

طبقه‌بندی تبخیر سالانه ایستگاه‌های تبخیرسنجی ایران با استفاده از محاسبات نرم (خوشه‌بندی فازی و شبکه‌عصبی کوهنن) بر اساس پارامترهای اقلیمی

رضا مکاریان^۱

حسین صدقی^۲

محمدعلی قربانی^۳

حسین بابازاده^۴

چکیده

تبخیر را می‌توان نقطه آغازین چرخه هیدرولوژیک آب به شمار آورد، که برآیند مجموعه عوامل اقلیمی و جغرافیایی منطقه هیدرولوژیک می‌باشد و مستقیماً بر منابع آب اثر بخش است. این پدیده از یک سیستم پیچیده و غیرخطی پیروی می‌کند، که تخمین دقیق آن بسیار دشوار است. از این رو استفاده از مدل‌های ریاضی مانند سیستم استنتاج فازی و شبکه عصبی کوهنن با درک رفتارهای غیر خطی سیستم برای حل این مشکل مناسب است. طبقه‌بندی پایگاه اطلاعات بزرگ نظیر ایستگاه‌های تبخیرسنجی موجب می‌گردد حجم زیادی از اطلاعات با اختصاص به چند دسته متجانس کوچکتر براحتی در روش‌های مختلف مدل‌سازی مورد استفاده قرار گیرد. خوشه‌بندی در این پژوهش با استفاده از داده‌های اقلیمی منجر به قرار گرفتن ایستگاه‌های تبخیرسنجی در ۷ خوشه گردیده‌است و بین مقادیر حداکثر شاخص RS و حداقل واریانس محاسباتی خوشه‌ها همخوانی

۱- مدرس دانشگاه

Email: mokarian4561@yahoo.com

۲- استاد گروه مهندسی آب- واحد علوم و تحقیقات تهران

۳- دانشیار گروه مهندسی آب دانشگاه تبریز

۴- دانشیار گروه مهندسی آب- واحد علوم و تحقیقات تهران

وجود دارد، بطوریکه نسبت به ضریب تعیین RS و واریانس خوشه‌ها روش شبکه عصبی کوهن نسبت به روش فازی نتایج بهتری را نشان می‌دهد.

واژگان کلیدی: تبخیر، خوشه‌بندی فازی، شبکه عصبی کوهن، ایستگاه‌های تبخیرسنجی، ایران

مقدمه

براساس آمارهای منتشر شده حدود سه چهارم نزولات جوی به صورت تبخیر از دسترس خارج می‌گردد. تبخیر نقش اساسی در بهره برداری مخازن، طراحی سیستم‌های آبیاری و زهکشی و مدیریت منابع آب ایفا می‌کند. این پدیده تحت تأثیر پارامترهایی چون تابش خورشید، دما، رطوبت و سرعت باد بوده، که پارامترهای اخیر نیز بر یکدیگر تأثیر متقابل می‌گذارند. از این رو تبخیر از یک سیستم پیچیده غیرخطی تبعیت می‌کند، که تخمین آن بسیار دشوار است (رحیمی خوب، ۲۰۰۶: ۷). روش‌های تحلیلی به داده‌های هواشناسی زیادی نیاز دارند و بعضی از این داده‌ها مثل تابش خورشید در همه جا و همیشه اندازه‌گیری نمی‌شود (لیناگر، ۱۹۹۴: ۴۳۲).

در تحقیق حاضر به خوشه‌بندی ایستگاه‌های تبخیرسنجی در کل کشور پرداخته خواهد شد. خوشه به مجموعه‌ای از داده‌ها گفته می‌شود، که به هم شباهت داشته باشند. در خوشه‌بندی سعی می‌شود، تا داده‌ها به خوشه‌هایی تقسیم شوند، که شباهت بین داده‌های درون هر خوشه حداکثر و شباهت بین داده‌های درون خوشه‌های متفاوت، حداقل گردد. بیابانکی و اسلامیان (۱۳۸۳: ۱۴)، از روش تحلیل خوشه‌ای درحوضه آبریز کرخه برای تعیین مناطق همگن براساس ۱۶ خصوصیت هیدرولوژیک استفاده کردند، سپس از روش تحلیل ممیزی صحت گروه‌بندی مورد آزمون قرار گرفت و در نتیجه منطقه به دو گروه همگن تقسیم بندی شد. رحیمی و همکاران (۱۳۹۰: ۱۶)، در پژوهشی به منظور شناخت پهنه‌های دمایی استان چهار محال و بختیاری، روش مجموعه‌ها و خوشه‌بندی فازی را مورد بررسی قرار دادند و منطقه مورد مطالعه را در سه گروه دمایی شامل ناحیه سرد، ناحیه بسیار سرد و ناحیه

معتدل و نیمه گرم طبقه‌بندی نمودند. مقدم‌نیا و همکاران (۲۰۰۹: ۸)، دو مدل شبکه عصبی و سیستم استنتاج فازی - عصبی تطبیقی را برای تعیین تبخیر روزانه از تشت با استفاده از داده‌های ایستگاه هواشناسی نزدیک مخزن چاه نیمه در سیستان شامل دمای هوا، سرعت باد، کمبود فشار بخار اشباع و رطوبت نسبی مورد مقایسه قرار دادند. فرسانیا و همکاران (۱۳۹۱: ۱۲)، 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. حسنعلی زاده و همکاران (۱۳۹۳: ۱۱)، با استفاده از داده‌های ۲۹ ایستگاه بارانسنجی و تبخیرسنجی استان گلستان و به روش خوشه‌بندی سلسله مراتبی وارد به تحلیل نواحی همگن بارش سالانه پرداخته و با روش گشتاورهای خطی همگنی خوشه‌ها را بررسی کردند. نتایج این پژوهش نشان داد، که می‌توان زیر حوضه‌های استان گلستان را به دو ناحیه همگن تقسیم کرد. جهانبخش اصل و همکاران (۱۳۹۴: ۲۳)، در پژوهشی به منظور بررسی توزیع زمانی و مکانی بارش در سطح شهرستان تبریز با استفاده از آمار روزانه ۱۱ ایستگاه تبریز و به روش تحلیل خوشه‌ای سلسله مراتبی ادغام وارد بر روی ماتریس داده‌ها، شهرستان تبریز را از لحاظ بارش به سه گروه مجزای بارش کم، متوسط و زیاد تفکیک کردند. فلاح قاهره و همکاران (۱۳۹۴: ۱۸) ناحیه بندی آب و هوایی استان گیلان را با روش‌های چند متغیره و بهره‌گیری از تحلیل خوشه‌ای روی دو عامل اقلیمی وجود ۳ ناحیه اقلیمی ناحیه معتدل و مرطوب، ناحیه کوهستانی و ناحیه نیمه مرطوب و سرد را در استان گیلان معین کردند. عساکره و دوستکامیان (۱۳۹۵: ۲۳) با بکارگیری داده‌های حاصل میانگین و ضریب تغییرات ماهانه آب قابل بارش به روش تجزیه خوشه‌ای ادغام وارد مورد بررسی و تجزیه و تحلیل قرار گرفت و سه پهنه اقلیمی، ناحیه با آب قابل بارش زیاد، آب قابل بارش متوسط و آب قابل بارش کم مشخص گردید. سی و اوپن شاو (۱۹۹۸: ۴)، به منظور افزایش دقت تخمین سیلاب در مناطق همگن انگلستان از روش خوشه‌بندی فازی استفاده نمودند، که نتایج حاکی از دقت خوب این روش در مقایسه با روش‌های دیگر همچون خوشه‌بندی طبقاتی بوده‌است. هولوا و داتر (۱۹۹۹: ۱۴)، از خوشه‌بندی فازی در طبقه‌بندی مکانی-زمانی بارش‌های روزانه بیش از ۴۰۰ ایستگاه در یک دوره ۲۰ ساله در استرالیا را مورد تجزیه و تحلیل قرار دادند، که نهایتاً منطقه مورد

