نشریه پژوهش‌های اقلیم شناسی
سال چهاردهم | شماره پنجاه و پنجم | پاییز ۱۴۰۲
وصول: ۱۴۰۱/۰۳/۱۲ پذیرش: ۱۴۰۲/۰۶/۲۶
صص ۲۸-۱۵

خوشه‌بندی ایستگاه‌های سینوپتیک ایران براساس پارامترهای هواشناسی و جغرافیایی

وجیهه محمدی ثابت[۱]، محمد موسوی بایگی[*۲]، مهدی جباری **نوقابی**[۳]، کامران داوری[٤]
۱- دانشجوی دکتری هواشناسی کشاورزی، دانشگاه فردوسی مشهد، ایران
۲- استاد هواشناسی کشاورزی، دانشگاه فردوسی مشهد، ایران
۳- دانشیار آمار ریاضی، دانشگاه فردوسی مشهد، ایران
٤- استاد علوم و مهندسی آب، دانشگاه فردوسی مشهد، ایران

چکیده

خوشه‌بندی ابزاری است که داده‌های موجود را در گروه‌های مختلفی قرار می‌دهد. عموماً تعداد خوشه‌ها بر اساس کمترین تغییرات درون گروهی و بیشترین تغییرات برون گروهی مشخص می‌شود. منطقه مورد نظر پهنای جغرافیایی و اقلیمی ایران می‌باشد. مختصات طول، عرض، ارتفاع جغرافیایی، میانگین دما، رطوبت نسبی و مجموع بارش ماهانه ۴۲۰ ایستگاه سینوپتیک از زمان تاسیس تا سال ۲۰۱۸ در این پژوهش به کار گرفته شد. پس از بررسی، پاک‌سازی و ترمیم داده‌ها تنها ۳۷۵ ایستگاه برای حضور در ادامه پژوهش باقی ماندند. با توجه به اینکه طول دوره آماری یک عامل مهم اثرگذار در خوشه‌بندی است، ایستگاه‌ها بر حسب دوره آماری به سه دوره کمتر از ۵ سال با ۴۲ ایستگاه، ۶-۱۰ سال با ۳۳ ایستگاه و بیشتر از ۱۰ سال با ۳۰۰ ایستگاه دسته‌بندی شدند. هفت روش خوشه بندی سلسله مراتبی (۵ زیرمجموعه) و افرازی (۲زیرمجموعه) در این پژوهش استفاده شد. از ضریب همبستگی کوفنتیک، آزمون عرض سیلوئت (سایه نما) به عنوان دومعیار برای انتخاب روش خوشه‌بندی استفاده شده است. کدنویسی ها در نرم افزار آماری R انجام شد. بر اساس شاخص‌های ضریب کوفنتیک و سیلوئت بهترین تعداد و روش خوشه برای داده‌های ۵-۱ سال ۴ خوشه با روش افرازی میانه‌محور، داده‌های ۱۰-۶ سال ۵ خوشه با روش سلسله مراتبی میانگین محور و برای ایستگاه‌ها با دوره آماری بیش از ۱۰سال ۴خوشه و روش افرازی میانگین محور می‌باشد. پهنه‌بندی خوشه‌ها بر نقشه جغرافیایی ایران با استفاده از نرم‌افزار ARCGIS برای هر سه دسته رسم شد. ایستگاه‌ها با دوره آماری کمتر از ۵ سال در ۴ خوشه به شرح؛ نواحی شمالی شامل ۴ ایستگاه، نواحی شرقی و مرکزی شامل ۵ ایستگاه، نواحی جنوب‌غربی شامل ۳ ایستگاه و در نهایت نواحی شمالی و شمال غربی در دامنه‌های رشته کوه البرز و زاگرس شامل ۲۰ ایستگاه قرار گرفتند. در ایستگاه‌ها با دوره آماری ۱۰-۶ سال ایستگاه شادگان در استان خوزستان در یک خوشه، آبدانان در استان ایلام در یک خوشه، ۱۳ ایستگاه در نواحی شمال غربی در یک خوشه، ۱۱ ایستگاه در نواحی جنوب و جنوب غربی و ۷ ایستگاه در نواحی شرقی در خوشه‌های مجزا قرار گرفتند. در پرتعدادترین دسته یعنی ایستگاه‌هایی با دوره آماری بیشتر از ۱۰ سال، ۱۱۸ ایستگاه در شمال‌غربی و دامنه‌های زاگرس در یک خوشه، ۵۶ ایستگاه در نواحی جنوب و جنوب غرب، ۲۹ ایستگاه در نواحی شمال و ۹۷ ایستگاه در شرق و مرکز ایران در خوشه‌های مجزا قرار گرفتند. خوشه‌بندی‌های حاصل تناسب خوبی با ویژگی‌های جغرافیایی و هواشناسی ایران دارد.

کلید واژه‌ها: خوشه‌بندی، مختصات جغرافیایی، سینوپتیک، ایران.

* نویسنده مسئول: ۰۹۳۵۲٤٥۲۹۳۱
Email: Mousavib@um.ac.ir

مقدمه

از خوشه‌بندی برای قرار دادن داده‌ها در گروه‌های مختلفی استفاده می‌شود و نوعی فرایند یادگیری می‌باشد که در طی آن نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگرند تقسیم می‌شوند. در خوشه‌بندی متغیر هدف وجود ندارد و طبقه‌بندی به مقدار متغیر هدف نمی‌پردازد (لاروس،۲۰۰۵). خوشه‌بندی نوعی فرایند یادگیری می‌باشد که در طی آن نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگرند تقسیم می‌شوند. داده‌ها عموماً بر اساس کمترین تغییرات درون-گروهی و بیشترین تغییرات برون‌گروهی درون خوشه‌ها قرار می‌گیرند. با توجه به مدل‌های مختلف خوشه‌بندی پژوهشگران متغیرها و روش‌های مختلفی را برای خوشه‌بندی مورد استفاده قرار می‌دهند. یوردانور و همکاران (Yurdanur et al. 2003) تعداد ۱۱۳ ایستگاه اقلیم-شناسی را بر اساس دما و بارش با استفاده از ۵ روش خوشه‌بندی تقسیم‌بندی کردند و روش ward مناسب‌ترین خوشه‌بندی را برای منطقه مورد نظر ارائه داد. فهمی و همکاران (Fahmi et al. 2011) با دو روش خوشه بندی سلسله مراتبی و فازی دمای ایستگاه‌های هواشناسی ترکیه در دوره آماری ۲۰۰۶-۱۹۵۰ را خوشه بندی کردند. فرزندی و همکاران (1396) تعداد ۱۴۹ ایستگاه را با روش‌های مختلف خوشه‌بندی و بر اساس دما، رطوبت، بارش و ۳ نمایه اقلیم‌بندی خوشه‌بندی کردند که در نهایت روش افرازی میانه محور برازش خوبی را نشان داد. پوربابک و همکاران (۱۳۹۵) تعداد ۱۱۲ ایستگاه سینوپتیک از کل ایستگاه‌های هواشناسی کشور را با سه روش خوشه‌بندی فازی، کلاسیک و شبکه عصبی مصنوعی کوهنن بر اساس دما و بارش طبقه‌بندی کردند. نتایج نشان داد که روش خوشه‌بندی فازی از دقت بالاتری برخوردار است. رورده و همکاران (۱۳۹۸) داده‌های روزانه بارش، رطوبت نسبی و دمای نقطه شبنم ۶۳ ایستگاه هواشناسی کشور را با استفاده از روش نگاشت SVD کاهش بُعد داده و سپس از طریق خوشه‌بندی فازی، خوشه‌بندی کرده‌اند. نتایج این تحقیق ۶ خوشه بارشی در ایران را نشان می‌دهد. این نواحی شامل نواحی ساحلی خزر، سواحل خلیج فارس و دریای عمان، نواحی بسیار خشک مرکزی، نواحی نیمه خشک، نواحی کوهستانی و نیمه کوهستانی می‌باشد.

هدف تحقیق مشخص کردن ایستگاه‌های همگن سینوپتیک از منظر ۶ شاخصه طول، عرض و ارتفاع جغرافیایی، میانگین دما، رطوبت نسبی و مجموع بارش ماهانه با توجه به طول آماری ایستگاه می‌باشد. داده‌های مذکور از زمان تاسیس ایستگاه تا سال ۲۰۱۸ در دسترس است.

