



Evaluating Soft Computing Models for Hydrologically Homogenizing Regional Floods (Case study: Karkheh river basin)

Sahar Safari¹ | Mohammad Sadegh Sadeghian² | Hooman Hajikandi³ |
S. Sajad Mehdizadeh⁴

1. Water Resources Engineering and Management, Department of civil Engineering, Faculty of Civil and Earth Resources Engineering, Central Tehran branch, Islamic Azad University, Tehran, Iran. E-mail: safari252@yahoo.com
2. Corresponding Author, Department of civil engineering, Faculty of civil and earth resources engineering, Central Tehran branch, Islamic Azad university, Tehran, Iran. E-mail: m.sadeghian@iauctb.ac.ir
3. Department of civil engineering, Faculty of civil and earth resources engineering, Central Tehran branch, Islamic Azad university, Tehran, Iran. E-mail: h_hajikandi@iauctb.ac.ir
4. Department of civil engineering, Faculty of civil and earth resources engineering, Central Tehran branch, Islamic Azad university, Tehran, Iran. E-mail: saj.mehdizadeh@iauctb.ac.ir

Article Info

Article type:
Research Article

Article history:

Received 8 April 2023
Received in revised form
23 May 2023
Accepted 22 July 2023
Published online 12 October 2023

Keywords:

Clustering
Flood
Homogeneous areas
Karkheh Basin
Self-organizing map

ABSTRACT

Identification of homologous hydrological groups is one of the fundamental topics in hydrology in both applied and research dimensions. One of the common methods to achieve homogeneous hydrological zones for estimating flood zones is the use of clustering methods. In this study, the using, evaluation and comparison of statistical methods and methods based on artificial intelligence for clustering Karkheh catchment have been investigated. SOM, K-means and hierarchical clustering were used for the study area. In the following, the study of hydrological homogeneity of the obtained areas was evaluated using the linear torque method and the heterogeneity adjustment was evaluated using the methods proposed by Husking and Wallis. The results show that the study area can be converted into two clusters. The values of homogeneity statistics for the first and second clusters were calculated to be 0.33 and 0.17, respectively, which indicates the homogeneity of each region. Due to the short statistical period in some stations, statistical shortcomings and the reliability of the results of regional frequency analysis, this method should be used in estimating floods in other catchments.

Cite this article: Safari, S., Sadeghian, M. S., Hajikandi, H., & Mehdizadeh, S. S. (2023). Evaluating Soft Computing Models for Hydrologically Homogenizing Regional Floods (Case study: Karkheh River Basin). *Journal of Water and Irrigation Management*, 13 (3), 817-835. DOI: <https://doi.org/10.22059/jwim.2023.354624.1054>





ارزیابی مدل‌های محاسبات نرم در همگن‌بندی هیدرولوژیکی سیلاب ناحیه‌ای (مطالعه موردی: حوضه آبریز کرخه)

سحر صفری^۱ | محمدصادق صادقیان^۲ | هومن حاجی‌کندی^۳ | سید سجاد مهدی‌زاده^۴

۱. گروه مهندسی عمران، دانشکده مهندسی عمران و منابع زمین، دانشگاه آزاد اسلامی، واحد تهران مرکزی، تهران، ایران. رایانامه: safari252@yahoo.com
۲. نویسنده مسئول، گروه مهندسی عمران، دانشکده مهندسی عمران و منابع زمین، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران. رایانامه: m.sadeghian@iauctb.ac.ir
۳. گروه مهندسی عمران، دانشکده مهندسی عمران و منابع زمین، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران. رایانامه: h_hajikandi@iauctb.ac.ir
۴. گروه مهندسی عمران، دانشکده مهندسی عمران و منابع زمین، دانشگاه آزاد اسلامی واحد تهران مرکزی، تهران، ایران. رایانامه: saj.mehdizadeh@iauctb.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

شناسایی گروه‌های همگن هیدرولوژیکی یکی از مباحث بنیادی هیدرولوژی در دو بعد کاربردی و تحقیقاتی است. یکی از روش‌های معمول به‌منظور دستیابی به مناطق همگن هیدرولوژیکی برای برآورد منطقه‌ای سیلاب، استفاده از روش‌های خوشه‌بندی است. در این پژوهش استفاده از ارزیابی و مقایسه روش‌های آماری و روش‌های مبتنی بر هوش مصنوعی به‌منظور خوشه‌بندی حوضه آبریز کرخه موردبررسی قرار گرفته است. از خوشه‌بندی SOM، K-means و سلسله‌مراتبی برای منطقه مورد مطالعه استفاده شد. در ادامه نیز بررسی همگنی هیدرولوژیکی مناطق به‌دست‌آمده با استفاده از روش گشتاورهای خطی و تعدیل ناهمگنی با استفاده از روش‌های پیشنهادی توسط هاسکینگ و والیس ارزیابی شد. نتایج نشان داده که منطقه مورد مطالعه قابلیت تبدیل به دو خوشه را دارد. مقادیر آماره همگنی برای خوشه‌های اول و دوم به ترتیب معادل $0/33$ و $0/17$ محاسبه گردید که نشان‌دهنده همگنی هر یک از مناطق می‌باشد. با توجه به کوتاهی دوره آماری در برخی از ایستگاه‌ها، نواقص آماری و قابل‌اعتماد بودن نتایج تحلیل فراوانی منطقه‌ای، در برآورد سیلاب حوضه‌های آبریز دیگر از این روش استفاده شود.

- تاریخ دریافت: ۱۴۰۲/۰۱/۱۹
- تاریخ بازنگری: ۱۴۰۲/۰۳/۰۲
- تاریخ پذیرش: ۱۴۰۲/۰۴/۳۱
- تاریخ انتشار: ۱۴۰۲/۰۷/۲۰

کلیدواژه‌ها:

حوضه آبریز کرخه

خوشه‌بندی

خوشه‌بندی خودسازماندهی

سیلاب

مناطق همگن

استناد: صفری، سحر؛ صادقیان، محمدصادق؛ حاجی‌کندی، هومن؛ مهدی‌زاده، سیدسجاد (۱۴۰۲). ارزیابی مدل‌های محاسبات نرم در همگن‌بندی هیدرولوژیکی سیلاب ناحیه‌ای (مطالعه موردی: حوضه آبریز کرخه). *نشریه مدیریت آب و آبیاری*، ۱۳ (۳)، ۸۱۷-۸۳۵.

DOI: <https://doi.org/10.22059/jwim.2023.354624.1054>



۱- مقدمه

از زمان‌های گذشته تاکنون تمدن‌ها و شهرهای بزرگ در اطراف رودخانه‌ها بنا شده‌اند. از طرفی محدودیت منابع آب، افزایش جمعیت و میل به توسعه در جوامع امروزی ساخت سازه‌های آبی در مسیر رودخانه‌ها را اجتناب‌ناپذیر ساخته است. بنابراین، وقوع سیلاب در این رودخانه‌ها یکی از خطراتی است که جوامع بشری و سازه‌های آبی همواره با آن روبه‌رو بوده است و هر ساله زیان‌های جانی و مالی فراوانی را در نقاط مختلف جهان به بار می‌آورد. اغلب اتفاق می‌افتد که پیش‌بینی سیلاب با دوره برگشت معینی در منطقه‌ای که مشاهدات آماری وجود ندارد، موردنیاز است (Lin and Wang, 2006). چون اکثر طرح‌های هیدرولوژیکی از یک سیل طرح یا بحرانی استفاده می‌نمایند، بنابراین پیش‌بینی مقادیر حداکثر سیلاب لحظه‌ای با روش‌های آماری ضرورت می‌یابد (Drissia et al., 2022).

تحلیل فراوانی سیل رودخانه و سایر کمیت‌های هیدرولوژیکی، اطلاعات مناسبی برای تحلیل‌های مختلف نظیر تعیین معیارهای ریسک و اطمینان‌پذیری در طراحی سازه‌ها به‌دست می‌دهد (Tripathi et al., 2022; Regier et al., 2022). هدف اولیه تحلیل فراوانی، ارتباط دادن بزرگی حوادث حادی به فراوانی وقوع آن‌ها از طریق استفاده از توزیع‌های آماری می‌باشد (Lin and Chen, 2006; Ley et al., 2011; Chavoshi et al., 2012). جهت بسط زمانی سری‌های حداکثر زمانی سیلاب از روش‌های مختلفی تحت عنوان تحلیل منطقه‌ای استفاده می‌گردد. در تحلیل نقطه‌ای سیلاب به دلایل مختلف، از جمله کوتاه‌بودن دوره زمانی آمار و خطای برداشت آمار، اطلاعات موجود نمی‌تواند دقت لازم را در جوابگویی به پیش‌بینی‌های درازمدت داشته باشد (Abedini et al., 2021; Zarei et al., 2022). از طرف دیگر رویدادهای نادر در سری زمانی برداشت‌شده نقطه‌ای، به دلیل محدودیت دوره زمانی، نمونه آماری را طوری تقسیم می‌کند که گویی هر بخش آن از جامعه جداگانه‌ای به‌وجود آمده است. لذا با قاطعیت نمی‌توان گفت که چه توزیعی برای آن‌ها مناسب است. راه‌حل منطقه‌ای بر این مشکل فائق آمده و قادر است نمونه مستقلی از رویدادهای نادر را فراهم سازد. مزایای تحلیل منطقه‌ای تخمین بهتر و قابل‌اطمینان سیل‌های طرح در ایستگاه‌های فاقد آمار مشاهداتی است (Kumar et al., 2015). تحلیل فراوانی منطقه‌ای سیلاب باید در نواحی همگن انجام شود. این نواحی دارای ایستگاه‌های هم رفتار از نظر تولید سیلاب بوده و لذا می‌توان آمار منطقه‌ای ناحیه همگن را به کل نقاط موجود در آن بسط داد و در نتیجه کلیه نقاط فاقد آمار که از نظر طراحی تأسیسات آبی و محاسبه بیلان آبی در کشاورزی نیاز مبرمی به آمار دارند، تحت سطره اطلاعات منطقه‌ای قرار می‌گیرد (Farsadnia and Moghaddamnia, 2014; Farsadnia and Moghaddamnia, 2013; Rostami, 2003).