مطالعه بصورت دو، سه و چهار خوشه‌ای پهنه‌بندی گردید. روگر و همکاران (۲۰۰۰: ۱۷)، برای تهیه نقشه مشخصات هیدرولوژیکی خاک در کشور فرانسه از روش خوشه‌بندی فازی استفاده نمودند. نتایج نشان داد، که روش متکی بر منطق فازی به دلیل اجرای خوب، هزینه پایینتر و عملکرد بهتر روش مناستری می‌باشد. جینگ یی و هل (۲۰۰۴: ۲۰)، طبقه‌بندی ۸۶ ایستگاه در حوضه گان از استان جیانگ سی و حوضه مینگ از استان فوجیان در جنوب شرقی هند را با استفاده از روش کلاستر وارد، روش فازی K-means و شبکه عصبی کوهن انجام دادند. نتایج حاصل نشان داد، که شبکه عصبی کوهن یک روش ممتاز در مقایسه با سایر روش‌هاست. لی و چن (۲۰۰۶: ۹)، از الگوریتم شبکه عصبی خود سامان برای خوشه‌بندی ۱۵۴ ایستگاه باران‌سنجی در تایوان به منظور تحلیل فراوانی 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، ۱۲*۱۲ انتخاب کردند و در نهایت منطقه مورد بررسی را به ۸ خوشه همگن تقسیم کردند. راجو و کومار (۲۰۰۷: ۱۲)، طبقه‌بندی ایستگاه‌های هواشناسی هند را با استفاده از خوشه‌بندی وارد، خوشه‌بندی فازی و شبکه عصبی مصنوعی کوهن مورد بررسی قرار دادند، که نهایتاً خوشه‌بندی فازی بهتر از دو روش دیگر نتیجه داد. پارامترهای مورد استفاده در این تحقیق عرض جغرافیایی، طول جغرافیایی، ارتفاع، دمای میانگین، رطوبت، سرعت باد، ساعات آفتابی و تابش خورشیدی بوده‌است. دیکباس و همکاران (۲۰۱۱: ۹)، خوشه‌بندی فازی را در طبقه‌بندی سری‌های سالانه بارش ۱۸۸ ایستگاه در ترکیه مورد تجزیه و تحلیل قرار دادند. پارامترهای مورد استفاده در این مطالعه طول جغرافیایی، عرض جغرافیایی، ارتفاع ایستگاه و بارش گزارش شده‌است. در این مطالعه کشور ترکیه به شش گروه همگن طبقه‌بندی گردیده‌است. لی و همکاران (۲۰۱۱: ۱۷)، از دو نوع ورودی الگوریتم شبکه عصبی خود سامان برای مقایسه شباهت‌های هیدرولوژیکی ۵۳ حوضه آبخیز در آلمان استفاده کردند. آنها گروه حوضه‌هایی که با رفتار پاسخ هیدرولوژیکی خوشه‌بندی شده‌بودند را با خوشه‌هایی که بر اساس 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 بهبود یابد.

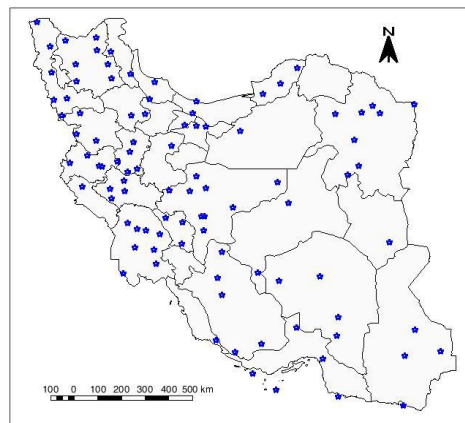
چاوشی و همکاران (۲۰۱۲: ۱۴)، به منظور مقایسه روش‌های سخت (سلسله مراتبی و k- میانگین) با روش‌های نرم (C- میانگین و کوهنن) با استفاده از داده‌های ۶۱ حوضه واقع در نوار ساحلی دریای خزر پرداختند. آنها برای آماده‌سازی داده‌های ورودی خوشه‌بندی از تحلیل مؤلفه اصلی استفاده کردند. مقایسه نتایج با استفاده از آزمون همگنی گشتاورهای خطی نشان داد، که خوشه‌های بدست آمده با روش نرم همگن تر هستند. رضوی و کولیایی (۲۰۱۳: ۱۳)، با استفاده از پنج مشخصه جریان آبراهه‌ای به شناسایی مناطق همگن حوضه‌های آبخیز استان انتاریو کانادا نمودند و ضمن بهره‌گیری از شاخص دیویس - بولدین و کاربرد روش‌های نگاشت ویژگی خود سامان و آنالیز غیرخطی مولفه‌های اصلی منطقه مطالعه را به چهار خوشه همگن تقسیم کردند. فرسادنیا و همکاران (۲۰۱۴: ۱۰)، از الگوریتم دومرحله‌ای خوشه‌بندی فازی و نگاشت ویژگی خود سامان برای شناسایی مناطق همگن هیدرولوژیک در حوضه‌های آبخیز استان مازندران استفاده کردند و استان مازندران را به سه ناحیه همگن هیدرولوژیک تفکیک کردند. با توجه به نیاز آبی کشور به منابع آبی و ضرورت مدیریت کارآمد مسایل مرتبط با منابع آب و هوا و اقلیم و پردازش حجم بالای اطلاعات موجود با اختصاص به دسته‌های کوچکتر، گسترش معلومات نقطه‌ای به اطلاعات منطقه‌ای برای مناطق فاقد آمار بیش از پیش احساس می‌شود. هدف از تحقیق حاضر طبقه‌بندی ایستگاه‌های تبخیرسنجی کشور و تعیین تعداد بهینه گروه‌های حاضر در طبقه‌بندی است، تا چشم انداز مناسبی از اتفاقات در حال وقوع پایگاه داده‌ها به مصرف کننده نهایی اطلاعات ارایه شود و با مقایسه نتایج حاصل از روش‌های خوشه‌بندی، بهترین روش انتخاب گردد.

