شاهرود، (نویسنده مسوول: ali.ahani66@yahoo.com)
۲- دانشیار، دانشگاه صنعتی شاهرود
۳- استادیار، دانشگاه شهید بهشتی

تاریخ دریافت: ۹۳/۳/۱ تاریخ پذیرش: ۹۴/۷/۲۲

چکیده

نگاشت‌های خودسازمانده یکی از انواع شبکه‌های عصبی مصنوعی هستند که قابلیت آن‌ها در تشخیص الگو و خوشه‌بندی داده‌ها، آن‌ها را به ابزاری قابل توجه در زمینه‌ی منطقه‌بندی حوزه‌های آبخیز به منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب تبدیل کرده است. در این مطالعه، توانایی نگاشت‌های خودسازمانده در منطقه‌بندی حوزه‌ی آبخیز سفیدرود به منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب با استفاده از الگوریتم گشتاورهای خطی مورد بررسی قرار گرفته است. نتایج به دست آمده بیانگر آن است که نگاشت‌های خودسازمانده می‌توانند به عنوان روشی قابل قبول در زمینه‌ی خوشه‌بندی داده‌ها و منطقه‌بندی حوزه‌های آبخیز به‌کار گرفته شوند. بررسی مقادیر شاخص‌های صحت خوشه نشان داد که این شاخص‌ها به‌تنهایی نمی‌توانند تعیین‌کننده‌ی منطقه‌بندی مطلوب برای تحلیل فراوانی منطقه‌ای باشند، بلکه این وضعیت همگنی مناطق است که عامل اساسی در تعیین منطقه‌بندی مطلوب است. بر اساس وضعیت همگنی مناطق و بزرگی مناطق تشکیل شده، حالت‌های دو منطقه‌ای حاصل از به‌کارگیری الگوریتم‌های وارد و نگاشت‌های خودسازمانده به عنوان حالت‌های بهینه به منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب برای حوزه‌ی آبخیز سفیدرود، انتخاب شدند. افزون بر این، نتایج حاصل از برآورد سیلاب در تحلیل نقطه‌ای و دو تحلیل منطقه‌ای گویای آن بود که برآوردهای حاصل از دو تحلیل منطقه‌ای بسیار به هم نزدیک بوده و اختلاف نسبی آن‌ها به‌طور میانگین در حدود ۱٪ است. هم‌چنین، اختلاف نسبی آن‌ها با برآوردهای حاصل از تحلیل نقطه‌ای نیز در هیچ‌یک از ایستگاه‌ها از ۱۷٪ تجاوز نمی‌کند و مقدار میانگین آن تقریباً برابر ۸٪ است.

واژه‌های کلیدی: نگاشت خودسازمانده، منطقه‌بندی، خوشه‌بندی، گشتاورهای خطی

مقدمه

تحلیل فراوانی سیلاب[۱]، روشی برای برآورد بزرگی سیلاب با دوره‌ی بازگشت معین و یا برآورد دوره‌ی بازگشت یک سیلاب با بزرگی معین در محل موردنظر است. تحلیل فراوانی سیلاب دارای دو رویکرد اصلی نقطه‌ای[۲] و منطقه‌ای[۳] است.

در تحلیل فراوانی منطقه‌ای سیلاب با گردآوری حوزه‌ها در مناطق همگن و افزایش داده‌های آماری مورد استفاده برای استخراج توزیع فراوانی، میزان اعتمادپذیری افزایش یافته و امکان برآورد سیلاب با دوره‌های بازگشت طولانی‌تر فراهم می‌شود. هم‌چنین رویکرد منطقه‌ای، امکان برآورد سیلاب برای نقاط فاقد داده‌های سیلاب را فراهم می‌کند.

پس از بررسی کیفیت و صحت داده‌ها، نخستین مرحله‌ی اساسی در تحلیل فراوانی منطقه‌ای سیلاب، منطقه‌بندی[۴] است که در آن هدف، تشکیل مناطقی از حوزه‌های دارای مکانیزم تولید سیلاب مشابه است. رویکردهای مختلفی برای اجرای منطقه‌بندی وجود دارند که یکی از پرکاربردترین آن‌ها استفاده از روش‌های تحلیل خوشه‌ای[۵] است. پژوهشگران مختلفی روی ابعاد گوناگون استفاده از روش‌های تحلیل خوشه‌ای برای اجرای منطقه‌بندی در تحلیل فراوانی منطقه‌ای سیلاب مطالعه کرده‌اند. روش‌های خوشه‌بندی را می‌توان به دو گروه خوشه‌بندی سخت[۶] و خوشه‌بندی فازی[۷] تقسیم کرد که الگوریتم‌های خوشه‌بندی سخت، خود به دو گروه الگوریتم‌های سلسله‌مراتبی[۸] و الگوریتم‌های افرازی[۹] قابل تقسیم‌بندی هستند. همچنین به منظور بهره‌گیری از نقاط قوت هر دو گروه الگوریتم‌های سلسله‌مراتبی و افرازی، از الگوریتم‌های خوشه‌بندی ترکیبی استفاده می‌شود (۱).

نوع خاصی از شبکه‌های عصبی مصنوعی به نام نگاشت‌های مشخصه‌ی خودسازمانده[۱۰] یا نگاشت‌های خودسازمانده[۱۱] دارای توانایی ویژه‌ای در زمینه‌ی یادگیری بدون نظارت[۱۲] و خوشه‌بندی داده‌ها هستند و ازاین‌رو می‌توانند برای منطقه‌بندی حوزه‌های آبخیز مورد استفاده قرار گیرند. دو مدل مختلف از شبکه‌های عصبی خود سازمانده وجود دارد (۶) که عبارتند از: ویلشاو- فون‌درمالزبرگ[۱۳] (۲۰) و مدل کوهونن[۱۴] (۱۰). مدل نخست به طور ویژه برای نگاشت در جایی که ابعاد فضای ورودی با ابعاد فضای خروجی یکسان است کاربرد دارد، در حالی که مدل دوم قادر به تولید نگاشت‌هایی از فضاهای ورودی با تعداد ابعاد بالا به فضاهای خروجی با تعداد ابعاد پایین‌تر است. در مطالعات هیدرولوژیک مدل کوهونن بیش‌تر مورد توجه قرار گرفته است. هال و مینز (۴،۳) از شبکه‌ی کوهونن برای منطقه‌بندی مجموعه‌ای از ۱۰۱ ایستگاه هیدرومتری در جنوب غربی انگلیس و ولز استفاده کردند.

هال و همکاران (۵) SOFM یک‌بعدی[۱۵] را برای سه مجموعه از جنوب غربی انگلیس و ولز، ولز و اسکاتلند و جزایر جاوه و سوماترا در اندونزی به کار گرفتند. این مطالعات در مورد صحت‌سنجی مناطق تشکیل شده با استفاده از شاخص‌های ناهمگنی[۱۶]، چیزی گزارش نکردند. با این حال،

1- Flood Frequency Analysis
2- At-Site
3- Regional
4- Regionalization
5- Cluster Analysis
6- Hard Clustering
7- Fuzzy Clustering
8- Hierarchical Algorithms
9- Partitional Algorithms
10- Self-Organizing Feature Maps (SOFM)
11- Self-Organizing Maps (SOM)
12- Unsupervised Learning
13- Willshaw – Von Der Malsburg
14- Kohonen
15- 1-D
16- Heterogeneity Measures

جینگیی و هال (۸) شاخص‌های ناهمگنی هاسکینگ و والیس (۷) را برای ارزیابی همگنی[1] مناطق مشخص شده به وسیله‌ی SOFM یک بعدی از ۸۶ ایستگاه هیدرومتری در چین به‌کار گرفتند.

لین و چن (۱۳)، SOFM را برای تشکیل مناطق به منظور تحلیل فراوانی منطقه‌ای سیلاب به کار گرفتند. آن‌ها اعلام کردند که SOFM نسبت به دیگر روش‌های خوشه‌بندی توانایی بالاتری در تشکیل مناطق همگن دارند و آن را به‌عنوان گزینه‌ای مناسب برای تشکیل مناطق همگن معرفی کردند. سرینیواس و همکاران (۱۷)، یک رویکرد دومرحله‌ای مبتنی بر SOFM را برای منطقه‌بندی حوزه‌های آبخیز معرفی کردند. در این رویکرد SOFM در ترکیب با الگوریتم فازی c-means به کار گرفته شد. نتایج به‌دست آمده نشان داد که عملکرد این روش در زمینه‌ی برآورد چندک‌های سیلاب در ایستگاه‌های فاقد آمار در مقایسه با روش‌های تحلیل رگرسیون و تحلیل همبستگی کانونی[2] بهتر است.

