استفاده از نگرش تحلیل مؤلفه‌های اصلی برای وزن‌دهی ویژگی‌های آماری، اقلیمی و جغرافیایی حداکثر بارندگی ۲۴ ساعته و تحلیل مکانی خوشه‌بندی (مطالعه موردی: حوضه دریاچه ارومیه)

زهره دهقان[1*]، سید سعید اسلامیان[1] و رضا مدرس[2]

(تاریخ دریافت: ۱۳۹۵/۱۲/۱۴ ؛ تاریخ پذیرش: ۱۳۹۶/۷/۳۰)

چکیده

ناحیه‌بندی یکی از ابزارهای مفید برای انجام تحلیل‌های مؤثر در مناطق فاقد داده، یا دارای داده‌های ناقص است. یکی از روش‌های ناحیه‌بندی که در مطالعات هیدرولوژیک بسیار مورد استفاده قرار می‌گیرد، خوشه‌بندی است. یکی از عوامل تأثیرگذار بر خوشه‌بندی، درجه اهمیت و میزان مشارکتی است که هر یک از این ویژگی‌ها می‌توانند داشته باشند. در این پژوهش، سعی شد از یک مجموعه گسترده‌ای از ویژگی‌ها برای مقایسه عملکرد آنها در ناحیه‌بندی استفاده شود. سپس، با توجه به درجه اهمیت و نقشی که هر ویژگی در ناحیه‌بندی دارد، وزن مناسب برای هر یک از گروه ویژگی‌ها با استفاده از خروجی روش تحلیل مؤلفه‌های اصلی تعیین و تأثیر وزن‌دهی در تشکیل ناحیه‌های همگن با استفاده از روش خوشه‌بندی وارد بررسی شد. بدین‌منظور، در این پژوهش از داده‌های حداکثر بارش ۲۴ ساعته ۶۳ ایستگاه باران‌سنجی واقع در حوضه دریاچه ارومیه که دارای طول دوره آماری ۳۰ سال (۱۳۸۷-۱۳۵۸) هستند، استفاده شد. علاوه بر این، هفت دسته ویژگی به‌منظور ناحیه‌بندی بارش تعریف شدند. نتایج نشان داد که با درنظر گرفتن ویژگی‌های مختلف و ترکیب آنها با یکدیگر، ناحیه‌بندی متفاوتی در تعداد نواحی تشکیل شده است و نیز تعداد ایستگاه‌های متفاوتی در هر ناحیه وجود خواهد داشت. از میان هفت گروه ویژگی مشخص شد که دسته ویژگی‌های جغرافیایی و اقلیمی- جغرافیایی خوشه‌بندی مناسب‌تری را در سطح حوضه نشان داده‌اند. همچنین، وزن دادن به ویژگی‌ها در بیشتر موارد می‌تواند تأثیر مثبتی در بهبود اندازه همگنی و تشکیل نواحی مستقل از لحاظ پراکندگی ایستگاه‌ها و چگونگی قرارگیری آنها در سطح حوضه داشته باشد.

واژه‌های کلیدی: تحلیل مؤلفه‌های اصلی، حوضه دریاچه ارومیه، روش خوشه‌بندی وارد، وزن‌دهی ویژگی‌ها

۱. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه صنعتی اصفهان
۲. گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه صنعتی اصفهان
*: مسئول مکاتبات: پست الکترونیکی: z.dehghan@ag.iut.ac.ir

www.SID.ir

Archive of SID

مقدمه

یکی از مشکلات هیدرولوژیست‌ها در انجام تحلیل‌های آماری، نبود داده‌هایی با طول دوره آماری مناسب یا عدم وجود داده در مناطق مورد مطالعه است. سوابق اطلاعات هیدرولوژیک (بارش و سیلاب) عمدتاً کوتاه بوده و یادر دسترس نیست. بنابراین، تحلیل نقطه‌ای (سیلاب، بارش و..) در بیشتر موارد به دلایل ذکر شده از جمله کوتاه بودن دوره زمانی آمار و خطای برداشت، اطلاعات موجود قادر به پاسخگویی به پیش‌بینی‌های دراز مدت با دقت لازم نیستند. از طرف دیگر، رویدادهای نادر در سری زمانی ثبت شده نقطه‌ای به‌دلیل محدودیت دوره زمانی، نمونه آماری را از جامعه مجزا نشان می‌دهد و انجام تحلیل در منطقه مورد نظر را با مشکل مواجه می‌سازد. در این شرایط، تفکیک منطقه مطالعاتی به مناطق همگن باعث دقت بیشتر و خطای کمتر در مدل‌ها برای مناطق فاقد آمار می‌شود.

بسته به هدفی که از ناحیه‌ای کردن منطقه وجود دارد و نیز کیفیت و کمیت داده‌های قابل دسترس روش‌های مختلفی وجود دارد (۲۶). یکی از انواع روش‌های ناحیه‌بندی روش خوشه‌بندی براساس مشخصات ایستگاه مورد نظر است که توسط هاسکینگ و والیس در سال ۱۹۹۷ ارائه شد (۱۶). الگوریتم‌های زیادی برای تحلیل خوشه‌بندی پیشنهاد شده است به‌طوری‌که سه روش اصلی خوشه‌بندی که به‌طور گسترده در مطالعات مختلف استفاده شده‌اند عبارتنداند: خوشه‌بندی سلسله مراتبی، خوشه‌بندی غیرسلسله مراتبی و خوشه‌بندی فازی. علیجانی (۷) به‌منظور تعیین نواحی بارش در ایران از برخی خصوصیات مانند بارش سالانه، تمرکز فصلی، حداکثر بارش روزانه استفاده و پنج ناحیه بارشی را در سرتاسر ایران شناسایی کرد. جهانبخش و ذوالفقاری (۲) نیز با استفاده از روش تحلیل مؤلفه‌های اصلی و به‌کارگیری داده‌های بارش روزانه ۲۲ ایستگاه سینوپتیک در غرب ایران، برای هشت ماه مرطوب (اکتبر تا می)، پنج ناحیه بارشی را در غرب ایران که مشابه با نتیجه علیجانی (۷) است، مشخص کردند. مسعودیان (۱۰) به‌منظور تعیین مناطق بارشی در ایران به تهیه نقشه رقومی ایران از داده‌های بارش ماهانه از سال ۱۹۵۱ تا ۱۹۹۹ پرداخت. وی از متغیرهای میانگین، کمینه و بیشینه بارش و تغییرپذیری چولگی در روش خوشه‌بندی استفاده کرد و سه رژیم بارش پاییزی، زمستانی و زمستانی- بهاری را شناسایی کرد. ایشان رژیم بارش در ایران را تابعی از عرض جغرافیایی معرفی کرد و نشان داد که مقدار بارش با پیکربندی ناهمواری ارتباط دارد. مدرس (۹) به‌منظور تحلیل فراوانی بارش در ایران از روش تحلیل خوشه‌بندی وارد با به‌کارگیری تعدادی ویژگی‌های فیزیوگرافی، جغرافیایی و آماری استفاده کرد. نتایج کار وی، هشت ناحیه همگن بارشی در ایران را شناسایی کرد. بعد از آن، مدرس و سرحدی (۲۵) در مطالعه دیگری به‌منظور ناحیه‌بندی بارش فصلی در کل ایران از داده‌های بارش سالانه ۱۳۷ ایستگاه با دوره آماری ۱۹۵۲-۲۰۰۳ از روش خوشه‌بندی وارد استفاده کردند و سراسر ایران را به هشت ناحیه همگن تقسیم‌بندی و نتایج تحقیق مدرس (۹) را تأیید کردند.

در تحقیقی در همین راستا، خام چین مقدم و همکاران (۳) به‌منظور پهنه‌بندی حداکثر بارش‌های روزانه از روش تحلیل خوشه‌بندی سلسله مراتبی استفاده کردند و ایران را به هفت منطقه همگن تقسیم کردند. آنها از بین ۲۱ ویژگی مورد بررسی، تعداد شش ویژگی که شامل ارتفاع ایستگاه، میانگین و انحراف معیار بارش‌های سالانه و حداکثر ۲۴ ساعته و نسبت میانگین بارش زمستانه به بهاره هستند را به‌عنوان ویژگی‌های منتخب برای انجام ناحیه‌بندی انتخاب کردند. شیروانی و نظام السادات (۶) با هدف پهنه‌بندی بارش در ایران و به‌کارگیری داده‌های مشاهداتی ۴۲ ایستگاه اندازه‌گیری بارش و به‌منظور کاهش متغیرهای ورودی، از روش تحلیل مؤلفه‌های اصلی استفاده کردند. آنها پس از نرمال‌سازی داده‌ها و کاهش داده‌های مشاهداتی با استفاده از روش خوشه‌بندی وارد، کشور ایران را به شش ناحیه همگن بارشی تقسیم کردند. پور اصغر و همکاران (۱) با استفاده از روش تحلیل مؤلفه‌های اصلی و به‌کارگیری روش خوشه‌بندی، بارش نواحی جنوبی ایران را با استفاده از شش پارامتر میانگین بارش، ضریب تغییرات، ضریب

www.SID.ir

Archive of SID

کشیدگی، ضریب چولگی، حداقل و حداکثر بارش به پنج ناحیه همگن تقسیم بندی کرد. پس از آن، رضیئی (۴) در تحقیق مشابه به‌منظور شناسایی مناطق همگن بارشی در ایران از داده‌های بارش ماهانه ۱۵۵ ایستگاه همدیدی پراکنده در سطح کشور برای دوره آماری ۱۹۹۰ تا ۲۰۱۴ استفاده کرد. وی با استفاده از تحلیل مؤلفه‌های اصلی پنج منطقه همگن بارشی خزری، شمال‌غربی، غربی، مرکزی- شرقی و مرکزی- شمال‌شرقی را برای کل ایران شناسایی کرد. دین‌پژوه و همکارن (۱۳) برای ناحیه‌ای کردن بارش در ایران با استفاده از تکنیک‌های خوشه‌بندی از داده‌های بارش ۷۷ ایستگاه استفاده کردند. آنها با به‌کارگیری روش تحلیل مؤلفه های اصلی، از بین ۵۷ متغییر ۱۲ متغیر نهایی از جمله عمق کل بارش سالانه، میانگین بارش ماه‌های خشک، میانگین بارش ماه‌های خشک تابستان، میانگین بارش ماه جولای، میانگین بارش ماه اکتبر و غیره را انتخاب کردند. آنها سرانجام کشور ایران را به هفت ناحیه همگن بارشی تقسیم‌بندی کردند. دارند و منصوری دانشور (۱۲) با استفاده از داده‌های بارش برای سال‌های ۱۹۹۵ تا ۲۰۰۷ با استفاده از روش تحلیل مؤلفه‌های اصلی، بهترین متغیرها را برای ناحیه‌ای کردن میانگین سالانه بارش ناحیه زاگرس در ایران انتخاب کردند و برای خوشه‌بندی سلسله مراتبی به‌کار گرفتند. نتایج کار آنها نشان داد که ۹ ناحیه همگن بارش فصلی در ایران وجود دارد.