مواد و روش‌ها

منطقه مورد مطالعه

شامل کشور ایران با وسعت ۱،۶۴۰،۱۹۵ کیلومتر مربع در نیمه جنوبی منطقه معتدل شمالی بین ۰۳'، ۲۵° و ۴۷'، ۳۹° عرض شمالی از خط استوا و ۱۴'، ۴۴°، ۶۳° طول شرقی از نصف النهار گرینویچ قرار دارد. در این پژوهش سعی شده‌است که ایستگاه‌های انتخابی تا حد امکان از نظر اقلیمی و آب و هوایی با یکدیگر متفاوت باشند، از این رو ۱۲۰

ایستگاه تبخیرسنجی از استان‌های مختلف کشور به نحوی که حداقل از هر استان ۲ ایستگاه در تحقیق حاضر باشند استفاده شده‌است. به ترتیب با شرایط آب و هوایی کوهستانی، خشک و مرطوب انتخاب گردید. در جدول ۱ اطلاعات مربوط به ایستگاه‌های تبخیرسنجی آورده شده‌است. از بین ۱۲۰ ایستگاه تبخیرسنجی تعداد ۲۳ ایستگاه تبخیرسنجی به دلیل کمبود آماری حذف و تعداد ۹۷ ایستگاه انتخاب گردید (شکل ۱).



شکل (۱) نمایش ایستگاه‌های تبخیرسنجی سازمان هواشناسی ایران

در بین ۹۷ ایستگاه تبخیرسنجی حداکثر تبخیر مربوط به ایستگاه ایرانشهر با مقدار $4281/81$ میلی متر در سال، که در استان سیستان و بلوچستان قرار دارد و حداقل آن مربوط به ایستگاه رشت با مقدار $858/082$ میلی متر در سال و همچنین میانگین حسابی سالانه میزان تبخیر ۹۷ ایستگاه تبخیرسنجی کشور $2260/46$ میلی متر است.

جدول (۱) اطلاعات مربوط به ایستگاه‌های تخییرسنجی مورد استفاده در تحقیق، داده‌های اقلیمی

ردیف	نام ایستگاه	سالانه (ن)	میانگین دمای باد (تات)	سرعت (درصد)	رطوبت نسبی (درصد)	ساعات آفتابی	تخییر سالانه (میلی متر)
۱	آبادان	۲۵/۵	۶/۴	۴۵	۳۰۵۳/۲	۳۹۸۰۰۴۲۶	
۲	آباده	۱۴/۴	۴/۶	۳۴	۳۳۵۸/۸	۲۲۲۶۰۰۶۶	
۳	آب علی	۸/۶	۴/۳	۴۸	۲۸۷۴	۱۳۶۷۰۶۷	
...	
...	
...	
۹۷	سیاه بیشه	۱۰/۷	۴/۵	۶۲	۱۹۱۸/۷	۸۲۶۰۲۸	

الگوریتم خوشه‌بندی فازی

خوشه‌بندی فازی مبتنی بر نظریه مجموعه‌های فازی چند ارزشی بوده و اساسا به منظور اقدام در شرایط ابهام ارائه گردیده‌است. در این نظریه، هر فرد هم زمان در کلاس‌های مختلف ولی با درجات متفاوت عضویت دارد. در نتیجه کلاس‌ها همپوشی داشته و پیوسته هستند. در این روش که مبتنی بر منطق فازی می‌باشد حل مسئله در سه مرحله فازی سازی، سیستم استنتاج فازی و غیرفازی سازی صورت می‌گیرد (شکاری و باقرنژاد ۱۳۸۴: ۱۴). کاربردی ترین الگوریتم خوشه‌بندی فازی، الگوریتم C-means می‌باشد. در این الگوریتم برای خوشه‌بندی n داده در C دسته یک تابع هدف J_m به صورت رابطه زیر تعریف می‌شود:

$$J_m(u, v) = \sum_{k=1}^n (\mu_{ik})^{m'} (d_{ik})^2 \quad (1)$$

که d_{ik} فاصله اقلیدسی بین مرکز خوشه i ام و داده k ام بوده و به شکل زیر تعریف می‌شود:

$$d_{ik} = d(x_i - v_i) = [\sum_{j=1}^m (x_{kj} - v_{im})^2]^{1/2} \quad (2)$$

d_{ik} برابر درجه عضویت داده k ام به خوشه i ام است. کمترین مقدار j_m مربوط به بهترین حالت خوشه‌بندی خواهد بود. پارامتر m' یک پارامتر وزنی است، که بازه تغییرات آن بصورت $m' \in [1, \alpha]$ بوده، این پارامتر میزان فازی بودن را در فرایند دسته‌بندی مشخص می‌نماید. همچنین v_i نشان دهنده مختصات مرکز i امین خوشه می‌باشد، بنابراین $v_i = \{v_{i1}, \dots, v_{im}\}$ که در آن m تعداد ابعاد v_i یا عبارتی تعداد معیارهای تشابه می‌باشد. مختصات مرکز خوشه‌ها از رابطه زیر بدست می‌آید:

$$u_i^j = \frac{\sum_{k=1}^n \mu_{ik}^{m'} x_k^j}{\sum_{k=1}^n \mu_{ik}^{m'}} \quad (3)$$

به طوری که $j=1, 2, \dots, m$ متغیری برای نشان دادن فضای معیارهاست (کوره پزان دزفولی، ۱۳۸۷: ۲۶۲).

