



شناسایی پهنه‌های همگن بارش سنگین و ابرسنگین ایران زمین به کمک نمایه‌های کنترل کیفیت

پراش درون‌گروهی

محمد دارند*^۱

تاریخ دریافت: ۱۳۹۲/۰۹/۲۴

تاریخ پذیرش: ۱۳۹۳/۱۱/۲۵

چکیده

هدف از این مطالعه ارائه الگویی برای خوشه‌بندی مکانی بارش‌های ایران می‌باشد. به این منظور، از داده‌های روزانه بارش میان‌یابی شده یک پایگاه داده موسوم به اسفزاری** طی بازه زمانی ۱۳۴۰/۱/۱ تا ۱۳۸۳/۱۰/۱۱ استفاده شده است. بر اساس محدوده جغرافیایی ایران و دوره زمانی مطالعه یک پایگاه داده در ابعاد ۱۵۹۹۲×۷۱۸۷ که بر روی سطرها زمان (روز) و بر روی ستون‌ها مکان قرار داشت، تشکیل شد. برای هر روز تقویمی از سال و هر سلول مکانی، مقادیر بارش سنگین و ابرسنگین به ترتیب بر پایه صدک ۹۵ و ۹۹ به صورت جداگانه محاسبه و دو پایگاه داده در ابعاد ۷۱۸۷×۳۶۶ برای دو آستانه یاد شده ایجاد شد. پهنه‌بندی به کمک تحلیل خوشه‌ای پایگانی با روش ادغام صورت گرفت. برای گزینش شمار خوشه‌ها از ۱۱ نمایه کنترل کیفیت پراش درون‌گروهی استفاده شد. تعیین شمار پهینه خوشه‌ها برای دو پایگاه داده، بر پایه برازش نمایه‌ها صورت گرفت. بر اساس نتایج حاصل از برازش ۱۱ نمایه بر روی خوشه‌بندی‌های مختلف، ایران زمین به ۵ ناحیه همگن بارش سنگین و ابرسنگین قابل پهنه‌بندی است. برای هر کدام از پهنه‌ها سری زمانی روزانه و ماهانه آستانه‌های بارش محاسبه گردید. یافته‌ها به‌طور کلی نشان داد که نمایه‌های کنترل کیفیت به‌کارگرفته شده از کارایی و دقت مناسبی جهت شناسایی و پایش مکانی مناطق همگن بارش سنگین و ابرسنگین ایران برخوردارند.

واژه‌های کلیدی: پهنه‌بندی، بارش سنگین و ابرسنگین، ایران، تحلیل خوشه‌ای

مقدمه

رویدادهای فرین نظیر سیلاب و خشکسالی می‌شود. برخی از پژوهش‌ها با شبیه‌سازی افزایش گازهای گلخانه‌ای تغییر در شدت بارش را تأیید کرده‌اند (میانرس^۴ و همکاران، ۱۹۹۰). آگاهی و شناخت هر پدیده‌ای ابزارهای مناسب و خاص خود را می‌طلبد. امروزه به توسعه نرم‌افزارهای زمین آمار و آماری پیشرفته پژوهشگران علوم جوی توانایی پاسخ به برخی از ناشناخته‌ها را پیدا کرده‌اند. پاسخ به این پرسش‌هاست که مرز دانش را به سمت جلو پیش می‌برد. امروزه استفاده از روش‌های پهنه‌بندی، گسسته‌سازی و تفکیک نواحی همگن در علوم زمین، جهت ارزیابی پاسخ عرصه‌های طبیعی به فرآیندهای هیدرولوژیک افزایش یافته است چراکه هر چه محدوده مورد واکاوی همگن‌تر باشد موجب بالارفتن دقت و

برآورد بزرگی و بسامد رویدادهای هیدرومتئورولوژیک مثل مقادیر فرین بارش روزانه در مرکز مطالعات ساختار هیدرولیکی، ناحیه‌بندی سیلاب و تخمین اقتصادی طرح‌های حفاظت از سیلاب قرار دارد (نوتو و لالوگیا^۲، ۲۰۰۹؛ سارکار^۳ و همکاران، ۲۰۰۹). اخیراً باور عمومی بر این است که هر گونه تغییری در اقلیم منجر به تغییر در بسامد و شدت

۱ استادیار گروه آب و هواشناسی، دانشکده منابع طبیعی، دانشگاه کردستان

**نویسنده مسئول: (m.darand@uok.ac.ir)

** برگرفته از نام یک اقلیم‌شناس ایرانی

² Noto & Laloggia

³ Sarkar

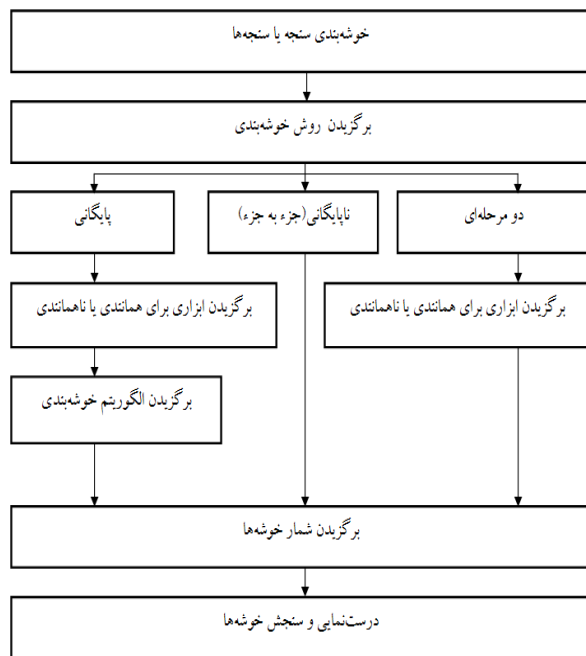
⁴ Mearns

هیدرولوژیک را به کمک روش میان‌مرکز فازی شناسایی کردند. در این پژوهش برای تفکیک نواحی همگن به لحاظ تولید رواناب به‌عنوان نمونه حوضه معرف والنات گولچ (WGEW) در جنوب‌شرقی ایالت آریزونا آمریکا به جهت کیفیت و کیفیت داده‌های هیدرولوژیک برگزیده شد. یافته‌ها نشان داد که استفاده از روش‌های مختلف و موجود اعتباریابی خوشه‌بندی، لزوماً منجر به تعیین تعداد بهینه خوشه‌ها به‌صورت واضح و منطقی نخواهد شد و برای این منظور می‌توان از یک روش منطقی پیشنهادی استفاده کرد. مسعودیان (۱۳۸۲) با به‌کارگیری تحلیل خوشه‌ای بر روی نمرات مؤلفه‌های مینا، ایران را به پانزده ناحیه اقلیمی همگن تقسیم کرد. مسعودیان (۱۳۸۸) در یک پژوهش دیگر نواحی بارشی ایران را شناسایی کرد و ایران‌زمین را به لحاظ مقدار و زمان دریافت بارش به ۸ ناحیه بارشی متمایز پهنه‌بندی کرد. دین‌پژوه و همکاران (۱۳۸۲) به کمک تحلیل‌های چند متغیره (تحلیل مؤلفه‌های مینا و خوشه‌ای)، به پهنه‌بندی اقلیمی ایران جهت استفاده در مطالعات کشاورزی پرداختند. نتایج نشان داد که دو عامل اول موسوم به دما و بارش بیش از ۸۸ درصد از تغییرات داده‌ها را توجیه می‌کنند. مقادیر عامل‌های اول و دوم، پس از چرخش، برای ۷۷ ایستگاه محاسبه و به‌عنوان ماتریس ورودی (۷۷×۲) در تحلیل خوشه‌ای مورد استفاده واقع شدند. کل ایستگاه‌ها با توجه به نمودار درختی حاصل از آن به هفت گروه تفکیک شدند. مسعودیان و دارند (۱۳۸۹) نواحی دمایی فرین سرد ایران را پهنه‌بندی کردند. یافته‌ها نشان داد که ایران را می‌توان به شش ناحیه متفاوت تقسیم کرد. مسعودیان و عطایی (۱۳۸۴) فصول بارشی ایران زمین را با بهره‌گیری از روش تحلیل خوشه‌ای شناسایی کردند و ایران را به پنج ناحیه بارشی با فصول تقریباً یکدست و متمایز تفکیک کردند. مسعودیان (۱۳۸۴) به واکاوی رژیم‌های بارش ایران به کمک روش تحلیل خوشه‌ای پرداخت و نشان داد که در ایران سه رژیم اصلی بارش وجود دارد. اکبری و مسعودیان (۱۳۸۸) به مطالعه رژیم دمایی و نواحی دمایی ایران پرداختند. برای شناسایی یکدست از روش تحلیل خوشه‌ای بهره گرفتند و ایران را به ۸ ناحیه دمایی و به لحاظ زمانی دو فصل دمایی نسبتاً متمایز دسته‌بندی کردند. منتظری و مسعودیان

سهولت در عملیات مدیریت منابع آب و یا کنترل خسارات طبیعی خواهد شد (رعیتی دماوندی و همکاران، ۱۳۹۰). طراحی کنترل مقادیر آب حاصل از رخداد بارش‌های فرین، شناخت جامع از پهنه برای انجام فعالیت‌های کشاورزی و تقویم زراعی مناسب (زمان برداشت و کاشت محصولات) جهت کاهش خسارات با توجه به رژیم زمانی رخداد بارش-های فرین، افزایش دقت سامانه‌های هشدار سیل، ارزیابی شدت بارش و مقایسه پهنه‌های مختلف با هم‌دیگر برپایه آستانه‌های رخداد بارش فرین، شناخت بهتر منطقه برای برنامه‌ریزی‌های بلندمدت و غیره نیازمند شناسایی مناطق همگن بارش‌های فرین منطقه است که از آن با عنوان پهنه-بندی یاد می‌کنند. انجام آن با بهره‌گیری از روش‌های طبقه-بندی صورت می‌پذیرد. هر چه تغییرات مکانی بارش بیشتر باشد، پهنه‌بندی بارش ضروری‌تر و پیچیده‌تر خواهد بود. تحلیل خوشه‌ای روشی مناسب برای شناسایی گروه‌های همگنی از افراد است. اشیائی که در یک خوشه قرار جای می‌گیرند، ویژگی‌های مشترک بسیار زیادی دارند درحالی که هیچ‌گونه شباهتی با افراد خارج از آن خوشه ندارند (وینگسل^۱ و همکاران، ۱۹۹۹). با بهره‌گیری از این روش، پژوهش‌های فراوانی با هدف‌های مختلف به ویژه در دانش اقلیم‌شناسی، در ایران و خارج از ایران‌زمین انجام شده است. برای نمونه می‌توان به کاربرد آن جهت شناسایی پهنه‌های اقلیمی، بارشی، دمایی، رژیم بارش و دمایی و غیره و یا شناسایی الگوها و تیپ‌های همدید هوا اشاره کرد. خام‌چین-مقدم و همکاران (۱۳۸۹) به‌کمک تحلیل خوشه‌ای پایگانی حداکثر بارش روزانه ایران را پهنه‌بندی کردند. آن‌ها از آزمون‌های گشتاورهای خطی برای تأیید همگنی نواحی شناسایی شده بهره گرفتند و کشور را به هفت ناحیه همگن تفکیک کردند. جهانبخش و ترابی (۱۳۸۳) به بررسی و پیش‌بینی تغییرات دما و بارش در ایران پرداختند. ایشان به منظور واکاوی تغییرات زمانی دما و بارش در ایران و پیش-بینی مقادیر آن‌ها برای آینده با بهره‌گیری از تحلیل خوشه-ای پنج ناحیه اقلیمی همگون را شناسایی کردند و سپس برای هر یک از مناطق مذکور یک ایستگاه معرف انتخاب شد. رعیتی دماوندی و همکاران (۱۳۹۰) نواحی همگن

¹ Weingessel

آن‌ها وجود دارد) در پایین درخت خوشه‌بندی با هم‌دیگر یک خوشه جدید را شکل می‌دهند. در گام بعدی جفت خوشه‌های دیگر با هم‌دیگر ترکیب می‌شوند و تا سطوح بالای پایگانی ادامه می‌یابد. خوشه‌بندی می‌تواند از بالا به پایین باشد که خوشه‌بندی شکافتی^۱ نام دارد. به این صورت که همه اشیاء با هم یک خوشه واحد را شکل می‌دهند، سپس به تدریج تفکیک می‌شوند. همانندی یا ناهمانندی افراد را به کمک روش‌های متفاوتی ارزیابی می‌کنند. بسته به نوع داده‌ها نیز روش‌های گوناگونی پیشنهاد شده است. هنگامی که داده‌ها کمی و عددی باشند از روش‌های بسیار رایج زیر برای محاسبه همانندی و ناهمانندی بهره گرفته می‌شود (گان^۲ و همکاران، ۲۰۰۷).



