



Regionalization of Aras Watershed by SOFM

A. Ahani^{1*} and S. S. Mousavi Nadoushani²

Abstract

Self-Organizing Feature Maps (SOFM) are a category of artificial neural networks with a special ability in pattern recognition and data clustering using attributes of data points. In this study, ability of Kohonen self-organizing feature maps has been assessed in regionalization of Aras watershed in order to perform regional flood frequency analysis. Results showed SOFM can be used as an efficient tool for data clustering and regionalization of watersheds. Furthermore the results showed that the cluster validity measure is equal to optimum value for regionalization of Aras watershed, when the number of regions is equal to 4. Also heterogeneity measures are equal to optimum values, when number of regions is equal to 2 or 4. However, when number of regions is equal to 2, largest regions can be formed.

Keywords: Kohonen SOFM, Cluster Analysis, L-moments, Flood Frequency Analysis.

Received: April 15, 2014

Accepted: October 22, 2014

منطقه‌بندی حوضه‌ی آبریز ارس با استفاده از
نگاشت‌های خود سازماندهعلی آهنی^{۱*} و سید سعید موسوی ندوشانی^۲

چکیده

نگاشت‌های خودسازمانده کوهونن گونه‌ای از شبکه‌های عصبی مصنوعی با توانایی ویژه در تشخیص الگو و خوشه‌بندی داده‌ها با استفاده از ویژگی‌های آن‌ها هستند. در این مطالعه، توانایی نگاشت‌های خودسازمانده کوهونن در منطقه‌بندی حوضه‌ی آبریز ارس به منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب با استفاده از الگوریتم گشتاورهای خطی مورد بررسی قرار گرفته است. نتایج به دست آمده نشان داد که نگاشت‌های خودسازمانده می‌توانند به عنوان ابزاری مفید در زمینه‌ی خوشه‌بندی داده‌ها و منطقه‌بندی حوضه‌های آبریز به کار گرفته شوند. افزون بر این، براساس شاخص صحت خوشه‌بندی، حالت ۴ منطقه‌ای از نظر کیفیت خوشه‌بندی به عنوان حالت بهینه برای منطقه‌بندی حوضه‌ی آبریز ارس شناخته شد. همچنین بر اساس شاخص‌های ناهمگنی این حالت در کنار حالت ۲ منطقه‌ای، برای اجرای تحلیل فراوانی منطقه‌ای سیلاب در حوضه‌ی آبریز ارس مناسب ارزیابی شدند. به علاوه، هنگامی که تعداد مناطق برابر ۲ باشد، تشکیل مناطق بزرگ‌تر امکان‌پذیر خواهد بود.

کلمات کلیدی: نگاشت‌های خودسازمانده کوهونن، تحلیل خوشه‌ای، گشتاورهای خطی، تحلیل فراوانی سیلاب

تاریخ دریافت مقاله: ۲۶ فروردین ۱۳۹۳

تاریخ پذیرش مقاله: ۳۰ آبان ۱۳۹۳

1- M.Sc. of Agricultural Engineering - Water Structures, School of Agricultural Engineering, Shahrood University of Technology, Shahrood, Iran. Email: ali.ahani66@yahoo.com

2- Assistant Professor, Faculty of Water and Environmental Engineering, Shahid Abbaspour Campus, Shahid Beheshti University, Tehran, Iran.

*- Corresponding Author

۱- دانش‌آموخته‌ی کارشناسی ارشد سازه‌های آبی، دانشگاه صنعتی شاهرود

۲- استادیار پردیس فنی و مهندسی شهید عباسپور، دانشگاه شهید بهشتی، تهران

*- نویسنده مسئول

۱- مقدمه

در طول دو دهه‌ی اخیر مدل‌های مبتنی بر شبکه‌های عصبی مصنوعی^۱ به شکل گسترده‌ای توسعه یافته و به منظور شبیه سازی رفتار نورون‌های بیولوژیک در مغز انسان مورد مطالعه قرار گرفته‌اند. غیرخطی بودن و انعطاف پذیری شبکه‌های عصبی مصنوعی، آن‌ها را در برخی از کاربردهای علوم فیزیکی مانند هیدرولوژی مفید می‌سازد (Govindaraju and Rao, 2000; ASCE Task Committee on Artificial Neural Networks in Hydrology, 2000a,b)

نگاشت‌های مشخصه خود سازمانده کوهون^۲، گونه‌ای از شبکه‌های یاد گیرنده‌ی رقابتی هستند (Kohonen, 1982). عملکرد این شبکه‌ها مبتنی بر یاد گیرندگی بدون نظارت است، یعنی هیچ خروجی هدفی برای دسته بندی داده‌های موجود مورد نیاز نیست. این شبکه‌ها می‌کوشند تا به وسیله‌ی نگاشت داده‌های موجود بر یک لایه‌ی مشخصه یا فضای مشخصه‌ی خروجی، ساختاری توپولوژیکی در داده‌های ورودی بیابند.

SOFM با لایه‌ی یک بعدی (1-D) کوهون (که شبکه‌ی کوهون خطی نیز خوانده می‌شود) در مطالعات منطقه‌بندی مورد استفاده قرار گرفته است (Hall and Minns, 1998; Hall and Minns, 1999; Hall et al., 2002).

Hall and Minns (1998; 1999) استفاده از شبکه‌ی کوهون را برای منطقه‌بندی با اعمال آن بر یک نمونه‌ی ۱۰۱ تایی از سایت‌های مجهز به تجهیزات هیدرومتری در جنوب غربی انگلیس و ولز در انگلستان مورد آزمایش قرار دادند. (Hall et al. (2002) نگاشت‌های خودسازمانده یک بعدی^۳ 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 صحت سنجی مناطق تشکیل شده با استفاده از شاخص‌های ناهمگنی، چیزی گزارش نکردند. با این حال، Jingyi and Hall (2004) شاخص‌های ناهمگنی معرفی شده توسط Hosking and Wallis (1997) را برای ارزیابی همگنی مناطق مشخص شده به وسیله‌ی SOFM یک بعدی از ۸۶ ایستگاه هیدرومتری در استان‌های جیانگزی و فوجیان چین مورد استفاده قرار دادند. (Srinivas et al. (2008) ایالت ایندیانا را با استفاده از شبکه‌ی کوهون و ترکیب آن با الگوریتم خوشه‌بندی فازی انجام دادند. آن‌ها از شاخص‌های صحت خوشه‌بندی و شاخص‌های ناهمگنی Hosking and Wallis (1997) برای قضاوت در مورد نتایج

منطقه‌بندی استفاده کردند. (Chavoshi et al. (2012) تشکیل مناطق همگن هیدرولوژیک در نوار ساحلی دریای خزر را با روش‌های مختلف خوشه‌بندی سخت و نرم بررسی کردند. (Razavi and Coulibaly (2013) نیز منطقه‌بندی حوضه‌های آبریز اونتاریو را بر اساس ویژگی‌های فیزیکی و سری‌های جریان آن‌ها انجام دادند. آنان از روش‌های غیرخطی خوشه‌بندی مانند نگاشت‌های خودسازمانده، تحلیل غیرخطی مؤلفه‌های اصلی و تحلیل غیرخطی فشرده‌ی مؤلفه‌های اصلویژگی‌های حوضه و جریان‌های روزانه‌ی ۹۰ زیرحوضه در اونتاریوی کانادا به منظور خوشه‌بندی زیرحوضه‌ها در چهار خوشه استفاده کردند. نتایج نشان داد که روش‌های غیرخطی خوشه‌بندی می‌توانند نتایج مناسبی را برای منطقه‌بندی حوضه‌های آبریز حاصل کنند. همچنین Farsadnia and et al. (2014) تعیین مناطق همگن برای منطقه‌بندی حوضه‌های آبریز به وسیله‌ی نگاشت‌های خودسازمانده دوسطحی را مورد مطالعه قرار دادند. نتایج آزمون‌های ناهمگنی نشان داد که چهار منطقه‌ی به دست آمده به وسیله‌ی نگاشت‌های خودسازمانده دوسطحی و الگوریتم وارد^۴ پس از اصلاح به اندازه‌ی کافی همگن هستند. یافته‌های این مطالعه همچنین بیانگر آن بود که ترکیب نگاشت‌های خودسازمانده و الگوریتم وارد، نسبت به ترکیبات نگاشت‌های خودسازمانده با الگوریتم فازی c-means و الگوریتم K-means نتایج بهتری را حاصل می‌کند.

