مجله‌ی علمی و پژوهشی شریف
دی ۱۳۸۸، شماره‌ی پنجاه و یکم، ص. ۱۹-۲۴

پیش‌بینی درازمدت بارش با استفاده از خوشه‌بندی سیگنال‌های هواشناسی با توجه به تغییرات بارش به‌روش K-Means اصلاح‌شده (مطالعه‌ی موردی: پیش‌بینی بارندگی استان سیستان و بلوچستان)

بنفشه زهرایی (دانشیار)
قطب علمی مهندسی و مدیریت زیرساخت‌ها، دانشگاه تهران
عباس روزبهانی (دانشجوی دکترا)
دانشکده مهندسی عمران، دانشگاه تهران

مطالعه‌ی تغییرات اقلیمی و شناسایی متغیرهای مختلف آب و هوایی (ازجمله بارش) کمک شایانی به پیش‌بینی این‌گونه متغیرها می‌کند. بیان ارتباط سیگنال‌های بزرگ مقیاس اقلیمی ـ مثل دمای سطح دریا (SST) ـ و متغیرهای مهم هیدرولوژیکی ـ مثل بارندگی ـ ازجمله تحقیقات صورت گرفته در سال‌های اخیر است. در این نوشتار روشی نوین با عنوان K-Means اصلاح شده برای خوشه‌بندی ارائه شده است. برای بررسی کارایی مدل پیشنهادی از مطالعه‌ی موردی بر روی استان سیستان و بلوچستان، واقع در منطقه‌ی جنوب شرق ایران، استفاده شده است. در این مطالعه از اطلاعات بارش ۲۰ ایستگاه مختلف این استان استفاده شده است و خوشه‌بندی سیگنال هواشناسی SST در مناطق جغرافیایی منتخب در دریای عمان، دریای عرب و اقیانوس هند با توجه به تغییرات بارش این ایستگاه‌ها صورت گرفته است. استفاده از روش K-Means اصلاح شده برای خوشه‌بندی سیگنال‌های هواشناسی و ارائه‌ی الگوریتمی پویا برای بیان ارتباط بارش منطقه مورد مطالعه در زمان‌های مختلف با الگوهای خوشه‌بندی زمان ـ مکان از نوآوری‌های این تحقیق محسوب می‌شود.

واژگان کلیدی: خوشه‌بندی، سیگنال‌های هواشناسی، K-Means اصلاح‌شده، خوشه‌بندی ژنتیک، دمای سطح دریا (SST).

bzahraie@ut.ac.ir
roozbahany@ut.ac.ir

۱. مقدمه

بحث پیش‌بینی متغیرهای هیدرولوژیکی مهم، نظیر بارش، همواره مورد توجه متخصصین علوم مهندسی منابع آب و هیدرولوژی بوده است. این پیش‌بینی‌ها شامل پیش‌بینی‌های بلندمدت و کوتاه‌مدت اقلیمی است. پیش‌بینی درازمدت بارندگی ازجمله مهم‌ترین اطلاعات مورد نیاز برای برنامه‌ریزی و مدیریت منابع آب در حوضه‌های آبریز است. تحقیقات اخیر نشان می‌دهد که موضوع یافتن ارتباط سیگنال‌های بزرگ‌مقیاس اقلیمی ـ همچون SST و SLP ـ با بارش و روان‌آب رو به افزایش است و در مقیاس‌های بزرگ زمانی و مکانی می‌توان با اتخاذ روش مناسب، به کشف ارتباط رفتار درازمدت این سیگنال‌ها با بارش یک منطقه در چند ماه آینده پرداخت. به غیر از روش‌های مرسوم مدل‌سازی آماری چندمتغیره، روش خوشه‌بندی سیگنال‌های هواشناسی ازجمله روش‌های سودمند به‌منظور یافتن این ارتباط است.

به‌طور کلی هدف تمام روش‌های خوشه‌بندی، دسته‌بندی اطلاعات و داده‌های در دسترس گروه‌های مشابه و استفاده از آنها برای مقاصد علمی دیگر است. روش‌هایی همچون K-Means، Fuzzy C-means، GA-Clustering، GKA[1]-Clustering و غیره، هرکدام دارای ویژگی‌ها و موارد استفاده‌ی مختلفی در مقیاس‌های زمانی و مکانی هستند و در علوم مختلف مورد استفاده قرار می‌گیرند. ازجمله کاربردهای مهم این روش‌ها در بحث‌های اقتصادی بوده، و تحقیقات متعددی درمورد کاربرد آن در اقتصاد انجام شده است.[۱] همچنین روش خوشه‌بندی فازی را با نگرش هم‌پوشانی خوشه‌های مختلف تعمیم داده‌اند.[۲] روش مؤثر و متداول K-Means توسط مک‌کوین ارائه شد و تا به امروز به‌تنهایی یا با تلفیق روش‌های دیگر در علوم مختلف مورد استفاده قرارگرفته است.[۳]

در سال ۱۹۹۹ روش GKA که تلفیقی از الگوریتم ژنتیک و K-Means است ارائه شد. پس از آن در سال ۲۰۰۵، روش خوشه‌بندی Ant K-Means با بهره‌گیری از الگوریتم مورچه و تلفیق آن با K-Means تدوین شد.[۵]

در سال ۲۰۰۵، از یک روش خوشه‌بندی چندمتغیره‌ی K-Means که در

تاریخ: دریافت ۱۳۸۶/۲/۳۰، داوری ۱۳۸۶/۹/۱۴، پذیرش ۱۳۸۶/۱۱/۲۴.

سال ۱۹۷۵ پیشنهاد شده بود به‌منظور خوشه‌بندی متغیرهای آب و هوایی بهره‌گیری شد.[۶] نتایج این روش شامل پیش‌بینی و مقایسه‌ی الگوهای مختلف آب و هوا در طی ۹۰ سال آینده بود که به‌نوبه‌ی خود یکی از بارزترین مطالعات انجام‌شده در استفاده از روش K-Means در بحث آب و هوا به شمار می‌رود.