شکل ۱- گام‌شمار تحلیل خوشه‌ای (موی و سارستدیت، ۲۰۱۱)

فاصله اقلیدسی: این روش رایج‌ترین روش برای محاسبه فاصله بین داده‌های عددی است. برای دو نقطه X و Y در محیط d بعدی فاصله اقلیدسی بین آن‌ها مطابق معادله ۱ است.

(۱۳۸۹) به واکاوی الگوهای فرارفت دمایی ایران طی سال‌های سرد پرداختند و با انجام تحلیل خوشه‌ای بر روی نمرات مؤلفه‌های مبنا دوازده الگوی فرارفتی را شناسایی کردند. هدف از انجام این پژوهش شناسایی پهنه‌های بارش سنگین و ابرسنگین ایران زمین به کمک نمایه‌های کنترل کیفیت پراش درون‌گروهی است.

مواد و روش‌ها

داده‌های روزانه بارش میان‌یابی شده پایگاه داده اسفزاری طی بازه زمانی ۱۳۴۰/۱/۱ تا ۱۳۸۳/۱۰/۱۱ جهت واکاوی و شناسایی پهنه‌های بارش سنگین و ابرسنگین به کار گرفته شد. یک پایگاه داده در ابعاد ۱۵۹۹۲×۷۱۸۷ ایجاد شد که بر روی سطرها زمان (روز) و بر روی ستون‌ها مکان (یاخته-ها) قرار داشت. برای هر روز تقویمی و هر یاخته، جداگانه مقادیر بارش سنگین و ابرسنگین به ترتیب بر پایه صدک ۹۵ و ۹۹م حساب شد. دو پایگاه داده در ابعاد ۷۱۸۷×۳۶۶ برای دو آستانه یاد شده ایجاد گردید. به یاری تحلیل خوشه‌ای پایگانی با روش ادغام وارد مناطق همگن بارش سنگین و ابرسنگین شناسایی شد. گام‌شمار فرآیند خوشه‌بندی در شکل ۱ آمده است. اولین گام در فرآیند خوشه‌بندی آن است که روش خوشه‌بندی داده‌ها را برگزینیم. سه روش متداول خوشه‌بندی شامل روش پایگانی، روش ناپایگانی و ترکیبی از هر دو روش می‌باشد. هر کدام از روش‌های یاد شده رویکردهای متفاوتی برای دسته‌بندی کردن و خوشه‌بندی اشیاء دارند. تفاوت خوشه‌بندی پایگانی با ناپایگانی این است که در روش پایگانی ابتدا دو مشاهده با هم‌دیگر پیوند داده می‌شوند و هرگز از هم جدا نمی‌شوند درحالی‌که در روش ناپایگانی هر مشاهده پس از پیوند با مشاهده دیگر از آن جدا شده و با مشاهدات دیگر پیوند داده می‌شود. در روش سوم از هر دو رویکرد استفاده می‌شود. در روش پایگانی طی مسیر تحلیل، یک درخت خوشه‌بندی ایجاد می‌شود. اغلب روش‌های پایگانی به یک خوشه ختم می‌شوند که خوشه‌بندی انباشتی نام دارد. نخست هر کدام از اشیاء به‌طور مجزا، یک خوشه در نظر گرفته می‌شود. سپس اشیاء برحسب همانندی با هم‌دیگر خوشه جدیدی را می‌سازند. ابتدا شبیه‌ترین دو فرد (آن‌هایی که کوچکترین فاصله بین

¹ Divisive

² Gan

³ Mooi & Sarstedt

$$D_{mah}(x, y) = \sqrt{(x-y)(x-y)'} \quad (7)$$

که V^{-1} ماتریس هم‌پراش سری داده‌ها و $(x-y)'$ ترانهاده آن می‌باشد. در مطالعات اقلیمی غالباً برای محاسبه درجه ناهمانندی (همانندی) از فاصله اقلیدسی استفاده می‌شود (مسعودیان، ۱۳۸۶). بعد از اینکه روشی را برای همانندی یا ناهمانندی برگزیدیم، نیازمند آن هستیم که الگوریتم خوشه‌بندی را انتخاب کنیم. روش‌های پایگانی انباشتی (پیوندی) مختلفی وجود دارد و از روشی که برای تعریف فاصله از خوشه جدید ایجاد شده برمی‌گزینند، قابل تشخیص هستند. رایج‌ترین روش‌های انباشتی به شرح زیر است.

پیوند تکی (نزدیکترین همسایه): نزدیکترین فاصله اعضای دو خوشه مبنای پیوند است.

پیوند کامل (دورترین همسایه): دورترین فاصله اعضای دو خوشه مبنای پیوند است.

پیوند متوسط: میانگین فاصله همه اعضای خوشه تا خوشه دیگر مبنای پیوند است.

پیوند مرکزی: در این رویکرد، مرکز هندسی هر خوشه نخست حساب می‌شود. فاصله بین دو خوشه برابر با فاصله بین مرکز دو خوشه است. هر کدام از الگوریتم‌های پیوندی نتایجی متفاوتی را برای داده‌های مشابه نشان می‌دهند. در پیوند تکی، داده‌ها در یک خوشه بزرگ در مقابل خوشه‌های دیگر که ممکن است یک یا چند فرد در هر کدام از آن‌ها جای داشته باشند، قرار می‌گیرند. این روش پیوند جهت شناسایی مقادیر پرت مناسب است و به‌عنوان یک روش و الگوریتم چندبعدی در نظر گرفته می‌شود. برخلاف آن پیوند کامل به شدت تحت تأثیر مقادیر پرت است چراکه این روش برپایه بیشینه فاصله است. روش‌های پیوند مرکزی و متوسط خوشه‌هایی با پراش درون گروهی کم با اندازه‌های مشابه ایجاد می‌کنند. هر دو روش از مقادیر پرت تأثیر می‌پذیرند ولی نه به اندازه پیوند کامل. یک روش رایج دیگر در خوشه-بندی پایگانی، روش وارد است. روش کمینه پراش وارد فراوان‌ترین تکنیک پایگانی برای طبقه‌بندی اقلیمی است (ویلیموت و ورنون^۱، ۱۹۸۰؛ وینکلر^۲، ۱۹۸۵؛ کالکستاین و

$$D_{enc}(x, y) = \left[\sum_{j=1}^d (X_j - y_j)^2 \right]^{\frac{1}{2}} \quad (1)$$

که در آن X_j و y_j به ترتیب مقادیر روی نقطه X و Y است.

فاصله مُنهایتان: نام دیگر این روش، فاصله بلوک شهری است و مجموع فواصل همه افراد (نقاط) مبنای محاسبه است. برای دو نقطه X و Y در محیط d بعدی فاصله مُنهایتان بین آن‌ها به کمک معادله ۲ به دست می‌آید.

$$D_{man}(x, y) = \sum_{k=1}^d |X_j - y_j| \quad (2)$$

اگر دو نقطه X و Y در برخی موقعیت‌ها نبود آماری داشته باشند فاصله مُنهایتان به کمک معادله ۳ حساب می‌شود.

$$D_{manw}(x, y) = \frac{\sum_{k=1}^d w_j |x_j - y_j|}{\sum_{k=1}^d w_j} \quad (3)$$

در صورتی که هم X و Y بر روی نقطه Z داده داشته باشند w_j برابر با یک است، در غیر اینصورت w_j برابر با صفر در نظر گرفته می‌شود.

بیشینه فاصله: این روش فاصله سوپ نیز نامیده می‌شود. بیشینه مقادیر فواصل بین نقاط مبنای محاسبه است.

$$D_{max}(x, y) = \max_{1 \leq k \leq d} |x_j - y_j| \quad (4)$$

فاصله مینکاووسکی: فاصله اقلیدسی، فاصله مُنهایتان و بیشینه فاصله سه روش خاصی از فاصله مینکاووسکی هستند که به کمک معادله ۵ به دست می‌آید.

$$D_{min}(x, y) = \left(\sum_{j=1}^d |x_j - y_j|^r \right)^{\frac{1}{r}} \quad r \geq 1 \quad (5)$$

که r مرتبه فاصله مینکاووسکی است. اگر r برابر با ۱، ۲ و ∞ باشد، به ترتیب فاصله اقلیدسی، فاصله مُنهایتان و بیشینه فاصله به دست خواهد آمد.

میانگین فاصله: روش میانگین فاصله، تغییر یافته روش فاصله اقلیدسی است. برای دو نقطه X و Y در محیط d بعدی میانگین فاصله بین آن‌ها به کمک معادله ۶ به دست می‌آید.

$$D_{ave}(x, y) = \left[\frac{1}{d} \sum_{j=1}^d (x_j - y_j)^2 \right]^{\frac{1}{2}} \quad (6)$$

فاصله ماهالانوبیس: این روش به کمک ترکیبی خطی از مقادیر بر روی نقاط، فاصله‌ها را از هم کم می‌کند.

¹ Willmott & Vernon

² Winkler

$$Z_{ij} = \frac{\binom{n_i+n_j}{2}}{\binom{\sum ijk}{2}}, Z_{ik} = \frac{\binom{n_i+n_k}{2}}{\binom{\sum ijk}{2}},$$

$$Z_{jk} = \frac{\binom{n_j+n_k}{2}}{\binom{\sum ijk}{2}}, \quad (10)$$

$$Z_{ijk} = \frac{\binom{n_i+n_j+n_k}{2}}{\binom{\sum ijk}{2}}$$

که $\sum ijk = n_i + n_j + n_k$ است و سرانجام مقادیر Z_i ، Z_j و Z_k به کمک معادله ۱۱ به دست می‌آید.

$$Z_i = \frac{\binom{n_i}{2}}{\binom{\sum ijk}{2}}, Z_j = \frac{\binom{n_j}{2}}{\binom{\sum ijk}{2}}, Z_k = \frac{\binom{n_k}{2}}{\binom{\sum ijk}{2}} \quad (11)$$

انتخاب متفاوت پارامترهای a_i ، a_j ، β ، γ ، δ_i ، δ_j ، δ_k منجر به خوشه‌بندی‌های جداگانه خواهد شد. مشکل اصلی در تحلیل خوشه‌ای، تصمیم بر سر این است که داده‌ها به چه تعداد خوشه دسته‌بندی شود و چه روشی برای درست‌نمایی خوشه‌های برگزیده مناسب است؟

این دوگام در بین گام‌شمار فرآیند خوشه‌بندی بسیار مهم به حساب می‌آیند چرا که اگر درست گزینش نکنیم دچار خطا خواهیم شد. معیارهای زیادی برای محاسبه کیفیت خوشه‌های حاصل از خوشه‌بندی توسط پژوهشگران مختلف معرفی و پیشنهاد شده است. کورماک و گوردون دو واژه همگنی و جدایی را برای توصیف خوشه‌ها پیشنهاد کردند که به ترتیب اولی اشاره به شباهت درون گروهی یا همگنی درون گروهی و دومی بیان‌گر فاصله بین گروهی است. هر چه شباهت درون گروهی یا همگنی درون گروهی و فاصله بین-گروهی بیشتر باشد، خوشه‌بندی مطلوب‌تر است (کورماک^۵، ۱۹۷۱؛ گوردون^۶، ۱۹۹۹).

