نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۲۶، شماره ۴، مهر – آبان ۱۳۹۱، ص. ۹۸۹-۹۷۹

Journal of Water and Soil
Vol. 26, No.4, Sept.-Oct. 2012, p. 979-989

پهنه بندی شمالغرب ایران بر مبنای بارش‌های روزانه، فواصل زمانی بارش ها و روزهای بارانی باکاربرد روش‌هایPCA,Ward, K-mean

بهاره فلاحی[1*]- احمد فاخری فرد[2]- یعقوب دین‌پژوه[3]- صابره دربندی[4]

تاریخ دریافت:۹۰/۸/۹
تاریخ پذیرش:۹۱/۳/۲۱

چکیده

روش‌های تحلیلی کاهش ابعاد داده‌ها مانند آنالیز مولفه‌های اصلی برای درک صحیحی از عوامل مؤثر بر تغییرات اقلیمی با توجیه قسمت مهمی از کل واریانس با تعداد محدودی از مولفه‌های اصلی، در برنامه ریزی منابع آب حائز اهمیت می‌باشند. در این تحقیق روش PCA به عنوان تصویرکننده فضای اطلاعات روی محورهای محدود و معین، K-mean به عنوان روش خوشه‌سازی گروهی و Wardبه عنوان خوشه‌سازی سلسله مراتبی مورد استفاده قرار گرفته است. در تحقیق حاضر با استفاده از این روش‌ها و کاربرد داده‌های بارش روزانه ۶۰ ایستگاه هواشناسی در طول دوره آماری ۳۵ ساله (۲۰۰۴-۱۹۷۰)،۴ نوع پهنه‌بندی بر مبنای بارش‌های روزانه، سری شاخص فاصله–مقدار، فواصل زمانی بارش و سری روزهای بارانی انجام گرفت. آماره S برای آزمون همگنی مناطق بدست آمده به کار گرفته شد. نتایج نشان داد که روش PCA به لحاظ ماهیت خود فضای داده‌ها را روی محورهای اصلی تصویر می‌کند و ساختار واقعی فضا را نشان خواهد داد اما در روش‌های سلسله مراتبی که در هر مرحله خوشه یا اعضا را به هم متصل می‌کنند، از نظر ماهیت فضایی خوشه‌ها بیانگر ساختار واقعی نمی‌باشند لذا انتظار می‌رود که خوشه‌های حاصل از PCA واقعی‌تر از خوشه‌های حاصل از دیگر روش‌ها باشد. اما از نقطه نظر گسترش اطلاعات خوشه‌بندی بر مبنای همگنی، روش‌های خوشه‌بندی سلسله مراتبی بر خوشه‌بندی واقعی ارجحیت پیدا می‌کند.

واژه‌های کلیدی: پهنه بندی، تجزیه به مولفه های اصلی، تجزیه خوشه ای، وارد، K-mean

مقدمه

بارش تقریبا تنها منبع تامین آب ایران است. تفاوت مقدار بارندگی از محلی به محل دیگر و از زمانی به زمان دیگر از جمله ویژگی‌های مهم بارش در ایران بوده که می‌تواند به عنوان موضوع تحقیقات محققین کشور قرار گیرد. علیرغم گسترش دانش و فناوری هنوز علل این نوسان‌ها کاملا مشخص نشده، لیکن تاثیر پذیری شرایط اقتصادی، روابط اجتماعی وسیاسی مردم از وضعیت بارش‌های جوی ایران انکارناپذیر است (۱۴). یکی از مسائل اساسی در هیدرولوژی، تخمین بارش در یک ایستگاه نامشخص با استفاده از اطلاعات داده‌های ایستگاه‌های هواشناسی اطراف آن می‌باشد. این مسئله به خصوص در مناطق کوهستانی و ناهموار مانند شمالغرب ایران، که دارای تعداد ایستگاه‌های محدودی بوده و به دلیل تاثیرات کوهستانی زیاد، دارای اهمیت زیادی است. به کمک پهنه بندی[5] می‌توان اطلاعات نقطه‌ای را به سطح منطقه گسترش داد و پدیده‌های مربوط به مناطق و ایستگاه‌های بدون اطلاعات و یا با اطلاعات ناکافی را مورد تجزیه وتحلیل قرار داد. اعتبار اطلاعات منطقه‌ای در اغلب موارد بیش از اطلاعات نقطه‌ای می باشد (۱۱).

طبقه بندی منطقه‌ای متغیرها به هیدرولوژیست‌ها در جهت ساده سازی پیچیدگی هیدرواقلیمی (آب و هوایی) و در نتیجه کاهش حجم اطلاعات، مشاهدات و متغیرها کمک می‌کند. معمولا روش‌های مختلفی برای پهنه‌بندی متغیرهای هیدرو اقلیمی وجود دارد (۱۳). از جمله این روش ها می‌توان به تجزیه به مولفه های اصلی[6] و تجزیه خوشه ای[7] اشاره کرد. تجزیه به مولفه‌های اصلی (PCA) و روش‌های

۱،۲،۳و۴- به ترتیب دانش آموخته کارشناسی ارشد، استاد و استادیاران گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
(*- نویسنده مسئول: Email: baharehfallahi@yahoo.com)

