

## تحلیل اکتشافی و تشخیص ناهمگنی سری‌های دما و بارش ایستگاه‌های هواشناسی ایران (دوره ۲۰۱۸-۱۹۵۹)

محسن رحمدل<sup>۱\*</sup>، زهره جوانشیری<sup>۲</sup> و سید حسین ثنائی‌نژاد<sup>۳</sup>

<sup>۱</sup>دانش‌آموخته دکتری هواشناسی کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران

<sup>۲</sup>استادیار، پژوهشکده اقلیم‌شناسی مشهد، مشهد، ایران

<sup>۳</sup>استاد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی، مشهد، ایران

(دریافت: ۱۴۰۰/۰۷/۱۵، پذیرش: ۱۴۰۰/۱۱/۰۵)

### چکیده

داده‌های دیده‌بانی شده در ایستگاه‌های هواشناسی زیربنای طیف وسیعی از برنامه‌ریزی‌ها، مطالعات کاربردی و مدل‌سازی‌ها در زمینه‌ها و علوم مختلف هستند. استفاده از این داده‌ها در مطالعات و برنامه‌ریزی‌ها، بدون اطمینان از صحت داده‌ها و همگن بودن سری زمانی آنها می‌تواند به نوعی عدم قطعیت در نتایج منجر شود؛ بنابراین با توجه به اهمیت پارامترهای دما و بارش در برنامه‌ریزی‌های محیطی، در این مقاله سری‌های زمانی دمای کمینه و بیشینه و بارش روزانه در ایستگاه‌های هواشناسی کشور با رویکرد تحلیل اکتشافی و بررسی همگنی در بازه زمانی ۶۰ ساله (۱۹۵۹-۲۰۱۸) ارزیابی شدند. در بررسی‌ها مشخص شد در سری زمانی ۶۰ ساله از بین ۳۷ ایستگاه به طور متوسط دمای بیشینه روزانه به میزان ۵٪، دمای کمینه بیش از ۷٪ و بارش روزانه، ۲٪ داده‌گم شده دارند. در همین دوره بدون احتساب داده‌های سال ۱۹۵۹، برای پارامتر دمای بیشینه و دمای کمینه هر کدام هفت مورد و برای بارش نیز هشت مورد داده پرت تشخیص داده شد. در بررسی سری‌های زمانی، بدون احتساب ایستگاه طیس و از بین ۳۶ ایستگاه، دمای روزانه ۱۶ ایستگاه همگن تشخیص داده شد و در ۲۰ ایستگاه نیز یک، دو یا سه نقطه شکستگی در سری زمانی مشخص شد. برای بارش نیز در همین دوره ۵ ایستگاه ناهمگن شناخته شدند.

**واژه‌های کلیدی:** بارش، دما، ناهمگنی، فراداده، بسته نرم‌افزاری کلایماتول

## ۱ مقدمه

داده‌های هواشناسی حاصل از اندازه‌گیری در انواع ایستگاه‌های هواشناسی، برای بسیاری از برنامه‌ها، مطالعات و مدل‌سازی‌ها در علوم مختلف نظیر هواشناسی، هوانوردی، اقلیم‌شناسی، محیط زیست، شهرسازی، کشاورزی، آب‌شناسی و... کاربرد دارند و بدون اطمینان از صحت و همگنی داده‌ها نمی‌توان به استفاده و استخراج نتایج از آنها اقدام کرد. وجود داده‌های گم‌شده (missing data) و پرت (outliers) در ایستگاه‌های هواشناسی و همگن نبودن سری‌های آنها از جمله مشکلاتی است که پژوهشگران در بخش‌های بزرگی از جهان به‌ویژه کشورهای کمتر توسعه‌یافته با آن مواجهند. جانمایی نامناسب ایستگاه‌ها، خطاهای انسانی در هنگام قرائت و ثبت داده‌ها، خطاهای مربوط به تجهیزات و ادوات اندازه‌گیری، جابه‌جایی ایستگاه‌های هواشناسی، یکسان نبودن روش‌های اندازه‌گیری و رعایت نشدن شرایط و استانداردهای تعریف‌شده، خارج شدن ایستگاه‌ها از شرایط استاندارد با گذشت زمان، تغییر در نوع ادوات و حسگرهای اندازه‌گیری پارامترهای جوئی در دوره آماری و... از مشکلاتی هستند که می‌توانند صحت و همگنی داده‌های هواشناسی را تحت تأثیر قرار دهند. در این زمینه سازمان هواشناسی جهانی (WMO) با همکاری کمیسیون اقلیم‌شناسی (Commission for Climatology)، مجموعه‌ای از شیوه‌نامه‌های لازم را برای همگن‌سازی داده‌ها تهیه کرده و در دسترس قرار داده است؛ برای نمونه می‌توان به سند شماره ۱۱۸۶ سازمان هواشناسی جهانی، با عنوان راهنمای فراداده‌های اقلیمی و همگن‌سازی (سازمان جهانی هواشناسی، ۲۰۰۳) و سند شماره ۱۲۰۳ این سازمان با عنوان شیوه‌نامه‌های سازمان جهانی هواشناسی برای محاسبه نرمال‌های اقلیمی (سازمان جهانی هواشناسی، ۲۰۱۷) اشاره کرد. در همین زمینه، سازمان همکاری‌های اروپا در علم و فناوری (European Cooperation in

