Archive of SID

نشریه دانش آب و خاک / جلد ۲۸ شماره ۳ صفحه‌های ۱۶۹ تا ۱۸۱ / سال ۱۳۹۷

پهنه‌بندی زمانی – مکانی اقلیم بارش ایران و انتخاب ایستگاه‌های شاخص با روش‌های آماری چند متغیره

محمد عیسی‌زاده*[۱]، یعقوب دین‌پژوه[۲]

۱-دانشجوی دکتری، گروه مهندسی آب دانشکده کشاورزی، دانشگاه تبریز

۲-دانشیار، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

* مسئول مکاتبات، پست الکترونیکی: Mohammadisazade@gmail.com

چکیده

در این مطالعه، جهت پهنه‌بندی نواحی بارش کشور، از اطلاعات ۳۱ ایستگاه هواشناسی همدید در دوره آماری ۲۰۱۰-۱۹۸۷ استفاده شد. برای این منظور، داده‌های هر ماه استاندارد شده و در ماتریس با ابعاد (n*m) که در آن n تعداد ایستگاه‌ها (۳۱) و m تعداد ماه‌ها (۱۲) است، نوشته شدند. تجزیه به مولفه‌های اصلی (PCA) روی ماتریس مذکور انجام شد. با توجه به معیار دارا بودن مقدار ویژه بالای یک، مولفه‌های اصلی انتخاب شدند. آنگاه مقادیر امتیازات مولفه‌های اصلی (PCS) برای مولفه‌های منتخب محاسبه شد. این مقادیر، به‌عنوان ورودی روش تجزیه خوشه‌ای با روش وارد استفاده شد. آنگاه با توجه به نمودار خوشه‌ای، اقلیم‌های بارش کشور تشخیص داده شدند. برای پاسخ به این سوال که کدام ایستگاه‌ها می‌توانند به‌عنوان نماینده اقلیم بارش هر منطقه باشند، از روش پروکراستس (PA) استفاده شد. افزون بر این، از همین روش برای انتخاب ماه‌هایی که بارش آن‌ها معرف بارندگی در کل سال است، استفاده شد. نتایج نشان داد که سه مولفه اصلی اول، ۹۷ درصد واریانس کل داده‌ها را توجیه می‌کند. براساس مولفه‌های منتخب، در کل کشور، شش ناحیه بارشی متمایز تشخیص داده شد. همچنین روش PA نشان داد که بارش ماه‌های مه، اوت و دسامبر، تقریبا کل اطلاعات بارش سالانه را دربر دارند. همچنین معلوم شد که هفت ایستگاه واقع در نقاط مختلف کشور شامل زاهدان، تهران، ارومیه، ایلام، یاسوج، گرگان و رشت می‌توانند به‌عنوان ایستگاه‌های شاخص در نظر گرفته شوند. این ایستگاه‌ها ۸۷٪ واریانس کل داده‌های ایستگاه‌های کشور را دربر داشتند.

واژه‌های کلیدی: اقلیم، ایران، بارش، تجزیه به مولفه‌های اصلی، تجزیه خوشه‌ای، روش‌های چند متغیره

www.SID.ir

Archive of SID

Spatio- Temporal delineation of Iran's Precipitation Climate and Selection of Indicator Stations Using the Multivariate Statistical Methods

M Isazadeh[*1], Y Dinpashoh[2]

[1-] Ph.D. Student, Dept. of Water Engineering, Faculty of Agric., Univ. of Tabriz, Iran
[2-] Assoc. Prof., Dept. of Water Engineering, Faculty of Agric., Univ. of Tabriz, Iran
[*] Corresponding Author, Email: mohammadisazade@gmail.com

Abstract

In the present study, the information of 31 synoptic weather stations in the period of 1987-2010 were used to delineate Iran's precipitation regions. For this purpose, each month's data were standardized and written in the (n*m) matrix, where n is the number of stations (31) and m is the number of months (12). Principal component analysis was conducted on the mentioned matrix. The main components were determined according to the criterion of having an eigen value greater than one. Then principal components scores were calculated for the selected components. These valeus were used as an input for Ward method of cluster analysis. Then according to the cluster diagram, precipitation climates of the country were identified. The Procrustes analysis (PA) was used to answer the question of which stations could be considered as the indicator of rainfall climates? Furthermore, from this method those months, which their precipitation can be recognized as an indicator of annual rainfall, were selected. Results showed that the first three components incorporated more than 97% of total variance. Based on the selected components the six distinct precipitation regions were distinguished across the country. Furthermore, PA results indicated that the precipitation of May, August and December approximately had whole of the annual precipitation information. Moreover, it was found that seven stations located in different points of the country namely Zahedan, Tehran, Urmia, Ilam, Yasooj, Gorgan and Rasht could be considered as the indicator stations. These stations incorporated over 87 percent of total variance of data of all selected stations in the country.

Keywords: Climate, Cluster analysis, Iran, Multivariate methods, Precipitation, Principal component analysis

مقدمه

ایران سرزمینی خشک با نزولات جوی بسیار نامنظم است، بطوریکه میانگین بارندگی سالانه ایران حدود یک سوم مقدار نظیر در سطح خشکی‌های کره زمین است. علاوه بر این، زمان ریزش نزولات جوی و محل ریزش آن‌ها نیز با نیازهای بخش کشاورزی، که مصرف کننده اصلی آب در کشور می‌باشد، مطابقت ندارد. بنابراین، باید پذیرفت که خشکی در ایران یک واقعیت اقلیمی است و این ما هستیم که باید خودمان را با آن سازگار کنیم (علیزاده ۱۳۸۹). دو مسئله مهم و اساسی در بررسی بارش و همچنین مدیریت اقتصادی و فنی ایستگاه‌های همدید کشور مطرح می‌باشد. مسئله اول تعیین مناطقی است که از لحاظ بارش همگن هستند که این امر می‌تواند در تصمیم‌گیری‌ها و سیاست گذاری‌های کلان کشور نقش مهمی ایفا کند. مسئله دوم که یکی از مسائل مهم و ضروری در مدیریت و برنامه‌ریزی شبکه ایستگاه‌های باران‌سنجی است، تعیین ایستگاه‌های شاخص می‌باشد. این ایستگاه‌ها دربرگیرنده بیشتر اطلاعات سایر ایستگاه-های منطقه مورد بررسی می‌باشند. ایستگاه‌های شاخص می‌توانند محل استقرار تجهیزات دیده‌بانی پیشرفته‌ای باشند که به‌دلیل بالا بودن هزینه آن‌ها، تهیه و نصب این تجهیزات در همه ایستگاه‌های کشور مقدور نمی‌باشد. در صورتی که موقعیت جغرافیایی ایستگاه‌های شاخص بدرستی انتخاب نگردد، مواردی مانند وجود اتلاف وقت و انرژی، از دست رفتن اطلاعات مورد نیاز، هدر رفت سرمایه گذاری‌های کلان مورد انتظار خواهد بود. مطالعات مختلفی در این زمینه‌ها صورت گرفته که در ادامه به بخشی از مهم‌ترین آن‌ها اشاره می‌گردد.