5- Regionalization
6- Principal Component Analysis(PCA)
7- Cluster Analysis

مختلف آنالیز خوشه‌ای به طور گسترده در ایران و جهان برای پهنه-بندی‌های اقلیمی مورد استفاده قرار گرفته است. یو و یانگ (۱۹) روش PCA و آنالیز خوشه‌ای را برای داده‌های جریان روزانه ۳۴ ایستگاه جریان‌سنجی به کار بردند. نتایج نشان داد که روش تجزیه خوشه‌ای سلسله مراتبی وارد و غیر سلسله مراتبی k-mean به طور قابل قبولی می‌تواند برای تفکیک مناطق همگن بکار رود و در نهایت سه ناحیه همگن برای منطقه بدست آوردند. باریزویل و ربتز (۶) روش PCA و روش خوشه‌بندی وارد را برای دو مجموعه از داده‌های بارش در سوئیس با هدف درک بیشتر توزیع مکانی رژیم‌های بارشی به کار گرفتند. نتایج آنالیز خوشه‌ای وارد نشان داد که در حالت اول ۷ ناحیه و در حالت دوم ۱۳ ناحیه بارشی متمایز در منطقه وجود دارد. نیل و فیلیپس (۱۵) بارش‌های روزانه پنج تابستان متوالی از سال ۱۹۹۴ تا سال ۱۹۹۸ را برای ۱۳۶ ایستگاه در شرق آنجلیا[1] برای بررسی تغییرات اقلیمی بارش به کار گرفتند. ایشان از روش آنالیز مولفه‌های اصلی S-mode برای شناسایی پهنه‌های مکانی بارش استفاده کردند. نتایج نشان داد که در حالیکه ۷ مولفه اصلی دارای مقادیر ویژه[2] بیشتر از یک بودند و در مجموع ۸۶/۳ درصد از کل واریانس را توصیف می کنند، ۵ مولفه اصلی[3] اول الگوهای بارشی را به خوبی توصیف می‌کند و می‌توان به آن‌ها الگوهای چرخشی خاصی را نسبت داد. در نهایت با کاربرد آنالیز خوشه‌ای وارد بر روی ضرایب مولفه‌های اصلی اول تا هفتم، هفت ناحیه بارشی متمایز را معرفی کردند. جهانبخش و ذوالفقاری (۱) نسبت به پهنه‌بندی غرب ایران از طریق بارش‌های روزانه ۲۲ ایستگاه سینوپتیک در ۸ ماه مرطوب سال (اکتبر تا می) با به کارگیری روش تجزیه به عامل‌ها اقدام کردند. آن‌ها یک دوره آماری ۲۰ ساله (۱۹۹۰-۱۹۷۱) را برای این منظور انتخاب نمودند که نتیجه تحقیقات آن‌ها پهنه‌بندی غرب ایران به پنج ناحیه متمایز بارشی بود. رضیئی و عزیزی (۳) برای شناخت مناطق همگن بارشی غرب ایران با استفاده از ۹ متغیر وابسته به بارش ۱۴۰ ایستگاه هواشناسی در دوره آماری (۲۰۰۰-۱۹۶۵) را مورد استفاده قرار دادند. متغیرها به کمک روش تجزیه به مولفه‌های اصلی به چهار مولفه کاهش و با استفاده از روش وریماکس چرخش داده شدند. سپس با بهره‌گیری از روش سلسله مراتبی وارد ایستگاه‌های مورد استفاده بر مبنای نمرات استاندارد مولفه‌های بدست آمده، گروه‌بندی شدند به این ترتیب غرب ایران به چهار زیر منطقه همگن بارشی تقسیم شد. مسعودیان (۵) به کمک داده‌های بارش روزانه ۳۳۳ ایستگاه سینوپتیک و کلیماتولوژی و باران سنجی در طول دوره آماری ۲۰ ساله از سال ۱۹۶۱ تا ۲۰۰۴ و با استفاده از تحلیل خوشه ای وارد هشت ناحیه بارشی متمایز برای ایران بدست آوردند. توزیع مکانی بارش در ایران کم و بیش مطالعه شده است اما اغلب مطالعات قبلی روی مقدار بارش تمرکز داشته‌اند واکثر آن‌ها بر اساس مقیاس زمانی ماهانه می‌باشد. هدف این مقاله بکارگیری متغیرهایی مانند فواصل زمانی بارش (فواصل زمانی وقوع بارندگی‌ها) و روزهای بارانی است که تاکنون بعنوان متغیرهای اقلیمی جهت پهنه‌بندی بکارگرفته نشده-اند در صورتیکه به نظر مولفین فواصل زمانی بارش بیان کننده کیفیت توزیع زمانی بارش در طول سال یا فصول مختلف می‌باشد. که بنحوی خشک بودن یا مرطوب بودن منطقه را می‌رساند و همینطور روزهای بارانی یعنی حذف صفرهای سری بارش روزانه که این صفرها عدم وجود بارش را می‌رساند، موجب می‌گردد که تغییرات مکانی بارش کاملا مشخص و پهنه‌بندی گویای این تغییرات شود. از طرف دیگر رفتار توزیع مقادیر بارش، ایستگاه را از نظر پرباران بودن یا کم باران بودن نشان می‌دهد.

مواد و روش‌ها

منطقه مورد مطالعه در این تحقیق شمالغرب ایران شامل استان-های آذربایجان شرقی، غربی و اردبیل است که دارای مساحت ۱۰۰۵۰۳ کیلومتر مربع می باشد. گستردگی این منطقه به طور تقریبی از عرض جغرافیایی ۳۷ تا ۴۰ درجه شمالی و طول جغرافیایی ۴۴ تا ۴۹ درجه شرقی است. شکل۲منطقه مورد مطالعه و موقعیت ایستگاه-های مورد بررسی رانشان می‌دهد.

آمار بارش‌های روزانه از سازمان هواشناسی استان‌های آذربایجان شرقی ، غربی و اردبیل اخذ گردید. و ۶۰ ایستگاه که دارای ۳۵ سال دوره آماری (۲۰۰۴-۱۹۷۰) بودند انتخاب شد. جهت بازسازی داده-های گمشده این ۶۰ ایستگاه کدی در محیط MATLAB نوشته شد که برای پیدا کردن داده ایستگاه مفقود از الگوریتم GRIDFIT استفاده کرده است (۸). GRIDFIT یک ابزار مدل‌بندی سطح می-باشد که سطحی را برای داده‌های پراکنده یا منظم برازش می‌کند. اساس این الگوریتم پیدا کردن داده‌های ایستگاه مفقود با استفاده از درونیابی و برونیابی بین ایستگاه‌های همسایه ایستگاه مفقود است.

روش تجزیه به مولفه های اصلی

روش تجزیه به مولفه‌های اصلی از روش‌های آماری چند متغیره است که می‌توان از آن برای سهولت تحلیل متغیرهای اولیه مسئله، در مواردی که با حجم زیادی از اطلاعات روبرو هستیم، استفاده نمود. با اعمال این روش متغیرهای اولیه به تعدادی مولفه‌های جدید و مستقل از هم تبدیل می‌شوند. مولفه های جدید ایجاد شده ترکیب خطی از متغیرهای اولیه هستند. با استفاده از این تکنیک ترکیباتی از p متغیر اولیه، $X_P \ldots, X_2, X_1$ برای ایجاد p مولفه مستقل (معادل با تعداد

1-Anglia
2-Eigen value
3 - Principal Component (PC)

متغیرهای اولیه مورد استفاده)، $Z_1, Z_2, \ldots Z_P$ ایجاد می‌شود. هر مولفه اصلی می‌تواند با دنباله زیر مشخص شود:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ij}X_p \quad (۱)$$

در آن Z_i معرف مولفه مورد نظر، a_{ij} ضریب مربوط به متغیرهای اولیه و X_i نیز متغیر اولیه می‌باشد. ضریب مربوط به متغیرهای اولیه از حل رابطه ۲ بدست می آید