مواد و روش‌ها

منطقه مورد مطالعه و داده‌ها

در این مطالعه اطلاعات ۴۲۰ ایستگاه هواشناسی همدیدی شامل مختصات طول جغرافیایی، عرض جغرافیایی، ارتفاع جغرافیایی، میانگین دما، رطوبت نسبی و مجموع بارش ماهانه از سازمان هواشناسی دریافت شد. طول دوره آماری ایستگاه‌ها از ابتدای تاسیس ایستگاه تا سال ۲۰۱۸ میلادی می‌باشد. با توجه به اینکه مقایسه رفتار ایستگاه‌ها با طول آماری مشابه، منجر به نتایج دقیق‌تر میشود، ایستگاه‌ها به شرح جدول (۱) به سه دوره زمانی تقسیم‌بندی شده‌اند. همچنین با توجه به بررسی تمامی ایستگاه‌های سینوپتیک ایران در این پژوهش و تازه تاسیس بودن برخی از این ایستگاه‌ها، به منظور کنترل اثر زمان در خوشه‌بندی مجبور به تفکیک ایستگاه‌ها از منظر طول دوره آماری هستیم. ایستگاه‌هایی با کمتر از ۵ سال ایستگاه‌های تازه تاسیس و فاقد اثرات اقلیمی ، ۱۰-۶ سال بیانگر بهتری از ویژگی‌های آب وهوایی منطقه و بیش از ۱۰ سال دارای مشخصات بهتر آب و هوایی و اقلیمی می‌باشند. ایستگاه ها در تمامی نقاط کشور پراکنده می‌باشند.

جدول ۱-دسته‌بندی آماری ایستگاه‌ها

ردیف	دوره آماری	تعداد ایستگاه
۱	۵-۱ سال	۴۲
۲	۱۰-۶ سال	۳۳
۳	بیش از ۱۰ سال	۳۰۰

در پژوهش حاضر از ۷ روش خوشه بندی سلسله مراتبی (۵ زیرمجموعه) و افرازی (۲زیرمجموعه) استفاده شده است. منطقه مورد نظر پهنای ایران می‌باشد. داده‌های هواشناسی همه ٤٢٠ ایستگاه سینوپتیک موجود در ایران از زمان تاسیس تا سال ۲۰۱۸ در این پژوهش به کار گرفته شده است. در صورت عدم پاکسازی و ترمیم داده‌ها، داده‌ها اریب محسوب می‌شوند و دقت نتایج را کم می‌کنند.۳۷۵ ایستگاه در نهایت پس از بررسی و کنترل و ترمیم داده‌ها برای ادامه کار این پژوهش تایید شدند. ایستگاه‌ها نیز برحسب طول دوره آماری به سه دوره کمتر از ۵سال، ٦-۱۰ سال و بیشتر از ۱۰ سال دسته‌بندی شدند.

پاک‌سازی، آماده‌سازی و ترمیم داده‌ها

داده‌ها اغلب همراه با خطا هستند. این خطا شامل ثبت نادرست، خرابی دستگاه، انتقال غلط و غیره است. بررسی این موارد به پاکسازی و آماده سازی معروف است. داده‌ها پس از پاکسازی و آماده سازی قابل اطمینان بوده و می‌توان آن‌ها را برای تحلیل‌های بعدی به کار برد. مشخص‌کردن وضعیت داده‌های مفقود نیز در همین مرحله انجام می‌شود. داده‌های مفقود باید ترمیم شوند. در غیر اینصورت داده‌ها اریب محسوب می‌شوند و دقت نتایج را کم می‌کنند (ارقامی و همکاران، ۱۳۸۰). ترمیم داده‌ها با توجه به آمار خود ایستگاه، ایستگاه‌های اطراف، مراجعه به پرونده داده‌ها و کارشناسان مربوط و همچنین به‌کارگیری روش‌های ترمیم داده‌ها انجام می‌شود. اگر نتوان هیچ روشی را به کار برد، بهترین گزینه میانگین دوره آماری است (رضایی پژند، ۱۳۸۰). با توجه به دسته‌بندی ایستگاه‌ها در سه دوره آماری ۱-۵ سال، ۱۰-٦ سال و بیشتر از ۱۰ سال، ترمیم و پاکسازی داده‌های ایستگاه‌ها، با توجه به دوره آماری‌شان انجام شد. دوره‌های آماری مفقود اگر بیشتر از یکسال در ابتدای دوره آماری بودند حذف شدند، برای داده‌های گم شده بین دوره آماری از میانگین دوره آماری همان ماه در مجموع کل دوره استفاده شد. پس از ترمیم و پاکسازی از مجموع ٤٢٠ ایستگاه، تنها ۳۷۵ ایستگاه نهایی شدند که ۳۰۰ ایستگاه دارای اطلاعاتی بیش از ۱۰ سال می‌باشند، ٤٢ ایستگاه دارای دوره آماری حدفاصل ۱۰-٦ سال و ۳۳ ایستگاه نیز دارای دوره آماری ۵-۱ سال می‌باشند.

متغیرهای خوشه‌بندی

در این پژوهش معیارهای تعیین کننده خوشه‌بندی اطلاعات هواشناسی ایستگاه‌ها شامل میانگین دمای ماهانه، مجموع بارش ماهانه و میانگین رطوبت نسبی ماهانه و مختصات جغرافیایی شامل طول جغرافیایی، عرض جغرافیایی و ارتفاع از سطح دریا می‌باشد. این متغیرها توصیف خوبی از یک ایستگاه را ارائه می‌دهند. طول و عرض جغرافیایی بیانگر موقعیت ایستگاه و ارتفاع یک شاخص توپوگرافی مهمی برای خوشه‌بندی می‌باشد که خود به عنوان یک متغیر مستقل نقش ایفا کرده و اثرگذار بر مقادیر سایر متغیرهای آب و هواشناسی می‌باشد. دما، بارش و رطوبت نسبی بیانگر وضعیت آب و هوای هر منطقه و مبنای تعریف و تولید شاخص‌های مختلف اقلیم شناسی می‌باشد. برای شروع کار خوشه بندی، مقادیر میانگین دوره آماری دما، بارش و رطوبت نسبی محاسبه شده و در فایل اکسل جداگانه با مشخصات ایستگاه‌ها ثبت شد. صحت سنجی مجدد محاسبه‌های صورت گرفته با فرمول نویسی در اکسل صورت گرفت.

مدل‌های خوشه‌بندی

در هر گروه از داده‌ها تعداد خوشه‌ها می‌تواند متغیر باشد. روش‌های گوناگونی برای خوشه‌بندی وجود دارد. معروف‌ترین و پرکاربردترین روش‌های خوشه بندی شامل روش‌های سلسله مراتبی، افرازی، پیوندی و فازی می‌باشد (James et.al, 2013). معیار در خوشه‌بندی فاصله متریک در فضای توپولوژی می‌باشد. خوشه بندی سلسله مراتبی خود به دو شاخه اصلی تجمیعی و تقسیمی تفکیک می‌شود. روش تجمیعی تحلیل را با خوشه‌های تک عضوی آغاز و خوشه‌های تک عضوی مشابه را با تابع فاصله مناسب ترکیب می‌کند. در روش تقسیمی ابتدا تمام داده‌ها را در یک خوشه جای می‌دهد و داده‌های مشابه را تفکیک و خوشه را تقسیم می‌کند، عمل خوشه‌بندی تا حصول نتیجه نهایی ادامه می‌یابد (کانتی ماردیا و همکاران، ۱۹۷۹). خوشه بندی سلسله مراتبی

تجمیعی دارای پنج روش تک محوری (single)، میانگین محور(average)، پیوند کامل (complete)، Ward و Ward2 می‌باشد (Romesburg, 2004). روش میانگین محور هر خوشه را با مرکز ثقل بررسی می‌کند، که میانگین بردارهای مشخصه درون خوشه است این روش به علت کفایت در خوشه بندی، برای داده‌هایی با حجم زیاد و با ویژگی های عددی شناخته می شود. این روش نسبت به داده های پرت حساسیت دارد. خوشه بندی افرازی به چند روش میانه محور، میانگین محور و غیره تقسیم می‌شود. روش میانه محور، میانه هرخوشه را به عنوان نماینده در نظر می‌گیرد. چون میانه یک برآوردگر استوار است لذا به داده‌های پرت حساسیت کمتری دارد. خوشه بندی پیوندی از ترکیب دو روش سلسله مراتبی و افرازی استفاده می کند. شاخصه مناسب بودن تعداد خوشه‌بندی آزمون عرض سیلوئت میباشد. مقدار سیلوئت معیار میزان شباهت یک شی به خوشه خودش در مقایسه با خوشه‌های دیگر است .