مشکل اساسی در استفاده از روش‌های آماری کمبود مکانی و زمانی آمارهای سیلاب است. در اکثر مطالعات منطقه‌ای به‌منظور دستیابی به مناطق همگن هیدرولوژیکی از روش‌های تحلیل خوشه‌ای k - میانگین^۱ و خوشه‌بندی سلسله‌مراتبی^۲ استفاده می‌کنند. Lin and Chen (2006) از الگوریتم نگاشت خود سامانده^۳ (SOM) برای خوشه‌بندی ۱۵۴ ایستگاه باران‌سنجی در تایوان به‌منظور تحلیل فراوانی منطقه‌ای استفاده کردند. آن‌ها از ۱۷ خصوصیت حوضه‌های آبخیز برای خوشه‌بندی استفاده کردند و دریافته‌اند که SOM توانایی بیشتری برای شناسایی مناطق همگن دارد. در پژوهش دیگری Chavoshi et al. (2012) به‌منظور مقایسه روش‌های خوشه‌بندی سخت (سلسله‌مراتبی و k - میانگین) با روش‌های نرم (C- میانگین و کوه‌ن) با استفاده از داده‌های ۶۱ حوضه واقع در نوار ساحلی دریای خزر پرداختند. آن‌ها برای آماده‌سازی داده‌های ورودی خوشه‌بندی از تحلیل مؤلفه اصلی استفاده کردند. مقایسه نتایج با استفاده از آزمون همگنی گشتاورهای خطی نشان داد که خوشه‌های به‌دست‌آمده با روش‌های نرم همگن‌تر هستند. علاوه بر روش‌های خوشه‌بندی ذکرشده پژوهش‌گران مختلفی از مدل‌های یادگیری ماشین برای تحلیل منطقه‌ای سیلاب استفاده کردند. Ghaderi et al.

(2022) به بررسی کاربرد روش‌های یادگیری ماشین برای تحلیل منطقه‌ای فراوانی سیلاب پرداختند (Nachappa *et al.*, 2020; Talukdar *et al.*, 2020) از ۱۸ پارامتر فیزیوگرافی، اقلیمی، سنگ‌شناسی و کاربری اراضی برای حوضه‌های بالادست ایستگاه‌های هیدرومتری حوضه‌های آبخیز کرخه و کارون برای تحلیل سیلاب استفاده کردند. نتایج تحلیل منطقه‌ای فراوانی سیلاب نشان داد که مدل بیزی با روش گشتاور خطی بهترین برآورد را در مقایسه با روش‌های دیگر دارد.

SOM تابع چگالی احتمال از داده‌های ورودی تحت یک الگوریتم یادگیری غیرنظارت‌شونده است، که یک روش مؤثر در خوشه‌بندی است که برای خلاصه‌سازی و بصری‌نمودن داده‌ها به کار می‌رود (Ahn *et al.*, 2021). SOM می‌تواند مجموعه داده‌های پیچیده مانند نقشه‌های طغیان سیلاب با ابعاد بسیار بالا را سازمان‌دهی کند (Towfiqul Islam *et al.*, 2021; Gao *et al.*, 2021; Chang *et al.*, 2021). هرچند SOM در خوشه‌بندی حوضه‌های آبخیز به نواحی همگن هیدرولوژیک با موفقیت استفاده شده است، اما تجزیه و تحلیل نقشه خروجی SOM دشوار است. به‌طور کلی، این الگوریتم به‌ندرت خوشه‌های واضحی در خروجی نشان می‌دهد. به‌منظور حل این مشکل پژوهش‌گران راهکارهای متفاوتی ارائه کردند. Srinivas *et al.* (2007) با استفاده از روش ترکیبی خوشه‌بندی SOM و فازی، مناطق همگن هیدرولوژیک در حوضه‌های آبخیز ایندیانا در ایالات متحده آمریکا را موردشناسایی قرار دادند و منطقه پژوهش را به هفت خوشه همگن تقسیم کردند. Inyang *et al.* (2020) یک رویکرد تحلیلی ترکیبی را از طریق روش‌های یادگیری بدون نظارت و نظارت‌شده برای کشف قطعات دانش، خوشه‌بندی و پیش‌بینی سطوح شدت سیل پیشنهاد کردند. آن‌ها یک یادگیری بدون نظارت دومرحله‌ای براساس K-means و نقشه‌های خودسازمان‌دهی (SOM) بر روی مجموعه داده سیل انجام دادند. در ادامه آن‌ها از مدل شبکه عصبی فازی و مدل توسعه داده‌شده آن توسط الگوریتم ژنتیک به پیش‌بینی سیلاب پرداخت. نتایج آن‌ها نشان داد که رویکرد ترکیبی پیشنهادی آن‌ها ابزاری مناسب جهت طبقه‌بندی ناحیه‌های مختلف سیل و پیش‌بینی آن است. به‌طور کلی، تلاش‌های محدودی برای کشف پتانسیل شبکه عصبی از نوع غیرنظارت‌شونده (نگاشت‌های خودسازمان‌ده) و خوشه‌بندی به‌منظور خوشه‌بندی حوضه‌های آبریز به مناطق همگن از لحاظ هیدرولوژیکی انجام شده است و استفاده از این روش در منطقه‌ای کردن حوضه‌های آبریز به‌عنوان روشی نوین در حوضه آبریز کرخه صورت می‌گیرد (Ghosh *et al.*, 2022; Arora *et al.*, 2022). نتایج نشان داد که محاسبات نرم همواره ابزار مناسبی جهت همگن‌سازی مناطق مستعد سیلاب بوده‌اند. لذا استفاده از آن‌ها برای مناطقی مختلفی که همواره در معرض خطر سیلاب هستند ضروری به‌نظر می‌رسد.

با توجه به موقعیت خاص و ویژگی‌های اقلیمی ایران، بروز حوادث حدی از جمله خشک‌سالی و سیلاب به‌ویژه به‌دلیل تأثیر توده‌های پرفشار هوایی در آن بسیار محتمل است (Asghari Sersekanroud and Saeedi Seta, 2023). حوضه کرخه یکی از مناطق پربارش و سیل‌خیز کشور محسوب می‌شود. در این استان به‌دلیل بارندگی‌های با شدت، مدت و مقدار زیاد و همچنین عوامل دیگر، هر از چند گاهی شاهد وقوع سیل و به‌دنبال آن وارد آمدن تلفات و خسارت‌های سنگینی به جان و مال ساکنان آن هستیم. غالباً حوضه‌ها فاقد ایستگاه‌ها اندازه‌گیری هستند یا طول دوره آماری آن‌ها در مقایسه با دوره بازگشت موردنیاز بسیار کوتاه می‌باشد. این امر مانع از استفاده روش‌های تحلیل فراوانی در هر ایستگاه برای تخمین بزرگی سیلاب با دوره‌های بازگشت بالا می‌شود. برای حل این مشکل، پژوهش‌گران در زمینه هیدرولوژی از روش‌های منطقه‌ای برای تفکیک حوضه‌های آبخیز موجود در یک منطقه به گروه‌های همگن استفاده می‌کنند. در یک خوشه (گروه)، از اطلاعات چندین ایستگاه اندازه‌گیری برای تخمین چندک‌های^۴ سیلاب در ایستگاه‌های اندازه‌گیری نشده، استفاده می‌شود. به همین منظور در این پژوهش به بررسی تحلیل منطقه‌ای و ناحیه‌ای کردن یک حوضه آبریز

پرداخته شده است. از ضرورت انجام این پژوهش می‌توان به پیش‌بینی سیلاب و کنترل خسارت‌های ناشی از آن و نیاز به تخمین سیلاب طرح در طراحی سازه‌های آبی اشاره کرد. هم‌چنین با توجه سیل‌خیزی جنوب کشور می‌توان از نتایج این مطالعه در سیستم مدیریت هشدار حوادث غیرمترقبه (احتمال وقوع سیل) استفاده نمود. لذا در این پژوهش از خوشه‌بندی SOM، K-means و سلسله‌مراتبی برای همگن‌بندی منطقه مورد مطالعه استفاده شده است که تاکنون در این محدوده چنین پژوهشی صورت نگرفته است. Vesanto and Alhoniemi (2000) گزارش دادند که ماتریس U نتوانست مرزهای واضحی روی خروجی SOM آموزش‌دیده پیدا کند، اما خوشه‌بندی سلسله‌مراتبی و میانگین k مرز بین خوشه‌ها را به‌طور واضح نشان داد. لذا در این مطالعه خروجی SOM با استفاده از روش خوشه‌بندی K-means و سلسله‌مراتبی تفکیک شد. درنهایت نیز آزمون همگنی خوشه‌ها مورد ارزیابی قرار گرفته و خوشه‌های ناهمگن به آزمون همگن تعدیل شده‌اند.

۲- مواد و روش‌ها

۲-۱- محدوده مطالعاتی

حوضه آبریز رودخانه کرخه در غرب ایران، در مناطق میانی و جنوبی کوه‌های زاگرس قرار دارد. مساحت این حوضه در حدود ۵۰۷۶۸ کیلومتر مربع می‌باشد که در حدود ۳۳۶۷۴ کیلومتر مربع آن را کوهپایه و حدود ۱۷۵۹۴ کیلومتر مربع آن را دشت‌ها تشکیل داده است (شکل ۱). به‌منظور استخراج خشک‌سالی هیدرولوژیک، پس از اخذ آمار دبی روزانه حوضه‌های آبریز سرشاخه‌های کرخه از سازمان مدیریت منابع آب ایران، با توجه به حداقل طول دوره آماری توصیه‌شده توسط Hosking and Wallis (1993)، تعداد ۱۷ ایستگاه با آمار بالای ۲۰ سال انتخاب گردیدند. در جدول (۱) نام و مشخصات ایستگاه‌های مورد استفاده در مطالعه ارائه شده است. خصوصیات حوضه‌های آبریز مورد مطالعه از شرکت مدیریت منابع آب ایران اخذ شد.

