خوشه بندی سیگنال‌های هواشناسی با توجه به تغییرات بارش در استان سیستان و بلوچستان

بنفشه زهرایی، استادیار دانشکده مهندسی عمران، دانشگاه تهران، bzahraie@ut.ac.ir
عباس روزبهانی، دانشجوی کارشناسی ارشد مهندسی آب، دانشکده مهندسی عمران، دانشگاه تهران، roozbahany@ut.ac.ir
تلفن : ۶۶۴۹۳۰۴۶-۰۲۱ نمابر : ۶۶۴۸۰۲۹۰- ۰۲۱ پست الکترونیکی: bzahraie@ut.ac.ir

چکیده

امروزه لزوم مطالعه و بررسی تغییرات آب و هوایی و شناخت رفتار متغیرهای مختلف هواشناسی مثل بارش، دما و فشار هوا در نقاط مختلف جهان بخصوص در کشورهایی که با تنوع آب و هوایی گوناگون و وقوع دوره‌های خشک و تر شدید مواجه هستند، اهمیت بسیار زیادی دارد. در منطقه جنوب شرق ایران و بخصوص استان سیستان و بلوچستان، ستادهای بحران سیل و خشکسالی بطور همزمان فعال هستند و این منطقه از جمله بخش هایی از کشور است که دارای تنوع آب و هوایی شدید و تغییرات بسیار قابل ملاحظه در سالهای مختلف است. در این مقاله، روشی نوین برای خوشه بندی اطلاعات ماهواره‌ای دمای سطح دریا (SST) در مناطق موثر بر آب و هوای استان سیستان و بلوچستان شامل دریای عمان، اقیانوس هند و دریای عرب، با استفاده از روش الگوریتم ژنتیک (GA-Clustering) ارائه شده است. مراحل مختلف این تحقیق شامل انتخاب ساختار مدل، تابع هدف، طول کروموزوم‌ها و عملگرهای تزویج و جهش می‌باشد. در تدوین الگوی خوشه بندی علاوه بر رفتار زمانی و مکانی اطلاعات سیگنال‌های هواشناسی، ارتباط این متغیرها با وقوع بارندگی‌های کمتر و بیشتر از نرمال، با استفاده از اطلاعات بارش در ایستگاه‌های مختلف استان سیستان و بلوچستان در نظر گرفته شده است. استفاده از روش الگوریتم ژنتیک جهت خوشه بندی سیگنالهای هواشناسی و ارائه الگوریتمی پویا برای بیان ارتباط بارش منطقه مورد مطالعه در زمان‌های مختلف با الگوهای خوشه بندی زمانی-مکانی، از جمله نو آوری‌های این تحقیق محسوب می‌شود. نتایج این تحقیق نشان دهنده پتانسیل استفاده از نتایج خوشه بندی در تعیین رفتار دراز مدت بارش در منطقه جنوب شرق ایران می‌باشد.

کلید واژه‌ها : ***الگوریتم ژنتیک، خوشه بندی، سیگنال‌های هواشناسی، دمای سطح دریا (SST).***