کورینگان^۱، ۱۹۸۶؛ به نقل از کالکستاین^۲ و همکاران، ۱۹۸۷). در هر گام خوشه‌بندی، مجموع مربعات درون گروهی کمینه است. این رویکرد دو اشیاء و اقلام متوالی مشابه را با هم ترکیب نمی‌کند بلکه به جای آن، آن دسته از اشیائی که پراش درون گروهی را به کوچکترین مقدار ممکن می‌رسانند با همدیگر ترکیب می‌شوند. در الگوریتم پایگانی فواصل بین خوشه قدیمی و خوشه جدید حساب می‌شود. لانس و ویلیامز^۳ (۱۹۶۷) یک معادله را پیشنهاد کردند که فاصله بین خوشه C_k و خوشه C_j را که ترکیبی از خوشه‌های C_i و C_j و غیره $C = C_i \cup C_j$ است، به کمک معادله پیشنهادی آن‌ها حساب می‌شود.

$$D(C_k, C_i \cup C_j) = a_i D(C_k, C_i) + a_j D(C_k, C_j) + \beta D(C_i, C_j) + \gamma |D(C_k, C_i) - D(C_k, C_j)| \quad (8)$$

که D فاصله بین دو خوشه است. اگر پارامترهای a_i ، a_j ، β و γ را به درستی انتخاب کرده باشیم فواصل درون گروهی متفاوتی را برای الگوریتم‌های خوشه‌بندی پایگانی به دست خواهیم آورد. در جدول ۱ مقادیر برخی از پارامترهایی را که معمولاً در معادله پیشنهادی لانس و ویلیامز به کار گرفته می‌شوند، آمده است. جامبو^۴ (۱۹۷۸)، به نقل از گان و همکاران، (۲۰۰۷) معادله (۹) را ارائه کرد که در آن از پارامترهای بیشتری برای محاسبه فاصله بین خوشه‌ها نسبت به معادله پیشنهادی لانس و ویلیامز (۱۹۶۷) استفاده شده است.

$$D(c_k, c_i \cup c_j) = a_i D(c_k, c_i) + a_j D(c_k, c_j) + \beta D(c_i, c_j) + \gamma |D(c_k, c_i) - D(c_k, c_j)| + \delta_i h(c_i) + \delta_j h(c_j) + \delta_k h(c_k) \quad (9)$$

که در آن، D فاصله بین دو گروه و $h(C_i)$ ارتفاع درخت خوشه‌بندی (دندروگرام C_i) است. در جدول ۱ نمادهای n_i ، n_j و n_k به ترتیب تعداد اعضا و افراد (نقاط) در گروه‌های C_i ، C_j و C_k هستند. همچنین نمادهای Z_{ij} ، Z_{ik} ، Z_{jk} و Z_{ijk} به کمک معادلات ۱۰ و ۱۱ محاسبه می‌شوند.

¹ Kalkstein & Corrigan

² Kalkstein

³ Lance & Williams

⁴ Jambu

⁵ Cormack

⁶ Gordon

جدول ۱- کمیت‌های پیشنهادی لانس و ویلیامز (۱۹۶۷) و جامبو (۱۹۷۸) به نقل از گان و همکاران (۲۰۰۷)

| نام روش | پارامترها | a_i | a_j | β | γ | δ_i | δ_j | ε | برگرفته از |
|-----------------------------------|-----------|--------------------------------|--------------------------------|---------|----------------|---------------------------|---------------------------|---------------------------|-------------------------------------|
| Single-link | | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $-\frac{1}{2}$ | 0 | 0 | 0 | (Sneath,1957) |
| Complete link | | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $-\frac{1}{2}$ | 0 | 0 | 0 | (McQuitty,1960) |
| Ward method (minimum variance) | | $\frac{n_i + n_k}{\sum_{ijk}}$ | $\frac{n_i + n_k}{\sum_{ijk}}$ | | 0 | 0 | 0 | 0 | (Ward Jr., 1963) (Wishart, 1969) |
| Group average (UPGMA) | | $\frac{n_i}{n_i + n_j}$ | $\frac{n_j}{n_i + n_j}$ | 0 | 0 | 0 | 0 | 0 | (McQuitty,1967) |
| Weighted groups average (WPGMA) | | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | 0 | 0 | 0 | 0 | (McQuitty,1966) (McQuitty,1967) |
| Centroid (UPGMC) | | $\frac{n_i}{n_i + n_j}$ | $\frac{n_j}{n_i + n_j}$ | | 0 | 0 | 0 | 0 | (Gower, 1967) |
| Median (WPGMC, weighted centroid) | | $\frac{1}{2}$ | $\frac{1}{2}$ | | 0 | 0 | 0 | 0 | (Gower, 1967) |
| Flexible | | $\frac{1-x}{2}$ | $\frac{1-x}{2}$ | | 0 | 0 | 0 | 0 | (Lance and Williams,1966) |
| Mean dissimilarity | | $\frac{Z_{ik}}{Z_{ijk}}$ | $\frac{Z_{ik}}{Z_{ijk}}$ | | 0 | 0 | 0 | 0 | (Podani,1989) |
| Sum of squares | | $\frac{n_i + n_k}{\sum_{ijk}}$ | $\frac{n_i + n_k}{\sum_{ijk}}$ | | 0 | $\frac{-n_k}{\sum_{ijk}}$ | $\frac{-n_j}{\sum_{ijk}}$ | $\frac{-n_i}{\sum_{ijk}}$ | (Jambu,1978,Podani,1989) |

★ سه کمیت پایانی جدول، مربوط به کمیت‌های پیشنهادی جامبو (۱۹۷۸) است.

متفاوت، جهت ارزیابی شمار خوشه‌های بهینه استفاده شده است. نمایه‌ها بر پایه پراش درون گروهی عمل می‌کنند. در زیر تعریف دقیق نمایه‌های مختلف آمده است. نام دیگر این نمایه‌ها، نمایه‌های کنترل کیفیت درون گروهی است که با نماد C نمایش داده می‌شوند.

۱- نمایه Ball Hall

میانگین پراکندگی یک خوشه برابر با میانگین مربع فواصل افراد از مرکز خوشه است. نمایه بال هال میانگین پراکندگی در داخل کلیه خوشه‌هاست و از معادله ۱۲ به دست می‌آید.

$$C = \frac{1}{K} \sum_{k=1}^K \frac{1}{n_k} \sum_{i \in I_k} \|M_i^{(k)} - G^{(k)}\|^2 \quad (12)$$

در یک دسته‌بندی کلی می‌توان روش‌ها و معیارهایی را که برای درست‌نمایی گزینش تعداد خوشه‌ها برگزیدیم را در سه دسته نمایه‌های درون گروهی، نمایه‌های برون گروهی و نمایه‌های نسبی جای داد. میلیگان و کوپر از ۳۰ روش پیشنهادی برای ارزیابی شمار واقعی خوشه‌های حاصل از الگوریتم خوشه‌بندی پایگانی بهره گرفتند. بر پایه یافته‌های این پژوهش نمایه Calinski_Harabasz کارایی بیشتری نسبت به سایر روش‌ها دارد (میلیگان و کوپر^۱، ۱۹۸۵). پژوهشگران دیگر به نتایج متفاوت دست یافته‌اند. به نظر می‌رسد که به‌کارگیری روش‌های متفاوت سنجش و درست‌نمایی شمار خوشه‌ها و مقایسه کردن نتایج آن‌ها با هم می‌تواند بسیار کارساز باشد. در این پژوهش از ۱۱ نمایه

¹ Milligan & Cooper

$$D_k = \max_{\substack{i \in I_k \\ j \in I_k}} \|M_i^{(k)} - M_j^{(k)}\| \quad (20)$$

پس d_{\max} بزرگترین مقدار D_k است.

$$d_{\max} = \max_{1 < k < K} D_k \quad (21)$$

شاخص Dunn حاصل تقسیم d_{\min} بر d_{\max} است.

$$C = \frac{d_{\min}}{d_{\max}} \quad (22)$$

۵- نمایه PBM

شاخص PBM^۱ برپایه فواصل بین افراد و مرکز خود خوشه‌ای که در آن قرار دارند و فواصل بین افراد از مرکز کل داده‌ها محاسبه می‌شود. اگر بزرگ‌ترین فاصله بین مرکز دو خوشه را با D_B نشان دهیم آنگاه:

$$D_B = \max_{k < k'} d(G^{(k)}, G^{(k')}) \quad (23)$$

و اگر مجموع فواصل افراد یک خوشه از مرکز را با E_W و مجموع فواصل همه افراد از مرکز (G) کل سری داده‌ها را با E_T نشان دهیم، آنگاه:

$$E_W = \sum_{k=1}^K \sum_{i \in I_k} d(M_i, G^{(k)}) \quad (24)$$

$$E_T = \sum_{i=1}^N d(M_i, G) \quad (25)$$

بنابراین شاخص PBM به صورت معادله ۲۶ تعریف می‌شود.

$$C = \left(\frac{1}{K} \times \frac{E_T}{E_W} \times D_B \right)^2 \quad (26)$$

مقدار E_T ثابت است و به تعداد خوشه‌ها و گروه‌بندی آن‌ها بستگی ندارد.

۶- نمایه Point-Biserial

به‌طور کلی در آمار ضریب دو رشته‌ای نقطه‌ای^۲ همبستگی بین متغیر پیوسته A و متغیر باینری (دودویی) B (متغیری که مقادیر آن ۰ و ۱ است) تعریف شده است. طول سری زمانی دو متغیر یک اندازه است. مقادیر ماتریس A بسته به مقادیر ۰ و ۱ گروه B ، به دو گروه A_0 و A_1 دسته‌بندی می‌شوند. اگر M_{A_0} و M_{A_1} میانگین A_0 و A_1 و n_{A_0} و n_{A_1} تعداد افراد هر گروه باشند، آنگاه مقدار ضریب

که K تعداد خوشه‌ها، n_k تعداد افراد داخل خوشه $M_i^{(k)}$ مشاهدات A از خوشه k و $G^{(k)}$ مرکز خوشه k است.

۲- نمایه C

این نمایه فاصله بین افرادی که در داخل یک خوشه قرار دارند را نشان می‌دهد.

$$C = \frac{S_W - S_{\min}}{S_{\max} - S_{\min}} \quad (13)$$

که در آن S_W مجموع فواصل N_W بین افراد داخل هر خوشه است. مقدار N_W به کمک معادله ۳ به دست می‌آید. S_{\min} مجموع کوچکترین فاصله N_W در بین همه افراد سری داده‌هاست. S_{\max} مجموع بزرگترین N_W در بین همه افراد سری داده‌هاست.

$$N_W = \sum_{k=1}^K \frac{n_k(n_k-1)}{2} \quad (14)$$

۳- نمایه Calinski_Harabasz

این نمایه به کمک معادله ۱۵ به دست می‌آید.

$$C = \frac{BGSS / (K-1)}{WGSS / (N-K)} = \frac{N-K(BGSS)}{K-1(WGSS)} \quad (15)$$

$$WGSS^{(k)} = \frac{1}{n_k} \sum_{i < j \in I_k} \|M_i^{(k)} - M_j^{(k)}\|^2 \quad (16)$$

$$BGSS = \sum_{k=1}^K n_k \|G^{(k)} - G\|^2 \quad (17)$$

۴- نمایه Dunn

فرض کنید که d_{\min} کمینه فاصله بین افراد خوشه‌های مختلف باشد و d_{\max} بیشینه فاصله درون خوشه‌ای باشد. فاصله بین خوشه‌های C_k و $C_{k'}$ به کمک فاصله نزدیک‌ترین افرادشان اندازه‌گیری می‌شود.

$$d_{kk'} = \min_{\substack{i \in I_k \\ j \in I_{k'}}} \|M_i^{(k)} - M_j^{(k')}\| \quad (18)$$

d_{\min} کمینه مقدار $d_{kk'}$ است.

$$d_{\min} = \min_{k \neq k'} d_{kk'} \quad (19)$$

برای هر خوشه C_k یک D_k که بزرگ‌ترین فاصله بین دو فرد در خوشه است به کمک معادله ۲۰ شناسایی می‌شود (برخی اوقات به آن قطر خوشه نیز گفته می‌شود).