در این مطالعه، هدف بررسی کارایی نگاشت‌های مشخصه خود سازمانده کوهون برای منطقه‌بندی حوضه‌های آبریز به منظور اجرای تحلیل فراوانی منطقه‌ای سیلاب است. بدین منظور منطقه‌بندی حوضه‌ی آبریز ارس با استفاده از این نگاشت‌ها اجرا شده و مورد بحث قرار گرفته است. همچنین کیفیت خوشه‌بندی اجرا شده با استفاده از نگاشت‌های خودسازمانده بر اساس یک شاخص صحت خوشه‌بندی با کیفیت خوشه‌بندی حاصل از الگوریتم خوشه‌بندی K-means به عنوان یک روش رایج برای عملیات منطقه‌بندی مورد مقایسه قرار گرفته است. در نهایت نیز تحلیل فراوانی منطقه‌ای سیلاب برای حوضه‌ی آبریز ارس با استفاده از الگوریتمی مبتنی بر تئوری گشتاورهای خطی^۵ که توسط ارائه Hosking and Wallis (1997) شده است، به اجرا در آمده و نتایج آن ارائه شده است.

۲- روش تحقیق

۲-۱- معرفی حوضه‌ی آبریز ارس

حوضه‌ی آبریز ارس دومین زیرحوضه‌ی بزرگ حوضه‌ی آبریز دریای

داده شدند و از ادامه‌ی فرآیند منطقه‌بندی و تحلیل فراوانی کنار گذاشته شدند.

۲-۵- خوشه‌بندی با استفاده از نگاشت‌های خودسازمانده

SOFM یکی از پرکاربردترین شبکه‌های عصبی مصنوعی برای تشخیص ساختار توپولوژیک در داده‌های مورد بررسی است. این نوع شبکه در زمینه‌های تشخیص الگو، مدل‌سازی بیولوژیک، فشرده‌سازی داده‌ها، پردازش سیگنال و داده‌کاوی کاربردهایی پیدا کرده است (Kohonen, 1997). SOFM دارای یک لایه‌ی ورودی و یک لایه‌ی خروجی است که هر یک شامل تعدادی گره است. تعداد گره‌ها در لایه‌ی ورودی SOFM برابر ابعاد بردار مشخصه است. لایه‌ی خروجی، که لایه‌ی رقابتی^۷ یا کوهون 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 (Fausett, 1994). برای یک SOFM یک بعدی، (Hall and Minns (1999) مقدار m را برابر حداقل $2C_{Exp}$ انتخاب کردند، در حالی که (Hall et al. (2002) مقدار آن را حداقل برابر $3C_{Exp}$ در نظر گرفتند.

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 (Reed et al., 1999). از این رو تعداد سایت‌های موجود در هر منطقه و سال‌های آماری موجود برای هر یک از آن‌ها عاملی مهم در انتخاب تعداد خوشه‌ها است. بر این اساس در این مطالعه، اختصاص سایت‌های موجود در حوضه‌ی آبریز ارس به ۲ تا ۵ خوشه یا منطقه مورد ارزیابی قرار گرفته است و بدین منظور از شبکه‌های عصبی کوهون با لایه‌ی خروجی شامل ۲ تا ۵ گره استفاده شده است.

یک طرح نمادین برای SOFM در شکل ۱ ارائه شده است. SOFM 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 به تمام گره‌ها لایه‌ی خروجی متصل می‌شود. به همراه هر اتصال، یک وزن اتصال وجود دارد.

خزر و دارای مساحتی برابر ۳۹۵۳۴ کیلومتر مربع (در ایران) است و قسمت‌های شمالی استان‌های آذربایجان غربی، آذربایجان شرقی و اردبیل را در بر می‌گیرد. در این مطالعه تعداد ۲۷ ایستگاه هیدرومتری این حوضه که دارای اطلاعات مورد نیاز جهت اجرای عملیات خوشه‌بندی بودند، انتخاب شدند.

۲-۲- زبان برنامه نویسی R

در این مطالعه، برای اجرای محاسبات و عملیات مربوط به خوشه‌بندی با استفاده از شبکه‌ی کوهون و تحلیل فراوانی منطقه‌ای از محیط زبان برنامه نویسی آماری R، نسخه ۳.۰.۲ و بسته‌های cluster_1.14.4 و lmomRFA_2.5 استفاده شده است.

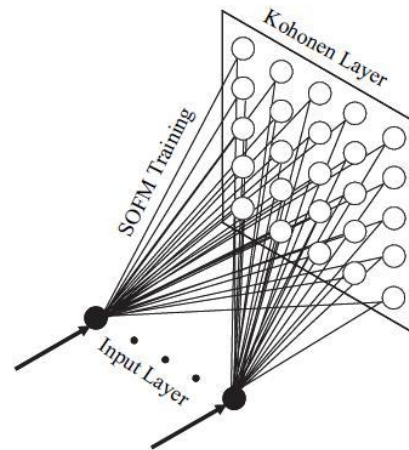
۲-۳- انتخاب ویژگی‌ها و آماده‌سازی بردارهای مشخصه

در این مرحله هدف انتخاب ویژگی‌های مؤثر بر واکنش سیلابی حوضه در منطقه‌ی مورد مطالعه است. از میان ویژگی‌های مؤثر بر واکنش سیلابی سایت‌های مورد نظر در حوضه، با توجه به اطلاعات در دسترس، طول و عرض جغرافیایی، مساحت سطح زهکشی، متوسط بارندگی سالانه، ضریب رواناب و ارتفاع از سطح دریا به عنوان ویژگی‌های جغرافیایی، فیزیوگرافیک و هواشناسی جهت استفاده در عملیات خوشه‌بندی انتخاب شدند.

داده‌های موجود برای هر ویژگی، برای خنثی کردن تفاوت‌ها در واریانس‌ها و بزرگی نسبی آن‌ها، تجدید مقیاس می‌شوند. تجدیدمقیاس ممکن است تبدیل مقادیر ویژگی‌ها با استفاده از تابع تبدیل مناسب و تقسیم مقادیر تبدیل یافته بر انحراف معیار، را به کار بگیرد. هر بردار مشخصه^۸ از ویژگی‌های تجدیدمقیاس شده‌ی (بدون بعد) یک حوضه‌ی آبریز تشکیل می‌شود. در میان ویژگی‌های منتخب در این مطالعه، تنها مساحت سطح زهکشی ایستگاه‌ها با لگاریتم آنها جایگزین شد و سپس استانداردسازی مقادیر مربوط به هر نوع از ویژگی‌ها انجام شد.