$$|R - \lambda I| = 0 \quad (۲)$$

که در آن I ماتریس واحد (۶۰×۶۰)، R ماتریس همبستگی (۶۰×۶۰) بین متغیرهای اولیه و λ نیز ماتریس مقادیر ویژه (۶۰×۶۰) می باشد. از این مقدار ویژه، بردار ویژه[1] بدست می‌آید. هر مقدار ویژه با اطلاعات مربوط به آن (بردارهای ویژه) ویژگی‌های یک مولفه را بدست می‌دهد (۱۲و۱۶). برای پهنه‌بندی با روش تجزیه به مولفه‌های اصلی از نرم افزار STATISTICA استفاده شد. برای بدست آوردن مقادیر ویژه و ضرایب مولفه‌ها (بردارهای ویژه) از ماتریس همبستگی بین ایستگاه‌ها به ابعاد (۶۰×۶۰) استفاده گردید، افزون بر این نمودار صخره‌ای[2] (نمودار مقادیر ویژه به عنوان تابعی از تعداد مولفه‌ها) رسم شد. آن دسته از مولفه‌های اصلی که مقادیر ویژه آن‌ها بیش از یک بود به عنوان مولفه‌های اصلی غالب انتخاب و ضرایب مولفه‌های اصلی غالب یا بارهای غالب نیز محاسبه شدند (۱۶). برای پهنه‌بندی نواحی بارشی از ضرایب مولفه‌ها استفاده گردید که به طور وسیعی در مطالعات اقلیمی استفاده شده است (۲، ۷ و ۱۰). جهت امکان تفکیک بهتر نواحی بارشی محورهای اصلی مورد چرخش متعامد از نوع وریماکس نرمالیزه واقع شدند با این کار امکان تسهیل در تفسیر-پذیری عامل‌ها نیز فراهم گردید. مقادیر درصد واریانس هر مولفه و ضرایب مولفه‌های اصلی پس از چرخش نیز محاسبه شد. در نهایت نواحی بارشی منطقه مورد مطالعه با توجه به این ضرایب مشخص گردید. در این تحقیق منطقه‌بندی جغرافیایی بر مبنای خوشه‌های حاصل شده، با استفاده از روش‌های آماری و سیستم اطلاعات جغرافیایی (GIS) انجام شده است. آزمون همگنی نواحی ایجاد شده توسط آماره S (رابطه۳) انجام و نقشه‌های حاصل از پهنه‌بندی با استفاده از نرم افزار ArcGis9.3 ترسیم شد.

$$S = \sum_{j}^{N} \frac{(cv_j - cv_0)^2}{U_j} \quad (۳)$$

که cv_0 مقدار وزنی میانگین منطقه ای cv، cv_j تغییرات نمونه در j امین محل و U_j تغییر نمونه گیری cv_j و N تعداد ایستگاه‌ها را نشان می‌دهد (۱۸).

روش خوشه بندی وارد

روش وارد دقیق‌ترین روش سلسله مراتبی است (۹). در این روش هر عضو در گروهی جای می‌گیرد که مجموع مربعات انحرافات درون گروهی به حداقل برسد. برای پهنه‌بندی با روش وارد ماتریس داده-های ورودی همان ماتریس‌های ورودی روش تجزیه به مولفه‌های اصلی بود. در مطالعه حاضر پس از انجام تجزیه خوشه‌ای نتایج به صورت نمودار درختی ارائه شد. پهنه‌بندی با استفاده از نمودار و تفکیک آن به خوشه‌های فرعی بدست آمد. مراحل تجزیه خوشه‌ای وارد در نرم افزار MINITAB و ترسیم نقشه‌های آن در نرم افزار ArcGis انجام شد.

روش K-mean

روش K-mean به عنوان روش کامل و رایج از نظر خصوصیات جداسازی و غیر سلسله مراتبی در اغلب کارهای عملی بکار گرفته شده است. معیار K-mean نیز حداقل کردن مربعات خطای بین خوشه‌ای می‌باشد و معمولا در این روش از فاصله اقلیدسی استفاده می‌شود (۱۷).

فرض می‌کنیم D مجموعه ای از داده‌ها با n عضو باشد. اگر $C_1, C_K, \ldots C_2$ خوشه‌های مجزا باشند. تابع خطا به شکل رابطه ۴ تعریف می شود:

$$E = \sum_{i=1}^{k} \sum_{x \in c_i} d(x_i, \mu(c_i)) \quad (۴)$$

که در آن:

$\mu(c_i)$ مراکز جرم خوشه ای C_i و $d(xi, \mu(c_i))$ فاصله بین x_i و $\mu(c_i)$ (معمولا فاصله اقلیدسی) است و k تعداد خوشه‌ها می باشد.

الگوریتم مرسوم K-mean به شکل زیر می‌باشد:

۱-تعداد k خوشه اولیه به صورت تصادفی انتخاب شده و مراکز خوشه‌ها بطور جداگانه تعیین می‌گردد.

۲-هر عضو به خوشه‌ای که نزدیک ترین فاصله را با مرکز آن دارد انتقال داده می‌شود.

۳-زمانی که همه اعضا به خوشه‌های مختلف اختصاص داده شدند، موقعیت مراکز دوباره تعیین می‌شود.

۴-گام ۲و۳ تا زمانی که اعضای خوشه ها تغییر نکند، یا خطا کمینه شود تکرار می شود (۹). ماتریس‌های ورودی برای الگوریتم K-mean نیز همان ماتریس ورودی روش‌های PCA و Wardبود. روش خوشه‌بندی K-mean در نرم افزارهای مختلف آماری مانند STATISTICA, SPSSو MINITAB وجود دارد ولی در مطالعه حاضر به دلیل ابعاد بزرگ ماتریس داده‌های بارش روزانه (۱۲۷۸۴×۶۰) انجام خوشه‌بندی با این نرم افزارها امکان پذیر نشد بنابراین برای انجام خوشه‌بندی با این روش برنامه ای در محیط

1- Eigen Vector
2- Scree Plot

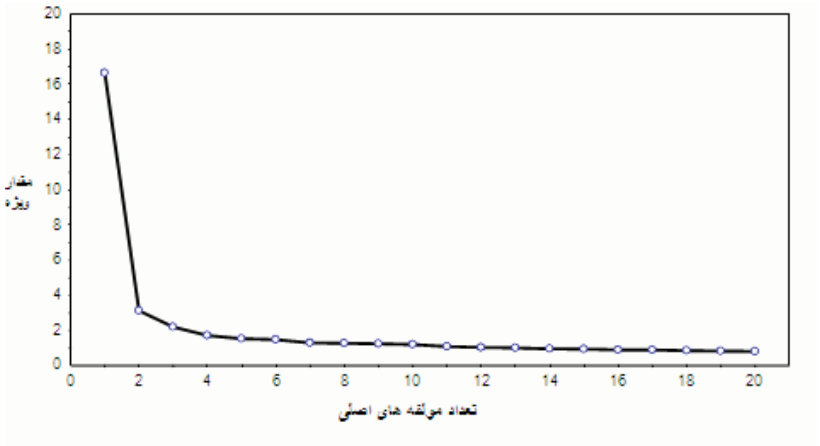
فرترن نوشته و اجرا شد.