^۱ Pakhira, Bandyopadhyay and Maulik

^۲ Point-Biserial

مقدار $s(x)$ بین ۱- و ۱ خواهد بود. مقادیر نزدیک به ۱ بیانگر آن است که فرد M_i مربوط به همان گروه است درحالی که مقادیر نزدیک به ۱- نشان می‌دهد که این فرد مربوط به گروه دیگری است. میانگین عرض silhouette برای هر گروه C_k میانگین silhouette گروه نامیده می‌شود و به کمک معادله ۳۷ به دست می‌آید.

$$S_k = \frac{1}{n_k} \sum_{i \in I_k} s(i) \quad (37)$$

سرانجام نمایه کلی silhouette برابر با متوسط میانگین-های silhouette کل خوشه‌ها است.

$$C = \frac{1}{K} \sum_{k=1}^K S_k \quad (38)$$

۹- نمایه Gamma

نمایه Gamma میزان سازش را در خوشه‌بندی می‌سنجد. نمایه Γ همبستگی بین دو بردار هم اندازه A و B را نشان می‌دهد. به طور کلی برای دو شاخص i و j که $a_i < a_j$ است، دو بردار هم‌نوا هستند اگر $b_i < b_j$. به بیانی دیگر اگر مقادیر با ترتیب مشابهی در دو بردار طبقه‌بندی شوند، آن دو بردار هم‌نوا هستند. تعداد s^+ جفت افراد ناهمنوای $\{i, j\}$ و s^- تعداد جفت‌های ناهمنوا را نشان می‌دهند. نمایه Γ به کمک معادله ۳۹ حساب می‌شود.

$$C = \Gamma = \frac{s^+ - s^-}{s^+ + s^-} \quad (39)$$

مقادیر آن بین ۱- و ۱ خواهد بود. هنگام گروه‌بندی اولین بردار A فواصل d_{ij} بین جفت افراد $\{M_i, M_j\}$ را نشان می‌دهد که $i < j$ است. دومین بردار B است که مقادیر آن دودویی است. در بردار B اگر دو فرد در یک گروه قرار گیرند، مقدار آن ۱ و در غیراینصورت برابر صفر خواهد بود. دو بردار طول یکسانی دارند و برابر با $NT = N(N-1)/2$ خواهد بود. تعداد s^+ بیانگر تعداد بارهایی که فاصله بین دو فرد که متعلق به یک گروه نیستند (که میزان آن برای بردار B برابر با صفر است) به شدت کوچکتر از فاصله بین دو فردی است که متعلق به یک گروه هستند (که میزان آن برای بردار B برابر با ۱ است). تعداد s^- بیانگر تعداد بارهایی که برای دو فرد شرایط عکس اتفاق می‌افتد. در مواردی که موازنه برابر قرار است شمارش نمی‌شوند. مقدار N_B برابر با فواصل بین

همبستگی دورشته‌ای نقطه‌ای به کمک معادله ۲۷ حساب می‌شود.

$$r_{pb}(A, B) = \frac{M_{A1} - M_{A0}}{s_n} \sqrt{\frac{n_{A0} n_{A1}}{n^2}} \quad (27)$$

که s_n انحراف معیار A است.

$$C = s_n \times r_{pb}(A, B) \quad (28)$$

۷- نمایه Ratkowsky_Lance

میانگین خارج قسمت BSGSS و TSS داده‌ها به کمک معادله ۲۹ حساب می‌شود.

$$BGSS_j = \sum_{k=1}^K n_k (\mu_j^{(k)} - \mu_j)^2 = b_{jj} \quad (29)$$

$$TSS_j = NVar(V_j) = \sum_{i=1}^N (a_{ij} - \mu_j)^2 \quad (30)$$

بنابراین

$$\bar{c}^{-2} = \bar{R} = \frac{1}{p} \sum_{j=1}^p \frac{BGSS_j}{TSS_j} \quad (31)$$

نمایه Ratkowsky_Lance (\bar{C} / \sqrt{k}) به کمک معادله ۳۲ قابل محاسبه است.

$$C = \sqrt{\frac{\bar{R}}{K}} = \frac{\bar{c}}{\sqrt{K}} \quad (32)$$

۸- نمایه Silhouette

فرض کنید که M_i میانگین فاصله هر فرد از گروه‌های مختلف باشد و $a(i)$ میانگین فاصله درون‌گروهی باشد، اگر $M_i \in C_k$ آن‌گاه معادله ۳۳ را خواهیم داشت.

$$a(i) = \frac{1}{n_k - 1} \sum_{i \in I_k} d(M_i - M_i) \quad (33)$$

میانگین فاصله M_i از افراد هر یک از خوشه‌های C_k به کمک معادله ۳۴ به دست می‌آید.

$$d(M_i, C_k) = \frac{1}{nk} \sum_{i \in I_k} d(M_i - M_i) \quad (34)$$

کوچکترین میانگین فواصل با $b(i)$ نشان داده می‌شود.

$$b(i) = \min_{k \neq k'} d(M_i, C_k) \quad (35)$$

برای هر یک از افراد M_i خارج قسمت حاصل تقسیم زیر در نظر گرفته می‌شود و عرض silhouette نامیده می‌شود.

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \quad (36)$$

$$C = \frac{s^+ - s^-}{\sqrt{N_B N_W \left(\frac{N_T (N_T - 1)}{2} \right)}} \quad (49)$$

۱۱- نمایه Xie_Beni

نمایه Xie_Beni یک نمایه برای خوشه‌بندی فازی است ولی برای سایر روش‌های خوشه‌بندی نیز کاربرد دارد. مقدار شاخص برابر با خارج‌قسمت میانگین درجه دوم خطا و کمینه مربع فاصله بین افراد یک خوشه است. میانگین خطای درجه دوم برای روش‌های خوشه‌بندی غیر از روش فازی به کمک $\frac{1}{N} WGSS$ به دست می‌آید. به بیانی دیگر میانگین مربع فاصله همه افراد از مرکز خوشه گروهی که در آن قرار دارند برابر با میانگین خطای درجه دوم خواهد شد. و نمایه Xie_Beni به کمک معادله‌های ۵۰ و ۵۱ به دست می‌آید.

$$\delta_1(C_k, C_{k'}) = \min_{\substack{i \in I_k \\ j \in I_{k'}}} d(M_i - M_j) \quad (50)$$

$$C = \frac{1}{N} \frac{WGSS}{\min_{k < k'} \delta_1(C_k, C_{k'})^2} \quad (51)$$

ملاک و معیار انتخاب بهینه‌ترین شمار خوشه‌ها

به منظور گزینش بهترین خوشه‌بندی داده‌ها برای هر یک از خوشه‌ها (از ۲ تا n) مقادیر نمایه‌هایی که در جدول ۱ آمده است، محاسبه شد. در ستون سوم جدول راهنمای انتخاب بهینه تعداد خوشه‌ها برای نمایه‌های مختلف آمده است. بر پایه آن‌ها گزینش شمار خوشه‌ها انجام می‌گیرد. لازم به ذکر است که این روش برای هر گونه تکنیک خوشه‌بندی (الگوریتم خوشه‌بندی پایگانی یا الگوریتم چند میانگین (k-means) و غیره) کارآیی مناسبی دارد. بر پایه جدول ۱ بهینه‌ترین شمار خوشه‌ها برگزیده می‌شود. برای نمونه اگر برای نمایه Calinski_Harabasz شاخص کنترل کیفیت برای خوشه‌بندی‌های مختلف داده‌ها محاسبه شود، بزرگترین مقدار شاخص، بهینه‌ترین شمار خوشه‌هاست. در جدول ۲، max و min به ترتیب به معنای بیشینه و کمینه مقدار شاخص است و بیشینه اختلاف max diff به معنای بهترین مقدار k (شمار خوشه‌ها) است که مربوط به

گروه‌ها است و هر یک از آن‌ها با مقادیر N_W فواصل درون گروهی مقایسه می‌شود. بنابراین تعداد مقایسه‌ها برابر با حاصل ضرب $N_W \times N_B$ است.

$$s^+ = \sum_{(r,s) \in I_B} \sum_{(u,v) \in I_W} 1_{\{d_{rs} < d_{uv}\}} \quad (40)$$

$$s^- = \sum_{(r,s) \in I_B} \sum_{(u,v) \in I_W} 1_{\{d_{rs} > d_{uv}\}} \quad (41)$$

اختلاف بین آن‌ها به کمک معادله ۴۲ به دست می‌آید.

$$s^+ - s^- = \sum_{(r,s) \in I_B} \sum_{(u,v) \in I_W} \text{sgn}(d_{uv} - d_{rs}) \quad (42)$$

۱۰- نمایه Tau

به کمک روابطی که برای نمایه Gamma ذکر شد، نمایه Tau کندال بین دو بردار داده با طول NT به کمک معادله ۴۳ به دست می‌آید.

$$\tau = \frac{s^+ - s^-}{N_T(N_T - 1)} \quad (43)$$

اگر فاصله بین گروهی و درون گروهی برابر باشد، تعداد s^+ و s^- شمارش نمی‌شود. برای اینکه مقادیر برابر نیز در نظر گرفته شود یک معادله تصحیح شده به شکل معادله ۴۴ تعریف می‌شود.

$$\tau_c = \frac{s^+ - s^-}{\sqrt{(v_0 - v_1)(v_0 - v_2)}} \quad (44)$$

$$v_0 = \frac{N_T(N_T - 1)}{2} \quad (45)$$

$$v_1 = \sum_i \frac{t_i(t_i - 1)}{2} \quad (46)$$

$$v_2 = \sum_j \frac{u_j(u_j - 1)}{2} \quad (47)$$

که t_i تعداد مقادیر برابر در بردار A و u_i تعداد مقادیر برابر در بردار B است. بردار B شامل مقادیر ۰ و ۱ است (بسته به نوع بین گروهی و درون گروهی به ترتیب مقادیر ۰ و ۱ را به خود اختصاص می‌دهد).

$$v_2 = \frac{N_B(N_B - 1)}{2} + \frac{N_W(N_W - 1)}{2} \quad (48)$$

با یک محاسبه آسان می‌توان به این نتیجه رسید که $v_0 - v_2 = N_B N_W$. بنابراین مقادیر اصلاح شده نمایه Tau به کمک معادله ۴۹ به دست خواهد آمد.

نتایج و بحث

میانگین سالانه بارش دریافتی ایران زمین در شکل ۳ الف نشان داده شده است. همانطور که ملاحظه می‌شود بین مناطق مختلف ایران زمین به لحاظ مقدار دریافت بارش تفاوت‌های بسیار زیادی وجود دارد. بر روی سواحل جنوبی دریای خزر و رشته کوه‌های زاگرس هسته‌های بیشینه بارش دیده می‌شود. بیشترین میزان بارش دریافتی مربوط به سواحل جنوب غربی دریای خزر است. میانگین بلندمدت بارش دریافتی ۱۷۴۰ میلی‌متر در سال است. کمترین مقدار بارش مربوط به مناطق مرکزی ایران است که به‌طور میانگین در طول سال در برخی نقاط حدود ۵۰ میلی‌متر بارش دریافت می‌شود. شکل (۳ ب) میزان صدک ۹۵م بارش سالانه را بر روی ایران زمین نشان می‌دهد.