۲-۴- تشخیص ایستگاه‌های ناجور

برای تشخیص ایستگاه‌هایی که از نظر آماری با مجموعه‌ی ایستگاه‌های هیدرومتری ناسازگار هستند، بنا بر توصیه‌ی (Hosking and Wallis (1997) از شاخص ناجوری D که بر اساس گشتاورهای خطی تعریف می‌شود، استفاده شد. از میان ۲۷ ایستگاه هیدرومتری برگزیده، تعداد ۲ ایستگاه که بر اساس شاخص ناجوری D از نظر آماری با سایر سایت‌های مورد بررسی ناسازگار تشخیص



شکل ۱- طرح نمادین نگاشت مشخصه‌ی خود سازمانده
کوهون؛ لایه‌ی ورودی شامل ویژگی‌های تجدید مقیاس
شده‌ی حوضه‌ی آبریز است. (Rao and Srinivas, 2008)

y_k نشان‌دهنده‌ی k امین بردار مشخصه در فضای n بعدی با محورهای مختصاتی به نام‌های (y_1, \dots, y_n) در نظر گرفته می‌شود، که $y_k = [y_{1k}, \dots, y_{nk}] \in R^n$ و y_{ik} معرف مقدار ویژگی i در k امین بردار مشخصه‌ی n بعدی y_k است.

ویژگی‌های بردار مشخصه‌ی y_k به صورت رابطه‌ی (۱) تجدید مقیاس می‌شوند

$$x_{ik} = \frac{w_i}{\sigma_i} [f(y_{ik})] \quad (1)$$

for $1 \leq i \leq n, 1 \leq k \leq N$

که در آن x_{ik} نشان‌دهنده‌ی مقدار تجدید مقیاس شده‌ی y_{ik} و w_i وزن مربوط به ویژگی i و مبتنی بر اهمیت نسبی آن است؛ σ_i معرف انحراف معیار ویژگی i است؛ $f(\cdot)$ تابع تبدیل تعیین شده را ارائه می‌کند؛ و N تعداد بردارهای مشخصه‌ی n بعدی را نشان می‌دهد. تجدید مقیاس ویژگی‌ها به دلیل تفاوت‌های موجود در واریانس، اندازه و اهمیت نسبی ضروری است.

به طور کلی، k امین بردار مشخصه‌ی تجدید مقیاس شده به صورت $x_k = [x_{1k}, \dots, x_{nk}]$ است. مجموعه‌ی N بردار مشخصه‌ی تجدید مقیاس شده می‌تواند به صورت ماتریس $N \times n$ داده‌های X نشان داده شود.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nN} \end{bmatrix} \quad (2)$$

بر این اساس، الگوریتم SOFM به صورت زیر ارائه می‌شود:
الف) وزن ابتدایی $\{w_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$ اتصالات از n گره ورودی به m گره خروجی به صورت تصادفی تعیین می‌شود.

این وزن‌های تصادفی عموماً از محدوده‌ی هم‌سانی از مقادیر به صورت مؤلفه‌های بردارهای ورودی انتخاب می‌شوند. $w_j = \{w_{1j}, w_{2j}, \dots, w_{nj}\}$ نشان دهنده‌ی بردار وزن بین گره خروجی j و گره‌های لایه‌ی ورودی در نظر گرفته شده و تکرار عددی به صورت $t = 0$ تنظیم می‌شود.
ب) X با \hat{X} جایگزین می‌شود.

ج) یک بردار ورودی x_k از \hat{X} (تصادفی بدون جایگزینی) ترسیم شده و فاصله‌ی آن از w_j با استفاده از معیار اقلیدسی محاسبه می‌شود. گره‌ی خروجی برنده ω را از طریق رابطه‌ی (۳) مشخص می‌شود:

$$\omega = \arg \min_j \|x_k(t) - w_j(t)\| \quad (3)$$

$j = 1, 2, \dots, m$

د) بردارهای وزن با استفاده از رابطه‌ی (۴) به هنگام می‌شوند:

$$w_j(t+1) = w_j(t) + \eta(t) h_{j,\omega}(t) [x_k(t) - w_j(t)] \quad (4)$$

که در آن $\eta(t)$ پارامتر میزان یادگیرندگی برای تکرار عددی t است، و $h_{j,\omega}(t)$ یک تابع همسایگی نامیده می‌شود. کمیت $\eta(t)$ برای کاهش یکنواخت با افزایش مقدار t به صورت رابطه‌ی (۵) انتخاب می‌شود:

$$\eta(t) = \eta(0) \exp\left(-\frac{t}{\tau_1}\right) \quad (5)$$

که در آن $\eta(0)$ به صورتی انتخاب می‌شود که مقداری نزدیک به ۰٫۱ را اختیار کند، τ_1 یک مقدار ثابت است که نوعاً برابر تعداد بیشینه‌ی تکرار عددی t_{max} تنظیم می‌شود (برای مثال ۱۰۰۰).

در رابطه‌ی (۴) تابع همسایگی $h_{j,\omega}(t)$ که حول گره برنده ω متمرکز می‌شود، توسط رابطه‌ی (۶) ارائه می‌شود:

$$h_{j,\omega}(t) = \exp\left(-\frac{d_{\omega,j}^2}{2\sigma^2(t)}\right) \quad (6)$$

که در آن $d_{\omega,j}$ فاصله‌ی توپولوژیک بین گره برنده‌ی ω و گره j همسایه‌ی آن در لایه‌ی خروجی است،

$$d_{\omega,j} = \|r_\omega - r_j\| \quad (7)$$

که در آن بردار گسسته‌ی r_ω موقعیت گره برنده‌ی ω و بردار گسسته‌ی r_j موقعیت گره j همسایه‌ی آن را تعریف می‌کند، که هر دو آن‌ها در فضای خروجی گسسته اندازه‌گیری می‌شوند.

در رابطه‌ی (۶)، پارامتر $\sigma(t)$ عرض مؤثر همسایگی توپولوژیک $h_{j,\omega}(t)$ در گام زمانی t است.

$$\sigma(t) = \sigma(0) \exp\left(-\frac{t}{\tau_2}\right) \quad (8)$$

که برابر شعاع شبکه در لایه‌ی خروجی SOFM است؛ یک مقدار ثابت است که به وسیله‌ی رابطه‌ی (۹) برآورد می‌شود:

۷-۲- شاخص صحت خوشه‌بندی و تعیین تعداد بهینه‌ی مناطق

خوشه‌های تشکیل شده به صورت 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 به دلیل عملکرد قابل قبول در مطالعات پیشین انتخاب شده است. این شاخص برای اطمینان بردار مشخصه در خوشه با $s(i)$ نشان داده شده و مطابق رابطه (۱۰) تعریف می‌شود:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (10)$$

که در آن $a(i)$ فاصله‌ی متوسط بردار مشخصه‌ی i ام نسبت به تمام بردارهای مشخصه‌ی دیگر موجود در خوشه و $b(i)$ حداقل فاصله‌ی متوسط بردار مشخصه‌ی i ام نسبت به تمام بردارهای مشخصه‌ی خوشه‌ی دیگر است. بر اساس رابطه (۱۰)، خواهیم داشت: $-1 \leq s(i) \leq 1$.