محدوده شاخص سیلوئت از ۱- تا ۱+ است که هرچه شاخص عرض سیلوئت به عدد ۱ نزدیکتر باشد ایده آل تر است و نشان از مطابقت خوب شی با خوشه خود دارد(Rao and Srinivas,2008).

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

a(i) متوسط فاصله بردار ویژگی iام از سایر بردارهای ویژگی در خوشه kام است. b(i) ام نیز کمترین فاصله بردار ویژگی iام از سایر بردارهای ویژگی در خوشه های دیگر است.

ضریب همبستگی کوفنتیک شاخصه بررسی مناسب بودن خوشه بندی در روش‌های سلسله مراتبی و ward میباشد. این ضریب برای اعتبارسنجی الگوریتم سلسله مراتبی به کار می‌رود و هرچه به عدد یک نزدیک تر باشد مناسب تر است و مشابه ضریب تعیین در رگرسیون عمل میکند (Romesburg, 2004).

$$CPCC = \frac{\left(1/M \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} d_{ij}^{P} c_{ij} - \mu P \mu_c\right)}{\sqrt{\left[\left(\frac{1}{M}\right)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(d_{ij}^{P}\right)^2 - \mu_P^2\right]\left[\left(\frac{1}{M}\right)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} c_{ij}^2 - \mu_c^2\right]}}$$

ضریب کوفنتیک از رابطه ذیل به دست می آید و در این رابطه μ_p و μ_c به ترتیب میانگین های عناصر مجاور ماتریس کوفنتیک در حالیکه d_{ij}^{P} و c_{ij} به ترتیب i, jامین عناصر مجاور ماتریس های کوفنتیک هستند. روش‌های خوشه‌بندی به کار گرفته شده در این پژوهش هفت روش سلسله مراتبی تجمیعی میانگین محور، تکی، پیوند کامل، افرازی میانه و میانگین محور، ward و ward.2 می‌باشد.

روش انجام

کدنویسی ها در نرم افزار آماری R 4.0.1 و بااستفاده از پکیج Hybridclust، Pvcclust و Clv انجام شد. (Everitt, and Hothorn, 2006). پس از تایید نهایی داده‌ها، فایل مورد نظر برای خوشه‌بندی بصورت میانگین متغیرها در دوره آماری برای تمامی ایستگاه‌ها آماده شد. این فایل به عنوان ورودی نرم افزار R برای خوشه بندی قرار گرفت. برای ترسیم خروجی خوشه‌بندی ها بر پهنه جغرافیایی ایران از نرم افزار ArcMap استفاده شده است.

با توجه به نمودار خوشه‌بندی بهترین عدد برای خوشه بندی مشخص شد. سپس ضرایب سیلوئت و کوفنتیک برای یک واحد بیشتر و یک واحد کمتر از عدد خوشه بندی مشخص شده مورد بررسی قرار می گیرد. هرچه ضرایب سیلوئت و کوفنتیک با توجه به روش‌های خوشه‌بندی به ۱ نزدیکتر باشند آن روش دقیق‌تر می‌باشد. البته لازم به ذکر است که در تعیین روش خوشه‌بندی علاوه بر استفاده از این ضرایب از اطلاعات آب و هواشناسی (بررسی شرایط آب و هوایی و توپوگرافی منطقه و ...) نیز استفاده شده است. لذا بر این اساس عدد صحیح خوشه بندی تعیین و روش‌های خوشه بندی انجام شد.

نتایج و بحث

خوشه بندی داده های ۵-۱ سال

با توجه به نمودارها و خروجی‌ها بهترین تعداد خوشه برای داده‌های ۵-۱ سال ٤ خوشه با عرض سیلوئت ۰/۳ می‌باشد. عرض سیلوئت برای تعداد خوشه های کمتر و بیشتر از ٤، کمتر از ۰/۳۰ شد، لذا تعداد خوشه‌ها برای این دوره مذکور برابر ٤ می‌باشد. سایر روش‌ها برای این ایستگاه‌ها مورد بررسی قرار گرفته است که نتایج آن در جدول (۳) آمده است. همانطور که مشخص است ضریب کوفنتیک در روش‌های سلسله مراتبی بالاست. با بررسی خوشه‌بندی‌های حاصل از روش سلسله مراتبی مشخص شد که از ٤۲ ایستگاه موجود تعداد کمی ایستگاه در سه خوشه و تعداد زیادی در یک خوشه قرار گرفته‌اند. به عنوان مثال در روش سلسله مراتبی تکی ۵ (۱،۲،۲) و میانگین محور ۷ ایستگاه (۱،۲،٤) در سه خوشه و مابقی ایستگاه‌ها در یک خوشه قرار گرفتند. تعداد کمی ایستگاه در یک خوشه نمی‌توانند بیانگر شرایط آب و هوایی یک منطقه باشند. لذا این خوشه‌بندی‌ها حذف شدند. در روش سلسله مراتبی کامل ایستگاه‌هایی از شمال و جنوب ایران در یک خوشه قرار گرفتند که از نظر آب و هواشناسی فاقد اعتبار است. لذا بهترین روش و تعداد خوشه برای ایستگاه‌ها با دوره آماری کمتر از ۵ سال ٤ خوشه با روش میانه محور می‌باشد. همانطور که در سایر تحقیقات (کدام تحقیقات و مطالعات) در آب و هواشناسی روش میانه محور نتایج دقیق‌تری را نیز ارائه داده است. در جدول (٤) نام ایستگاه‌ها به تفکیک خوشه‌ها مشخص شده است. آمار توصیفی متغیرهای مورد بررسی به تفکیک خوشه‌ها محاسبه و در جدول (۵) آورده شده است.

برای نمایش بهتر و واضح تر موقعیت ایستگاه‌ها نسبت به یکدیگر و خوشه بندی بر پهنه جغرافیایی ایران از نرم افزار ArcMap استفاده شده است.

همانطور که از جدول (۵) مشخص است بیشینه بارش در این خوشه مربوط به ایستگاه شول آباد در استان چهارمحال بختیاری با ارتفاع ۱۵۹۱ متر و کمینه بارش مربوط به ایستگاه زابلی در استان سیستان و بلوچستان با میانگین بارش سالانه ۱۵.۲٤ میلیمتر. و ارتفاع ۱۲۷۱ متر می‌باشد. همچنین کمینه و بیشینه میانگین دمای ماهانه مربوط به ایستگاه‌های چاراویماق در استان آذربایجان شرقی (۱۰.٦) و بندرگناوه در استان بوشهر(۲۷.٤۹) درجه سانتی‌گراد می باشد. در این خوشه-بندی کمترین و بیشترین ارتفاع به ترتیب مربوط به ایستگاه بندرگز در استان گلستان با ۱٤- و چادگان در استان اصفهان با ۲۱۷۵ متر می باشد.