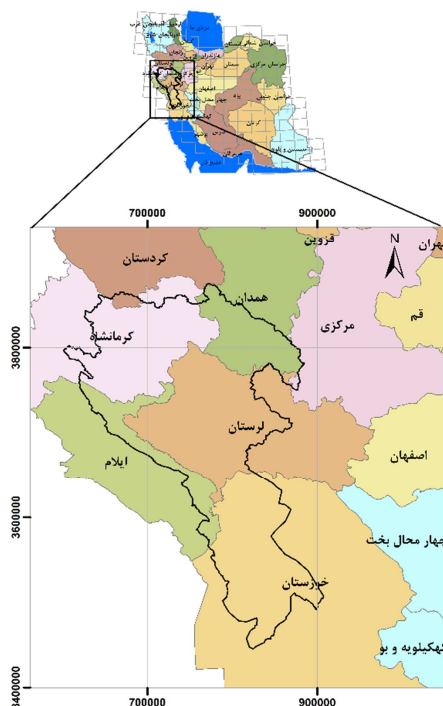


Figure 1. The location of Karkheh basin and selected water measuring stations

Table 1. Specifications of selected water measuring stations

No.	Station	Average height of basin (m)	Latitude (Decimal degrees)	Longitude (Decimal degrees)	Statistical period length (years)	Basin area (km ²)
1	Kaka Reza	2026	33.72	48.25	56	1152
2	Tang Sazin	1739	33.55	46.84	30	2147
3	Cham Anjir	1652	33.45	48.25	56	1636
4	Afarine-Kashkan	1240	33.33	42.89	56	1812
5	Qoorbaghestan	1552	34.23	47.25	57	667
6	Noorabad-BadAvar	2047	34.08	47.97	42	610
7	Halilan-simrere	1808	33.73	47.25	49	1599
8	Polchehr	1871	34.33	47.43	57	1090
9	Aran Gharb	1773	34.42	47.92	57	2041
10	Sarab Seyed	2104	33.79	48.21	57	777
11	Afarineh-Chelhol	1644	33.31	47.89	56	808
12	Poldokhtar-Kashkan	1252	33.16	47.72	56	9253
13	Dartoot	1531	33.74	46.68	42	2588
14	Doab Mark	1542	34.55	46.78	57	1225
15	FiroozAbab-Firoozan	1531	34.35	48.12	57	836
16	Jelogir-Majhin	1125	32.97	47.8	54	3476
17	Doab-Gamasiab	1601	33.48	47.95	41	6359

۲-۲- تعیین نواحی همگن

تحلیل فراوانی منطقه‌ای سیلاب باید در نواحی همگون (همرفتار) صورت گیرد. تعیین نواحی همگون سیلاب منتهی به کاربرد روش‌های تحلیل الگوها^۵ با استفاده از متغیرهای تولیدکننده سیلاب می‌شود. این نواحی دارای ایستگاه‌های همرفتار از نقطه نظر تولید سیلاب بوده و لذا می‌توان آمار منطقه‌ای ناحیه همگون را به کل نقاط موجود در آن تعمیم داده در نتیجه کلیه نقاط فاقد آمار که از نظر طراحی تأسیسات آبی و بیابان‌بندی کشاورزی نیاز مبرمی به آمار دارند تحت سیطره اطلاعات منطقه‌ای قرار می‌گیرند (Zarei *et al.*, 2022).

در روش‌های سنتی، مناطق همگن هیدرولوژیک مجاور و منطبق بر مرزهای جغرافیایی در نظر گرفته شده است، درحالی که حوضه‌های غیرمجاور ممکن است دارای خصوصیات هیدرولوژیک مشابه نباشند. از این رو، برای رفع این مشکل روش‌های دیگری به منظور تدوین مناطق همگن و منطقه‌ای نمودن برآورد سیلاب پیشنهاد گردید (Eslamian *et al.*, 1999). شکل (۲) فلوچارت پژوهش پیش‌رو را بیان می‌کند. در ابتدا اطلاعات موردنظر از محدوده مطالعاتی جمع‌آوری گردید. به منظور انتخاب خصوصیات تأثیرگذار در تحلیل خوشه‌ای و تشکیل مناطق همگن هیدرولوژیک، همبستگی بین میانگین سیلاب سالانه (به عنوان متغیر وابسته) و H_e (تولید آبراهه اصلی (L)، مساحت حوضه (A)، متوسط شیب حوضه (S)، میانگین ارتفاع حوضه (Have)، محیط حوضه (P)، شیب آبراهه اصلی (LS)، بیش‌ترین ارتفاع آبراهه اصلی (Hmax)، کم‌ترین ارتفاع آبراهه اصلی (Hmin)، ارتفاع مرکز ثقل حوضه (Ho)) استفاده گردید. در ادامه و مطابق شکل مذکور از یک تکنیک خوشه‌بندی دو مرحله‌ای برای منطقه‌ای کردن حوضه آبخیز کرخه استفاده شده است. در مرحله اول، از SOM برای شکل‌دهی نقشه ویژگی دوبعدی استفاده شد. سپس گره‌های خروجی از SOM با استفاده از الگوریتم خوشه‌بندی سلسله‌مراتبی، خوشه‌بندی شد. در نهایت خوشه‌بندی از نظر همگنی مورد ارزیابی قرار گرفت. در این گام اگر خوشه‌بندی انجام شده همگن باشد مراحل به پایان می‌رسد و خوشه‌بندی هیدرولوژیک سیلاب تعیین خواهد شد در غیر این صورت مرحله قبل دوباره تکرار خواهد شد. تاکنون برای تفکیک نقشه آموزش دیده SOM به چند خوشه (گروه)، از روش‌های خوشه‌بندی فازی و خوشه‌بندی سلسله‌مراتبی و خوشه‌بندی میانگین k استفاده شده است.

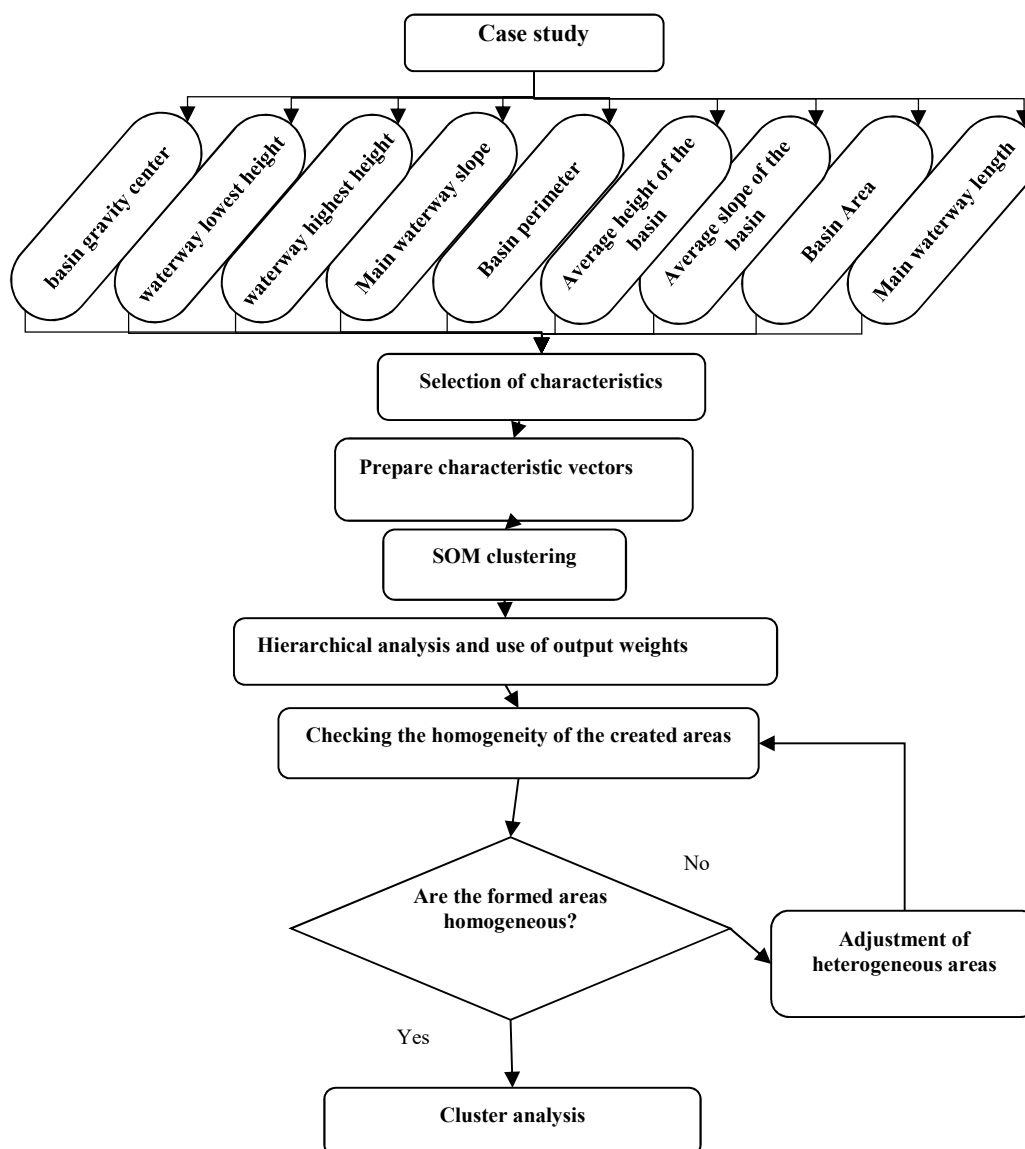


Figure 2. Schematic view of the stages of regional flood frequency analysis

۲-۳- نکاشت ویژگی خود سامانده (SOM)

نگاشت خودسامانده تابع چگالی احتمالاتی از داده‌های ورودی است، که یک روش مؤثر در خوشه‌بندی است. این روش که از نوع یادگیری غیرنظارت‌شونده است، برای خلاصه‌سازی و بصری‌نمودن داده‌ها به کار می‌رود (Kohonen, 2001). این الگوریتم دارای خصوصیات حفاظت از همسایگی و تجزیه و تحلیل فضای ورودی متناسب با توزیع داده‌ها را دارا است (Kohonen, 1982, 2001). SOM شامل دولایه است؛ یک لایه ورودی شکل‌گرفته از مجموعه گره‌ها و یک لایه خروجی (لایه کوهونن) که توسط گره‌هایی که در یک شبکه دو بعدی قرار گرفته‌اند، تشکیل شده است. تعداد نرون‌های خروجی در یافتن انحراف داده‌ها مهم است. اگر اندازه نقشه خیلی کوچک باشد، نمی‌تواند برخی تفاوت‌های مهم را آشکار کند. در مقابل، اگر اندازه نقشه خیلی بزرگ باشد، تفاوت‌ها خیلی کم می‌شود (Wilppu, 1997).