اگر مقدار $s(i)$ نزدیک به ۱ باشد، می‌توان این طور نتیجه‌گیری کرد که بردار مشخصه‌ی i ام در خوشه‌ای مناسب جای گرفته است. از سوی دیگر چنانچه مقدار $s(i)$ به ۱- نزدیک باشد، می‌توان این گونه استنتاج کرد که اطمینان بردار مشخصه به خوشه‌ی مناسبی تعلق نیافته است. از میانگین تمام مقادیر $s(i)$ برای قضاوت کلی در مورد خوشه‌بندی انجام گرفته استفاده می‌شود (Rao and Srinivas, 2008).

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 (Rao and Srinivas, 2008). از این رو 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 نتایج حاصل از اختصاص سایت‌های موجود در حوضه‌ی آبریز سفیدرود بزرگ به ۲ تا ۵ خوشه یا منطقه با استفاده از الگوریتم شبکه عصبی کوهونن مورد ارزیابی قرار گرفته است.

۸-۲- 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

$$\tau_2 = \frac{t_{max}}{\ln \sigma(0)} \quad (9)$$

از روابط (۸) و (۹) برای $t = 0$ ، $\sigma(t) = \sigma(0)$ است در حالی که برای $t = t_{max}$ ، $\sigma(t) = 1$ است. به عبارت دیگر، در $t = 0$ تقریباً تمام گره‌ها در لایه‌ی خروجی متمرکز روی گره برنده به هنگام می‌شوند، در حالی که در پایان تکرارهای عددی، تنها تعداد اندکی گره‌های همسایه حول یک گره برنده به هنگام می‌شوند. افزون بر این، از رابطه‌ی (۶) می‌توان نتیجه گرفت که $h_{j,\omega}(t)$ با افزایش مقدار t ، به صورت نمایی کاهش می‌یابد.

ه) اگر X خالی است، گام (و) آغاز می‌شود، در غیر این صورت باید بازگشت به گام (ج) انجام شود.

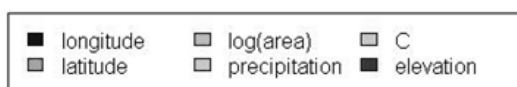
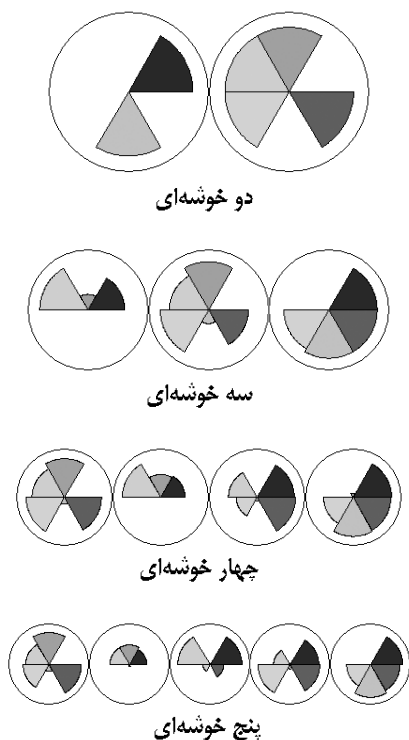
و) اگر $t \geq t_{max}$ گام (ز) اجرا می‌شود، در غیر این صورت t را به $t + 1$ افزایش داده و گام (ب) تا زمانی که هیچ تغییر قابل توجهی در نگاشت‌ی مشخصه مشاهده نشود، ادامه می‌یابد.

ز) به هر بردار ورودی x_k نام گره خروجی برنده‌ی j ($j = 1, \dots, m$) آن با استفاده از رابطه‌ی (۳) اختصاص می‌یابد. تعداد m گره خروجی برنده ($m \leq m$) به عنوان پیش الگو (نمونه) شناخته می‌شوند. (Rao and Srinivas, 2008)

۶-۲- الگوریتم خوشه‌بندی K-means

الگوریتم خوشه‌بندی K-means یک الگوریتم خوشه‌بندی افزاری است که به دلیل کارایی خود در خوشه‌بندی مجموعه‌های بزرگ داده‌ها با ویژگی‌های عددی مشهور است. این الگوریتم به دلیل محاسبات و چارچوب نظری ساده و بازدهی مناسب، کاربرد فراوانی در مطالعات مرتبط با منطقه‌بندی یافته است.

در انجام عملیات تحلیل خوشه‌ای با استفاده از الگوریتم K-means، تابع هدفی تعریف می‌شود که 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. انتخاب اولیه‌ی مراکز خوشه‌ها به صورت بهینه، یکی از مسائل مهم در استفاده از این الگوریتم است. هر یک از این مراکز، یک بردار مشخصه است که اختصاص داده‌ها به خوشه‌های مختلف بر اساس دوری و نزدیکی به آن بردار مشخصه 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 (Rao and Srinivas, 2008).



شکل ۲- وزن ویژگی‌های تجدید مقیاس شده‌ی گره‌های لایه‌های خروجی در حالت‌های ۲ تا ۵ خوشه‌ای

برای نمونه، در حالت دو خوشه‌ای، خوشه‌ی اول به طور متوسط (سمت چپ) در برگرفته‌ی ایستگاه‌هایی است که دارای مقادیر استاندارد شده‌ی مثبت و نسبتاً بزرگ طول جغرافیایی و ضریب رواناب هستند و در مقابل خوشه‌ی دوم (سمت راست) به طور میانگین مشتمل بر ایستگاه‌هایی است که دارای مقادیر استاندارد شده‌ی مثبت و نسبتاً بزرگ عرض جغرافیایی، لگاریتم مساحت سطح زهکشی، متوسط بارندگی سالانه و ارتفاع از سطح دریا هستند. وضعیت متوسط ویژگی‌های استاندارد شده در سایر حالت‌های خوشه‌بندی نیز به همین صورت قابل تفسیر است.

همچنین بر اساس نتایج مندرج در جدول ۱، در میان حالت‌های مورد بررسی، حالت چهار خوشه‌ای که در آن شاخص صحت خوشه‌بندی عرض silhouette نزدیک‌ترین مقدار به ۱ را اختیار می‌کند، از نظر کیفیت خوشه‌بندی می‌تواند به عنوان حالت بهینه در نظر گرفته شود. در شکل ۳ نیز تعداد ایستگاه‌های اختصاص یافته به هر یک از خوشه‌ها، در حالت‌های ۲ تا ۵ خوشه‌ای مربوط به خوشه‌بندی با استفاده از نگاشت‌های خودسازمانده به نمایش در آمده است.

$1 \leq H < 2$ 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 (Hosking and Wallis, 1997). البته پیش از بررسی وضعیت همگنی مناطق شاخص ناچوری D برای شناسایی سایت‌هایی که از نظر آماری با سایر سایت‌های مورد بررسی ناسازگار هستند، محاسبه می‌شود و چنانچه سایتی ناسازگار تشخیص داده شد، از مجموعه‌ی سایت‌های مورد بررسی حذف می‌شود و شاخص‌های H می‌توانند مجدداً محاسبه شوند.