جدول ۲- خوشه‌بندی ایستگاه‌های ۵-۱ سال با شاخص عرض سیلوئت

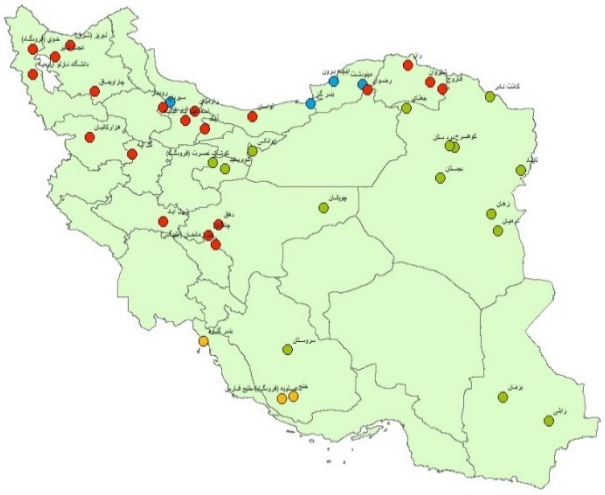
روش/ خوشه**k**	K=۴	K=۵
افرازی میانگین محور (k-means)	۰/۲۴	۰/۲۹
افرازی میانه محور(k-med)	۰/۳۰	۰/۲۷

جدول ۳- نتایج سایر روش‌های خوشه‌بندی برای ایستگاه‌ها با دوره آماری ۵-۱ سال

ردیف	مدل	ضریب کوفنتیک	متوسط عرض سیلوئت
۱	**افرازی میانگین محور (k-means)**	--	۰/۲٤
۲	**افرازی میانه محور (k-med)**	--	۰/۳۰
۳	**سلسله مراتبی میانگین محور(average)**	**0/80**	**0/30**
٤	**سلسله مراتبی (single)**	**0/749**	**0/30**
۵	**سلسله مراتبی کامل (complete)**	**0/70**	**0/30**
٦	**Ward.D**	۰/۵۵۰	۰/۳۰
۷	**Ward.D2**	۰/۵۹۹	۰/۳۰

جدول ٤- خوشه بندی داده های ۵-۱ سال با روش میانه محور

شماره خوشه	تعداد خوشه	نام خوشه ها
۱	۲۰	گل تپه، تبریز (شرق)، چاراویماق، عجب شیر، هزارکانیان، دانشگاه نازلو ارومیه، اسماعیل آباد اقبالیه، لواسان، خوی(فرودگاه)، سیردان، راز، رضوان، چادگان، فاروج، آبیک، رازمیان، شیروان، پل زمانخان(ایلبیگی)، دهق، شول آباد.
۲	۳	بندرگناوه، خنج، فرودگاه عسلویه
۳	۱۵	سروستان، بزمان، زابلی،کلات نادر، بجستان، بردسکن، درمیان، چوپانان، ایوانکی، چغتای، کوه سفید، کوشک نصرت(فرودگاه)، زهان، کوهسرخ، تایباد.
٤	٤	اینچه برون، رودبار، مینودشت، بندرگز.



شکل ۱- نمایش موقعیت جغرافیایی و خوشه‌بندی ایستگاه‌ها با دوره آماری کمتر از ۵ سال

جدول ۵- آمارتوصیفی ٤۲ ایستگاه با دوره آماری ۵-۱ سال در طول دوره آماری

آماره/ متغیر	دما	بارش	رطوبت نسبی	ارتفاع
Mean	۱۷/٤۹	۲۵۵/۵۱	٤۵/۸۲	۱۱۵٦/٦۷
Std	۵/۲۸	۱٦۹/۱۸۸	۱۵/۰٦	۵۹۳/۳٤
Min	۱۰/٦ (چاراویماق)	۱۵/۲٤(زابلی)	۱٤/۱۳(بزمان)	۱٤- (بندرگز)
Max	۲۷/٤۹ (بندرگناوه)	۹۱۲/۸٤ (شول اباد)	۷٤/۲٤ (بندرگز)	۲۱۷۵ (چادگان)

خوشه بندی داده های ۱۰-٦ سال

با توجه به تحلیل‌های صورت گرفته از نمودارها و خروجی‌ها بهترین تعداد خوشه برای داده‌های ۱۰-٦ سال ۵ خوشه می باشد. عرض سیلوئت برای تعداد خوشه‌های کمتر و بیشتر از ۵، کمتر از ۰.۳۸ می‌باشد، لذا تعداد خوشه‌ها برای این ایستگاه‌ها و با دوره آماری مذکور برابر ۵ خوشه می‌باشد. مقادیر عرض سیلوئت با دو روش میانگین و میانه محور و تعداد خوشه ٤ و ۵ در جدول (٦) مشخص شده است.

با توجه به انتخاب ۵ خوشه برای ایستگاه‌ها با دوره آماری ٦-۱۰ سال و بررسی روش میانه محور و سلسله مراتبی میانگین محور در این ایستگاه‌ها، روش سلسله مراتبی میانگین محور با ضریب کوفنتیک ۰/۷۵۸ نتایج دقیق‌تری را ارائه داد.

جدول٦– **خوشه‌بندی ایستگاه‌های ۱۰- ٦ سال با شاخص عرض** سیلوئت

روش/ خوشهk	K=٤	K=٥
میانگین محور (k-means)	۰/۳۲	۰/۳۷
میانه محور (k-med)	۰/۳۵	۰/۳۸

جدول۷– **خوشه‌بندی ایستگاه‌های ۱۰- ٦ سال با شاخص عرض** سیلوئت **و ضریب کوفنتیک**

ردیف	مدل	ضریب کوفنتیک	متوسط **عرض** سیلوئت
۱	افرازی میانگین محور	--	۰/۳۵
۲	افرازی میانه محور	--	۰/۳۸
۳	سلسله مراتبی میانگین محورaverage	۰/۷۵۸	۰/۳۸
٤	سلسله مراتبی تکی single	۰/۷۰۸	۰/۳۸
٥	سلسله مراتبی کاملcomplete	۰/٦٦٦	۰/۳۸
٦	سلسله مراتبی Ward.D	۰/٦۳۸	۰/۳۸
۷	سلسله مراتبی Ward.D2	۰/٦۸۸	۰/۳۸

خوشه‌بندی داده‌های ۱۰-٦ سال با روش سلسله مراتبی میانگین محور

شادگان در استان خوزستان با ارتفاع ۳ متر و میانگین بارش سالانه ۸۵۲ میلیمتر و آبدانان در استان ایلام با ارتفاع ۹۲۰ متر و میانگین بارش سالانه ٦۱۱ میلیمتر از سایر ایستگاه‌ها مجزا شده اند (در شکل ۳ با رنگ های آبی و قرمز مشخص شده‌اند).

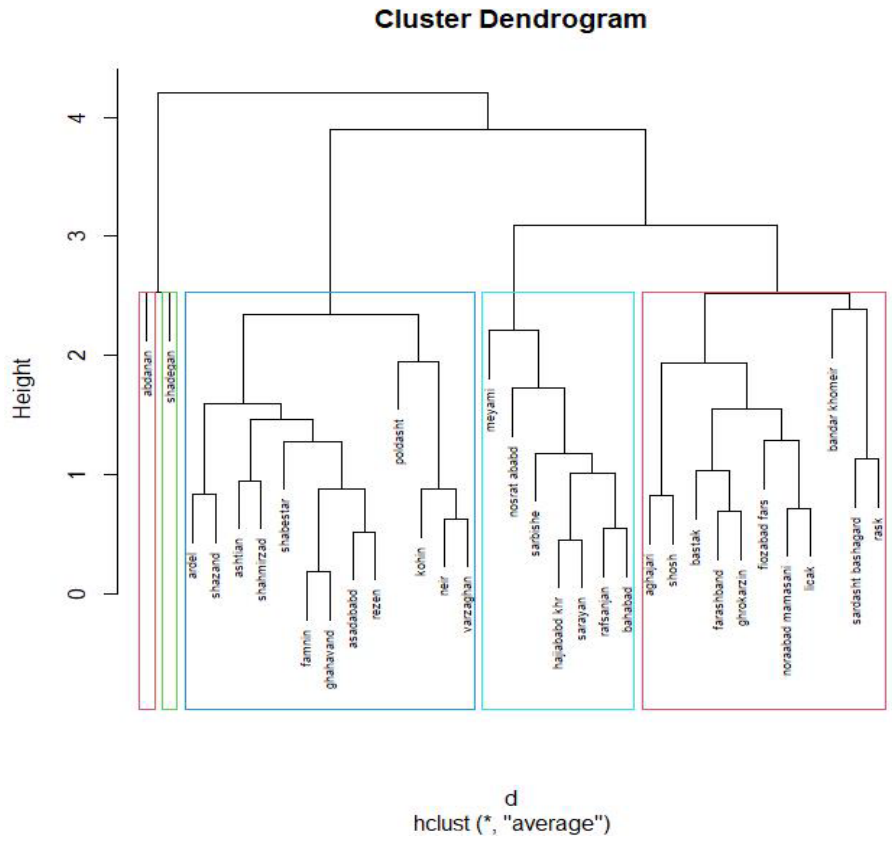
خوشه بندی داده ها با دوره آماری بیشتر از ۱۰ سال