Archive of SID
www.SID.ir

۱- مقدمه

در علوم هیدرولوژی و سایر زمینه‌های مرتبط با منابع آب، همواره لزوم توجه به مطالعات اقلیمی و بررسی تغییرات آب و هوا و شناخت سیگنال های هواشناسی مورد توجه بوده است. در حقیقت بررسی این تغییرات کمک شایانی به پیش بینی بارش و برنامه‌ریزی دراز مدت منابع آب هر منطقه خواهد نمود. در بین بحث های مربوط به اقلیم و پیش بینی متغیرهای هیدرولوژیکی از جمله بارش، بحث خوشه بندی سیگنال های هواشناسی و ارتباط آن با این متغیرها بخصوص با مسئله بارندگی اهمیت بسیار زیادی دارد. امروزه روشهای متعددی جهت خوشه بندی انواع متغیرها مثل روشهای معمول K-means، Map to map، Fuzzy c-means و غیره وجود دارد که از بین این روشها روش خوشه بندی با استفاده از الگوریتم ژنتیک روشی نوین و کارآمدی است و بوسیله آن می‌توان بهترین ساختار خوشه بندی را تعیین نمود. خوشه بندی سیگنال های هواشناسی مثل SST (دمای سطح دریا) و SLP (فشار سطح دریا) توسط روشهای ذکر شده در بالا صورت گرفته است اما تا کنون از الگوریتم ژنتیک برای خوشه بندی این متغیرهای هواشناسی استفاده نشده است و بیشتر کاربرد آن در مسائل اقتصادی و اجتماعی بوده است. توسعه روش خوشه بندی با استفاده از الگوریتم ژنتیک به اوایل دهه ۹۰ باز می‌گردد. Alippi & Cucchiara (1992) و Bandyopadhyay & Maulik (2002) تلاش‌های متعددی را در این مسیر صرف نمودند و روش مناسبی را برای خوشه بندی با الگوریتم ژنتیک با استفاده از ساختار اعداد واقعی ژن‌ها و با فرض نامشخص بودن تعداد خوشه‌ها ارائه دادند. روش نوین دیگری برای خوشه بندی خودکار توسط Chaudhuri & Garai (2004) ارائه شد. در سال Lin et al. (2005) یک روش موثرتر بر مبنای کد گذاری دو دویی برای خوشه بندی با تعداد نامشخص خوشه‌ها ارائه دادند.

در مقاله‌ای که پیش رو قرار دارد روشی ارائه شده است که علاوه بر نوین بودن آن در استفاده از الگوریتم ژنتیک در خوشه بندی سیگنال‌های هواشناسی، ساختار مدل به نحوی تدوین شده است که خوشه بندی متغیرهایی مثل SST با توجه به متغیر دیگری مانند بارندگی صورت گیرد و به نوعی ارتباط بین این دو متغیر درون مدل گنجانده شده است. این قابلیت کمک شایانی به پیش بینی و تقسیم بندی وضعیت بارش یک منطقه به سالهای کمتر از نرمال، نرمال و بیش از نرمال با در نظر گرفتن تغییرات سیگنال های هواشناسی در مناطق موثر بر بارش منطقه مورد نظر نموده است که جزئیات آن در بخش‌های بعدی ارئه شده است.

۲- ساختار مدل

الگوریتم ژنتیک یک روش نوین بهینه سازی است که ساختار آن شامل تابع هدف، کروموزوم ها، عملگرهای جهش و تزویج است. در مدل پیشنهادی ارائه شده در این مقاله طول کروموزم‌ها بر حسب تعداد نواحی انتخاب شده برای بررسی سیگنال‌های هواشناسی و همچنین تعداد خوشه ها، متفاوت است و برای این منظور از یک نوع VLGA (الگوریتم ژنتیک با کروموزوم با طول متغیر) استفاده شده است. در این مدل، تعداد نواحی مورد بررسی ۳ و تعداد حداکثر خوشه‌ها نیز ۳ در نظر گرفته شده و بنابراین طول کروموزم بالغ بر ۹ ژن خواهد بود که نمونه آن در شکل (۱) دیده می‌شود.