مناطق که بیش‌تر به حالت ناهمگنی نزدیک هستند، به منظور بهبود وضعیت همگنیشان می‌توانند با حذف یک یا چند ایستگاه که تأثیر بیش‌تری در افزایش ناهمگنی دارند یا جا به جایی محدود برخی سایت‌ها بین خوشه‌ها و یا باز تعریف مناطق، اصلاح شوند. (Hosking and Wallis, 1997)

۲-۹- برآورد چندک‌های سیلاب

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، در نهایت برای به دست آوردن برآوردهای چندک 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 نشان دهنده‌ی بهینه بودن انتخاب توزیع مورد نظر باشد (Hosking and Wallis, 1997). در این مطالعه چگونگی برازش توزیع‌های سه پارامتری لجستیک تعمیم یافته^۱، مقادیر حدی تعمیم یافته^۲، نرمال تعمیم یافته^۳، پیرسون تیپ III^۴ و پارتوی تعمیم یافته^۵ با استفاده از شاخص نکویی برازش Z مورد بررسی قرار گرفته است. توزیع منتخب منطقه‌ای معمولاً با اعمال ضریبی مشخص برای هر سایت که می‌تواند دبی متوسط سایت باشد، تبدیل به توزیع ویژه‌ی آن سایت می‌گردد.

۳- نتایج و بحث

در شکل ۲ وزن ویژگی‌های استاندارد شده‌ی گره‌های لایه‌های خروجی برای حالت‌های ۲ تا ۵ خوشه‌ای نشان داده شده است.

جدول ۱- مقادیر میانگین شاخص صحت خوشه‌بندی
عرض silhouette برای حالت‌های ۲ تا ۵ خوشه‌ای

تعداد خوشه‌ها	۲	۳	۴	۵
مقدار میانگین شاخص عرض silhouette	۰,۲۰۰	۰,۲۶۷	۰,۲۸۲	۰,۲۵۲

متوسط خوشه‌بندی اجرا شده در همین حالات با استفاده از الگوریتم K-means است. اما در حالت‌های ۲ و ۵ خوشه‌ای، عکس این وضعیت برقرار است.

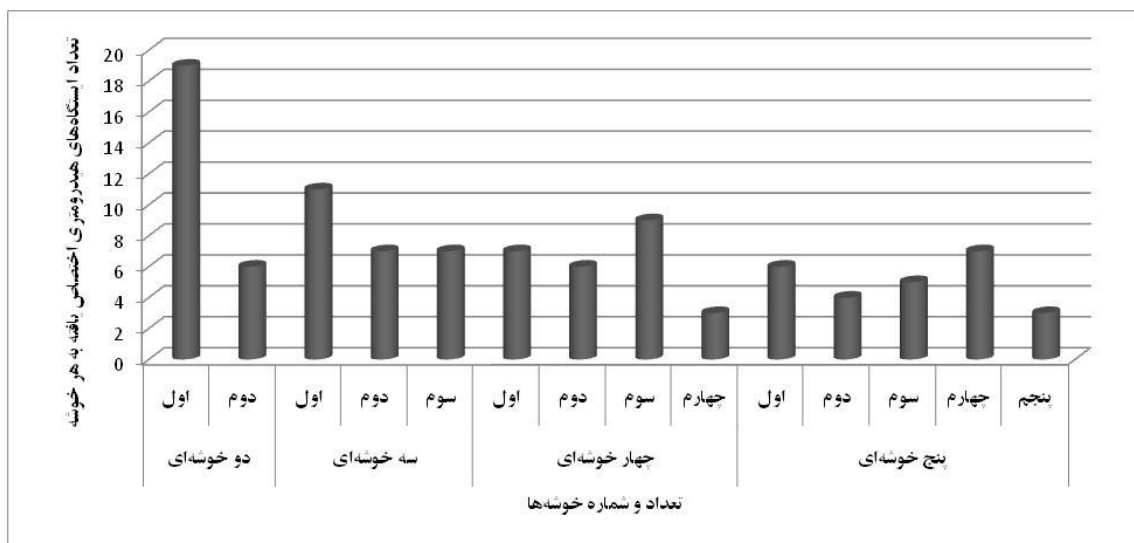
در شکل ۶ مقادیر محاسبه شده‌ی شاخص ناهمگنی H_1 برای مناطق حاصل از منطقه‌بندی با استفاده از نگاشت‌های خودسازمانده کوهونن در حالت‌های ۲ تا ۵ منطقه‌ای نشان داده شده است. مطابق شکل، بر اساس شاخص ناهمگنی H_1 در حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای تمام مناطق حاصل دارای وضعیت همگن هستند، در حالی که در هر یک از حالت‌های ۳ و ۵ منطقه‌ای، یک منطقه‌ی ناهمگن وجود دارد. بر این اساس، دو حالت ۲ منطقه‌ای و ۴ منطقه‌ای می‌توانند به عنوان گزینه‌های بهینه برای ادامه‌ی تحلیل فراوانی منطقه‌ای سیلاب در نظر گرفته شوند. در میان این دو گزینه، امتیاز حالت ۲ منطقه‌ای ایجاد مناطقی است که شامل تعداد ایستگاه‌های بیشتری هستند (به ویژه منطقه‌ی اول) و می‌توانند امکان برآورد سیلاب با دوره‌های بازگشت طولانی‌تری را فراهم کنند. در مقابل، بالاتر بودن کیفیت خوشه‌بندی حالت ۴ منطقه‌ای بر اساس شاخص عرض silhouette می‌تواند به عنوان عامل انتخاب این حالت به عنوان گزینه‌ی بهینه در نظر گرفته شود.

در این مطالعه برای کاهش تأثیر انتخاب مراکز اولیه، ۱۰۰ مرتبه انتخاب مراکز اولیه‌ی خوشه‌ها به صورت تصادفی انجام گرفته و سپس برای هر انتخاب تصادفی، الگوریتم K-means با ۱۰۰۰ مرتبه تکرار اجرا شد. در هر یک از ۱۰۰ مرتبه اجرای کامل الگوریتم، مقادیر شاخص عرض silhouette برای حالت‌های ۲ تا ۵ خوشه‌ای مورد محاسبه قرار گرفت.

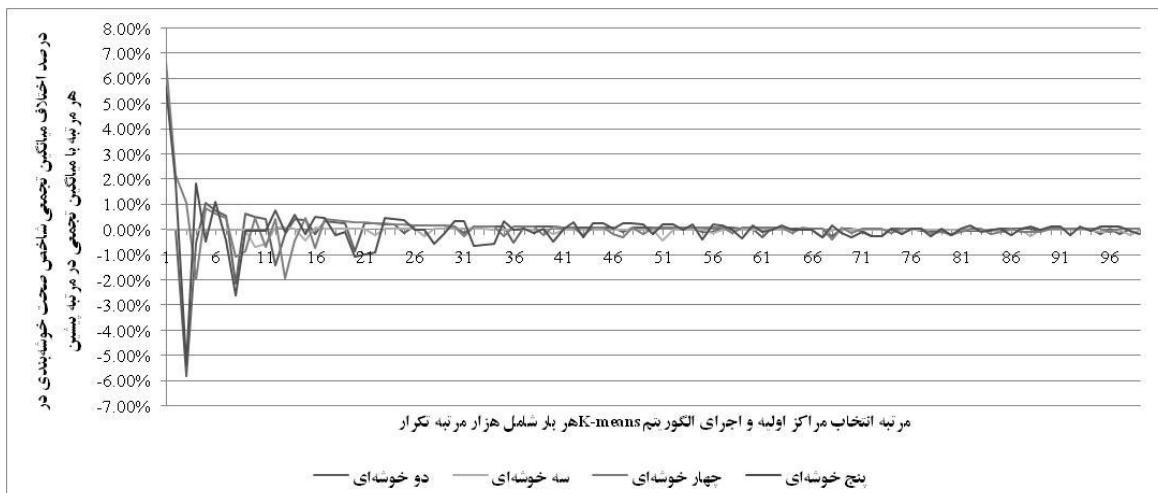
آن گونه که در شکل ۴ دیده می‌شود، میزان تغییرات مقدار میانگین جمعی شاخص صحت خوشه‌بندی در مراتب مختلف اجرای الگوریتم، پس از مرتبه‌ی بیست و دوم به کم‌تر از ۱٪ کاهش یافته است، با این وجود محاسبات تا صدمین مرتبه اجرای کامل الگوریتم ادامه یافته و در نهایت میانگین شاخص صحت خوشه‌بندی برای حالت‌های ۲ تا ۵ خوشه‌ای در ۱۰۰ مرتبه اجرای کامل الگوریتم K-means با انتخاب تصادفی مراکز اولیه محاسبه شده است.