در سال ۱۹۹۸ نشان داده شد که تغییرات دمای سطح آب درخلیج فارس تأثیر معنی‌داری بر تغییرات بارندگی در مناطق جنوب و جنوب غرب ایران دارد.[۷] این مطالعات نشان می‌دهد که بارش زمستانی (ژانویه تا مارچ) در مناطق مزبور با SST خلیج‌فارس رابطه‌ی معکوس دارد. یعنی زمانی که SST خلیج فارس بیشتر (کم‌تر) از مقدار زمان آن باشد، انتظار می‌رود که مناطق یادشده با خشک‌سالی (ترسالی) مواجه شوند. لازم به ذکر است که هواشناسان معتقدندکه با افزایش دما، قابلیت تزریق رطوبت به سامانه‌های عبوری از روی خلیج فارس بیشتر شده و احتمال بارندگی افزایش می‌یابد.

پیش‌تر، در سال ۲۰۰۱، همبستگی همراه با تأخیر بین SST و بارندگی مانسون تابستانیِ شرق چین مورد بررسی قرار گرفت.[۸] برای این کار از دمای سطحی مناطقی از اقیانوس هند، آرام و اطلس استفاده کرده‌اند. در سال ۲۰۰۴ نیز از طریق مدل رایانه‌یی REMO، که به مدل‌سازی وضعیت جوی اختصاص دارد، به بررسی زمانی ـ مکانی اثر سیگنال SST بر تداوم و شدت بارش‌های شدید در منطقه‌ی وسیعی از اروپا پرداخته شد و نتایج جالبی ازجمله تأثیر چشم‌گیر بالا رفتن مقدار SST در دریای بالتیک بر افزایش بارش‌های سنگین در مناطق اطراف به‌دست آمد.[۹]

در ایران در سال ۱۳۸۳، اثر تغییرات فشار سطح دریا (SLP) در نقاط مختلفِ اقیانوس اطلس، دریای مدیترانه، سیاه، خزر، عرب و خلیج فارس بر بارش حوزه‌ی آبریز کارون و دز مورد بررسی قرار گرفت.[۱۰] برای این منظور تعدادی قانون اولیه برای پیش‌بینی بارش استخراج شد و در ادامه برای تدقیق و کمی‌کردن رابطه‌ی بارش با شاخص‌های SLP نام برده شده از نظریه‌ی مجموعه‌های فازی استفاده شد. همچنین در پیش‌بینی روان‌آب از روش K - نزدیک‌ترین همسایه استفاده شد.

در سال ۱۳۸۵ نیز نگارندگان روش نوین دیگری برای به‌کارگیری الگوریتم ژنتیک در خوشه‌بندی سیگنال‌های هواشناسی و با توجه به تغییرات بارش ارائه کردند که در آن با مطالعه‌ی بارش منطقه‌ی سیستان و بلوچستان و SST دریاهای جنوب ایران، ساختارکروموزوم‌ها به‌گونه‌یی طراحی شده که در خوشه‌بندی SST، ارتباط آن با عمق متوسط فصلی بارش نیز در نظرگرفته شود؛ این مطالعه نتایج قابل توجهی برای کاربرد آن در پیش‌بینی بارش فصلی به‌همراه داشته است.[۱۱و۱۲] در این مقاله و در راستای مطالعات صورت‌گرفته‌ی قبلی، برای توسعه‌ی روش K-Means به‌منظور رسیدن به هدف مشابه تلاش فراوان کرده‌ایم. اساس کار روش پیشنهادی، تدوین مدلی بوده است که قادر به پیش‌بینی تغییرات درازمدت بارش براساس سیگنال‌های هواشناسی باشد. در ادامه، اشاره‌ی مختصری به روش معمول K-Meansدر خوشه‌بندی می‌شود و سپس روش K-Means اصلاح‌شده به تفصیل مورد بررسی قرار می‌گیرد.

۲. تعریف مسئله

روش K-Means روشی متداول برای خوشه‌بندی است که ساختار نسبتاً ساده و مؤثری دارد. به‌طور خلاصه مراحل این روش به‌ترتیب عبارت‌اند از:

۱. تعیین تعداد خوشه‌های مورد نظر (K) و انتخاب تصادفی K مرکز خوشه از بین داده‌های موجود؛

۲. محاسبه‌ی فاصله‌ی اقلیدسی داده‌های موجود از مراکز انتخاب شده؛

۳. تعیین نزدیک‌ترین مرکز خوشه به داده‌های موجود و خوشه‌بندی اولیه؛

۴. محاسبه‌ی میانگین داده‌های قرارگرفته در هر خوشه و جایگزینی این میانگین با مرکز قبلی خوشه؛

۵. تکرار مراحل ۲ تا ۴ تا رسیدن به هم‌گرایی در جواب‌ها.