جدول ۲- معیارگزینش بهینه شمار خوشه‌ها

| شماره | نمایه | بهینه گزینش |
|-------|-------------------|---------------|
| ۱ | Ball Hall | بیشینه اختلاف |
| ۲ | C_Index | کمینه |
| ۳ | Calinski Harabasz | بیشینه |
| ۴ | Dunn | بیشینه |
| ۵ | PBM | بیشینه |
| ۶ | Point biserial | بیشینه |
| ۷ | Ratkowsky Lance | بیشینه |
| ۸ | Silhoutte | بیشینه |
| ۹ | Gamma | بیشینه |
| ۱۰ | Tau | بیشینه |
| ۱۱ | Xie Beni | کمینه |

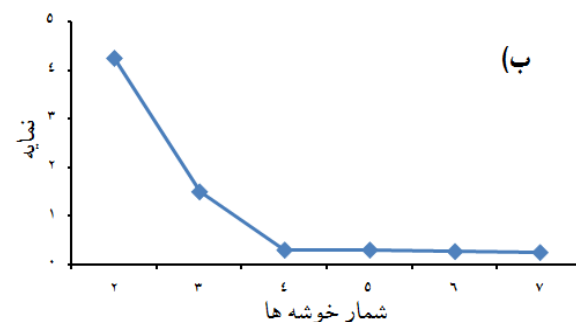
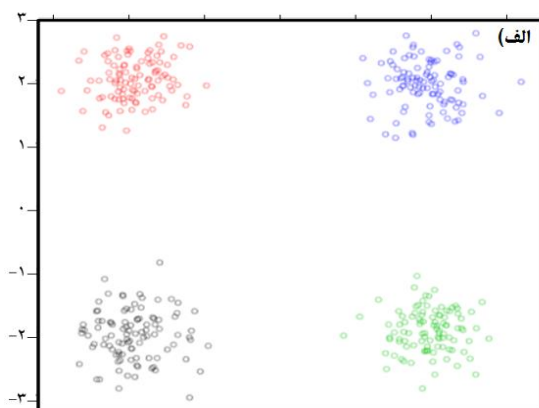
همانند نقشه پراکنش بارش بر روی سواحل جنوبی دریای خزر و رشته کوه زاگرس هسته‌های بیشینه مقادیر صدک بارش مشاهده می‌شود. میزان صدک ۹۹م بارش سالانه در جنوب غرب دریای خزر ۴۲ میلی‌متر و بر روی گستره وسیعی از ایران مرکزی برابر با ۳/۵ میلی‌متر است (شکل ۳ ج). برپایه نگاره‌های پراکنش مکانی صدک‌های بالای بارش ایران زمین می‌توان گفت که شدت رخداد بارش سنگین در مناطق مختلف ایران تفاوت‌های بسیار آشکاری با همدیگر دارد. برای نمونه در جنوب غرب دریای خزر بارش سنگین، بارشی است که به‌طور میانگین مقدار بارش دریافتی روزانه بیش از ۲۵ میلی‌متر باشد درحالی‌که در

بزرگترین اختلاف بین دو مقدار شاخص کیفیت پیاپی است. اگر بر روی یک نمودار مقادیر شاخص را در مقابل تعداد خوشه‌ها قرار دهیم شیب بین دو نقطه پیاپی نمودار بیانگر بهترین شمار خوشه‌هاست (دسکراوپس^۱، ۲۰۱۳).

$$V_i = Q_{i+1} - Q_i \quad (52)$$

$$K = \arg \max_{K_m < i < K_M} (Q_{i+1} - Q_i) \quad (53)$$

شمار خوشه‌های بهینه داده‌ها به کمک معادله ۵۳ قابل تشخیص است. برای نمونه برای نمایه Ball Hall بیشینه اختلاف max diff در شکل ۲ آمده است.



شکل ۲- پراکنش داده‌ها (الف) و نمایه Ball Hall برای خوشه‌بندی‌های مختلف (ب)

بیشترین خمیدگی (elbow) در شکل ۲ (ب) بیانگر تعداد خوشه‌های بهینه است. شکل ۲ (الف) مقادیر نمایه Ball Hall را برای خوشه‌بندی‌های مختلف داده‌های شکل ۲ (ب) نمایش می‌دهد. بر پایه آن می‌توان گفت که در مثال یاد شده، خوشه‌بندی داده‌ها به چهار خوشه، بهینه‌ترین حالت برای گزینش شمار خوشه‌هاست.

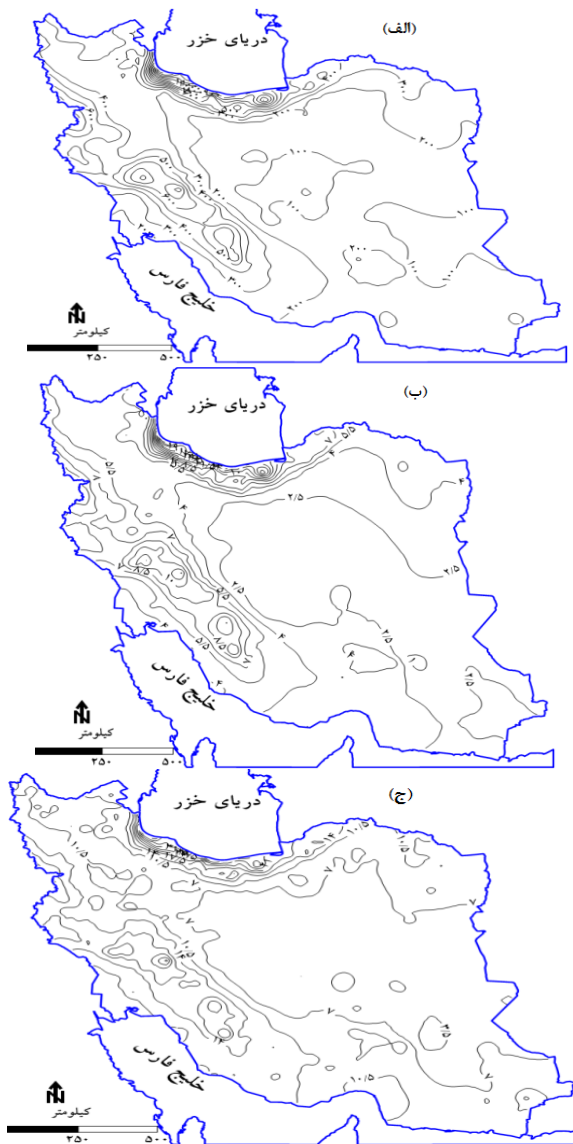
¹ Desgraupes

را می‌توان در داخل یک گروه قرار داد که با سایر اعضای داخل آن گروه شباهت بسیار زیادی داشته باشد.

پهنه‌بندی ایران برپایه صدک ۹۵م بارش (سنگین)

برای آن که تصمیم درستی در مورد شمار خوشه‌ها (تعداد نواحی) داشته باشیم، ابتدا ایران برپایه صدک ۹۵م بارش به ۲ گروه (ناحیه) تفکیک شد. سپس به ۳ گروه (ناحیه) و این کار تا ۲۰ گروه (ناحیه) ادامه داده شد. برای خوشه‌بندی‌های مختلف (از ۲ تا ۲۰ گروه)، مقادیر حاصل از برازش نمایه‌های به‌کارگرفته شده محاسبه شد (جدول ۳). برپایه جدول ۱ بهترین‌ترین شمار خوشه‌ها برای نمایه‌های مختلف گزینش شد. همانطور که قابل ملاحظه است در اغلب نمایه‌ها بهترین‌ترین شمار خوشه‌ها، ۵ خوشه (ناحیه) است. بنابراین پهنه‌بندی ایران زمین برپایه صدک ۹۵م بارش به ۵ خوشه (ناحیه) نسبت به سایر شمار خوشه‌ها مناسب‌تر است و از همگنی و همبستگی درون‌گروهی قابل قبول‌تری برخوردار است. این به آن معنی است که اگر شمار خوشه‌ها را بیشتر و یا کمتر از ۵ گروه در نظر بگیریم پراکنش مکانی نواحی پهنه‌بندی بر روی ایران زمین در شکل ۴ آمده است. به لحاظ مکانی پهنه ۱ شمال‌غرب و بلندی‌های البرز تا خراسان شمالی را شامل می‌شود. این پهنه حدود ۱۷ درصد از گستره ایران زمین را در بر دارد (جدول ۴). سری زمانی روزانه مقادیر صدک ۹۵م بارش نشان می‌دهد که در برخی از روزها در ماه‌های فروردین و اردیبهشت برپایه آستانه بارش سنگین مقدار بارش حدود ۱۴ میلی‌متر در روز است. آستانه بارش سنگین در ماه‌های فروردین و اردیبهشت بیشتر از سایر ماه‌های سال است (شکل ۵). به‌طور میانگین در ماه‌های یاد شده صدک ۹۵م بارش برابر با حدود ۹ میلی‌متر در روز است و این بیانگر مقدار و شدت بارش در ماه‌های یاد شده در این پهنه است. آستانه رخداد بارش سنگین با توجه صدک ۹۵م بر روی این پهنه در فصل تابستان نسبت به سایر پهنه‌های دیگر (به غیر از پهنه ۴) بیشتر است.

ایران مرکزی ۲/۵ میلی‌متر بارش در روز یک بارش سنگین تلقی می‌شود.

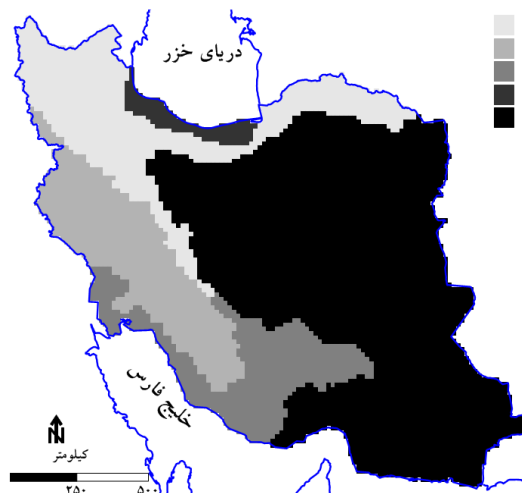


شکل ۳- میانگین (الف)، صدک ۹۵م (ب) و صدک ۹۹م (ج) بارش

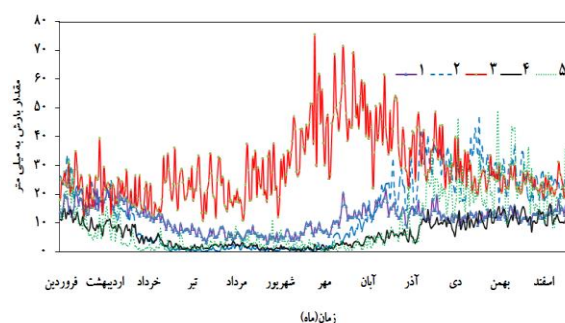
ایران زمین

اگر ایران را برپایه مقادیر بارش صدک ۹۵م (سنگین) و ۹۹م (ابرسنگین) پهنه‌بندی کنیم مناطق همگن شناسایی خواهد شد. اما پرسش اصلی آن است که ایران را برپایه صدک ۹۵م و صدک ۹۹م، به چند ناحیه همگن دسته‌بندی کنیم. همانطور که در بحث داده و روش‌شناسی گفته شد برای جواب دادن به این پرسش از نمایه‌هایی بهره گرفته شد که برپایه آن‌ها شباهت افراد داخل گروه و ناحیه مورد نظر را مورد واکاوی قرار می‌دهند. به‌طور کلی افرادی (یاخته‌هایی)