پانسرا و همکاران (۲۰۱۳) با استفاده از روش‌های سلسله مراتبی و ترکیب آن‌ها با روش K-means، خوشه‌بندی ۲۲۷ ایستگاه باران‌سنجی واقع در ایالت پارانا (کشور برزیل) را انجام داده‌اند. نتایج نشان داد که در بین روش‌های

www.SID.ir

Archive of SID

سلسله مراتبی، روش وارد[1]بهترین خوشه‌بندی را انجام داده است. همچنین در ترکیب روش‌های سلسله مراتبی با روش K-means نیز ترکیب روش‌های وارد و K-means نسبت به سایر ترکیب‌ها عملکرد بهتری داشته است. در نهایت، نیز با توجه به آزمون‌های همگنی، ترکیب روش وارد و K-means به‌عنوان کارآمدترین روش خوشه‌بندی ایستگاه‌های هواشناسی معرفی گردید. گوجیک و تراجکویک (۲۰۱۴) الگوهای مکانی- زمانی بارش صربستان را با استفاده از داده‌های بارش ماهانه ۲۹ ایستگاه همدید مورد ارزیابی قرار دادند. در این تحقیق، تجزیه مولفه‌های اصلی (PCA) با چرخش واریماکس[2]بر روی هفت متغیر بارش انجام شد. سپس براساس روش خوشه‌بندی وارد و به‌ازای مولفه‌های اصلی اول و دوم چرخش یافته، خوشه‌بندی صورت گرفت. در نهایت، صربستان به سه ناحیه همگن بارشی تفکیک گردید. نام و همکاران (۲۰۱۵) آنالیز فراوانی منطقه‌ای حداکثر بارش سالانه را با استفاده از داده‌های ۶۷ ایستگاه هواشناسی در کشور کره‌جنوبی انجام دادند. ایشان با استفاده از روش پروکراستس[3](PA) از کل ۴۲ متغیر در نظر گرفته شده، تعداد ۳۳ متغیر به‌عنوان متغیرهای مهم انتخاب کردند. سپس از این متغیرهای منتخب و روش‌های خوشه-بندی وارد و فازی- C-means جهت تقسیم ایستگاه‌ها استفاده گردید که این روش‌ها پنج ناحیه همگن را برای بارش کره‌جنوبی مشخص نمود. عثمان و همکاران (۲۰۱۵) جهت شناسایی الگوی بارش روزانه، از داده‌های بارش ۸۹ ایستگاه در کشور مالزی و روش تجزیه مولفه‌های اصلی (PCA) استفاده کردند. شش مولفه اول با ۵۳/۴۳ درصد واریانس کل، توسط روش تجزیه مولفه‌های اصلی انتخاب گردیدند. نتایج نشان داد که مولفه‌های اول و دوم، مناطق شمال شرقی و جنوب غربی را که تحت تاثیر بادهای موسمی هستند را شامل می‌گردند. کاروالهو و همکاران (۲۰۱۶) با استفاده از اختلاف داده‌های حداقل و حداکثر دما و بارش روزانه، قاره اروپا را به کلاس‌های مختلف تقسیم کردند. در این تحقیق از روش K-means با تعداد خوشه‌های متفاوت استفاده شد که در نهایت تعداد شش خوشه به‌عنوان بهترین تعداد خوشه تعیین گردید. نتایج نشان داد هر یک از خوشه‌ها در یکی از سه پارامتر مورد استفاده دارای اختلاف معنی‌داری هستند.

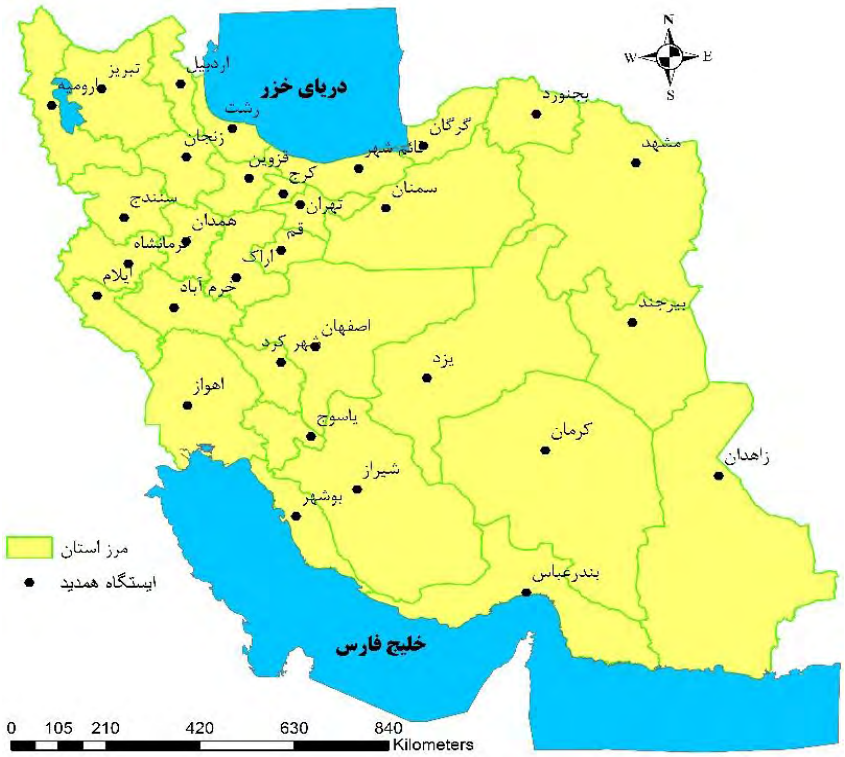
ناظم‌السادات و همکاران (۱۳۸۲) بارندگی زمستانه استان-های بوشهر، فارس، کهگیلویه و بویراحمد را با استفاده از روش PCA پهنه‌بندی کرده‌اند. با توجه به دو مولفه اصلی اول منطقه مورد مطالعه به دو قسمت تقسیم گردید که علت این تقسیم‌بندی را نیز تفاوت در منشا تولید بارش عنوان کرده‌اند. دین‌پژوه و همکاران (۲۰۰۴) جهت پهنه‌بندی اقلیم بارش کشور ایران از ۵۷ متغیر مرتبط با بارش بدست آمده از ۷۷ ایستگاه هواشناسی استفاده کرده‌اند. سپس با استفاده از روش PA تعداد ۱۲ متغیر مهم از بین همه متغیرهای کاندیدا انتخاب کردند. آنگاه با اعمال روش‌های PCA و تجزیه خوشه‌ای وارد، هفت گروه مشابه اقلیم بارش را در ایران تشخیص دادند. نتایج آزمون همگنی نشان داد که شش ناحیه کاملا همگن و یک ناحیه غیر همگن می‌باشد. فلاحی و همکاران (۱۳۹۱) پهنه‌بندی بارش‌های روزانه شمال غرب ایران را با استفاده از روش‌های K-means، وارد و PCA انجام داده‌اند. روش‌های ذکر شده منطقه مورد مطالعه را به‌ترتیب به دو، دو و نه منطقه همگن تقسیم کرده‌اند. نتایج نشان داد که خوشه‌های حاصل از روش PCA، واقعی‌تر از خوشه‌های حاصل از سایر روش‌ها می‌باشد. شیروانی و ناظم‌السادات (۱۳۹۱) برای گروه‌بندی ایستگاه‌ها از نظر بارش ماهانه ایران به نواحی همگن از روش‌های PCA و خوشه‌ای استفاده کرده‌اند. پس از انجام PCA، تحلیل خوشه‌ای وارد استفاده شد که در آن کشور ایران به شش ناحیه مختلف پهنه‌بندی شد. نادی و خلیلی (۱۳۹۲) با استفاده از داده‌های دراز مدت ۸۱ ایستگاه همدید، ۹۸ متغیر ترکیبی محاسبه کرده-اند. سپس با روش حذف پسرو ۳۹ متغیر جهت تحلیل انتخاب گردید که این مجموعه نیز با روش تحلیل عاملی به شش متغیر کاهش یافت. در نهایت، گروه‌بندی امتیازات عاملی با الگوریتم خوشه‌بندی سلسله مراتبی وارد انجام شد و شش ناحیه اقلیم بارش در گستره کشور ایران نتیجه گردید. خسروی و همکاران (۱۳۹۳) نیز براساس ۱۸۰ ایستگاه همدیدی، بارش و دمای ایران را خوشه‌بندی کرده‌اند. تجزیه خوشه‌ای وارد انجام شده بروی داده‌های دما و بارش، پنج پهنه اقلیمی برای ایران مشخص نمود. تحقیقات دیگری نیز در بعضی استان‌های ایران با روش خوشه‌بندی وارد صورت گرفته است که از آن میان می‌توان به مطالعات مسعودیان و همکاران (۱۳۹۰)، خورشیددوست و شیرزاد (۱۳۹۳)، آروین و همکاران

[1]- Ward method
[2]- Varimax
[3]- Procrustes

www.SID.ir

Archive of SID

(۱۳۹۱) و جهان‌بخش‌اصل و همکاران (۱۳۹۴) اشاره نمود. ناظری تهرودی و همکاران (۱۳۹۵) تحلیل روند تغییرات ایستگاهی و منطقه‌ای بارش نیم قرن اخیر کشور ایران را انجام دادند. جهت بررسی روند در مقیاس منطقه‌ای، از روش خوشه‌بندی فازی استفاده گردیده که طی آن ایستگاه‌های کشور به پنج خوشه اصلی تقسیم شده است. آماره کندال منطقه‌ای، بیشترین روند کاهشی بارش را در مناطق شمال غرب کشور تشخیص داده است.