الگوریتم شبکه عصبی کوهنن

شبکه خود سازمانده کوهنن یک شبکه عصبی بدون سرپرست می‌باشد، که هدفش کاهش بعد و خوشه‌بندی است. کوهنن به علت استفاده از تابع همسایگی به منظور حفظ ویژگی‌های مکانی فضای ورودی با سایر شبکه‌های عصبی متفاوت است. هر شبکه کوهنن از تعدادی نود (نرون) تشکیل شده‌است. هر نود دارای برداری از وزن‌ها می‌باشد. ابعاد این بردار با ابعاد فضای ورودی برابر است. پس از آموزش شبکه هر ناحیه از نودهای شبکه به الگوهای خاصی از داده‌های ورودی واکنش نشان می‌دهند. روش آموزش شبکه یادگیری رقابتی است. وقتی نمونه آموزشی جدید به شبکه اعمال می‌شود، فاصله اقلیدسی آن از بردار وزن تمام نودهای شبکه حساب می‌شود. نرونی که بردار وزنش بیشترین شباهت را با بردار ورودی داشته باشد برنده‌است. وزن‌های هر نرون مطابق فرمول زیر به روز رسانی می‌گردند.

$$W_i(t+1) = W_i(t) + \theta(v, t)\alpha(t)(X(t) - W_i(t)) \quad (4)$$

که $\alpha(t)$ نرخ یادگیری کاهشی و $X(t)$ بردار ورودی می‌باشد. تابع همسایگی بستگی $\theta(v, t)$ وابسته به فاصله بین نرون i و BMU در شبکه است. همچنین شعاع همسایگی و نرخ یادگیری در طول زمان توسط فرمول‌های زیر کاهش می‌یابد.

$$\alpha(t) = \alpha_0 \exp\left(-\frac{t}{\lambda}\right), t=1,2,3,\dots \quad (5)$$

که t شماره تکرار و λ مقدار ثابت و α_0 نرخ یادگیری اولیه می‌باشد.

$$\Theta(t) = \exp\left(-\frac{\text{dist}^2}{2\sigma(t)^2}\right) \quad (6)$$

که dist فاصله نود از BMU و $\sigma(t)$ با استفاده از فرمول زیر محاسبه می‌شود.

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\lambda}\right) \quad t=(1,2,3,\dots) \quad (7)$$

در این فرمول σ_0 عرض شبکه در زمان t_0 و λ مقدار ثابت و t شماره تکرار جاریست (سلطانی و همکاران، ۱۳۸۹: ۷).

شاخص اعتبارسنجی (RS (R Square)

این شاخص ضریب تعیین و یا ضریب تشخیص اطلاق می‌گردد. دامنه مقادیر این نسبت بین ۰ و ۱ است. هر چقدر این نسبت به یک نزدیکتر باشد، بیانگر آن است که پراکندگی داده‌ها داخل یک خوشه کم و فاصله بین خوشه‌ها زیاد است RS از روابط زیر بدست می‌آید:

$$RS = \frac{SS_b}{SS_t} = \frac{SS_t - SS_w}{SS_t} \quad (8)$$

$$SS_w = \sum_{i=1}^k \sum_{x \in G_j} \sum_{j=1}^d (XX_{ij} - \bar{X}_{ij})^2 \quad (9)$$

$$SS_t = \sum_{i=1}^n \sum_{j=1}^d (XX_{ij} - \bar{X}_{ij})^2 \quad (10)$$

که در روابط فوق G_j بیانگر خوشه j ، d تعداد متغیرها و n نشان دهنده تعداد داده‌ها است و همچنین X_{ij} مقدار شی i از نظر شاخص j ، \bar{X}_j میانگین داده‌های شاخص j و \bar{X}_{ij} میانگین شاخص j در خوشه i است.

واریانس خوشه‌ها

این شاخص قابلیت ارزیابی نتایج سایر تکنیک‌های خوشه‌بندی را نیز دارا می‌باشند. در شاخص اعتبارسنجی $RMSSDT$ از واریانس خوشه‌ها استفاده می‌شود، که به شکل رسمی می‌توان از رابطه زیر برای محاسبه آن استفاده کرد.

(۱۱)

$$RMSSDT = \sqrt{\frac{\sum_{j=1..d} \sum_{i=1..n_j} (x_k - \bar{x}_j)^2}{\sum_{j=1..d} (n_{ij} - 1)}}$$

با توجه به رابطه بالا و این که این معیار میزان همگنی خوشه‌ها را اندازه می‌گیرد، می‌توان دریافت که هرچه مقدار آن کمتر باشد نشان دهنده خوشه‌بندی بهتر داده‌ها است (مومنی، ۱۳۹۰: ۲۹۶).

یافته‌ها و بحث

نرمال سازی داده‌ها و نرم افزار خوشه‌بندی

مجموعه داده‌ها برای هر ۵ پارامتر (تبخیر، دما، سرعت باد، رطوبت نسبی و تعداد ساعات آفتابی) برای ۹۷ ایستگاه تبخیرسنجی نرمال شده است، تا داده‌ها را بی بعد سازد. نرمال سازی معیار j برای ایستگاه تبخیرسنجی i با رابطه زیر تعریف شده است که X_{ij} ، j امین معیار برای ایستگاه تبخیرسنجی i ام و X_{jmax} و X_{jmin} مقادیر بیشترین و کمترین معیار j

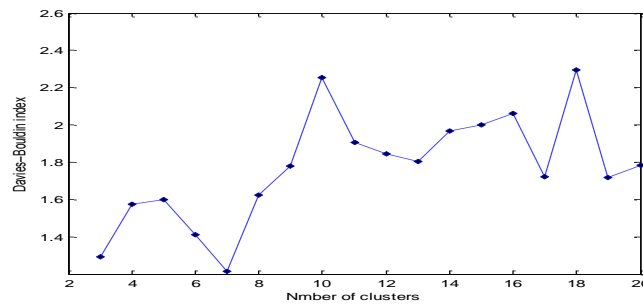
ام در میان ۹۷ ایستگاه تبحیرسنجی هستند. مقادیر نرمال شده بدین صورت بدست آورده شده‌اند و سپس برای طبقه‌بندی ایستگاه‌های تبحیرسنجی استفاده شده‌اند.

$$y_{ij} = \frac{x_{ij} - x_{jmin}}{x_{jmax} - x_{jmin}} \quad (12)$$

الگوریتم خوشه‌بندی فازی در محیط متلب با استفاده از تابع fcm برنامه شده است. ورودی در این سیستم یک فایل اکسل شامل دو صفحه از داده‌ها بوده‌است. صفحه اول شامل تعداد ایستگاه‌های تبحیرسنجی و ضوابط، صفحه دوم شامل داده‌های نرمال شده متناظر با همه ۵ پارامتر برای ۹۷ ایستگاه تبحیرسنجی و ورودی‌های دیگر شامل تعداد بیشترین و کمترین خوشه‌ها برای پیدا کردن تعداد بهینه تکرارها و مقادیر خطا است. داده‌های خروجی در فایل‌های متفاوتی نگهداری می‌شوند. بر مبنای تعداد اعضاء خوشه انتخاب شده، که پیش از این شامل اطلاعات ورودی مانند مقادیر نرمال شده ۵ پارامتر برای هر ایستگاه تبحیرسنجی بود، نمره عضویت هر ایستگاه تبحیرسنجی و ایستگاه تبحیرسنجی نماینده برای هر گروه درج می‌شود.