در روشِ یادشده امکان خوشه‌بندی با دانستن تعداد اولیه‌ی خوشه‌ها فراهم می‌شود. اما در حالت خاص، ما به دنبال راهی هستیم که بتوانیم هم‌زمان با خوشه‌بندی یک یا چند متغیر، ارتباط این متغیرها را با متغیر دیگری ـ نظیر بارش ـ بررسی کنیم. درحقیقت هدف این تحقیق تدوین الگوریتمی برای خوشه‌بندی براساس نظریه‌ی K-Means است که قابل استفاده در کمی‌کردن ارتباط دو متغیر مختلف باشد. این دو متغیر در این تحقیق عبارت‌اند از: SST به‌عنوان متغیر مورد خوشه‌بندی، و بارش به‌عنوان متغیر مرتبط در تعیین خوشه‌ها. درحقیقت برای خوشه‌بندی متغیری همچون SST الگوریتم معمولی K-Means به‌تنهایی قابل استفاده است، اما زمانی که بخواهیم تأثیرات متغیر دیگری مانند بارش را نیز در این خوشه‌بندی دخالت دهیم، استفاده از روش متداول دیگر پاسخ‌گو نخواهد بود. لذا برای از بین بردن این نقص به ارائه‌ی روش خوشه‌بندی K-Means اصلاح‌شده می‌پردازیم. روش پیشنهادی بدین‌صورت است که در ابتدا مانند روش مرسوم تعداد K دسته عدد از بین داده‌های SST انتخاب می‌شود. منظور از دسته عدد این است که در صورتی که تعداد مناطق جغرافیایی که تغییرات SST در آنها مورد بررسی قرار می‌گیرد، بیش از یک منطقه باشد، مجموعه‌یی از متغیرهای SST در مناطق مذکور مورد استفاده قرار می‌گیرد. درواقع هر سریِ زمانی SST ورودی به مدل مورد نظر در حالت کلی به شکل زیر است:

$$(SST_{1,t},\ SST_{2,t},\ \ldots,\ SST_{j,t}) \qquad (۱)$$

که در آن j معرف تعداد کل مناطق مورد نظر برای بررسی تغییرات SST است. در گام بعدی، فاصله‌ی SST هر سال از مرکز انتخاب شده به‌وسیله‌ی رابطه‌ی ۲ محاسبه می‌شود:

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

که در آن $d_{i,t}$نشان‌دهنده‌ی فاصله‌ی اقلیدسی SST مشاهده‌شده در سال tام از مرکز انتخاب‌شده برای خوشه‌ی iام، $SST_{t,j}$ دمای سطح آب مشاهده شده در سال tام در ناحیه‌ی جغرافیایی jام و $SST^{c}_{i,j}$ دمای سطح آب مرکز خوشه‌ی iام در ناحیه‌ی jام است که به‌صورت تصادفی از بین داده‌های موجود انتخاب شده است. بدین‌ترتیب همانند K-meansمعمولی، خوشه‌بندی اولیه با توجه به کم‌ترین فاصله‌ی هر سال از K مرکز منتخب انجام می‌شود. تا این مرحله بارش در خوشه‌بندی تأثیری ندارد، اما از این مرحله به بعد مقدار بارش تأثیرگذار خواهد بود. قبل از شروع مرحله بعدی، میانگین درازمدت عمق متوسط بارش فصلی منطقه‌ی تحت مطالعه (P) محاسبه می‌شود و تعداد سال‌های با بارش بیشتر (NP_{wet}) و کم‌تر از میانگین (NP_{dry}) مشخص می‌شوند:

$$NP_{wet} = \left\{ P_i \middle| P_i > \overline{P} \right\}, \qquad NP_{dry} = \left\{ P_i \middle| P_i < \overline{P} \right\} \qquad (۳)$$

اینک تک‌تک خوشه‌های اولیه را تحت بررسی قرار می‌دهیم. ابتدا بارش متناظر با سال‌های قرارگرفته در هر خوشه را در نظرگرفته، تعداد سال‌های با بارش بیشتر و

کم‌تر از میانگین درازمدت بارش را محاسبه می‌کنیم و آنها را $NP_{wet,i}$ و $NP_{dry,i}$ می‌نامیم. در ادامه نسبت‌های $\frac{NP_{wet,i}}{NP_{wet}}$ و $\frac{NP_{dry,i}}{NP_{dry}}$ را تعیین می‌کنیم. بیشترین مقدار این دو نسبت، مبنای انتخاب مرکز جدید قرار می‌گیرد:

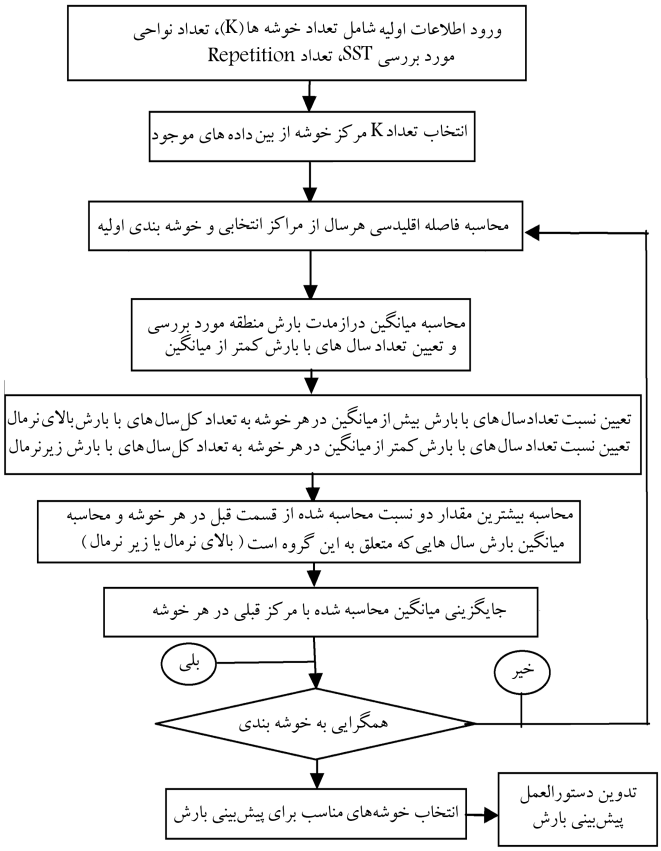
$$MAX\left\{\frac{NP_{wet,i}}{NP_{wet}}, \frac{NP_{dry,j}}{NP_{dry}}\right\}$$