دی‌پرینیزیو و همکاران (۲)، SOFM را برای منطقه‌بندی بیش از ۳۰۰ حوزه در ایتالیا به‌کار گرفتند. آن‌ها SOFM را برای منطقه‌بندی بر اساس ویژگی‌های حوزه‌ها و هم‌چنین متغیرهای حاصل از تحلیل مؤلفه‌های اصلی[3] و تحلیل همبستگی کانونی مورد استفاده قرار دادند و اعلام کردند که استفاده از PCA و CCA در بهبود عملکرد SOFM در زمینه‌ی منطقه‌بندی به منظور تحلیل فراوانی منطقه‌ای تأثیر قابل توجهی دارد. هم چنین لی و همکاران (۱۲)، SOFM را برای خوشه‌بندی ۵۳ حوزه در آلمان مورد استفاده قرار دادند. آن‌ها خوشه‌بندی را یک مرتبه بر اساس ویژگی‌های حوزه‌ها و بار دیگر بر مبنای پاسخ هیدرولوژیک حوزه‌ها اجرا کردند و بیان کردند که خوشه‌بندی‌های حاصل از این دو رویکرد در ۶۷٪ موارد دارای مشابهت هستند.

کار و همکاران (۹) نیز عملکرد روش‌های مختلف خوشه‌بندی از جمله SOFM را در مطالعه‌ای بر روی یک حوزه‌ی آبخیز در هندوستان مورد بررسی قرار داده و از مقایسه‌ی نتایج آن‌ها برای تعیین تعداد بهینه‌ی مناطق استفاده کردند.

تاث (۱۹) از SOFM برای گروه‌بندی ۴۴ حوزه‌ی آبخیز در ایتالیا با استفاده از ویژگی‌های سری‌های زمانی بارندگی و جریان استفاده کرد و نتیجه گرفت که این روش می‌تواند یک راه مطمئن برای نشان دادن بهتر ارتباط ویژگی‌های هیدرولوژیک و اقلیمی حوزه‌ها باشد.

رضوی و کولیبالی (۱۵)، کاربرد تکنیک‌های تحلیل خوشه‌ای غیرخطی شامل SOFM را برای تشکیل خوشه‌های همگن هیدرولوژیک حوزه‌های آبخیز اونتاریو در کانادا مورد مطالعه قرار دادند و اعلام کردند که این روش‌های غیرخطی می‌توانند ابزار توانمندی برای منطقه‌بندی حوزه‌های فاقد آمار باشند.

از دیگر کاربردهای SOFM در پژوهش‌های مرتبط با مهندسی آب می‌توان به مطالعه‌ی طباطبائی و همکاران (۱۸) اشاره کرد که در آن از SOFM برای اجرای خوشه‌بندی به منظور برآورد غلظت رسوب معلق روزانه ایستگاه هیدرومتری سیرا استفاده کردند.

هدف این مطالعه، بررسی کارایی نگاشت‌های خودسازمانده در منطقه‌بندی حوزه‌ی آبخیز سفیدرود به‌منظور تحلیل فراوانی منطقه‌ای سیلاب مورد مطالعه قرار گرفته است. بدین منظور کیفیت خوشه‌بندی اجرا شده توسط این روش با استفاده از شاخص صحت خوشه‌بندی ارزیابی می‌شود.

همگنی مناطق تشکیل شده نیز با استفاده از شاخص‌های ناهمگنی مورد بررسی قرار می‌گیرند. هم‌چنین نتایج به‌دست آمده با نتایج حاصل از به‌کارگیری تعدادی از الگوریتم‌های خوشه‌بندی سلسله‌مراتبی، مقایسه می‌شوند تا میزان کارایی این روش به نسبت روش‌های معمول سنجیده شود.

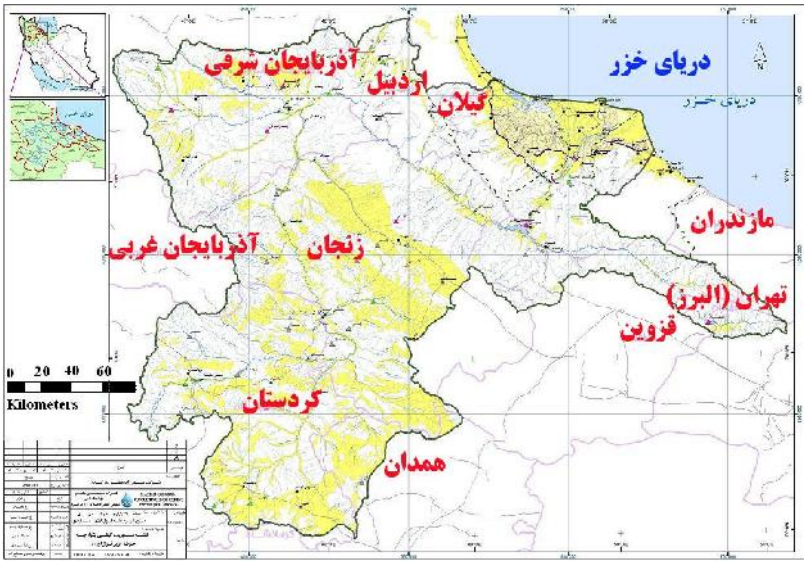
مواد و روش‌ها

معرفی حوزه‌ی آبخیز سفیدرود

حوزه‌ی آبخیز سفیدرود بعد از حوزه‌ی دریاچه‌ی نمک از نظر اراضی قابل آبیاری بزرگ‌ترین حوزه‌ی کشور است .مساحت این حوزه ۶۳۹۴۵ کیلومتر مربع می‌باشد 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 و در تقسیم‌بندی‌های طرح جامع آب به عنوان حوزه‌ی سوم از منطقه‌ی اول مطالعاتی مشخص شده است. این حوزه‌ی آبخیز از دو شاخه‌ی رودخانه‌ای اصلی به نام قزل اوزن و شاهرود تشکیل یافته است که در محل سد سفیدرود به هم می‌پیوندند و رودخانه سفیدرود را تشکیل می‌دهند. ذینفعان این حوزه، مطابق شکل یک استان‌های آذربایجان شرقی، اردبیل، تهران، زنجان، قزوین، کردستان، گیلان و همدان می‌باشند. در این تحقیق تعداد ۳۹ ایستگاه هیدرومتری این حوزه که دارای اطلاعات مورد نیاز جهت اجرای خوشه‌بندی ترکیبی بودند، انتخاب شدند.

1- Homogeneity
2- Canonical Correlation Analysis (CCA)
3- Principal Component Analysis (PCA)

[DOR: 20.1001.1.22516174.1395.7.14.11.3]
[DOI: 10.29252/jwmr.7.14.118]



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

Figure 1. Geographical location of Sefidroud (Ghezel-Owzan) basin (Mahab Ghodss Consulting Engineering Company)

نرم‌افزار و زبان برنامه‌نویسی R

در این مطالعه، برای اجرای محاسبات و عملیات مربوط به خوشه‌بندی با استفاده از شبکه‌ی کوهونن و تحلیل فراوانی منطقه‌ای از محیط زبان برنامه‌نویسی آماری R (نسخه ۳.۰.۲) و بسته‌های kohonen، cluster و lmomRFA استفاده شده است.

غربال کردن داده‌ها

پیش از تشکیل مناطق، شاخص ناجوری[1] D (۷) برای شناسایی ایستگاه‌هایی که از نظر آماری با سایر ایستگاه‌های مورد بررسی ناسازگارند، محاسبه می‌شود. این شاخص بر اساس نسبت‌های گشتاورهای خطی محاسبه می‌شود و برای مجموعه‌ای شامل ۱۵ ایستگاه یا بیش‌تر، ایستگاه‌های دارای D برابر ۳ یا بزرگ‌تر، به عنوان ایستگاه ناجور در نظر گرفته می‌شوند (۷). جزئیات بیش‌تر برای مطالعه در مورد گشتاورهای خطی در منبع (۷) در دسترس است.

انتخاب ویژگی‌ها و تشکیل بردارهای مشخصه

از میان ویژگی‌های مؤثر بر واکنش سیلابی ایستگاه‌های مورد نظر در حوزه، با توجه به آمار و اطلاعات مورد اعتماد در دسترس، طول و عرض جغرافیایی به عنوان ویژگی‌های جغرافیایی، مساحت سطح زهکشی و ارتفاع حوزه به عنوان ویژگی‌های فیزیوگرافیک، ضریب رواناب به‌عنوان نماینده‌ای از ویژگی‌های پوشش سطح، جنس خاک و نفوذپذیری حوزه و متوسط بارندگی سالانه به عنوان یک ویژگی هواشناسی جهت استفاده در عملیات خوشه‌بندی انتخاب شدند. مشخصات آماری کلی این ویژگی‌ها برای ایستگاه‌های مورد مطالعه در جدول یک ارائه شده است.

هر بردار مشخصه[2] از ویژگی‌های تجدید مقیاس شده‌ی[3] (بدون بعد) یک حوزه‌ی آبخیز تشکیل می‌شود. در میان ویژگی‌های منتخب در این مطالعه، مساحت سطح زهکشی ایستگاه‌ها با لگاریتم آن‌ها جایگزین شد.

داده‌های موجود برای هر ویژگی، برای خنثی کردن تفاوت‌ها در واریانس‌ها و بزرگی نسبی آن‌ها، تجدید مقیاس می‌شوند (۱۴). در این مطالعه برای تجدید مقیاس ویژگی‌ها از روش استانداردسازی[4] داده‌ها استفاده شده است.

