



Regional Flood Frequency Analysis by Self-Organizing Feature Maps and Fuzzy Clustering Approach

F. Farsadnia^{1*} and A. Moghaddamnia²

Abstract

One of the methods for estimation of flood quantiles in ungauged watersheds or watersheds with short data records is using the regional frequency analysis method. In regional studies, the clustering methods are used to achieve homogeneous regions. Self-Organization Feature Map (SOFM) is recently used in several researches for clustering the watersheds. However the interpretation of the SOFMs output units is one of the SOFMs problems. Consequently, the trained SOFM units are used as input to the other clustering algorithms. In this study, SOFM method is used to form a two- dimension feature map, and then output nodes are fed to fuzzy c-mean clustering to form the required regions for flood frequency analysis. The optimum number of the clusters is determined by Xie-Beni and Kwon indices. The results showed that this approach has a good performance to determine homogeneous regions in Mazandaran province, northern Iran.

Keywords: Self-Organization Feature Maps, Fuzzy c-means clustering, Regionalization, hydrologic homogeneity, Mazandaran province.

Received: October 7, 2012

Accepted: July 7, 2013

تحلیل فراوانی منطقه‌ای سیلاب با استفاده از روش خوشه‌بندی ترکیبی نگاشت ویژگی خود سامان و فازی

فرهاد فرسادنیا^{۱*} و علیرضا مقدم‌نیا^۲

چکیده

یکی از روش‌های تخمین چندک‌های سیلاب در حوضه‌های فاقد آمار یا با طول دوره آماری کوتاه، استفاده از روش تحلیل فراوانی منطقه‌ای است. در مطالعات منطقه‌ای، به منظور دستیابی به مناطق همگن هیدرولوژیک از تکنیک‌های خوشه‌بندی استفاده می‌شود. اخیراً در چند تحقیق از نگاشت ویژگی خود سامان (Self-Organizing Feature Maps) استفاده شده است. اما مشکل اصلی SOFM تفسیر نقشه خروجی از این روش به منظور یافتن مناطق همگن هیدرولوژیک است. به این دلیل از SOFM به عنوان ورودی الگوریتم‌های خوشه‌بندی دیگر به کار می‌رود. در این مطالعه، ابتدا از نگاشت ویژگی خود سامان برای شکل‌گیری یک نقشه ویژگی دو بعدی استفاده شد، سپس گره‌های خروجی از نگاشت ویژگی خود سامان توسط الگوریتم خوشه‌بندی c- میانگین فازی برای شکل‌گیری مناطق مورد نیاز در تحلیل فراوانی منطقه‌ای سیلاب استفاده شد. تعداد بهینه خوشه‌های فازی براساس شاخص‌های زی-بنی (Xie-Beni) تعمیم یافته و وون (Kwon) تعیین شد. نتایج نشان داد که این تکنیک کارایی قابل قبولی در تفکیک حوضه‌های آبخیز استان مازندران به مناطق همگن هیدرولوژیک دارد.

کلمات کلیدی: نگاشت ویژگی خود سامان؛ خوشه‌بندی فازی؛ منطقه‌ای کردن؛ همگنی هیدرولوژیک؛ استان مازندران.

تاریخ دریافت مقاله: ۱۶ مهر ۱۳۹۱

تاریخ پذیرش مقاله: ۱۶ تیر ۱۳۹۲

1- Ph.D Student of Irrigation and Drainage, Faculty of Agriculture, Ferdowsi University, Mashhad, Iran, Email: farhadfarsad@gmail.com

2- Associate Professor, Faculty of Natural Resources, Mountainous and Dry Region Restoration Group, University of Tehran, Iran Email: ali.moghaddamnia@live.com

*- Corresponding Author

۱- دانشجوی دکتری آبیاری و زهکشی، دانشکده کشاورزی، گروه آبیاری و زهکشی، دانشگاه فردوسی مشهد، مشهد، ایران.

۲- دانشیار هیدرولوژی، دانشکده منابع طبیعی، گروه احیای مناطق خشک و کوهستانی، دانشگاه تهران، تهران، ایران.

*- نویسنده مسئول

نسبت به دو روش دیگر از توانایی‌های بیشتری در تعیین زیر منطقه‌های همگن هیدرولوژیک برخوردار است. Lin and Chen (2006) از الگوریتم SOFM برای خوشه‌بندی ۱۵۴ ایستگاه باران‌سنجی در تایوان به منظور تحلیل فراوانی منطقه‌ای استفاده کردند. آنها از ۱۷ خصوصیت حوضه‌های آبخیز برای خوشه‌بندی استفاده کردند و اندازه نقشه خروجی را به منظور مطمئن شدن از حداکثر تعداد خوشه، 12×12 انتخاب کردند و در نهایت منطقه مورد بررسی را به ۸ خوشه همگن تقسیم کردند. آنها خوشه‌های بدست آمده از SOFM را با دو الگوریتم سنتی وارد و میانگین k مقایسه کردند و دریافتند که SOFM نسبت به دو روش دیگر توانایی بیشتری برای شناسایی مناطق همگن دارد. Ley et al., (2011) از دو نوع ورودی SOFM برای مقایسه شباهت‌های هیدرولوژیکی ۵۳ حوضه آبخیز در آلمان استفاده کردند. آنها گروه حوضه‌هایی که با رفتار پاسخ هیدرولوژیکی خوشه‌بندی شده بودند را با خوشه‌هایی که براساس خصوصیات حوضه‌های آبخیز خوشه‌بندی شده‌اند مقایسه کردند. نتایج آنها یک همپوشانی ۶۷٪ بین این دو نوع خوشه‌بندی نشان داد که می‌تواند با استفاده از توپولوژی SOFM بهبود یابد. Chavoshi et al., (2012) به منظور مقایسه‌ی روش‌های خوشه‌بندی سخت (سلسله مراتبی و k -میانگین) با روش‌های نرم (C -میانگین و کوهونن) 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 بدست آمده با روش‌های نرم همگن‌تر هستند. فرسنادیا و همکاران (۱۳۹۱) از تحلیل خوشه‌ای فازی برای تعیین مناطق همگن استان مازندران استفاده کردند و روند منطقه‌ای بارندگی‌ها در سه منطقه‌ی بدست آمده از الگوریتم خوشه‌بندی فازی را مورد مطالعه قرار دادند.

هرچند SOFM در خوشه‌بندی حوضه‌های آبخیز به نواحی همگن هیدرولوژیک با موفقیت استفاده شده است، اما تجزیه و تحلیل نقشه خروجی SOFM دشوار است. به طور کلی این الگوریتم به ندرت خوشه‌های واضحی در خروجی نشان می‌دهد. به منظور حل این مشکل پژوهشگران راهکارهای متفاوتی ارائه کردند. Lampinen and Oja (1992) یک الگوریتم خوشه‌بندی دو مرحله‌ای ارائه کردند که خروجی SOFM اول را به‌عنوان ورودی SOFM دوم استفاده کردند. آزمایشات آنها روی داده‌های طبیعی و مصنوعی نشان داد که الگوریتم SOFM ترکیبی از قابلیت بیشتری نسبت به الگوریتم میانگین k کلاسیک و SOFM معمولی برای تفکیک کلاس‌های مختلف داده‌ها برخوردار است.

در هیدرولوژی برآورد فراوانی و بزرگی مقادیر حد طبیعی مانند سیلاب، جریان کمینه و حداکثر بارندگی اهمیت فراوانی دارد. به دلیل نادر بودن مقادیر حدی و کوتاه بودن طول دوره آماری، تخمین فراوانی وقوع این حوادث دشوار است، بنابراین از تحلیل فراوانی منطقه‌ای برای تخمین واقع بینانه‌ی چندک‌ها^۱ استفاده می‌شود. در روش تحلیل فراوانی منطقه‌ای، باید ایستگاه‌ها به مناطق همگن اختصاص یابند، زیرا همگنی منطقه برای اطمینان از قابل اعتماد بودن تحلیل منطقه‌ای الزامی است (Hosking and Wallis, 1997). زمانی که در منطقه تعداد ایستگاه‌های زیادی برای تحلیل فراوانی منطقه‌ای وجود نداشته باشد، شناسایی مناطق همگن هیدرولوژیکی دشوار می‌باشد. به این دلیل هیدرولوژیست‌ها از تحلیل خوشه‌ای برای دستیابی به مناطق همگن هیدرولوژیک استفاده می‌کنند. روش‌های خوشه‌بندی متعددی برای طبقه‌بندی حوضه‌های آبخیز ارائه شده است. از جمله این روش‌ها می‌توان به خوشه‌بندی میانگین k (Burn and Goel, 2000) و خوشه‌بندی سلسله مراتبی تراکمی (Hosking and Wallis, 1997) و خوشه‌بندی هیبرید (Rao and Srinivas, 2006) اشاره نمود.