در شکل ۷ تعداد مجموع داده‌های آماری مربوط به هر یک از مناطق در حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای نشان داده شده است که بر این اساس تعداد داده‌های آماری مناطق اول و دوم حالت ۲ منطقه‌ای به ترتیب برابر ۴۰۰ و ۱۶۲ ایستگاه-سال، و تعداد داده‌های آماری مناطق اول، دوم، سوم و چهارم حالت ۴ منطقه‌ای به ترتیب مساوی ۱۸۱، ۱۴۷، ۱۹۹ و ۳۵ ایستگاه-سال است.

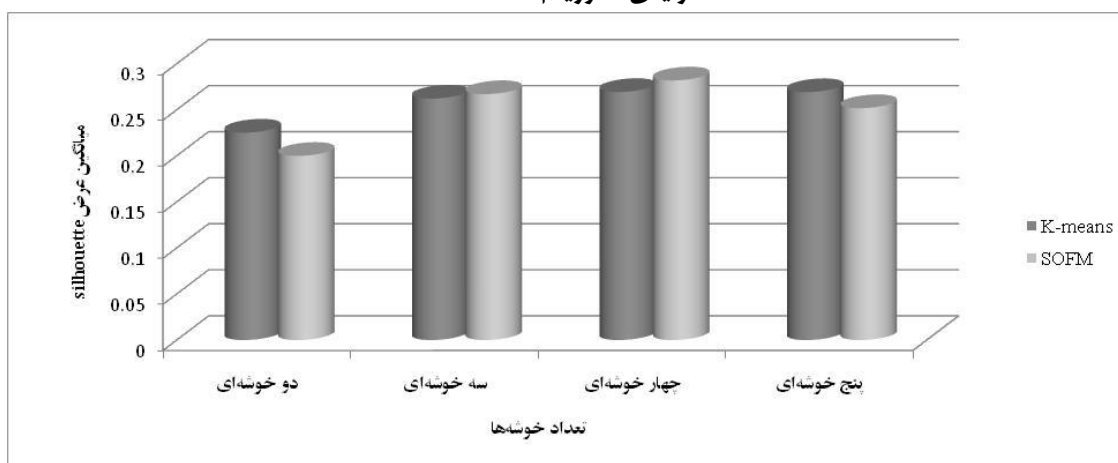
مطابق شکل ۵ که مقادیر میانگین شاخص صحت خوشه‌بندی برای حالت‌های ۲ تا ۵ خوشه‌ای برای این ۱۰۰ حالت را در کنار مقادیر شاخص مذکور برای حالت‌های متناظر خوشه‌بندی با استفاده از نگاشت‌های خودسازمانده نشان می‌دهد، کیفیت خوشه‌بندی حاصل از به کارگیری SOFM در حالت‌های ۳ و ۴ خوشه‌ای بهتر از کیفیت



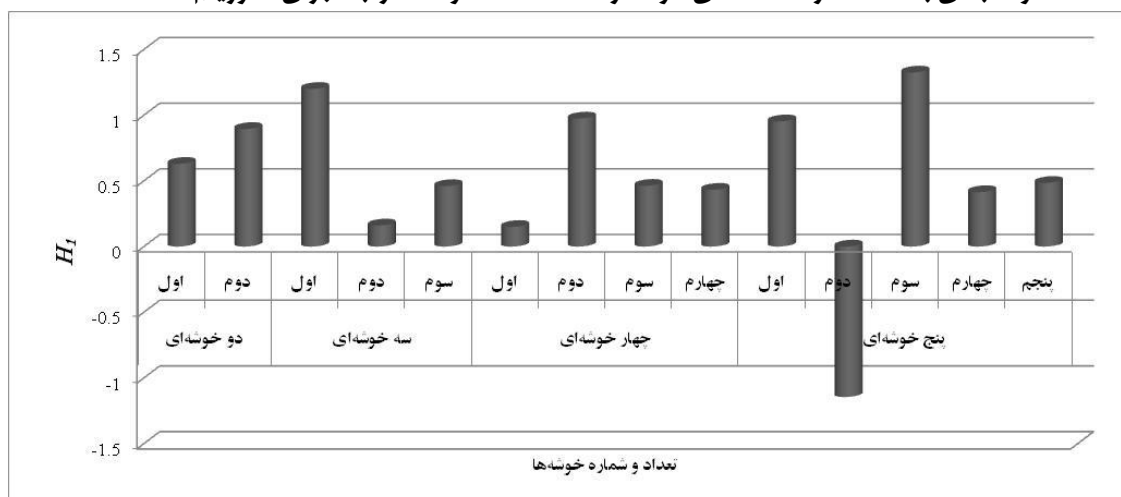
شکل ۳- تعداد ایستگاه‌های هیدرومتری اختصاص یافته به هر یک از خوشه‌ها در حالت ۲ تا ۵ خوشه‌ای



شکل ۴- نمایش تغییر اختلاف بین دو میانگین تجمعی متوالی شاخص میانگین عرض silhouette در ۱۰۰ انتخاب تصادفی مراکز اولیه‌ی الگوریتم K-means



شکل ۵- نمایش مقادیر محاسبه شده‌ی میانگین شاخص عرض silhouette برای حالت‌های ۲ تا ۵ خوشه‌ای حاصل از خوشه‌بندی با استفاده از نگاشت‌های خودسازمانده (SOFM) و صد مرتبۀ اجرای الگوریتم K-means



شکل ۶- مقادیر شاخص ناهمگنی H_1 برای مناطق حاصل از منطقه‌بندی با استفاده از نگاشت‌های خودسازمانده (SOFM) در حالت‌های ۲ تا ۵ منطقه‌ای

عنوان توزیع منطقه‌ای مناطق دوم و چهارم حالت ۴ منطقه‌ای برگزیده شدند.

۴- خلاصه و جمع‌بندی

بنا بر نتایج به دست آمده در این مطالعه، نگاهت‌های مشخصه‌ی خود سازمانده کوهون می‌تواند به عنوان ابزاری کارآمد در زمینه‌ی منطقه‌بندی حوضه‌های آبریز به منظور اجرای تحلیل فراوانی منطقه‌ای مورد توجه قرار گیرند.