اگر y_k نماینده‌ی kامین بردار مشخصه در فضای n بعدی با محورهای مختصاتی به نام‌های $(y_1, \ldots, y_n)$ باشد، که $yk=[y1k, \ldots, y_{nk}] \in R^n$ و y_{ik} معرف مقدار ویژگی i در kامین بردار مشخصه‌ی n بعدی y_k است، که برای استانداردسازی داده‌ها از رابطه یک استفاده می‌شود:

رابطه (۱)
$$x_{ik} = \frac{y_{ik} - «_{ik}}{\pm_i} \qquad for \quad 1 \le i \le n, 1 \le k \le N$$

که در آن x_{ik} : نشان‌دهنده مقدار تجدید مقیاس شده y_{ik} است، $\pm_i$: معرف انحراف معیار ویژگی i است و N : تعداد بردارهای مشخصه‌ی n بعدی را نشان می‌دهد.

الگوریتم‌های سلسله‌مراتبی

الگوریتم‌های خوشه‌بندی سلسله‌مراتبی که گروهی از الگوریتم‌های پرکاربرد در زمینه‌ی منطقه‌بندی هستند، خود به دو زیرگروه الگوریتم‌های متراکم‌کننده[5] و الگوریتم‌های تقسیم‌کننده[6] دسته‌بندی می‌شوند که از این میان، الگوریتم‌های متراکم‌کننده در منطقه‌بندی کاربرد بیش‌تری یافته‌اند.

برای مجموعه‌ای از N بردار مشخصه، روش‌های خوشه‌بندی سلسله‌مراتبی متراکم کننده با N خوشه‌ی تکی آغاز می‌شوند. یک معیار فاصله مانند فاصله‌ی اقلیدسی برای ارزیابی اختلاف بین دو خوشه و یا بردارهای مشخصه انتخاب می‌شود. خوشه‌هایی که کم‌ترین اختلاف را با یکدیگر دارند مشخص شده و به هم می‌پیوندند. این امر به تشکیل $N-2$

1- discordancy measure
2- Feature Vector
3- Rescaled
4- Standardization
5- Agglomerative
6- Divisive

[DOR: 20.1001.1.22516174.1395.7.14.11.3]
[DOI: 10.29252/jwmr.7.14.118]

خوشه‌ی تکی و یک خوشه شامل دو بردار مشخصه منجر می‌شود. فرآیند تشخیص و ادغام دو خوشه‌ی نزدیک تا رسیدن به تعداد مطلوب خوشه‌ها تکرار می‌شود. به طور کلی تعداد خوشه‌های موجود پس از n مرتبه ادغام برابر $N\text{-}n$ است (۱۴).

جدول ۱- مشخصات آماری ویژگی‌های برگزیده‌ی ایستگاه‌های مورد مطالعه

Table 1. Statistical characteristics of selected attributes of the interested stations

ویژگی	میانگین	انحراف معیار
طول جغرافیایی (dd)	۴۸/۷۰	۱/۳۰
عرض جغرافیایی (dd)	۳۶/۵۹	۰/۷۴
سطح زهکشی (km^2)	۵۸۱۷/۵۱	۱۲۰۰۴/۵۱
متوسط بارندگی سالانه (mm)	۴۶۷/۷۰	۳۲۳/۶۲
ضریب رواناب	۰/۴۷	۰/۳۳

در این مطالعه از میان الگوریتم‌های خوشه‌بندی سلسله‌مراتبی چهار الگوریتم تک‌پیوند[1]، تمام‌پیوند[2]، پیوندمتوسط[3] و الگوریتم وارد[4] مورد استفاده و بررسی قرار گرفته‌اند. در الگوریتم تک‌پیوند، فاصله‌ی بین دو خوشه، عبارت است از فاصله‌ی بین نزدیک‌ترین جفت بردارهای مشخصه که هر یک از آن‌ها متعلق به یکی از دو خوشه هستند. در الگوریتم تمام‌پیوند، فاصله‌ی میان دو خوشه به صورت فاصله‌ی دو بردار مشخصه از دو خوشه تعریف می‌شود که بیش‌ترین فاصله را با یکدیگر دارند. در الگوریتم پیوند متوسط، فاصله‌ی بین دو خوشه به صورت فاصله‌ی متوسط بین آن‌ها بیان می‌شود. این فاصله‌ی متوسط می‌تواند به شکل میانه‌ی فواصل بین بردارهای مشخصه یا میانگین‌های وزنی یا غیروزنی فواصل بردارهای مشخصه‌ی دو خوشه تعریف شود (۱۴).

الگوریتم وارد یک روش پرکاربرد برای مطالعات منطقه‌بندی در هیدرولوژی 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 (۱۴). روابط مختلفی برای سنجش فاصله‌ی میان هر دو بردار مشخصه وجود دارد که در این مطالعه از تعریف فاصله‌ی اقلیدسی استفاده شده است.

نگاشت‌های خودسازمانده کوهونن

SOFM کوهونن دارای یک لایه‌ی ورودی و یک لایه‌ی خروجی است که هر یک شامل تعدادی گره است. تعداد گره‌ها در لایه‌ی ورودی مساوی تعداد ویژگی‌های در نظر گرفته شده برای منطقه‌بندی است. هر گره در لایه‌ی ورودی به وسیله‌ی اتصالات سیناپتیک[5] به تمام گره‌ها لایه‌ی خروجی متصل می‌شود. به همراه هر اتصال، یک وزن اتصال وجود دارد (۶).

تعداد گره‌ها در لایه‌ی ورودی SOFM برابر ابعاد بردار مشخصه، یعنی n است. لایه‌ی خروجی، که لایه‌ی رقابتی[6] یا لایه‌ی کوهونن نیز خوانده می‌شود، دارای m گره است که در یک شبکه که معمولاً یک یا دو بعدی است، سازماندهی شده است. مقدار m می‌تواند به صورت حداکثر تعداد مورد نظر برای تشکیل خوشه‌ها انتخاب شود (۱۴). در زمینه‌ی منطقه‌بندی در هیدرولوژی، مقدار m عموماً به گونه‌ای انتخاب می‌شود که بزرگ‌تر از تعداد خوشه‌های مورد انتظار C_{EXP} باشد. برای یک SOFM یک بعدی، هال و مینز (۴) مقدار m را برابر حداقل $2C_{EXP}$ انتخاب کردند، در حالی که هال و همکاران (۵) مقدار آن را حداقل برابر $3C_{EXP}$ در نظر گرفتند. برای مطالعه‌ی بیش‌تر در مورد الگوریتم نگاشت خودسازمانده کوهونن می‌توان به منبع (۶) مراجعه کرد.

در این مطالعه، شبکه‌های کوهونن خطی مورد بررسی قرار گرفته‌اند.

شاخص‌های صحت خوشه

خوشه‌های تشکیل شده به صورت بصری و با استفاده از شاخص‌های صحت خوشه[7] برای تعیین کیفیت خوشه‌بندی و تعداد بهینه‌ی خوشه‌ها تفسیر می‌شوند. در مطالعه‌ی حاضر از میان شاخص‌های سنجش صحت خوشه، چهار شاخص عرض نیم‌رخ[8]، شاخص دیویس- بولدین[9]، شاخص دان[10] و شاخص کالینسکی- هاراباز[11] به دلیل عملکرد قابل قبول در مطالعات پیشین انتخاب شده‌اند (۱۴).

مقدار شاخص عرض نیم‌رخ در محدوده‌ی ۱- و ۱+ جای می‌گیرد. چنان چه مقدار این شاخص نزدیک به یک باشد، می‌توان این‌طور نتیجه‌گیری کرد که خوشه‌بندی از کیفیت خوبی برخوردار است و از سوی دیگر چنان چه مقدار آن به منهای یک نزدیک باشد، می‌توان استنتاج کرد که خوشه‌بندی به شکل مناسبی صورت نگرفته است (۱۴). یک مقدار کوچک برای شاخص دیویس- بولدین، نشان‌دهنده‌ی یک افراز خوب است که متناظر با خوشه‌هایی فشرده است که مراکز آن‌ها از یکدیگر دور هستند. در مورد شاخص دان، تعداد بهینه‌ی خوشه‌ها، گزینه‌ای است که مقدار این شاخص به ازای آن بیشینه می‌شود.

در مورد شاخص کالینسکی- هاراباز نیز، مقدار بیشینه‌ی شاخص، معرف افراز بهینه است (۱۴).

آزمون همگنی مناطق

در این تحقیق همگنی مناطق حاصل از عملیات خوشه‌بندی، با استفاده از شاخص‌های ناهمگنی H مورد ارزیابی قرار گرفتند. 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 $1\le H<2$ باشد، منطقه نسبتاً ناهمگن و اگر $H\ge 2$ باشد، منطقه کاملاً ناهمگن است (۷). مناطقی که بیش‌تر به حالت ناهمگنی نزدیک هستند، به منظور بهبود وضعیت همگن‌شان می‌توانند با حذف یک یا چند ایستگاه که تأثیر بیش‌تری در افزایش ناهمگنی

1- Single Linkage
2- Complete Linkage
3- Average Linkage
4- Ward's Algorithm
5- Synaptic
6- Competitive Layer
7- Cluster Validity Measure
8- Silhouette Width
9- Davies - Bouldin Index
10- Dunn's Index
11- Calinski - Harabasz index

[DOR: 20.1001.1.22516174.1395.7.14.11.3]
[DOI: 10.29252/jwmr.7.14.118]

دارند یا جا به جایی محدود برخی ایستگاه‌ها بین خوشه‌ها و یا باز تعریف مناطق در صورت نیاز، اصلاح شوند (۷).