با تعیین مقدار میانگین SST، سال‌هایی که بارش در آنها در گروه منتخب فوق قرار می‌گیرد محاسبه، و به‌عنوان مرکز جدید خوشه محاسبه می‌شود:

$$SST^{c}_{i,new} = average\left\{SST \middle| SST_i \in MAX \left\{\frac{NP_{wet,i}}{NP_{wet}}, \frac{NP_{dry,j}}{NP_{dry}}\right\}\right\} \quad (۴)$$

لازم به ذکر است که میانگین جدید شامل چند بعد است که هر بعد، میانگین SST در منطقه‌ی جغرافیایی منتخب است. با استفاده از رابطه‌ی ۴ سعی شده تا در تعیین مراکز خوشه‌های جدید در فرآیند K-Means سال‌هایی که متناظر با وقوع بارش‌های بیشتر یا کم‌تر از میانگین بوده‌اند، مورد توجه قرار گیرد. بدیهی است که در صورت بررسی یک ناحیه‌ی جغرافیایی، میانگین تنها یک عدد است.

در گام بعدی، فاصله‌ی اقلیدسی هر سال از میانگین جدید محاسبه می‌شود و مجدداً خوشه‌بندی انجام می‌گیرد. در واقع در هر مرحله از محاسبات مراکز هر خوشه با روش ذکرشده محاسبه شده و فواصل تمام داده‌ها از این مراکز محاسبه، و



شکل ۱. الگوریتم مدل پیشنهادی K-Means برای خوشه‌بندی SST با استفاده از اطلاعات بارش.

ترکیب جدید خوشه‌بندی حاصل می‌شود؛ در ادامه روش K-Means معمولی تا همگراشدن به یک ترکیب خوشه‌بندی ثابت ادامه می‌یابد. شکل ۱ الگوریتم روش پیشنهادی را نشان می‌دهد. لازم به ذکر است که منظور از خوشه‌بندی ثابت، همگرایی به ترکیبی از خوشه‌بندی است که با افزایش تعداد مراحل اجرای مدل، این ترکیب دچار تغییر نشود و مقادیر میانگین هر خوشه تغییر نکند. درنتیجه این ترکیب به‌عنوان خوشه‌بندی نهایی پذیرفته می‌شود.

ازجمله مشکلاتی که در روش K-Means وجود دارد الزام آن برای مشخص‌بودن تعداد خوشه‌ها از ابتدای حل است. در این مقاله مبنای مقایسه‌ی عملکردِ روش پیشنهادی مطالعه‌ی انجام‌شده‌ی قبلی، یعنی خوشه‌بندی الگوریم ژنتیک است. مزیت این روش نه‌تنها برخورداری از یک ساختار جست‌وجوی هوشمند است، بلکه چنانچه فرضاً تعداد خوشه‌های اولیه در این روش ۳ باشد این الگوریتم از بین ۱ تا ۳ خوشه بهترین تعداد خوشه را انتخاب می‌کند؛ چگونگی این مقایسه در بخش نتایج مدل ارائه می‌شود.

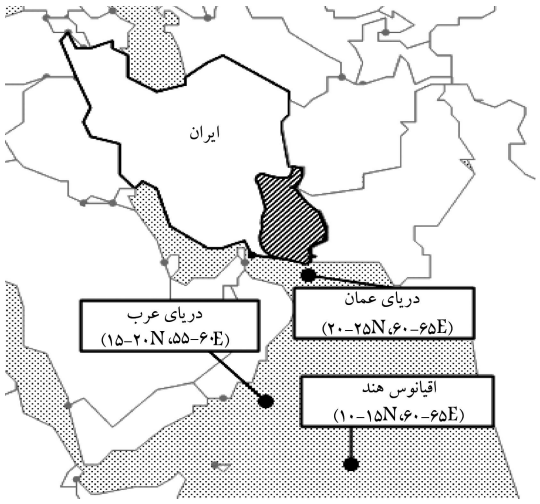
۳. مطالعه‌ی موردی

برای بررسی کارآیی مدل خوشه‌بندی سیگنال‌های هواشناسی 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 سالانه‌ی ریزش جوی حدود ۱۹۹۵۴ میلیون متر مکعب است. براساس تقسیم‌بندی وزارت نیرو تمامی یا بخشی از زیرحوزه‌های 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 به‌صورت رگبارهای سیل‌آسا است. کمبود نزولات و پراکندگی و نامشخص‌بودن زمان بارش از یک‌سو، و بالابودن سطح میزان تبخیر از سوی دیگر باعث شده تا منابع آب در دسترس با کمبود شدیدی مواجه شود. به‌طور کلی مشکلات متعدد منابع آبی استان مانند نوسانات شدید منابع آبی، محدودیت منابع آب زیرزمینی، بارندگی اندک، بادهای موسمی مخرب، پدیده‌ی مانسون و غیره اهمیت پیش‌بینی بارش به‌منظور برنامه‌ریزی و مدیریت صحیح منابع آبی در آن را بیشتر نمایان می‌سازد. بالغ بر ۷۰ ایستگاه باران‌سنجی توسط وزارت نیرو در این استان نصب شده که در این تحقیق پس از بررسی ایستگاه‌های گوناگون، ۲۰ ایستگاه باران‌سنجی برای بررسی کارایی مدل انتخاب شده‌اند. ۱۰ ایستگاه از این تعداد، ایستگاه‌های شاخص باران‌سنجی استان سیستان و بلوچستان هستند که توسط شرکت سهامی آب منطقه‌یی استان سیستان و بلوچستان معرفی شده‌اند و عبارت‌اند از: ایستگاه‌های چابهار، ایرانشهر، سراوان، زاهدان، زابل، لادیز، نصرت‌آباد، خاش، قصر قند و پیشین که برخی از این ایستگاه‌ها در برخی سال‌ها مشکل کمبود اطلاعات دارند. همچنین تعداد ۱۰ ایستگاه باران‌سنجی دیگر نیز که دارای طول دوره‌ی آماری بیشتری بوده‌اند مورد تحلیل قرار گرفتند. علاوه بر این برای بررسی تأثیر و ارتباط بین سیگنال‌های بزرگ مقیاس اقلیمی، نقشه‌های تغییرات سیگنال SST(دمای سطح آب دریا) در دریاهای اطراف منطقه در ماه‌های مختلف سال مورد بررسی قرار گرفت و در نهایت سه منطقه‌ی منتخب در دریای عمان، خلیج فارس و اقیانوس هند به‌منظور بررسی و خوشه‌بندی SST انتخاب شد؛ موقعیت این مناطق در شکل ۲ نشان داده شده است. لازم به ذکر است که اطلاعات این سیگنال از بانک اطلاعات مرکز تحقیقات ملی اتمسفری کشور آمریکا (NCAR[۲]) اخذ شده است.