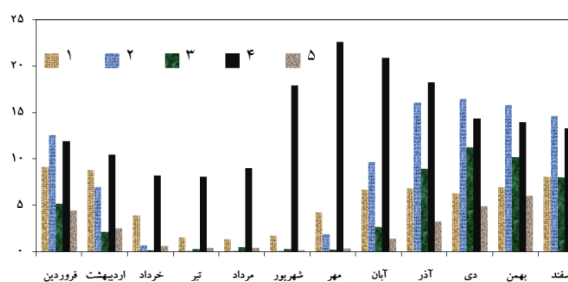
پهنه به طور میانگین صدک ۹۵م بارش طی ماه‌های آذر، دی، بهمن و اسفند بیش از ۱۵ میلی‌متر در روز است. در برخی از روزها در ماه‌های آذر و دی صدک ۹۵م بارش به ۲۵ میلی‌متر در روز می‌رسد. میزان صدک ۹۵م بارش در ماه‌های فروردین و اردیبهشت نیز قابل توجه است. در فصل تابستان بر روی این گستره از ایران زمین صدک ۹۵م بارش صفر است و هرگونه رخنمود بارشی طی این فصل از سال یک بارش سنگین نامگذاری می‌شود. در ماه‌های خرداد تا مهر میزان صدک ۹۵م بارش بر روی این پهنه در مقایسه با پهنه ۱ کمتر است و این بیانگر اهمیت بارش‌های ماه‌های یاد شده برای پهنه ۱ و مشارکت کم‌رنگ بارش‌های این ماه‌ها بر روی پهنه ۲ است. پهنه ۳ شامل مناطق پست و کم‌ارتفاع اهواز، بوشهر و بلندی‌های کرمان است. این پهنه حدود ۱۱ درصد از گستره ایران زمین را در بر دارد. مقادیر آستانه بارش سنگین بر پایه صدک ۹۵م بارش بر روی این پهنه کمتر از پهنه ۲ است. در این پهنه صدک ۹۵م بارش در ماه دی برابر با حدود ۱۳ میلی‌متر در روز است. به بیانی دیگر رخداد بارش بیشتر از ۱۳ میلی‌متر در روز در این ماه از سال بر روی این پهنه یک رویداد فرین و سنگین و کمتر از آن یک بارش عادی به حساب می‌آید. از اوایل خرداد تا اوایل آبان میزان صدک ۹۵م بارش بسیار کوچک و حدود ۰/۵ میلی‌متر در روز است. پهنه ۴ درصد بسیار کوچکی از گستره ایران زمین (۱/۹ درصد) را در بر دارد و شامل پست سواحل جنوبی دریای خزر است که به صورت نوار باریکی از رشت تا ساری است. هسته بیشینه بارش ایران زمین در این پهنه قرار دارد. آستانه‌های رخداد بارش سنگین بر پایه صدک ۹۵م بارش بر روی این پهنه بزرگتر از سایر پهنه‌های دیگر است. آستانه بارش سنگین در تمام ماه‌های سال مقدار نسبتاً بزرگی است. برخلاف سایر پهنه‌های دیگر اوج آستانه صدک ۹۵م بارش در فصل پاییز (مهرماه) است. به طور میانگین آستانه بارش سنگین در مهرماه به ۲۳ میلی‌متر در روز می‌رسد. در این ماه از سال در برخی روزها آستانه بارش حدود ۳۵ میلی‌متر در روز است. در ماه‌های گرم سال آستانه بارش سنگین بیشتر از آستانه بارش سنگین در ماه‌های بارشی (زمستان) سایر پهنه‌های دیگر است. رخداد بارش‌های سنگین به‌ویژه در پهنه‌های ۱ و ۵ بر روی این گستره از



شکل ۴- پهنه‌های بارش سنگین ایران زمین بر پایه صدک ۹۵م

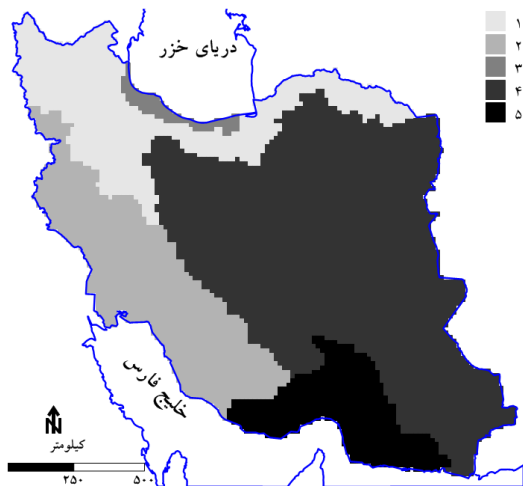


شکل ۵- مقادیر صدک ۹۵م بارش روزانه پهنه‌های مختلف طی ماه‌های مختلف سال

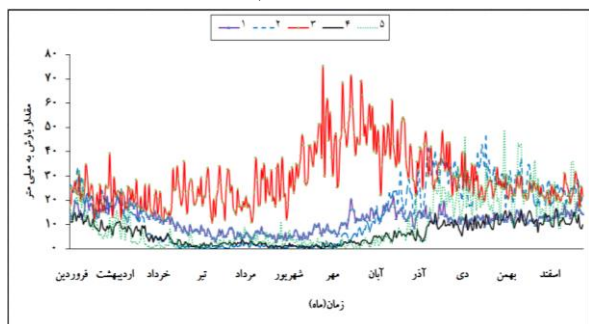


شکل ۶- میانگین بارش پهنه‌های مختلف طی ماه‌های مختلف سال به میلی‌متر

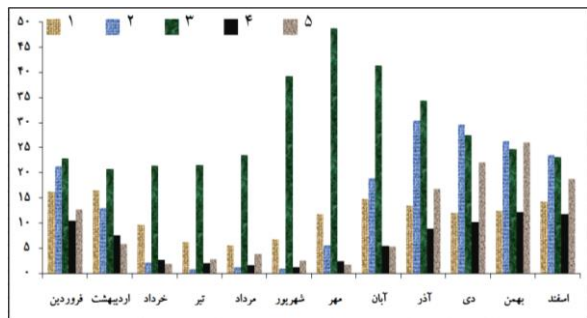
طی فصل تابستان به طور میانگین حدود ۲ میلی‌متر بارش روزانه بر روی این گستره از ایران زمین یک بارش سنگین تلقی می‌شود (شکل ۶). پهنه ۲ رشته‌کوه‌های زاگرس از شمال شهرستان بانه تا جنوب شیراز را در بر دارد. حدود ۱۴ درصد از گستره ایران زمین را شامل می‌شود. این پهنه منطبق بر هسته‌های بارش روی زاگرس است. بر روی این



شکل ۷- پهنه‌های بارش اثر سنگین ایران زمین برپایه صدک ۹۹ام



شکل ۸- مقادیر صدک ۹۹ام بارش روزانه پهنه‌های مختلف طی ماه‌های مختلف سال



شکل ۹- میانگین صدک ۹۹ام بارش پهنه‌های مختلف طی ماه‌های مختلف سال به میلی‌متر

بیشترین آستانه بارش ابرسنگین برپایه صدک ۹۹ام در بین ماه‌های مختلف سال مربوط به ماه‌های فروردین و اردیبهشت است. به‌طور میانگین آستانه اینگونه بارش‌ها در ماه‌های یاد شده به ۱۷ میلی‌متر در روز می‌رسد (شکل ۸). در برخی روزهای این دو ماه آستانه بارش ابرسنگین به ۲۴ میلی‌متر در روز می‌رسد (شکل ۷). در ماه‌های آبان تا اسفند آستانه رخداد اینگونه بارش‌ها تا حدودی یکسان و برابر است.

ایران زمین یک رویداد بارشی عادی و معمول به حساب می‌آید. پهنه ۵ حدود ۵۶ درصد از گستره ایران زمین را شامل می‌شود. هسته‌های کم‌بارش ایران در این پهنه قرار دارند. مقدار دریافتی بارش بر روی این پهنه به‌طور میانگین در سال برابر و کوچک‌تر از ۱۰۰ میلی‌متر است. بیشترین آستانه بارش سنگین بر روی این پهنه مربوط به ماه‌های بهمن و اسفند است و مقدار آن حدود ۶ میلی‌متر در روز است. به بیانی دیگر بارش‌های برابر و بیش‌تر از مقدار یاد شده در ماه‌های یاد شده برای این گستره یک رویداد فرین است. در سایر ماه‌های دیگر سال آستانه صدک ۹۵ام بارش کمتر از ۵ میلی‌متر در روز است. در ماه‌های گرم سال آستانه بارش سنگین بسیار کوچک است و به بیانی دیگر رخداد هرگونه بارشی با شدت بیشتر از ۱ میلی‌متر در روز یک پدیده فرین بارشی تلقی می‌شود.

پهنه‌بندی ایران برپایه صدک ۹۹ام بارش (ابرسنگین)

مراحلی را که بر روی مقادیر صدک ۹۵ام بارش یاد شد بر روی مقادیر صدک ۹۹ام بارش نیز انجام شد. برپایه نتایج حاصل از انجام ۱۱ نمایه مورد بررسی بر روی خوشه‌بندی‌های مختلف می‌توان ایران را به ۵ ناحیه همگن متفاوت دسته‌بندی کرد (جدول ۵). به لحاظ پراکنش مکانی تفاوت بسیار زیادی بین نواحی مختلف حاصل از پهنه‌بندی ایران برپایه صدک ۹۵ام و صدک ۹۹ام بارش دیده نمی‌شود (شکل ۷). در جنوب ایران پهنه ۲ و ۵ام صدک ۹۹ام، به لحاظ مکانی با پهنه‌بندی بارش برپایه صدک ۹۵ام متفاوت است. مقادیر بارش برپایه صدک ۹۹ام بیشتر از مقادیر صدک ۹۵ام است و می‌توان این دسته از بارش‌ها را در گروه بارش ابرسنگین قرار داد. پهنه ۱ شمال‌غرب و بلندی‌های البرز تا خراسان شمالی را به‌صورت نوار باریکی در بر دارد. این پهنه ۱۷ درصد از گستره ایران زمین را پوشانده است (جدول ۴).

طی فصل تابستان بر روی این پهنه آستانه رخداد بارش‌های ابرسنگین به‌طور میانگین ۵ میلی‌متر در روز است. پهنه ۲ شامل مناطق جنوب ارومیه و شمال غرب استان کردستان تا جنوب شیراز است و ۱۹/۵ درصد از گستره ایران‌زمین را در بر دارد. در ماه‌های آذر و دی آستانه رخداد بارش ابرسنگین حدود ۳۰ میلی‌متر در روز است. در ماه‌های گرم سال آستانه رخداد بارش ابرسنگین از همه نواحی دیگر کمتر است. به

بیانی دیگر در این ماه‌های سال رخداد بارش اندک است و هر گونه بارشی با شدت بیش از ۱ میلی‌متر در روز یک بسامد سنگین و ابرسنگین بارشی به حساب می‌آید. سواحل جنوبی دریای خزر تا بلندی‌های البرز در پهنه ۳ جای می‌گیرند. آستانه بارش ابرسنگین بر روی این گستره از ایران-زمین در ماه‌های مختلف سال با هم متفاوت است.