برآورد چندک‌های سیلاب

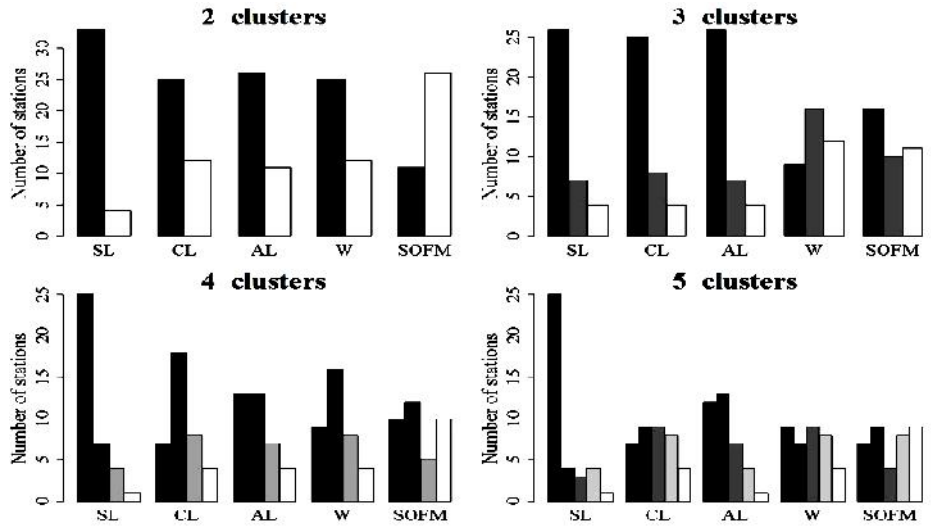
هدف این مرحله اجرای آزمون‌های نکویی برازش منطقه‌ای برای تشخیص و برازش یک توزیع فراوانی سیلاب مناسب بر داده‌های سیلاب ایستگاه‌ها در یک منطقه است. توزیع برازش یافته، در نهایت برای به دست آوردن برآوردهای چندک[1] سیلاب برای طراحی هیدرولوژیک استفاده می‌شوند. در این مطالعه به منظور شناسایی توزیع بهینه برای 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 به عنوان توزیع منطقه‌ای انتخاب شود. نزدیک‌تر بودن این مقدار به صفر می‌تواند نشان‌دهنده بهینه بودن انتخاب توزیع مورد نظر باشد (۷). در این مطالعه چگونگی برازش توزیع‌های سه پارامتری لجستیک تعمیم یافته[2]، مقادیر حدی تعمیم یافته[3]، نرمال تعمیم یافته[4]، پیرسون تیپ III[5] و پارتوی تعمیم یافته[6] با استفاده از شاخص نکویی برازش Z مورد بررسی قرار گرفته است. توزیع منتخب منطقه‌ای معمولاً با اعمال ضریبی مشخص برای هر ایستگاه، تبدیل به توزیع ویژه‌ی آن ایستگاه می‌گردد. این ضریب می‌تواند متوسط دبی حداکثر لحظه‌ای هر ایستگاه باشد.

نتایج و بحث

پیش از بررسی وضعیت همگنی مناطق شاخص ناجوری D برای شناسایی ایستگاه‌های ناجور محاسبه شد که از میان ۳۹ ایستگاه هیدرومتری مورد مطالعه، مقدار این شاخص برای دو ایستگاه از مقدار مجاز $D=3$ تجاوز کرد و این دو ایستگاه از ادامه فرآیند منطقه‌بندی و تحلیل فراوانی سیلاب کنار گذاشته شدند.

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 (۱۴). از این رو تعداد ایستگاه‌های موجود در هر منطقه و سال‌های آماری موجود برای هر یک از آن‌ها عاملی تعیین‌کننده در انتخاب تعداد خوشه‌ها است. در این مطالعه، نتایج حاصل از اختصاص ایستگاه‌های موجود در حوزه‌ی آبخیز سفیدرود بزرگ به دو تا پنج خوشه یا منطقه مورد ارزیابی قرار گرفته است.

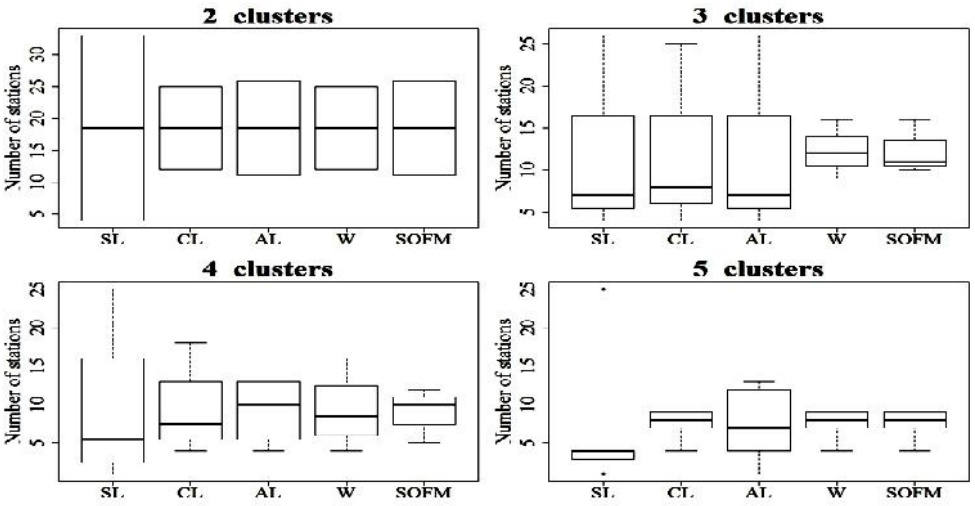
در ادامه خوشه‌بندی ۳۷ ایستگاه هیدرومتری مورد نظر با استفاده از چهار الگوریتم سلسله مراتبی معرفی شده و شبکه‌های کوهونن با دو تا پنج گره در لایه‌ی خروجی برای حالت‌های دو تا پنج منطقه‌ای اجرا شد. در شکل دو اندازه‌ی هریک از خوشه‌ها یا مناطق تشکیل شده با استفاده از الگوریتم‌های خوشه‌بندی مورد استفاده در مراحل دو تا پنج خوشه‌ای نشان داده شده است. منظور از اندازه‌ی خوشه‌ها یا مناطق، همان تعداد ایستگاه‌های جای گرفته در هر خوشه یا منطقه است. شکل سه نیز نمودار جعبه‌ای اندازه‌ی مناطق تشکیل شده به‌وسیله‌ی الگوریتم‌های خوشه‌بندی در حالت‌های دو تا پنج خوشه‌ای را نشان می‌دهد. آن‌چنان که در شکل‌های دو و سه مشاهده می‌شود، SOFM و الگوریتم سلسله‌مراتبی وارد نسبت به سایر الگوریتم‌ها گرایش بیش‌تری به تشکیل خوشه‌هایی با تعداد اعضای نزدیک به هم دارند. هم‌چنین به‌نظر می‌رسد الگوریتم تک‌پیوند بیش‌تر متمایل به تشکیل یک خوشه‌ی بزرگ و تعدادی خوشه‌ی کوچک است. در نمودارهای جعبه‌ای مربوط به الگوریتم تک‌پیوند دامنه‌ی تغییرات اندازه‌ی خوشه‌های تشکیل شده اغلب بزرگ‌تر از دامنه‌ی تغییرات اندازه‌ی خوشه‌های سایر روش‌ها است.



شکل ۲- اندازه‌ی خوشه‌ها یا مناطق تشکیل شده به‌وسیله‌ی الگوریتم‌های خوشه‌بندی، SL، CL، AL، W و SOFM به‌ترتیب معرف الگوریتم‌های تک‌پیوند، تمام‌پیوند، پیوندمتوسط، الگوریتم وارد و نگاشت خودسازمانده هستند.