تعداد نرون‌های خروجی در یک SOM را با استفاده از روش ابتکاری پیشنهادشده توسط Vesanto *et al.* (2000) انتخاب کرد.

هر گره در لایه ورودی به هر گره در لایه خروجی توسط شبکه سیناپسی متصل است. هر گره خروجی دارای یک بردار ضرایب متصل به داده‌های ورودی است. بردار ضرایب یک وزن (یا شدت اتصال) را با نام W بین ورودی و لایه خروجی بازمی‌گرداند. وزن‌ها یک شبکه بین واحدهای ورودی (بردار مشخصه‌ها) و واحدهای خروجی وابسته به آن‌ها (گروه‌هایی از بردار مشخصه‌ها) برقرار می‌کند. زمانی که یک بردار مشخصه ورودی X' به SOM ارائه شد، گره‌ها در لایه خروجی با یکدیگر رقابت می‌کنند و گره برنده (گره‌ای که فاصله‌ی تمامی وزن‌هایش از بردار ورودی حداقل است) انتخاب می‌شود. براساس قاعده یادگیری SOM بردار وزن گره برنده و همسایه‌های از پیش تعریف‌شده‌اش در الگوریتم، براساس رابطه (۱) به‌روزرسانی می‌شوند:

$$W_{ij}(t+1) = W_{ij} + \alpha(t) \cdot h_{jc}(t) [X_i(t) - W_{ij}(t)] \quad \text{رابطه (۱)}$$

که در آن، $X_i(t)$ بردار مشخصه در گره i و در تکرار t می‌باشد. $w_{ij}(t)$ وزن بین گره i در لایه ورودی و گره j در لایه خروجی در زمان تکرار t است و $\alpha(t)$ فاکتور سرعت یادگیری است که یک تابع نزولی از زمان تکرار t است و $h_{jc}(t)$ تابع همسایگی (هسته اصلی هموارسازی تعریف‌شده روی نقاط شبکه) است که مقدار همسایگی از گره برنده (c) در طی فرایند یادگیری به‌روزرسانی می‌شود. این فرایند یادگیری تا زمانی که یک معیار توقف معرفی شود (معمولاً زمانی که بردار وزن ثابت شود و یا زمانی که تعداد تکرارها کامل شود) ادامه می‌یابد. ماتریس وزن‌های نهایی بعد از مرحله SOM یک ماتریس $n \times m'$ به نام ماتریس W' است.

$$W' = \begin{bmatrix} W_{11} & \dots & W_{1m'} \\ \vdots & & \vdots \\ W_{n1} & \dots & W_{nm'} \end{bmatrix} \quad \text{رابطه (۲)}$$

۲-۴ - خوشه‌بندی k-means

خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت و فرایند خودکاری است که طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر می‌باشند تقسیم می‌شوند که به این دسته‌ها خوشه گفته می‌شود (Kardan moghadam *et al.*, 2021). هدف از الگوریتم خوشه‌بندی K-means کمینه کردن تابع هدف J است که با توجه به رابطه (۳) تعیین می‌شود (Dehariya, 2010).

$$J = \sum_{j=1}^k \sum_{i=1}^n \|X_{ij} - c_j\|^2 \quad \text{رابطه (۳)}$$

در رابطه یادشده، $\|X_{ij} - c_j\|^2$ فاصله اقلیدسی بین X_{ij} و c_j است و X_{ij} نقطه داده و c_j مرکز خوشه است. مراحل خوشه‌بندی شامل چهار مرحله به شرح ذیل می‌باشد (MacQueen, 1967).

- ۱- k خوشه اولیه به صورت تصادفی انتخاب شده و مراکز خوشه‌ها به طور جداگانه تعیین می‌شوند؛
- ۲- هر نمونه داده به خوشه‌ای که مرکز آن خوشه کم‌ترین فاصله تا آن داده را داراست، نسبت داده می‌شود؛
- ۳- پس از تعلق همه داده‌ها به خوشه‌ها، برای هر خوشه یک نقطه جدید به عنوان مرکز محاسبه می‌شود (میانگین نقاط متعلق به هر خوشه).

۴- مراحل دو و سه تکرار می‌شوند تا زمانی که دیگر تغییری در مرکز خوشه‌ها به وجود نیاید و تابع هدف (رابطه ۳) کمینه شود. برای صحت‌سنجی مقدماتی که شامل حذف داده‌های غیرواقعی، داده‌های پرت، انتخاب تعداد خصوصیت‌ها و

ستون ماتریس‌ها انجام گردید. در ادامه از روش نرمال‌کردن کمینه- بیشینه برای نرمال‌سازی داده‌ها استفاده گردید. مطابق رابطه (۴) این روش بر فاصله میان نقاط کمینه و بیشینه تأکید دارد و مقیاس‌سازی را براساس تفاوت این دو مقدار انجام می‌دهد.

$$X^* = \frac{X_i - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} \quad \text{رابطه ۴}$$

که در آن، X^* متغیر نرمال‌سازی شده، X_i متغیر اصلی است. با توجه به رابطه ذکر شده تمام مقادیر داده‌ها در محدوده صفر تا یک قرار خواهند گرفت. تعداد خوشه‌های بهینه انتخاب شده باید مورد ارزیابی و صحت‌سنجی قرار گیرد، بنابراین از شاخص دیویس- بولدین استفاده گردید.

۲-۴-۱- شاخص دیویس- بولدین

Davies and Bouldin (1979) معیاری براساس میزان پراکندگی درونی و میزان فاصله بین خوشه‌های ایجاد شده، در روش خوشه‌بندی کلاسیک تعریف نمودند (رابطه ۵). شاخص DB برابر با میانگین مقادیر R_j تعریف می‌شود که برابر R_j حداکثر فاصله R_{jk} به دست آمده از مقایسه خوشه j با سایر خوشه‌ها می‌باشد (روابط ۶ و ۷). با توجه به رابطه ذکر شده هرچقدر خوشه‌های ایجاد شده دارای شکلی کروی، متراکم و خوشه‌ها قابل تفکیک از هم باشند، خوشه‌بندی مناسبی ایجاد می‌گردد و در نتیجه تعداد خوشه‌های انتخاب شده نیز مناسب خواهد بود.

$$I_{DB} = \frac{1}{g} \sum_{j=1}^g R_j. \quad \text{رابطه ۵}$$

$$R_j = \text{Max} R_{jk}. \quad k = 1, 2, \dots, g; k \neq j \quad \text{رابطه ۶}$$

$$R_{jk} = \frac{\sigma_j + \sigma_k}{\|\mu_j - \mu_k\|}, \quad j, k = 1, 2, \dots, g; k \neq j \quad \text{رابطه ۷}$$

که در این رابطه، μ_j برابر میانگین فواصل تمامی ایستگاه‌ها واقع شده در خوشه j و σ_j میزان پراکندگی درونی خوشه j می‌باشد و طبق رابطه (۸) محاسبه می‌گردد. مقدار R_{jk} زمانی به حداقل مقدار خود می‌رسد که فاصله بین دو خوشه از یکدیگر زیاد و پراکندگی درونی خوشه‌ها به کم‌ترین مقدار خود برسد.

$$\sigma_j = \sqrt{\frac{1}{n_j} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2} \quad \text{رابطه ۸}$$

در این رابطه، C_j مجموعه داده‌های متعلق به خوشه j و n_j تعداد این داده‌ها می‌باشد.

۲-۴-۲- شاخص میانگین عرض سیلهوت

یکی دیگر از شاخص‌های ارزیابی خوشه‌بندی، شاخص میانگین عرض سیلهوت است. این شاخص برای بردار مشخصه \mathbf{a} در خوشه k به صورت زیر تعیین می‌شود:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad \text{رابطه ۹}$$

که در آن، $a(i)$ میانگین فاصله از بردار مشخصه \mathbf{a} به تمام بردار مشخصه‌های خوشه k ، $b(i)$ نیز حداقل میانگین

فاصله از بردار مشخصه نام به تمام بردار مشخصه‌های دیگر است. مقدار این شاخص بین یک و منفی یک است. در صورتی که مقدار این شاخص به یک نزدیک‌تر باشد نشان از انتخاب بهینه خوشه‌ها می‌باشد.

۲-۵- خوشه‌بندی سلسله‌مراتبی

ماتریس وزن‌های خروجی SOM به‌عنوان ورودی در الگوریتم خوشه‌بندی سلسله‌مراتبی برای مشخص کردن خوشه‌ها روی نقشه آموزش یافته به کار رفت. اساس این روش بر این فرض استوار است که اگر دو خوشه ترکیب شوند، تغییر در مقدار تابع هدف در نتیجه از دست دادن اطلاعات، تنها به رابطه بین دو خوشه ترکیب شده بستگی داشته و به رابطه خوشه‌های دیگر بستگی ندارد. جزئیات الگوریتم وارد را می‌توان در (Roa and Srinivas 2008) جستجو کرد.