جدول ۳- مقادیر نمایه‌ها برای خوشه‌بندی‌های مختلف پهنه‌بندی ایران برپایه صدک ۹۵ام بارش

| شمار خوشه‌ها | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ | ۱۰ | ۱۱ | ۱۲ | ۱۳ | ۱۴ | ۱۵ | ۱۶ | ۱۷ | ۱۸ | ۱۹ | ۲۰ |
|-------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Ball Hall C_Index | ۶۱۴۸ | ۶۲۵۵ | ۶۸۷۹ | ۶۸۱۴ | ۶۸۹۶ | ۶۰۰۲ | ۵۲۷۴ | ۵۳۲۰ | ۶۷۵۸ | ۴۳۶۶ | ۴۴۴۰ | ۴۲۷۵ | ۴۰۰۲ | ۳۸۴۵ | ۳۷۷۱ | ۳۵۵۹ | ۳۸۰۰ | ۳۶۹۸ | ۳۶۸۵ | ۳۶۲۱ |
| Calinski Harabasz | ۳۳۸۷ | ۳۳۸۷ | ۳۳۳۷ | ۱۹۱۳ | ۲۰۶۲ | ۱۷۹۰ | ۱۵۹۱ | ۱۵۳۴ | ۱۴۰۴ | ۱۳۸۹ | ۱۲۵۳ | ۱۱۸۱ | ۱۱۰۶ | ۱۰۵۴ | ۱۰۰۷ | ۹۵۳ | ۹۶۱ | ۹۲۸ | ۹۰۵ | ۸۸۰ |
| Dunn | ۳۶۱۹ | ۳۶۱۹ | ۳۹۷۵ | ۲۵۱۶ | ۴۹۷۰ | ۳۶۹۵ | ۳۰۹۱ | ۳۴۹۲ | ۲۰۴۱ | ۱۷۰۶ | ۱۴۵۵ | ۱۲۵۴ | ۱۰۹۳ | ۹۶۵ | ۸۵۶ | ۸۱۸ | ۱۶۷۵ | ۱۵۲۱ | ۱۳۸۹ | ۱۲۷۶ |
| PBM | -۲۱ | -۲۱ | -۲۱ | -۳۲ | -۳۳ | -۲۶ | -۱۹ | -۲۰ | -۲۰ | -۱۹ | -۲۰ | -۲۰ | -۱۹ | -۱۹ | -۱۹ | -۱۶ | -۱۶ | -۱۶ | -۱۷ | -۱۷ |
| Point biserial | ۰/۳۰۶ | ۰/۳۰۶ | ۰/۳۰۲ | ۰/۲۹۲ | ۰/۳۲۵ | ۰/۳۰۴ | ۰/۳۸۴ | ۰/۳۷۰ | ۰/۴۵۷ | ۰/۳۴۵ | ۰/۳۳۶ | ۰/۳۲۷ | ۰/۳۱۹ | ۰/۳۱۴ | ۰/۳۰۸ | ۰/۳۰۳ | ۰/۲۹۹ | ۰/۲۹۴ | ۰/۲۹ | ۰/۱۸۶ |
| Raikowsky Lance | ۰/۳۲۵ | ۰/۳۲۵ | ۰/۳۳۶ | ۰/۱۵۲ | -۰/۱۷۳ | ۰/۱۳۱ | ۰/۱۰۸ | ۰/۱۰۷ | ۰/۱۱۸ | ۰/۱۱۹ | ۰/۱۱۹ | ۰/۱۱۷ | ۰/۱۱۵ | ۰/۱۱۳ | ۰/۱۱۱ | ۰/۱۰۹ | ۰/۱۱۹ | ۰/۱۲۳ | ۰/۱۱۹ | ۰/۱۱۶ |
| Silhouette | -۰/۵۲۱ | -۰/۵۲۱ | -۰/۷۳۷ | -۰/۸۰۳ | -۰/۸۲۷ | -۰/۷۴۱ | -۰/۶۱۱ | -۰/۶۷۳ | -۰/۶۸۱ | -۰/۶۷۳ | -۰/۷ | -۰/۷۲۵ | -۰/۷۱۱ | -۰/۷۴۳ | -۰/۷۸۲ | -۰/۷۲۹ | -۰/۷۳۱ | -۰/۷۴۳ | -۰/۷۶۱ | -۰/۷۷۲ |
| Gamma | -۰/۳۶۸ | -۰/۳۶۸ | -۰/۵۱۵ | -۰/۵۵۲ | -۰/۵۶۷ | -۰/۴۵۷ | -۰/۳۲۴ | -۰/۳۴۹ | -۰/۳۴۴ | -۰/۳۲۸ | -۰/۳۳۷ | -۰/۳۴۲ | -۰/۳۲۵ | -۰/۳۲۹ | -۰/۳۳۴ | -۰/۳۸۵ | -۰/۳۸۶ | -۰/۳۸۸ | -۰/۳۹۲ | -۰/۳۹۵ |
| Tau | ۳۳ | ۳۳ | ۶۸ | ۶۱ | ۵۱ | ۱۳۹ | ۳۲۰ | ۳۹۹ | ۴۹۱ | ۲۸۵ | ۲۷۲ | ۲۶۵ | ۲۶۲ | ۲۵۶ | ۲۵۲ | ۲۵۷ | ۲۴۵ | ۲۴۰ | ۲۳۵ | ۲۳۱ |

در مهرماه آستانه رخداد بارش ابرسنگین بیشینه است و به‌طور میانگین حدود ۵۰ میلی‌متر در روز است. در برخی روزهای اوایل و اواخر این ماه از سال آستانه رخداد اینگونه

بارش‌ها به ۷۰ میلی‌متر نیز می‌رسد. بارش‌های ابرسنگین اواخر ماه شهریور منجر به افزایش آستانه رخداد بارش‌ها شده است. در ماه شهریور به‌طور میانگین آستانه رخداد بارش‌های

داده اسفزاری طی بازه زمانی ۱۳۴۰/۱/۱ تا ۱۳۸۳/۱۰/۱۱ (۱۵۹۹۲×۷۱۸۷) بهره گرفته شد. برای هر روز تقویمی از سال بر روی هر یاخته، میانگین بلندمدت آستانه‌های صدک ۹۵م و ۹۹م بارش محاسبه گردید. برای محاسبه همانندی و ناهمانندی مقادیر بارش سنگین و ابرسنگین بر روی نقاط مختلف بر روی پهنه ایران زمین از فاصله اقلیدسی بهره بردیم. از تحلیل خوشه‌ای پایگانی به روش ادغام وارد جهت شناسایی نواحی همگن و برای شناسایی شمار پهنه‌های بهینه، ۱۱ نمایه کنترل کیفیت خوشه‌بندی با ارزیابی پراش درون‌گروهی به‌کارگرفته شد. برازش نمایه‌های یاد شده بر روی خوشه‌بندی‌های مختلف (از ۲ تا ۲۰ خوشه) نشان داد که پهنه‌بندی ایران بر پایه آستانه‌های بارش سنگین و ابرسنگین به ۵ ناحیه همگن بهینه است. به نظر می‌رسد که شمار خوشه‌های برگزیده شده با واقعیت هماهنگی مناسبی نشان می‌دهند. مقادیر آستانه‌های ۹۵م و ۹۹م صدکی بارش روزانه و میانگین ماهانه آن‌ها در پهنه‌ها طی ماه‌های مختلف سال گویای این تفاوت‌هاست. مناطقی که پربارش هستند مقادیر آستانه‌های بارش فرین بیشتر و مناطق کم‌بارش مقادیر آستانه‌ها کوچک‌تر است. در ماه‌های مختلف سال آستانه‌های بارش فرین متفاوت است. اوج آستانه‌های فرین بارش پهنه‌های مختلف نیز به لحاظ زمانی یکسان نیست. در پهنه خزری بیشینه آستانه فرین بارشی مربوط به مهرماه است درحالی‌که در پهنه شمال‌غرب- شمال‌شرق اوج آستانه فرین بارشی در فروردین رخنمود دارد. بنابراین شدت بارش بر روی مناطق مختلف ایران در ماه‌های مختلف سال بار معنایی یکسانی ندارد. برای سواحل جنوبی دریای خزر در ماه‌های خرداد تا شهریور بارشی سنگین نامگذاری می‌شود که مقدار آن حداقل ۸ میلی‌متر در روز باشد درحالی‌که طی ماه-های یاد شده برای مناطق مرکزی، جنوبی و غربی کشور، بارشی سنگین در نظر گرفته می‌شود که شدت آن برابر و بیشتر از ۱ میلی‌متر در روز باشد. بنابراین داوری و قضاوت در مورد شدت رخداد بارش هنگامی برحق است که به زمان و مکان رخداد آن توجه داشته باشیم.

ابرسنگین بیشتر از ماه‌های فصل زمستان است و این خود بیانگر ساز و کار متفاوت رخداد بارش در ماه‌های یاد شده بر روی پهنه یاد شده است.

جدول ۴- گستره پهنه‌های مختلف بر حسب درصد

| | گستره به درصد | |
|--------|-----------------|-----------------|
| | بر پایه صدک ۹۹م | بر پایه صدک ۹۵م |
| پهنه ۱ | ۱۶/۹ | ۱۷ |
| پهنه ۲ | ۱۲/۸ | ۱۹/۵ |
| پهنه ۳ | ۱۱/۱ | ۱/۳ |
| پهنه ۴ | ۱/۹ | ۵۲/۵ |
| پهنه ۵ | ۵۶/۳ | ۹/۸ |

پهنه ۴ از جنوب‌شرق تا شمال‌شرق کشور و از شرق تا پایکوه‌های شرقی رشته کوه‌های زاگرس را در بر دارد. گستره آن ۵۲/۵ درصد از وسعت ایران را می‌پوشاند. میزان بارش دریافتی این پهنه بسیار اندک است. بنابراین آستانه‌های رخداد بارش ابرسنگین نیز کوچک است. به بیانی دیگر بارش‌هایی با شدت اندک نیز برای این پهنه یک رویداد ابرسنگین به شمار می‌روند. اوج آستانه بارش ابرسنگین در ماه‌های بهمن و اسفند است که به‌طور میانگین میزان آن حدود ۱۱ میلی‌متر در روز است. از ماه خرداد تا آبان آستانه رخداد اینگونه بارش‌ها بسیار کوچک است و حدود ۳ میلی-متر در روز است. در پهنه ۵ آستانه رخداد بارش‌های ابرسنگین نسبت به پهنه ۴ بالاتر است. این پهنه بلندی‌های کرمان تا بندرعباس و سواحل جنوبی ایران را در بر دارد. گستره مکانی این پهنه ۹/۸ درصد از وسعت ایران زمین است. آستانه رخداد بارش ابرسنگین بر روی این پهنه همانند سایر پهنه‌های دیگر طی ماه‌های مختلف سال با هم متفاوت است. اوج آستانه برای بارش‌های یاد شده در ماه بهمن دیده می‌شود و میزان آن حدود ۲۷ میلی‌متر در روز است. در برخی روزهای ماه یاد شده آستانه رخداد بارش ابرسنگین به بیش از ۴۵ میلی‌متر در روز نیز می‌رسد.