شاخص دیویس- بولدین^۶ به عنوان مبنایی جهت یافتن مقدار بهینه خوشه‌ها برای گروه‌بندی ایستگاه‌های مورد مطالعه مورد استفاده قرار می‌گیرد (رابطه ۱۳). در این شاخص، $\delta(X_i, X_j)$ مقدار فاصله درون خوشه‌ای خوشه‌های $X_i, X_j, \Delta(X_i)$ قطر خوشه X_i و C تعداد خوشه‌های قسمت‌های U می‌باشد (راجو و کمار، ۲۰۰۷: ۱۱).

$$DB(U) = \frac{1}{c} \sum_{i=1}^c \max_{i \neq j} \left[\frac{\Delta(X_i) + \Delta(X_j)}{\delta(X_i, X_j)} \right] \quad (13)$$



شکل (۲) مقادیر شاخص دیویس-بولدین برای تعداد خوشه‌های مختلف بر اساس پارامترهای اقلیمی

مقادیر شاخص اعتبار D-B برای خوشه‌های ۳ تا ۲۰ محاسبه می‌شود و در شکل ۲ کمترین مقدار شاخص D-B مطلوب است. می‌توان از شکل ۲ مشاهده کرد، که طبقات اقلیمی شامل ۷ خوشه است. شاخص D-B، ۱/۲۱۲ است که برای خوشه ۷ اتفاق افتاده است. بنابراین اینطور نتیجه شده است، که تعداد بهینه خوشه‌ها از نظر تخبیر براساس پارامترهای اقلیمی ۷ خوشه است. پس از تعیین تعداد بهینه خوشه‌ها به منظور بررسی معیار شباهت اعضای واقع در هر گروه از مجموع مربعات خطا که در رابطه ۱۴ نشان داده شده استفاده گردید. در واقع منظور، تعیین مجموع فاصله (فاصله اقلیدسی) نقاط از مرکز دسته‌های ایجاد شده می‌باشد (دیکباش، ۲۰۱۱: ۹).

$$J(U, V; X) = \sum_{i=1}^c \sum_{k=1}^N (U_{ik})^m d_{ik}^2(x_k, v_i) \quad (14)$$

در این رابطه، N تعداد ایستگاه‌ها C تعداد دسته‌ها، U_{ik} نشان دهنده عدد عضویت ایستگاه k ام در دسته i ام و m یک عدد ثابت می‌باشد که در اینجا برابر ۲ در نظر گرفته شده است. X نشان دهنده نقاط یا ایستگاه و V مرکز دسته‌ها می‌باشد. مربع فاصله اقلیدوسی نیز با استفاده از رابطه ۱۵ بدست آمده است:

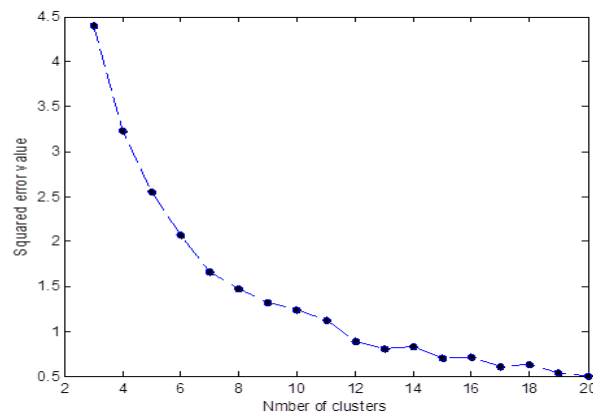
(۱۵)

$$d_{ik}^2(x_k, v_i) = (x_{k,1} - v_{i,1})^2 + (x_{k,2} - v_{i,2})^2 + (x_{k,3} - v_{i,3})^2 + \dots + (x_{k,n} - v_{i,n})^2$$

مقادیر مربعات خطای هر کدام از گروه‌ها بین ۳ تا ۲۰ خوشه در شکل ۳ نشان داده شده است. از این نمودار مشاهده می‌شود، که مقادیر مربعات خطا از ۴/۳۹۷۲ تا ۰/۵۱ کاهش پیدا کرده‌است.

مقدار عضویت در هر گروه، احتمال قرار گرفتن ایستگاه تخبیرسنجی گروه‌بندی شده در آن گروه مخصوص را نشان می‌دهد. خلاصه مقادیر عضویت هر کدام از ۹۷ ایستگاه تخبیرسنجی در هر کدام از ۷ گروه در جدول ۲ نشان داده شده است.

تعداد ایستگاه‌های تخبیرسنجی قرار گرفته در خوشه‌های ۱ تا ۷ به ترتیب ۶ ایستگاه در خوشه اول، ۱۴ ایستگاه در خوشه دوم، ۱۷ ایستگاه در خوشه سوم، ۲۰ ایستگاه در خوشه چهارم و ۱۹ ایستگاه در خوشه پنجم، ۶ ایستگاه در خوشه ششم، ۱۶ ایستگاه در خوشه هفتم می‌باشد. در جدول ۳ اسامی ایستگاه‌های متعلق به گروه‌های ۷ گانه با روش خوشه‌بندی فازی آورده شده است.