همانطور که گفته شد تعداد ۳۰۰ ایستگاه در دوره آماری بیشتر از ۱۰ سال قرار گرفته اند. همانند دو دسته بندی قبلی پراکندگی ایستگاه‌ها در تمامی استان‌ها می‌باشد. خروجی نرم افزار بهترین عدد برای خوشه بندی را ٤ و ۱۱ مشخص می‌کند، لذا خوشه بندی با تعداد ۳، ٤ ،٥ ، ۱۰، ۱۱ و ۱۲ خوشه مورد آزمایش قرار گرفتند. هرگاه تعداد خوشه ها افزایش پیدا می‌کند عرض سیلوئت کاهش می‌یابد. همچنین با افزایش تعداد خوشه‌ها تعداد زیادی از ایستگاه ها در یک خوشه و تعداد کمی در چند خوشه دیگر قرار می‌گیرد. برای بررسی بهتر ضرایب تصمیم گیرنده تعداد خوشه‌ها در جدول ذیل با یکدیگر مقایسه شده اند.

همانطور که از جدول (۱۰) مشخص است، عرض سیلوئت در خوشه‌بندی با تعداد ۳ خوشه با عرض سیلوئت با تعداد ٤ خوشه یکسان می‌باشد. با پذیرش ۳ خوشه ۱٦۸ ایستگاه در یک خوشه و ۱۰۳ ایستگاه در یک خوشه دیگر قرار می‌گیرند. به عبارتی ویژگی‌های زیادی از ایستگاه‌ها smooth شده‌اند و دقت خوشه‌بندی کم شده است. در خوشه‌بندی با ٤ خوشه، ویژگی‌های خوشه‌ها نقش متمایزکننده‌تری دارند. همچنین با توجه به بررسی روش افرازی میانگین محور و سلسله مراتبی میانگین محور در این ایستگاه‌ها، روش افرازی میانگین محور نتایج دقیق‌تری را ارائه داد.

خوشه بندی داده ها با دوره آماری بیشتر از ۱۰ سال

با توجه به انتخاب مدل خوشه‌بندی و تعداد خوشه‌های متناسب برای داده‌ها با طول دوره آماری بیش از ۱۰ سال و اعمال این روش خوشه‌بندی بر داده‌ها، در جدول ۱۲ تعداد ایستگاه‌ها و عرض سیلوئت حاصل برای هر دسته از خوشه‌ها مشخص شده است.



شکل ۲- نمودار خوشه‌بندی سلسله مراتبی میانگین محور برای داده‌های ۱۰-٦ سال

جدول ۸- **خوشه‌بندی ایستگاه‌های ۱۰- ٦ سال با شاخص عرض** سیلوئت **و ضریب کوفنتیک**

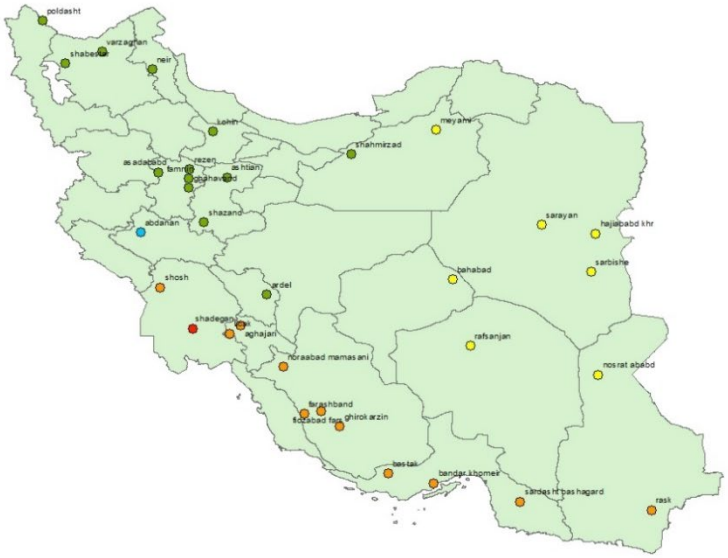
شماره خوشه	تعداد خوشه	نام خوشه ها
۱	۱	شادگان
۲	۱	آبدانان
۳	۱۳	شبستر، اردل، اسدآباد، رزن، فامنین، قهاوند، آشتیان، شهمیرزاد، شازند، نیر، پلدشت، ورزقان، کوهین
۴	۱۱	فراشبند، فیروزآباد فارس، قیروکارزین، نورآباد ممسنی، بستک، بندرخمیر، سردشت بشاگرد، آغاجاری، شوش، راسک، لیکک.
۵	۷	میامی، سرایان، سربیشه، رفسنجان، نصرت آباد، بهاباد، حاجی آباد(خراسان جنوبی).

در این جدول ایستگاه‌ها با دوره آماری ۱۰-٦ سال در ۵ دسته نهایی خوشه‌بندی شده‌اند.

جدول ۹ - آمار توصیفی ۳۳ ایستگاه با دوره آماری ۱۰-٦ سال در طول دوره آماری

آماره/ متغیر	دما	بارش	رطوبت نسبی	ارتفاع
Mean	۲۰/۲۲	۲۳/۸۹	۳۹/۹۵	۱۱٦۹/۵۸
Std	۷/۱۰	۱۳/۸۷	۱۲/۸۸	٦۳٦/۳۳
Min	۹/٦٤ (ورزقان)	۵/۷٦ (نصرت آباد)	۱۹/۹۸ (نصرت آباد)	۳ (شادگان)
Max	۳۰/٦۰ (راسک)	۷۱ (شادگان)	٦٦/۰۷ (پلدشت)	۲۰۹۷ (آشتیان)

برای نمایش بهتر و واضح تر موقعیت ایستگاه‌ها نسبت به یکدیگر و خوشه بندی ایستگاه‌ها با دوره آماری ۱۰-٦ سال بر پهنه جغرافیایی ایران از نرم افزار Arcmap استفاده شده است.



شکل ۳ - نمایش موقعیت جغرافیایی و خوشه‌بندی ایستگاه‌ها با دوره آماری ۱۰-٦ سال

جدول ۱۰- مقایسه ضرایب عرض سیلوئت **وکوفنتیک با تعداد و مدلهای خوشه‌بندی مختلف برای داده‌ها با دوره آماری بیش از ۱۰ سال**