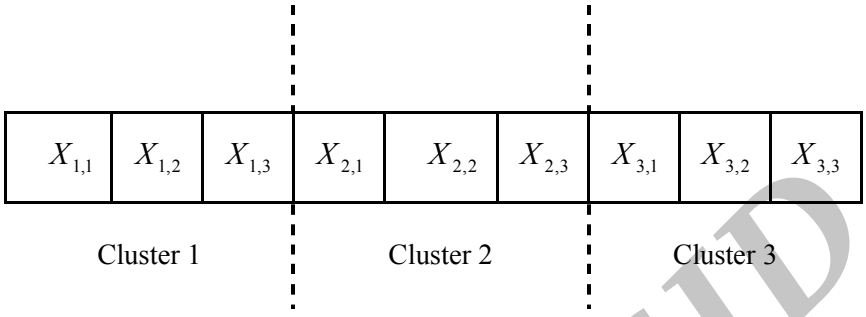
روش پیشنهادی بدین صورت است که به دلیل ساده تر کردن محاسبات، تنها از مراکز خوشه‌ها به عنوان اعداد واقعی داخل ژن‌ها استفاده می‌شود. همانطور که در شکل (۱) دیده می‌شود، $X_{i,j}$ معرف مقدار متغیر مورد نظر برای خوشه بندی در خوشه iام از ناحیه جغرافیایی jام است. در این تحقیق، متغیر مورد خوشه بندی دمای سطح دریا در سه ناحیه جغرافیایی مختلف

Archive of SID

www.SID.ir

می‌باشد که در بخش‌های بعدی مقاله شرح داده شده اند. روش کار بدین صورت است که در ابتدا با تعیین محدوده تغییرات دمای سطح آب دریا (SST) در هر ناحیه به تعداد اعداد ژنهای هر کروموزم، عدد تصادفی از بین مقادیر تاریخی مشاهده شده تولید می‌شود. به عنوان مثال برای شکل (۱) داریم:

$$SST_{\min,j} \le X_{i,j} \le SST_{\max,j} \qquad (i,j = 1,2,3) \qquad (۱)$$



شکل(۱) : *ساختار پیشنهادی برای کروموزوم نمونه در مدل GA برای خوشه بندی*

در رابطه فوق، $SST_{\max,j}$ و $SST_{\min,j}$ به ترتیب مقادیر مشاهده شده حداکثر و حداقل دمای سطح آب دریا در محدوده جغرافیایی j است. در گام بعدی با استفاده از جدول داده‌ها که شامل سری زمانی SST و بارش می‌باشد، به خوشه بندی سالها می‌پردازیم. در حقیقت هر سطر اطلاعات این جدول که مربوط به یک سال خاص می‌باشد، به صورت مقابل است :

$(SST_{1,t}, SST_{2,t}, SST_{3,t}, P_t)$: سال tام

$SST_{j,t}$ در عبارت فوق نشان دهنده مقدار SST مشاهده شده در منطقه جغرافیایی jام در سال t ام می‌باشد. P_t مقدار متغیر کمکی خوشه بندی می‌باشد که در این تحقیق میزان بارندگی در فصل مورد نظر در سال t ام می‌باشد. برای خوشه بندی، فاصله اقلیدسی SST هر سال از مختصات مراکز خوشه‌ها که قبلاً به صورت تصادفی انتخاب شده بودند، از رابطه (۲) بدست می‌آید:

$$d_{i,t} = \sqrt{\sum_{j=1}^{3} \left(SST_{t,j} - SST_{i,j}{}^{c} \right)^2} \qquad (۲)$$