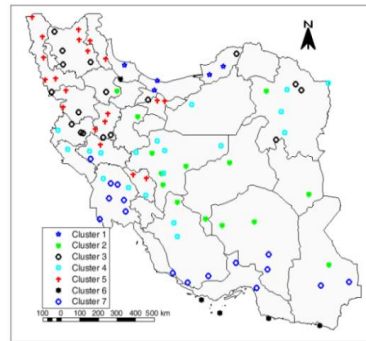
شکل (۳) مقادیر خطای مربع شده برای تعداد خوشه‌های مختلف بر اساس پارامترهای اقلیمی

جدول (۲) درجه عضویت تعلق گرفته به ایستگاه‌های تبخیرسنجی در روش خوشه‌بندی فازی با استفاده از داده‌های اقلیمی

ایستگاه‌های تبخیرسنجی	شماره خوشه‌ها							خوشه اختصاصی
	۱	۲	۳	۴	۵	۶	۷	
آبادان	۰/۰۳۳۹	۰/۱۱۴۰	۰/۰۶۱۲	۰/۰۹۲۸	۰/۰۴۱۸	۰/۱۳۳۹	۰/۵۲۱۵	۷
آباده	۰/۰۲۲۳	۰/۳۳۱۱	۰/۱۹۶۰	۰/۳۰۶۸	۰/۰۷۹۸	۰/۰۲۱۴	۰/۰۴۲۶	۲
آب علی	۰/۰۵۷۵	۰/۰۵۵۶	۰/۲۲۱۶	۰/۰۷۷۱	۰/۵۴۴۸	۰/۰۲۰۷	۰/۰۲۲۶	۵
جزیره ابوموسی	۰/۰۰۵۱	۰/۰۰۵۸	۰/۰۰۵۷	۰/۰۰۶۵	۰/۰۰۴۴	۰/۹۶۲۲	۰/۰۱۰۳	۶
امیدیه اهواز	۰/۰۴۹۵	۰/۱۰۵۲	۰/۰۷۸۳	۰/۱۱۷۱	۰/۰۵۲۳	۰/۲۸۰۵	۰/۳۱۷۱	۷
اهر	۰/۱۰۱۷	۰/۰۵۴۱	۰/۲۵۸۵	۰/۰۷۱۵	۰/۴۵۸۹	۰/۰۳۰۲	۰/۰۲۵۰	۵
اهواز	۰/۰۰۹۸	۰/۰۳۶۰	۰/۰۱۹۲	۰/۰۳۷۵	۰/۰۱۲۵	۰/۰۳۶۸	۰/۸۴۸۶	۷
...
...
شرق اصفهان	۰/۰۱۹۳	۰/۳۸۲۶	۰/۲۵۷۹	۰/۲۱۶۷	۰/۰۶۷۸	۰/۰۲۰۵	۰/۰۳۴۹	۲
شیراز	۰/۰۱۱۴	۰/۱۳۶۷	۰/۰۷۷۸	۰/۶۹۴۳	۰/۰۳۰۶	۰/۰۱۴۸	۰/۰۳۴۳	۴
شمال تهران	۰/۰۸۴۰	۰/۰۸۰۴	۰/۲۲۵۹	۰/۲۶۱۶	۰/۲۶۷۴	۰/۰۳۴۱	۰/۰۴۶۵	۵
شوشتر	۰/۰۱۷۱	۰/۰۷۲۴	۰/۰۳۴۶	۰/۰۶۴۶	۰/۰۲۲۳	۰/۰۶۴۲	۰/۷۲۴۸	۷
سیاه بیشه	۰/۳۳۳۸	۰/۰۵۴۹	۰/۱۵۴۲	۰/۰۷۷۰	۰/۳۰۶۰	۰/۰۴۱۱	۰/۰۳۳۰	۱

جدول (۳) اسامی ایستگاه‌های متعلق به گروه‌های ۷ گانه با روش خوشه‌بندی فازی براساس پارامتر اقلیمی

شماره گروه	اسامی ایستگاه‌ها
۱	گنبد کاووس، گرگان، نوشهر، پارس آباد مغان، رشت، سیاه بیشه
۲	آباده، کرمان، خاش، ... شهر بابک، شهر رضا، شرق اصفهان
۳	فردوس، گل‌مکان چناران، جلفا، ... سنندج، سرارود کرمانشاه، سردشت
۴	گلپایگان، گناباد، ایلام، ... سرپل ذهاب، سمنان، شیراز
۵	آبعلی، اهر، الشتر، ... سراب، شهرکرد، شمال تهران
۶	جزیره ابوموسی، جاسک، کنارک چاپهار، جزیره کیش، منجیل
۷	آبادان، امیدیه آقاجری، اهواز، ... رامهرمز، سراوان، شوشتر



شکل (۴) ایستگاه‌های تبحیرسنجی متعلق به گروه‌های ۷ گانه با روش خوشه‌بندی فازی بر اساس پارامترهای اقلیمی

از آنجائیکه مبنای انتخاب تعداد بهینه خوشه‌ها، روش خوشه‌بندی فازی است. بنابراین در این روش هم تعداد بهینه خوشه‌ها را ۷ خوشه در نظر می‌گیریم. در روش شبکه عصبی کوهنن مراکز دسته‌ها همان وزن نرون‌ها می‌باشد. در شبکه عصبی کوهنن لایه ورودی شامل پنج پارامتر تبحیر، دما، سرعت باد، رطوبت نسبی و مجموع ساعات آفتابی و لایه خروجی شامل تعداد گروه‌ها (۷-۱) است. پارامترهایی که برای آموزش الگوریتم استفاده شده است تعداد گروه‌ها که در اینجا ۷ گروه (تعداد بهینه خوشه‌ها که از روش دیویس-بولدین بدست آمده) است، نرخ یادگیری ۰/۰۱، نرخ هوش ۰/۰۰۱ و تعداد تکرار ۱۰۰۰ است. در جدول ۴ می‌بینیم که تمامی ایستگاه‌های تبحیرسنجی در ۷ گروه قرار گرفته‌اند. در نتیجه برای انتخاب نرخ یادگیری مناسب مقادیر خطای مربع شده را مقایسه کرده و کمترین مقدارش را در نظر می‌گیریم که مربوط به نرخ یادگیری ۰/۴ است. تعداد ایستگاه‌های قرار گرفته در این ۷ گروه با نرخ یادگیری ۰/۴، طبق جدول زیر به ترتیب ۱۵، ۲۱، ۱۵، ۶، ۱۳، ۱۵ و ۱۲ است.

بطور مشترک ۶۴ ایستگاه تبخیرسنجی توسط خوشه‌بندی فازی و شبکه عصبی کوهنن در گروه‌های مشابه تشخیص داده شده‌اند. در شبکه عصبی کوهنن نرخ یادگیری برای یک نرخ هوش داده شده و تعداد تکرارها یک نقش بزرگ را بازی می‌کند.