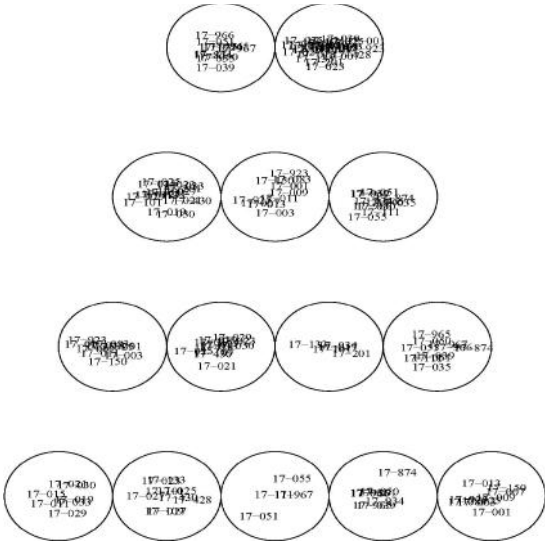
Figure 2. Sizes of clusters or regions formed by clustering algorithms; SL, CL, AL, W and SOFM denote Single Linkage, Complete Linkage, Average Linkage, Ward's algorithm and Self-Organizing Feature Map, respectively

1- Quantile
2- Generalized Logistic (GLO)
3- Generalized Extreme Values (GEV)
4- Generalized Normal (GNO)
5- Pearson type III (PE3)
6- Generalized Pareto (GPA)

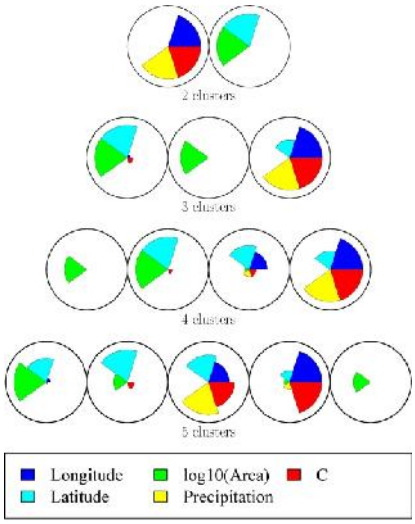


شکل ۳- نمودار جعبه‌ای اندازه‌ی مناطق تشکیل شده به‌وسیله‌ی الگوریتم‌های خوشه‌بندی
Figure 3. Boxplot for sizes of regions formed by clustering algorithms

شکل چهار نحوه‌ی اختصاص ایستگاه‌های مورد بررسی به گره‌های لایه‌ی خروجی نگاشت خود سازمانده در هر یک از حالت‌های دو تا پنج منطقه‌ای را نشان می‌دهد که هر گره متناظر با یک منطقه است. هم‌چنین شکل پنج نشان‌دهنده‌ی وزن ویژگی‌های تجدید مقیاس شده‌ی ایستگاه‌های اختصاص یافته به هر گره در هر یک از مراحل دو تا پنج خوشه‌ای است. برای نمونه، در حالت دو منطقه‌ای، ایستگاه‌های گره اول دارای وزن مقادیر تجدید مقیاس شده‌ی عرض جغرافیایی و لگاریتم مساحت بزرگ هستند. اما گره دوم دارای وزن مقادیر جدید مقیاس شده‌ی طول جغرافیایی، بارندگی متوسط و ضریب رواناب بزرگ و وزن مقادیر تجدید مقیاس شده‌ی عرض جغرافیایی و لگاریتم طبیعی مساحت کوچک است. چگونگی وزن‌دهی ویژگی‌های تجدید مقیاس شده برای هر یک از گره‌های لایه‌ی خروجی در حالت‌های سه تا پنج منطقه‌ای نیز به‌صورت مشابه قابل بررسی و تفسیر است.



شکل ۴- اختصاص ایستگاه‌ها به گره‌های لایه‌ی خروجی نگاشت خودسازمانده. شماره‌ها معرف کد ایستگاه‌ها هستند. (عدد ۱۷ کد حوزه است)
Figure 4. Station assignment to output-layer of SOFM. The numbers represent station codes (The number 17 denotes the basin code.)



شکل ۵- وزن ویژگی‌های تجدید مقیاس شده در گره‌های لایه‌ی خروجی
Figure 5. Weights of rescaled attributes at output-layer nodes

در ادامه مقدار چهار شاخص صحت خوشه معرفی شده برای حالت‌های دو تا پنج خوشه‌ای محاسبه شد که نتایج به‌دست آمده در شکل شش منعکس شده است. بنا بر نتایج مندرج در این شکل، در حالت دو خوشه‌ای مقادیر شاخص‌های عرض نیم‌رخ، دان و کالینسکی- هاراباز حاکی از برتری خوشه‌های حاصل از الگوریتم‌های پیوندمتوسط و SOFM هستند، در حالی که بر اساس شاخص دیویس- بولدین، الگوریتم تک‌پیوند در این حالت بهترین نتایج را حاصل می‌کند. در حالت سه خوشه‌ای، سه شاخص عرض نیم‌رخ، دان و دیویس- بولدین گویای برتری خوشه‌بندی انجام شده به‌وسیله‌ی الگوریتم‌های تک‌پیوند و پیوندمتوسط هستند، در حالی که شاخص کالینسکی- هاراباز، خوشه‌بندی مربوط به SOFM را برتر از سایر گزینه‌ها معرفی می‌کند. برای حالت چهار خوشه‌ای، شاخص‌های عرض نیم‌رخ و کالینسکی- هاراباز، گویای برتری خوشه‌های تشکیل شده به‌وسیله‌ی الگوریتم وارد هستند. در این حالت شاخص دیویس- بولدین، خوشه‌های ایجاد شده به وسیله‌ی الگوریتم پیوند متوسط و شاخص دان خوشه‌های تشکیل شده به‌وسیله‌ی الگوریتم تک‌پیوند را به‌عنوان خوشه‌های بهینه شناسایی می‌کنند. در نهایت، در حالت پنج خوشه‌ای، بر اساس شاخص‌های عرض نیم‌رخ و کالینسکی- هاراباز، الگوریتم وارد و SOFM خوشه‌بندی بهینه را حاصل می‌کنند، در حالی که بر مبنای مقادیر شاخص دیویس- بولدین، بهترین خوشه‌ها به‌وسیله‌ی الگوریتم‌های تمام‌پیوند و وارد ایجاد می‌شوند. در این حالت، شاخص دان، الگوریتم تک‌پیوند را تشکیل‌دهنده‌ی خوشه‌های بهینه معرفی می‌کند.

بنا بر نتایج مربوط به شاخص‌های صحت خوشه، امکان معرفی یکی از الگوریتم‌های مورد استفاده، به عنوان روش بهینه‌ی کلی برای خوشه‌بندی بردارهای مشخصه‌ی ایستگاه‌های مورد مطالعه وجود ندارد. به این معنا که مقادیر به‌دست آمده برای شاخص‌های مختلف صحت خوشه، یکدیگر را به‌طور کامل تأیید نمی‌کنند. هم‌چنین اگر انتخاب یک حالت خوشه‌بندی از نظر تعداد خوشه‌ها و الگوریتم خوشه‌بندی به‌عنوان حالت بهینه مدنظر باشد، باز هم نتایج مربوط به شاخص‌های مختلف، مؤید یکدیگر نیستند.

در حالی که بهترین مقدار شاخص عرض نیم‌رخ مربوط به حالت سه خوشه‌ای الگوریتم‌های تک‌پیوند و پیوندمتوسط است، مقدار بهینه‌ی شاخص دیویس- بولدین متعلق به حالت پنج خوشه‌ای الگوریتم‌های تمام‌پیوند و وارد است، بهترین مقدار شاخص دان به حالت دو خوشه‌ای الگوریتم‌های پیوندمتوسط و SOFM و حالت سه خوشه‌ای الگوریتم‌های تک‌پیوند و پیوند متوسط اختصاص دارد و در نهایت مقدار بهینه‌ی شاخص کالینسکی- هاراباز به حالت پنج خوشه‌ای الگوریتم‌های وارد و SOFM مربوط است.

پس از محاسبه‌ی شاخص‌های صحت خوشه، مقادیر شاخص‌های ناهمگنی H_1، H_2 و H_3 برای تمامی مناطق تشکیل شده در حالت‌های دو تا پنج منطقه‌ای محاسبه شد. بر اساس نتایج به‌دست آمده، چنان چه شاخص H_1 مبنای قضاوت در مورد همگنی و ناهمگنی مناطق تشکیل شده قرار گیرد، مطابق شکل هفت شرط همگنی $H_1<1$ برای تمامی مناطق تشکیل شده برقرار است و تنها منطقه‌ی اول تشکیل شده به‌وسیله‌ی الگوریتم تک‌پیوند در حالت دو منطقه‌ای و مناطق دوم ایجاد شده به‌وسیله‌ی الگوریتم‌های تک‌پیوند و تمام‌پیوند در حالت پنج منطقه‌ای هستند که این شرط را ارضا نمی‌کنند. با توجه به مقادیر شاخص H_2 که در شکل هشت به نمایش درآمده است، در حالت دو منطقه‌ای، هر دو منطقه‌ای تشکیل شده به‌وسیله‌ی الگوریتم تک‌پیوند ناهمگن هستند، اما در مورد تمامی مناطق تشکیل شده به‌وسیله‌ی سایر الگوریتم‌ها شرط $H_2<1$ برقرار است. در حالت سه منطقه‌ای، تنها در مورد الگوریتم وارد، شرط همگنی برای هر

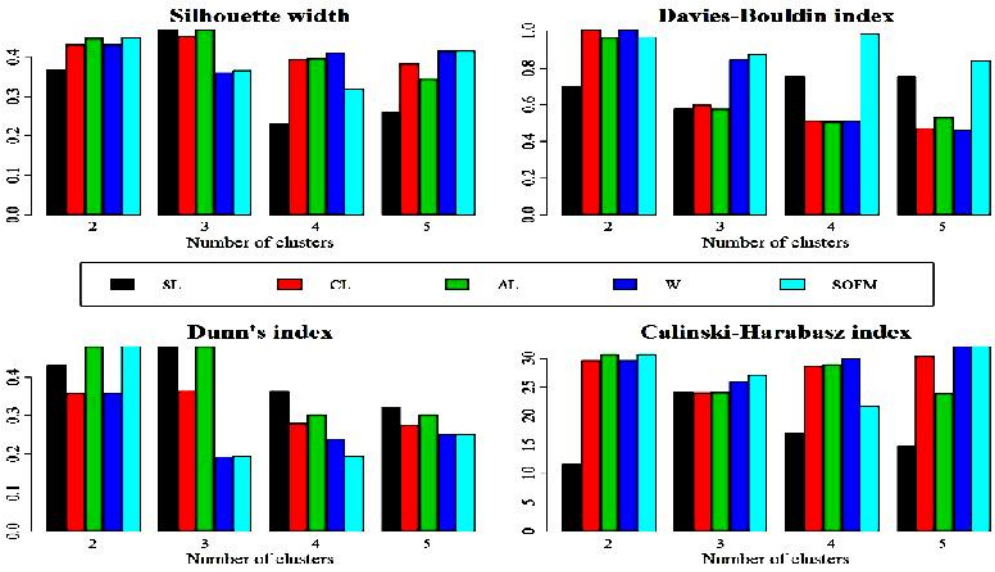
[DOR: 20.1001.1.22516174.1395.7.14.11.3]
[DOI: 10.29252/jwmr.7.14.118]

سه منطقه برقرار است و در مورد سایر الگوریتم‌ها، همواره یک منطقه‌ی ناهمگن تشکیل می‌شود، اگر چه میزان این ناهمگنی شدید نیست. در حالت‌های چهار و پنج منطقه‌ای همواره برای تمامی الگوریتم‌ها، حداقل یک منطقه با ناهمگنی جزئی دیده می‌شود.