۲-۶- استانداردسازی داده‌ها

داده‌های ورودی باید قبل از ورود به الگوریتم‌ها می‌باید استانداردسازی شود، زیرا داده‌های خام شامل تعداد زیادی خصوصیات هستند که هر یک از این متغیرها واحدها و بزرگی متفاوتی نسبت به یکدیگر دارند. اگر هر یک از این متغیرها به‌صورت غیراستاندارد سازی و خام به الگوریتم SOM تغذیه شوند، متغیرهایی که دارای مقادیر بزرگ‌تری هستند نقش مهم‌تری در فرایند به‌روزرسانی وزن‌ها ایفا می‌کنند. بنابراین استانداردسازی کردن بردار ورودی می‌تواند در بهبود مدل کمک کند. اگر X_i ($i=1, 2, \dots, m$) یک متغیر ورودی باشد، مقادیر X تغییر مقیاس یافته می‌تواند از رابطه (۹) به‌دست آید:

$$X' = (X_i - \bar{X}) / \sigma_x \quad \text{رابطه ۹}$$

که در آن، \bar{X} میانگین مقادیر و σ_x انحراف معیار می‌باشند. این تغییر شکل داده‌ها باعث می‌شود تا میانگین مجموعه داده‌ها صفر و انحراف معیار آن‌ها یک شود.

۲-۷- آزمون همگنی ناحیه‌ای

Hosking and Wallis (1993) روش‌هایی را بر پایه ویژگی‌های گشتاورهای خطی برای آزمون‌های همگنی و ناهمگنی مجموعه‌ای از نواحی محتمل که از تحلیل خوشه‌ای منتج شده است، ارائه داده‌اند. در این پژوهش از این روش‌ها استفاده شد. عمده‌ترین معیار ناهمگنی بر پایه گشتاورهای خطی، ضریب تغییرات (CV-L) بنا نهاده شده است (رابطه ۱۰):

$$H = (V - \mu_v) / \sigma_v \quad \text{رابطه ۱۰}$$

۳- نتایج و بحث

۳-۱- تحلیل سیلاب منطقه‌ای و همگنی ناحیه‌ای

تحلیل منطقه‌ای، نیازمند شناسایی مناطق همگن است. لذا در این بخش ابتدا همگنی منطقه مورد مطالعه با استفاده از آزمون‌های گشتاورهای خطی بررسی گردید. سپس به‌منظور یافتن مناطق همگن از لحاظ هیدرولوژیکی، از الگوریتم خوشه‌بندی ترکیبی منطقه استفاده گردید. مطابق جدول (۳) مقادیر ضرایب خطی چولگی و کشیدگی و تغییرات و آماره غیریکنواختی (ناهمگونی) محاسبه شده با روش گشتاورهای خطی برای تمام ایستگاه‌ها از مقدار سه کم‌تر است. با توجه به نتایج جدول مذکور هیچ ایستگاهی ناهمگن با ایستگاه‌های دیگر نیست. بنابراین می‌توان گفت شرط اصلی تحلیل فراوانی منطقه‌ای که همگن بودن منطقه است تأیید می‌گردد.

Table 3. Some statistical parameters and non-uniformity statistics of Hosking and Wallis

No.	Station	Statistical period length (years)	linear variation Coefficient	Linear skewness coefficient	Linear Kurtosis coefficient	Non-uniform statistic
1	Kaka Reza	37	0.3723	0.2535	0.1046	0.7
2	Tang Sazin	33	0.4829	0.5492	0.3721	2.74
3	Cham Anjir	35	0.357	0.2797	0.1941	0.42
4	Afarine-Kashkan	39	0.326	0.2294	0.1257	0.97
5	Qoorbaghestan	29	0.4524	0.4069	0.2973	1
6	Noorabad-BadAvar	31	0.3778	0.2189	0.0941	0.58
7	Halilan-simrere	40	0.3833	0.3385	0.1961	0.48
8	Polchehr	34	0.382	0.321	0.2631	1.14
9	Aran Gharb	31	0.4205	0.3209	0.2534	1.59
10	Sarab Seyed	26	0.382	0.375	0.2187	1.21
11	Afarineh-Chelhol	38	0.4672	0.243	0.0447	1.53
12	Poldokhtar-Kashkan	40	0.3632	0.2452	0.1745	0.77
13	Dartoot	24	0.5613	0.3576	0.1261	2.53
14	Doab Mark	28	0.48	0.3825	0.1942	0.44
15	FiroozAbab-Firoozan	28	0.4766	0.3825	0.1914	0.45
16	Jelogir-Majhin	52	0.4031	0.2706	0.1429	0.16
17	Doab-Gamasiab	33	0.4035	0.3353	0.1791	0.31

به‌منظور بررسی همگنی کل حوضه آبریز کرخه نمودار گشتاورهای خطی^۷ MRDs ایستگاه‌های منطقه پژوهش ترسیم گردید (شکل ۳- الف و ب). همان‌طور که مشاهده می‌شود پراکندگی نقاط ضریب تغییرات خطی و ضریب چولگی خطی حول میانگین در شکل (۳- الف) مناسب نیست، هم‌چنین پراکندگی نقاط ضریب چولگی خط و ضریب کشیدگی خطی در شکل (۳- ب) حول میانگین مناسب نیست. به‌عبارت دیگر، چندین ایستگاه در شکل پراکنش قابل‌توجهی نسبت به میانگین از خود نشان نمی‌دهند. این نکته لزوم استفاده از خوشه‌بندی برای ایجاد همگنی‌کردن مناطق مختلف حوضه را نشان می‌دهد.

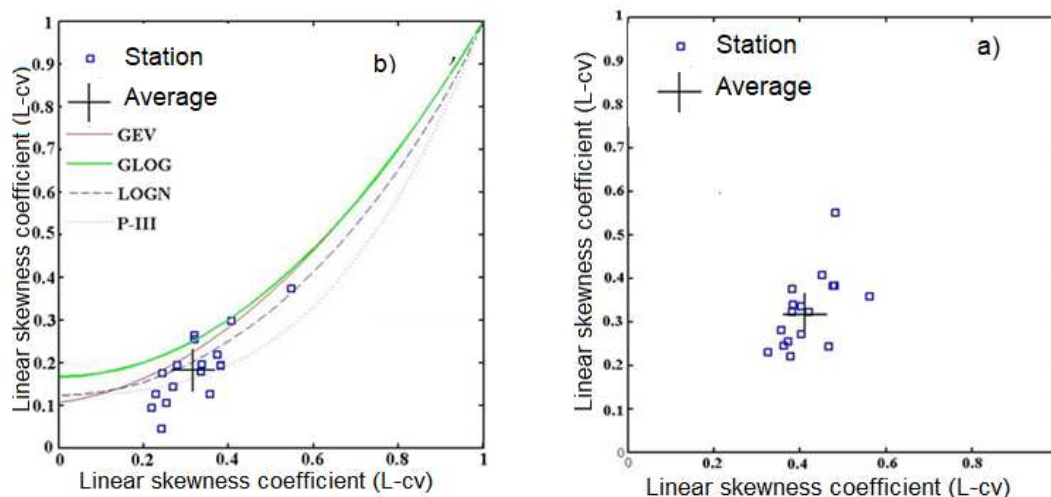


Figure 3. Diagram of linear moments of the studied stations, a) Line skewness coefficient, b) Line skewness coefficient and linear elongation coefficient

با توجه به نتایج مقدار آماره ناهمگنی برای کل حوضه کرخه برابر $H1=1/31$ و $H2=-0/17$ و $H3=-0/77$ به‌دست آمد که نشان‌دهنده ناهماهنگی حوضه آبریز کرخه است که نمی‌توان کل حوضه آبریز کرخه را به‌عنوان یک منطقه همگن هیدرولوژیک در نظر گرفت. در ادامه نتایج از روش خوشه‌بندی برای ایجاد منطقه‌های همگن استفاده شده است.

۳-۲- انتخاب پارامترهای ورودی الگوریتم خوشه‌بندی

به منظور انتخاب خصوصیات تأثیرگذار در تحلیل خوشه‌ای و تشکیل مناطق همگن هیدرولوژیکی، همبستگی بین میانگین سیلاب سالانه (به عنوان متغیر وابسته) و نه خصوصیت فیزیوگرافی (طول آبراهه اصلی (L)، مساحت حوضه (A)، متوسط شیب حوضه (S)، میانگین ارتفاع حوضه (Have)، محیط حوضه (P)، شیب آبراهه اصلی (LS)، بیش‌ترین ارتفاع آبراهه اصلی (Hmax)، کم‌ترین ارتفاع آبراهه اصلی (Hmin)، ارتفاع مرکز ثقل حوضه (Ho)) در ایستگاه جلوگیری‌ماژین به عنوان نمونه در مورد مطالعه به عنوان متغیرهای مستقل محاسبه شد (شکل ۴). نتایج نشان می‌دهد که ضریب همبستگی بین میانگین سیلاب سالانه و متوسط شیب حوضه ناچیز ($R^2=0.07$) می‌باشد، بنابراین این پارامتر از محاسبات حذف شد. مطابق شکل مذکور مساحت حوضه، طول آبراهه اصلی به همراه محیط حوضه به ترتیب با ضریب همبستگی ۰/۹۶، ۰/۹۵ و ۰/۹۳ بیش‌ترین همبستگی را با میانگین سیلاب داشتند. هم‌چنین طول و عرض جغرافیایی به این دلیل در تحلیل‌ها مورد استفاده قرار گرفتند تا تضمین‌کننده نزدیکی جغرافیایی حوضه‌های آبخیز واقع در مناطق شکل‌گرفته پس از تحلیل خوشه‌ای باشند (شباهت بیش‌تر در فواصل کمتر) (Chavoshi et al., 2012). در نهایت طول و عرض جغرافیایی به همراه هشت خصوصیات فیزیکی حوضه ذکر شده به عنوان ورودی الگوریتم خوشه‌بندی استفاده شدند.