با بررسی تحقیقات انجام شده مشخص می‌شود که با توجه به کیفیت و کمیت داده‌ها و ویژگی‌های در دسترس، دسته‌ای از ویژگی‌ها به‌منظور انجام ناحیه‌بندی انتخاب می‌شوند. از جمله دیگر ویژگی‌های استفاده شده در تحقیقات مختلف می‌توان به ویژگی‌های فیزیکی حوضه مانند مساحت زهکشی، شیب میانگین حوضه، شیب رودخانه اصلی حوضه، طول رودخانه (۱۵، ۲۴، ۲۸ و ۲۹)، ویژگی‌های موقعیت جغرافیایی مانند طول جغرافیایی، عرض جغرافیایی و طول جغرافیایی از مرکز حوضه (۳۰)، عوامل هواشناسی مانند جهت باد، میانگین بارندگی سالانه، شدت بارندگی، دما، تابش خورشیدی، تبخیر و تعرق (۱۱، ۲۷ و ۳۳) و مقادیر آماری محاسبه شده از سری‌های زمانی در دسترس در ایستگاه مورد نظر (۲۳ و ۳۲) اشاره کرد. چاهوکی و همکاران (۵) از دسته‌ای از شاخص‌های هیدرولوژیکی حوضه به‌منظور منطقه‌ای کردن منحنی‌های تداوم جریان در حوضه‌های خشک و بدون آمار در ایران مرکزی استفاده کردند. آنها با کاربرد روش تحلیل مؤلفه‌های اصلی، تأثیرگذارترین شاخص‌ها از قبیل مساحت حوضه، طول آبراهه، شیب آبراهه اصلی و متوسط بارندگی سالانه به‌عنوان عوامل اصلی و اختلاف ارتفاع را به‌عنوان متغیر کمکی به‌منظور پیش‌بینی دبی در مناطق مورد نظر انتخاب کردند. در تحقیقات بررسی شده مشخص شد که در اکثر موارد، تحلیل‌ها براساس تعداد محدودی از ویژگی‌ها (اقلیمی، جغرافیایی و ...) انجام شده است که در این پژوهش این محدودیت حذف شده است. همچنین، در هیچ یک از تحقیقات وزن مناسبی برای ویژگی‌ها اعمال نشد و همهٔ ویژگی‌ها با درجه اهمیت یکسانی در ناحیه‌بندی مورد استفاده قرار گرفته‌اند. واضح است که هیچگاه تمام ویژگی‌ها در گروه‌های تعریف شده، دارای درجه اهمیت و سهم مشارکت یکسانی در ناحیه‌بندی نیستند. بنابراین هدف از پژوهش حاضر، ناحیه‌ای کردن حداکثر بارش ۲۴ ساعته حوضه دریاچه ارومیه بر اساس نگرش خوشه‌بندی و بررسی تأثیر نوع ویژگی‌های منتخب آماری، اقلیمی و جغرافیایی ایستگاه‌های باران‌سنجی با تعریف وزن مناسب برای هر یک از ویژگی‌ها در خوشه‌بندی است. در تحقیق حاضر در ابتدا با توجه به اهداف مورد نظر، ویژگی‌های مورد استفاده انتخاب می‌شوند و سپس، با استفاده از خروجی مدل تحلیل مؤلفه‌های اصلی، برای هر یک از آنها وزن مناسبی تعیین می‌شود. در ادامه، با استفاده از هر یک از گروه ویژگی‌ها و وزن‌های تعریف شده، ناحیه‌بندی در منطقه انجام می‌شود و نتایج خوشه‌بندی مورد بررسی قرار خواهند گرفت.

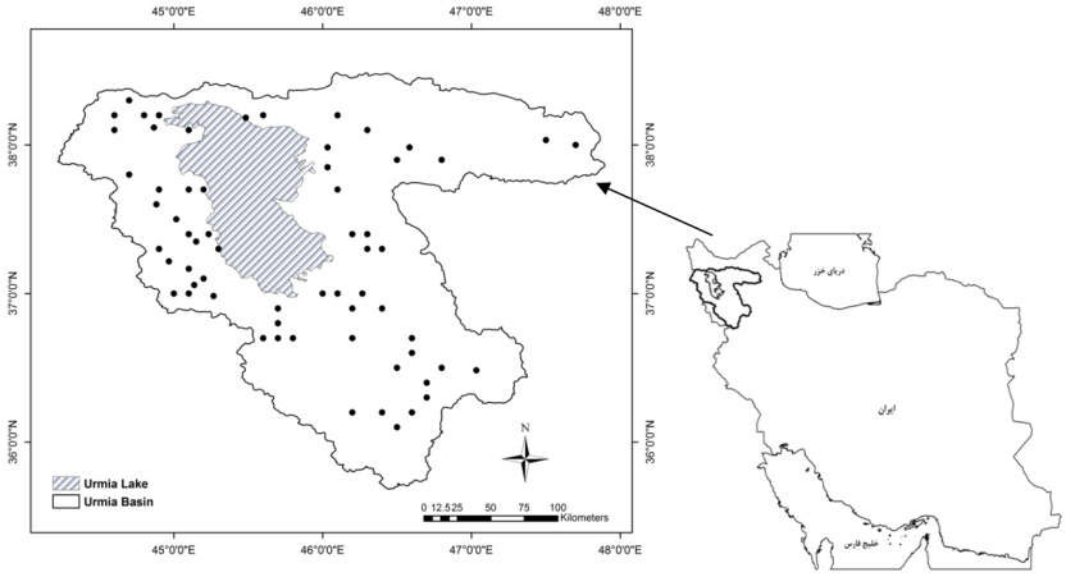
مواد و روش‌ها

منطقه مطالعاتی و داده‌ها

در این پژوهش به‌منظور دستیابی به اهداف مورد نظر، حوضه دریاچه ارومیه (Urmia Lake Basin) به عنوان منطقه مطالعاتی

www.SID.ir

Archive of SID



شکل ۱. موقعیت حوضه آبریز دریاچه ارومیه و موقعیت جغرافیایی ایستگاهای باران‌سنجی

درنظر گرفته شد. حوضه آبریز دریاچه ارومیه با مساحت حدود ۵۲ هزار کیلومتر مربع، در شمال غرب ایران قرار گرفته است و قسمت‌هایی از دامنه شمالی کوه‌های زاگرس و دامنه‌های جنوبی کوه سبلان و نیز دامنه‌های شمالی، غربی و جنوبی کوه سهند، حوضه را احاطه می‌کنند. مختصات جغرافیائی حوضه بین طول ′۱۳ °۴۴ تا ′۵۳ °۴۷ درجه شرقی و عرض ′۴۱ °۳۵ تا ′۳۰ °۳۸ درجه شمالی واقع شده است. ارتفاع ایستگاه‌های مورد مطالعه بین ۱۲۸۰ تا ۲۱۵۰ متر تغییر می‌کند. شکل ۱ موقعیت حوضه آبریز را به همراه موقعیت ایستگاه‌ها و چگونگی پراکندگی آنها در سطح حوضه نشان می‌دهد. پس از بررسی داده‌های جمع‌آوری شده، به‌دلیل نقص‌های فراوان موجود در داده‌های در دسترس، در نهایت تعداد ۶۳ ایستگاه باران‌سنجی با یک دوره آماری مشترک ۳۰ ساله از سال ۱۳۵۸ تا ۱۳۸۷ به‌عنوان ایستگاه‌های منتخب مورد استفاده قرار گرفتند.

انتخاب ویژگی‌ها

نوع ویژگی‌های مورد استفاده نقش بسیار اساسی در ناحیه‌بندی دارد. به‌منظور دستیابی به یک ناحیه‌بندی نسبتاً همه جانبه و قوی برای حداکثر بارش ۲۴ ساعته در حوضه دریاچه ارومیه از گروه گسترده‌ای از اطلاعات آماری (Statistical)، اقلیمی (Climatic) و جغرافیایی (Geographical) و ترکیبات مناسب آنها در پژوهش حاضر استفاده شد. در ادامه با استفاده از سه گروه ویژگی نام برده شده، هفت دسته ویژگی به‌منظور ناحیه‌بندی حداکثر بارش ۲۴ ساعته تعریف شد. گروه اول که شامل همه ویژگی‌ها (all attr)، گروه دوم شامل ویژگی‌های کاملاً اقلیمی (clim)، گروه سوم شامل ویژگی‌های جغرافیایی (geo)، گروه چهارم شامل ویژگی‌های آماری (stat)، گروه پنجم شامل ترکیبی از ویژگی‌های اقلیمی و جغرافیایی (clim&geo)، گروه ششم شامل ویژگی‌های اقلیمی و آماری (clim&stat) و گروه هفتم شامل ویژگی‌های جغرافیایی و آماری (geo&stat) است. در جدول ۱ هفت گروه تعریف شده و ویژگی‌های قرار گرفته شده در هر یک از گروه‌ها مشخص شده است.