الگوریتم نگاشت ویژگی خود سامان (Kohonen, 1982) یک مدل اکتشافی برای مجسم ساختن و کشف روابط خطی و غیر خطی در مجموعه داده‌ها با ابعاد زیاد است. SOFM کاربرد زیادی در کارهای صنعتی مانند شناسایی گفتار، مدل سازی زیستی، فشرده سازی داده‌ها، پردازش سیگنال و داده کاوی دارد (Kohonen, 2001). اخیراً نگاشت ویژگی خود سامان به عنوان یک ابزار انفورماتیک مدرن در شناسایی مناطق همگن هیدرولوژیک به کار رفته است. (Hall and Minns (1999) از SOFM برای منطقه‌ای کردن ایستگاه‌های آب‌سنجی در جنوب غربی انگلستان و ولز با استفاده از مساحت حوضه آبخیز، طول آبراهه اصلی، شیب آبراهه اصلی، میانگین بارندگی سالانه و شاخص خاک استفاده کردند. آنها نرون‌های خروجی از SOFM را به سه گروه مشخص به منظور دستیابی به سه منطقه‌ی همگن تقسیم کردند. (Lin and Wang (2006) یک روش تک مرحله‌ای را برای انجام تحلیل خوشه‌ای براساس SOFM ارائه کردند. آنها این روش را برای فاکتورهای هیدرولوژیکی تأثیر گذار در منحنی‌های مدت جریان کمینه در جنوب تایوان به کار بردند. (Jingyi and Hall (2004) روش‌های خوشه‌بندی وارد، خوشه‌بندی فازی و شبکه عصبی کوهونن را برای طبقه‌بندی حوضه‌ی رودخانه گان-مینگ در جنوب شرق چین بکار بردند. نتایج آنها نشان داد که شبکه عصبی کوهونن

(2000) Vesanto and Alhoniemi از الگوریتم خوشه‌بندی سلسله مراتبی و میانگین k برای خوشه‌بندی خروجی SOFM استفاده کردند. آنها مهمترین مزیت مدل ارائه شده را کاهش نسبتاً زیاد بارگذاری انجام شده بوسیله الگوریتم‌ها، امکان خوشه‌بندی مجموعه داده‌های بزرگ و بررسی چندین استراتژی پردازش متفاوت در زمان محدود اعلام کردند. (Srinivas et al., 2007) از روش ترکیبی خوشه‌بندی SOFM و خوشه‌بندی فازی برای شناسایی مناطق همگن هیدرولوژیک در حوضه‌های آبخیز ایندیانا در ایالات متحده آمریکا استفاده کردند و منطقه مورد مطالعه را به 7 خوشه همگن تقسیم کردند.

در این مقاله، از یک تکنیک خوشه‌بندی دو مرحله‌ای برای منطقه‌ای کردن حوضه‌های آبخیز استان مازندران استفاده شده است. در مرحله اول، از SOFM برای شکل‌دهی نقشه ویژگی دو بعدی استفاده شد. سپس گره‌های خروجی از SOFM با استفاده از الگوریتم c -میانگین فازی (FCM) خوشه‌بندی شد. در مرحله بعد پس از تعدیل همگنی مناطق به دست آمده از خوشه‌بندی، تابع توزیع منطقه‌ای برای هر یک از مناطق همگن بدست آمد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه و داده‌های مورد استفاده

منطقه مورد مطالعه، استان مازندران در شمال ایران می‌باشد (شکل ۱). این استان با مساحتی بالغ بر ۲۴ هزار کیلومتر مربع در محدوده ۳۵ درجه و ۴۷ دقیقه تا ۳۶ درجه و ۵۸ دقیقه عرض شمالی و ۵۰ درجه و ۳۴ دقیقه تا ۵۴ درجه و ۱۴ دقیقه طول شرقی قرار دارد. ارتفاع استان مازندران از زیر صفر در کناره‌های دریای خزر تا ۵۶۷۰ متر در قله دماوند متغیر است (وزارت نیرو، ۱۳۸۷). میانگین بارندگی در غرب استان مازندران بیشتر از ۱۰۰۰ میلی‌متر و در شرق منطقه بیش از ۳۰۰ میلی‌متر است (وزارت نیرو، ۱۳۸۷). قسمت مرطوب غربی ارتفاع کمی دارد اما بارندگی زیاد در آن باعث مناسب شدن آن منطقه برای فعالیت‌های کشاورزی شده است. برای تحلیل آماری ایستگاه‌های آب‌سنجی منطقه، ابتدا از طریق آب منطقه‌ای استان مازندران آمار حداکثر دبی سالانه و خصوصیات حوضه‌های آبخیز استان مازندران تهیه شد. سپس با توجه به رعایت حداقل طول دوره آماری مورد نیاز، از ۷۸ ایستگاه موجود ایستگاه‌ها با طول دوره آماری کمتر از ۱۲ سال حذف و تعداد ۴۷ ایستگاه برای ادامه مطالعات مورد استفاده قرار گرفت (جدول ۱).

(Hosking and Wallis (1997) اظهار نمودند که می‌توان میانگین

بارندگی سالانه را حتی برای حوضه‌هایی که اندازه‌گیری در آنها انجام نشده را از نقشه خطوط هم‌بارش به طور قابل اطمینانی تخمین زد. بر این اساس میانگین بارندگی سالانه‌ی هر حوضه، از نقشه‌ی هم‌باران تهیه شده توسط شرکت مدیریت منابع آب ایران استخراج گردید.

۲-۲- آماده سازی داده‌ها

داده‌های ورودی باید قبل از ورود به الگوریتم SOFM نرمالیزه 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 به صورت غیر نرمالیزه و خام به الگوریتم SOFM تغذیه شوند، متغیرهایی که دارای مقادیر بزرگتری هستند نقش مهمتری در فرآیند به روزرسانی وزن‌ها ایفا می‌کنند. بنابراین نرمالیزه 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 زیر بدست آید:

$$X' = (X_i - \bar{X}) / \sigma_x \quad (1)$$

که \bar{X} میانگین مقادیر و σ_x انحراف معیار می‌باشند. این تغییر شکل داده‌ها باعث می‌شود تا میانگین مجموعه داده‌ها ۰ و انحراف معیار آنها ۱ شود.

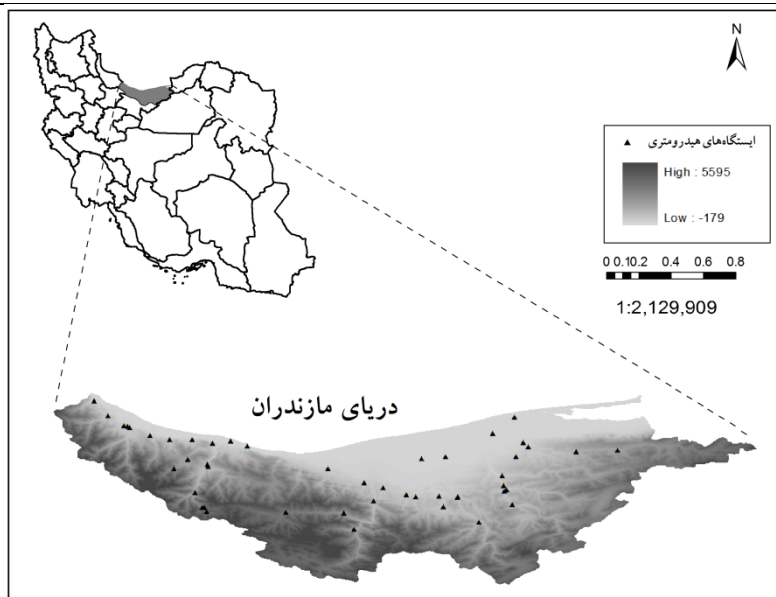
۲-۳- نگاشت ویژگی خود سامان (SOFM)

نگاشت خود سامان تابع چگالی احتمال از داده‌های ورودی تحت یک الگوریتم یادگیری غیر نظارت شونده است، که یک روش موثر در خوشه‌بندی است که برای خلاصه سازی و بصری نمودن 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, 2001, 1982). SOFM شامل دو لایه است: یک لایه ورودی شکل گرفته از مجموعه گره‌ها^۴ (یا نرون‌هایی که واحدهای^۵ محاسباتی دارند) و یک لایه خروجی (لایه کوهونن) که توسط گره‌هایی که در یک شبکه دو بعدی قرار گرفته‌اند، تشکیل شده است (شکل ۲).

تعداد نرون‌های خروجی SOFM (یا اندازه نقشه) 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). می‌توان تعداد نرون‌های خروجی در یک SOFM را با استفاده از روش ابتکاری پیشنهاد شده توسط Vesanto et al. (2000) انتخاب کرد.