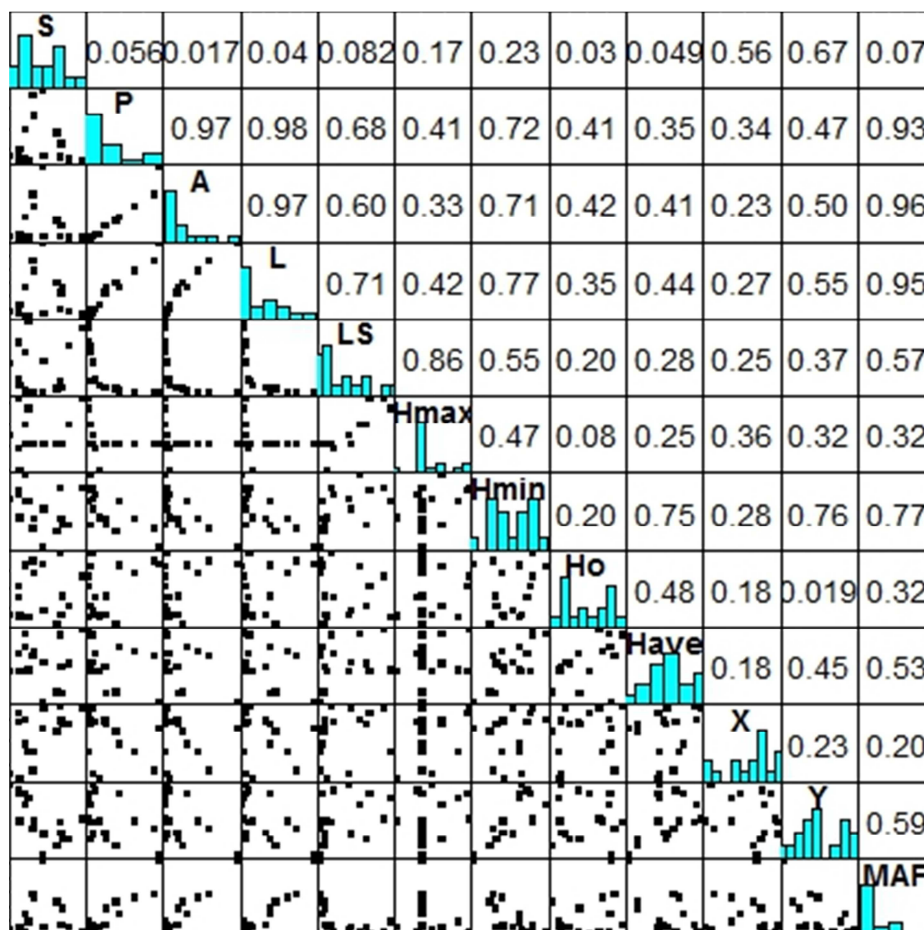


Figure 4. Matrix of correlation coefficients, histogram and dot diagram of the characteristics of the studied watersheds (length of the main waterway (L), area of the basin (A), average slope of the basin (S), average height of the basin (Have), perimeter of the basin (P), slope Main channel (LS), the highest height of the main channel (Hmax), the lowest height of the main channel (Hmin), the height of the center of gravity of the basin (Ho), geographical position X, Y).

۳-۳- خوشه‌بندی نکاشت ویژگی خود سامانده

هشت خصوصیات فیزیوگرافی تأثیرگذار در مرحله قبل به‌همراه سیستم مختصات به‌عنوان ورودی SOM مورد استفاده قرار گرفت. به‌واسطه فرایند یادگیری SOM، بردار وزن متناسب با چگالی احتمال داده‌ها (برای چگالی احتمال داده‌ها) تخمین زده شد. شکل (۵) توزیع خصوصیات تأثیرگذار در سیلاب حوضه را در نقشه آموزش‌یافته SOM نشان می‌دهد. مقدار هر خصوصیت در واحدهای خروجی SOM (هر یک از شش ضلعی‌ها) اهمیت آن‌ها در هر واحد نشان می‌دهد. در شکل (۵) رنگ تیره نشان‌دهنده مقادیر زیاد هر خصوصیت حوضه آبریز و رنگ سفید کم‌ترین مقدار را نشان می‌دهد. به‌عنوان نمونه شیب آبراهه اصلی در قسمت بالای شکل کم‌ترین مقادیر و در قسمت پایین، بیش‌ترین مقادیر را داد. این موضوع قابلیت SOM در خوشه‌بندی و بصری‌نمودن نتایج را نشان می‌دهد. به‌عبارت‌دیگر، الگوریتم SOM مقادیر مشابه را در نقشه خروجی SOM کنار هم قرار می‌دهد.

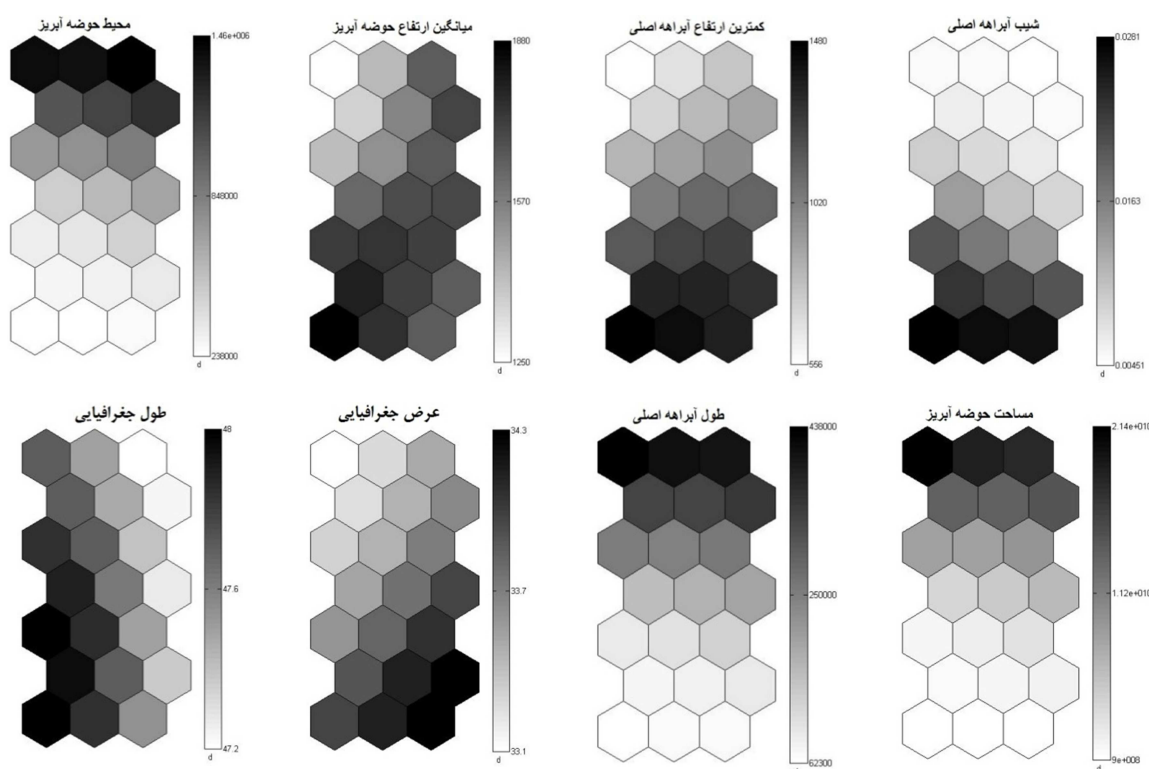


Figure 5. Distribution of physiographic characteristics effective in flood formation in the trained SOM map

بعد از آموزش SOM و به‌دست‌آمدن وزن‌های نهایی، برای خوشه‌بندی واحدها از میانگین K و از بردار وزن‌های خروجی SOM استفاده شد.

۳-۴- نتایج خوشه‌بندی k-mean

برای تعیین تعداد خوشه بهینه شاخص‌های میانگین عرض سیلهوت، دیویس-بولدین به ازای تعداد دو تا هشت خوشه محاسبه شد. تعداد خوشه بهینه در شاخص میانگین عرض سیلهوت زمانی حاصل می‌شود که مقدار این شاخص به یک نزدیک باشد. با توجه به نمودار (۶) تعداد دو خوشه به‌عنوان خوشه بهینه انتخاب شد. بیش‌ترین مقدار شاخص سیلهوت

برابر با $0/39$ بود که متناسب با تعداد دو خوشه است. در شاخص دان تعداد خوشه بهینه برای بیشترین مقدار حاصل می‌شود. بیشترین مقدار $1/56$ ، که متناسب با تعداد دو خوشه است.

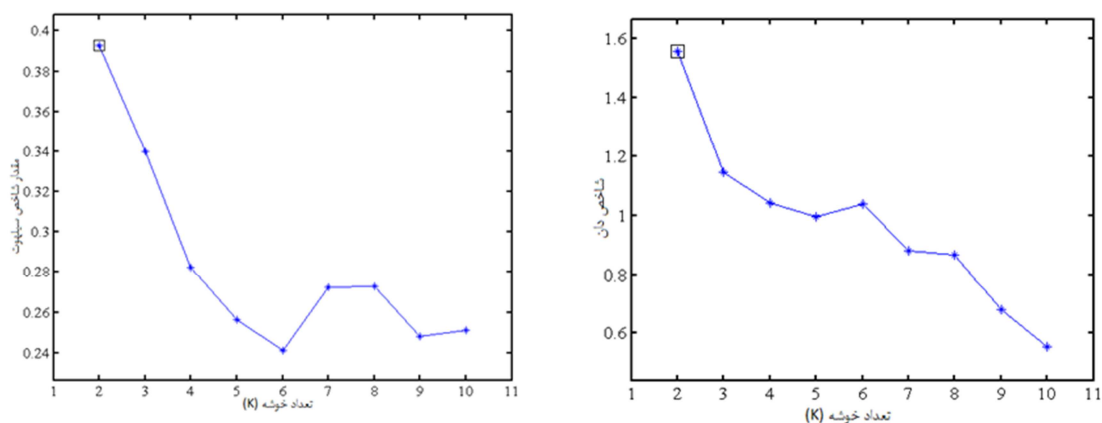


Figure 6. Determining the number of optimal clusters

نتایج دو شاخص خوشه‌بندی میانگین k در شکل (۷) اعداد درون شش‌ضلعی‌ها نشان‌دهنده تعداد ایستگاه‌های اختصاص داده‌شده به واحدهای SOM بین صفر تا دو است.