جدول (۴) اطلاعات خوشه‌بندی با شبکه عصبی کوهنن بر اساس پارامترهای اقلیمی

ردیف	نرخ آموزش	واریانس	تعداد ایستگاه‌های درون خوشه‌ها							تعداد کل	تعداد خوشه‌ها
			۱	۲	۳	۴	۵	۶	۷		
۱	۰/۰۱	۳۲/۱۴۲۹	۵	۱۶	۱۴	۹	۲۳	۱۵	۱۵	۹۷	۷
۲	۰/۱	۲۷/۱۴۲۹	۱۸	۵	۱۴	۱۴	۱۸	۱۹	۹	۹۷	۷
۳	۰/۲	۱۲۳/۱۴۲۹	۱۳	۱۸	۱۴	۰	۳۵	۵	۱۲	۹۷	۶
۴	۰/۳	۵۱/۴۷۶۲	۱۳	۱۱	۱۴	۹	۱۴	۷	۲۹	۹۷	۷
۵	۰/۴	۲۰/۱۴۲۹	۱۵	۲۱	۱۵	۶	۱۳	۱۵	۱۲	۹۷	۷
۶	۰/۵	۱۰۶/۸۰۹۵	۱۲	۵	۲۶	۹	۳۱	۷	۷	۹۷	۷
۷	۰/۶	۹۴/۸۰۹۵	۷	۲۴	۶	۷	۱۱	۳۱	۱۱	۹۷	۷
۸	۰/۷	۸۱/۱۴۲۹	۸	۲۷	۸	۱۲	۲۵	۱۴	۳	۹۷	۷
۹	۰/۸	۸۸/۱۴۲۹	۷	۱۴	۸	۷	۲۵	۲۹	۷	۹۷	۷
۱۰	۰/۹	۶۶/۸۰۹۵	۱۳	۶	۳۱	۱۱	۱۵	۱۳	۸	۹۷	۷

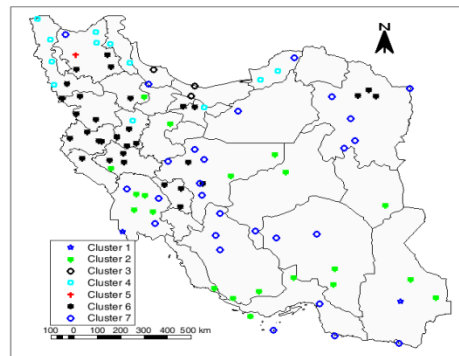
در شبکه عصبی کوهنن توزیع ایستگاه‌ها در ۷ گروه با استفاده از روش خوشه‌بندی فازی مناسب است. همانطور که مشاهده می‌شود تعداد ایستگاه‌های تبخیرسنجی در هر گروه متفاوت است. بنابراین تعداد ایستگاه‌های تبخیرسنجی مشترک با ایستگاه‌هایی که از روش خوشه‌بندی فازی بدست آمده‌اند متفاوت خواهد بود. همچنین مشاهده می‌شود که تاثیر نرخ

یادگیری بر روی مقادیر خطای مربع شده مهم است، ولی هیچ گرایش ثابتی مشاهده نشده است.

جدول (۵) اسامی ایستگاه‌های متعلق به گروه‌های ۷ گانه با روش خوشه‌بندی شبکه عصبی کوهنن بر اساس

پارامترهای اقلیمی

شماره گروه	اسامی ایستگاه‌ها
۱	آبادان، ایرانشهر
۲	اهواز، حاجی آباد هرمزگان، کهنوج، کنگان جم، ... سراوان، ساوه، شوشتر،
۳	نوشهر، رشت، سیاه بیشه
۴	آبعلی، اهر، گنبد کاووس، ... ارومیه، پارس آباد مغان، پیرانشهر
۵	سهند
۶	الشت، گلمکان چناران، همدان (فرودگاه)، ... سر پل ذهاب، شهرکرد، شمال تهران
۷	آباد، جزیره ابوموسی، امیدیه آقاجری، ... شهر رضا، شرق اصفهان، شیراز



شکل (۵) ایستگاه‌های تیخیرسنجی متعلق به گروه‌های ۷ گانه با روش خوشه‌بندی کوهنن بر اساس پارامترهای اقلیمی

یافته‌ها و بحث

خوشه‌بندی با استفاده از داده‌های اقلیمی منجر به قرار گرفتن ایستگاه‌های تبخیرسنجی در ۷ خوشه گردیده‌است و بین مقادیر حد اکثر شاخص RS و حداقل واریانس محاسباتی خوشه‌ها همخوانی وجود دارد. در جدول ۵ مقایسه شاخص‌ها درج گردیده‌است، بطوریکه نسبت به شاخص RS و واریانس خوشه‌ها روش شبکه عصبی کوهنن نتایج بهتری را نشان می‌دهد.

جدول (۵) مقایسه شاخص‌های انتخاب روش خوشه‌بندی

روش خوشه‌بندی	تعداد خوشه‌ها	شاخص RS	شاخص واریانس خوشه‌ها
Fuzzy	۷	۰/۷۹	۳۱۷/۱
ANN	۷	۰/۹۳	۱۸۳/۷

نتیجه‌گیری

پنج پارامتر اقلیمی شامل داده‌های تبخیر، دما، سرعت باد، رطوبت نسبی و تعداد ساعات آفتابی برای طبقه‌بندی استفاده شده‌اند. داده‌ها برای طبقه‌بندی نرمال شدند، تعداد بهینه گروه‌ها بر اساس شاخص دیویس-بولدین انتخاب شدند و نتایج مربوط به خوشه‌بندی و خوشه‌بندی فازی و شبکه عصبی مصنوعی کوهنن آنالیز شده و مقایسه شدند. نتایج خوشه‌بندی هر دو روش قابل استناد است، با این حال روش شبکه عصبی کوهنن در مقایسه با روش فازی با توجه به معیارهای ارزشیابی ضریب تعیین، واریانس خوشه‌ها و دامنه تغییرات تعداد ایستگاه‌ها عملکرد بهتری را نشان می‌دهد. با خوشه‌بندی ایستگاه‌های تبخیرسنجی می‌توان مناطقی که در یک خوشه قرار می‌گیرند، برنامه ریزی مدیریت منابع آبی مشترکی را ارایه داد. آب مجازی را می‌توان برای مناطق با تبخیر بالا و زراعت و کشاورزی را برای منطق با تبخیر پایین پیشنهاد داد. آگاهی از مقدار تبخیر به عنوان یکی از متغیرهای هیدرولوژیکی در تحقیقات کشاورزی و حفاظت آب و خاک و مدل‌سازی آن اهمیت زیادی دارد. نتایج حاصل از این مطالعه با نتایج حاصل از مطالعه جینگ یی و هل در

سال ۲۰۰۴، لی و همکاران در سال ۲۰۱۱، چاوشی و همکاران در سال ۲۰۱۲ و فرسادنیا و همکاران در سال ۲۰۱۴ که در آنها روش‌های نرم خوشه‌بندی و به ویژه روش شبکه عصبی مصنوعی کوهنن بعنوان بهترین روش انتخاب گردیده‌اند، مطابقت دارد.