شکل ۱- موقعیت منطقه مورد مطالعه و ایستگاه‌های همدید مورد استفاده.

تحقیق حاضر، با استفاده از بارش‌های ماهانه ایستگاه‌های همدید و به‌ازای ترکیب روش‌های تجزیه به مولفه‌های اصلی (PCA) و تجزیه پروکراستس (PA) با روش خوشه-بندی سلسله مراتبی وارد، چهار هدف اصلی زیر را دنبال می‌کند. هدف اول، تعیین مولفه‌های اصلی میانگین بارش ماهانه و سپس خوشه‌بندی ایستگاه‌ها با روش وارد می-باشد. خروجی این مرحله، ایستگاه‌های با اقلیم بارش مشابه در ایران را مشخص کرده و در خوشه‌های متمایز قرار می‌دهد. هدف دوم، تعیین ماه (ماه‌هایی) است که میانگین بارش‌های ماهانه آن‌ها، گویای اقلیم بارش خوشه است. این کار (کاهش تعداد ورودی) با استفاده از روش PA انجام شده و در ادامه جهت خوشه‌بندی ایستگاه‌ها به روش وارد استفاده می‌گردد. هدف سوم، مقایسه خوشه‌های حاصل از روش‌های وارد- PCA و وارد- PA می‌باشد که در آن میزان صحت و دقت روش PA در تعیین ورودی‌های موثر و در شناخت خوشه‌های هم اقلیم مشخص می‌گردد. هدف چهارم، تعیین ایستگاه‌های شاخص (ایستگاه‌های حاوی حداکثر اطلاعات اقلیم بارش) با استفاده از روش PA است.

مواد و روش‌ها

منطقه و داده‌های مورد استفاده

منطقه مورد مطالعه کل کشور ایران است. تعداد ۳۱ ایستگاه، جهت مطالعه حاضر انتخاب شدند. شکل ۱ موقعیت ایستگاه‌ها را در کل کشور نشان می‌دهد.

در این تحقیق، از داده‌های بارش ماهانه استفاده شده است. به غیر از ایستگاه قائم‌شهر، سایر ایستگاه‌ها مربوط به مراکز استان‌ها می‌باشند. دوره آماری ۱۹۸۷ تا ۲۰۱۰ به‌عنوان دوره مشترک کلیه ایستگاه‌ها انتخاب گردید. متوسط بارش سالانه کشور ایران برابر با ۲۵۰ میلی‌متر می‌باشد.

روش تجزیه به مؤلفه‌های اصلی (PCA)

www.SID.ir

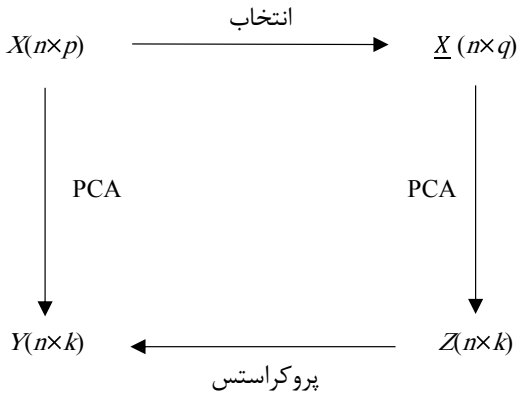
Archive of SID

روش PCA شاید قدیمی‌ترین روش چند متغیره بوده است که اولین بار در سال ۱۹۰۱ توسط پیرسون و بعدا در سال ۱۹۳۳ توسط هتلینگ معرفی شد (جولیف ۱۹۸۶). ایده اصلی PCA کاهش ابعاد داده‌هایی است که همبستگی نسبی بالایی (از نظر قدر مطلق) با همدیگر دارند. کاهش ابعاد داده‌ها به‌کمک تبدیل این داده‌ها به متغیرهای جدید (PCها یا مولفه‌های مستقل از هم) آن‌چنان صورت می‌گیرد که اولا چند مولفه اول دربرگیرنده بخش زیادی از تغییرات کل داده‌ها باشند و ثانیا تغییرات مولفه اول بیشتر از دوم، مولفه دوم بیشتر از سوم و ... باشد (جانسون و هانسون ۱۹۹۵).

روش تجزیه پروکراستس (PA)

اگر بخواهیم سیمای کلی یک سری از داده‌های چند متغیره را با یکی از روش‌های کاهش ابعاد داده‌ها، مانند PCA نشان دهیم، لازم است تعداد مولفه‌های غالب (k) مشخص شده و سری مقادیر PCS منتخب در مورد این مولفه‌ها محاسبه شوند. مقادیر مولفه‌ها ترکیب خطی از همه متغیرها (به تعداد p) می‌باشند. اگر p زیاد باشد، تعداد ضرایب نیز زیاد خواهد بود. بنابراین، نام‌گذاری و تفسیر مولفه‌ها با مشکل روبه‌رو می‌گردد. یکی از راه‌های حل این مشکل، انتخاب تعداد معقولی متغیر (به تعداد q) از بین همه متغیرها (p) می‌باشد، به‌طوریکه، اطلاعات همه متغیرهای کاندیدا (بطور تقریبی) در بطن متغیرهای انتخاب شده موجود باشد.

در این مطالعه، از روش پیشنهادی کرزانوفسکی (۱۹۸۷) استفاده شد. روش مذکور در ایران توسط دین‌پژوه (۱۳۸۲) برای تشخیص متغیرهای مهم در اقلیم بارش ایران مورد استفاده قرار گرفت. در این روش، لازم است اولاً تعداد عناصر زیر مجموعه منتخب نسبت به مجموعه مرجع خیلی کم باشد ($q<<p$) و ثانیاً تعداد این عناصر برابر یا بیشتر از تعداد مولفه‌های اصلی منتخب باشد ($q\geq k$). در این روش، انتخاب متغیرها براساس روش بهینه‌سازی با کمینه کردن تابع هدف که با "M^2" نمایش داده می‌شود، انجام می‌گیرد. شکل (۲) مراحل کلی تجزیه پروکراستس را نشان می‌دهد.



شکل ۲- چرخه انتخاب متغیرها با روش پروکراستس (دین‌پژوه و همکاران ۱۳۸۲).

در شکل ۲، X ماتریس کل داده‌های استاندارد شده و $\underline{X}$ ماتریس داده‌های استاندارد شده پس از حذف متغیرها (به‌صورت مرحله‌ای) می‌باشند. ماتریس Y که بر اساس همه متغیرهای کاندیدا محاسبه می‌شود را اصطلاحا آرایش واقعی[4] و ماتریس Z که مبتنی بر متغیرهای منتخب می‌باشد را به عنوان آرایش تقریبی[5] داده‌ها می‌نامند. تجزیه پروکراستس، اختلاف مجموع مربعات نقاط متناظر این دو آرایش را ارزیابی می‌کند. هرچه متغیرهای منتخب دقیق‌تر انتخاب شوند، این دو آرایش مشابه هم بوده و تفاوت بین آرایش تقریبی و واقعی کمتر خواهد بود. حاصل نتایج تجزیه پروکراستس محاسبه اختلاف بین مجموعه مربعات دو آرایش تقریبی و واقعی می‌باشد که از معادله ۱ بدست می‌آید.

$$M^2 = Trace\{YY' + ZZ' - 2ZQ'Y'\} \qquad [۱]$$

که در آن، Q از رابطه زیر بدست می‌آید:

$$Q = VU' \qquad [۲]$$

که در آن، ماتریس U و V از عملیات تجزیه مقدار منفرد[6] ماتریس $Z'Y$ به ابعاد $k \times k$ حاصل می‌شود که شکل ریاضی آن چنین است:

$$Z'Y = U\sum V' \qquad [۳]$$

که در آن روابط $UU' = I_k$، $VV' = V'V = I_k$ و $\sum = diag(\sigma_1, \sigma_2, ..., \sigma_k)$ برقرار می‌باشند. I_k ماتریس همانی به ابعاد $k \times k$ است.