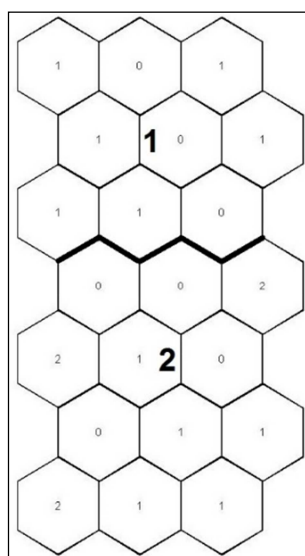


Figure 7. Clustering of 17 stations on the trained SOM map using K-means

۳-۵- نتایج خوشه‌بندی سلسله‌مراتبی

بر پایه نتایج دندوگرام (شکل ۸)، ایستگاه‌ها بر روی نقشه SOM به دو گروه تقسیم شدند. می‌توان دریافت که در خوشه دوم محیط و مساحت حوضه آبریز و همچنین طول آبراهه اصلی کمترین مقادیر را دارا بوده و حوضه‌های این خوشه دارای آبراهه‌ها با شیب زیاد می‌باشند. در ضمن خصوصیت کمترین آبراهه اصلی نیز از خصوصیات فیزیوگرافی تأثیرگذار در ایجاد این خوشه می‌باشند (شکل‌های ۸ و ۹).

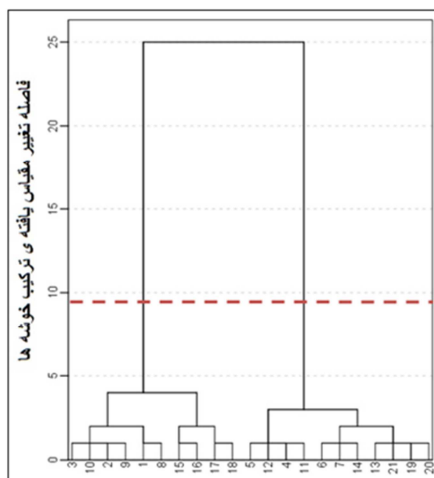


Figure 8. Dendrogram resulting from Kohonen layer clustering using Ward hierarchical clustering

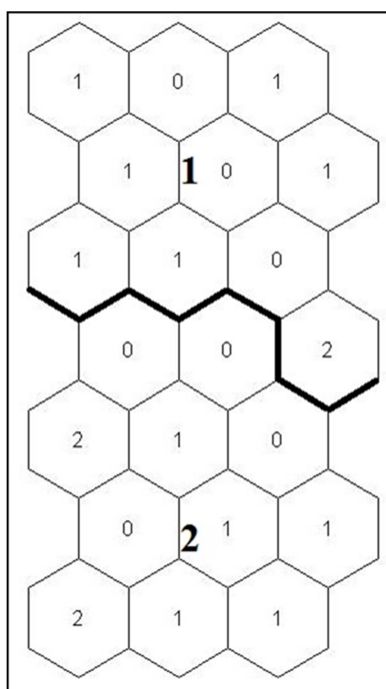


Figure 9. Clustering of the SOM output map by Ward's hierarchical clustering algorithm (the bold numbers indicate the cluster number and the numbers inside each of the hexagons indicate the number of stations assigned to that unit.)

جدول (۴) نشان می‌دهد که در هر دو روش خوشه‌بندی به‌دلیل عدم وجود همه متغیرهای تأثیرگذار در تحلیل فراوانی منطقه‌ای سیلاب، برای تحلیل فراوانی منطقه‌ای سیلاب همگن نشده است و عدم همگنی با توجه به آزمون آماره‌ها مشهود است (مقدار $1/11$ و $1/77$ به ترتیب در خوشه‌بندی SOM+Kmeans و SOM+ Hierarchical. این امر نیازمند تعدیل برای بهبود همگنی دارد. روش‌های مختلفی برای تعدیل همگنی توسط (1997) برای اصلاح مناطق ایجاد، پیشنهاد شده است که می‌توان به ۱- حذف یک یا چند ایستگاه از مجموعه داده‌ها؛ ۲- انتقال یک یا چند ایستگاه از یک منطقه به منطقه دیگر؛ ۳- تفکیک منطقه به دو یا چند خوشه جدید؛ ۴- اجازه‌دادن به یک ایستگاه برای مشارکت در دو

یا چند منطقه؛ ۵- انحلال مناطق با انتقال ایستگاه‌هایش به دیگر مناطق؛ ۶- الحاق یک منطقه به منطقه یا مناطق دیگر؛ ۷- الحاق دو یا چند منطقه و تعریف دوباره گروه‌ها؛ ۸- کسب داده‌های بیش‌تر و تعریف دوباره گروه‌ها اشاره نمود. سه مورد اول از موارد ذکرشده در کاستن مقدار شاخص ناهمگنی (H) مفیدتر هستند، و پنج گزینه آخر تضمین‌کننده حداقل اندازه هر خوشه است. آماره ناهمگنی (Hosking and Wallis, 1993) اطلاعات هر ایستگاه را متناسب با طول دوره آماری آن می‌سنجد. در نتیجه، تأثیر ایستگاه‌ها با طول دوره آماری بالاتر، بیش‌تر از ایستگاه‌ها با طول دوره آماری کوتاه‌تر است. مطابق جدول (۴) نتایج خوشه‌بندی با روش ترکیب نگاشت ویژگی خودسامانده و روش خوشه‌بندی سلسله‌مراتبی منجر به دستیابی به مناطق همگن‌تری می‌گردد. لذا برای ادامه محاسبات و تعدیل همگنی از نتایج این روش استفاده و با استفاده از روش‌های پیشنهادهایی که در بالا اشاره شد، مناطق همگن هیدرولوژیک استخراج گردید. به عبارت دیگر، ایستگاه ناهمگون "دارتوت" که به میزان معنی‌داری از کم‌شدن آماره ناهمگنی منطقه‌ای (H) در خوشه دو جلوگیری می‌کنند، شناسایی شد از تحلیل منطقه‌ای حذف گردید. نتایج جدول (۴) نشان می‌دهد که هر دو منطقه پس از تعدیل به‌طور قابل‌قبولی همگن هستند ($H_i > 1$). هم‌چنین بعد از تعدیل و همگنی‌کردن در خوشه اول و دوم به ترتیب هشت و نه ایستگاه هواشناسی قرار گرفت.

Table 4. Heterogeneity statistics for each region before homogeneity adjustment

Clustering Method	No. Zone	Num. stations		Before adjustment			After adjustment		
		Before	After	H1	H2	H3	H1	H2	H3
SOM+Kmeans	1	6	6	0.68	-0.43	-0.40	-	-	-
	2	11	11	1.77*	0.38	-0.06	-	-	-
SOM+Hierarchical	1	8	9	0.33	-0.25	-0.40	0.33	-0.52	-0.42
	2	9	8	1.11*	0.22	-0.8	0.71	-0.52	0.52

* منطقه ناهمگن از لحاظ هیدرولوژیک

نتایج پژوهش نشان داد که منطقه مورد مطالعه قابلیت تقسیم به دو خوشه همگن را دارد که خروجی نتایج هر دو روش خوشه‌بندی نیز بیانگر این موضوع بود. از ۱۷ ایستگاه آب‌سنجی مورد مطالعه هشت و نه ایستگاه به ترتیب در خوشه اول و دوم قرار گرفتند که بعد از تعدیل و همگن‌کردن خوشه‌ها نه ایستگاه در خوشه اول و هشت ایستگاه در خوشه دوم قرار گرفتند. نتایج این پژوهش همسو با نتایج پژوهش‌های عطایی و همکاران (۱۳۹۰) و Ghaderi et al. (2022) می‌باشد. استفاده از محاسبات نرم در این مطالعه نتایج مثبتی به همراه داشت که نشان داد این مدل‌ها همواره ابزار مناسبی جهت استفاده در زمینه مشکلات هیدرولوژیکی هستند. نتایج این پژوهش همواره استفاده از مدل‌های یادگیری ماشین در ناحیه‌ای کردن سیلاب و پیش‌بینی سیلاب به همراه خطرات احتمالی آن را پیشنهاد می‌کند که همسو با نتایج پژوهش‌های Eslaminezhad et al. (2021)، Zarei et al. (2022)، Lu et al. (2023) و Li et al. (2023) در زمینه استفاده از محاسبات نرم و مدل‌های یادگیری ماشین در ناحیه‌ای کردن مناطق سیلابی، پیش‌بینی و خطرات احتمالی سیل است.