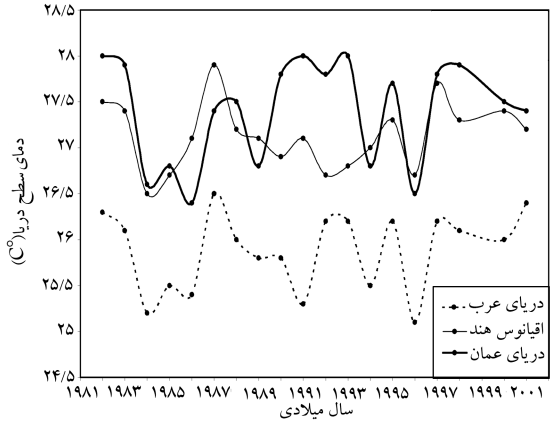
همچنین پس از بررسی نقشه‌های SST در منطقه و اطلاعات بارش به این نتیجه رسیدیم که برای ورودی‌های مدل از اطلاعات عمق متوسط بارندگی در فصل زمستان که فصل پربارش در سال آبی منطقه‌ی مورد مطالعه هستند (در سال‌های ۱۳۶۱ تا ۱۳۸۱) و همچنین از اطلاعات SST در ماه‌های آگوست، سپتامبر و اکتبر (مصادف با ماه‌های مرداد، شهریور و مهر) سال‌های ۱۹۸۲ تا ۲۰۰۱ برای خوشه‌بندی استفاده شود. لازم به ذکر است که سعی شده انتخاب ماه‌ها به‌نحوی انجام شود که اطلاعات SST برای پیش‌بینی ماه‌های پربارش منطقه‌ی مورد نظر قابل استفاده باشد تا بتوان درمورد وضعیت آبی منطقه از لحاظ کم‌بارش یا پربارش بودن با تأخیر زمانی مشخص تصمیم‌گیری کرد. شکل ۳ سری زمانی دمای سطح دریا در این ۳ منطقه را نشان می‌دهد.

لازم به ذکر است که مدل مورد نظر به‌راحتی قابلیت تغییر دادن تعداد نواحی مورد بررسی از ۱ به ۳ ناحیه را دارد. لذا تأثیر SST هریک از نواحی به‌تنهایی، دو به دو و با هم بر روی بارش ایستگاه‌های مورد نظر بررسی شده، و در نهایت نتایج قابل توجهی درمورد برخی ایستگاه‌ها به دست آمده است. درمورد سایر ایستگاه‌ها نتایج نشان‌گر عدم همبستگی مناسب بین بارش این ایستگاه‌ها و SST مناطق جغرافیایی منتخب است.

همچنین درمورد ایستگاه‌هایی که نتایج قابل توجه بوده است نیز تعداد نواحی که SST آنها بر بارش مؤثر است، از ۱ تا ۳ منطقه‌ی جغرافیایی متغیر است.



شکل ۲. مناطق تحت مطالعه برای در نظرگرفتن اطلاعات SST.

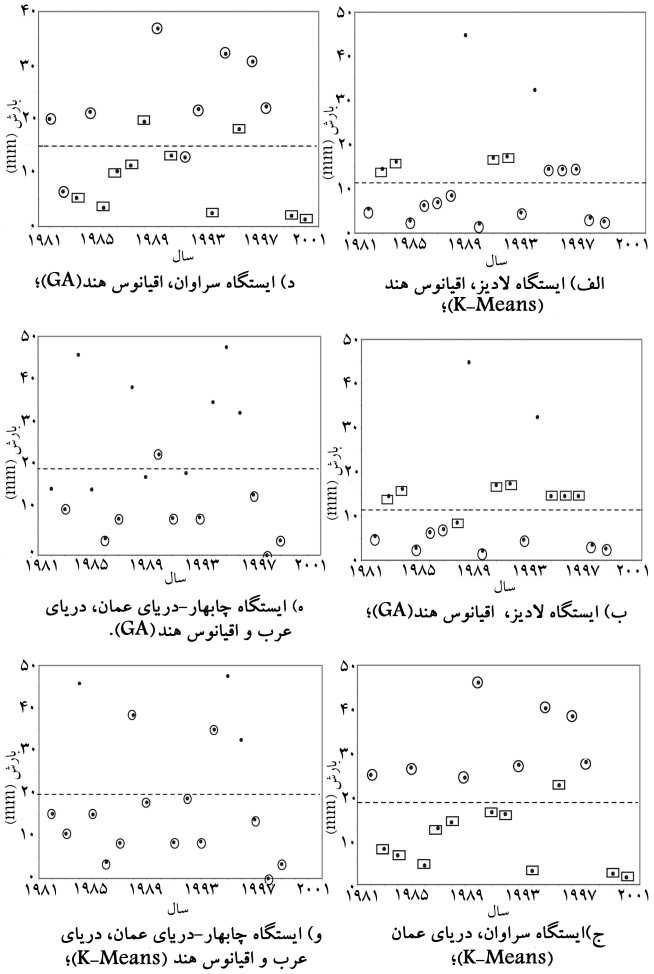


شکل ۳. متوسط دمای سطح دریا در ماه‌های آگوست، سپتامبر و اکتبر.