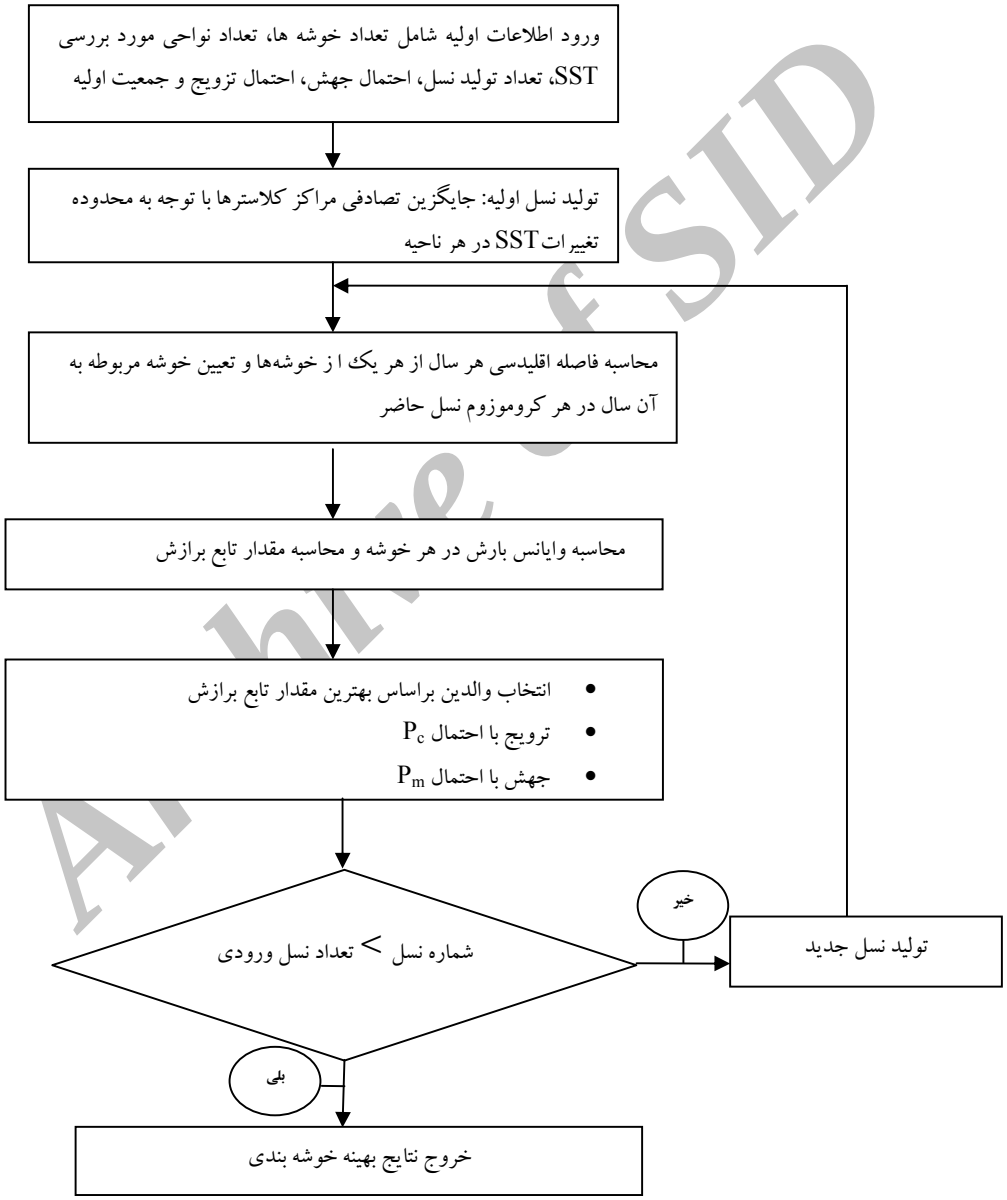
که در آن $d_{i,t}$ نشان دهنده فاصله اقلیدسی SST مشاهده شده در سال t ام از مرکز انتخاب شده برای خوشه i ام ، $SST_{t,j}$ دمای سطح آب مشاهده شده در سال t ام در ناحیه جغرافیایی jام و $SST_{i,j}{}^{c}$ دمای سطح آب مرکز خوشه i ام در ناحیه jام می‌باشد که به صورت تصادفی در GA تولید شده است.