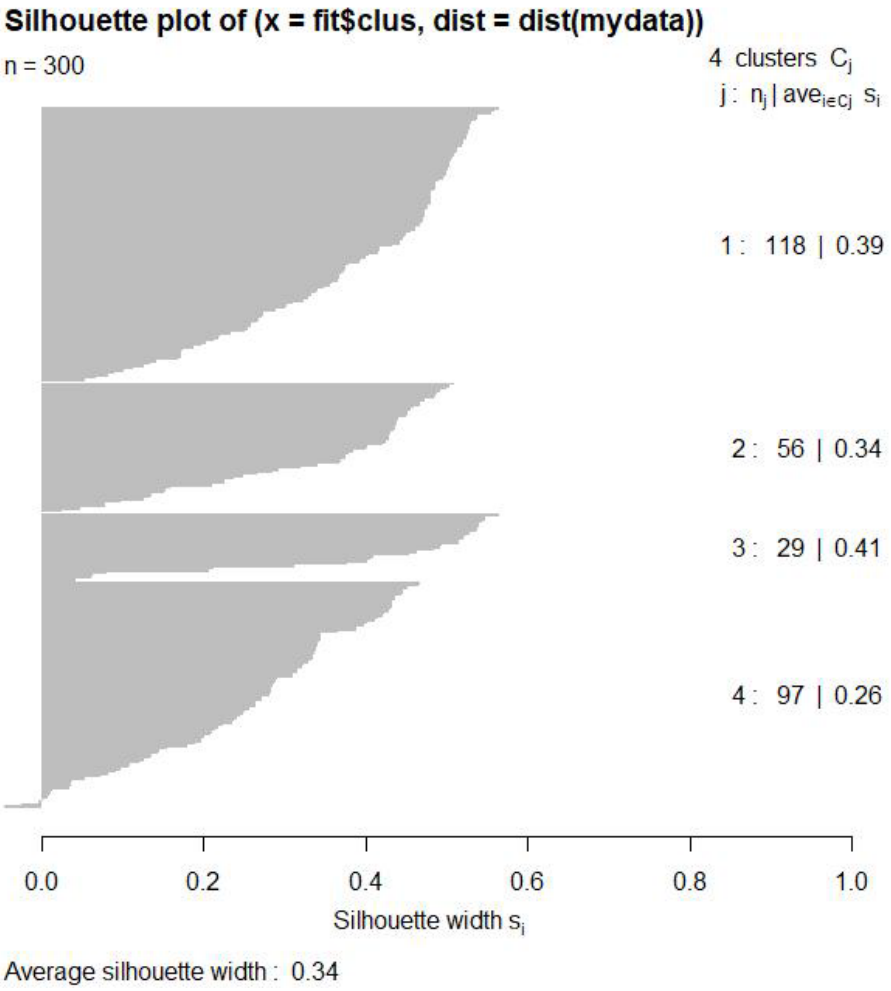
تعداد خوشه	۱۲	۱۱	۱۰	۵	٤	۳
عرض سیلوئت	۰/۲۸	۰/۳۰	۰/۳۱	۰/۲۷	۰/۳٤	۰/۳٤
ضریب کوفنتیک در روش سلسله مراتبی میانگین محور average	۰/۷٤۳	۰/۷٤۳	۰/۷٤۳	۰/۷٤۳	۰/۷٤۳	۰/۷٤۳
ضریب کوفنتیک در روش سلسله مراتبی تکی single	۰/٥۳٥	۰/٥۳٥	۰/٥۳٥	۰/٥۳٥	۰/٥۳٥	۰/٥۳٥
ضریب کوفنتیک در روش سلسله مراتبی تراکمی (کامل) complete	۰/٦۰۸	۰/٦۰۸	۰/٦۰۸	۰/٦۰۸	۰/٦۰۸	۰/٦۰۸
ضریب کوفنتیک در روش Ward.D	۰/٥۲٤	۰/٥۲٤	۰/٥۲٤	۰/٥۲٤	۰/٥۲٤	۰/٥۲٤
ضریب کوفنتیک در روش Ward.D2	۰/٦٦٤	۰/٦٦٤	۰/٦٦٤	۰/٦٦٤	۰/٦٦٤	۰/٦٦٤

جدول ۱۱- مقایسه ضرایب عرض سیلوئت وکوفنتیک با ٤ خوشه در روش‌های مختلف خوشه‌بندی برای داده‌ها با دوره آماری بیش از ۱۰ سال

ردیف	مدل	ضریب کوفنتیک	متوسط عرض سیلوئت
۱	افرازی میانگین محور	--	۰/۳٤
۲	افرازی میانه محور	--	۰/۳۱
۳	سلسله مراتبی میانگین محور average	۰/۷٤۳	۰/۳۱
٤	سلسله مراتبی تکی single	۰/٥۳٥	۰/۳٤
٥	سلسله مراتبی کامل complete	۰/٦۰۸	۰/۳٤
٦	سلسله مراتبی Ward.D	۰/٥۲٤	۰/۳٤
۷	سلسله مراتبی Ward.D2	۰/٦٦٤	۰/۳٤

جدول ۱۲– تعداد ایستگاه و میانگین عرض سیلوئت در هر خوشه برای داده‌ها با دوره آماری بیش از ۱۰ سال

شماره خوشه	تعداد ایستگاه	میانگین عرض سیلوئت
۱	۱۱۸	۰/۳۹
۲	٥٦	۰/۳٤
۳	۲۹	۰/٤۱
٤	۹۷	۰/۲٦



شکل ٤ – نمودار خوشه‌بندی افرازی بر اساس عرض سیلوئت **با ٤ خوشه، به ترتیب با تعداد ۱۱۸، ٥٦، ۲۹ و ۹۷**

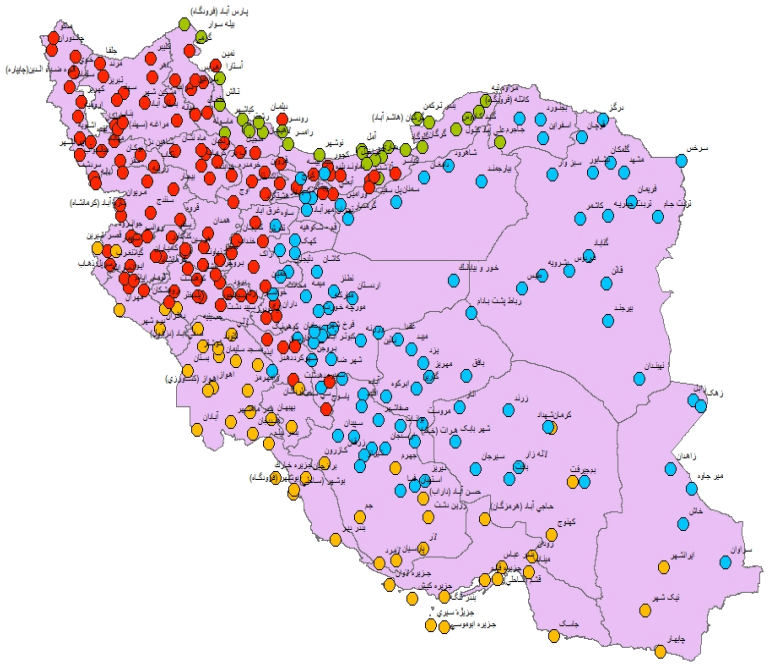
آمار توصیفی ایستگاه‌ها در هر خوشه‌بندی در جدول (۱۳) نمایش داده شده است. همانطور که مشخص است بیشینه بارش در این خوشه مربوط به ایستگاه بندرانزلی در استان گیلان با ارتفاع ۲۳/٦– متر و کمینه بارش مربوط به ایستگاه شهداد در استان کرمان با میانگین بارش سالانه ۳۲/۱٦ میلیمتر و ارتفاع ٤۸۲ متر می‌باشد. همچنین کمینه و بیشینه میانگین دمای ماهانه مربوط به ایستگاه‌های فیروزکوه در استان تهران (۵/۵۸) و رودان در استان هرمزگان(۳۰/۰۳) درجه سانتی گراد میباشد. در این خوشه بندی کمترین و بیشترین ارتفاع به ترتیب مربوط به ایستگاه بندرانزلی در استان گیلان با ۲۳/٦– و فیروزکوه در استان تهران با ۲۹۸۵/۷ متر می‌باشد.

جدول ۱۳– آمار توصیفی داده‌ها با دوره آماری بیش از ۱۰ سال در طول دوره آماری

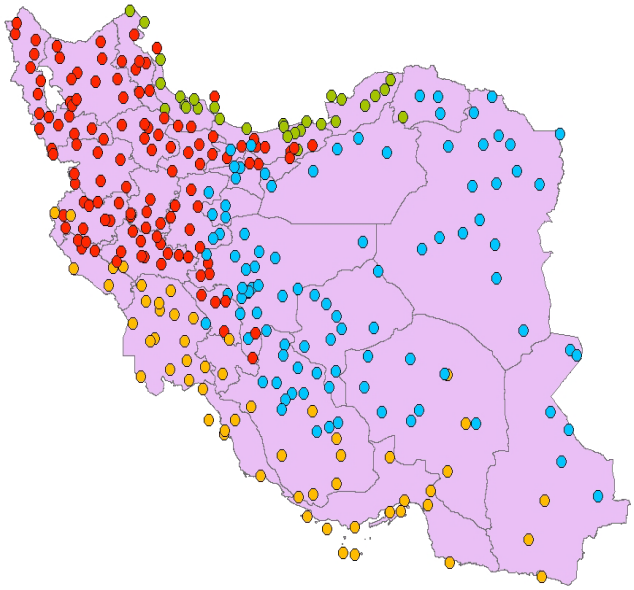
آماره/ متغیر	دما	بارش	رطوبت نسبی	ارتفاع
Mean	۱۷/۸۷	۳٤٤/۰۸	٤٦/۸٥	۱۱۳۷/۳۷
Std	٥/۲۳	۲٦۲/۱٤	۱٤/۱۳	٦۹٥/۹٦
Min	٥/٥۸ (فیروزکوه)	۳۲/۱٦ (شهداد)	۱۷/۹٦(میرجاوه)	۲۳/٦– (بندرانزلی)
Max	۳۰/۰۳ (رودان)	۱۸۳۰/٤۸ (بندرانزلی)	۸۳/۳۹ (بندرانزلی)	۲۹۸٥/۷ (فیروزکوه)