فاصله اقلیدسی (Euclidean distance)

در فضای متریک، مشابهت اغلب بر اساس معیار فاصله تعریف می‌شود. فاصله اقلیدسی در واقع، اندازه خط راست بین دو شی را اندازه‌گیری می‌کند و به‌صورت رابطه (۱) تعریف می‌شود:

www.SID.ir

Archive of SID

جدول ۱. مشخصات هفت گروه ویژگی تعریف شده

ویژگی	اقلیمی- جغرافیایی- آماری (all attr)	اقلیمی (clim)	جغرافیایی (geo)	آماری (stat)	اقلیمی- جغرافیایی (clim&geo)	اقلیمی- آماری (clim&stat)	جغرافیایی- آماری (stat&geo)
y	√		√		√		√
x	√		√		√		√
h	√		√		√		√
d	√		√		√		√
Sd	√			√		√	√
Lcv	√			√		√	√
Lcs	√			√		√	√
Lck	√			√		√	√
Lcv/Lcs	√			√		√	√
Lcv/Lck	√			√		√	√
MR25th	√			√		√	√
MR75th	√			√		√	√
MAR	√	√			√	√	
Rs/Rw		√			√	√	
Dd/Dw	√	√			√	√	

عرض جغرافیایی (درجه): y، طول جغرافیایی (درجه): x، فاصله از سطح آب آزاد (متر): h و فاصله ایستگاه‌ها از مرکز دریاچه ارومیه (کیلومتر): d (نقطه ثقل دریاچه به‌عنوان مرکز دریاچه درنظر گرفته شد)، ضریب پراکندگی خطی بارندگی: LCv، ضریب چولگی خطی بارندگی: LCs، ضریب کشیدگی خطی بارندگی: LCk، انحراف معیار حداکثر بارش روزانه (میلی‌متر): Sd، نسبت ضریب پراکندگی خطی به ضریب چولگی خطی: LCv/LCs، نسبت ضریب پراکندگی به ضریب کشیدگی خطی: LCv/LCk، مقدار چارک اول حداکثر بارش روزانه: MR_{25th}، مقدار چارک سوم حداکثر بارش روزانه: MR_{75th}، میانگین حداکثر بارش روزانه (میلی‌متر در روز): MAR، نسبت میانگین بارش تابستانه به زمستانه: Rs/Rw، نسبت روزهای خشک به تر: Dd/Dw. (با فرض بارش صفر در روز خشک)

$$d_{ij} = \left[\sum_{m=1}^{M} \left(X_m^i - Y_m^j \right)^2 \right]^{\frac{1}{n}} \qquad (1)$$

فاصله اقلیدسی وزن‌داده شده نیز به‌صورت رابطه (۲) تعریف می‌شود:

$$d_{ij} = \left[\sum_{m=1}^{M} W_m \left(X_m^i - Y_m^j \right)^2 \right]^{\frac{1}{n}} \qquad (2)$$

که در این روابط، d_{ij} فاصله اقلیدسی وزن داده شده بین ایستگاه‌های مورد نظر i و j، W_m وزن تخصیص داده شده برای m امین ویژگی ایستگاه مورد نظر، X_{ij} مقدار کمی m امین ویژگی در ایستگاه i و M تعداد ویژگی‌ها است. مقدار X_m عبارتست از مقدار استاندارد شده ویژگی‌های ایستگاه مورد نظر که قبل از محاسبه d_{ij} نرمال‌سازی می‌شوند. به عبارت دیگر، ویژگی‌ها دارای میانگین صفر و واریانس واحد هستند (۱۱).

فاصله اقلیدسی بیشترین کاربرد را در اندازه‌گیری میزان تشابه در خوشه‌بندی دارد. به‌طوری‌که بیشتر پژوهش‌های نام برده شده در بالا، از فاصله اقلیدسی برای اندازه‌گیری میزان تشابه استفاده کرده‌اند، به همین دلیل از این فاصله در پژوهش حاضر استفاده شده است. نرمال‌سازی داده‌ها با استفاده از روش واریانس و میانگین داده‌ها با رابطه (۳) انجام پذیرفت که در آن $\overline{X}$ میانگین و σ_X انحراف معیار ویژگی X است (۲۲).

$$X^* = \frac{X - \overline{X}}{\sigma_X} \qquad (3)$$

روش تحلیل مؤلفه‌های اصلی (Principal Component Analysis)

این روش متداول‌ترین روش چند متغییره آماری است که براساس کار اصلی پیرسون هتلینگ، در سال ۱۹۳۳ توسعه یافته است. در این روش، تحلیل مؤلفه‌های اصلی نوعی از تحلیل

www.SID.ir

Archive of SID

آماری است که تعداد کمتری از عوامل را به نام مؤلفه‌های اصلی از میان عوامل اولیه گزینش می‌کند. در این پژوهش از ماتریس دورانی واریمکس (Varimax) برای دوران مؤلفه‌های اصلی استفاده شد. در این دوران، بیشترین بار عاملی در هر متغیر را در مؤلفه‌های اصلی استخراج می‌شود (برای اطلاعات بیشتر از روش PCA به مطالعات ۱۸، ۱۹، ۲۰ و ۲۱ مراجعه شود). در تحقیق حاضر نیز، از روش تحلیل مؤلفه‌های اصلی به‌منظور پیدا کردن درجه اهمیت و میزان سهم مشارکت هر یک از متغیرها در مؤلفه‌های اصلی نهایی دوران داده شده استفاده شد. بارهای عاملی از جنس همبستگی و مجذور مقدار بارهای عاملی برای هر متغیر در مؤلفه‌های اصلی نهایی چنانچه بیان شد، برابر با مقدار ویژه آن متغیر است که درصد سهم واریانس متغیر را در مؤلفه اصلی نشان می‌دهد. بنابراین به‌منظور وزن‌دهی ویژگی‌ها با روش PCA، مقدار ویژه به‌دست آمده از ماتریس واریمکس برای هر یک از ویژگی‌ها در مؤلفه‌های اصلی دوران یافته (Rotated Principal Component: RPC) و به‌صورت مقادیر حداکثر و متوسطی از بارهای عاملی (از جنس همبستگی) مورد استفاده قرار می‌گیرد. بنابراین، ۳ سناریوی وزن‌دهی شامل ۱- ویژگی‌ها با وزن‌هایی برابر و مساوی یک (w=1)، ۲- متوسطی از مقادیر ویژه هر متغیر در مؤلفه‌های نهایی ($w=PCA_{ave}$) و ۳- حداکثر مقدار ویژه هر متغیر در مؤلفه‌های نهایی ($w=PCA_{max}$) تعریف شدند.

خوشه‌بندی وارد (Ward)

روش وارد از جمله روش‌های خوشه‌بندی است که هدف آن یافتن خوشه‌های کروی و متراکم است. روش وارد اغلب در تحلیل فروانی برای تشریح همگنی خوشه‌ها مورد استفاده قرار می‌گیرد. این روش دارای الگوریتم خوشه‌بندی سلسله مراتبی بوده که در ابتدا هر ایستگاه را به‌عنوان یک خوشه درنظر می‌گیرد و سپس خوشه‌ها با استفاده از فرایند تحلیل واریانس ادغام می‌شوند و میزان تشابه بین آنها اندازه‌گیری می‌شود (۳۱). روش وارد برای حداقل کردن واریانس درون خوشه‌ای استفاده می‌شود و در هر گام هر دو خوشه با حداقل فاصله با یکدیگر ادغام می‌شوند که برای یک خوشه r با s ایستگاه و p ویژگی به‌صورت ($X=X_1,X_2,...,X_p$) با رابطه (۴) تعریف می‌شود.

$$ESS_r = \sum_{j=1}^{s} \left(X_j - \overline{X_r}\right)' \left(X_j - \overline{X_r}\right) \qquad (۴)$$

که طبق این رابطه، $X_j = \left(X_{j1}, X_{j2}, ..., X_{jp}\right)$ شامل برداری از ویژگی‌ها در ایستگاه j و X_r شامل برداری از میانگین ویژگی‌ها در خوشه r است، ESS_r برابر مجموع مربعات خطا است که هدف از تعیین خوشه‌ها در روش وارد، به‌دست آوردن حداقل مقدار خطا در مرحله بعدی است (برای اطلاعات بیشتر در مورد این روش به تحقیق وارد (۳۱) رجوع شود). روش وارد نسبت به سایر روش‌ها نتایج بهتری ارائه کرده است، به همین دلیل به‌عنوان روش نهایی برای خوشه‌بندی در این تحقیق انتخاب شد (۹).

آزمون ناهمگنی منطقه‌ای

از آزمون ناهمگنی (Heterogeneity) گشتاور خطی به‌منظور برآورد درجه غیرهمگنی در یک گروه از ایستگاه‌ها و ارزیابی اینکه آیا آنها می‌توانند به‌طور معقول به‌عنوان یک گروه همگن قرار گیرند یا نه، استفاده شده است. معیار غیرهمگنی، تغییرات بین ایستگاهی گشتاورهای خطی نمونه را برای گروهی از ایستگاه‌ها با مقدار مورد انتظار برای یک ناحیه همگن مقایسه می‌کند (۱۶ و ۱۷). برای آزمودن همگنی، سه آزمون مختلف وجود دارد: ۱- آزمون بر پایه معیار پراکنش خطی (L-CV)؛ ۲- آزمون بر پایه پراکنش و چولگی خطی و ۳- آزمون بر پایه چولگی و گشیدگی خطی. آماره این آزمون‌ها V نامیده شده است و هاسکینگ و والیس (۱۷) مشاهده کردند که آماره‌های H_2 و H_3 بر اساس معیارهای V_2 و V_3 فاقد توان لازم برای تفکیک بین نواحی همگن و ناهمگن بوده و اینکه آماره H_1 بر اساس معیار V_1 توان خیلی بهتری برای تفکیک دارد. بنابراین، آماره H_1 به‌عنوان شاخص اساسی برای ناهمگنی توصیه شده

www.SID.ir

Archive of SID

است و به‌صورت رابطه (۵) و (۶) تعریف می‌شوند (۱۶):

$$H_i = \frac{(V_i - \mu_{V_i})}{\sigma_{V_i}};\quad i = ۱, ۲, ۳ \qquad (۵)$$

$$V_i = \left\{ \sum_{i=۱}^{N} n_i (\tau^{(i)} - \tau^R)^۲ \Big/ \sum_{i=۱}^{N} n_i \right\}^{۰/۵} \qquad (۶)$$

طبق این رابطه، μ_{V_i} و σ_{V_i} به‌ترتیب میانگین و انحراف معیار مقادیر شبیه‌سازی شده V_i است. در پژوهش حاضر شاخص همگنی با استفاده از شبیه‌سازی مونت کارلو برای هر یک از مناطق مصنوعی با ۵۰۰۰ تکرار بر اساس داده‌های واقعی ساخته و محاسبه شد. علاوه بر این، مقادیر میانگین منطقه‌ای نسبت‌های گشتاور خطی از رابطه (۷) محاسبه می‌شود:

$$t^R = \frac{\sum_{i=۱}^{N} n_i t^{(i)}}{\sum_{i=۱}^{N} n_i} \qquad (۷)$$

که N تعدادایستگاه‌ها در منطقه، n_i طول دوره آماری هر ایستگاه و $\tau^{(i)}$ نسبت‌های گشتاور خطی برای ایستگاه i ام است.