برای هر یک از زیر مجموعه‌های مربوط به مجموعه مرجع متغیرها که حداقل q عضو داشتند، کمیت M^2 تعیین شد. بهترین زیر مجموعه مرجع متغیرها چنان انتخاب گردید

[4]- True Configuration

[5]- Approximate Configuration

[6]- Singular value decomposition (SVD)

www.SID.ir

Archive of SID

که کمترین مقدار کمیت M^2 متعلق به آن زیر مجموعه بود (کرزانوفسکی ۱۹۸۷). کلیه محاسبات مربوط به روش‌های PA و PCA براساس کدنویسی در محیط برنامه‌نویسی متلب انجام گرفته است.

روش خوشه‌بندی سلسله مراتبی وارد

تجزیه خوشه‌ای با روش حداقل واریانس وارد، یکی از انواع روش‌های تجزیه خوشه‌ای سلسله مراتبی است که در هر مرحله از الحاق خوشه‌ها به یکدیگر، امکان جابجایی فرد (ایستگاه) از یک خوشه به خوشه دیگر وجود ندارد. هدف مشترک همه روش‌های تجزیه خوشه‌ای این است که شباهت افراد (ایستگاه‌ها) را در داخل خوشه بیشینه، ولی در بین گروه‌ها کمینه شود. در هر گام از اعمال روش وارد برای خوشه‌بندی، هدف کمینه کردن میانگین مربعات انحراف افراد (ایستگاه‌ها) از میانگین گروه در درون یک خوشه معین به‌عنوان شاخص تشابه می‌باشد. در درون یک خوشه، W (شاخص تشابه) به‌عنوان تابع هدف می‌باشد که از رابطه ۴ محاسبه می‌گردد (کالکستین و همکاران ۱۹۸۷):

$$W = \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{N_k}(X_{ijk} - X_{.jk})^2 \qquad [۴]$$

در فرمول ۴، X_{ijk} مقدار متغیر jام در فرد iام از خوشه kام، N_k تعداد افراد (ایستگاه‌ها) در خوشه kام، J تعداد کل متغیرهای مورد بررسی و K تعداد کل خوشه‌ها می‌باشد. X_{jk} میانگین ایستگاه‌ها (به تعداد N_K) برای متغیر jام از خوشه kام است.

با فرض ملحق شدن خوشه ۱ به خوشه ۲ در روش وارد، تغییر در مقدار W (ΔW) فقط وابسته به رابطه بین دو خوشه ملحق شده می‌باشد و به رابطه بین دیگر خوشه‌ها بستگی ندارد. با استفاده از رابطه ۴ برای دو خوشه (با N_1 و N_2 مشاهده) مقدار ΔW از رابطه ۵ محاسبه می‌گردد.

$$\Delta W = W' - W = \qquad (۵)$$
$$N_1\sum_{j=1}^{J}(X_{.j.} - X_{.j1})^2 + N_2\sum_{j=1}^{J}(X_{.j.} - X_{.j2})^2$$

در فرمول ۵، $W=W$ قبل از الحاق و $W' = W$ بعد از الحاق می‌باشد. روش وارد، خوشه‌های با کمترین مقدار ΔW را به یکدیگر ملحق می‌کند، که شامل مربعات فاصله بین مرکز جدید ($X_{.j.}$) و دو مرکز اولیه ($X_{.j1}$ و $X_{.j2}$) می‌باشد که با تعداد مشاهدات (N_1 و N_2) وزن‌دار شده است.

پس از رسم نمودار درختی، تعیین تعداد خوشه‌ها به‌صورت چشمی و با نگاه کردن به خط راست افقی که نمودار درختی حاصله را به تعداد مناسب خوشه‌ها تفکیک می‌کند، انجام گرفته است. در این تحقیق از نرم افزار SPSS22 برای خوشه‌بندی وارد استفاده شده است.

دو سری داده، جهت ارزیابی و خوشه‌بندی بارش استان‌های کشور مورد استفاده قرار گرفت. دسته اول شامل متوسط بارش هر ماه و دسته دوم شامل سری زمانی بارش ماهانه ایستگاه همدید هر استان‌ها می‌باشد. در بخش اول، میانگین دراز مدت بارش هر ماه و در کلیه ایستگاه‌های هواشناسی منتخب که به‌عنوان ورودی روش‌های PCA و PA درنظر گرفته شدند. ابعاد ماتریس ورودی این بخش ($n*m$) بود که در آن m تعداد ماه‌ها (۱۲) و n تعداد ایستگاه‌های هواشناسی منتخب (۳۱) می‌باشد. با استفاده از روش PCA و روش خوشه‌بندی وارد، کلاس‌بندی ایستگاه‌ها براساس میانگین دراز مدت بارش-های ماهانه صورت گرفت. در مرحله بعدی، با استفاده از روش PA، ماه‌هایی که بارش آن‌ها بیشترین تاثیر را در تمایز بین کلاس‌ها داشت، تعیین گردید.

در بخش دوم، ایستگاه‌های هواشناسی به‌عنوان متغیر ورودی روش‌های PCA و PA معرفی گردید. ابعاد ماتریس ورودی این بخش ($n'*m'$) بود که در آن m' تعداد ایستگاه-های هواشناسی (۳۱) و n' تعداد مشاهدات مربوط به سری زمانی بارش ماهانه‌های هر ایستگاه (۲۸۸) بود. در این بخش، با استفاده از روش PCA، تعداد مولفه اصلی معنی‌دار تعیین می‌گردد. در مرحله بعد، بوسیله روش PA، هویت ایستگاه‌هایی که باعث تفکیک هرچه بهتر خوشه‌ها می‌شدند، مشخص گردید. این ایستگاه‌ها در واقع نماینده کل اقلیم بارشی ایستگاه‌های منتخب شناخته شدند. کلیه داده‌ها قبل از استفاده استاندارد شدند.

نتایج و بحث

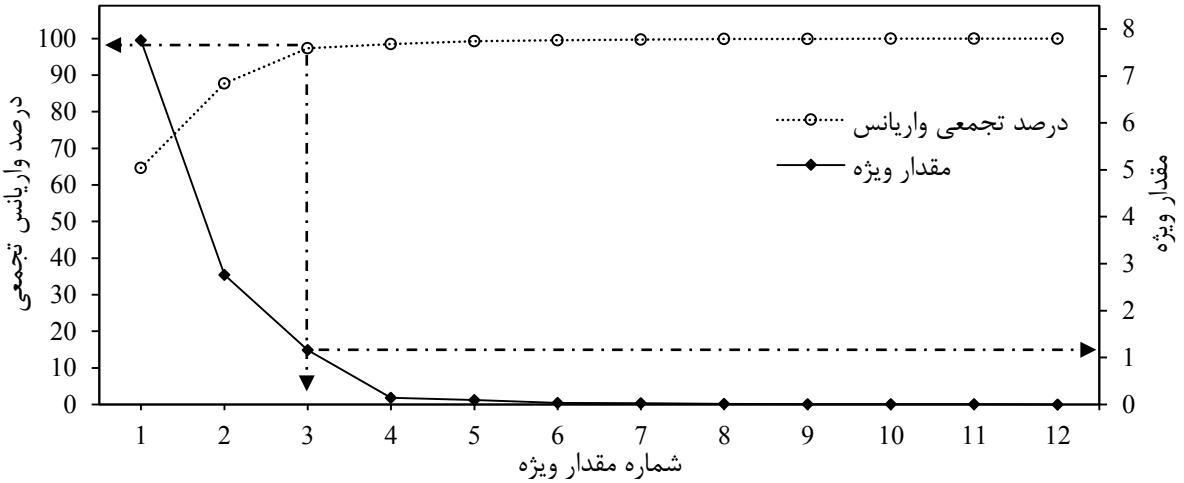
تعیین مؤلفه‌های اصلی متوسط بارش ماهانه

بطوریکه قبلا اشاره شد، تعیین مولفه‌های اصلی با استفاده از روش PCA صورت پذیرفته است. براساس داده‌های میانگین بارش هر ماه در کل دوره آماری و کلیه ایستگاه-های هواشناسی، ماتریس ورودی برای روش PCA ایجاد گردید. ابعاد این ماتریس ($n*m$) بود که در آن تعداد m=۱۲ ستون (به تعداد ماه‌های سال) و n=۳۱ ردیف (به تعداد ایستگاه‌های هواشناسی منتخب) می‌باشند. پس از انجام

www.SID.ir

Archive of SID

PCA نمودار صخره‌ای[7] رسم شد. در این مطالعه، تعدادی از مولفه‌های اصلی اول که مقدار ویژه آن‌ها بیش از واحد بود به‌عنوان مولفه‌های اصلی در نظر گرفته شدند.
شکل ۳ واریانس مربوط به هر یک از دوازده مولفه اصلی، به‌همراه درصد واریانس تجمعی آن‌ها را نشان می‌دهد. با توجه به این شکل سه مؤلفه اصلی اول (k=۳) بیش از ۹۷ درصد کل واریانس داده‌ها را تامین می‌کنند. بنابراین، از مقادیر PCS این سه مؤلفه اصلی در روش PA استفاده شد.