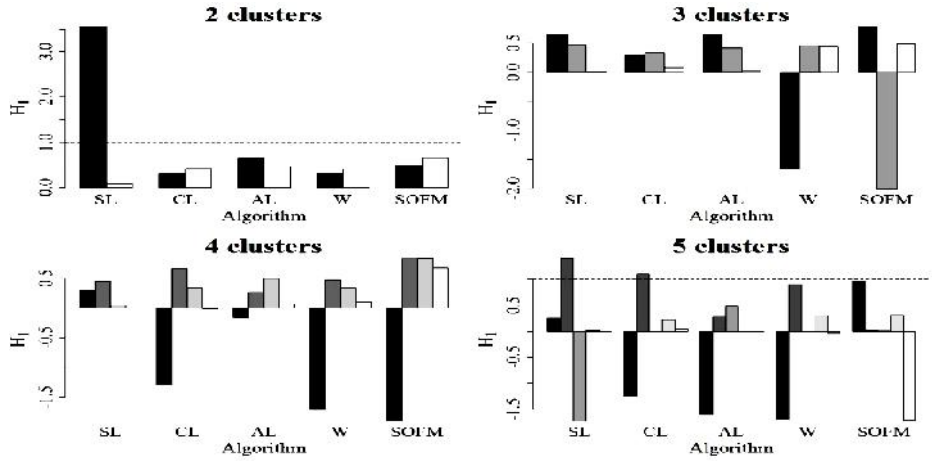
با نگاه به شکل نه که مقادیر شاخص H_3 در آن نشان داده شده است، در حالت دو منطقه‌ای، یکی از مناطق تشکیل شده به‌وسیله‌ی الگوریتم تک‌پیوند ناهمگن است، اما در مورد کلیه‌ی مناطق ایجاد شده توسط سایر الگوریتم‌ها شرط $H_3<1$ برقرار است. در حالت سه منطقه‌ای، شرط همگنی برای تمام مناطق تشکیل شده توسط الگوریتم‌های وارد و SOFM برقرار است، اما در مورد سایر الگوریتم‌ها، همواره یک منطقه‌ی ناهمگن تشکیل می‌شود. در حالت چهار منطقه‌ای، تنها برای SOFM شرط همگنی هر چهار منطقه ارضا می‌شود و در مورد سایر الگوریتم‌ها همواره یک منطقه‌ی ناهمگن مشاهده می‌شود. در حالت پنج منطقه‌ای همواره برای تمامی الگوریتم‌ها، حداقل یک منطقه با ناهمگنی جزئی دیده می‌شود.

با توجه به نتایج مربوط به شاخص‌های ناهمگنی، می‌توان چنین نتیجه گرفت که به‌طور کلی الگوریتم وارد و SOFM در تشکیل مناطق همگن نسبت به سایر الگوریتم‌ها عملکرد بهتری دارند و در مقابل، الگوریتم تک‌پیوند ضعیف‌ترین عملکرد را در زمینه‌ی ایجاد مناطق همگن در بین الگوریتم‌های مورد مطالعه از خود به نمایش می‌گذارد. البته باید توجه داشت که همگنی و ناهمگنی مناطق تشکیل شده، گذشته از الگوریتم خوشه‌بندی تا حد زیادی متأثر از ویژگی‌های مورد استفاده در منطقه‌بندی و عوامل مؤثر بر مکانیزم تولید سیلاب در حوزه‌های آبخیز مختلف است.

در مجموع در میان کلیه‌ی مناطق تشکیل شده به وسیله‌ی الگوریتم‌های مختلف در حالت‌های دو تا پنج منطقه‌ای، مناطق مربوط به حالت دو منطقه‌ای چهار الگوریتم تمام‌پیوند، پیوند متوسط، وارد و SOFM بر اساس هر سه شاخص H، همگن ارزیابی می‌شوند. این وضعیت در مورد هر سه منطقه‌ی حالت سه منطقه‌ای الگوریتم وارد نیز صادق است، بنابراین، حالت‌های مذکور می‌توانند برای ادامه‌ی اجرای فرآیند تحلیل فراوانی منطقه‌ای سیلاب مورد استفاده قرار گیرند. با این حال، از آن‌جا که مناطق تشکیل شده در حالت دو منطقه‌ای به‌طور متوسط نسبت به مناطق حالت سه منطقه‌ای بزرگ‌تر بوده و شامل داده‌های بیش‌تری هستند، لذا استفاده از آن‌ها، امکان برآورد سیلاب با دوره‌های بازگشت طولانی‌تر را فراهم می‌کند و در نتیجه نسبت به استفاده از مناطق حالت سه منطقه‌ای برتری دارد. بررسی بیش‌تر مناطق تشکیل شده در حالت دو منطقه‌ای نشان می‌دهد که در این حالت مناطق تشکیل شده توسط دو الگوریتم تمام‌پیوند و وارد یکسان بوده و مناطق ایجاد شده به‌وسیله‌ی دو الگوریتم پیوندمتوسط و SOFM نیز همسان هستند (فقط شماره‌ی مناطق تشکیل شده توسط دو الگوریتم اخیر متفاوت است). از این رو، مناطق حاصل از به‌کارگیری دو الگوریتم وارد و SOFM به‌عنوان نمایندگان حالت‌های مطلوب برای تکمیل فرآیند تحلیل فراوانی منطقه‌ای سیلاب مورد استفاده قرار گرفتند.

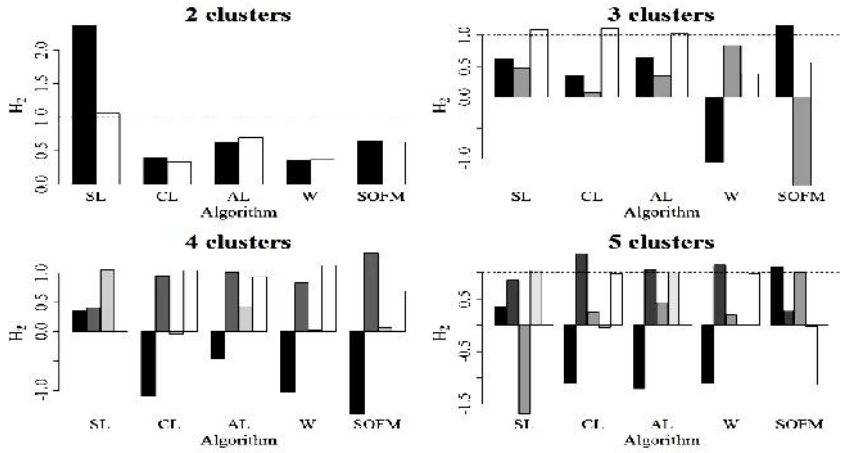


شکل ۶- مقادیر متوسط شاخص‌های صحت خوشه برای تعداد دو، سه، چهار و پنج خوشه
Figure 6. Average values of cluster validity measures for two, three, four and five clusters



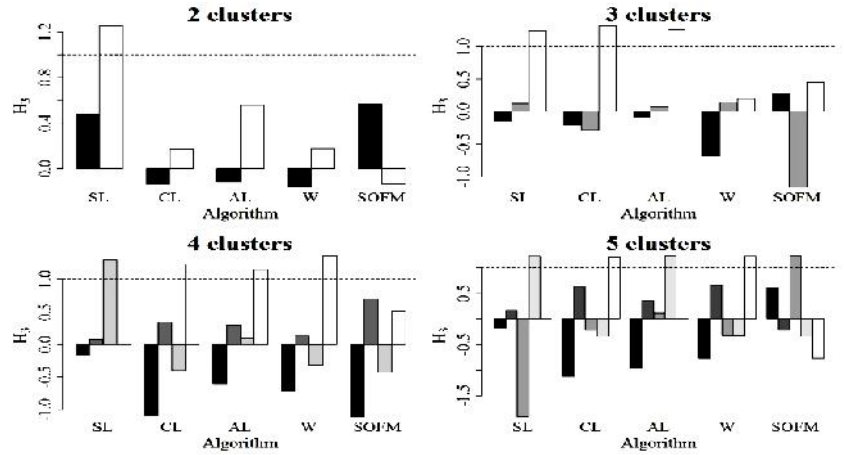
شکل ۷- مقدار شاخص ناهمگنی H_1 برای مناطق تشکیل شده