نتایج و بحث

نتایج روش PCA

با بررسی ویژگی‌های آماری شرایط موجود برای درصد واریانس اختصاص یافته به هر ویژگی، علاوه بر سناریوی وزن‌دهی که در آن وزن تمام ویژگی‌ها یکسان درنظر گرفته می‌شود $w = ۱$، دو سناریوی متوسط و حداکثر مقادیر ویژه هر متغیر در مؤلفه‌های نهایی به‌عنوان دو سناریو دیگر برای وزن‌دهی ویژگی‌ها انتخاب شدند. نتایج حاصل از وزن‌های نهایی برای هر هفت گروه ویژگی‌ها در جدول ۲ آورده شده است.

براساس نتایج ارائه شده، از میان سه حالت مختلف وزن‌دهی، روش میانگین هندسی کمترین مقدار وزن را به هر ویژگی اختصاص داده است. همان‌طور که مشاهده می‌شود، زمانی که از همه ویژگی‌ها (دسته اول ویژگی‌ها) برای خوشه‌بندی استفاده می‌شود، وزن‌ها کاهش می‌یابند. در سناریوی دوم وزن‌دهی (PCA_{ave})، وزن هر ویژگی به یکدیگر نزدیک‌تر است. در سناریوی سوم وزن‌دهی (PCA_{max})، کاهش وزن هر ویژگی بیشتر می‌شود. نکته قابل توجه این است که در هر گروه ویژگی، با افزایش تعداد ویژگی‌ها سهم هر ویژگی کمتر شده است. همچنین، برخی ویژگی‌ها سهم بیشتری را نسبت به سایرین به خود اختصاص داده‌اند که کاملاً منطقی و قابل قبول است. بنابراین، با اعمال این وزن‌ها هر ویژگی به میزان سهم خود در خوشه‌بندی تأثیرگذار است.

نتایج خوشه‌بندی به روش وارد

به‌منظور ناحیه‌بندی حوضه دریاچه ارومیه، پس از انتخاب ویژگی‌ها و وزن‌های مناسب برای آنها، با استفاده از فاصله اقلیدسی ماتریس تشابه برای هر دسته ویژگی تشکیل شد. در ادامه، خوشه‌بندی وارد برای هر گروه از ویژگی‌ها انجام شد و تعداد نواحی به‌دست آمده در حوضه از روش خوشه‌بندی وارد در جدول ۳ آورده شده است.

با توجه به نتایج جدول ۳، در تعداد خوشه‌ها در سه سناریوی وزن‌دهی تغییر چندانی مشاهده نمی‌شود. در برخی گروه‌ها، تعداد نواحی با اعمال وزن ویژگی‌ها افزایش یافته و در برخی حالت‌های دیگر کاهش یافته است، درنتیجه قانون کلی را نمی‌توان ارائه داد. اما با توجه به نتایج به‌دست آمده در سناریوی $w = ۱$ به‌دست آمده در مورد در دسته اول، مشاهده می‌شود که کمترین تعداد خوشه‌ها در اکثر مواقع در دسته با بیشترین تعداد ویژگی‌ها رخ داده و می‌توان نتیجه گرفت که کاهش ویژگی‌ها در بیشتر مواقع موجب افزایش خوشه‌ها شده است. نتایج حاصل از تعداد خوشه‌های به‌دست آمده در سناریوی وزن‌دهی اول $w = ۱$ با تحقیق مدرس (۹) برای بارش حداکثر در کل ایران مطابقت داشته و نتایج این بخش از پژوهش را تأیید می‌کند.

نتایج معیار همگنی بر اساس وزن برابر ویژگی‌ها

بنابر توصیه هاسکینگ و والیس یک ناحیه وقتی می‌تواند ناحیه

www.SID.ir

Archive of SID

جدول ۲. وزن هر یک از ویژگی‌ها در گروه‌ها

	Dd/Dw	Rs/Rw	AMR	MR75th	MR25th	L-CV/L-CK	L-CV/L-CS	L-CK	L-CS	L-CV	Sd	d	h	y	x
All attr															
1	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱
PCA-ave	۰/۲۲	۰/۱۷	۰/۱۷	۰/۲۱	۰/۱۸	۰/۱۹	۰/۱۹	۰/۱۹	۰/۱۹	۰/۲	۰/۲۲	۰/۱۵	۰/۲۶	۰/۲۱	۰/۲۲
PCA-max	۰/۷۳	۰/۹۴	۰/۸۴	۰/۹۲	۰/۹۱	۰/۹	۰/۸۱	۰/۷۸	۰/۹۱	۰/۹۵	۰/۵	۰/۸۷	۰/۴۷	۰/۶۹	۰/۳۹
clim															
1	۱	۱	۱												
PCA-ave	۰/۴	۰/۳۵	۰/۳۹												
PCA-max	۰/۹۷	۰/۹۹	۰/۹۷												
geo															
1												۱	۱	۱	۱
PCA-ave												۰/۴۹	۰/۳۹	۰/۴	۰/۴۱
PCA-max												۰/۵۲	۰/۹۱	۰/۹۱	۰/۹۱
stat															
1				۱	۱	۱	۱	۱	۱	۱	۱				
PCA-ave				۰/۳۱	۰/۲۴	۰/۲۷	۰/۲۷	۰/۲۷	۰/۲۷	۰/۲۶	۰/۳۴				
PCA-max				۰/۱۹	۰/۱۹	۰/۲	۰/۲۲	۰/۱۵	۰/۲۶	۰/۲۱	۰/۲۲				
clim&geo															
1	۱	۱	۱									۱	۱	۱	۱
PCA-ave	۰/۲۸	۰/۲۵	۰/۲۵									۰/۲۶	۰/۲۹	۰/۳۶	۰/۳۳
PCA-max	۰/۹۱	۰/۹۸	۰/۸۹									۰/۹۲	۰/۹۱	۰/۵۸	۰/۵۶
clim&stat															
1	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱				
PCA-ave	۰/۲۱	۰/۱۸	۰/۱۹	۰/۲۴	۰/۲۰	۰/۲۱	۰/۲۰	۰/۲۰	۰/۲۲	۰/۲۱	۰/۲۵				
PCA-max	۰/۹۱	۰/۹۸	۰/۹۵	۰/۸۴	۰/۹۲	۰/۹۱	۰/۹۵	۰/۸۳	۰/۷۸	۰/۹۳	۰/۴۹				
clim&stat															
1				۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	۱
PCA-ave				۰/۲۳	۰/۱۸	۰/۲۰	۰/۲۲	۰/۲۱	۰/۲۰	۰/۲۱	۰/۲۵	۰/۱۹	۰/۲۰	۰/۲۴	۰/۲۳
PCA-max				۰/۸۱	۰/۹۱	۰/۶۹	۰/۵۵	۰/۸۱	۰/۹۶	۰/۹۳	۰/۹۳	۰/۸۱	۰/۸۱	۰/۹۲	۰/۴۵

جدول ۳. تعداد خوشه‌های نهایی در هر یک از سناریوهای وزن‌دهی

سناریوهای وزن‌دهی	اقلیم- جغرافیایی- آماری	اقلیمی	جغرافیایی	آماری	اقلیمی- جغرافیایی	اقلیمی- آماری	جغرافیایی- آماری
1	۳	۴	۵	۵	۷	۵	۴
PCAave	۳	۴	۶	۴	۴	۵	۴
PCAmax	۳	۵	۶	۴	۴	۳	۳

www.SID.ir

Archive of SID

جدول ۴. مقادیر H برای سناریوی وزن‌دهی در هر گروه از ویژگی‌ها

سناریوها	خوشه ۱ (C6)	خوشه ۲ (C5)	خوشه ۳ (C4)	خوشه ۴ (C3)	خوشه ۵ (C1)	خوشه ۶ (C1)	خوشه ۷
				all attr			
1	۰/۷۵	−۰/۴۴	۰/۲۱				
PCA_{ave}	−۱/۵۴	−۰/۰۹	−۰/۰۹				
PCA_{max}	۰/۵۴	۰/۶۶	−۰/۲۴				
				clim			
1	۰/۸۸	۰/۹۲	−۰/۶۹	۰/۱۲			
PCA_{ave}	۰/۶۲	۰/۹۱	−۰/۷۴	۰/۱۲			
PCA_{max}	۰/۹۸	۰/۸۸	۰/۵۷	−۰/۷۸	۰/۳۵		
				geo			
1	۰/۲۲	۰/۴۶	−۰/۲۹	۰/۳۵	−۰/۲۴		
PCA_{ave}	۰/۶۲	۰/۷۸	۰/۴۶	−۰/۳۶	۰/۵۸	۰/۸۷	
PCA_{max}	۰/۵	۰/۸۵	۰/۴	−۰/۳۵	−۰/۱۴	۰/۹۵	
				stat			
1	۰/۰۸	۰/۳۶	−۰/۱۱	−۱/۵۸	−۰/۷۴	۰/۰۸	
PCA_{ave}	۰/۴۵	۰/۲۱	−۱/۶	−۰/۸۸	۰/۴۵		
PCA_{max}	۰/۹۶	۰/۷۷	−۱/۵۶	−۰/۷۵			
				clim&geo			
1	۰/۰۷	−۰/۲۱	۰/۸۶	۰/۴۳	۰/۱	۰/۴۴	−۰/۰۹
PCA_{ave}	۰/۷۲	−۰/۲۴	۰/۴۸	۰/۷۳			
PCA_{max}	۰/۱۸	−۰/۲۸	−۰/۶۹	۰/۶			
				clim&stat			
1	۰/۵۶	−۲	−۱/۴۷	−۰/۴	−۱/۲۵		
PCA_{ave}	−۰/۵۴	−۱/۴۷	−۱/۴۴	۰/۶	−۰/۵۷		
PCA_{max}	۰/۱۲	۰/۵۵	۰/۸۸				
				geo&stat			
1	۰/۴۱	۰/۴۲	−۰/۱,۷۵	−۲/۲۳			
PCA_{ave}	−۰/۲۱	−۰/۱	۰/۸۵				
PCA_{max}	۰/۷۱	−۰/۶۴	۰/۶۶				

همگن قابل قبول باشد که $H < ۱$ باشد و می‌تواند ناحیه نسبتاً ناهمگن باشد، اگر $۱ \le H < ۲$ و به‌عنوان ناحیه ناهمگن تلقی شود زمانی که $H \ge ۲$ باشد (۲۱). میزان مقادیر همگنی بر اساس روابط ارائه شده در بخش مواد و روش‌ها، برای سناریوی وزن برابر و برای گروه‌های مختلف از ویژگی‌های بیان شده، در جدول ۴ آورده شده است.