نتایج و بحث

از آنجایی که بررسی بارش در مقیاس روزانه می‌تواند تغییرات مکانی اقلیم بارش را نسبت به مقیاس‌های زمانی دیگر مانند ماهانه و فصلی بهتر نمایش دهد، بنابراین در این تحقیق مقیاس روزانه برای مطالعه حاضر انتخاب شد. از طرفی با توجه به این که مقدار بارش و فاصله زمانی (فاصله زمانی وقوع بارش ها) دو متغیر موثر و مهم در پهنه‌بندی می‌باشند از این رو در این تحقیق شاخص جدیدی معرفی گردید که این شاخص در برگیرنده هر دو متغیر مذکور به طور همزمان باشد و از تقسیم مقدار بارش بر فاصله زمانی آن در هر روز بدست آمده است. به دلیل اینکه بارش روزانه دارای صفرهای زیادی می‌باشد و پیش بینی می‌گردد این صفرها سبب ایجاد همگنی کاذب شوند، لذا با حذف صفرها یک نوع پهنه بندی، تنها بر اساس روزهای بارانی انجام گرفت و به نظر می‌رسد پهنه بندی بر اساس روزهای بارانی واقعی‌تر باشد. لذا هدف از پهنه‌بندی بر مبنای روزهای بارانی، پهنه بندی بر مبنای رفتار منطقه از نقطه نظر تغییرات در مقدار بارش است که در این مورد مقدار همگنی و غیر همگنی مناطق بسیار مهم می‌باشد. نتایج تجزیه به مولفه‌های اصلی سری بارش‌های روزانه نشان داد که سیزده مولفه اول دارای مقدار ویژه بیش از یک بوده و در مجموع ۵۷/۹۴ درصد از واریانس کل داده‌ها را توجیه می‌کند. شکل ۱ نمودار صخره‌ای مقادیر ویژه تابعی از تعداد مولفه‌های اصلی و جدول ۱ مقادیر ویژه و درصد واریانس قبل و بعد از چرخش سیزده مولفه اصلی را برای بارش‌های روزانه با ماتریس همبستگی بین داده-ها نشان می‌دهد.

جدول۱-مقادیر ویژه و درصد واریانس تجمعی قبل و بعد از چرخش حاصل از ماتریس همبستگی سری بارش روزانه

شماره مولفه	مقدار ویژه قبل از چرخش	درصد واریانس تجمعی قبل از چرخش	مقدار ویژه بعد از چرخش	درصد واریانس تجمعی بعد از چرخش
۱	۱۶/۶۴	۲۷/۷۴	۶/۹۴	۱۱/۵۷
۲	۳/۱۲	۳۲/۹۴	۲/۴۹	۱۵/۷۳
۳	۲/۱۹	۳۶/۵۹	۴/۶۷	۲۳/۵۱
۴	۱/۷۱	۳۹/۴۴	۱/۶۹	۲۶/۳۳
۵	۱/۵۳	۴۱/۹۹	۲/۷۳	۳۰/۸۹
۶	۱/۴۷	۴۴/۴۴	۳/۱۲	۳۶/۰۹
۷	۱/۲۹	۴۶/۵۹	۲/۵۶	۴۰/۳۵
۸	۱/۲۷	۴۸/۷۰	۱/۴۷	۴۲/۸۰
۹	۱/۲۴	۵۰/۷۶	۱/۶۹	۴۵/۶۱
۱۰	۱/۲۰	۵۲/۷۶	۱/۹۱	۴۸/۷۹
۱۱	۱/۰۸	۵۴/۵۶	۲/۱۰	۵۲/۳۰
۱۲	۱/۰۳	۵۶/۲۷	۱/۶۴	۵۵/۰۲
۱۳	۱/۰۰	۵۷/۹۴	۱/۷۵	۵۷/۹۴

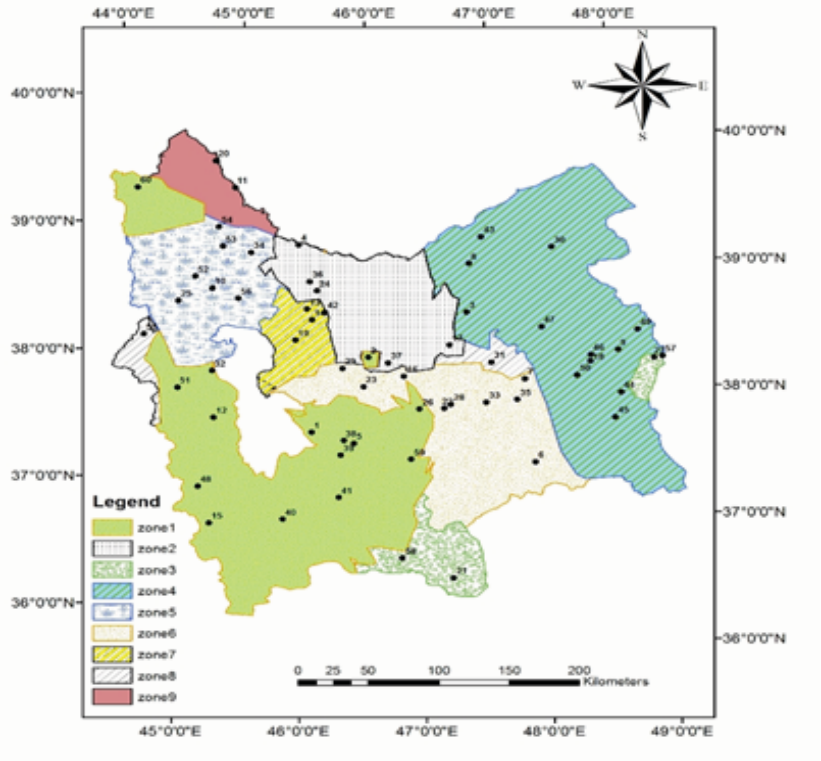


شکل۱-نمودار صخره‌ای حاصل ازPCA سری بارش روزانه

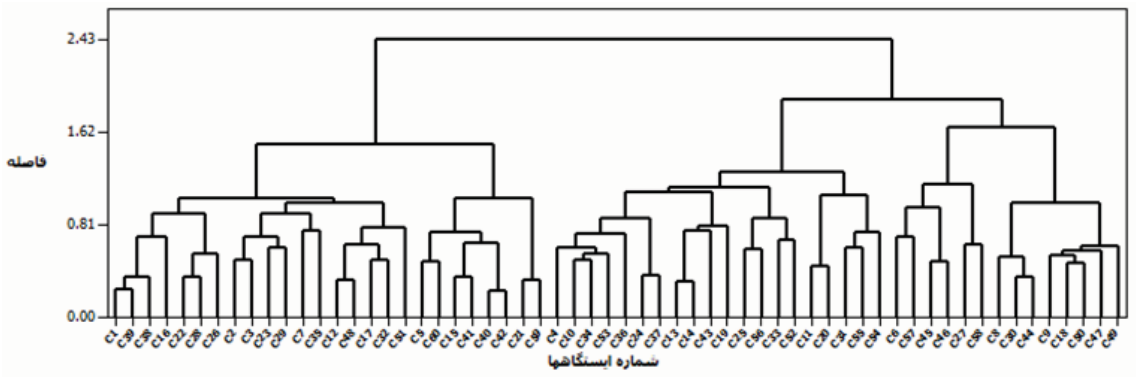
شکل ۲- منطقه و موقعیت ایستگاه‌های مورد بررسی

شکل ۳ نواحی همگن ایجاد شده با روش PCA و شکل‌های ۴، ۵ و ۶ به ترتیب دندروگرام نواحی همگن حاصل شده از روش وارد ونواحی همگن حاصل شده با استفاده از روش k-mean سری بارش‌های روزانه را نشان می‌دهد. در شکل ۳ مشاهده می‌شود که ۹ ناحیه حاصل شده در صورتی که ما ۱۳ مولفه را انتخاب کرده بودیم دلیل آن را می توان این طور بیان کرد که پس از رسم ضرایب مولفه‌ها در نرم افزار ArcGis و روی هم انداختن نقشه ها، از بین ضرایبی که هم پوشانی داشته‌اند ضریب بزرگتر انتخاب شده، بنابراین برخی نواحی حذف شده اند.