جدول ۱- خصوصیات ایستگاه های مورد بررسی و برخی پارامترهای آماری

ردیف	ایستگاه	مساحت حوضه آبخیز (کیلومتر مربع)	طول شاخه اصلی (کیلومتر)	عرض جغرافیایی	طول جغرافیایی	میانگین بارندگی سالانه (میلی متر)	ردیف	ایستگاه	مساحت حوضه آبخیز (کیلومتر مربع)	طول شاخه اصلی (کیلومتر)	عرض جغرافیایی	طول جغرافیایی	میانگین بارندگی سالانه (میلی متر)
۱	پایین زرنندین	۲۶۱.۹۲	۱۶.۳۹	۳۶-۳۷	۵۳-۱۹	۷۶۰	۲۵	کره سنگ	۳۹۸۶.۶۶	۴۶.۳۸	۳۶-۱۶	۵۲-۲۲	۷۰۰
۲	سفید چاه	۱۰۳۶.۹۲	۱۸.۵۲	۳۶-۳۵	۵۳-۵۲	۴۲۴	۲۶	بلیران	۸۲.۰۶	۱۲.۱۲	۳۶-۲۱	۵۲-۲۵	۷۵۰
۳	گلورد	۱۴۲۷.۷۴	۵۱.۸۵	۳۶-۳۵	۵۳-۳۷	۹۲۴	۲۷	آغوزکنی	۱۴۰.۶۲	۱۸.۴۷	۳۶-۲۹	۵۲-۰۵	۷۵۶
۴	آبلو	۱۹۰۵.۶۱	۴۵.۲۲	۳۶-۳۸	۵۳-۱۷	۷۱۷	۲۸	نوشهر	۷۵.۵۰	۱۳.۴۳	۳۶-۳۹	۵۱-۲۸	۱۱۹۰
۵	نوذر آباد	۲۰۱۷.۳۲	۱۱۲.۳۴	۳۶-۴۸	۵۳-۱۴	۶۹۰	۲۹	خیرود کنار	۲۱۰.۹۲	۱۹.۶۷	۳۶-۳۷	۵۱-۳۴	۱۳۵۰
۶	سلیمان تنگه	۱۲۴۸.۲۵	۴۱.۴۱	۳۶-۱۵	۵۳-۱۳	۶۰۱	۳۰	پل مرگن	۵۹.۳۱	۹.۳۰	۳۶-۱۴	۵۱-۱۹	۵۵۰
۷	واستان	۱۳۶.۳۵	۱۰.۳۰	۳۶-۲	۵۳-۱	۷۵۰	۳۱	دره هریجان	۸۴.۸۴	۶.۹۹	۳۶-۱۳	۵۱-۱۸	۵۷۰
۸	ورند	۱۲۰۲.۳۸	۴۱.۵۹	۳۶-۲	۵۳-۱۱	۷۵۰	۳۲	ولی آباد	۱۸۱.۱۰	۹.۳۰	۳۶-۱۴	۵۱-۱۸	۵۷۰
۹	ریگ چشمه	۲۱۱۵.۲۵	۳۳.۹۳	۳۶-۲۲	۵۳-۱	۷۷۳	۳۳	آبشار	۵۸۶.۱۹	۳۹.۸۶	۳۶-۱۹	۵۱-۱۵	۴۷۰
۱۰	گرم رود	۸۷۶.۷۹	۵۳.۸۵	۳۶-۲۶	۵۳-۰۹	۷۶۰	۳۴	دوآب	۶۲۷.۴۳	۲۲.۸۵	۳۶-۲۹	۵۱-۲	۳۴۰
۱۱	کردخیل	۴۰۲۶.۵۷	۸۰.۲۰	۳۶-۴۲	۵۳-۰۶	۶۸۱	۳۵	پل ذغال	۱۵۸۳.۳۶	۳۹.۸۶	۳۶-۳	۵۱-۲	۳۱۸
۱۲	داراب کلا	۲۷.۵۷	۵.۷۵	۳۶-۳۳	۵۳-۱۵	۷۷۵	۳۶	کلاردشت	۱۹۰.۴۷	۱۲.۰۰	۳۶-۲۸	۵۱-۰۷	۵۶۰
۱۳	کریکلا	۵۶۶.۶۴	۲۳.۲۲	۳۶-۰۸	۵۳-۰۱	۶۸۰	۳۷	والت	۳۳۰.۵۱	۱۵.۸۰	۳۶-۳۲	۵۱-۱۲	۴۵۰
۱۴	کیاکلا	۲۳۸۶.۷۹	۶۷.۹۶	۳۶-۳۳	۵۲-۴۸	۷۱۲	۳۸	زوات	۴۱۹.۰۱	۲۹.۲۵	۳۶-۳۸	۵۱-۲۱	۱۸۰
۱۵	شیرگاه	۳۴۲.۹۰	۲۲.۲۵	۳۶-۱۸	۵۲-۵۳	۱۰۲۹	۳۹	چارز	۱۸.۷۱	۶.۰۸	۳۶-۴	۵۱-۱۴	۱۴۵۰
۱۶	شیرگاه	۱۷۷۶.۱۶	۸۵.۰۵	۳۶-۱۷	۵۲-۵۳	۱۰۲۹	۴۰	ماشالله آباد	۱۵۱.۳۵	۱۱.۹۳	۳۶-۴	۵۱-۰۶	۱۵۹۳
۱۷	پاشاکلا	۲۱۱.۳۴	۱.۹۲	۳۶-۱۴	۵۲-۴۸	۹۸۰	۴۱	هراتیر	۷۷۶.۲۰	۲۹.۱۵	۳۶-۴۵	۵۰-۵	۱۳۴۸
۱۸	قران تالار	۴۰۶.۶۲	۴.۲۷	۳۶-۱۸	۵۲-۴۶	۱۰۸۴	۴۲	قلعه گردن	۱۶۱.۹۵	۱۵.۶۲	۳۶-۴۴	۵۰-۵	۱۳۵۰
۱۹	کشتارگاه	۱۶۲۵.۱۵	۳۹.۸۲	۳۶-۳۲	۵۲-۳۹	۶۸۶	۴۳	گانگسر	۴۰۹.۲۳	۱۷.۵۰	۳۶-۴۹	۵۰-۴۳	۱۰۸۲
۲۰	دیوا	۱۲۹.۷۱	۱۱.۳۴	۳۶-۱۹	۵۲-۳۴	۸۹۵	۴۴	رامسر	۱۳۵.۸۲	۱۸.۴۰	۳۶-۵۴	۵۰-۳۷	۱۳۲۱
۲۱	گلوگاه-بندی	۲۵۹.۰۴	۳۷.۰۹	۳۶-۱۸	۵۲-۳۷	۱۰۱۰	۴۵	دینارسرا	۲۲۴.۶۹	۱۹.۵۳	۳۶-۴۱	۵۰-۵۸	۱۴۷۴
۲۲	بلده	۷۵۲.۷۵	۱۳.۰۴	۳۶-۱۲	۵۱-۴۹	۳۰۰	۴۶	اسکو محله	۸۱.۳۰	۹.۷۵	۳۶-۲۳	۵۲-۱۸	۸۳۰
۲۳	رزن	۱۱۸۲.۳۱	۳۸.۹۶	۳۶-۱۱	۵۲-۱	۳۳۵	۴۷	رضایط	۱۰۸.۱۹	۱۲.۴۳	۳۶-۴۵	۵۰-۴۸	۱۲۵۰
۲۴	پنجاه	۲۳۵.۵۲	۱۵.۲۵	۳۶-۰۵	۵۲-۱۴	۲۲۰							



شکل ۱- موقعیت استان مازندران و ایستگاه‌های مورد مطالعه

از گره برنده (c) در طی فرآیند یادگیری به روز رسانی می‌شود. این فرآیند 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. برای جزئیات بیشتر در مورد الگوریتم SOFM می‌توان به (2003) Haykin مراجعه کرد.

ماتریس وزن‌های نهایی بعد از مرحله SOFM یک ماتریس $n \times m'$ به نام ماتریس W' است.

$$W' = \begin{bmatrix} W_{11} & \dots & W_{1m'} \\ \vdots & & \vdots \\ W_{n1} & \dots & W_{nm'} \end{bmatrix} \quad (3)$$

۲-۴- خوشه‌بندی واحدهای SOFM

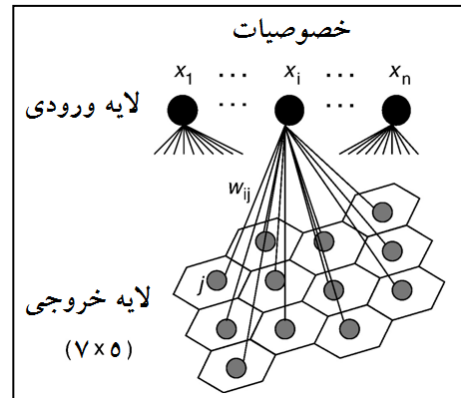
در این مطالعه خروجی SOFM با استفاده از ماتریس U و روش خوشه‌بندی فازی، تفکیک شد. هرچند روش‌های دیگری نیز به این منظور پیشنهاد شده است، اما هر یک مزایا و معایب خاص خود را در تعیین خوشه‌های نهایی دارد. به‌عنوان مثال اخیراً استفاده از ماتریس U برای تعیین خوشه‌ها بسیار رایج شده است اما گاهی اوقات مرز بین خوشه‌ها روی ماتریس U واضح نیست (Kohonen, 2001).