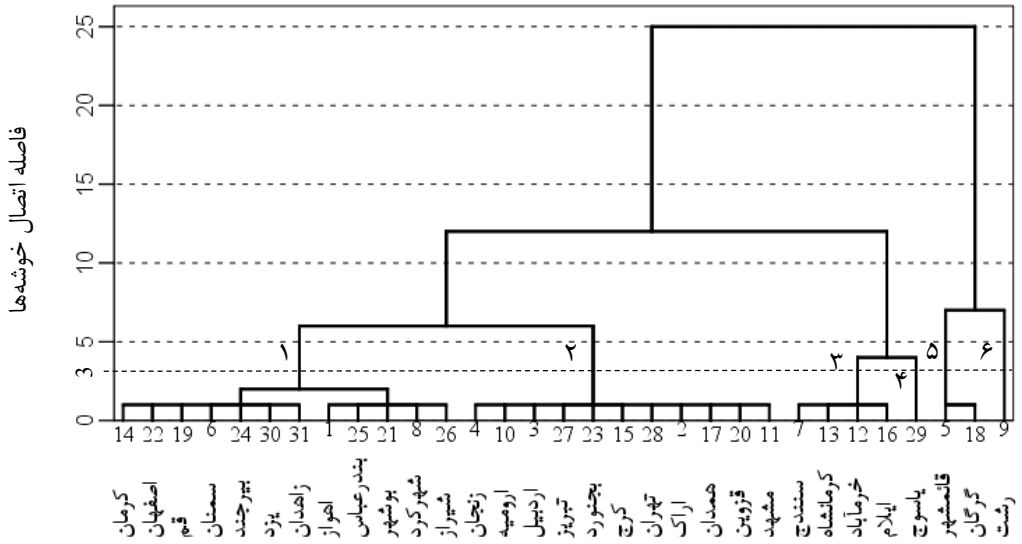
شکل ۳- نمودار صخره‌ای حاصل از PCA در بخش اول مطالعه حاضر.

پس از تعیین مولفه‌های اصلی، خوشه‌بندی ایستگاه‌های همدید با استفاده از روش وارد انجام شد. دندروگرام خوشه‌های حاصل، به‌همراه شماره هر خوشه، در شکل ۴ نشان داده شده‌اند. با توجه به شکل مذکور و با در نظر گرفتن فاصله بین خوشه‌ای برابر با ۳، ایستگاه‌ها به شش خوشه متمایز تقسیم شدند. ایستگاه‌هایی که در یک خوشه قرار گرفته‌اند، اقلیم مشابهی در میانگین بارش ماهانه دارند.
در ادامه، هویت ماه‌هایی که بارش آن‌ها تأثیر بیشتری در خوشه‌بندی ایستگاه‌های هواشناسی دارند، با روش PA شناسایی گردید. در این مطالعه هر دو مقدار k و q برابر با ۳ در نظر گرفته شدند. پس از تجزیه داده‌ها به مولفه‌های اصلی، دوازده واریانس به تعداد ۱۲ ماه سال بدست آمده است. مقادیر $\text{Min } M^2$ به‌ازای حذف ستون داده‌ها در هر مرحله از محاسبات، در شکل ۵ نشان داده شده است. چون مقدار q=۳ فرض شده است، بنابراین لازم است از کل ۱۲ ماه سال فقط ۳ ماه (که بارش آن‌ها نشان دهنده بخش اعظم واریانس بارش کل ۱۲ ماه سال است) معرفی گردد. به‌عبارت دیگر، ۹ ماه از سال (در ۹ مرحله پیاپی روش PA) که اطلاعات بارش آن‌ها از کل واریانس ناچیز است، حذف گردد.

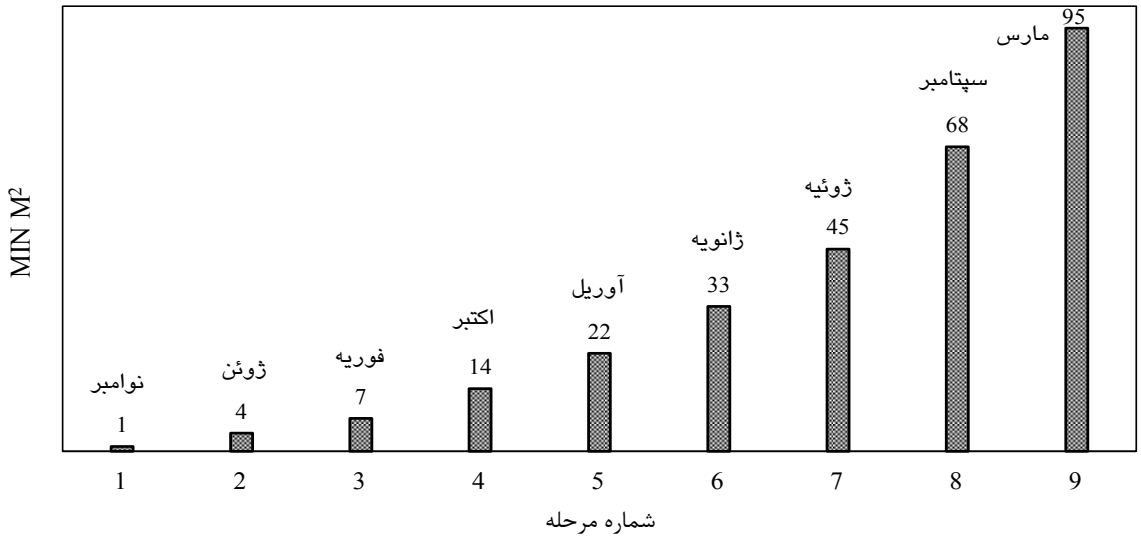
[7]- Scree plot

www.SID.ir

Archive of SID



شکل ۴- دندروگرام حاصل از روش PCA ایستگاه‌های مورد مطالعه با الگوریتم خوشه‌بندی سلسله مراتبی وارد.



شکل ۵- مقادیر Min M^2 به‌ازای حذف متغیرها در هر مرحله.