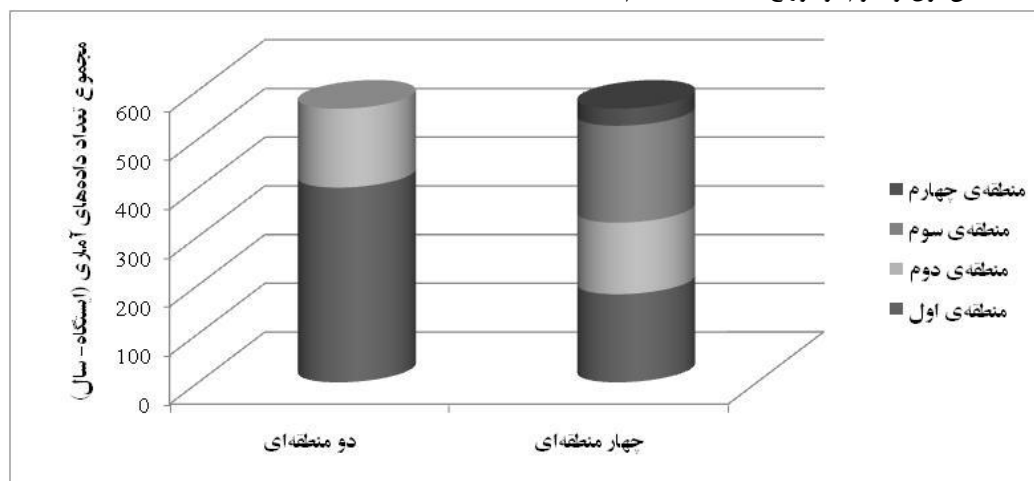
با توجه به مقادیر به دست آمده برای شاخص صحت خوشه‌بندی عرض silhouette به نظر می‌رسد که حالت ۴ منطقه‌ای از نظر کیفیت خوشه‌بندی، حالت بهینه برای اجرای منطقه‌بندی حوضه‌ی آبریز ارس بر اساس ویژگی‌های مورد استفاده در این مطالعه است. همچنین مقایسه‌ی مقادیر محاسبه شده‌ی شاخص عرض silhouette برای خوشه‌بندی اجرا شده با استفاده از نگاهت‌های خودسازمانده با مقادیر محاسبه شده‌ی متوسط همین شاخص برای ۱۰۰ خوشه‌بندی اجرا شده با استفاده از الگوریتم K-means نشان می‌دهد که استفاده از نگاهت‌های خودسازمانده در مواردی می‌تواند نتایجی هم‌تراز یا بعضاً بهتر از سایر روش‌های معمول خوشه‌بندی ارائه کند.

افزون بر آنچه ذکر شد، بر مبنای شاخص ناهمگنی H_1 ، در این مطالعه حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای از نظر همگنی مناطق حاصل، به عنوان حالت‌های بهینه برای اجرای تحلیل فراوانی منطقه‌ای سیلاب ارزیابی شدند. اجرای تحلیل فراوانی منطقه‌ای در حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای و بررسی برآوردهای سیلاب ۱۰ ساله و ۲۵ ساله حاصل از آن‌ها حاکی از نزدیکی قابل توجه مقادیر برآوردها در این دو حالت است.

از آنجا که 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، برای مناطق حالت ۲ منطقه‌ای، برآورد سیلاب به ترتیب با دوره بازگشت تقریبی ۸۰ و ۳۲ سال و برای مناطق حالت ۴ منطقه‌ای، برآورد سیلاب به ترتیب با دوره بازگشت تقریبی ۳۶، ۳۰، ۴۰ و ۷ سال قابل اعتماد خواهد بود.

با در نظر گرفتن این مطلب که در محاسبات و طراحی‌های هیدرولوژیک معمولاً دوره‌های بازگشت با طول‌های مشخصی مانند ۱۰، ۲۵، ۵۰، ۱۰۰، ۵۰۰، ۱۰۰۰ و ۱۰۰۰۰ سال از اهمیت بیشتری برخوردار هستند، در این مطالعه نیز برآورد سیلاب با دوره‌های بازگشت ۱۰ و ۲۵ سال برای ایستگاه‌های مورد مطالعه در دو حالت ۲ منطقه‌ای و ۴ منطقه‌ای انجام گرفت، اگرچه امکان برآورد سیلاب با دوره بازگشت ۵۰ سال برای منطقه‌ی اول حالت ۲ منطقه‌ای وجود دارد و در عین حال برآورد سیلاب حتی با دوره بازگشت ۱۰ سال نیز برای منطقه‌ی چهارم حالت ۴ منطقه‌ای کاملاً قابل اعتماد نیست.

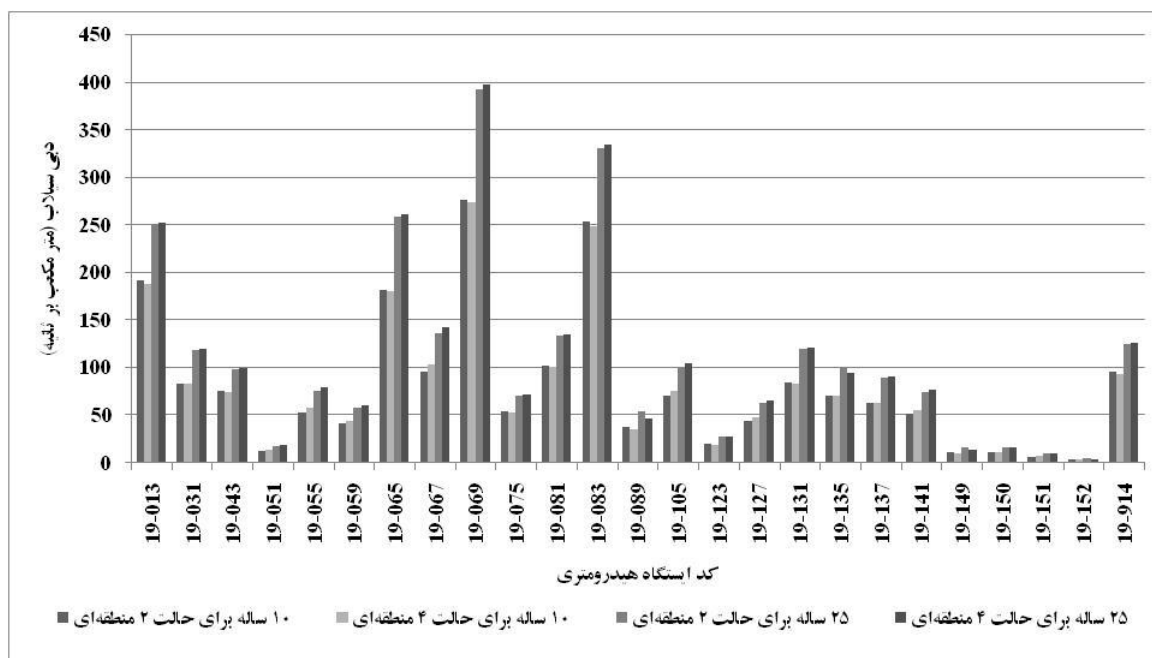
پس از تعیین وضعیت همگنی مناطق، آزمون نکویی برازش منطقه‌ای برای تشخیص و برازش یک توزیع فراوانی سیلاب مناسب بر داده‌های سیلاب سایت‌ها در هر یک از مناطق با استفاده از شاخص نکویی برازش Z اجرا شد. در جدول ۲ مقادیر محاسبه شده‌ی شاخص نکویی برازش Z مربوط به هر یک از توزیع‌های مورد نظر برای هر یک از مناطق حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای ارائه شده است. بر این اساس، توزیع‌های لجستیک تعمیم یافته و پیرسون تیپ ۳ به ترتیب به عنوان توزیع‌های منطقه‌ای مناطق اول و دوم حالت ۲ منطقه‌ای انتخاب شدند. همچنین توزیع نرمال تعمیم یافته به عنوان توزیع منطقه‌ای مناطق اول و سوم، و توزیع لجستیک تعمیم یافته به