تاکنون برای تفکیک نقشه آموزش دیده SOFM به چند خوشه (گروه)، از روش‌های خوشه‌بندی فازی و خوشه‌بندی سلسله مراتبی و خوشه‌بندی میانگین k استفاده شده است. (2000) Vesanto and Alhoniemi گزارش دادند که ماتریس U نتوانست مرزهای واضحی روی خروجی SOFM آموزش دیده پیدا کند اما خوشه‌بندی سلسله مراتبی و میانگین k مرز بین خوشه‌ها را به طور واضح نشان داد.

در این مطالعه ابتدا الگوریتم ماتریس فواصل متحد شده یا ماتریس U (Ultsch, 1993; Ultsch and Siemon, 1990) به کار رفت. ماتریس U فواصل بین واحدهای همسایه در نقشه خروجی را محاسبه می‌کند. این فواصل می‌تواند برای ارائه خوشه‌ها با استفاده از مقیاس خاکستری^۷ روی نقشه نشان داده شود (Kohonen, 2001). همچنین ماتریس وزن‌های خروجی SOFM به‌عنوان ورودی در الگوریتم خوشه‌بندی فازی (Bezdek, 1981) برای مشخص کردن خوشه‌ها روی نقشه آموزش یافته به کار رفت.

۲-۵- خوشه بندی فازی

الگوریتم خوشه‌بندی c -میانگین فازی (FCM) یک تکنیک تحلیل داده‌های چند متغیره است که یک مجموعه داده



شکل ۲- طرح شماتیک از ساختار نگاشت ویژگی خود سامان

تقریباً تعداد بهینه‌ی واحدهای نقشه برابر $5 \times \sqrt{N}$ می‌باشد که N تعداد نمونه‌ها در مجموعه داده‌ها هستند. در این مطالعه تعداد گره‌ها در لایه ورودی SOFM (یعنی خصوصیات انتخاب شده برای تحلیل خوشه‌ای که در جدول ۱ آورده شده‌اند) برابر ۵ است. لایه خروجی از S گره خروجی در شبکه‌ی شش وجهی (در این مطالعه ۳۵ گره به صورت شبکه 7×5) برای فراهم کردن تجسم بهتر بدست می‌آید. باید توجه کرد که شبکه‌ی شش وجهی نسبت به شبکه مستطیلی ارجح‌تر است، زیرا از هیچ یک از جهت‌های عمودی یا افقی جانبداری نمی‌کند (Kohonen, 2001).

هر گره در لایه ورودی به هر گره در لایه خروجی توسط شبکه‌ی 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 باز می‌گرداند. وزن‌ها یک شبکه بین واحدهای ورودی (بردار مشخصه‌ها) و واحدهای خروجی وابسته به آنها (گروه‌هایی از بردار مشخصه‌ها) برقرار می‌کند.

عملکرد الگوریتم به شرح زیر می‌باشد: زمانی که یک بردار مشخصه ورودی X' به SOFM 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 SOFM بردار وزن گره برنده و همسایه‌های از پیش تعریف شده‌اش در الگوریتم، براساس معادله زیر به روز رسانی می‌شوند:

$$W_{ij}(t+1) = W_{ij} + \alpha(t) h_{jc}(t) [X_i(t) - W_{ij}(t)] \quad (2)$$

که $W_{ij}(t)$ 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 هموار سازی تعریف شده روی نقاط شبکه) است که مقدار همسایگی

محدودیت اول (معادله ۵) برطرف می‌شود.

$$u_{ik} = \frac{u_{ik}^{init}}{\sum_{i=1}^c u_{ik}^{init}} \quad \text{for } 1 \leq i \leq c, 1 \leq K \leq N \quad (7)$$

اگر الگوریتم FCM که با ماتریس مرکز خوشه V (شامل c مرکز خوشه فازی $v_1^{init}, \dots, v_c^{init}$) مقدار دهی اولیه شده، میزان تعلق (u_{ik}^{init}) را با استفاده از جایگذاری v_i^{init} به جای v_i در معادله (۹) تعیین می‌کند.

iii- مرکز خوشه فازی (v_i) به ازا $i=1, 2, \dots, c$ توسط معادله (۸) محاسبه می‌شود.

$$V_i = \frac{\sum_{K=1}^N (u_{iK})^m X_K}{\sum_{K=1}^N (u_{iK})^m} \quad (8)$$

iv- میزان تعلق (u_{ik}) با استفاده از معادله (۹) بهنگام سازی می‌شود.

$$u_{ik} = \frac{\left(\frac{1}{d^2(X_K, V_i)} \right)^{1/(m-1)}}{\sum_{i=1}^c \left(\frac{1}{d^2(X_K, V_i)} \right)^{1/(m-1)}} \quad (9)$$

$$\text{for } 1 \leq i \leq c, 1 \leq K \leq N$$

مراحل (iv), (iii) تا زمانی که مقدار ماتریس عضویت بین دو تکرار متوالی به حد کافی کوچک شود، تکرار می‌شود. در این هنگام، روش‌های مرسوم تحلیل خوشه‌ای فازی توصیه می‌کنند که ماتریس عضویت فازی را غیر فازی کنند، تا سر انجام بردار خصوصیات خوشه‌ها اختصاص یابند.

ماتریس عضویت فازی می‌تواند با استفاده از روش حداکثر تعلق غیر فازی یا سخت شود (Ross, 1995). در روش حداکثر عضویت، به بزرگ‌ترین المان در هر ستون ماتریس عضویت فازی، مقدار تعلق واحد اختصاص می‌یابد (معادله ۱۰). به عبارت دیگر، بردار خصوصیت به خوشه‌ای که بیشترین تعلق را دارد اختصاص می‌یابد.

$$\text{if } d_{ik} = \min_{1 \leq i \leq C} \{d_{ik}\} = \min_{1 \leq i \leq C} \|V_i - X_K\| \quad (10)$$

$$u_{jk} = \max_{1 \leq i \leq c} \{u_{ik}\} = 1; \quad u_{ik} = 0 \quad \text{for all } i \neq j$$

۲-۶- شاخص‌های ارزیابی خوشه‌بندی فازی

یکی از مهمترین مسایل در خوشه‌بندی انتخاب تعداد خوشه‌ی مناسب می‌باشد.

$X = \{x_1, \dots, x_n\} \subset R^P$ (در فضای P بعدی اقلیدسی) را به $V_i (i=1, \dots, c)$ خوشه فازی که با مراکز خوشه $c \in \{2, \dots, n-1\}$ شناخته می‌شوند، تفکیک می‌کند. این الگوریتم، براساس بهینه سازی تابع هدف زیر با تکرار است، که N حوضه آبخیز موجود در یک ناحیه را به c خوشه‌ی فازی تفکیک می‌کند.

$$\text{Minimize } J(U, V : X) = \sum_{i=1}^C \sum_{K=1}^N (u_{ik})^m d^2(X_K, V_i) \quad (4)$$

مشروط به دو محدودیت زیر،

$$\sum_{i=1}^C u_{ik} = 1 \quad \forall k \in \{1, \dots, N\} \quad (5)$$

$$0 < \sum_{K=1}^N u_{ik} < N \quad \forall i \in \{1, \dots, c\} \quad (6)$$

که u_{ik} درجه عضویت خصوصیت k در خوشه i است و U ماتریس عضویت فازی است. $m \in [1, \infty]$ توان وزنی یا پارامتر فازی سازی است که درجه فازی بودن نتایج خوشه‌بندی را کنترل می‌کند و درجه همپوشانی بین خوشه‌ها است. با حداقل مقدار معنی‌دار $m=1$ خوشه‌بندی FCM تبدیل به خوشه‌بندی میانگین k می‌شود. با میل m به سمت بی‌نهایت به بالاترین درجه از فازی شدن میل می‌کند (Bezdek, 1981). معمولاً مقدار پیش فرض $m=2$ به کار می‌رود (Hathaway and Bezdek., 2001). $d^2(x_k, v_i)$ فاصله k بردار مشخصه تا مرکز i خوشه (v_i) است. اولین محدودیت تابع هدف J (معادله ۵) بیانگر این است که ماتریس عضویت^۱ در محدوده $[0, 1]$ می‌باشد.

دومین محدودیت (معادله ۶) مجموع درجه عضویت در یک خوشه فازی را بین 0 و N نشان می‌دهد. اگر مجموع درجه عضویت‌ها در یک خوشه فازی مساوی صفر باشد، ایجاب می‌کند که خوشه حاوی هیچ ایستگاهی نباشد. در مقابل، اگر مجموع برابر N شود، خوشه فازی تمام بردار مشخصات را در برمی‌گیرد. بنابراین، الگوریتم FCM در افراز بهینه، حداقل یک بردار مشخصه را به هر خوشه اختصاص می‌دهد.