Science and Technology)، پروژه ES0601 را با نام روش‌های پیشرفته همگن‌سازی سری‌های اقلیمی یا HOME با هدف ارزیابی روش‌های مختلف همگن‌سازی و ترکیب کاراترین روش‌ها و ارائه روشی یکپارچه و کلی، بین سال‌های ۲۰۰۷ تا ۲۰۱۱ اجرا کرد. نتیجه این پروژه، نرم‌افزار هومر (Homer) بود که نرم‌افزاری بسیار قوی برای تشخیص ناهمگنی و همگن‌سازی داده‌های سالانه و ماهانه است (مستره و همکاران، ۲۰۱۳). علاوه بر آن، گویچارو (۲۰۱۴) الگوریتمی را برای کنترل کیفیت و همگن‌سازی داده‌های روزانه ارائه کرد. در تحقیق مینگ و فدحی ... (۲۰۱۲) در شبه‌جزیره (peninsular) مالزی که جهت صحت‌سنجی داده‌های بارش انجام گرفت، با استفاده از سری‌های زمانی روزانه بارش، آزمون همگنی برای سه متغیر میانگین بارش سالانه، بیشینه بارش سالانه و میانه بارش سالانه با چهار روش آزمون دامنه بیشاند (Buishand range test)، آزمون همگنی پتیت (Pettitt test)، آزمون نسبت ون‌نیومن (Von Neumann ratio test) و آزمون نرمال استاندارد اجرا شد. این تحقیق نشان داد برای میانگین بارش سالانه و بیشینه بارش سالانه در منطقه مورد مطالعه، تمامی ایستگاه‌ها همگن و استفاده‌شده بودند. لین و همکاران (۲۰۱۵)، بارش حوزه رودخانه کلاتان (Kelantan) مالزی را مطالعه کردند. در آن مطالعه، همگنی داده‌ها - که بارش‌های دامنه بیشاند، آزمون پتیت، آزمون نسبت ون‌نیومن و آزمون نرمال استاندارد برای داده‌های ماهانه، فصلی و روزانه بارش بررسی شده بود - نشان داد ۹۴/۷٪ داده‌های بارش روزانه، ۷۰٪ سری‌های بارش سالانه و ۹۷/۵٪ از سری‌های بارش فصلی همگن هستند. در تحقیق عمر و همکاران (۲۰۱۷) آزمون همگنی برای اطلاعات بارندگی فصلی و سالیانه شمال عراق با روش‌های دامنه بیشاند، آزمون پتیت، آزمون نسبت ون‌نیومن و آزمون نرمال استاندارد انجام شد. ایشان

اقليمی، تقاضای تبخیری جوّ (atmospheric evaporative demand) و خشکسالی هواشناسی بین سال‌های ۱۹۸۸ تا ۲۰۱۸ در جنوب غرب ایران در حوزه کرخه با استفاده از سری زمانی ماهانه متغیرهای جوّی (دما، بارش، سرعت باد و ساعت آفتابی) بررسی شد. این پژوهش نشان داد تغییرات اقلیمی معنی‌داری در محدوده مورد مطالعه وجود دارد، از جمله افزایش دمای هوا و سرعت باد و کاهش رطوبت نسبی که باعث روند مثبت در تقاضای تبخیری جوّ شده است. کوثری و همکاران (۱۴۰۰) کارایی روش‌های بازسازی نواقص آماری بارش در مناطق خشک ایران را با استفاده از روش‌های نسبت نرمال، رگرسیون خطی، رگرسیون چندگانه و عکس مجذور فاصله (IDW) بررسی کردند و نشان دادند هنگامی که نقص داده‌ها زیاد نباشد، روش رگرسیون چندمتغیره از دقت بیشتری در بازسازی داده‌های روزانه برخوردار است، ولی در کل به میزان نسبت داده‌های گم‌شده حساس است. روش نسبت نرمال در بازسازی نواقص بارش روزانه مناسب نیست ولی نسبت به میزان نقص داده‌ها از سایر روش‌ها پایدارتر است. در سری‌های زمانی ماهانه، عملکرد IDW و سپس نسبت نرمال مناسب است. در سری‌های سالانه به ترتیب، روش همبستگی خطی، نسبت نرمال و IDW عملکرد بهتری دارند.