شکل ۷- تعداد مجموع داده‌های آماری مربوط به هر یک از مناطق در حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای

جدول ۲- مقادیر شاخص نکویی برازش Z مربوط به توزیع‌های مورد نظر برای مناطق حالت‌های ۲ منطقه‌ای و ۴ منطقه‌ای

توزیع آماری					شماره منطقه	تعداد مناطق
GPA	PE3	GNO	GEV	GLO		
-۳,۲۲	-۳,۰۴	-۱,۷۰	-۰,۹۳	-۰,۱۳	۱	۲
-۰,۷۵	۰,۲۶	۱,۱۴	۱,۶۳	۲,۵۶	۲	
-۱,۴۵	-۰,۸۴	۰,۰۳	۰,۵۳	۱,۲۸	۱	۴
-۲,۵۱	-۲,۶۵	-۱,۸۱	-۱,۳۳	-۰,۹۵	۲	
-۱,۳۹	-۰,۹۴	۰,۱۳	۰,۷۳	۱,۵۱	۳	
-۱,۰۸	-۰,۶۲	-۰,۳۵	-۰,۲۰	۰,۱۵	۴	



شکل ۸- چندک‌های سیلاب با دوره‌های بازگشت ۱۰ و ۲۵ سال برای ایستگاه‌های هیدرومتری در حالت‌های ۲ و ۴ منطقه‌ای

توصیه می‌شود در مطالعات آینده، استفاده از ویژگی‌های مختلف فیزیوگرافی، هواشناسی، زمین‌شناسی، ویژگی‌های خاک حوضه، ویژگی‌های مربوط به پوشش گیاهی و دیگر ویژگی‌هایی که احتمالاً می‌توانند بر فرآیند تولید سیلاب اثر بگذارند، برای اجرای خوشه‌بندی مورد استفاده قرار گرفته و نتایج حاصل مورد بررسی قرار گیرند. همچنین استفاده از نگاشت‌های خودسازمانده با معماری شبکه‌ی و ابعاد متفاوت و مطالعه‌ی نتایج به‌دست آمده می‌تواند تأثیر عامل ساختار این شبکه‌ها را بر عملیات خوشه‌بندی آشکار سازد. استفاده از روش‌های مرکب از نگاشت‌های خودسازمانده و دیگر تکنیک‌های خوشه‌بندی برای یافتن روش‌هایی با کارایی بهتر، از دیگر زمینه‌هایی است که مطالعه‌ی آن مفید به‌نظر می‌رسد.

باین‌حال، حالت ۲ منطقه‌ای به جهت فراهم کردن یک منطقه‌ی بزرگ‌تر و امکان برآورد سیلاب با دوره‌های بازگشت طولانی‌تر برای ۱۹ ایستگاه از میان ۲۵ ایستگاه، می‌تواند نسبت به حالت ۴ منطقه‌ای برای اجرای تحلیل فراوانی منطقه‌ای ترجیح داده شود.

از جمله محدودیت‌های موجود در مسیر این مطالعه می‌توان به وجود نواقص آماری و دشواری دسترسی به برخی ویژگی‌های فیزیوگرافی حوضه‌ی مورد مطالعه اشاره کرد. همچنین باید توجه داشت که نگاشت‌های خودسازمانده همچون دیگر شبکه‌های عصبی مصنوعی به‌صورت یک جعبه‌ی سیاه عمل کرده و روابط کشف شده میان ورودی و خروجی برای اجرای عملیات خوشه‌بندی برای پژوهشگر به‌طور دقیق مشخص نمی‌شود.

- Fausett LV (1994) Fundamentals of neural networks: Architectures, algorithms, and applications. Englewood Cliffs, Prentice Hall, NJ, 461p.
- Govindaraju RS, Rao AR (2000) Artificial neural networks in hydrology. Kluwer Academic Publishers, Holland, 329p.
- Hall MJ, Minns AW (1998) Regional flood frequency analysis using artificial neural network. In: Proc. of the Third International Conference on Hydroinformatics, Vol. 2. Copenhagen, Denmark, 759-763.
- Hall MJ, Minns AW (1999) The classification of hydrologically homogeneous regions. Hydrological Sciences Journal 44(5): 693-704.
- Hall MJ, Minns AW, Ashrafuzzaman AKM (2002) The application of data mining techniques for the regionalization of hydrological variables. Hydrology and Earth System Sciences 6(4): 685-694.
- Hosking JRM, Wallis JR (1997) Regional frequency analysis: An approach based on L-moments. Cambridge University Press, New York, 224p.
- Jingyi Z, Hall MJ (2004) Regional flood frequency analysis for the Gan-Ming river basin in China. Journal of Hydrology 296: 98-117.
- Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biological Cybernetics 43: 59-69.
- Kohonen T (1997) Self-organizing maps, second edition. Springer Verlag, Berlin, 511p.
- Rao AR, Srinivas VV (2008) Regionalization of watersheds - An approach based on cluster analysis, Series: Water Science and Technology Library, Vol. 58, Springer Publishers, 248p.
- Razavi T, Coulibaly P (2013) Classification of Ontario watersheds based on physical attributes and streamflow series. Journal of Hydrology 493: 81-94.
- Reed DW, Jakob D, Robson AJ (1999) Selecting a pooling group. In: Statistical procedures for flood frequency estimation, Flood estimation handbook, Vol. 3, Institute of Hydrology, Wallingford, UK (chapter 6, pp. 28-39).
- Srinivas VV, Tripathi S, Rao AR, Govindaraju RS (2008) Regional flood frequency analysis by combining self organizing feature maps and fuzzy clustering. Journal of Hydrology 348(1-2):148-166.

- 1-Artificial Neural Networks (ANN)
- 2-Kohonen Self-Organizing Feature Maps (SOFM)
- 3-1-D
- 4-Ward algorithm
- 5-L-moments
- 6-Feature vector
- 7-competitive layer
- 8-Generalized Logistic (GLO)
- 9-Generalized Extreme Values (GEV)
- 10-Generalized Normal (GNO)
- 11-Pearson type III (PE3)
- 12-Generalized Pareto (GPA)

۵- تشکر و قدردانی

نسخه خلاصه‌ای از این مقاله در پنجمین کنفرانس مدیریت منابع آب ایران ارائه شده است.

۶- مراجع

- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial neural networks in hydrology, I: Preliminary concepts. Journal of Hydrologic Engineering, ASCE 5(2): 115-123.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial neural networks in hydrology, II: Hydrologic applications. Journal of Hydrologic Engineering, ASCE 5(2): 124-137.
- Chavoshi S, Azmin Sulaiman WN, Saghafian B, Sulaiman MDNB, Latifah AM (2012) Soft and hard clustering methods for delineation of hydrological homogeneous regions in the southern strip of the Caspian Sea Watershed. Journal of Flood Risk Management 5(4):282-294.
- Farsadnid F, Rostami Kamrood M, Moghaddam Nia A, Modarres R, Bray MT, Han D, Sadatinejad J (2014) Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. Journal of Hydrology 509:387-397.