Figure 7. The values of the heterogeneity measure H_1 for the formed region



شکل ۸- مقدار شاخص ناهمگنی H_2 برای مناطق تشکیل شده

Figure 8. The values of the heterogeneity measure H_2 for the formed region



شکل ۹- مقدار شاخص ناهمگنی H_3 برای مناطق تشکیل شده

Figure 9. The values of the heterogeneity measure H_3 for the formed region

[DOR: 20.1001.1.22516174.1395.7.14.11.3]

[DOI: 10.29252/jwmr.7.14.118]

در ادامه توزیع منطقه‌ای مناسب برای هر منطقه با توجه به شاخص نکویی برازش Z انتخاب شد. برای هر منطقه، توزیعی که نزدیک‌ترین مقدار Z به صفر را اختیار کرد، به‌عنوان توزیع منطقه‌ای انتخاب شد. در جدول دو، توزیع برگزیده برای هر منطقه ارائه شده است. سپس توزیع فراوانی سیلاب برای هر حوزه یا ایستگاه با ضرب کردن میانگین داده‌های حداکثر سیلاب لحظه‌ای آن ایستگاه در توزیع فراوانی مربوط به منطقه‌ای که ایستگاه در آن قرار گرفته است، حاصل شد. با استفاده از توزیع‌های ویژه‌ی ایستگاه‌ها، امکان برآورد سیلاب با دوره‌ی بازگشت مشخص برای هر ایستگاه فراهم شد. به‌علاوه، یک مرتبه نیز برای هر ایستگاه تحلیل فراوانی نقطه‌ای تنها با استفاده از داده‌های سیلاب همان ایستگاه صورت گرفت. با توجه به طول داده‌های آماری موجود، برآورد سیلاب با دوره‌ی بازگشت ده سال به عنوان نمونه برای هر یک از ایستگاه‌ها انجام شد. این فرآیند یک مرتبه با استفاده از مناطق ایجاد شده توسط الگوریتم وارد و توزیع‌های متناظر این مناطق و یک بار هم بر اساس مناطق تشکیل شده به‌وسیله‌ی SOFM و توزیع‌های مربوط به آن‌ها به اجرا درآمد. نتایج تحلیل فراوانی نقطه‌ای برای دوره‌ی بازگشت مشابه نیز برای هر ایستگاه محاسبه شد. مقادیر برآورد شده‌ی سیلاب با دوره‌ی بازگشت ده سال با استفاده از هر یک از این سه تحلیل، در شکل ده نشان داده شده است. از آنجا که مناطق تشکیل شده به‌وسیله‌ی دو الگوریتم وارد و SOFM بسیار شبیه به هم بوده و اختلاف آن‌ها تنها در حد جابه‌جایی یک ایستگاه است، آن‌چنان که در شکل ده مشاهده می‌شود، اختلاف بین برآوردهای دو تحلیل منطقه‌ای اغلب بسیار ناچیز است، به‌طوری که میانگین قدر مطلق اختلاف نسبی بین برآوردهای دو تحلیل منطقه‌ای کمی بیش از ۱٪ است. عمده‌ی این اختلاف نیز مربوط به ایستگاه با کد ۲۰۱-۱۷ است که نقطه‌ی اختلاف دو الگوریتم در منطقه‌بندی است و اختلاف نسبی مربوط به آن نیز در برآورد سیلاب در دو تحلیل منطقه‌ای در حدود ۲۴٪ است.

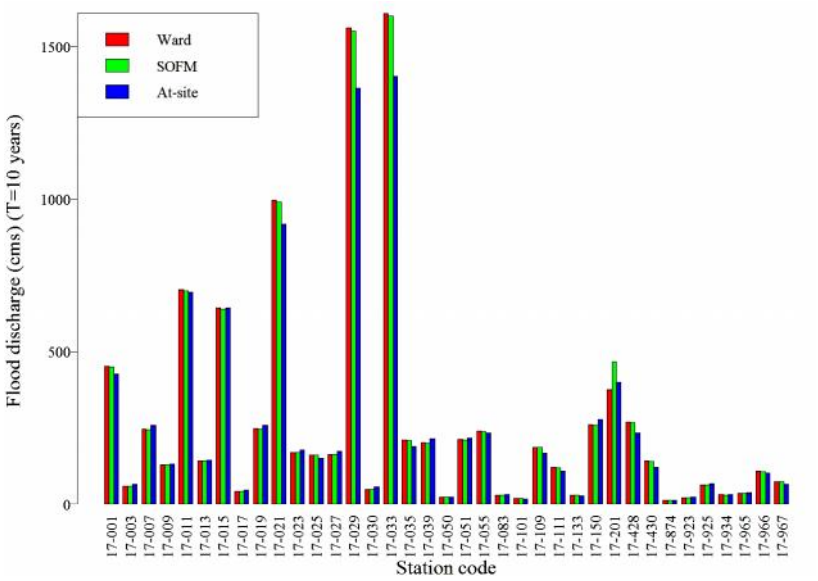
هم‌چنین در شکل ۱۱ میزان اختلاف نسبی برآوردهای سیلاب هر یک از تحلیل‌های منطقه‌ای با تحلیل نقطه‌ای در هر یک از ایستگاه‌ها نشان داده شده است که بنا بر نتایج به‌دست آمده مقدار میانگین قدر مطلق این اختلاف نسبی با برآوردهای نقطه‌ای برای برآوردهای هر یک از تحلیل‌های منطقه‌ای تقریباً برابر ۷٪ است. مطابق نتایج نشان داده شده در شکل ۱۱، در ۱۷ ایستگاه برآوردهای منطقه‌ای مقادیر بزرگ‌تری اختیار می‌کنند و در ۱۸ ایستگاه برآوردهای نقطه‌ای مقادیر بزرگ‌تری دارند. در دو ایستگاه نیز یک برآورد منطقه‌ای بزرگ‌تر و یک برآورد منطقه‌ای کوچک‌تر از برآورد نقطه‌ای است. لذا از این حیث نمی‌توان یک حکم کلی در مورد مقایسه‌ی بزرگی برآوردهای منطقه‌ای و نقطه‌ای صادر کرد. با این حال میانگین اختلاف نسبی بین برآوردهای منطقه‌ای و نقطه‌ای، حاکی از بزرگ‌تر بودن برآوردهای منطقه‌ای به میزان ۸٪ نسبت به برآوردهای نقطه‌ای است. شایان ذکر است که اختلاف بین برآوردهای منطقه‌ای و نقطه‌ای، چه در مقادیر مثبت و چه در مقادیر منفی، در هیچ‌یک از ایستگاه‌های مورد مطالعه از ۱۷٪ تجاوز نکرد.

جدول ۲- توزیع‌های منطقه‌ای برگزیده بر اساس شاخص نکویی برازش Z، برای مناطق تشکیل شده به‌وسیله‌ی SOFM و الگوریتم وارد در حالت دو منطقه‌ای

Table 2. Selected regional distributions based on the goodness-of-fit measure Z for the regions formed by SOFM and Ward's algorithm for 2-region state

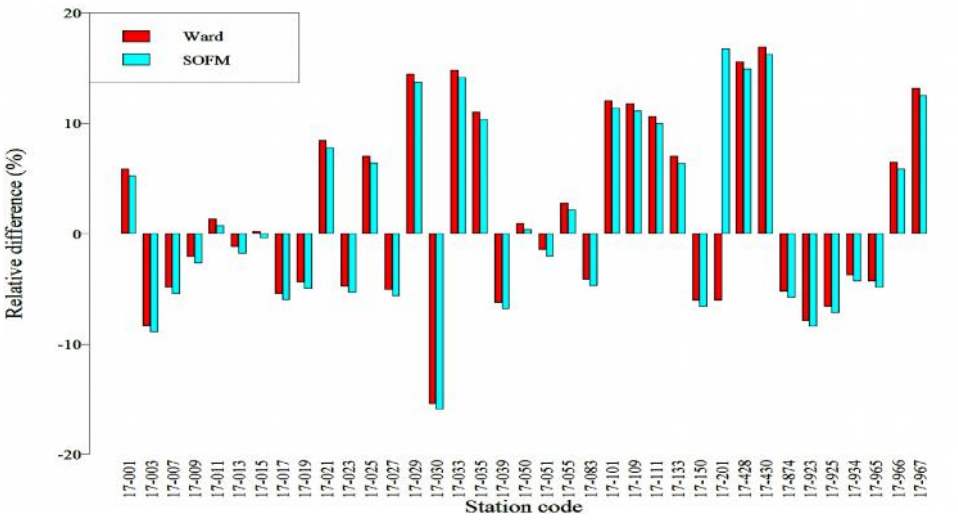
الگوریتم	شماره‌ی منطقه	توزیع منطقه‌ای
SOFM	۱	پیرسون تیپ ۳
	۲	مقادیر حدی تعمیم‌یافته
Ward	۱	نرمال تعمیم‌یافته
	۲	پیرسون تیپ ۳