پراکندگی جغرافیایی ایستگاه در هرخوشه بر پهنه جغرافیایی کشور ایران در دو تصویر زیر با نام و بدون نام ایستگاه (برای وضوح بیشتر) نمایش داده شده است. همانطور که مشخص است پهنه شرقی و مرکزی در یک خوشه با تعداد ۹۷ ایستگاه، پهنه شمالی با تعداد ۲۹ ایستگاه با رنگ سبز، پهنه شمال غربی و تا حدودی غرب با رنگ قرمز و ۱۱۸ ایستگاه و پهنه جنوبی با رنگ زرد و ٥٦ ایستگاه مشخص شده‌اند. در این خوشه‌بندی ها ایستگاه دیلمان (نزدیک لاهیجان) در استان گیلان برخلاف سایر ایستگاه‌های این استان دارای رنگ قرمز می‌باشد. بررسی ها نشان داد که این ایستگاه دارای ارتفاع ۱٤٤۷/٦ متر و میانگین بارش ۳۱/۲٤ میلیمتر و متفاوت با ارتفاع سایر ایستگاه‌های این استان که پربارش و کم ارتفاع هستند می‌باشد، لذا در خوشه‌بندی رنگ قرمز قرار گرفته است. این شرایط برای ایستگاه نمین (نزدیک آستارا) با ارتفاع متر۱٤۸۰ و میانگین بارش ۲۱/۱٦ میلیمتر در استان گیلان همچنان برقرار است. رطوبت نسبی در این دو ایستگاه بالا و دما پایین می‌باشد. میانگین بارش در این دو ایستگاه بسیار نزدیک به میانیگن بارش دوره آماری در خوشه‌بندی رنگ قرمز(نوار غربی ایران) ۳۳/۷۳ میلیمتر می‌باشد.

در جدول (۱٤) اسامی ایستگاه‌های قرار گرفته به تفکیک هر خوشه در این دوره آماری آورده شده است. در جدول (۱٥) آمار توصیفی متغیرهای مورد بررسی به تفکیک هر خوشه در دوره آماری بالای ۱۰ سال محاسبه شده است.



شکل ٥ – نمایش موقعیت جغرافیایی و خوشه‌بندی ایستگاه‌ها با دوره آماری بیش از ۱۰ سال با نام ایستگاه



شکل ٦ – نمایش موقعیت جغرافیایی و خوشه‌بندی ایستگاه‌ها با دوره آماری بیش از ۱۰ سال بدون نام ایستگاه

جدول ١٤– خوشه‌بندی ایستگاه‌ها با دوره آماری بیش از ۱۰ سال

شماره خوشه	تعداد خوشه	نام خوشه ها
۱	۱۱۸	کرج – هشتگرد – اردبیل (فرودگاه) – خلخال – سرعین – فیروز آباد (اردبیل) – گرمی – مشکین شهر – نمین – ارومیه – اشنویه – بوکان – پیرانشهر – تکاب – چالدوران – خوی – سردشت – سلماس – شاهین دژ – قره ضیاء الدین(چاپیاره) – کهریز – ماکو – مهاباد – میاندوآب – نقده – اهر – بستان آباد – بناب – تبریز – ملکان – جلفا – سراب – سهند – کلیبر – مراغه (سهند) – مرند – میانه – هریس – شهرکرد – فارسان – کوهرنگ – لردگان – خوانسار – داران – فریدونشهر – جیرنده – دیلمان – منجیل – تویسرکان – ملایر – نهاوند – همدان – ایلام – ایوان – سرابله – لومار – اسلام اباد غرب – جوانرود – روانسر – سرآرود (کرمانشاه) – سرپل ذهاب – تازه آباد (کرمانشاه) – سنقر – گیلانغرب – هرسین – کرمانشاه – کنگاور – سی سخت – یاسوج – بانه – بیجار – زرینه – سقز – سنندج – قروه – مریوان – کامیاران – ازنا – الشتر – سیلاخور – الیگودرز – ایمان آباد – بروجرد – خرم اباد – درود – رومشکان – سپید دشت – نورآباد (لرستان) – کوهدشت – اراک – تفرش – خمین – خنداب – غرق آباد – کمیجان – آلاشت – بلده – سیاه بیشه – کجور – کیاسر – آوج – بوئین زهرا – تاکستان – قزوین – معلم کلایه – آبعلی – تهران (شمیران) – دماوند – فیروزکوه – فیروزکوه (آلودگی زمینه جو) – آب بر – خدابنده – خرمدره – خیر آباد – زنجان – زنجان (فرودگاه) – گرماب – ماه نشان
۲	۵۶	برازجان – بندر دیر – بندر دیلم – بوشهر (ساحلی) – بوشهر (فرودگاه) – جزیره خارک – زرین دشت – جهرم – حسن آباد (داراب) – کازرون – لار – لامرد – جزیره قشم – بندر عباس – بندر لنگه – پارسیان – جاسک – جزیره لاوان – جزیره کیش – جزیره ابوموسی – جزیره سیری – حاجی آباد (هرمزگان) – رودان – قشم (ساحلی) – میناب – دره شهر – دهلران – مهران – جیرفت – شهداد – کهنوج – سومار – قصر شیرین – آبادان – امیدیه (آغاجاری) – اهواز – اهواز (کشاورزی) – ایذه – بستان – بندر ماهشهر – بهبهان – حسینیه – رامهرمز – شوشتر – صفی اباد (دزفول) – گتوند – لالی – مسجد سلیمان – هندیجان – دو گنبدان(گچساران) – دهدشت – پلدختر – ایرانشهر – چابهار – نیک شهر – جم

ادامه جدول ١٤– خوشه‌بندی ایستگاه‌ها با دوره آماری بیش از ۱۰ سال

شماره خوشه	تعداد خوشه	نام خوشه ها
۳	۲۹	بیله سوار – پارس آباد (فرودگاه) – آستارا – بندر انزلی – تالش – رشت – رشت (کشاورزی) – رودسر – لاهیجان – ماسوله – کیاشهر – بندر ترکمن – علی آباد کتول – گرگان – گرگان (هاشم آباد) – گنبد کاووس – مراوه تپه – کلاله (فرودگاه) – آمل – بابلسر – بندر امیر آباد – پل سفید – رامسر – ساری (فرودگاه دشت ناز) – ساری – قراخیل – گلوگاه – نوشهر – جاجرم
۴	۹۷	بروجن – سامان – فرخ شهر – اردستان – اصفهان (ازن سنجی) – اصفهان (فرودگاه) – خور و بیابانک – زرین شهر – سمیرم – شهر ضا – کاشان – کبوتر آباد – گلپایگان – مبارکه – مورچه خورت – میمه – نائین – نجف آباد – نطنز – ورزنه – آباده – ارسنجان – استهبان – اقلید – ایزد خواست – بوانات – تخت جمشید – درود زن – زرقان – سپیدان – شیراز – صفاشهر – فسا – نیریز – بافت – بم – زرند – سیرجان – انار – شهر بابک – لاله زار – کرمان – تربت جام – تربت حیدریه – درگز – سبز وار – سرخس – فریمان – قوچان – گلمکان – گناباد – مشهد – نیشابور – کاشمر – دهدز – دلیجان – ساوه – محلات – اسفراین – بجنورد – مانه و سملقان(آشخانه) – سلفچگان – قم – شکوهیه – کهک – بیارجمند – دامغان – سمنان – شاهرود – گرمسار – خاش – زابل – زاهدان – سراوان – زهک – میر جاوه – بشرویه – بیرجند – طبس – فردوس – قائن – نهبندان – تهران (ژئوفیزیک) – تهران مهرآباد – چیتگر – شهریار – فرودگاه امام خمینی – ورامین – ابرکوه – بافق – رباط پشت بادام – عقدا – گاریز – مروست – مهریز – میبد – هرات (خاتم) – یزد