این مقدار برای هر سال نسبت به هر یک از خوشه‌ها محاسبه می‌شود و کوچکترین مقدار این فاصله نشان دهنده تعلق سال مورد نظر به خوشه مورد نظر است. نکته مهم در این روش آن است که بارش منطقه مورد پیش بینی تاثیری در نحوه خوشه بندی ندارد بلکه تاثیر اصلی آن در محاسبه مقدار تابع هدف و انتخاب بهترین کروموزوم‌ها برای تولید نسل بعدی خواهد بود. پس از خوشه بندی سالها، روند محاسبه مقدار تابع هدف به صورت رابطه (۳) است:

$$\text{Min}\,[\text{Max}(\,Var(\underline{P})\,)\,] \qquad (۳)$$

Archive of SID
www.SID.ir

در این رابطه، Var نشان دهنده واریانس و $\bar{P}$ نشان دهنده مجموعه مقادیر بارندگی سال هایی است که در هر خوشه قرار گرفته اند. رابطه (۳) نشان دهنده انتخاب کروموزوم هایی به عنوان کروموزوم برتر است که بارندگی متناظر با سال‌های قرار گرفته در هر خوشه، دارای حداقل واریانس باشند. در نهایت در هر نسل، کروموزومی به عنوان کروموزوم برتر شناخته می‌شود که تابع هدف کوچکتری داشته باشد. لازم به تذکر است که با این روش تعداد حداکثر خوشه‌ها مشخص می‌شود و بنابراین در صورتی که حداکثر تعداد خوشه‌ها سه فرض شود پاسخهای مدل، خوشه بندی مناسب از یک تا سه خوشه را پیشنهاد خواهد نمود. برای ساده تر شدن درک نحوه کار مدل ، الگوریتم آن در شکل (۲) نشان داده شده است.



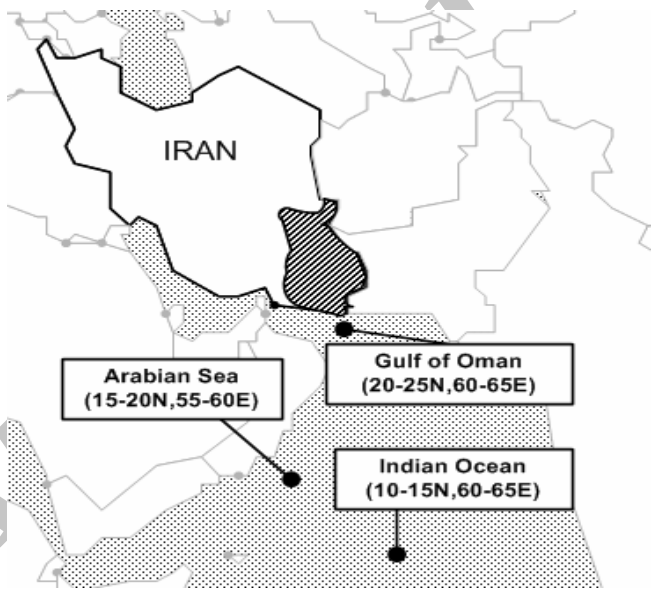
شکل(۲) : *الگوریتم مدل پیشنهادی برای خوشه بندی SST با استفاده از اطلاعات بارش*

www.SID.ir

۳- مطالعه موردی

برای بررسی کارآیی مدل خوشه بندی با استفاده از بارش، از اطلاعات بارندگی استان سیستان و بلوچستان استفاده شده است. استان سیستان و بلوچستان با مساحتی در حدود ۱۸۱۰۰۰ کیلومتر مربع و جمعیت ۱/۷ میلیون نفر، دارای میانگین سالانه ریزش جوی حدود ۱۹۹۵۴ میلیون متر مکعب می‌باشد. بر اساس تقسیم بندی وزارت نیرو تمامی یا بخشی از زیر حوزه‌های آبریز هیرمند، کویر لوت، جازموریان، ماشکید و دریای عمان در این استان قرار دارند. این استان دارای آب و هوای خشک با بارندگی ناچیز است و رژیم بارندگی آن اغلب به صورت رگبارهای سیل آسا است. به طور کلی مشکلات متعدد منابع آبی استان مانند نوسانات شدید منابع آبی، محدودیت منابع آب زیر زمینی، بارندگی اندک، بادهای موسمی مخرب و غیره اهمیت برنامه ریزی و مدیریت صحیح منابع آبی را در آن بیشتر نمایان می‌سازد.