[DOR: 20.1001.1.22516174.1395.7.14.11.3]
[DOI: 10.29252/jwmr.7.14.118]



شکل ۱۰- برآورد سیلاب با دوره بازگشت ده سال با استفاده از تحلیل فراوانی نقطه‌ای و تحلیل فراوانی منطقه‌ای بر اساس دو منطقه‌ی تشکیل شده به وسیله‌ی الگوریتم وارد و دو منطقه‌ی تشکیل شده به وسیله‌ی SOFM

Figure 10. Flood estimates with ten-year return period by using at-site frequency analysis and regional frequency analysis based on two regions formed by Ward's algorithm and two regions formed by SOFM



شکل ۱۱- مقایسه‌ی برآوردهای سیلاب با دوره بازگشت ده سال برای ایستگاه‌های مورد مطالعه

Figure 11. Comparison of the flood estimates with ten-year return period for the interested stations

در نهایت بنا بر نتایج به دست آمده در این مطالعه، می‌توان چنین نتیجه گرفت که نگاشت‌های خودسازمانده کوهونن در مقایسه با روش‌های معمول خوشه‌بندی توانایی حصول نتایج قابل قبولی را در زمینه‌ی منطقه‌بندی حوزه‌های آبخیز دارا هستند و از این‌رو می‌توانند به عنوان ابزاری کارآمد در فرآیند تحلیل فراوانی منطقه‌ای مورد توجه قرار گیرند. هم‌چنین همان‌طور که مشاهده شد، برآوردهای سیلاب حاصل از تحلیل فراوانی منطقه‌ای مبتنی بر منطقه‌بندی با استفاده از نگاشت‌های خودسازمانده، نتایج نسبتاً نزدیکی را نسبت به برآوردهای سیلاب حاصل از تحلیل فراوانی نقطه‌ای حاصل کردند و لذا به نظر می‌رسد در صورت وجود ایستگاه‌های فاقد آمار هیدرومتری، می‌توان با استفاده از تحلیل فراوانی منطقه‌ای بر اساس این روش منطقه‌بندی، سیلاب با دوره‌های بازگشت مورد نظر را به شکلی نسبتاً قابل اطمینان برای آن ایستگاه‌ها برآورد کرد. افزون بر این موارد، نتایج به‌دست آمده گویای آن است که شاخص‌های صحت

خوشه به تنهایی نمی‌توانند تعیین‌کننده‌ی منطقه‌بندی مطلوب برای اجرای تحلیل فراوانی منطقه‌ای باشند و عامل اصلی در تعیین منطقه‌بندی مطلوب، همگنی مناطق تشکیل شده است که وضعیت آن بر اساس شاخص‌های ناهمگنی مشخص می‌شود. با توجه به موارد ذکر شده، برای مطالعات آینده آزمودن تکنیک‌های خوشه‌بندی از جمله نگاشت‌های خودسازمانده برای منطقه‌بندی حوزه‌های آبخیز با استفاده از ویژگی‌های متعدد و گسترده‌تر هواشناسی، فیزیوگرافیک، زمین‌شناسی، پوشش گیاهی، کاربری اراضی و غیره و بررسی تأثیر انتخاب این ویژگی‌ها توصیه می‌شود. چرا که عدم دسترسی به اطلاعات قابل اطمینان از ویژگی‌های متعدد حوزه‌ها از جمله محدودیت‌های مطالعه‌ی حاضر بود. هم‌چنین استفاده از پیکربندی‌های دیگر شبکه‌ی کوهونن و به‌ویژه حالت دوبعدی و مطالعه‌ی اثر آن بر تغییر مناطق تشکیل شده به ویژه در نواحی مطالعاتی بزرگ‌تر و شامل زیرحوزه‌ها و ایستگاه‌های بیش‌تر می‌تواند محل تحقیق و پژوهش بیش‌تر باشد.

منابع

1. Ahani, A., S. Emamgholizadeh, S.S. Mousavi Nadoushani and K. Azhdari. 2015. Regional Flood Frequency Analysis by Hybrid Cluster Analysis and L-moments. Journal of Watershed Management Research, 6: 11-20.
2. Di Prinzio, M., A. Castellarin and E. Toth. 2011. Data-Driven Catchment Classification: Application to the Pub Problem. Hydrology and Earth System Sciences, 15: 1921-1935.
3. Hall, M.J. and A.W. Minns. 1998. Regional Flood Frequency Analysis Using Artificial Neural Network. In: Babovic, V. and L.C. Larsen (eds.), Proceedings of the Third International Conference on Hydro informatics (Copenhagen, Denmark), 2: 759-763.
4. Hall, M.J. and A.W. Minns. 1999. The Classification of Hydrologically Homogeneous Regions. Hydrological Sciences Journal, 44: 693-704.
5. Hall, M.J., A.W. Minns and A.K.M. Ashrafuzzaman. 2002. The Application of Data Mining Techniques for the Regionalization of Hydrological Variables. Hydrology and Earth System Sciences. 6: 685-694.
6. Haykin, S. 2009. Neural Networks: A Comprehensive Foundation. Pearson Education, Inc. 906 pp.
7. 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.
8. Jingyi, Z. and M.J. Hall. 2004. Regional Flood Frequency Analysis for the Gan-Ming River Basin in China. Journal of Hydrology, 296: 98-117.
9. Kar, A.K., N.K. Goel, A.K. Lohani and G.P. Roy. 2012. Application of Clustering Techniques Using Prioritized Variables in Regional Flood Frequency Analysis - Case Study of Mahanadi Basin. Journal of Hydrologic Engineering- ASCE, 17: 213-223.
10. Kohonen, T. 1982. Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics, 43: 59-69.
11. Kohonen, T. 2001. Self-Organizing Maps. 3rd Extended Edn. Springer Verlag, Berlin, GERMANY, 501 pp.
12. Ley, R., M.C. Casper, H. Hellebrand and R. Merz. 2011. Catchment Classification by Runoff Behavior with Self-Organizing Maps (SOM). Hydrology and Earth System Sciences 15: 2947-2962.
13. Lin, G.F. and L.H. Chen. 2006. Identification of Homogeneous Regions for Regional Frequency Analysis Using the Self-Organizing Map. Journal of Hydrology, 324: 1-9.
14. Rao, A.R. and V.V. Srinivas. 2008. Regionalization of Watersheds-An Approach Based on Cluster Analysis, Series: Water Science and Technology Library, 248 pp.
15. Razavi, T. and P. Coulibaly. 2014. Classification of Ontario Watersheds Based on Physical Attributes and Streamflow Series. Journal of Hydrology, 493: 81-94.
16. Rousseeuw, P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20: 53-65.
17. Srinivas, V.V., S. Tripathi, A.R. Rao and R.S. Govindaraju. 2008. Regional Flood Frequency Analysis by Combining Self-Organizing Feature Map and Fuzzy Clustering. Journal of Hydrology, 348: 148-166.
18. Tabatabaei, M., K. Solaimani, M. Habibnejad Roshan and A. Kavian. 2014. Estimation of Daily Suspended Sediment Concentration using Estimation of Daily Suspended Sediment Concentration by Self-Organizing Map (Case Study: Sierra Hydrometry Station- Karaj Dam Watershed). Journal of Watershed Management Research, 5: 98-116.
19. Toth, E. 2013. Catchment Classification Based on Characterisation of Streamflow and Precipitation Time Series. Hydrology and Earth System Sciences, 17: 1149-1159.
20. Willshaw, D.J. and C. Von Der Malsburg. 1976. How Patterned Neural Connections Can be Set up by Self Organization. Proceedings of Royal Statistical Society London B, 194: 431-445.

Regionalization of Watersheds Using a Type of ANNs to Regional Flood Frequency Analysis

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

1- M.Sc. Student, Shahrood University of Technology,
(Corresponding author: ali.ahani66@yahoo.com)
2 - Associate Professor, Shahrood University of Technology
3- Assistant Professor, Shahid Beheshti University
Received: May 22, 2014 Accepted: October 14, 2015

Abstract

Self-Organizing Feature Maps (SOFM) are a variety of artificial neural networks that their applications in the areas of pattern recognition and data clustering makes them noticeable tools to perform regional flood frequency analysis (RFFA). In this study, ability of Self-Organizing Feature Maps for regionalization of Sefidrood watershed in order to perform regional flood frequency analysis using L-moment algorithm is assessed. Results of this study show that SOFMs may be used as an acceptable method for data clustering and regionalization of watersheds. Evaluation of values of cluster validity measures showed that they can't be a determining factor to identify suitable number of regions for regional flood frequency analysis, but homogeneity of regions is main factor to determine desirable number of regions. According to homogeneity of regions and sizes of formed regions, regionalizations including two regions that formed by Ward's algorithm and SOFM were chosen as optimum choices to regional flood frequency analysis on Sefidrood watershed. Furthermore, based on results of flood estimation by at-site FFA and two RFFA, regional estimates are very close to each other and their average relative difference is equal to 1% nearly. Also relative difference between regional and at-site estimates doesn't exceed 17% in any station and its mean value is about 8%.

Keywords: Clustering, L-moments, Regionalization, SOFM

[DOR: 20.1001.1.22516174.1395.7.14.11.3]

[DOI: 10.29252/jwmr.7.14.118]