۴- نتیجه‌گیری

در این پژوهش قابلیت خوشه‌بندی دومرحله‌ای نگاشت ویژگی خودسامانده با میانگین K و سلسله‌مراتبی در تفکیک حوضه‌های آبریز برای دستیابی به مناطق همگن هیدرولوژیکی مورد بررسی قرار گرفت. به منظور تعیین ویژگی‌های مؤثر حوضه آبریز در ایجاد سیلاب، مقادیر همبستگی عوامل تأثیرگذار در سیلاب به دست آمد. بدین منظور همبستگی خصوصیات فیزیوگرافی، هواشناسی و موقعیت جغرافیایی با میانگین سیلاب به دست آمد و خصوصیات تأثیرگذار بر سیلاب انتخاب گردید. در ادامه با استفاده از الگوریتم خوشه‌بندی دومرحله‌ای، منطقه پژوهش به دو خوشه تقسیم شد و همگنی هر یک از مناطق با استفاده از

آزمون همگنی براساس گشتاورهای خطی موردبررسی واقع گردید. با استفاده از آزمون همگنی بر مبنای گشتاورهای خطی، مناطق همگن از لحاظ هیدرولوژیکی تشخیص داده شد. نتایج نشان داد که مناطق شکل گرفته با الگوریتم خوشه‌بندی ترکیبی پیشنهادی به منظور دستیابی به مناطق همگن، از کارایی قابل قبولی برخوردار می‌باشند و الگوریتم خوشه‌بندی مورد استفاده در این مطالعه خروجی واضحی نشان می‌دهد. نتایج نشان داد که هر دو روش خوشه‌بندی به دلیل عدم وجود همه متغیرهای تأثیرگذار در تحلیل فراوانی منطقه‌ای سیلاب، برای تحلیل فراوانی منطقه‌ای سیلاب همگن نشده است. بنابراین پس از تعدیل به طور قابل قبولی همگن تبدیل شدند ($Hi > 1$). همچنین بعد از تعدیل و همگنی کردن در خوشه اول و دوم به ترتیب هشت و نه ایستگاه هواشناسی قرار گرفت. عدم وجود اطلاعات به روز شده هیدرولوژیکی از محدوده، وجود عدم قطعیت در داده‌های موجود از جمله محدودیت‌های این پژوهش به شمار می‌آید. استفاده از سایر مدل‌های خوشه‌بندی، ترکیب مدل‌های خوشه‌بندی با الگوریتم‌های تکاملی، در نظر گرفتن عدم قطعیت با استفاده از منطق فازی و مقایسه نتایج آن‌ها با نتایج این پژوهش از جمله مواردی است که توسط نویسندگان به عنوان چشم‌انداز این پژوهش می‌توان در نظر گرفت.

۵- پی‌نوشت‌ها

1. K-means
2. Hierarchical clustering
3. Self-organizing map
4. Quantile
5. Pattern analysis
6. Nodes
7. Moment Ratio Diagram

۶- تعارض منافع

هیچ‌گونه تعارض منافی توسط نویسندگان وجود ندارد.

۷- منابع

- Abedini, M., Pour Farrash Zadeh, F., & Gharachorlu, -. M. (2021). Analysis and Modeling of the Relationship between Monthly Discharge and Geomorphometric Characteristics (Case Study: Kashafrood Watershed). *Geography and Environmental Planning*, 32(4), 29-44. (In Persian)
- Ataei, H., & Shiran, M. (2011). Identifying homogeneous hydrological basins base on effective geomorphologic variants on flood by cluster analysis Introduction. *Geography and Environmental Planning*, 22(2), 79-98. (In Persian)
- Ahn, H., Shin, J. Y., Jeong, C., & Heo, J. H. (2018). Assessing applicability of self-organizing map for regional rainfall frequency analysis in South Korea. *Journal of Korea Water Resources Association*, 51(5), 383-393.
- Arora, A., Arabameri, A., Pandey, M., Siddiqui, M. A., Shukla, U. K., Bui, D. T., ... & Bhardwaj, A. (2021). Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. *Science of the Total Environment*, 750, 141565.
- Asghari Sersekanroud, S., & Saeedi Seta, A. (2023). Investigating the Effects of Land Use Changes on the Runoff of Qara Chai River Basin Using the SWAT Model. *Geography and Environmental Planning*, 34(3), 95-118. (In Persian)

- Chang, L. C., Wang, W. H., & Chang, F. J. (2021). Explore training self-organizing map methods for clustering high-dimensional flood inundation maps. *Journal of Hydrology*, 595, 125655.
- Chavoshi, S., Azmin Sulaiman, W. N., Saghafian, B., Sulaiman, M. N., & Latifah, A. M. (2012). Soft and hard clustering methods for delineation of hydrological homogeneous regions in the southern strip of the C aspian S ea W atershed. *Journal of Flood Risk Management*, 5(4), 282-294.
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227.
- Dehariya, V. K., Shrivastava, S. K., & Jain, R. C. (2010, November). Clustering of image data set using K-means and fuzzy K-means algorithms. In 2010 International conference on computational intelligence and communication networks (pp. 386-391). *IEEE*.
- Drissia, T. K., Jothiprakash, V., & Sivakumar, B. (2022). Regional flood frequency analysis using complex networks. *Stochastic Environmental Research and Risk Assessment*, 36(1), 115-135.
- Eslamian, S. S., Abrishamchi, A., & Farzamnia, K. (1999). Regionalizing Flood Frequency Estimation by Cluster Analysis. *Journal of Water and Wastewater; Ab va Fazilab*. (In Persian)
- Eslaminezhad, S. A., Eftekhari, M., Mahmoodizadeh, S., Akbari, M., & Haji Elyasi, A. (2021). Evaluation of Tree-Based Artificial Intelligence Models to Predict Flood Risk using GIS. *Iran-Water Resources Research*, 17(2), 174-189. (In Persian)
- Farsadnia, F., & Moghaddamnia, A. (2014). Regional Flood Frequency Analysis by Self-Organizing Feature Maps and Fuzzy Clustering Approach. *Iran-Water Resources Research*, 9(3), 24-36. (In Persian)
- Farsadnia, F., & Ghahreman, B. (2015). Using Hierarchical Clustering in Order to Increase Efficiency of Self-Organizing Feature Map to Identify Hydrological Homogeneous Regions for Flood Estimation. *Water and Soil*, 29(5), 1207-1218. (In Persian)
- Gao, Q., Li, G., Bao, J., & Wang, J. (2021). Regional Frequency Analysis Based on Precipitation Regionalization Accounting for Temporal Variability and a Nonstationary Index Flood Model. *Water Resources Management*, 35(13), 4435-4456.
- Ghaderi K, Motamedvaziri B, Vafakhah M, Deghani A. Comparison of Artificial Neural Network, Decision Tree and Bayesian Network Models in Regional Flood Frequency Analysis using L-moments and Maximum Likelihood Methods in Karkheh and Karun Watersheds. *JWSS*, 25 (4), 313-326 . (In Persian)
- Ghosh, S., Saha, S., & Bera, B. (2022). Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin. *Natural Hazards Research*.
- Hosking, J.R.M., & Wallis, J.R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2), 271-281. (Correction: 31(1), 1995, p. 251).
- Inyang, U. G., Akpan, E. E., & Akinyokun, O. C. (2020). A hybrid machine learning approach for flood risk assessment and classification. *International Journal of Computational Intelligence and Applications*, 19(02), 2050012.
- Kardan Moghaddam, H., Ghordoyee Milan, S., Kayhomayoon, Z., Rahimzadeh kivi, Z., & Arya Azar, N. (2021). The prediction of aquifer groundwater level based on spatial clustering approach using machine learning. *Environmental Monitoring and Assessment*, 193, 1-20.
- Kohonen T. (2001). Self-Organizing Maps. *Springer*, Berlin, Germany.
- Kumar, R., Goel, N. K., Chatterjee, C., & Nayak, P. C. (2015). Regional flood frequency analysis using soft computing techniques. *Water Resources Management*, 29(6), 1965-1978.
- Ley, R., Casper, M.C., Hellebrand, H., & Merz, R. (2011). Catchment classification by runoff behavior with self-organizing maps (SOM). *Hydrology and Earth System Sciences*, 15(9), 2947-2962.
- Lin, G., & Chen, L. (2006). Identification of homogenous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*, 324, 1-9.

- Lin, G., & Wang, C. (2006). Performing cluster analysis and discrimination analysis of hydrological factors in one step. *Advances in Water Resources*, 29, 1573-1585.
- Li, Y., & Hong, H. (2023). Modelling flood susceptibility based on deep learning coupling with ensemble learning models. *Journal of Environmental Management*, 325, 116450.
- Luu, C., Ha, H., Bui, Q. D., Luong, N. D., Khuc, D. T., Vu, H., & Nguyen, D. Q. (2023). Flash flood and landslide susceptibility analysis for a mountainous roadway in Vietnam using spatial modeling. *Quaternary Science Advances*, 11, 100083.
- MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1 (14), 281-297.
- Nachappa, T. G., Piralilou, S. T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of hydrology*, 590, 125275.
- Rao, R.A., & Srinivas, V.V. (2005). Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, 318, 37-56.
- Regier, E., Naughton, J., & McDonald, W. (2022). Transposing flood risk from extreme rainfall events: A case study of Hurricane Harvey. *Journal of Flood Risk Management*, 15(2), e12778.
- Rostami, R. (2003). Regional flood frequency analysis using linear moments (case study of Halil River basin and West Azerbaijan province basins). Master's Thesis, Faculty of Agriculture, Shahid Bahonar Kerman University. (In Persian)
- Srinivas, V.V., Tripathi, S., Ramachandra Rao, A. and Govindaraju, R. S. 2007. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *Journal of Hydrology*, (348), 148-166.
- Talukdar, S., Ghose, B., Shahfahad, Salam, R., Mahato, S., Pham, Q. B., ... & Avand, M. (2020). Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stochastic Environmental Research and Risk Assessment*, 34, 2277-2300.
- Towfiqul Islam, A.R., Talukdar, S., Mahato, S., Kundu, S., UddinEibek, K., Bao Pham, Q., Kuriqi, A., Nguyen, & ThuyLinh, T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models, *Geoscience Frontiers*, 12(3), 101075.
- Tripathi, G., Pandey, A. C., & Parida, B. R. (2022). Flood Hazard and Risk Zonation in North Bihar Using Satellite-Derived Historical Flood Events and Socio-Economic Data. *Sustainability*, 14(3), 1472.
- Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). SOFM Toolbox for Matlab 5. Technical Report A57. *Neural Networks Research Centre*, Helsinki University of Technology, Helsinki, Finland.
- Vesanto, J., & Alhoniemi, R. (2000). Clustering of the self-organizing map, *IEEE Trans. Neural, Netw*, 11 (3), 586-600.
- Wilppu, R. (1997). The Visualisation Capability of Self Organizing Maps to Detect Deviation in Distribution Control. TUCS Technical Report No. 153, *Turku Centre for Computer Science*, Finland.
- Zarei, M., Zandi, R., & Naemitabar, M. (2022). Assessment of Flood Occurrence Potential using Data Mining Models of Support Vector Machine, Chaid and Random Forest (Case study: Frizi watershed). *JWMMR*, 13 (25), 133-144. (In Persian)