بالغ بر ۷۰ ایستگاه باران سنجی توسط وزارت نیرو در این استان نصب شده است که در این تحقیق پس از بررسی ایستگاه های گوناگون، ۱۰ ایستگاه باران سنجی که دارای طول دوره آماری بیشتری بوده اند برای بررسی کارایی مدل انتخاب گردیده اند. همچنین برای بررسی تاثیر و ارتباط بین سیگنال های بزرگ مقیاس اقلیمی، سیگنال SST (دمای سطح آب دریا) در سه منطقه منتخب در دریای عمان، خلیج فارس و اقیانوس هند مورد بررسی واقع شده است که موقعیت این مناطق در شکل (۳) و همچنین مشخصات جغرافیایی این ۳ منطقه در جدول (۱) آورده شده است.



شکل(۳) : *مناطق تحت مطالعه برای در نظر گرفتن اطلاعات SST*

جدول (۱): *مشخصات جغرافیایی مناطق مورد بررسی برای در نظر گرفتن اطلاعات SST*

ردیف	نواحی جغرافیایی	عرض جغرافیایی	طول جغرافیایی
1	دریای عمان	20N-25N	60E-65E
2	دریای عرب	15N-20N	55E-60E
3	اقیانوس هند	10N-15N	60E-65E

www.SID.ir

پس از بررسی نقشه های SST در منطقه و اطلاعات بارش به این نتیجه رسیدیم که از اطلاعات بارندگی در ماههای دی، بهمن و اسفند که ماه‌های پربارش منطقه مورد مطالعه هستند در سال‌های ۱۳۶۱ تا ۱۳۸۱ و همچنین اطلاعات SST در ماه‌های October,September,August سال‌های ۱۹۸۲ تا ۲۰۰۱، جهت خوشه بندی استفاده شود. لازم به ذکر است که سعی شده انتخاب ماه‌ها به نحوی انجام شود که اطلاعات SST برای پیش بینی ماه‌های پر بارش منطقه مورد نظر قابل استفاده باشد. مدل مورد نظر به راحتی قابلیت تغییر دادن تعداد نواحی مورد بررسی از ۱ به ۳ ناحیه را داراست. لذا تاثیر SST هر یک از نواحی به تنهایی و با هم بر روی بارش این دو ایستگاه بررسی شده است که نتایج خوشه بندی آن مطابق شکل (۴) است. لازم بذکر است که تعداد نسل اولیه الگوریتم ژنتیک مورد نظر ۵۰ عدد و بهترین مقادیر احتمال جهش و تزویج به ترتیب ۰/۰۶ و ۰/۶ بر اساس تحلیل حساسیت انتخاب شده اند.

بررسی نتایج نشان داد که از بین ایستگاه های انتخاب شده، بارندگی ثبت شده در ایستگاه های سراوان و کجدار تاثیر بیشتری از تغییرات SST گرفته اند که نتایج آن در شکل (۴) نشان داده شده است. زمانی که تاثیر SST دریای عمان بر ایستگاه سراوان بررسی می شود، مدل خوشه بندی دو خوشه را انتخاب می کند که دسته اول شامل سال‌های ۶۱ و ۶۲ و ۶۴ و ۶۹ و ۷۱ و ۷۲ و ۷۴و ۷۶و ۷۷ با احتمال ۷۸٪ دارای بارندگی بیش از میانگین و مابقی سالها در دسته دوم و با احتمال ۸۰٪ دارای بارش کمتر از میانگین دراز مدت می‌باشند. همچنین خوشه بندی بر اساس SST دریای عرب نیز تحت بررسی قرار گرفت که همانطور که مشاهده می‌شود، سالهای ۶۱ و ۷۱ و ۷۲ و ۷۳ و ۷۴ در یک خوشه قرار گرفته اند که نشان دهنده بارش بیش از میانگین با احتمال ۸۰٪ می‌باشند.