نتایج اندازه همگنی در دسته اول که شامل ویژگی‌های اقلیمی، جغرافیایی و آماری است و تعداد نواحی تشکیل شده در هر سه سناریو یکسان است، کمترین مقادیر H در سناریوی $w=PCA_{ave}$ مشاهده می‌شود که حاکی از بهبود همگنی در حوضه در این دسته است. در دسته اقلیمی نیز نسبت به سناریوی بدون وزن $w = ۱$ مشاهده می‌شود که همگنی در

www.SID.ir

Archive of SID

سناریوی $w=PCA_{ave}$ بهبود یافته است. به‌طورکلی، می‌توان نتیجه گرفت که در سناریوهای وزن داده شده مقادیر H حاکی از بهبود همگنی در حوضه است. همان‌طور که در جدول ۴ مشاهده می‌شود، نتایج اندازه همگنی حاصل در سناریوی $w=PCA_{max}$ که وزن‌های پایینی را شامل می‌شوند نسبت به دو سناریوی دیگر ضعیف‌تر بوده است و نتیجه حاکی از کم بودن میزان همگنی نواحی تشکیل شده نسبت به دو سناریوی دیگر در سطح حوضه است.

نتایج الگوی مکانی خوشه‌بندی حداکثر بارش ۲۴ ساعته

الف) براساس سناریو $w=1$

با توجه به نتایج در شکل ۲ (الف) تا (د)، در بیشتر دسته ویژگی‌ها، ایستگاه‌های جنوبی حوضه و جنوب دریاچه (تقریباً مرکز حوضه) خوشه‌های همگن‌تری را تشکیل داده‌اند. همچنین کاملاً مشهود است که خوشه‌بندی ایستگاه‌های مورد مطالعه، بسته به نوع ویژگی‌ها درنظر گرفته شده در هر یک از دسته‌ها الگوی بسیار متفاوتی را در سطح حوضه ایجاد کرده‌اند. با توجه به نتایج حاصله در این سناریو، ویژگی‌های جغرافیایی، اقلیمی و اقلیمی- جغرافیایی توانایی بسیار بالایی را در تشکیل نواحی با الگوی مکانی مشخص از خود نشان دادند. به‌طوری‌که در این دسته ویژگی‌ها، خوشه‌ها در سطح حوضه با الگویی بسیار یکپارچه و قابل تفکیک پراکنده شده‌اند. چنانچه در شکل ۲ (الف) تا (د) نیز مشاهده می‌شود، پراکندگی خوشه‌ها در شرق دریاچه بیشتر بوده و با دور شدن از دریاچه خوشه‌بندی نظم بیشتری را به خود می‌گیرد. شمال، غرب و به‌ویژه در جنوب دریاچه خوشه‌های تشکیل شده با الگوی مکانی یکپارچه‌تر قابل تفکیک هستند که در همه دسته ویژگی‌ها (به‌جز دسته آماری و اقلیمی- آماری) به‌ویژه دسته جغرافیایی و جغرافیایی- اقلیمی این نتیجه مشاهده شد. علاوه بر این، مشاهده می‌شود که با دور شدن از دریاچه خوشه‌ها متراکم‌تر و همگن‌تر شده‌اند.

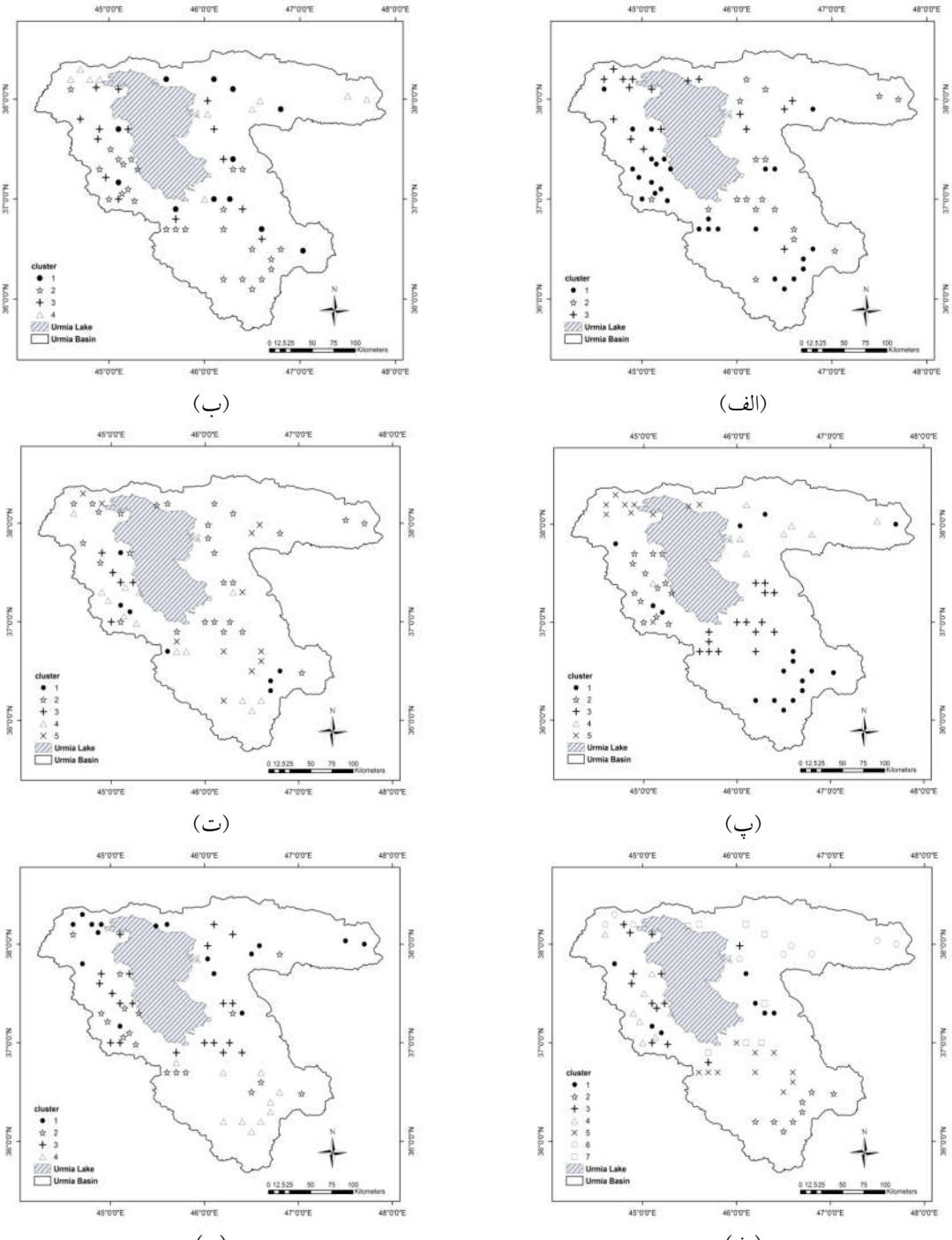
البته لازم به ذکر است که تشکیل خوشه‌هایی منسجم و متراکم که به‌خوبی در سطح حوضه قابل تفکیک باشند، دلیلی بر وجود نتایج بهتر و برآوردهایی دقیق از مقادیر حدی نیست. در بحث ناحیه‌بندی، تشکیل این خوشه‌ها در صورت همجواری و یا قرارگیری جغرافیایی این خوشه‌ها با ایستگاه‌های فاقد آمار یا با کمبود آمار که نزدیک به این خوشه‌ها باشند، می‌تواند شرایطی را ایجاد کند که از داده‌های ناحیه‌ای یا ایستگاهی واقع شده در این خوشه‌ها برای بازسازی و بهبود یا برآورد داده برای ایستگاه مورد نظر (با درنظر گرفتن سایر شرایط) استفاده کرد. به توجه به‌دلیل ذکر شده در این پژوهش، یک الگوی پسندیده و قابل قبول برای تحلیل الگوی مکانی منسجم و قابل تفکیک برای خوشه‌ها در سطح حوضه ذکر شده است. چنانچه در نتایج جدول ۴ نیز قابل مشاهده است، برای سناریوی $w=1$ مقدار همگنی به‌دست آمده در هر یک هفت گروه ویژگی مورد بررسی با نتایج الگوی مکانی خوشه‌ها در سطح حوضه به‌ویژه در خصوص H_2 و H_3 مطابقت زیادی نشان می‌دهد. در این سناریو، با توجه به آنکه تمام مقادیر H در محدوده همگن قرار دارند، در حالاتی که بهترین الگوی مکانی در سطح حوضه مشاهده می‌شود (گروه‌های اقلیمی، جغرافیایی و جغرافیایی- آماری) مقدار اندازه همگنی تا حدودی در اندازه H_1 و بیشتر در مقادیر H_2 و H_3 این نتیجه را تأیید می‌کنند.