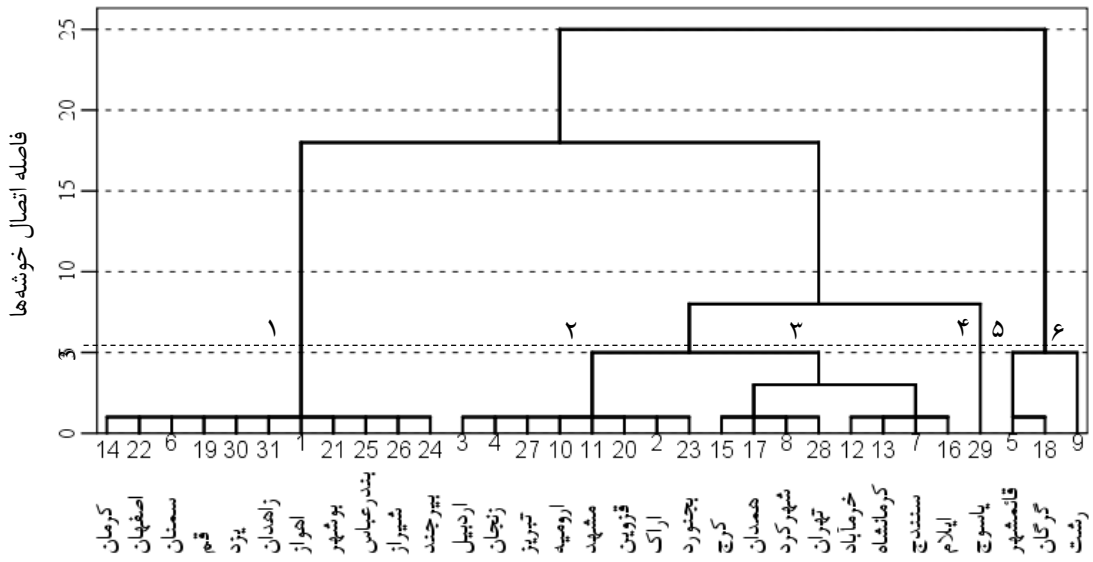
با توجه به شکل ۵، مجموع مجذور اختلاف (M^2) بین روش‌های PCA (به‌ازای ۱۲ ستون داده میانگین ماهانه ورودی) و PA (به‌ازای ۳ ستون داده ماهانه ورودی) برابر ۹۵ بدست آمده است.

نتایج نشان داد که داده‌های میانگین ماهانه حذف شده در هر مرحله به‌ترتیب، شامل ماه‌های نوامبر، ژوئن، فوریه، اکتبر، آوریل، ژانویه، ژوئیه، سپتامبر و مارس بودند. پس طبق روش PA، مولفه‌های اصلی بدست آمده از داده‌های بارش ماهانه سه ماه (مه، اوت و دسامبر) بیش‌ترین مطابقت را با مولفه‌های اصلی حاصل از کل داده‌ها داشته‌اند. بنابراین، در گام بعدی خوشه‌بندی به‌ازای مقادیر PCS بدست آمده از داده‌های بارش ماهانه سه ماه مذکور با استفاده از روش وارد صورت گرفت. شکل ۶ نمودار خوشه‌ای حاصل را نشان می‌دهد. با توجه به این شکل و با در نظر گرفتن فاصله بین خوشه‌ای برابر با ۳، ایستگاه‌های همدید کشور به شش خوشه متمایز تفکیک شد. این شش خوشه، شباهت بسیار زیادی به خوشه‌های بدست آمده از مولفه‌های اصلی کل داده‌ها (شکل ۴) دارند. این شباهت زیاد حاکی از مناسب و دقیق بودن روش PA در تشخیص هویت موثرترین ماه‌ها در تعیین مولفه‌های اصلی می‌باشد. با توجه به شکل‌های ۴ و ۶، خوشه‌بندی-

www.SID.ir

Archive of SID

های صورت گرفته با مولفه‌های اصلی حاصل از روش‌های PCA و PA در شش کلاس مجزا قرار داشته‌اند. جدول ۱ میزان شباهت این خوشه‌ها را نشان می‌دهد. این جدول، براساس شماره ایستگاه‌ها که در دندروگرام‌ها مشخص گردیده، تنظیم شده است.



شکل ۶- نمودار خوشه‌ای حاصل از خوشه‌بندی وارد با اطلاعات ۳ ماه معرفی شده با روش تجزیه پروکراستس.

جدول ۱- مقایسه نتایج خوشه‌های حاصل از خوشه‌بندی وارد با روش‌های PCA و PA.

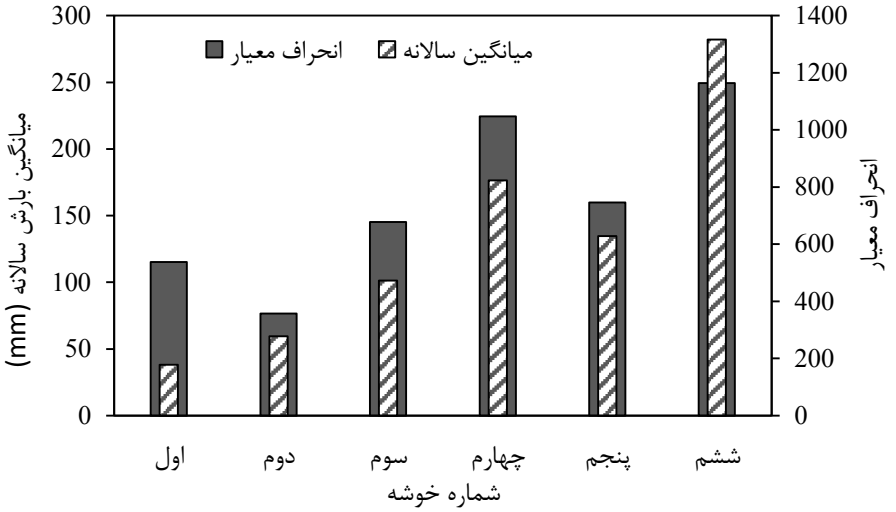
شماره ایستگاه همدید در خوشه‌های شش گانه با روش PA					
اول	دوم	سوم	چهارم	پنجم	ششم
۱	۲	۷	۲۹	۵	۹
۶	۳	۸		۱۸	
۱۴	۴	۱۲			
۱۹	۱۰	۱۳			
۲۱	۱۱	۱۵			
۲۲	۲۰	۱۶			
۲۴	۲۳	۱۷			
۲۵	۲۷	۲۸			
۲۶					
۳۰					
۳۱					

شماره ایستگاه همدید در خوشه‌های شش گانه با روش PCA					
اول	دوم	سوم	چهارم	پنجم	ششم
۱	۲	۷	۲۹	۵	۹
۶	۳	۱۲		۱۸	
۸	۴	۱۳			
۱۴	۱۰	۱۶			
۱۹	۱۱				
۲۱	۱۵				
۲۲	۱۷				
۲۴	۲۰				
۲۵	۲۳				
۲۶	۲۷				
۳۰	۲۸				
۳۱					

با توجه به جدول فوق، به‌غیر از ۴ ایستگاه همدید با شماره‌های ۸، ۱۵، ۱۷ و ۲۸، سایر اعضای خوشه‌های حاصل از روش PA با اعضای خوشه‌های حاصل از روش PCA هم‌خوانی کامل دارد. خوشه اول PCA یک ایستگاه بیشتر از خوشه اول PA دارد و همچنین خوشه دوم PCA نیز سه ایستگاه بیشتر از خوشه دوم PA دارد. در عوض، خوشه سوم PA، علاوه بر ایستگاه‌های خوشه سوم PCA، چهار ایستگاه بیشتر در خود جای داده است. اعضای خوشه‌های چهارم، پنجم و ششم، PCA و PA کاملا مشابه یکدیگر می‌باشند. این مشابهت زیاد، حاکی از تشخیص درست روش PA در تعیین ماه‌های موثرتر در تعیین مولفه‌های اصلی موثر می‌باشد.