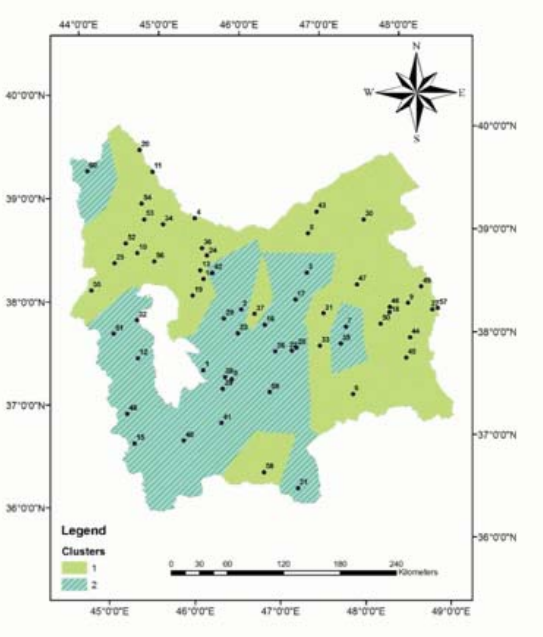
تمامی مراحل ذکر شده برای بارش روزانه، برای سری‌های شاخص فاصله–مقدار، فواصل زمانی بارش و روزهای بارانی انجام شده ونواحی همگن ایجادشده با سه روش PCA، Wardو k-mean در شکل‌های ۷ تا ۱۵نشان داده شده است.