منابع

- بیابانکی، م؛ اسلامیان، س. (۱۳۸۳)، «استفاده از روش خوشه‌بندی در تعیین همگنی هیدرولوژیک و ارزیابی آن توسط روش‌های تحلیل ممیزی و منحنی‌های *Andrew* در حوضه آبریز کرخه»، مجله کشاورزی، جلد ۶ شماره ۲.
- جهانبخش اصل، س؛ ابطی، و؛ قربانی، م. ع؛ تدینی، م.؛ والا، ا. (۱۳۹۴)، «بررسی توزیع زمانی و مکانی بارش شهرستان تبریز با روش تحلیل خوشه‌ای»، فصلنامه فضای جغرافیایی، سال ۹۴، شماره ۵۰: ۵۹-۸۱.
- حسنعلی زاده، ن.؛ مساعدی، ا.؛ ظهیری، ع. ر.؛ بابانژاد، م. (۱۳۹۳)، «تعیین نواحی همگن توزیع بارش سالانه در سطح استان گلستان با استفاده از تحلیل خوشه‌ای و روش گشتاورهای خطی»، نشریه آب (علوم و صنایع کشاورزی)، سال ۹۳، جلد ۲۸: ۱۰۶۱-۱۰۷۱.
- رحیمی، د.؛ ولی پور، ق.؛ یزدانپناه، ح. (۱۳۹۰)، «کاربرد سامانه و مجموعه‌های فازی در پهنه‌بندی دمایی استان چهارمحال و بختیاری»، مجله جغرافیا و برنامه‌ریزی محیطی، سال ۹۰، شماره ۴۱: ۸۵-۹۶.
- رحیمی خوب، ع؛ بهبهانی، م.؛ نظری فر، م. (۱۳۸۵)، «بررسی استفاده از حداقل داده‌های هواشناسی در معادله پنمن ماتیت- مطالعه موردی استان خوزستان»، مجله علوم کشاورزی، سال ۱۲.
- سلطانی، آ؛ صدوقی یزدی، ه؛ اشک زری طوسی، س؛ روحانی، م. (۱۳۸۹)، «بهبود شبکه خود سازمانده کوهنن با هدف خوشه‌بندی داده‌های فازی»، دهمین کنفرانس سیستم‌های فازی ایران، دانشگاه شهید بهشتی.
- شکاری، پ؛ باقرنژاد، م. (۱۳۸۴)، «بررسی کاربرد روش فازی (Fuzzy) در طبقه‌بندی خاک‌ها. مطالعه موردی: چشمه سفید کرمانشاه»، مجله علوم و فنون کشاورزی و منابع طبیعی، سال نهم، شماره ۴، تهران.
- عساکره، ح.؛ دوستکامیان، م. (۱۳۹۵)، «ناحیه بندی اقلیمی آب قابل بارش جو ایران زمین»، نشریه جغرافیا و برنامه‌ریزی، سال بیستم، شماره ۵۸: ۱۸۱-۲۰۲.
- فرسادنیا، ف؛ رستمی کامرود، م.؛ مقدم نیا، ع. (۱۳۹۱)، «تحلیل روند بارندگی در استان مازندران با استفاده از روش من-کندال منطقه‌ای»، تحقیقات منابع آب ایران، سال ۸۷، شماره ۲: ۶۰-۷۰.

- فلاح قالمهری، غ.؛ اسدی، م.؛ انتظاری، ع. (۱۳۹۴)، « ناحیه بندی آب و هوایی استان گیلان با روش های چند متغیره»، *نشریه جغرافیایی و برنامه ریزی*، سال نوزدهم، شماره ۵۴: ۲۳۵-۲۵۱.
- کوره یزان دزفولی، ا. (۱۳۸۷)، « اصول تئوری مجموعه‌های فازی و کاربردهای آن در مدل‌های مسائل منابع آب». انتشارات جهاد دانشگاهی (دانشگاه صنعتی امیر کبیر).
- مومنی، م. (۱۳۹۰)، « خوشه‌بندی داده‌ها (تحلیل خوشه‌ای)». انتشارات دانشگاه تهران.
- Chavoshi, S., Azmin Sulaiman W.N. Saghafian B. Sulaiman MD NB. Latifah, A.M. (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.
- Dikbas, F., Firat, M., Koc, A.C., Gungor, M. (2011), " Classification of precipitation series using fuzzy cluster Method", *International Journal of Climatology*, DOI: 10.1002/joc.2350.
- Farsadnia, F., Rostami Kamrood, M., Moghaddam Nia, A., Modarres, R. Bray, M.T. and Han, D., Sadatinejad, J. (2014), "Identification of homogeneous regions for regionalization of watersheds by two-level selforganizing feature maps", *Journal of Hydrology*, 509: 387-397.
- Holawe, F., Dutter, R. (1999), " Geostatistical study of precipitation series in Austria", *Journal of Hydrology*, Vol: 219.
- Jingyi, Z. and M.J. Hall. (2004), " Regional flood frequency analysis for the Gan-Ming river basin in China", *Journal of Hydrology* 296: 98-117.
- Ley, R. M.C., Casper, H. Hellebrand and R. Merz. (2011), "Catchment classification by runoff behavior with self-organizing maps (SOM) ", *Hydrology and Earth System Sciences* 15: 2947-2962.
- Lin, G.F. and L.H. Chen. (2006), " Identification of homogeneous regions for regional frequency analysis using the self-organizing map", *Journal of Hydrology*, 324: 1-9.
- Linacre, E. and G. Bart. (1997), "*Climate and weather explained*" , First published, London and New York, pp: 1-432.

- Moghaddamnia, A., Ghafari Gousheh, M., Piri, J., Amin, S., and Han, D. (2009), "Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques", *Advances in Water Resources*. 32(1): 89-97
- Raju, K.S., Kumar, D.N. (2007), "Classification of Indian meteorological stations using cluster and fuzzy cluster analysis, and Kohonen artificial neural networks", *Nordic Hydrology*, 38(3), 303-314.
- Razavi, T., Coulibaly, P. (2013), "Classification of Ontario watersheds based on physical attributes and streamflow series", *Journal of Hydrology*, 493: 81-94.
- Roger, M.C., Durk, R., Cazemier, P. (2000), "Representing and processing uncertain soil information for mapping soil hydrological properties", *Computers and electronics in agriculture*, Vol:29.
- See, L., Openshaw, S. (1998), "Using Soft Computing Techniques to Enhance Flood Forecasting, International Conference on The River Ouse, Proceeding Hydroinformatics", Copenhagen, Denmark, pp. 24-26.