۴. نتایج حاصله و تحلیل آن

مدل خوشه‌بندی K-Means پیشنهادی برای حالات مختلف ۱ تا ۳ خوشه و ۱ تا ۳ منطقه‌ی جغرافیایی برای در نظرگرفتن SST و بارش یک منطقه‌ی جغرافیایی اجرا شد. اکثر مدل‌ها در زمانی کم‌تر از چند ثانیه و درکم‌تر از ۱۰ مرحله سعی و خطا به جواب نهایی همگرا شدند. بررسی نتایج نشان داد که از بین ایستگاه‌های انتخاب‌شده، بارندگی ثبت‌شده در ایستگاه‌های سراوان، کجدار، لادیز و چابهار تأثیر بیشتری از تغییرات SST گرفته‌اند. به‌منظور بررسی کارایی مدل، بین نتایج روش پیشنهادی و مطالعه‌ی قبلیِ انجام‌شده در مورد خوشه‌بندی با الگوریتم ژنتیک صورت مقایسه‌یی صورت گرفته که بخشی از نتایج آن در شکل ۴ ارائه شده است. در این شکل، نقاط مشخص‌شده با علائم دایره و مربع، نشان‌دهنده‌ی نقاط واقع بر خوشه‌های منتخب است. لازم به ذکر است که داده‌های مورد استفاده از SST به‌صورت استاندارد مورد استفاده قرار گرفته‌اند.

در شکل ۴ الف، وقتی تأثیر اقیانوس هند بر بارش ایستگاه لادیز بررسی می‌شود، در روش K-Means، ۸ سال با بارندگی کم‌تر نرمال در یک خوشه و ۸ سال دیگر در یک خوشه قرار دارند که ۷ عدد از آنها بالای میانگین بارش هستند و نمایانگر بارش بالای نرمال با احتمال ۸۴٪ هستند. این میزان دقت برای روش GAمطابق



شکل ۴. نتایج خوشه‌بندی SST باتوجه به بارندگی ایستگاه‌های منتخب با روش‌های K-Means پیشنهادی و GA.

شکل ۴ ب، به‌ترتیب ۱۰۰٪ و ۸۰٪ برای بالای نرمال و پایین نرمال است. زمانی که تأثیر SST دریای عمان بر ایستگاه سراوان بررسی می‌شود، مطابق شکل ۴ ج، مدل خوشه‌بندی K-Means دو خوشه را انتخاب می‌کند که دسته‌ی اول شامل ۸ سال با بارندگی بیش از میانگین و مابقی سال‌ها در دسته‌ی دوم و با احتمال ۹۲٪ دارای بارش کم‌تر از میانگین درازمدت هستند.

نتایج مدل K-Means پیشنهادی درمورد ایستگاه چابهار در شکل ۴ و، در حالتی که تأثیر هر سه منطقه‌ی جغرافیایی بررسی می‌شود، نشان می‌دهد که ۸ سال در یک خوشه قرار دارد که ۷ سال آن با بارش زیر نرمال است که نشان‌دهنده‌ی دقت ۸۳٪ است؛ در این حالت نتایج بسیار نزدیک به روش خوشه‌بندی ژنتیک (شکل ۴ ه) است.

به‌طورکلی با توجه به نتایج کسب‌شده، در نتایج روش K-Means پیشنهادی و روش GA-Clustering نزدیکی بسیار زیادی دیده می‌شود و در بعضی مواقع نیز نتایج روش پیشنهادی نسبت به روش قبلی بهتر بوده است. در جدول ۱ مقایسه‌ی دقت خوشه‌بندی هرکدام از موارد بین سه روش K-Means پیشنهادی، الگوریتم ژنتیک و K-Means معمولی ارائه شده است. اعداد نشان داده شده در جدول ۱ نشان‌دهنده‌ی آن است که چه‌درصدی از داده‌های قرارگرفته در هر خوشه نماینده‌ی داده‌های با بارش بالای نرمال و پایین نرمال هستند. قسمت‌هایی که با علامت (-) نشان داده شده‌اند و فاقد عددند، دقت کم‌تر از ۵۰ درصد دارند که بدین دلیل برای تدوین دستورالعمل پیش‌بینی بارش قابل استفاده نیستند. همان‌طور که مشخص است روش K-Means معمولی همواره از دقت پایینی برخوردار است در حالی که نتایج دو روش دیگر در اکثر مواقع قابل قبول‌اند.

نتایج ارائه‌شده در مقاله نشان می‌دهد که کم‌ترین دقت پیش‌بینی با استفاده از روش K-means مربوط به پیش‌بینی بارش در ایستگاه سراوان با استفاده از SST دریای عمان بوده است که در ۱۶ درصد سال‌ها، دوره‌های پربارش و کم بارش را درست تشخیص نداده است. این میزان در روش GA، ۲۲ درصد بوده است.

جدول ۱. مقایسه‌ی دقت نتایج خوشه‌بندی K-Means با روش‌های خوشه‌بندی GA-Clustering و K-Means معمولی.