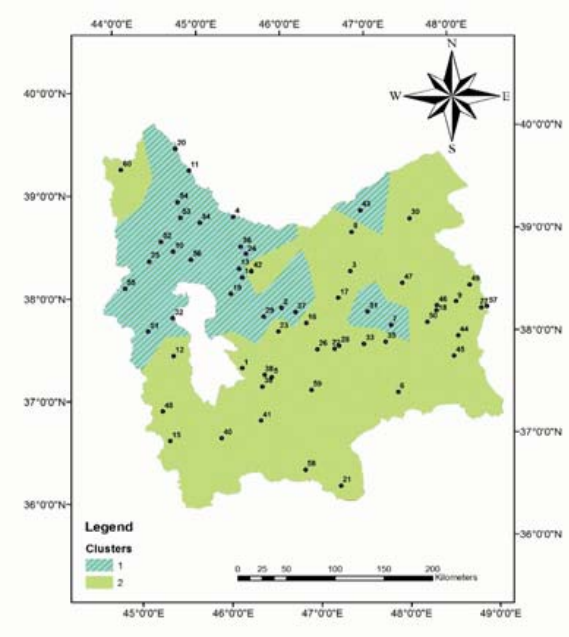
شکل ۳- نواحی همگن ایجاد شده با روش PCA



شکل ٤-دندروگرام حاصل از روش وارد سری بارش‌های روزانه شمالغرب ایران

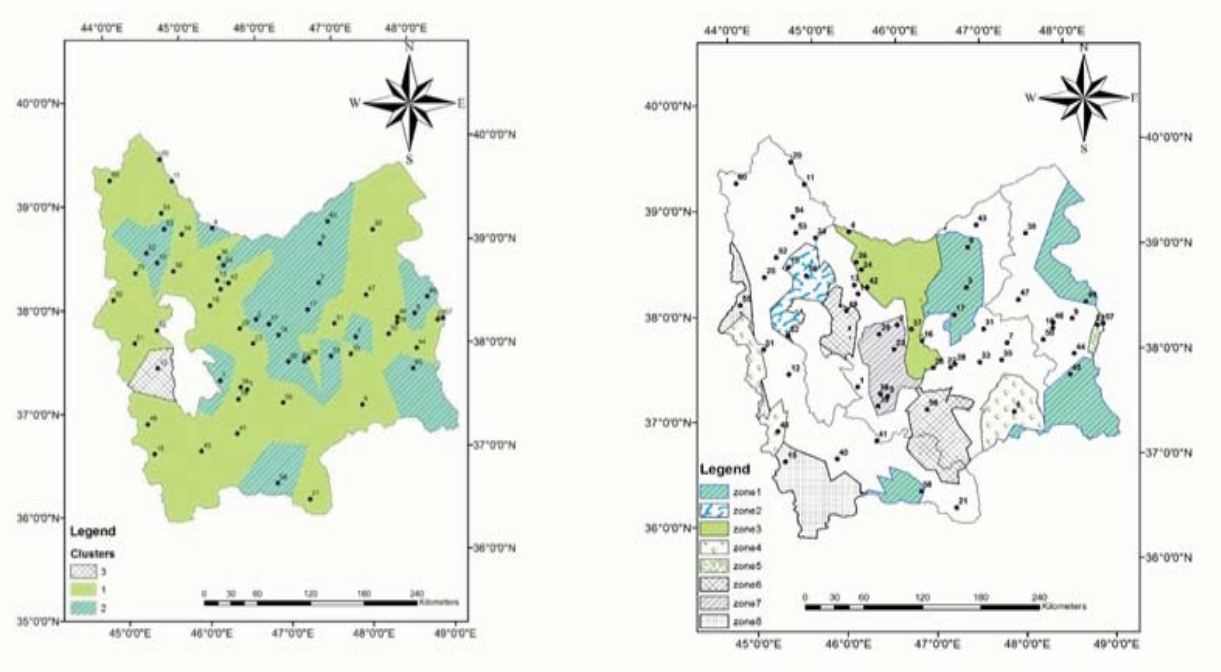


شکل۵-نواحی همگن بر مبنای بارش روزانه باروش وارد

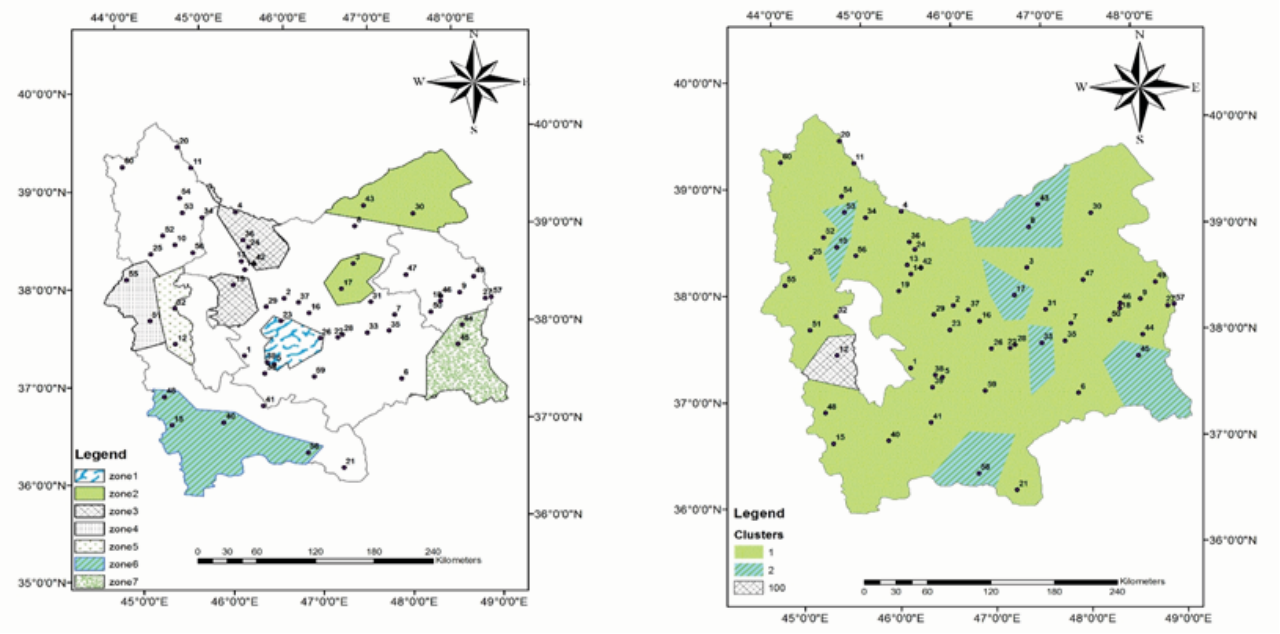


شکل٦-نواحی همگن بر مبنای بارش‌های روزانه باروشk-mean

با استفاده از روش تجزیه به مولفه‌های اصلی بارش های روزانه شمال غرب ایران، منطقه به ۹ ناحیه همگن تقسیم شد (شکل۳). این پهنه‌بندی تا حدودی با پهنه‌بندی دین‌پژوه و همکاران (۲) شباهت دارد. ایشان ناحیه شمالغرب ایران را بر مبنای بارش‌های روزانه و با روش تجزیه به مولفه‌های اصلی به ۷ ناحیه بارشی مستقل کاملاً همگن و یک ناحیه متمایل به غیرهمگن بودن تفکیک کردند همچنین برای بررسی همگنی مناطق ایجاد شده از آماره H استفاده کردند. جهانبخش و ذوالفقاری (۱) نیز غرب ایران را از طریق بارش-های روزانه با به کارگیری روش تجزیه به عامل‌ها به ۵ ناحیه متمایز تفکیک کردند. کاسب‌پور (۴) روش تجزیه به مولفه‌های اصلی را برای سری بارش ماهانه در استان آذربایجان شرقی به کار برد و ناحیه مذکور را به ۵ ناحیه همگن تقسیم نمود ایشان نیز برای آزمون همگنی از آماره S استفاده کردند. روش خوشه‌بندی وارد با همین مجموعه داده‌ها، منطقه مذکور را به ۲ ناحیه مجزا تفکیک کرده است (شکل۵). با این تفاوت که در روش وارد ناحیه ۱ (شمال و شرق منطقه مورد مطالعه) تقریبا از ترکیب ۵ پهنه (موسوم به پهنه های ۳، ۴، ۷، ۸ و۹) روش تجزیه به مولفه‌های اصلی تشکیل یافته است. کاسب‌پور (۴) با کاربرد روش خوشه‌بندی وارد برای سری بارش ماهانه، استان آذربایجان شرقی را به ۳ ناحیه متمایز تفکیک کرده است. روش خوشه‌بندی k-mean نیز منطقه مذکور را به دو ناحیه متمایز تفکیک کرد (شکل۶). به طوریکه ناحیه۱ (شمال و شمال غرب منطقه) تقریبا ترکیبی از سه پهنه (۴، ۶ و ۹) روش تجزیه به مولفه‌های اصلی است.



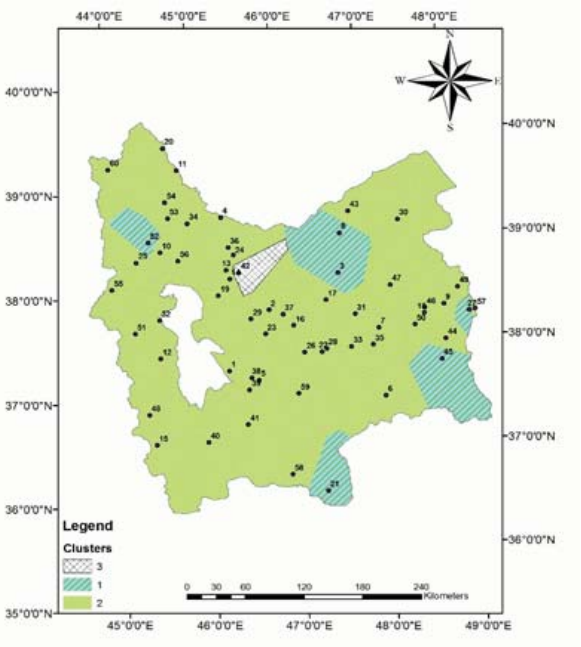
شکل۷-نواحی همگن برمبنای سری شاخص فاصله-مقدار با روش PCA شکل۸-نواحی همگن بر مبنای سری شاخص فاصله-مقدار با روش وارد



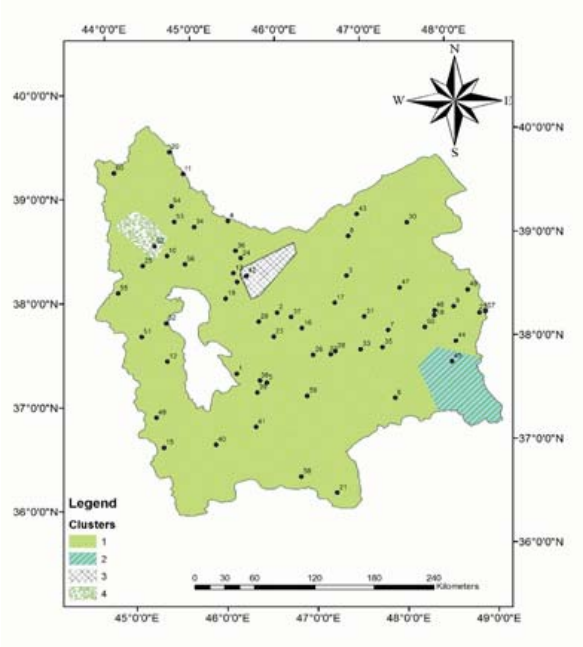
شکل ۹-نواحی همگن شاخص فاصله-مقدار با روشk-mean شکل ۱۰-نواحی همگن فواصل زمانی بارش باروش PCA

کاربرد روشPCA برای سری شاخص فاصله-مقدار منطقه را به ۸ ناحیه همگن و کوچک تقسیم کرد (شکل۷). هچنین پهنه‌بندی با دو روش وارد و k-mean منطقه را به دو ناحیه همگن و یک ایستگاه ناسازگار (ارومیه) تفکیک نمود (شکل ۸ و ۹). روشk-mean بیشترین مساحت منطقه را تنها با یک پهنه (ناحیه۱) توجیه می‌کند (شکل۹). در روش PCA هیچ یک از نواحی موجود شمالغرب منطقه مورد مطالعه را پوشش ندادند (شکل۷). کاربرد روش PCA برای سری فواصل زمانی بارش منطقه را به ۷ ناحیه همگن و کوچک تقسیم کرد (شکل۱۰).