جدول ١٥– آمار توصیفی خوشه‌ها با دوره آماری بیش از ۱۰ سال

Cluster		(نوار غربی)۱	(نوارجنوبی)۲	(نوار شمالی)۳	(نوارشرقی)٤
تعداد ایستگاه		۱۱۸	٥٦	۲۹	۹۷
دمای ماهیانه	mean	۱۳/۸۸	۲٦/٤٤	۱۷/۰۳	۱۸/۰٥
	max	۲۱/۹۰	۳۰/۰۳	۱۹/٥۷	۲٦/٥٦
	min	٥/٥۸	۲۲/۸۰	۱۲/۰٦	۱۲/۱۱
بارش سالانه	mean	۳۳/۷۳	۲۰/۰۳	۷۱/۲٥	۱٤/۷۸
	max	۱۱۱/۱۹	٥۱/۷۳	۱٥۲/٥٤	٥۱/٦٤
	min	۱٦/۸٤	۲/٦۸	۲۳/٥۳	۳/۲۰
رطوبت نسبی ماهانه	mean	٤۹/٥٥	٤٥/۲٦	۷٤/۸۳	۳٦/۱۳
	max	۷۱/۷۷	۷۲/۱٥	۸۳/۳۹	٥۸/۷۹
	min	۳٤/٦۹	۲۰/۰۰	٤٤/٦٦	۱۷/۹٦
ارتفاع	mean	۱٥٤٥/٥۷	۲۹۸/۷۳	۱۲٥/۲٦	۱٤۲۷/٥٦
	max	۲۹۸٥/۷۰	۱۰۹۸/۲۰	۱۰۸۱/۰۰	۲۷۷٥/۰۰
	min	۳۳۸/۳۰	-۱/۲۰	-۲۳/٦۰	۲۷۸/۰۰

نتیجه‌گیری

این پژوهش به خوشه‌بندی تمامی ایستگاه‌های سینوپتیک ایران از منظر هواشناسی و جغرافیایی پرداخته است. در این پژوهش اطلاعات ٤۲۰ ایستگاه سینوپتیک شامل مختصات طول، عرض، ارتفاع جغرافیایی، میانگین دما، رطوبت نسبی و مجموع بارش ماهانه از زمان تاسیس تا سال ۲۰۱۸ مورد بررسی قرار گرفت. پس از بررسی، پاک‌سازی و ترمیم داده‌های٤۲۰ ایستگاه سینوپتیک ایران، ۳۷٥ ایستگاه برای حضور در ادامه پژوهش مورد پذیرش قرار گرفتند. از طرفی با توجه به اهمیت رفتار متغیرهای هواشناسی در طول دوره

آماری و بررسی رفتار ایستگاه‌ها با طول دوره آماری مشابه و کنترل اثر زمان بر خوشه‌بندی، ایستگاه‌ها برحسب دوره آماری به سه دوره کمتر از ۵ سال با ٤۲ ایستگاه، ٦-۱۰ سال با ۳۳ ایستگاه و بیشتر از ۱۰ سال با ۳۰۰ ایستگاه دسته‌بندی شدند. سپس خوشه‌بندی درون هر گروه به صورت مجزا صورت گرفت. دو شاخصه ضریب کوفنتیک و سیلوئت به منظور یافتن بهترین تعداد خوشه و روش خوشه‌بندی برای هر گروه داده استفاده شده است. هفت روش خوشه‌بندی شامل پنج روش سلسله مراتبی و دو روش افرازی در این پژوهش به کار گرفته شده است. بر این اساس بهترین تعداد و روش خوشه برای داده‌های ۵-۱ سال ٤ خوشه با روش افرازی میانه‌محور با عرض سیلوئت ۰/۳۰، برای داده‌های ٦-۱۰ سال ۵ خوشه با روش سلسله مراتبی میانگین محور با عرض سیلوئت ۰/۳۸ و برای ایستگاه‌ها با دوره آماری بیش از ۱۰سال ٤خوشه و روش افرازی میانگین محور با عرض سیلوئت ۰/۳٤ می‌باشد. روش‌های خوشه‌بندی پذیرفته شده در این پژوهش به نتایج فرزندی و همکاران و رورده نزدیک می‌باشد. لازم به ذکر است در این پژوهش تمامی ایستگاه‌های سینوپتیک کشور مورد بررسی قرار گرفتند و با توجه به متفاوت بودن سال تاسیس ایستگاه، در ابتدا از نظر زمانی گروه بندی شدند. به‌عبارت دیگر اثر زمان در خوشه‌بندی کنترل شد. همچنین خوشه‌بندی همزمان با در نظر گرفتن پارامترهای جغرافیایی و هواشناسی انجام شده است. پهنه‌بندی خوشه‌ها بر نقشه جغرافیایی ایران با استفاده از نرم‌افزار ARCGIS برای هر سه دسته رسم شد که در هر گروه داده نشان از خوشه‌بندی دقیق در جغرافیای ایران دارد. پیشنهاد میشود خوشه‌بندی این مجموعه ایستگاه‌ها با درنظر گرفتن عوامل جغرافیایی و سایر عوامل هواشناسی و همچنین دیگر روش‌های خوشه‌بندی در تحقیقات بعدی انجام شود.

منابع

1. Arghami.N.R.Sanjari.D.Bozorgnia.A.2001. Elementary survey sampling. Ferdowsi university of Mashhad. Mashhad.
2. Alizadeh.A.2011.Climate and Agriculture Meteorology. Imam Reza University of Mashhad. Mashhad.
3. Everitt B.S. and Hothorn, T.. A Handbook of Statistical Analyses Using R. Chapman & Hall/CRC, Taylor & Francis Group, pp269.
4. Farzandi et al. 2017. Approach, recognition and development of Iran's clustering with main meteorological parameters. National conference of new knowledge and technology in engineering sciences in the age of technology; November 17, 2016-Tehran. Tehran.
5. Fahmi, Fidan M. et al, 2011. Determining the Climate Zones of Turkey by Center-Based Clustering Methods, Middle East Technical University, Ankara, Turkey.
6. Gareth J. Daniela W. Trevor H. Robert T. 2013. An Introduction to Statistical Learning with Applications in R. Springer.
7. Khatibi Rasool, Soltani Saeed, Khodagholi Morteza. Bioclimatic Classification of South East of Iran Using Multivariate Statistical Methods. Applied Ecology and Environmental Sciences. 2019; 7(5):190-205.
8. Larose, D. T. (2005). Discovering Knowledge in Data An Introduction To Data Mining, John Wiley And Sons.
9. Pourbabak et al.2015. Classification of temperature and annual precipitation of meteorological stations of Iran using fuzzy clustering. Journal of Geography and Planning.NO 55.
10. Rezaee Pazhand.H.2001. Application of statistics and probability in water resources. The first edition of Sokhongostar Publications of Islamic Azad University of Mashhad. Mashhad.
11. Rourdeh et al.2018. Rainfall clustering in Iran using a new method based on the application of SDV mapping and FCM fuzzy clustering. Scientific Research Quarterly of Golestan University, No. 31.
12. Rao, A.R. and Srinivas V.V. 2006. Regionalization of watersheds by hybrid cluster analysis", Journal of Hydrology, 318: 37-56.
13. Romesburg H. C., 2004. Cluster Analysis for Researchers. Lifetime Learning Publications, Belmont, C.A, pp169.
14. Tabatabaei.M.M. 1997.Multivariate Analysis. University Publishing Center of Tehran. Tehran.
15. The website of the National Meteorological Organization (www.irimo.ir).
16. Yurdanur U. Tayfun K. and Mehmet K. 2003. Redefining the climate zones of turkey using cluster analysis. International Journal of Climatology. 23: 1045–1055.