ایستگاه باران سنجی	ناحیه جغرافیایی	K-Means معمولی	GA-Clustering	K-Means پیشنهادی
سراوان	دریای عمان	بالای نرمال: - زیر نرمال: -	بالای نرمال: ٪۷۸ زیر نرمال: ٪۸۰	بالای نرمال: ٪۱۰۰ زیر نرمال: ٪۹۲
سراوان	دریای عرب	بالای نرمال: - زیر نرمال: -	بالای نرمال: ٪۸۵ زیر نرمال: -	بالای نرمال: ٪۸۵ زیر نرمال: -
کجدار	دریای عمان	بالای نرمال: - زیر نرمال: -	بالای نرمال: - زیر نرمال: ٪۱۰۰	بالای نرمال: - زیر نرمال: ٪۱۰۰
لادیز	اقیانوس هند	بالای نرمال: - زیر نرمال: -	بالای نرمال: ٪۱۰۰ زیر نرمال: ٪۸۰	بالای نرمال: ٪۹۰ زیر نرمال: ٪۱۰۰
چابهار	دریای عمان	بالای نرمال: - زیر نرمال: -	بالای نرمال: ٪۱۰۰ زیر نرمال: ٪۸۴	بالای نرمال: - زیر نرمال: ٪۹۰
چابهار	دریای عمان، دریای عرب و اقیانوس هند	بالای نرمال: - زیر نرمال: -	بالای نرمال: - زیر نرمال: ٪۸۵	بالای نرمال: - زیر نرمال: ٪۸۴

۵. تدوین دستورالعمل پیش‌بینی بارش

از مهم‌ترین کاربردهای خوشه‌بندی سیگنال‌های هواشناسی، علاوه بر تعیین رفتار درازمدت آنها و تأثیرشان بر بارندگی، استفاده از آن در پیش‌بینی بارش است. درمورد تحقیق صورت‌گرفته در این نوشتار، با استفاده از نتایج به‌دست آمده به‌وسیله‌ی مدل خوشه‌بندی K-Means پیشنهادی در هر ایستگاه و تعیین سال‌های با بارندگی بیش از نرمال و کم‌تر از نرمال می‌توان با توجه به میزان تغییرات SST در مناطق جغرافیایی انتخاب شده، وقوع بارش‌های بیشتر یا کم‌تر از نرمال را پیش‌بینی کرد. پس از اجرای مدل خوشه‌بندی مورد نظر، دستورالعملی برای پیش‌بینی بارش در ایستگاه‌های منتخب استان سیستان و بلوچستان در ماه‌های دی و بهمن و اسفند براساس مقدار SST در ماه‌های آگوست، سپتامبر و اکتبر براساس نتایج خوشه‌بندی تدوین می‌شود که نتایج آن در جدول ۲ نشان داده شده است.

روش کار بدین صورت است که با دانستن این که به‌عنوان مثال بارش حوزه‌ی ایستگاه لادیز تحت تأثیر SST اقیانوس هند، دارای چه ترکیبی از خوشه‌بندی خواهد بود، با در دست داشتن میانگین مقدار SST دریای عمان در ماه‌های نام برده شده و با مراجعه به خوشه‌یی که SST مورد نظر در آن قرار دارد می‌توان به وضعیت کلی متوسط بارش به‌صورت زیر نرمال یا بالای نرمال در ماه‌های دی و بهمن و اسفند همان سال پی برد و وضعیت به دست آمده را با وضعیت واقعی مقایسه کرد. با تکرار این

جدول ۲. دستور العمل پیش بینی بارش با استفاده از خوشه بندی SSTبه روش K-Means پیشنهادی.

نام ایستگاه	ناحیه جغرافیایی پیش‌بینی کننده (SST)	محدوده تغییرات SST(C°)	وضعیت پیش‌بینی بارش
سراوان	دریای عمان	$۲۷/۷ \le SST_O \le ۲۸$	$P > ۱۵/۳\,mm$
		$۲۶/۴ \le SST_O \le ۲۷/۷$	$P < ۱۵/۳\,mm$
سراوان	دریای عرب	$۲۶/۲ \le SST_A \le ۲۶/۳$	$P > ۱۵/۳\,mm$
کجدار	دریای عمان	$۲۷/۴ \le SST_O \le ۲۷/۷$	$P < ۳۰/۰۸\,mm$
لادیز	اقیانوس هند	$۲۶/۵ \le SST_I \le ۲۶/۸$	$P > ۱۲/۷\,mm$
		$۲۶/۹ \le SST_I \le ۲۸$	$P < ۱۲/۷\,mm$
چابهار	دریای عمان	$۲۷/۷ \le SST_O \le ۲۸$	$P < ۱۹/۲\,mm$
چابهار	دریای عمان، دریای عرب و اقیانوس هند	$۲۶/۲ \le SST_O \le ۲۶/۳$ $۲۵/۴ \le SST_A \le ۲۶/۵$ $۲۶/۸ \le SST_I \le ۲۸$	$P < ۱۹/۲\,mm$

SST_I: دمای سطح اقیانوس هند

SST_O: دمای سطح دریای عمان

SST_A: دمای سطح دریای عرب

آزمایش برای چند سال مختلف و مقایسه با وضعیت حقیقی بارش آن سال می‌توان به صحت عملکرد مدل تدوین شده پی برد و به پیش‌بینی وضعیت درازمدت بارش در سال‌های آینده و در حوزه‌های مختلف استان پرداخت. در مثال یادشده و با توجه به جدول ۲ اگر متوسط اندازه‌گیری شده‌ی SST اقیانوس هند در ماه‌های آگوست تا اکتبر بین ۲۶٫۵ و ۲۶٫۸ درجه سانتی‌گراد باشد، وضعیت بارش ایستگاه لادیز در زمستان همان سال شمسی بیشتر از میانگین ($P > ۱۲٫۷\ mm$) و چنانچه SST بین ۲۶٫۹ و ۲۸ درجه سانتی‌گراد باشد، بارش کم‌تر از میانگین ($P > ۱۲٫۷\ mm$) مورد انتظار خواهد بود. مشابه همین پیش‌بینی برای ایستگاه‌های دیگری که نتایج خوشه‌بندی آنها قابل قبول بوده است، در جدول ۲ مشاهده می‌شود.