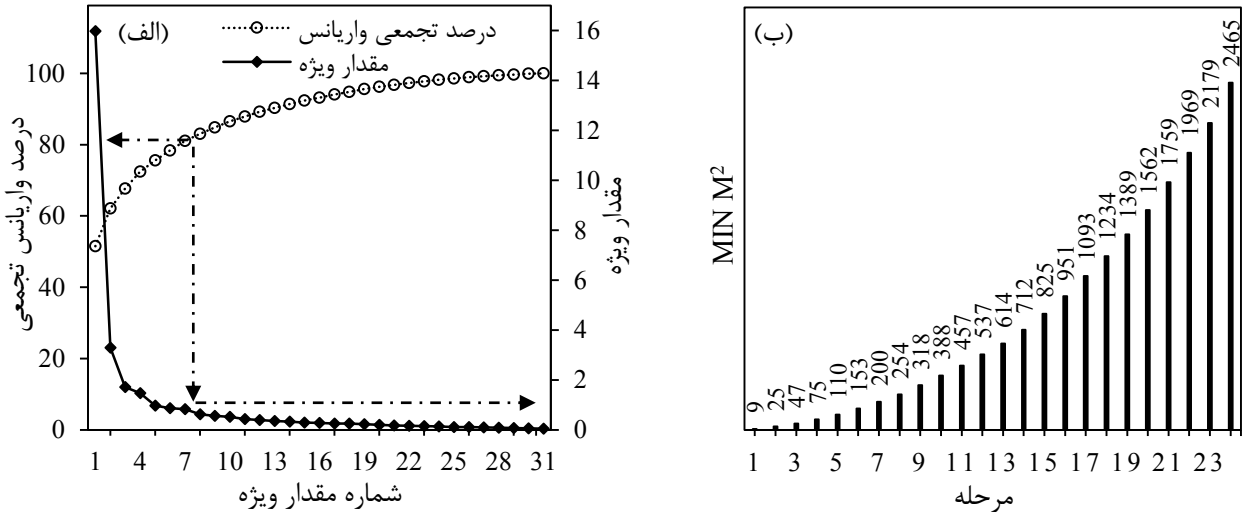
www.SID.ir

Archive of SID



شکل ۷- مقادیر میانگین و انحراف معیار بارش‌های سالانه خوشه‌های شش گانه.

جهت مقایسه خوشه‌ها، میانگین بارش سالانه هر یک از خوشه‌ها و همچنین انحراف معیار آن‌ها محاسبه گردید (شکل ۷). حداقل اختلاف میانگین بارش‌های سالانه، برابر با ۹۹ میلی‌متر و بین خوشه‌های اول و دوم می‌باشد و همچنین بیشترین اختلاف، مربوط به خوشه‌های ششم و اول می‌باشد که برابر ۱۱۳۸ میلی‌متر است. سایر خوشه‌ها نیز بین ۹۹ تا ۱۱۳۸ میلی‌متر در میانگین بارش سالانه با یکدیگر اختلاف دارند. خوشه‌ها از لحاظ انحراف معیار نیز تفاوت قابل توجهی با یکدیگر داشته‌اند. مقادیر انحراف معیار بارش‌های سالانه خوشه‌ها بین ۷۷ و ۲۴۹ می‌باشد که بیشترین آن‌ها به‌ترتیب مربوط به خوشه ششم (ایستگاه رشت) و چهارم (ایستگاه یاسوج) می‌باشد. قطعا یکی از علل اصلی قرار گرفتن این ایستگاه‌ها در خوشه‌های تکی، میزان انحراف معیار آن‌ها می‌باشد. همانطور که مشخص است، دو علت اصلی تفکیک خوشه‌ها از یکدیگر، اختلاف میانگین بارش و انحراف معیار بارش‌های خوشه-ها هستند. بطوریکه قبلا اشاره شد، بخش دوم این مطالعه مربوط به تعیین ایستگاه‌های شاخص است. جهت دستیابی به این ایستگاه‌ها از روش PA استفاده شده است. پس از اعمال روش PCA، مقدار واریانس و درصد واریانس تجمعی بدست آمد که در شکل ۸-الف نشان داده شده است. با توجه به این شکل از روش PA به‌ازای p=۷ و q=۷ جهت تعیین ایستگاه‌های شاخص استفاده شده است. با اعمال PA بر روی ماتریس ورودی، مقادیر مربوط به Min M^2 در شکل ۸-ب نشان داده شده است.



شکل ۸- (الف) نمودار صخره‌ای حاصل از PCA و (ب) مقادیر Min M^2 به‌ازای حذف متغیرها در هر مرحله.

www.SID.ir

Archive of SID

پس از حذف هر یک از ایستگاه‌های غیر مهم در هر یک از ۲۴ مرحله PA، هفت ایستگاه همدید رشت، ارومیه، ایلام، گرگان، تهران، یاسوج و زاهدان به‌عنوان ایستگاه-های شاخص کشور ایران تعیین شدند. با توجه به این نتایج، هر یک از شش خوشه نشان داده شده در شکل ۳ حداقل دارای یک ایستگاه شاخص می‌باشند.

ایستگاه‌های همدید زاهدان، تهران، ارومیه، ایلام، یاسوج، گرگان و رشت به‌ترتیب، به‌عنوان ایستگاه‌های شاخص خوشه‌های اول، دوم، دوم، سوم، چهارم، پنجم و ششم تعیین شدند.

شکل ۹- خوشه‌های یک تا شش، بدست آمده از وارد- PCA و ایستگاه‌های شاخص حاصل از وارد- PA در ایران.

شکل ۹ خوشه‌های ایجاد شده به‌همراه ایستگاه‌های شاخص آن‌ها را نشان می‌دهد. از شاخص دومارتن جهت تعیین اقلیم هر یک از خوشه‌های شش‌گانه استفاده گردید. مقدار این شاخص برای هر یک از خوشه‌های اول تا ششم به‌ترتیب، برابر ۶، ۱۲، ۱۸، ۳۳، ۲۳ و ۵۰ بدست آمده است که نشان دهنده اقلیم‌های خشک، نیمه خشک، نیمه خشک، مرطوب، مدیترانه‌ای و بسیار مرطوب برای هر یک از خوشه‌ها می‌باشند.

خسروی و همکاران (۱۳۹۳) نیز کشور ایران را با استفاده از روش خوشه‌بندی وارد و به‌ازای دما و بارش به پنج ناحیه همگن اصلی تقسیم کردند. خوشه‌بندی‌های حاصل از تحقیق حاضر با نتایج خوشه‌بندی خسروی و همکاران (۱۳۹۳) در بعضی نواحی تطابق کامل ندارد. علت این عدم تطابق را می‌توان استفاده مجموعه داده‌های مختلف دانست. شیروانی و ناظم‌السادات (۱۳۹۱) نیز به-ازای سری زمانی بارش‌های ماهانه، ۴۲ ایستگاه همدیدی و روش خوشه‌بندی وارد، کشور ایران را به شش منطقه همگن طبقه‌بندی کرده‌اند. لازم به‌ذکر است که روش کار تحقیق حاضر و شیروانی و ناظم‌السادات (۱۳۹۱) تقریبا یکی می‌باشد، ولی متغیرهای مورد استفاده دو تحقیق با یکدیگر متفاوت هستند. به‌طوری‌که شیروانی و ناظم-السادات (۱۳۹۱) سری زمانی داده‌های بارش ماهانه و تحقیق حاضر میانگین بارش‌های ماهانه را به‌عنوان متغیر ورودی در نظر گرفته است. بنابراین، اختلافی که در برخی نواحی خوشه‌بندی ملاحظه می‌گردد به‌علت استفاده از متغیرهای مختلف می‌باشد. نتایج خوشه‌بندی مسعودیان و همکاران (۱۳۹۰) که غرب و شمال غرب ایران را مورد بررسی قرارداده‌اند، همچون این تحقیق استان‌های اردبیل، ارومیه، آذربایجان شرقی، همدان و قزوین را در یک

www.SID.ir

Archive of SID

خوشه و ایستگاه‌های سنندج و کرمانشاه را در خوشه دیگر طبقه‌بندی کرده است. خورشیددوست و شیرزاد (۱۳۹۳) نیز ساختار همگنی غرب و شرق ناحیه شمالی ایران (استان‌های گیلان، مازندران و گرگان) را جدا از یکدیگر معرفی کرده‌اند که این موضوع با نتایج این تحقیق مطابقت دارد.

نتیجه‌گیری کلی

در این تحقیق، از داده‌های بارش ماهانه ۳۱ ایستگاه همدید جهت خوشه‌بندی استفاده شده است. با توجه به بررسی‌های انجام شده، ترکیب روش‌های PCA و PA با روش خوشه‌بندی وارد جهت کلاس‌بندی بارش و همچنین تعیین ایستگاه‌های شاخص کشور ایران صورت نگرفته است. بنابراین، سعی گردید با استفاده از روش‌های مذکور، در گام اول خوشه‌بندی بارش ماهانه کشور صورت پذیرد و سپس در گام دوم ایستگاه‌های شاخص کشور تعیین گردند.