ب) وزن‌دهی ویژگی‌ها بر اساس w=PCAave

چنانچه در شکل۳ (الف) تا (د) مشاهده می‌شود، با درنظر نگرفتن یک یا دو ایستگاه که از لحاظ مکانی ناجور به نظر می‌رسند، در تمام نواحی موقعیت قرارگیری ایستگاه‌ها در هر یک از ناحیه‌های مشخص شده در منطقه، از لحاظ موقعیت جغرافیایی و الگوی مکانی در خوشه‌بندی رعایت شده و کاملاً قابل توجیه است. در این سناریو نیز بهترین الگوی مکانی خوشه‌بندی در شکل ۳ (ث) شامل ویژگی‌ها اقلیمی- جغرافیایی و سپس شکل ۳ (د) شامل ویژگی‌ها جغرافیایی- آماری و شکل ۳ (الف) تمام ویژگی‌ها مشاهده می‌شود. لازم به ذکر است که در سایر نواحی نیز می‌توان الگوی مکانی نسبتاً منظمی را تشخیص داد. نتایج این سناریو وزن‌دهی (PCAave) حاکی از

www.SID.ir

Archive of SID



(الف) (ب)

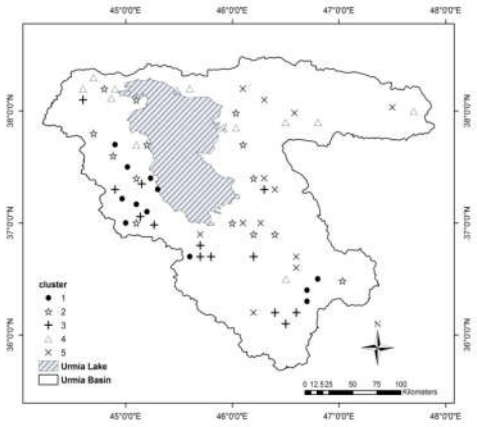
(پ) (ت)

(ث) (ج)

شکل ۲. خوشه‌بندی حداکثر بارش ۲۴ ساعته برای سناریوی وزن‌دهی برابر: الف) گروه ویژگی‌ها all attr، ب) گروه ویژگی‌ها clim، پ) گروه ویژگی‌ها geo، ت) گروه ویژگی‌ها stat، ث) گروه ویژگی‌ها clim&geo، ج) گروه ویژگی‌ها clim&stat و د) گروه ویژگی‌ها geo&stat

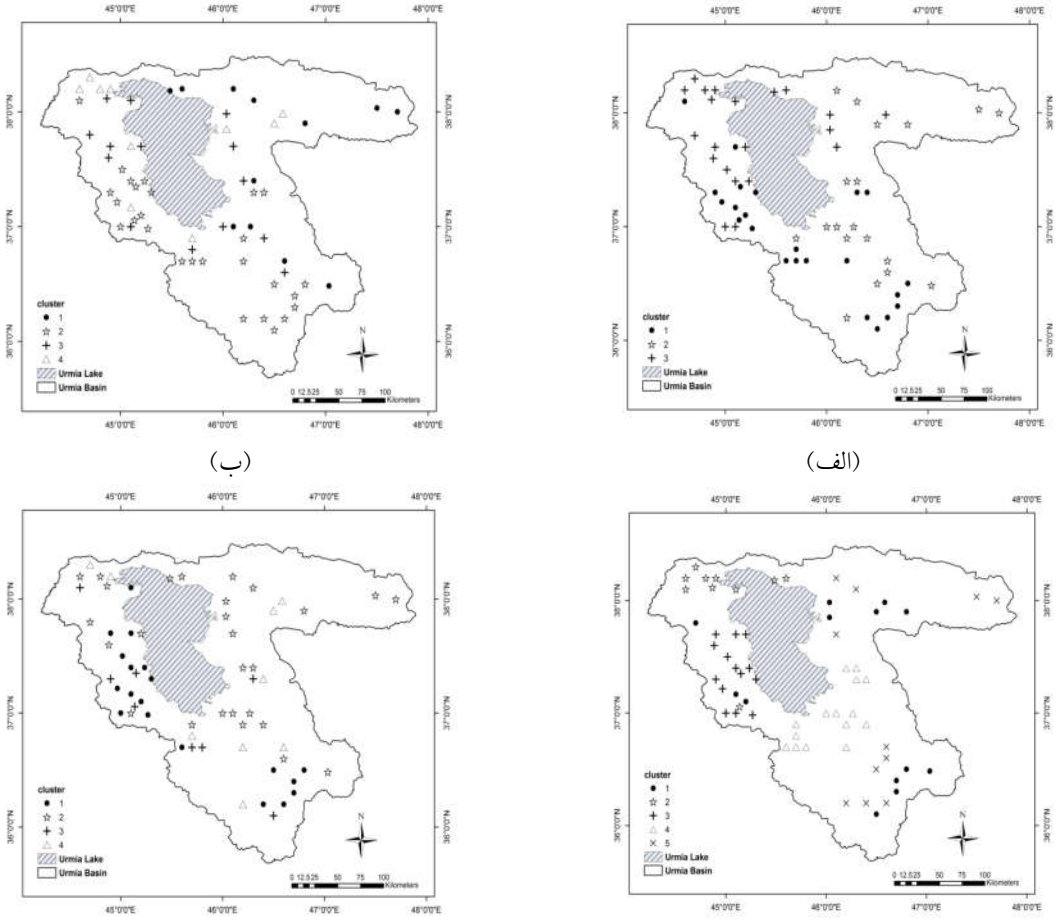
www.SID.ir

Archive of SID



(د)

ادامه شکل ۲.



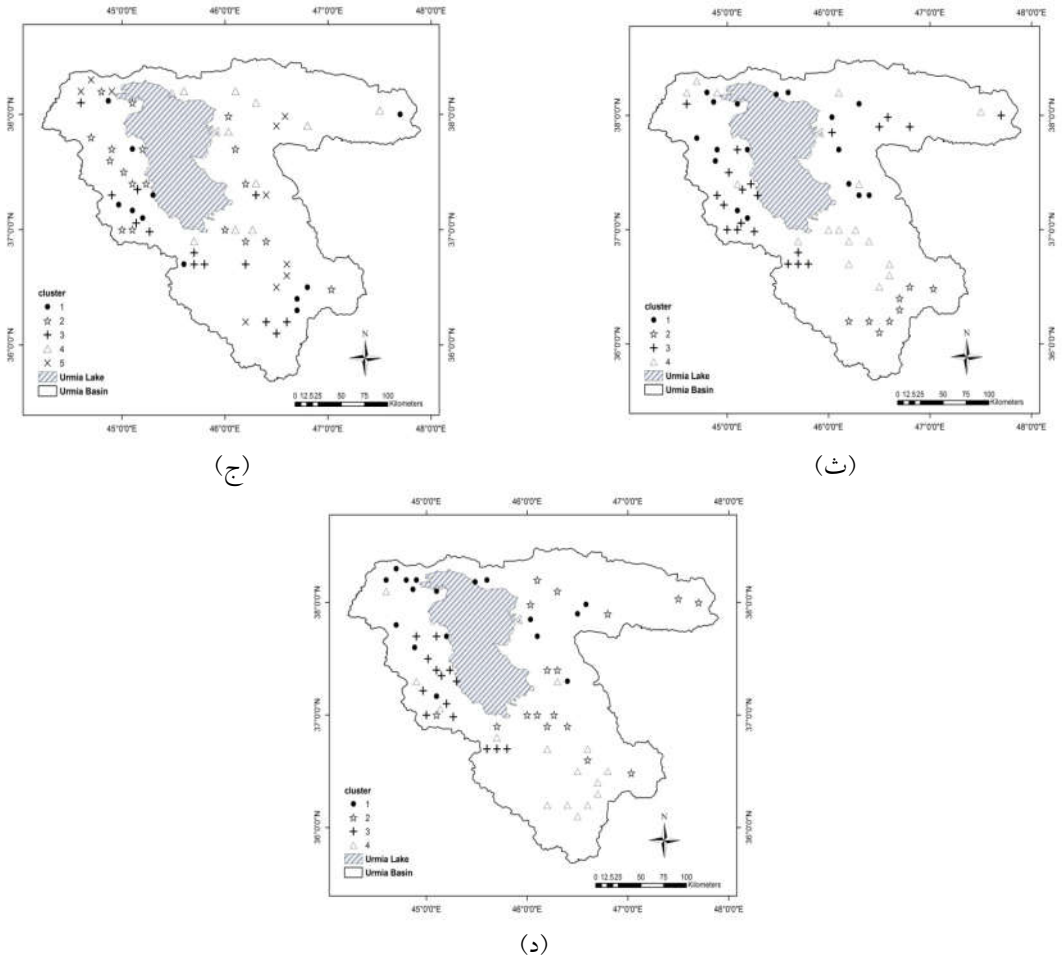
(ب) (الف)

(ت) (پ)

شکل ۳. خوشه‌بندی حداکثر بارش ۲۴ ساعته برای سناریوی وزن‌دهی ave: الف) گروه ویژگی‌ها all attr، ب) گروه ویژگی‌ها clim، پ گروه ویژگی‌ها geo، ت) گروه ویژگی‌ها stat، ث) گروه ویژگی‌ها clim&geo، ج) گروه ویژگی‌ها clim&stat و د) گروه ویژگی‌ها geo&stat

www.SID.ir

Archive of SID



(ج)

(ث)

(د)

ادامه شکل ۳.

آن بود که وزن‌دهی ویژگی‌ها تا حدی توانسته است خوشه‌بندی در اطراف دریاچه را نسبت به سناریوی قبل یکپارچه‌تر و از پراکندگی و بی‌نظمی خوشه‌ها در اطراف دریاچه بکاهد. چنانچه مشاهده می‌شود، در این سناریو دریاچه خود به‌عنوان یک عامل طبیعی توانسته ایستگاه‌های موجود در حوضه را به ایستگاه‌های شرقی، غربی، شمالی و جنوبی تفکیک کند. در این سناریو وزن‌دهی ویژگی‌ها به وضوح موجب بهبود پراکندگی خوشه‌ها و کاهش بی‌نظمی‌های الگوهای مکانی در اطراف دریاچه در سطح حوضه شده است.

تأثیر وزن ویژگی‌ها در مقادیر اندازه همگنی در جدول ۴ نیز قابل مشاهده است به‌نحوی که، کاهش اندازه همگنی خود دلیلی بر بهبود همگنی خوشه‌های تشکیل شده و تأثیر مثبت اعمال وزن بر ویژگی‌ها نسبت به حالت بدون وزن است.