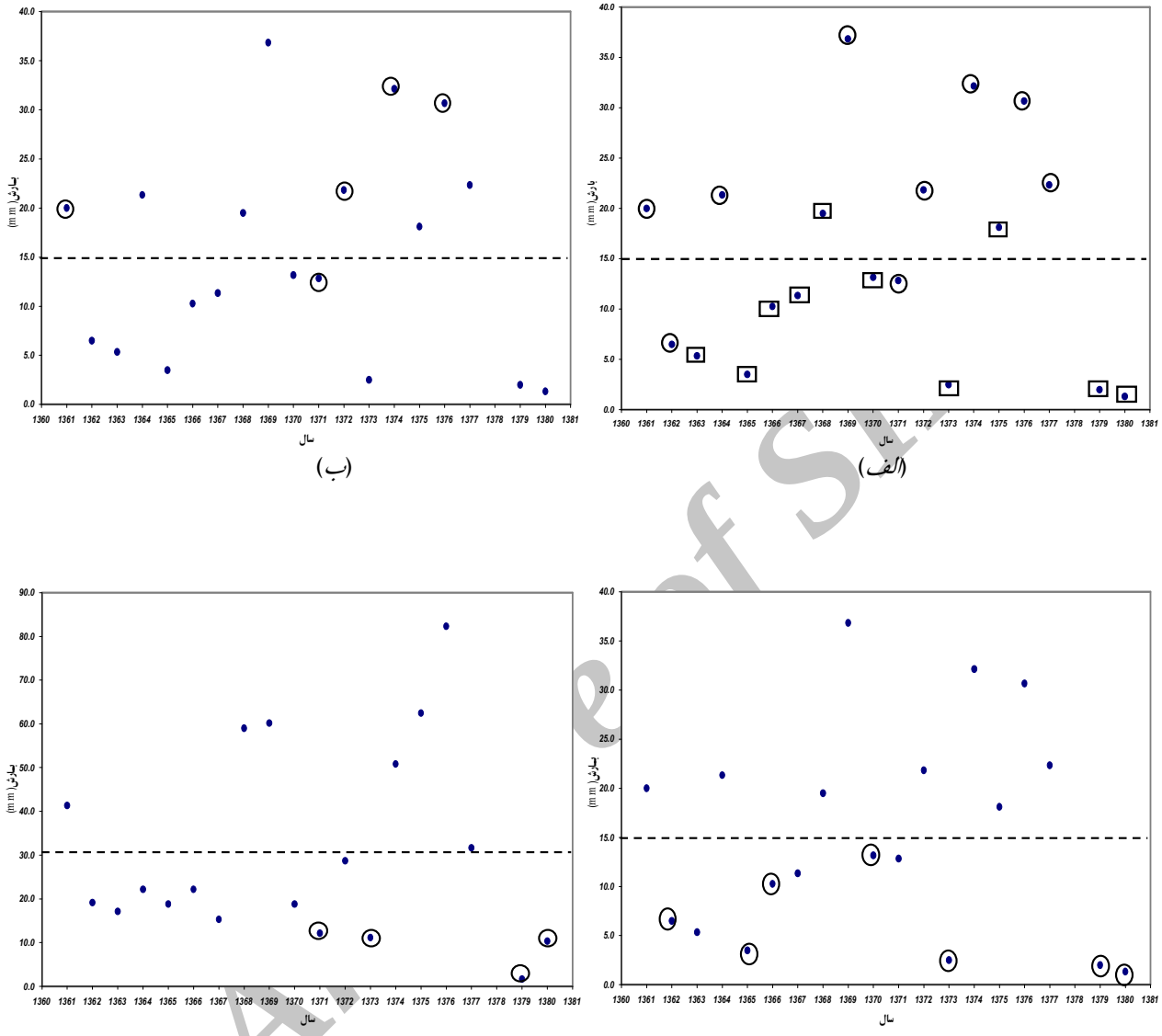
در حالتی که ۳ ناحیه با هم در نظر گرفته شده اند نیز سالهای ۶۲ و ۶۵ و ۶۶ و۷۰ و ۷۳ و ۷۹ و۸۰ در یک خوشه قرار دارند که همگی دارای بارش زیر میانگین و نشان دهنده شرایط کم آبی است. در مورد ایستگاه کجدار (شکل۴- د) نیز تنها تاثیر منطقه دریای عمان قابل توجه بوده و در آن سالهای ۷۱ و ۷۳ و ۷۸ و ۷۹ در یک خوشه واقع شده و همگی دارای بارش کمتر از میانگین هستند.