۶. نتیجه‌گیری

روش‌های متعدد آماری، تحلیلی و تجربی برای بیان ارتباط بین سیگنال‌های بزرگ مقیاس اقلیمی و اطلاعات دما و فشار سطح دریا با متغیرهای مهم هیدرولوژیکی همچون بارش به کارگرفته شده، اما استفاده از روش‌هایی مثل K-Means و الگوریتم ژنتیک به‌ندرت صورت گرفته است. در این نوشتار در الگوریتم روش K-Means معمولی، نحوه‌ی میانگین‌گیری چنان اصلاح شده که بتوان با در نظرگرفتن تغییرات بارش سال‌های مختلف به‌صورت زیر نرمال و بالای نرمال به خوشه‌بندی سیگنال SST در همان سال‌ها پرداخت و در واقع بدون استفاده از روش‌های رگرسیونی و آماری در داخل مدل خوشه‌بندی، می‌توان به ارتباط بین میانگین SST ماه‌های آگوست تا اکتبر (مصادف با ماه‌های مرداد تا آبان) با بارش میانگین فصل زمستان بعد از آن پی برد. این رویکرد تحت عنوان روش K-Means اصلاح‌شده نامگذاری شده است. نتایج این روش نشان‌گر قابلیت استفاده از آن برای پیش‌بینی فصلی بارش در منطقه‌ی مورد مطالعه بوده است. با استفاده از این روش می‌توان در دهم آبان ماه (انتهای ماه اکتبر) وقوع بارش بیشتر یا کم‌تر از نرمال را در زمستان همان سال پیش‌بینی کرد. برنامه‌ریزی و پیش‌بینی درازمدت منابع آب سطحی در جهت بهره‌برداری صحیح از مخازن سدها در استان سیستان و بلوچستان به‌علت دارابودن شرایط نسبتاً بحرانی در زمینه‌ی منابع آب موجود و شرایط خشک‌سالی حاکم بر آن بسیار حائز اهمیت است. لذا این منطقه به‌عنوان مطالعه‌ی موردی در این تحقیق انتخاب شد که نتایج مدل نیز نشان‌دهنده‌ی قابلیت آن در پیش‌بینی بارش منطقه و تأثیرپذیری آن از آب‌های اطراف در بخش‌هایی از دریای عمان، دریای عرب و اقیانوس هند است. شناسایی این مناطق در دریاهای ذکرشده با استفاده از نظرکارشناسی و با درنظرگرفتن ارتباط تغییرات داده‌های بارش با SST در محدوده‌ی جغرافیایی نسبتاً وسیعی در جنوب ایران صورت گرفته است که در تحقیقات بعدی، با توسعه‌ی مدل‌های مبتنی بر تغییرات مکانی ممکن است به‌صورت دقیق‌تری صورت گیرد.

نتایج این تحقیق و به‌کارگیری آن در مناطق دیگرکشور و همچنین در نظرگرفتن داده‌های مهم و مؤثر دیگر مانند SLP(فشار سطح دریا)، کمک شایانی به بهبود روش‌های پیش‌بینی و خوشه‌بندی موجود و برنامه‌ریزی سودمند منابع آبی منطقه توسط کارشناسان مرتبط با علوم مختلف مهندسی آب خواهد کرد.

پانوشت

1. genetic k-means algorithm
2. national center for atmospheric research

منابع

1. Green, P.E.; Frank, R.E and Robinson, P.J. "Cluster analysis in test market selection", *Journal of management science*, **13**(8), pp. 387-400 (1967).
2. Levrat, E.; Bombardier, V.; Lamotte, M. and Bremont, J. "Multi-level image segmentation using fuzzy clustering and local membership variations detection", *IEEE International Conference on Fuzzy Systems* **Volume, Issue**, 8-12, pp. 221-228 (1992).
3. MacQueen, J. "Some methods for classification and analysis of multivariate observations" *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* **Volume I,** Statistics. Edited by Lucien M., Le Cam and Jerzy Neyman. University of California Press (1967).
4. Krishna, K. and Murty Narasimha, M. "Genetic k-means algorithm", *IEEE Transactions on Systems Man And Cybernetics*, Part B: Cybernetics, **29**(3), pp.433-439 (1999).
5. Kuo, R.J.; Wang, H.S.; Lai Hu, T. and Chou, S.H. "Application of ant k-means on clustering analysis", *Journal of Computers and Mathematics with applications*, **50**, pp. 1709-1724 (2005).
6. Hoffman, F.M.; Hargrove, W. and Erickson, D. "Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models ", *Journal of Earth Interactions*, **9**, (2005).
7. Nazemosadat, M.J. "Persian gulf sea surface temperature as a drought diagnostic for southern parts of Iran", *J. of Drought News Network*, **10**, pp. 12-14 (1998).
8. Semmler, S. and Jacob, J. "Modeling extreme precipitation events-a climate change simulation for Europe", *J. of Global and Planetary Change*, **44**, pp.119-127 (2004).
9. Rucong, Y.; Minghua, Z.; Yongqiang, Y. and Yimin, L. "Summer monsoon rainfalls over mid-Eastern China lagged correlated with global SSTs", *J. of Advances in Atmospheric Sciences*, **18**(2), pp.179-196.
10. Eghdami, S."Survey of climate signals effects on rainfall and runoff in kareen and Dez river basins", M.Sc thesis, School of Civil Engineering, University of Tehran, (2004).
11. Zahraie, B. and Roozbahani, A. "Climate Signals Clustering with respect to precipitation variations in Sistan & Balouchestan province", *Proceedings of second Conference on water resources Management in Iran*, Isfahan, (2006).
12. Zahraie, B. and Roozbahani, A. "Climate signal clustering using genetic algorithm for precipitation forecasting: A case study of southeast of Iran", *Proceedings of the Word Environmental & Water Resources Congress* (ASCE),Tampa, Florida, USA, (2007).