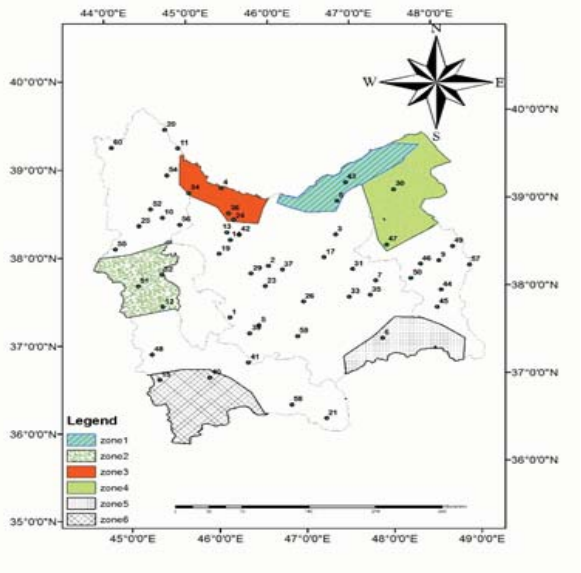
پهنه‌بندی با روش وارد منطقه را به دو ناحیه همگن و یک ایستگاه ناسازگار (بناب) تفکیک نمود (شکل۱۱).



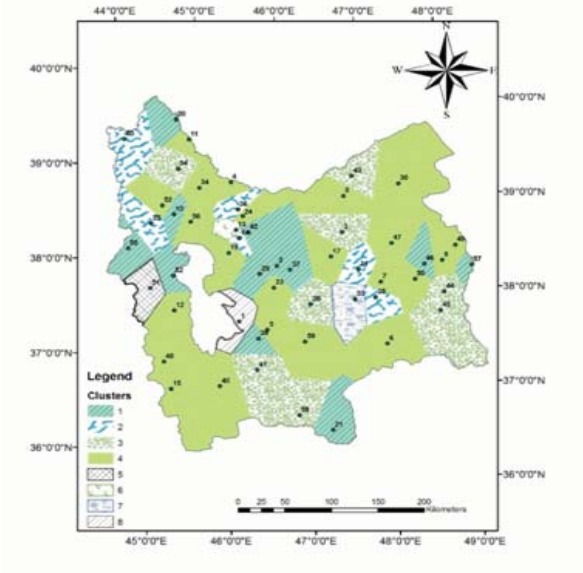
شکل۱۱-نواحی همگن فواصل زمانی بارش با روش وارد



شکل ۱۲-نواحی همگن فواصل زمانی بارش باروشkmean

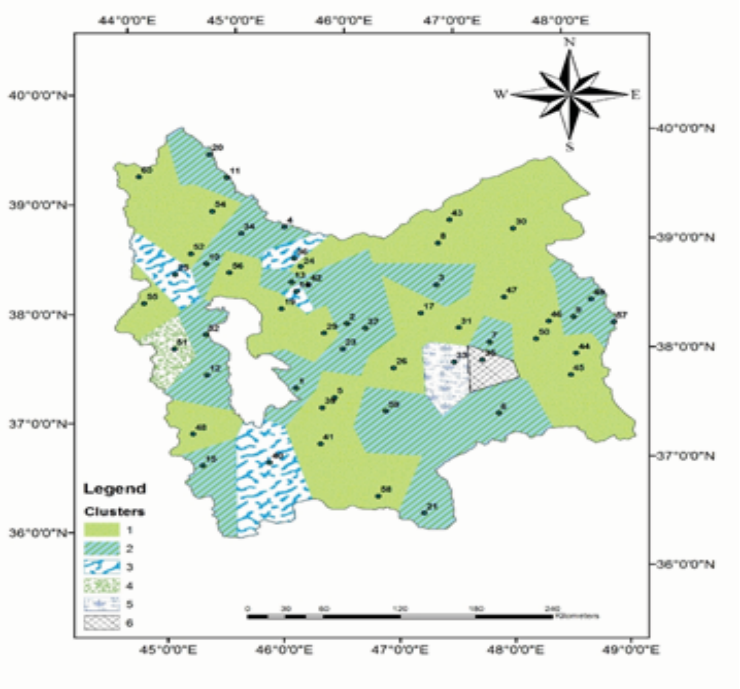


شکل۱۳-نواحی همگن روزهای بارانی باروش PCA



شکل۱٤-نواحی همگن روزهای بارانی با روش وارد

این بار نیز روش k-mean بیشترین مساحت منطقه را تنها با یک پهنه (ناحیه۱) توجیه می‌کند این روش منطقه را به یک خوشه بزرگ و یک خوشه تک عضوی (کیوی علیا) و دو ایستگاه ناسازگار (بناب و قشلاق فیرورق) تفکیک نمود (شکل۱۲). در روش PCA هیچ یک از نواحی قسمت‌های شمالغرب، شرق، جنوب شرق و مراکز منطقه مورد مطالعه را پوشش ندادند. کاربرد روش PCA برای سری روزهای بارانی (سری بارش روزانه با حذف صفرهای آن) منطقه را به ۶ ناحیه همگن و کوچک تقسیم کرد به طوری که شمالغرب، شرق، جنوب و جنوب شرق و مراکز منطقه مورد مطالعه بدون پوشش باقی ماندند (شکل۱۳). پهنه بندی با روش وارد منطقه را به ۴ ناحیه همگن و ۴ ایستگاه ناسازگار تفکیک نمود (شکل۱۴). روش k-mean نیز منطقه را به سه ناحیه همگن و سه ایستگاه ناسازگار تفکیک نمود (شکل۱۵). در هر دو روش وارد و k-mean ایستگاه‌های راوند ارومیه و قوشچی سراب به صورت ایستگاه‌های مجزا مشخص شدند.