۴- جمع بندی و نتیجه گیری

همانطورکه ملاحظه شد روش ارائه شده نه تنها قابلیت خوشه بندی و جستجوی کامل فضای جوابها را ممکن می‌سازد، بلکه قادر خواهد بود طوری عمل خوشه بندی را صورت دهد که ارتباط SST سالهایی که در یک خوشه قرار می‌گیرند با بارش همان سال‌ها در منطقه مورد مطالعه حفظ گردد و بتوان از این طریق به سالهای با بارش کمتر از نرمال، نرمال و بیش از نرمال دست یافت. این نتیجه مهم فواید زیادی را در پی خواهد داشت از جمله این که با اندازه گیری و محاسبه دمای سطح آب می‌توان به پیش بینی فصلی بارش در مناطق تحت تاثیر قرار گرفته پرداخت. همچنین می‌توان در ادامه کار با دخالت دادن دیگر سیگنال های مهم هواشناسی همچون SLP (فشار سطح دریا) و همچنین مقایسه روش پیشنهادی با روش های مرسوم خوشه‌بندی مثل K-means به امتیازات روش پیشنهادی در این مقاله پی برد. همچنین با توجه به قابلیت متغیر بودن طول کروموزوم‌ها در مدل می‌توان به بررسی نواحی بیشتری در سطح اقیانوس هند و دریاهای اطراف منطقه مورد مطالعه بارش پرداخت و همچنین با توجه به محدوده تغییرات بارش در منطقه مورد نظر تعداد خوشه‌ها را اضافه یا کسر نمود. نتیجه این تحقیق کمک شایانی در پیش بینی دراز مدت اقلیمی و در نهایت برنامه ریزی سودمند منابع آبی کشور خواهد نمود.

Archive of SID

www.SID.ir



شکل (۴) : نتایج خوشه بندی SST با توجه به بارندگی ایستگاه‌های منتخب

الف) دریای عمان- ایستگاه سراوان، ب) دریای عرب- ایستگاه سراوان،

ج) دریای عمان و دریای عرب و اقیانوس هند- ایستگاه سراوان، د) دریای عمان- ایستگاه کجدار

(توضیح : خطوط خط چین نشان دهنده میانگین درازمدت بارندگی و نقاط مشخص شده با اشکال دایره و مستطیل، خوشه های برگزیده می‌باشند.)

www.SID.ir

۶- مراجع

[1] Alippi, C., and Cucchiara, R.(1992), "Cluster partitioning in image analysis and classification: A genetic algorithm approach", Proc.CompEuro 1992 on Comput. System Software Eng., IEEE Computer Society Press, Silver Spring, MD, pp. 139-144.

[2] Bandyopadhyay, S., and Maulik, U. (2002), "Genetic algorithm-based clustering technique", *Journal of Pattern Recognition Society*., Vol. 33, pp. 1455-1465.

[3] Garai, G., and Chaudhuri, B.B. (2004), "A novel genetic algorithm for automatic clustering", *Journal of Pattern Recognition Letters*, Vol. 25, pp. 173-187.

[4] Lin H., F. Yang, and Y. Ta Kao (2005), " An Efficient GA-based Clustering Technique", *Journal of Sience and Engineering*., 8(2), pp. 113-122.

Archive of SID

www.SID.ir