روش تکرار الگوریتم FCM (Bezdek, 1981) به صورت زیر خلاصه شده است:

i- توسط مولد اعداد تصادفی به ماتریس عضویت فازی U (یا ماتریس مرکز خوشه‌ها V) مقدار دهی اولیه می‌شود.

ii- اگر الگوریتم FCM که با ماتریس عضویت فازی U مقدار دهی اولیه شده، تعلق اولیه U_{ik}^{init} را با استفاده از معادله (۷) تعدیل کند،

برای رفع این مشکل، Kwon (1998) شاخص ارزیابی جدید V_k را معرفی کرد.

$$V_k(U, V : X) = \frac{\sum_{i=1}^C \sum_{K=1}^N (u_{ik})^m \|V_i - X_K\|^2 + \frac{1}{C} \sum_{i=1}^C \|V_i - V\|^2}{N \min_{i \neq k} \|v_i - v_k\|^2} \quad (13)$$

۷-۲- گشتاورهای وزنی

گشتاور وزنی توسط Greenwood et al. (1979) به صورت زیر تعریف شد:

$$M_{p,r,s} = \int_0^1 [(F)]^p F^r (1-F)^s dF \quad (14)$$

معمولاً گشتاور وزنی به دو صورت زیر مورد بررسی قرار می‌گیرند:

$$\alpha_s = \sum_{k=0}^s \binom{s}{k} (-1)^k \beta_k \quad (15)$$

$$\beta_r = \sum_{k=0}^r \binom{r}{k} (-1)^k \alpha_k \quad (16)$$

که در آن‌ها α_s و β_s به صورت زیر است:

$$M_{1,0,s} = \alpha_s = \int_0^1 X(F)(1-F)^s dF \quad (17)$$

$$M_{1,r,0} = \beta_r = \int_0^1 X(F)F^r dF \quad (18)$$

همچنین، گشتاور خطی یک ترکیب خطی از گشتاور وزنی (PWMs) است که به داده پرت غیر حساس است (Vogel and Fennessey, 1993) و توسط Hosking (1986) این گشتاور که فرم اصلاح شده گشتاور وزنی است، به صورت زیر تعریف شده است:

$$\lambda_1 = \beta_0 \quad (19)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad (20)$$

$$(21)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$\lambda_4 = 20\beta_3 - 20\beta_2 + 12\beta_1 - \beta_0 \quad (22)$$

که λ_r گشتاور خطی و β_r گشتاور وزنی است و به صورت زیر تعریف شده است:

$$\beta_r = \int_0^1 x(F)F^r dF \quad (23)$$

در روابط فوق λ_1 میانگین گشتاور خطی یا معیار تمایل به مرکز و λ_2 انحراف معیار گشتاور خطی یا معیار پراکندگی است.

بدین منظور باید هر خوشه بیشترین فشردگی را داشته باشند و تا حد امکان بین خوشه‌ها حداکثر فاصله وجود داشته باشد. برای یک خوشه‌بندی مناسب باید هر دو معیار با هم ارضا شوند چرا که اگر تنها معیار فشردگی مورد استفاده قرار گیرد، هر داده می‌تواند به صورت یک خوشه در نظر گرفته شود چرا که هیچ خوشه‌ای فشردتر از خوشه‌ای با یک داده نمی‌باشد و اگر تنها معیار جدایی در نظر گرفته شود در آن صورت بهترین خوشه‌بندی این می‌باشد که کل داده‌ها را یک خوشه بگیریم با این توضیح که فاصله هر خوشه از خودش صفر است. بنابراین باید از ترکیب دو معیار فوق استفاده شود. اگر چه در روش FCM تعداد خوشه‌ها از قبل مشخص شده است ولی در ابتدای کار تعداد خوشه‌ها برای هیدرولوژیست‌ها مشخص نیست و با روش سعی و خطا تعداد مناسب خوشه‌ها تعیین می‌شود. برای مشخص کردن تعداد درست خوشه‌ها توابع ارزیابی مختلفی تعریف شده است که می‌توان با استفاده از آنها تعداد خوشه‌ها را مشخص کرد. در ادامه دو شاخص ارزیابی خوشه‌بندی فازی معرفی شده است.

تابع زی-بنی: شاخص پیشنهاد شده توسط Xie and Beni (1991) تابعی از مجموعه داده و مراکز خوشه‌ها است. که از نسبت فشردگی کل بر جدایی یک c -افراز فازی تشکیل شده است. این شاخص به صورت زیر تعریف شده است:

$$XB(U, V : X) = \frac{\sum_{i=1}^C \sum_{K=1}^N (u_{ik})^2 \|V_i - X_K\|^2}{N \min_{i \neq K} \|V_i - V_K\|^2} \quad (11)$$

که جزء صورت کسر مجموع مربعات اختلاف فازی هر بردار خصوصیت $x_i (i=1, \dots, c)$ از مرکز هر خوشه فازی $x_k (k=1, \dots, N)$ است. مقدار این جزء با افزایش فشردگی خوشه کاسته می‌شود. جزء مخرج کسر، حداقل اندازه جدایی بین مراکز خوشه‌ها است، برای خوشه‌هایی که به خوبی مجزا شده‌اند مقدار بزرگتری دارد. مقدار حداقل XB به یک افراز خوب اشاره دارد، که به فشردگی و نیز جدایی خوب خوشه‌ها بستگی دارد.

Xie and Beni (1991) توصیه کردند زمانی که در معادله (۴)، $m \neq 2$ باشد در معادله (۱۱) به جای $(u_{ik})^2$ مقدار $(u_{ik})^m$ جایگزین شود. شاخص زی-بنی گسترش یافته از رابطه زیر بدست می‌آید:

$$XB, m(U, V : X) = \frac{\sum_{i=1}^C \sum_{K=1}^N (u_{ik})^m \|V_i - X_K\|^2}{N \min_{i \neq K} \|V_i - V_K\|^2} \quad (12)$$

زمانی که c زیاد می‌شود مقدار XB به طور یکنواخت کاهش می‌یابد.

۲-۸-۲-آزمون ناهمگنی

اگر تغییر پذیری یا فضای پراکنش ایستگاه‌ها بزرگ باشد، احتمال تعلق این ایستگاه‌ها به یک مجموعه واحد را می‌توان به وسیله آزمون ناهمگنی گشتاورهای خطی بررسی کرد که شامل سه آماره H_1, H_2, H_3 است، برای بررسی ضریب تغییرات خطی (LCV)، H_2 برای بررسی ترکیبی از ضریب تغییرات خطی و ضریب چولگی (LCS) و آماره H_3 برای بررسی ترکیبی از ضریب چولگی خطی و ضریب کشیدگی خطی (LCK) می‌باشد. اگر مقدار این آماره‌ها کم‌تر از یک باشد، منطقه همگن و اگر بین ۱ تا ۲ باشد، منطقه احتمالاً ناهمگن و اگر بزرگ‌تر از ۲ باشد، منطقه کاملاً ناهمگن است. در عمل عنوان می‌شود که معیار H_1 به منظور بررسی این آزمون مناسب‌تر است (Hosking, 1986). مقدار این آماره‌ها از روابط زیر به دست می‌آید:

$$H_i = (V_{obs} - \mu_V) / \sigma_V \quad (29)$$

$$V_1 = \sum_{i=1}^{NS} N_i (t^{(i)} - \bar{t})^2 / \sum_{i=1}^{NS} N_i \quad (30)$$

$$V_2 = \sum_{i=1}^{NS} N_i \left\{ (t^{(i)} - \bar{t})^2 (t_3^{(i)} - \bar{t}_3)^2 \right\}^{1/2} / \sum_{i=1}^{NS} N_i \quad (31)$$

$$V_3 = \sum_{i=1}^{NS} N_i \left\{ (t_3^{(i)} - \bar{t}_3)^2 (t_4^{(i)} - \bar{t}_4)^2 \right\}^{1/2} / \sum_{i=1}^{NS} N_i \quad (32)$$

$$\bar{t} = \left(\sum_{i=1}^{NS} N_i t^{(i)} \right) / \left(\sum_{i=1}^{NS} N_i \right) \quad (33)$$

که N_i اندازه نمونه در ایستگاه i ، τ_2^i گشتاور خطی نمونه LCV، τ_2^R میانگین نقطه‌ای گشتاور خطی نمونه LCV، μ_V میانگین مقادیر V و σ_V انحراف معیار مقادیر V می‌باشد.

۲-۹-آزمون نکویی برازش

آزمون نکویی برازش تابع توزیع منطقه‌ای به منظور انتخاب بهترین تابع توزیع منطقه‌ای انجام می‌شود (Tallaksen et al., 2004) و شامل محاسبه آماره Z^{DIST} است. تابع توزیع مناسب تابعی است که $|Z^{DIST}| < 1/64$ باشد. این تابع به عنوان تابع منطقه‌ای شناخته می‌شود و به صورت زیر تعریف می‌شود:

$$Z^{DIST} = (\tau_4^{DIST} - \bar{\tau}_4 + \beta_4) / \sigma_4 \quad (34)$$

$$\beta_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (\bar{\tau}_{4m} - \bar{\tau}_4) \quad (35)$$

به منظور مستقل ساختن گشتاورهای خطی از واحد اندازه‌گیری متغیر تصادفی (X) ، آن را اغلب به صورت نسبت گشتاورهای خطی نشان می‌دهند:

$$\tau_r = \frac{\lambda_r}{\lambda_2} \quad r = 3, 4 \quad (24)$$

$$\tau = \frac{\lambda_2}{\lambda_1} \quad (25)$$

که در آن τ_3 پارامتر ضریب چولگی خطی LCS^1 و τ_4 ضریب کشیدگی خطی LCK^1 و τ ضریب تغییرات خطی LCV^2 است.