در گام اول، با استفاده از روش تجزیه به مولفه‌های اصلی، سه مولفه اصلی اول با داشتن ۹۷ درصد کل واریانس داده‌ها (به‌عنوان معیارهای اصلی خوشه‌بندی) درنظر گرفته شدند. سپس با استفاده از روش خوشه‌بندی سلسله مراتبی وارد و سه مولفه اصلی اول ذکر شده، ایستگاه‌های همدید کشور خوشه‌بندی گردیدند. نتایج نشان داد که با در نظر گرفتن فاصله اقلیدسی برابر با سه، کشور ایران دارای شش خوشه اصلی می‌باشد.

سپس با استفاده از روش PA، سه ماه از سال (که مولفه‌های اصلی آن‌ها با مولفه‌های اصلی حاصل از کل داده‌ها هم‌خوانی بیشتری دارد) انتخاب گردید. نتایج محاسبات نشان داد که داده‌های بارش ماه‌های مه، اوت و دسامبر، بیش‌ترین مطابقت را با مولفه‌های اصلی بدست آمده از کل داده‌ها دارند. بنابراین، با استفاده از مولفه‌های اصلی بدست آمده از این سه متغیر و روش خوشه‌بندی وارد، ایستگاه‌های همدید کشور ایران، برای بار دوم خوشه‌بندی شدند. در این خوشه‌بندی نیز به‌ازای فاصله اقلیدسی برابر با سه، شش خوشه با ناحیه اقلیمی مشابه ایجاد گردید. خوشه‌های ایجاد شده به‌ازای سه متغیر (روش PA) با خوشه‌بندی به‌ازای دوازده متغیر (روش PCA) مشابهت بسیار زیادی با یکدیگر دارند.

مقایسه این دو روش خوشه‌بندی نشان داد که روش PA با استفاده از سه متغیر توانسته است، ۲۷ ایستگاه همدید (۸۷ درصد ایستگاه‌ها) را دقیقا همانند روش PCA (با دوازده متغیر) در خوشه‌های مشابه طبقه-بندی کند. این روش، فقط در تعیین خوشه چهار ایستگاه همدید با روش PCA تفاوت دارد. در واقع، جهت تعیین ایستگاه‌های هم اقلیم نیازی به استفاده از داده‌های کل ماه-ها نیست و فقط با داده‌های سه ماه مذکور می‌توان ایستگاه‌های هم اقلیم را مشخص نمود.

در گام دوم، از روش PCA جهت تعیین تعداد مولفه‌های اصلی مناسب که بیشترین اطلاعات کل ایستگاه-های همدید را دارند، استفاده گردید. سپس با استفاده از روش PA نیز هفت ایستگاه شاخص تعیین گردید.

ایستگاه‌های همدید شاخص شامل زاهدان (خوشه اول)، تهران (خوشه دوم)، ارومیه (خوشه دوم)، ایلام (خوشه سوم)، یاسوج (خوشه چهارم)، گرگان (خوشه پنجم) و رشت (خوشه ششم) می‌باشند. همانطور که در شکل ۸ نشان داده شده است، هر یک از شش خوشه تعیین شده حداقل دارای یک ایستگاه شاخص می‌باشند. در واقع، این هفت ایستگاه شاخص حاوی بیشتر اطلاعات کل ایستگاه‌های همدید کشور می‌باشند. بنابراین، ایستگاه‌های شاخص می‌توانند محل نصب تجهیزات گران قیمتی که استفاده از آن‌ها در کلیه ایستگاه‌ها مقدور نیست، باشند. با توجه به نتایج بدست آمده می‌توان روش PA را به‌عنوان یکی از روش‌های مناسب جهت شناسایی متغیرهای موثر در کاهش حجم داده‌های ورودی، همراه با حداکثر اطلاعات موجود در کل داده‌ها معرفی نمود.

منابع مورد استفاده

Alizadeh A, 2010. Principles of Applied Hydrology. 30th edition. Astan Quds Razavi, Mashhad. (In Persian)

Arvin AA, Mofidi Khaje AM and Mazini F, 2012. Selecting Place and Time Patterns in Golestan Province Rainfall by Clustering Analysis. Geographical planning of space quarterly journal 2(6): 117-132. (In Persian)

Carvalho MJ, Melo-Gonçalves P, Teixeira JC and Rocha A, 2016. Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation. Physics and Chemistry of the Earth, Parts A/B/C, 94: 22-28.

www.SID.ir

Archive of SID

Dinpashoh Y, 2003. Meteorological drought analysis using the pattern analysis. Ph.D. dissertation, University of Tabriz, Tabriz. (In Persian)

Dinpashoh Y, Fakheri-Fard A, Moghaddam M, Jahanbakhsh S and Mirnia M, 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. Journal of Hydrology 297(1): 109-123.

Fallahi B, Fakheri Fard A, Dinpajooh Y and Darbandi S, 2012. Regionalization of Northwest Iran Based on Daily Rainfalls and Rain's Time Intervals Using PCA, Ward and K-mean Methods. Journal of Water and Soil 26(4): 979-989. (In Persian)

Gocic, M and Trajkovic S, 2014. Spatio-temporal patterns of precipitation in Serbia. Theoretical and Applied Climatology 117(3-4): 419-431.

Jahanbakhsh Asl S, Abtahi V, Ghorbani MA, Tadayyoni M and Valayi A, 2015. Temporal and Spatial Distribution of Rainfall in Tabriz County Using Hierarchical Cluster Analysis. Journal of geographic space 15(50): 59-81. (In Persian)

Johnson GL and Hanson CL, 1995. Topographic and atmospheric influences on precipitation variability over a mountainous watershed. Journal of Applied Meteorology 34(1): 68-87.

Jolliffe IT, 1986. Principal Component Analysis. Springer-Verlag, 271p.

Kalkstein LS, Tan G and Skindlov JA, 1987. An evaluation of three clustering procedures for use in synoptic climatological classification. Journal of Climate and Applied Meteorology 26(6): 717-730.

Khorshiddoust AM and Shirzad AA, 2014. The Study of Precipitation in North of Iran Using Cluster and Discriminative Function Analyses. Journal of geography and planning, 18(49): 110-118. (In Persian)

Khosravi M, Doostkamiyan M, Mirmosavi SH, Bayat A and Beygrezae A, 2014. Classification temperature and precipitation in Iran using geostatistical methods and cluster analysis. Journal of Management System 4(13): 121-132. (In Persian)

Krzanowski WJ, 1987. Selection of variables to preserve multivariate data structure, using principal components. Applied Statistics 36(1): 22-33.

Masoudian A, Darand M and Karsaz S, 2011. Rainfall zoning of west and northwest of Iran by cluster analysis. Journal of Physical Geography 4(11): 35-44. (In Persian)

Nadi M and Khalili A, 2013. Classification of Iran's precipitation climate using factor-cluster analysis method. Journal of Irrigation Science and Engineering 44(3): 235-242. (In Persian)

Nam W, Shin H, Jung Y, Joo K and Heo JH, 2015. Delineation of the climatic rainfall regions of South Korea based on a multivariate analysis and regional rainfall frequency analyses. International Journal of Climatology 35(5): 777-793.

Nazemosadat SMJ, Baigi B and Amin S, 2004. Application of the Principal Component Analysis for the Regionalization of Winter Precipitation over Boushehr, Fars, and Kohgiloye & Boyerahmad Provinces. Journal of Science and Technology of Agriculture and Natural Resources 7(1): 61-72. (In Persian)

Nazeri Tahroudi M, Khalili K and Ahmadi F, 2016. Spatial and Regional Analysis of Precipitation Trend over Iran in the Last Half of Century. Journal of Water and Soil 30(2): 643-654. (In Persian)

Othman M, Ash'aari ZH and Mohamad ND, 2015. Long-term daily rainfall pattern recognition: Application of principal component analysis. Procedia Environmental Sciences 30: 127-132.

Pansera WA, Gomes BM, Boas AV and Mello EL, 2013. Clustering rainfall stations aiming regional frequency analysis. Journal of Food, Agriculture & Environment 11(2): 877-885.

Shirvani A and Nazemosadat SMJ, 2012. Regionalization of precipitation in Iran using principal components and cluster analysis. Iran Water Resource Research 8(1): 81-85.

www.SID.ir