شکل ۱۵- نواحی همگن روزهای بارانی با روش k-mean

نتیجه گیری

نتایج پهنه‌بندی با روش PCA نشان داد که پهنه‌های حاصل از سری فواصل زمانی بارش، شاخص فاصله-مقدار و روزهای بارانی بسیار کوچک بوده و مناطق زیادی را پوشش ندادند. می‌توان نتیجه گرفت علت آن تصادفی بودن و عدم همبستگی این متغیرها و در واقع عدم تشابه منطقه مورد مطالعه در این متغیرها می‌باشد، بنظر می‌رسد دلیل آن خشک بودن منطقه مورد مطالعه است. زیرا اگر منطقه دارای بارش‌های موسمی منظم بود احتمالا پهنه‌های حاصله بر مبنای این متیرها وسیع‌تر می‌شد و مناطق بیشتری تحت پوشش قرار می‌گرفت. روش PCA به لحاظ ماهیت خود فضای داده‌ها را روی محورهای اصلی تصویر می‌کند و ساختار واقعی فضا را نشان خواهد داد اما در روش‌های سلسله مراتبی که در هر مرحله خوشه یا اعضا را به هم متصل می‌کنند، از نظر ماهیت فضایی خوشه‌ها بیانگر ساختار واقعی نمی‌باشند. بنابراین، انتظار می‌رود که خوشه‌های حاصل از PCA واقعی‌تر از خوشه‌های حاصل از دیگر روش‌ها باشد. اما انتخاب خوشه‌بندی بر مبنای تست‌های آماری ممکن است گویای فضای واقعی نباشد هر چند که همگن باشد. زیرا فواصل کوتاه خوشه‌ها غیر همگنی معنی‌دار ایجاد نخواهد نمود. از نقطه نظر گسترش اطلاعات خوشه‌بندی بر مبنای همگنی از خوشه‌بندی واقعی بهتر می‌باشد زیرا خوشه‌هایی با اعضای بیشتر را در بر می‌گیرد که نتیجه آن در بر گرفتن اطلاعات زیاد و گسترش دقیق و قابل قبول آن خواهد بود.

منابع

۱-جهانبخش س. و ذوالفقاری ح. ۱۳۸۱. بررسی الگوهای سینوپتیک بارش‌های روزانه در غرب ایران. فصلنامه تحقیقات جغرافیایی . ۶۴: ۲۵۸-۲۳۴.

۲-دین‌پژوه ی. ۱۳۸۲. تحلیل خشکسالی‌های هواشناختی با استفاده از آنالیز الگوها. پایان‌نامه دکترا. دانشکده کشاورزی. دانشگاه تبریز.

۳-رضیئی ط. و عزیزی ق. ۱۳۸۸ . شناخت مناطق همگن بارشی در غرب ایران . مجله جغرافیا و برنامه ریزی محیطی. ۲۰:۸۶-۶۵.

۴-کاسب پور ف.۱۳۸۷. پهنه‌بندی خشکسالی هواشناسی استان آذربایجان شرقی باروش تجزیه به مولفه‌های اصلی. پایان‌نامه کارشناسی ارشد. دانشکده کشاورزی. دانشگاه تبریز.

۵-مسعودیان ا. ۱۳۸۸. نواحی بارشی ایران. مجله جغرافیا و توسعه.۳: ۹۱-۷۹.

6-Baeriswyl P.A., and Rebetez M. 1997.Regionalization of precipitation in Switzerland by means of principal

component analysis.Theor. Appl. Climatol, 58: 31-41.
7-Comrie A.C., and Glenn E.C. 1998. Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability.J.of Climate Research, 10: 201–215.
8-D´Erico J. 2006. Understanding GRIDFIT. 2006. Available for download at http://www.mathworks.com/matlabcentral/fileexchange/8998 (lastaccessed 20090414).
9-Gan G., Chaoqun M., and Jianhong W. 2007. Data clustering: Theory, Algorithms, and Applications, ASA-SIMA Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria,VA, ISBN 978-0-898716-23-8.
10-Kolivras K.N., and Comrie A.C. 2007. Regionalization and variability of precipitation in Hawaii.J.of Physical Geography, 28(1): 76-96.
11-Lee C. 2005. Application of rainfall frequency analysis on studying rainfall distribution characteristics of Chia-Nan plain area in Southern Taiwan.J.of Climatology,9:145-167.
12-Legates D.R., and McCabe G.J. 1999. Evaluating the use of "goodness-of-fit" measures inhydrologic and hydroclimatic model validation. Water Resour. Res, 35: 233-241.
13-Modarres R., and Sarhadi A. 2010.Statistically-based regionalization of rainfall climates of Iran.J.of Global and Planetary Change, 75: 67-75.
14-National Drought Mitigation Center of United States (NDMC). 1998. Report on drought related effects of El Ninofor February.
15-Neal R.A., and Phillips I.D. 2009. Summer daily precipitation variability over the East Anglian region of Great Britain. Int. J. of Climatol, 29:1661-1679.
16-Rencher A.C. 2002. Methods of Multivariate Analysis.John Wiley & Sons, INC.
17-Vejmelka M., Musilek P., Palus M., and Pelikan E. 2009. K-means clustering for problems with periodic attributes.International Journal of pattern Recognition and Artificial Intelligence ,23:721-743.
18-Wiltshire S.E. 1986. Identification of homogeneous regions for flood frequency analysis. J. of Hydrology, 84: 287-302.
19-Yu P.S.H., and Yung T.C.H. 1996. Synthetic regional flow duration curve for Southern Taiwan.J.of Hydrological Processes, 10: 373-391

Journal of Water and Soil
Vol. 26, No.4, Sept.-Oct. 2012, p. 979-989

نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۲۶، شماره ۴، مهر – آبان ۱۳۹۱، ص. ۹۸۹-۹۷۹

Regionalization of Northwest Iran Based on Daily Rainfalls and Rain's Time Intervals Using PCA, Ward and K-mean Methods

B. Fallahi[1*] - A. Fakheri Fard[2] - Y. Dinpajooh[3] - S. Darbandi[4]

Received: 31-10-2011
Accepted: 10-6-2012

Abstract

Having a correct view of the effective factors on climatic changes by explanation of a considerable part of the total variance in data with limited number of principal components the analytical methods of decreasing data dimensions, such as PCA are important tools in water resources planning. In this study PCA method as a projection tool for projecting the information space on the limited and specific axes, ward's method as a hierarchical clustering and k-mean as partitioning clustering method has been applied in this research. Using this methods and application of daily precipitation data of 60 meteorological stations during a 35 years period (1970-2004), 4 types of delineated regions were come out on the basis of daily precipitations, distance-quantity index, time intervals and rainy days series. S statistic test algorithm was used for homogeneity test of the regions. Results showed the nature of the PCA method is such that projects the data space on the main axes and shows the real space. But in the hierarchical methods, clusters do not describe the real structure. Therefore we do expect that the resulting clusters of PCA would be more realistic than that of methods. But hierarchical methods have the advantage of containing the wider clustering information on the basis of homogeneity than the others.

Keywords: Regionalization, Principal Components Analysis, Cluster Analysis, Ward, K-mean

1,2,3,4- Former MSc Student, Professor and Assistant Professors, Department of Water Resources Engineering, Faculty of Agriculture, University of Tabriz, Respectively
(* - Corresponding Author Email: baharehfallahi@yahoo.com)