استفاده از منحنی نسبت گشتاورها (MRD)³ یک روش مقایسه بصری برای تشخیص همگنی منطقه‌ای است (Stedinger et al., 1993) و همواره رسم این نمودار بر آزمون نکویی برازش مقدم است (Vogel and Fennessey, 1993). منحنی نسبت گشتاورها به اندازه کافی برای فهمیدن همگنی منطقه مناسب نیست، به این دلیل Hosking and Wallis (1993) دو آماره را برای آزمون همگنی استنتاج کردند که شامل آماره ناهمگونی (ناجوری) D^4 و آماره ناهمگنی H^5 است.

۲-۸-۲-آزمون گشتاورهای خطی

۲-۸-۱-آزمون ناهمگونی

آزمون ناهمگونی، مکان‌های ناجور را نسبت به کل گروه مشخص می‌کند. اگر مقدار D بزرگ‌تر از ۳ باشد ایستگاه ناجور بوده و از گروه حذف می‌شود. این شاخص به اندازه نمونه وابسته نیست چون این آماره بر اساس گشتاورهای خطی محاسبه می‌شود و نسبت به اندازه نمونه حساسیت کم‌تری دارد. در حالی که گشتاورهای معمولی مانند میانگین و انحراف معیار وابستگی شدیدی به اندازه نمونه، به ویژه داده پرت دارند.

آماره ناجوری به شکل زیر نوشته می‌شود:

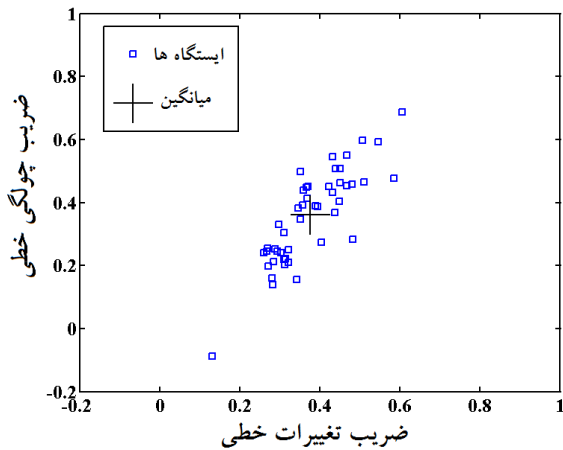
$$D_i = \frac{1}{3} (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \quad (26)$$

$$S = (NS - I)^{-1} \sum_{i=1}^{NS} (u_i - \bar{u})(u_i - \bar{u})^T \quad (27)$$

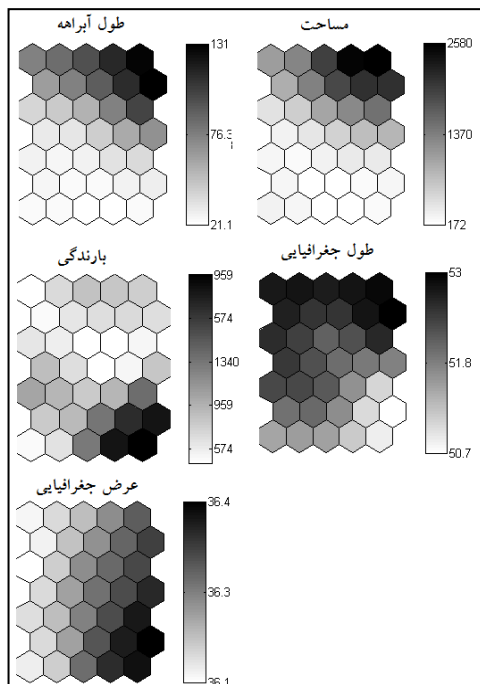
$$\bar{u} = \frac{1}{NS} \sum_{i=1}^{NS} u_i \quad (28)$$

که $u_i = [\tau_1^i, \tau_3^i, \tau_4^i]^T$ ماتریس ترانهاده گشتاورهای خطی در ایستگاه i تعداد ایستگاه و S ماتریس کوواریانس نمونه است.

پارامتر فازی سازی $m=2$ و مقادیر مختلف c بین ۲ تا ۱۰ محاسبه شد (شکل ۶). همانطور که مشاهده می‌شود هر دو شاخص Xie-Beni و Kwon در $c=3$ کمترین مقدار را دارد. بنابراین ایستگاه‌ها روی نقشه SOFM به سه گروه تقسیم شدند (شکل ۷).



شکل ۳- نمودار نسبت گشتاورهای خطی برای ۴۷ ایستگاه آب‌سنجی در استان مازندران



شکل ۴- توزیع هر یک از خصوصیات در نقشه آموزش یافته SOFM

شکل ۷ تعداد ایستگاه اختصاص داده شده به هر واحد SOFM را نشان می‌دهد. با توجه به شکل‌های (۴ و ۷) می‌توان دریافت که مساحت حوضه آبخیز در خوشه‌ی یک بیشترین تاثیر را دارد و تاثیر طول آبراهه اصلی در خوشه سه کمتر از دو خصوصیت مساحت

$$\sigma_4 = \left\{ (N_{sim} - 1)^{-1} \sum_{m=1}^{N_{sim}} (\bar{\tau}_{4m} - \bar{\tau}_4)^2 - N_{sim} \beta_4^2 \right\}^{1/2} \quad (36)$$

در اینجا Z^{DIST} به مفهوم توزیع، τ_4^{DIST} گشتاور خطی برجستگی جامعه (LCK)، $\bar{\tau}_4$ میانگین ناحیه‌ای گشتاور خطی برجستگی نمونه، β_4 مقدار اریبی ناحیه‌ای از گشتاور فوق، σ_4 انحراف معیار ناحیه-ای گشتاور فوق است و N_{sim} تعداد نواحی شبیه سازی شده و برابر ۵۰۰ است (Hosking and Wallis, 1993). لازم به ذکر است که محاسبات خوشه‌بندی توسط نرم‌افزار Matlab و محاسبه گشتاورهای خطی توسط الگوریتم نوشته شده توسط Hosking به زبان Fortran انجام شده است.

۳- نتایج و بحث

ابتدا نمودار گشتاورهای خطی ایستگاه‌های انتخاب شده برای بررسی اولیه منطقه رسم شد (شکل ۳). نمودار ضریب تغییرات خطی نسبت به ضریب چولگی خطی نشان داد که گشتاورهای خطی پراکنده‌اند و استان مازندران نمی‌تواند بعنوان یک منطقه‌ی همگن در نظر گرفته شود.

بنابراین، ۵ خصوصیت طول آبراهه اصلی، مساحت حوضه آبخیز، میانگین بارندگی سالانه، طول جغرافیایی و عرض جغرافیایی پس از نرمالیزه شدن توسط معادله‌ی (۱)، به عنوان ورودی SOFM مورد استفاده قرار گرفت. بواسطه‌ی فرآیند یادگیری SOFM، بردار وزن متناسب با چگالی احتمال داده‌ها تخمین زده شد. شکل ۴ توزیع خصوصیات را در نقشه آموزش یافته SOFM با ۵ خصوصیت را نشان می‌دهد. مقدار هر خصوصیت در واحدهای خروجی SOFM اهمیت آنرا در هر واحد نشان می‌دهد. در شکل ۴ رنگ تیره نشان دهنده‌ی مقایر زیاد هر خصوصیت و رنگ سفید کمترین مقدار را نشان می‌دهد.