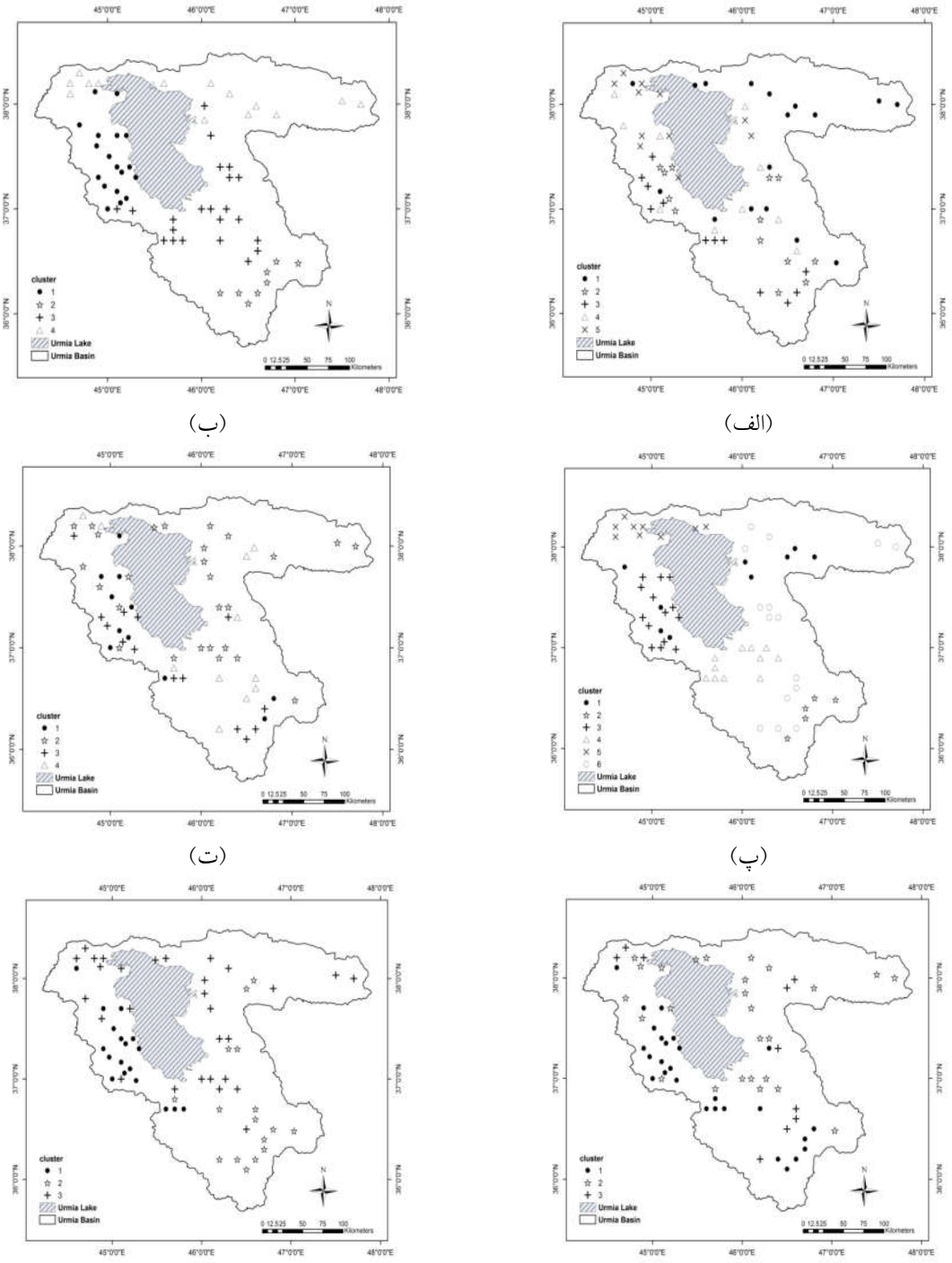
به‌طورکلی مشاهده می‌شود که مقادیر H نسبت به سناریوی بدون وزن قبلی کاهش نشان داده است. این نتیجه به‌ویژه در گروه‌هایی که الگوی مکانی بسیار منسجم و متراکمی را در حوضه از خود نشان داده‌اند (دسته اقلیمی، جغرافیایی و دسته جغرافیایی- آماری) تأثیر بیشتری در کاهش H داشته است. همچنین، در دسته اقلیمی در این سناریو خوشه‌بندی به‌ویژه در جنوب و غرب حوضه منسجم‌تر شده است.

ج) براساس سناریو w=PCAmax

نتایج حاصل از خوشه‌بندی در این سناریوی وزن‌دهی w=PCAmax در شکل ۴ (الف) تا (د) نشان داده شده است. چنانچه مشاهده می‌شود، نواحی خوشه‌بندی شده در

www.SID.ir

Archive of SID



(الف) (ب)

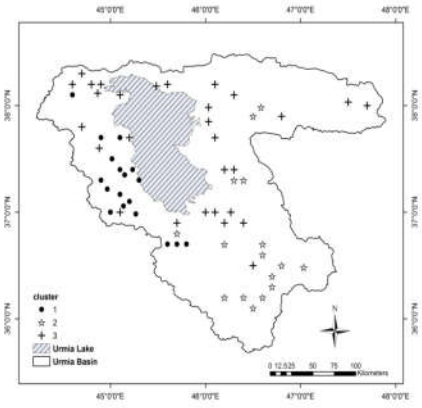
(پ) (ت)

(ث) (ج)

شکل ۴. خوشه‌بندی حداکثر بارش ۲۴ ساعته برای سناریوی وزن‌دهی gm: الف) گروه ویژگی‌ها all attr، ب) گروه ویژگی‌ها clim، پ) گروه ویژگی‌ها geo، ت) گروه ویژگی‌ها stat، ث) گروه ویژگی‌ها clim&geo، ج) گروه ویژگی‌ها clim&stat و د) گروه ویژگی‌ها (geo&stat)

www.SID.ir

Archive of SID



(د)

ادامه شکل ۴.

شکل‌های ۴ (الف)، (پ)، (ث)، (ج) و (د) الگوی مکانی بسیار خوبی، از لحاظ پراکندگی ایستگاه‌ها در هر یک از نواحی در سطح حوضه نشان داده‌اند. فقط در شکل ۴ (ب) که شامل ویژگی‌ها اقلیمی است، الگوی مکانی ایستگاه‌ها چندان منظم و یکپارچه به نظر نمی‌رسد.

نتایج حاصل از این سناریوی وزن (w=PCAmax) نشان داد که خوشه‌بندی انجام شده در هر یک از دسته ویژگی‌ها در سطح حوضه، به‌صورت محسوسی منسجم‌تر و یکپارچه‌تر از دو سناریو قبل است. همچنین، ناحیه‌بندی ایستگاه‌ها موجب تفکیک ایستگاه‌ها در سطح حوضه به‌صورت ایستگاه‌های شرقی، غربی، مرکزی و جنوب حوضه شده است. به‌طوری‌که، در دسته اقلیمی- جغرافیایی- آماری در جنوب دریاچه، جنوب و مرکز حوضه شمال و غرب دریاچه، افزایش یکپارچگی مکانی خوشه‌های تشکیل شده به وضوح دیده می‌شود. در دسته آماری غرب و مرکز حوضه و جنوب غربی دریاچه خوشه‌ها منسجم‌تر و با قدرت تفکیک بیشتری تشکیل شده‌اند. بهبود و افزایش تراکم و انسجام الگوی مکانی ایستگاه‌های هر خوشه در سطح حوضه در این سناریو، در دسته ویژگی‌های اقلیمی، جغرافیایی و اقلیمی- جغرافیایی که در سناریوهای قبل نیز مشاهده شده بود، در این سناریو بیشتر شده است. تا حدود زیادی نتایج جدول ۴ نیز با توجه به مقادیر H این بهبود را نشان داده است. نکته مهم در این سناریو، افزایش چشمگیر نظم و یکپارچگی در دسته آماری و اقلیمی- آماری است که در دو سناریو قبل مشاهده نشد، به‌طوری‌که با توجه به شکل خوشه‌ها در سطح حوضه با یک الگوی مکانی قابل تفکیک پراکنده شده‌اند. این نتیجه نشان می‌دهد که با استفاده از این سناریو می‌توان به‌طور نسبی با انتخاب هر دسته از ویژگی‌ها و وزن‌دهی آنها، انتظار ناحیه‌بندی منظم و قابل قبولی را در سطح حوضه داشت. نتایجی مشابه با تحقیق حاضر در تحقیق فاضل و همکاران (۱۴) مشاهده شد. آنها در نتایج خودشان ایستگاه‌های سطح حوضه را به سه زیر گروه جنوبی، غربی و مرکزی تقسیم کردند که بخشی از نتایج حاصل از این تحقیق را تأیید می‌کند.

نتیجه‌گیری

با توجه به خوشه‌بندی انجام شده در هر یک از سه دسته سناریوی وزن‌دهی، مشاهده شد که تعداد چهار خوشه در حوضه بیشترین فراوانی را در بین سایر تعداد خوشه‌ها داراست. نتایج نشان داد که ایستگاه‌ها با فاصله گرفتن از دریاچه به سه بخش اصلی شامل ایستگاه‌های جنوبی، شرقی و غربی تفکیک شده‌اند. چنانچه که در اشکال الگوی مکانی نیز مشاهده شد، وجود دریاچه به‌عنوان یک عامل تأثیرگذار در مرکز حوضه موجب شده است که ایستگاه‌ها در اطراف دریاچه به چهار

www.SID.ir

Archive of SID

دسته ایستگاه‌های شمالی، جنوبی، غربی و شرقی تقسیم‌بندی شوند. نتایج حاکی از آن است که در خوشه‌بندی، نوع ویژگی‌ها و ترکیب‌بندی آنها در گروه‌های متفاوت، نقش به‌سزایی را در تشکیل ناحیه‌ها از لحاظ یکپارچگی و افزایش همگنی از لحاظ مکانی در پی خواهد داشت. در هر سه سناریوی وزن‌دهی مشاهده شد که دسته ویژگی‌های جغرافیایی و ترکیب این دسته ویژگی‌ها با سایر ویژگی‌ها (آماری و اقلیمی) می‌توانند الگوهای مکانی منظم و یکپارچه‌ای را در هر خوشه (ناحیه) تشکیل دهند. همچنین مشخص شد که وزن‌دهی ویژگی‌ها در خوشه‌بندی می‌تواند تا حد زیادی از میزان ناهمگنی‌ها و پراکندگی مکانی ایستگاه‌ها در هر یک از خوشه‌ها و نیز در هر یک از دسته ویژگی‌های به‌کار برده شده، بکاهد. نتایج نشان داد که بر اساس سناریوی دوم وزن‌دهی (w=PCAave)، ایستگاه‌هایی که در یک خوشه قرار گرفته‌اند، نسبت به سناریوی وزن‌های برابر (w = ۱)، الگوی مکانی همگن‌تر و قابل توجیه‌تری را از خود نشان داده‌اند و خوشه‌ها با تفکیک بهتری مشاهده می‌شوند. همچنین، همین روند مشابه خوشه‌بندی در سناریوی سوم وزن‌دهی (w=PCAmax) نیز مشاهده شد. به‌طورکلی با توجه به نتایج به‌دست آمده، عملکرد سناریوی دوم و سوم وزن‌دهی در تفکیک حوضه از لحاظ الگوی مکانی بسیار بهتر و چشمگیرتر از سناریوی اول بوده است و عملکرد سناریوی سوم نیز نسبتاً بهتر از سناریوی دوم به‌نظر می‌رسد. این نتایج حاکی از آن است که با درنظر گرفتن وزن‌های مناسب برای هر یک از ویژگی‌ها، بهبود نتایج ناحیه‌بندی در منطقه مورد مطالعه را می‌توان انتظار داشت.