نتیجه‌گیری

در این پژوهش برای شناسایی پهنه‌های همگن بارش سنگین و ابرسنگین ایران زمین از داده‌های یاخته‌ای پایگاه

جدول ۵- مقادیر نمایه‌ها برای خوشه‌بندی‌های مختلف پهنبندی ایران بر پایه صدک ۹۹ام بارش

| نمایه | Ball Hall | C ₁ Index | Galinski Harabasz | Dunn | PBM | Point Biserial | Ratkowsky Lance | Silhouette | Gamma | Tau | Xie Beni |
|-------|-----------|----------------------|-------------------|-------|-------|----------------|-----------------|------------|--------|--------|----------|
| ۲ | ۲۲۷۹۹ | ۰/۲۱۵ | ۱۵۳۹ | ۰/۰۱۵ | ۹۰۹۳ | -۴۶ | ۰/۳۳۵ | ۰/۱۹۴ | -۰/۶۵۳ | -۰/۴۵۴ | ۱۳۳ |
| ۳ | ۲۰۳۸۱ | ۰/۱۸۵ | ۱۳۶۶ | ۰/۰۱۵ | ۵۳۹۶ | -۴۵ | ۰/۳۷۷ | ۰/۱۰۹ | -۰/۵۹۴ | -۰/۴۱۰ | ۱۱۹ |
| ۴ | ۲۸۵۴۲ | ۰/۱۳۱ | ۱۱۴۲ | ۰/۰۱۵ | ۴۸۸۶ | -۵۱ | ۰/۳۶۴ | ۰/۱۱۱ | -۰/۷۰۱ | -۰/۴۷۵ | ۱۰۹ |
| ۵ | ۲۵۲۱۵ | ۰/۱۱۳ | ۱۱۲۰ | ۰/۰۰۲ | ۱۶۵۰۲ | -۵۴ | ۰/۲۷۲ | ۰/۱۳۳ | -۰/۷۳۳ | -۰/۴۸۸ | ۹۹ |
| ۶ | ۲۱۱۴۲ | ۰/۱۷۸ | ۹۶۳ | ۰/۰۱۳ | ۱۴۹۰۰ | -۳۶ | -۰/۲۵۶ | ۰/۰۹۴ | -۰/۵۴۲ | -۰/۳۱۵ | ۲۲۵ |
| ۷ | ۲۰۹۱۵ | ۰/۱۳۹ | ۹۰۸ | ۰/۰۱۳ | ۹۸۷۷ | -۴۰ | ۰/۳۴۵ | ۰/۰۹۲ | -۰/۶۳۴ | -۰/۳۵۶ | ۲۱۴ |
| ۸ | ۲۸۵۶۹ | ۰/۱۲۸ | ۸۳۴ | ۰/۰۱۳ | ۸۷۴۲ | -۳۹ | ۰/۳۳۵ | ۰/۰۸۴ | -۰/۶۶ | -۰/۳۵۳ | ۲۰۸ |
| ۹ | ۲۷۱۲۲ | ۰/۱۲۲ | ۷۶۹ | ۰/۰۱۳ | ۶۳۵۶ | -۴۰ | ۰/۳۲۴ | ۰/۰۹۴ | -۰/۶۷۸ | -۰/۳۵۹ | ۲۰۳ |
| ۱۰ | ۲۷۲۶۹ | ۰/۱۰۷ | ۷۳۶ | ۰/۰۱۳ | ۵۲۸۲ | -۴۱ | ۰/۲۱۶ | ۰/۰۹۶ | -۰/۷۱۵ | -۰/۳۷۳ | ۱۹۶ |
| ۱۱ | ۲۶۴۰۸ | ۰/۱ | ۶۹۶ | ۰/۰۱۳ | ۴۴۶۴ | -۴۲ | ۰/۲۰۸ | ۰/۱۰۶ | -۰/۷۳۳ | -۰/۳۸ | ۱۹۱ |
| ۱۲ | ۲۴۵۶۴ | ۰/۱۳۸ | ۶۶۷ | ۰/۰۱۳ | ۴۰۰۵ | -۳۳ | ۰/۲ | ۰/۰۹ | -۰/۶۳۱ | -۰/۲۹۷ | ۲۴۰ |
| ۱۳ | ۲۳۲۵۱ | ۰/۱۳۷ | ۶۰۹ | ۰/۰۱۳ | ۳۴۸۱ | -۳۲ | ۰/۱۹۳ | ۰/۰۹۴ | -۰/۶۲۸ | -۰/۲۸۴ | ۲۳۷ |
| ۱۴ | ۲۳۵۵۷ | ۰/۱۲ | ۵۸۹ | ۰/۰۱۳ | ۳۰۵۷ | -۳۳ | ۰/۱۹ | ۰/۰۸۲ | -۰/۶۷۴ | -۰/۲۹۶ | ۲۳۲ |
| ۱۵ | ۲۲۲۷۰ | ۰/۱۴۶ | ۵۵۶ | ۰/۰۱۳ | ۲۷۱۱ | -۳۷ | ۰/۱۸۴ | ۰/۰۷۹ | -۰/۵۹۴ | -۰/۳۲۸ | ۲۳۰ |
| ۱۶ | ۲۱۳۸۵ | ۰/۱۴۳ | ۵۳۴ | ۰/۰۱۳ | ۲۴۲۷ | -۳۷ | ۰/۱۷۹ | ۰/۰۷۷ | -۰/۶۰۲ | -۰/۳۳۵ | ۲۲۶ |
| ۱۷ | ۲۰۵۹۶ | ۰/۱۴ | ۵۱۶ | ۰/۰۱۳ | ۲۱۹۰ | -۳۷ | ۰/۱۷۵ | ۰/۰۸۹ | -۰/۶۰۷ | -۰/۳۳۶ | ۲۲۳ |
| ۱۸ | ۲۱۰۰۰ | ۰/۱۳۲ | ۵۰۵ | ۰/۰۱۳ | ۱۹۸۵ | -۳۷ | ۰/۱۷۱ | ۰/۰۹۲ | -۰/۶۳ | -۰/۳۴۲ | ۲۱۸ |
| ۱۹ | ۲۰۴۶۸ | ۰/۱۲۹ | ۴۸۸ | ۰/۰۱۳ | ۱۸۰۵ | -۳۷ | ۰/۱۶۸ | ۰/۰۹۵ | -۰/۶۳۷ | -۰/۳۴۱ | ۲۱۵ |
| ۲۰ | ۲۰۰۹۳ | ۰/۱۱۸ | ۴۷۲ | ۰/۰۱۲ | ۱۶۵۱ | -۳۶ | ۰/۱۶۵ | ۰/۰۹ | -۰/۶۶۳ | -۰/۳۳۵ | ۲۱۳ |

منابع

دین پژوه، ی.، فاخری، ا.، مقدم، م.، میرنیا، ک.، جهانبخش اصل، س. ۱۳۸۲. پهنبندی اقلیمی ایران با استفاده از تحلیل‌های چند متغیره برای استفاده در مطالعات کشاورزی، نشریه دانش کشاورزی، ۱۳: ۹۰-۷۱.

رعیتی دماوندی، م.، غروی، م.، ثقفیان، ب. ۱۳۹۰. شناسایی نواحی همگن هیدرولوژیک با استفاده از روش میان-مرکزی فازی. نشریه تحقیقات منابع آب ایران، ۲: ۴۹-۳۷.

مسعودیان، س. ا. ۱۳۸۲. نواحی اقلیمی ایران. نشریه جغرافیا و توسعه، (۲): ۱۸۵-۱۷۱.

اکبری، ط.، مسعودیان، س. ا. ۱۳۸۸. شناسایی نقش الگوهای پیوند از دور نیمکره شمالی بر دمای ایران. نشریه پژوهشی دانشگاه اصفهان، (۱) (پیاپی ۲۲): ۱۳۲-۱۱۷.

جهانبخش، س.، ترابی، س. ۱۳۸۳. بررسی و پیش‌بینی تغییرات دما و بارش در ایران. فصلنامه تحقیقات جغرافیایی، ۷۴: ص. ۱۰۴-۱۲۵.

خام‌چین مقدم، ف.، صدقی، ح.، کاوه، ف.، منشوری، م. ۱۳۸۹. پهنبندی حداکثر بارش روزانه ایران، نشریه آب و خاک (علوم و صنایع کشاورزی)، (۱) ۲۴: ۱۰۶-۹۷.

- Noto, L. V., LaLoggia, G. 2009. Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. *Water Resour. Manag.*, 23: 2207–2229.
- Kalkstein, L. S., Tan, G., Skindlov, J. A. 1987. An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Clim. Appl. Meteorol*, 26: 717-730.
- Sarkar, S., Goel, N. K., Mathur, B. S. 2009. Development of isopluvial map using L-moment approach for Tehri-Garhwal Himalaya. *Stoch. Environ. Res. Risk Assess.*, 24: 411–423.
- Weingessel, A., Dimitriadou, E., Dolnicar, S. 1999. An examination of indexes for determining the number of clusters in binary data sets. *Working Paper*, 29: 1-19.
- مسعودیان، س. ا. ۱۳۸۴. شناسایی رژیم‌های بارش ایران به روش تحلیل خوشه‌ای. نشریه پژوهش‌های جغرافیایی، ۵۲: ۴۷-۶۱.
- مسعودیان، س. ا. ۱۳۸۸. نواحی بارشی ایران. نشریه جغرافیا و توسعه، ۷(۱۳): ۷۹-۹۱.
- مسعودیان، س. ا.، دارند، م. ۱۳۸۹. شناسایی و پهنه‌بندی نواحی دمای فرین سرد ایران. نشریه مطالعات جغرافیایی مناطق خشک، ۲: ۴۳-۵۴.
- مسعودیان، س. ا.، عطایی، ه. ۱۳۸۴. شناسایی فصول بارشی ایران زمین به روش تحلیل خوشه‌ای. نشریه پژوهشی علوم انسانی دانشگاه اصفهان، ۱۸: ۱-۱۲.
- منتظری، م.، مسعودیان، س. ا. ۱۳۸۹. شناسایی الگوهای فرارفت دمایی ایران در سال‌های سرد. نشریه پژوهش‌های جغرافیای طبیعی، ۷۴: ۷۹-۹۴.
- Gordon, A. D. 1999. *Classification* Chapman and Hall/CRC, Boca Raton, FL, 2nd edition.
- Cormack, R. 1971. A review of classification. *J. Stat. Soc. Series A (General)*, 134(3): 321–367.
- Desgraupes, B. 2013. *Clustering indices*, University Paris Ouest, Lab Modal'X .
- Milligan, G. W., Cooper, M. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2): 159–179.
- Gan, G., Ma, Ch., Wu, J. 2007. *Data clustering theory, algorithms and applications*, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.
- Lance, G. N. Williams, W. T. 1967. A General Theory of Classificatory Sorting Strategies. *Computer J.*, 9(4): 373-380.
- Mearns, L. O., Schneider, S. H., Thompson, S. L., McDaniel, L. R. 1990. Analysis of climate variability in general circulation models: comparison with observations and changes in variability in $2 \times \text{CO}_2$ experiments. *J. Geophys. Res.*, 95: 20469–20490.
- Mooi, E., Sarstedt, M. 2011. *A Concise Guide to Market Research*, Springer-Verlag Berlin Heidelberg, DOI 10.1007/978-3-642-12541-6_9.



Recognition of homogeneous regions of heavy and super heavy precipitation in Iran by intergroup variance quality control indices

M. Darand^{*1}

Received: 15/12/2013

Accepted: 14/02/2015

Abstract

The aim of this study is spatial clustering of the rainfall patterns of Iran. In order to perform this research daily interpolated precipitation from Asfazari^{**} database during 21/03/1961 to 31/12/2004 has been used. Considering the country's area and study period, a database with dimension 15992×7187 has been created with time (days) as the rows and spatial location as the columns' pixels. For every calendar day of year and each pixel, heavy and super heavy precipitation have been calculated based on 95 and 99 percentiles. Besides, two data base with dimension 7187×366 obtained. Regionalization was performed using cluster analysis with Ward linkage method. For determining number of clusters, 11 variance inter groups quality control indices have been applied on the different clustering. The results showed that Iran can be clustered to five homogeneous regions of heavy and super heavy precipitation. Daily and monthly precipitation thresholds were also calculated for each region. In general, the results indicated that applied quality control indices have are suitable for spatial analysis and detecting number of homogeneous regions of heavy and super heavy precipitation across Iran.

Keywords: Regionalization, heavy and super heavy precipitation, Iran, cluster analysis

¹ Assistant professor, Dep. of climatology, Collage of Natural Resources, University of Kurdistan
(*Corresponding author email address: m.darand@uok.ac.ir)

** Named after an Iranian Climatologist.