بعد از آموزش SOFM، الگوریتم ماتریس U برای خوشه‌بندی واحدهای آموزش یافته بکار رفت. همانطور که در شکل ۵ مشاهده می‌شود ماتریس U مرز مشخصی بین واحدهای نقشه نشان نمی‌دهد. البته بصورت تقریبی ماتریس U ، نقشه خروجی از SOFM را به سه خوشه با مرزهای نامشخص و مبهم تقسیم نموده است (مناطق پررنگ‌تر نشان‌دهنده مرز خوشه‌ها و مناطق روشن‌تر دورن خوشه‌ها هستند). بنابراین، بردار وزن SOFM برای خوشه‌بندی واحدها با الگوریتم خوشه‌بندی c - میانگین فازی استفاده شد. برای مشخص کردن تعداد خوشه‌ی بهینه شاخص‌های ارزیابی خوشه‌بندی فازی Xie-Beni تعمیم یافته (XB, m) و $Kwon (Vk)$ به ازاء

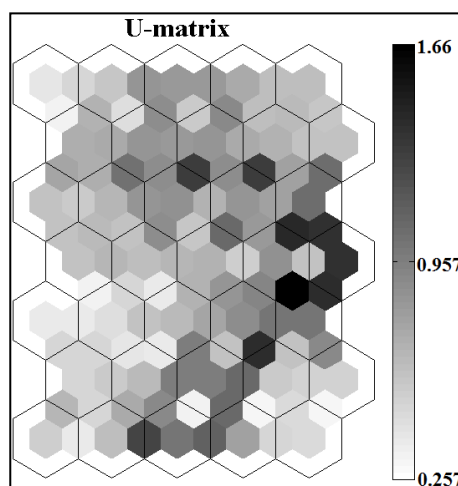
(۱) حذف یک یا چند ایستگاه از مجموعه داده‌ها؛ (۲) انتقال یک یا چند ایستگاه از یک منطقه به منطقه دیگر؛ (۳) تفکیک منطقه به دو یا چند خوشه جدید؛ (۴) اجازه دادن به یک ایستگاه برای مشارکت در دو یا چند منطقه؛ (۵) انحلال مناطق با انتقال ایستگاه‌هایش به دیگر مناطق؛ (۶) الحاق یک منطقه به منطقه یا مناطق دیگر؛ (۷) الحاق دو یا چند منطقه و تعریف دوباره گروه‌ها؛ (۸) کسب داده‌های بیشتر و تعریف دوباره گروه‌ها.

در ابتدا ۵ ایستگاه که توسط آماره ناچوری (D) هاسکینگ و والیس ناهمگون شناخته شدند، حذف گردیدند. اگر چه Hosking and Wallis (1993,1997) مقادیر بحرانی آماره ناهمگونی ($D > 3$) را برای شناسایی ایستگاه‌های غیرهمگن معرفی کردند، اما بهتر است با شناسایی تمام ایستگاه‌ها با مقادیر آماره ناچوری بالا شناسایی شوند. سپس آماره ناهمگونی منطقه‌ای (H) با حذف و یا تعویض ایستگاه‌ها تعدیل شد. در ادامه، ایستگاه‌های ناچور که به میزان معنی داری از کم شدن آماره ناهمگونی منطقه‌ای (H) ممانعت می‌کنند، شناسایی شده و بعد از حصول اطمینان از زیاد نبودن عضویت فازی این ایستگاه‌ها (ایستگاه‌های سفیدچاه، گلورد، داراب‌کلا، پاشاکلا، ره هریجان)، حذف شدند. مقادیر شاخص ناهمگونی قبل و بعد از تعدیل در جدول ۲ ارائه شده است.

جدول ۲- آماره ناهمگونی برای مناطق حاصل از الگوریتم خوشه بندی									
قبل از تعدیل				بعد از تعدیل				مناطق	N
H_1	H_2	H_3	N	H_1	H_2	H_3	N		
۲/۱۱**	۱/۴۰*	۰/۸۷	۱۲	۰/۴۳	۰/۹۲	۰/۷۸	۱۰	۱	
۲/۸۶**	۱/۷۸*	۰/۷۳	۲۵	۱/۱*	۰/۷۱	-۰/۱	۲۱	۲	
۰/۶۱	-۰/۳۹	-۰/۹۳	۱۰	۰/۸۷	-۰/۰۵	-۰/۷۳	۱۱	۳	

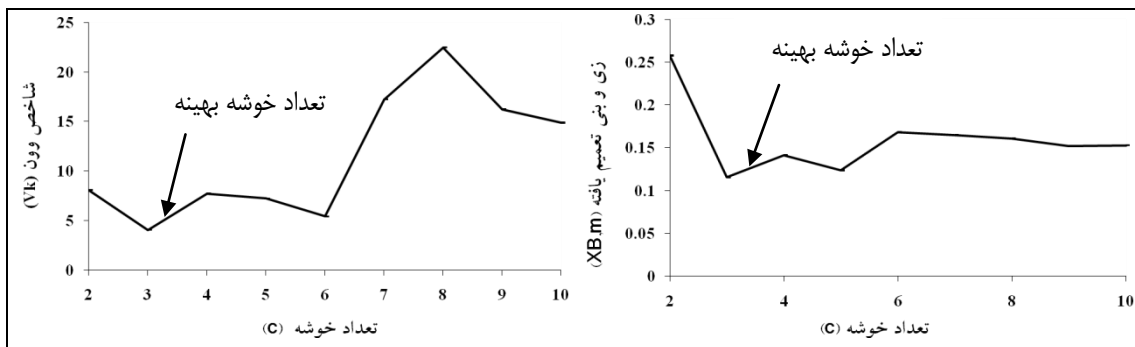
N: تعداد ایستگاه‌ها؛ H_i : شاخص ناهمگونی
 ** ناهمگن؛ * تقریباً همگن

حوضه آبخیز و طول آبراهه اصلی است. در ضمن مقادیر میانگین بارندگی سالانه در خوشه‌ی دو متوسط است. با توجه به نقشه خروجی SOFM (شکل ۴) بنظر می‌رسد عامل تاثیر گذار در شکل‌گیری خوشه‌ی یک مساحت زیاد حوضه‌های شرق استان مازندران و موثرترین عامل در شکل‌گیری خوشه سه میانگین بارندگی بیشتر در غرب استان مازندران است.



شکل ۵- ماتریس فواصل متحد شده (ماتریس U)

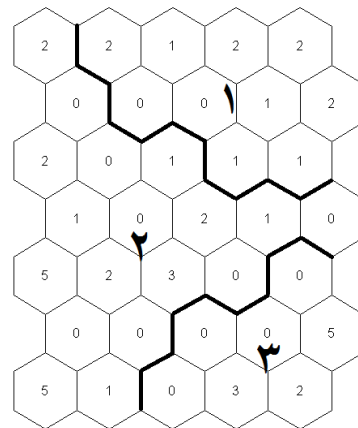
زمانی که مجموعه جامعی از متغیرهای تاثیر گذار در تحلیل فراوانی منطقه‌ای سیلاب وجود ندارد، معمولاً مناطق شکل گرفته برای تحلیل فراوانی منطقه‌ای سیلاب همگن نخواهند بود و نیاز به تعدیل برای بهبود همگنی آن وجود دارد (Hosking and Wallis, 1997). این حقیقت در هیدرولوژی کاملاً شناخته شده است، از این رو هیدرولوژیست‌ها پیشنهاد می‌کنند با تعدیل، این مناطق به گروه‌هایی همگن اصلاح شوند. گزینه‌های پیشنهاد شده توسط Hosking and Wallis (1993) برای اصلاح مناطق شکل گرفته با تحلیل خوشه‌ای عبارتند از:



شکل ۶- نمودار شاخص‌های زی-بنی و وون در مقابل تعداد خوشه‌ها

تعمیم یافته، مقادیر حدی تعمیم یافته، لوگ نرمال، پیرسون نوع ۳ و پارتو تعمیم یافته محاسبه شد (جدول شماره ۳).

مقادیر Z^{DIST} در تمامی مناطق، برای دو توزیع لوجستیک تعمیم یافته و مقادیر حدی تعمیم یافته کمتر از $1/64$ است، اما توزیع لوجستیک تعمیم یافته کمترین مقدار را بین دو توزیع فوق دارد. بر این اساس توزیع لوجستیک تعمیم یافته به عنوان بهترین توزیع برای هر سه منطقه شناخته شد. پارامترهای توزیع برای هر منطقه در جدول شماره ۴ بیان شده است.



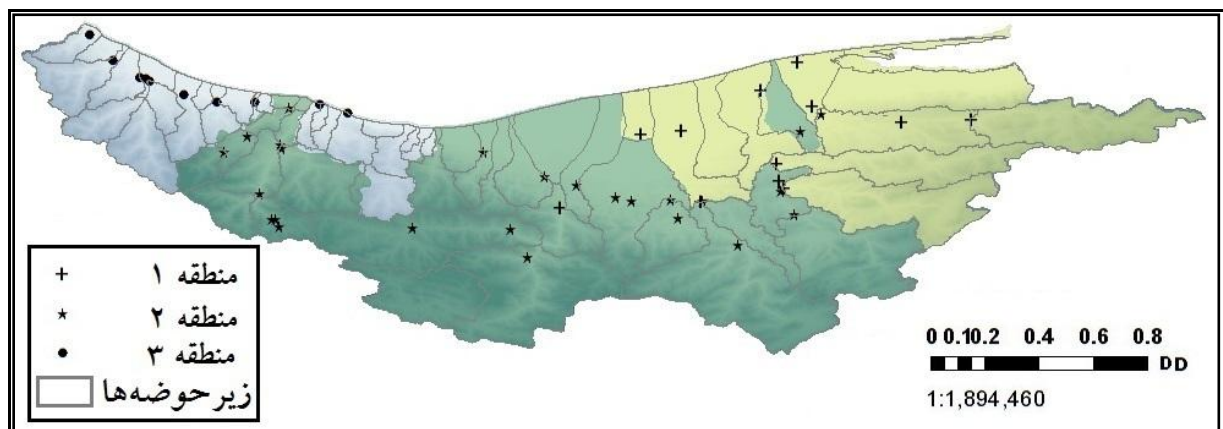
شکل ۷- خوشه‌بندی نقشه خروجی SOFM توسط الگوریتم خوشه‌بندی فازی

مناطق همگن هیدرولوژیک پس از تعدیل در شکل ۸ نشان داده شده است. همانطور که در شکل ۸ دیده می‌شود استان مازندران از غرب به شرق به سه خوشه تقسیم شده است که با روند کاهش بارندگی از غرب به شرق استان نیز مطابقت نشان می‌دهد. با توجه به شکل‌های (۱ و ۸) مشخص است که حوضه‌های واقع در ارتفاعات استان مازندران دارای شباهت هیدرولوژیکی هستند، به نحوی که ارتفاعات در خوشه دو قرار گرفته‌اند و بیشتر مناطق پست در خوشه ۱ واقع شده‌اند. تقسیم حوضه‌های آبخیز استان مازندران به سه منطقه‌ی همگن هیدرولوژیکی توسط روش خوشه‌بندی دو مرحله‌ای SOFM و FCM نتایج (Chavoshi et al. (2012 را تایید می‌کند.