منابع مورد استفاده

۱. پور اصغر. ف.، س. جهانبخش، ب. ساری صراف، ه. قائمی و م. تدینی. ۱۳۹۲. پهنه‌بندی رژیم بارش در نیمه جنوبی ایران. *مجله جغرافیا و برنامه‌ریزی* ۴۴ (۱۷). ۴۶-۲۷.

۲. جهانبخش اصل، س. و ح. ذوالفقاری. ۱۳۸۱. بررسی الگوهای سینوپتیک بارش‌های روزانه در غرب ایران. *فصلنامه تحقیقات جغرافیایی* ۶۴. ۲۵۸-۲۳۴.

۳. خام‌چین‌مقدم، ف.، ح. صدقی، ف، کاوه و م. منشوری. ۱۳۸۹. پهنه‌بندی حداکثر بارش روزانه ایران. *نشریه آب و خاک (علوم و صنایع کشاورزی)* ۲۴: ۱۰۶-۹۱.

۴. رضیئی. ط. ۱۳۹۵. شناسایی مناطق همگن بارشی ایران با استفاده از روش تحلیل مؤلفه‌های اصلی. *مجله ژئوفیزیک ایران* ۱۰(۳): ۱۴۴-۱۲۸.

۵. زارع چاهوکی، ا.، ع. سلاجقه، م. مهدوی، ش. سیگارودی و س. اسدی. ۱۳۹۲. مدل منطقه‌ای منحنی تداوم جریان حوزه‌های آبخیز بدون آمار مناطق خشک (مطالعه موردی: ایران مرکزی). *مرتع و آبخیزداری (منابع طبیعی ایران)* ۶۶ (۲): ۲۶۵-۲۵۱.

۶. شیروانی، ا. و س. م. ج. ناظم السادات. ۱۳۹۱. پهنه بندی بارش در ایران با استفاده از تحلیل مؤلفه‌های اصلی و تحلیل خوشه‌ای. *مجله تحقیقات منابع آب ایران* ۸(۱): ۸۵-۸۱.

۷. علیجانی، ب. ۱۳۷۴. *آب وهوای ایران*. انتشارات دانشگاه پیام نور. ایران.

۸. فیضی ح. و س. اسلامیان. ۱۳۸۴. مقایسه روش‌های ایستگاهی و منطقه‌ای گشتاورهای خطی در برآورد بارندگی‌های حداکثر ماهانه حوضه زاینده رود. *مجله آب و فاضلاب* ۱۶(۲): ۶۴-۵۴.

۹. مدرس، ر. ۱۳۸۵. توابع توزیع منطقه ای بارش در ایران. *مجله منابع طبیعی* ۷۵: ۹۱-۸۶.

۱۰. مسعودیان، س. ا. ۱۳۸۴. شناسایی رژیم‌های بارش در ایران به روش تحلیل خوشه‌ای. *مجله علوم انسانی دانشگاه اصفهان* ۱(۱۸): ۱۳-۱.

www.SID.ir

Archive of SID

11. Bulygina N., N. McIntyre and W. Howard. 2011. Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments. and under land use changes. *Journal of Water Resource Research* 47(2):1-13.
12. Darand M. and M. R. Mansouri Daneshvar. 2014. Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Journal of Environmental Processess* 1: 517–532.
13. Dinpashoh, Y., A. Fakheri-Fard, M. Moghaddam, S. Jahanbakhsh and M. Mirnia. 2004. Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology* 297: 109–123
14. Fazel, N., R. Berndtsson, C. Bertacchi Uvo, K. Madani and B. Kløve. 2017. Regionalization of precipitation characteristics in Iran's lake urmia basin. *Theoretical Applied Climatology Journal* 132(1-2): 363-373.
15. Heuvelmans, G., B. Muys and J. Feyen, 2006. Regionalization of the parameters of a hydrological model. Comparison of linear regression models with artificial neural nets. *Journal of Hydrology* 319(1–4): 245–265.
16. Hosking, J. R. M. and J. R. Wallis. 1997. Regional frequency analysis: an approach based on L-moments, Cambridge University, England.
17. Hosking, J. R. and J. R. Wallis. 1991. Some Statistics Useful in Regional Frequency Analysis. Res. Rep. RC 17096, IBM Research Division, Yorktown Heights, New York.
18. Hotelling H. H. 1933. Analysis of complex statistical variables into principal components. *Journal of Educational Psychology* 24: 417-520
19. Hutcheson, G. and S. Nick. 1999. The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models, Sage Publications, Thousand Oaks, CA.
20. Jiang, Y., H. Guoa, Y. Jia, Y. Cao and C. Hu. 2014. Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia. *Journal of Chemie der Erde-Geochemistry* 75(2):197–205.
21. Jollife, I. T. 2002. Principal Component Analysis. Part 2nd ed. Springer, Verlag, New York.
22. Johnson, R. A. and D. W. Wichern. 1982. Applied Multivariate Statistical Analysis, 3rd Ed, Prentice-Hall Inc., Englewood Cliffs, SA.
23. Lima, C. H. R. and U. Lall. 2010. Spatial scaling in a changing climate. A hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *Journal of Hydrology* 383(3–4): 307–318.
24. Merz, B. and G. Bloeschl 2004. Regionalization of catchment model parameters. *Journal of Hydrology* 287: 95–123.
25. Modarres, R. and A. Sarhadi. 2011. Statistically-based regionalization of rainfall climates of Iran. *Global and Planetary Change* 75: 67–75.
26. Olden, J. D., M. J. Kennard and B. J. Pusey. 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Journal of Ecohydrology* 5:503–518.
27. Post, D. A. 2009. Regionalizing rainfall–runoff model parameters to predict the daily streamflow of ungauged catchments in the dry tropics. *Journal of Hydrology Research* 40(5): 433–444.
28. Razavi, T. and P. Coulibaly. 2013. Classification of Ontario watersheds based on physical attributes and streamflow series. *Journal of Hydrology* 493: 81–94.
29. Samaniego, L., A. Bardossy and R. Kumar. 2010. Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. *Journal of Water Resource Research* 46(2): 1-22.
30. Samuel, J., P. Coulibaly and R. A. Metcalfe 2011. Estimation of continuous streamflow in Ontario ungauged basins. Comparison of regionalization methods. *Journal of Hydrology Engineering* 16(5):447–459.
31. Ward, J. and H. Joe. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301): 236–244.
32. Wazneh, H., F. Chebana and T. B. M. J. Ouarda. 2013. Depth-based regional index-flood model. *Journal of Water Resources Research* 49(12): 7957–7972.
33. Zhang Z., T. Wagener, P. Reed and R. Bhushan. 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. *Journal of Water Resource Reserch* 44(12).

www.SID.ir

Archive of SID

Journal of Water and Soil Science (Science and Technology of Agriculture and Natural Resources)
Vol. 22, No. 4, Winter 2019, Isfahan University of Technology, Isfahan, Iran.

Using the Principal Component Analysis Approach for Weighting Statistical, Climatic and Geographical Attributes of the Maximum 24-hour Rainfall and Spatial Clustering Analysis (A Case Study: Urmia Lake Basin)

Z. Dehghan[1*], S. S. Eslamian[1] and R. Modarres[2]

(Received: March 4-2017 ; Accepted: October 22-2017)

Abstract

Regionalization is one of the useful tools for carrying out effective analyses in regions lacking data or with having only incomplete data. One of the regionalization methods widely used in the hydrological studies is the clustering approach. Moreover, another effective factor on clustering is the degree of importance and participation level for each of these attributes. In this study, it was tried to use a broad range of attributes to compare their performance in regionalization. Then, according to the importance and role of each attribute in regionalization, the appropriate weight for each of the attributes in each category was determined using the principal component analysis (PCA) method, and the effect of this weighting in forming the homogenous regions was investigated by the Ward's clustering method. In this regard, the maximum 24-hour rainfall data of 63 meteorological stations located in Urmia Lake Basin (ULB) was used in this study during a time period of 30 years (1979-2008). Furthermore, seven categories of attributes were defined in order to regionalize the rainfall. The results showed that by considering different attributes and combining them with each other, a different clustering is obtained in each category in terms of the number of clusters and stations. Among seven categories of attributes, it was found that the geographical and climatic-geographical categories of attributes showed a more appropriate clustering over the ULB. Additionally, the weighting of attributes could have more effect on improving homogeneity and forming the independent clusters in most cases in terms of the scattering of station and how to locate over the basin.

Keywords: Principle component analysis, Urmia Lake basin, Ward's clustering method, Attributes' weighting

1. Department of Water Engineering, Faculty of Agriculture, Isfahan University of Technology, Isfahan, Iran.
2. Department of Rangeland and Watershed Management, College of Natural Resources, Isfahan University of Technology, Isfahan, Iran.

*: Corresponding Author, Email: z.dehghan@ag.iut.ac.ir

www.SID.ir