به منظور انتخاب بهترین توزیع منطقه‌ای، آماره Z^{DIST} برای ایستگاه‌های هر منطقه با استفاده از برنامه نوشته شده به زبان فرترن توسط Hosking (1991) برای توزیع‌های ۳ پارامتری لوجستیک

جدول ۳- مقادیر آماره آزمون نکویی برازش برای منطقه مورد مطالعه			
مقادیر آماره Z^{DIST}			توزیع آماری
منطقه ۱	منطقه ۲	منطقه ۳	
* -۰/۱۹	* -۰/۴۷	* -۰/۶۸	لوجستیک تعمیم یافته
* -۰/۸۸	* -۱/۳۱	* -۱/۱۰	مقادیر حدی تعمیم یافته
-۱/۶۶	-۲/۲۵	-۱/۸۵	لوگ نرمال
-۳/۰۱	-۳/۸۵	-۳/۱۵	پیرسون نوع ۳
-۲/۹۳	-۳/۸	-۲/۵۱	پارتو تعمیم یافته
* توزیع‌های آماری قابل قبول به عنوان توزیع منطقه‌ای			

جدول ۴- پارامترهای توزیع لوجستیک تعمیم یافته برای هر منطقه همگن پارامترها			
مکان	مقیاس	شکل	
شرق (خوشه ۱)	۰/۸۱۴	۰/۲۹	-۰/۳۳۹
وسط (خوشه ۲)	۰/۸۲۲	۰/۲۸۱	-۰/۳۳۶
غرب (خوشه ۳)	۰/۷۴۵	۰/۳۲۲	-۰/۳۹۵



شکل ۸- مناطق همگن هیدرولوژیک

۴- نتیجه گیری

الگوریتم دو مرحله‌ای خوشه‌بندی فازی و SOFM برای شناسایی مناطق همگن هیدرولوژیک برای تحلیل فراوانی منطقه‌ای در حوضه‌های آبخیز استان مازندران در شمال ایران به کار رفت. همگنی مناطق بدست آمده از الگوریتم خوشه‌بندی توسط آزمون ناهمگنی براساس گشتاورهای خطی مورد بررسی قرار گرفت. بهترین تعداد خوشه‌ها برای تقسیم استان مازندران به نواحی همگن هیدرولوژیک، $C=3$ بود.

شاخص‌های Xie-Beni و Kwon تعداد خوشه‌ی بهینه را بدرستی برابر $C=3$ بدست آوردند، زیرا نتایج حاصل از سه خوشه بر اساس آزمون همگنی به مناطق همگن هیدرولوژیک نزدیکتر بود.

هرچند ماتریس U یک ابزار بصری مناسب برای برآورد ابتدایی تعداد خوشه‌ها در خروجی SOFM آموزش یافته تلقی می‌شود اما روش چندان مناسبی برای مشخص کردن مرز خوشه‌ها روی نقشه‌ی آموزش یافته نیست و به روش‌های دیگری همچون FCM برای خوشه‌بندی نقشه‌ی SOFM مورد نیاز است.

در پایان براساس آزمون نکویی برازش، توزیع لوجستیک تعمیم یافته با پارامترهای متفاوت بهترین توزیع برای هر سه منطقه شناخته شد.

۵- تشکر

این مقاله برگرفته از پروژه‌ای است که بخشی از هزینه‌های آن توسط شرکت آب منطقه‌ای استان مازندران تامین شده است. بدینوسیله از شرکت فوق‌قدردانی می‌گردد.

پی‌نوشت‌ها

- 1- Quantile
- 2- Self-Organizing Feature Map
- 3- Fuzzy c-mean
- 4- Nodes
- 5- Units
- 6- Map size
- 7- Grey scale
- 8- Membership Matrix
- 9- probability weighted moments
- 10- Linear skewness
- 11- Linear kurtosis
- 12- Linear coefficient of variation
- 13- Moment Ratio Diagram
- 14- Discordancy
- 15- Heterogeneity

۶- مراجع

فرسادنیا ف، رستمی کامرود م، مقدم‌نیا ع (۱۳۹۱) تحلیل روند بارندگی در استان مازندران با استفاده از روش من-کندال منطقه‌ای. تحقیقات منابع آب ایران، سال ۸، شماره ۲: ۶۰-۷۰. مطالعات جامع مهندسی رودخانه‌های استان مازندران، وزارت نیرو، شرکت سهامی آب منطقه‌ای استان مازندران، (۱۳۸۷).

Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York.

Burn DH and Goel NK (2000) The formation of groups for regional flood frequency analysis. Hydrological Sciences Journal 45(1): 97-112.

Chavoshi S, Azmin Sulaiman WN, Saghafian B, Sulaiman MD NB, 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.

Greenwood JA, Landwehr JM, Matalas NC, Wallis JR (1979) Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research. 15: 1049-1054.

Hall MJ, Minns AW (1999) The classification of hydrologically homogeneous regions. Hydrological Sciences Journal. 44 (5): 693-704.

Hathaway R J, Bezdek J C (2001) Fuzzy c-means clustering of incomplete data, IEEE Trans. Syst. Man Cybern. B. 31: 735- 744.

Haykin S (2003) Neural networks: A comprehensive foundation. Fourth Indian Reprint, Pearson Education, Singapore, p. 842.

Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. Water Resources Research. 29: 271-281.

Hosking JRM, Wallis JR (1996) Regional frequency analysis: an approach based on Lmoments. Cambridge University Press: Cambridge.

Hosking JRM (1986) The theory of probability weighted moments. Res. Rep. RC 12210, IBM Research Division, Yorktown Heights, NY.

Hosking JRM (1991) Fortran routines for use with the method of L-moments, Version 2, Res. Rep. RC 17097, IBM Research Division, York Town Heights, NY 10598.

- Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993) Frequency analysis of extreme Event, Handbook of Hydrology. McGraw-Hill: New York.
- Tallaksen LM, Madsen H, Hisdal H (2004) Frequency analysis, hydrological drought – Processes and Estimation Methods for Stream flow and Groundwater, Developments in Water Sciences 48. Elsevier Science Publisher: The Netherlands.
- Ultsch A, Siemon HP (1990) Kohonen's self organizing feature maps for exploratory data analysis. In: Proceedings of INNC'90, International Neural Network Conference. Kluwer Academic Publishers, Dordrecht, Netherlands: 305-308.
- Ultsch A (1993) Self-organizing neural networks for visualization and classification. In: Opitz, O., Lausen, B., Klar, R. (Eds.), Information and Classification. Springer, Berlin: 307-313.
- Vesanto J, Alhoniemi R (2000) Clustering of the self organizing map. IEEE Trans. Neural. Netw. 11 (3): 586-600.
- 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.
- Vogel RM, Fennessey NM (1993) L-moment diagram should replace product moment diagram. Water Resources Research. 29: 1745–1752.
- Wilppu R (1997) The visualisation capability of self organizing maps to detect of deviation in distribution control. TUCS Technical Report No. 153. Turku Centre for Computer Science, Finland.
- Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence 13 (8): 841–847.
- 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 (2001) Self-organizing maps. Springer, Berlin, Germany.
- Kwon SH (1998) Cluster validity index for fuzzy clustering. Electronics Letters 34 (22): 2176–2177.
- Lampinen J, Oja E (1992) Clustering properties of hierarchical self-organizing maps. Journal of Mathematical Imaging and Vision 2 (2–3): 261–272.
- Ley R, Casper MC, Hellebrand H, Merz R (2011) Catchment classification by runoff behaviour 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.
- Rao AR, Srinivas VV (2006) Regionalization of watersheds by fuzzy cluster analysis. Journal of Hydrology 318: 57-79.
- Ross TJ (1995) Fuzzy logic with engineering applications. McGraw-Hill, New York.
- Srinivas VV, Tripathi S, Rao AR, Govindaraju RS (2007) Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering Journal of Hydrology 348: 148– 166.