دما و بارش از مهم‌ترین عناصر جوّی و اقلیمی هستند که وضعیت جوّی، آب‌شناختی، پوشش گیاهی، کشاورزی، منابع طبیعی و محیط زیست را در یک منطقه تحت تاثیر قرار می‌دهند. آشکار است که قبل از هر مطالعه‌ای درباره این پارامترها، کنترل کیفیت و همگن‌سازی سری زمانی مربوط به آنها ضروری است. اگر داده‌ها کیفیت مناسبی نداشته باشند، نتایج مطالعات اعتبار کافی را نخواهند داشت؛ از این رو در این مقاله، کنترل کیفیت و همگن‌سازی برای داده‌های مربوط به سری‌های زمانی پارامترهای دما و بارش برای دوره ۱۹۵۹-

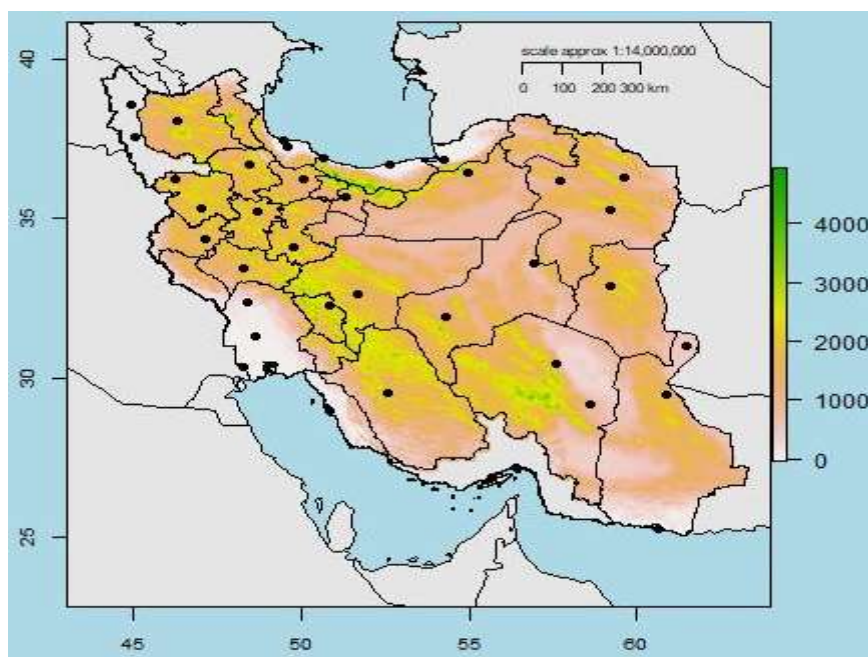
از مجموع سری‌های زمانی بارش نه ایستگاه، سری زمانی دو ایستگاه برای بارش سالانه، سه ایستگاه برای بارش زمستانه و یک ایستگاه را نیز برای بارش بهار به عنوان سری‌های مشکوک به ناهمگنی شناسایی کردند. در ایران نیز رحیم‌زاده و نساجی‌زواره (۲۰۱۳) سری زمانی دمایی ۱۵۵ ایستگاه هواشناسی ایران را در بازه زمانی ۲۰۱۰-۱۹۶۰ به لحاظ همگن بودن و روند تغییرات دمایی ارزیابی کردند و با توجه به نتایج مطالعات گذشته دین‌پژوه و همکاران (۱۳۸۲)، مسعودیان (۱۳۸۲)، علیجانی (۱۳۹۰) و توکلی و محمودیان (۲۰۱۱) که بیشتر با استفاده از روش‌های کمی و تجربی شامل تحلیل مؤلفه‌های اصلی و تحلیل خوشه‌ای صورت گرفته بود، مناطق مختلف کشور را خوشه‌بندی کردند و به این نتیجه رسیدند ناپیوستگی‌های دمایی در سری‌های زمانی اغلب به دلیل جابه‌جایی و تغییر در شرایط محیطی ایستگاه‌ها بوده است. در پژوهش رضیعی (۱۳۹۵) به منظور شناسایی مناطق همگن بارشی ایران با استفاده از داده‌های بارش ۱۵۵ ایستگاه همدیدی و با انجام دادن آرایه S، مؤلفه‌های اصلی برای ماتریس بارش ماهانه ایستگاه‌های استفاده‌شده تحلیل شد و پنج منطقه همگن بارشی خزری، منطقه بارشی شمال غربی، منطقه بارشی غربی، منطقه بارشی مرکزی-شرقی و منطقه بارشی مرکزی-شمال‌شرقی برای ایران شناسایی شدند. در تحقیق بازگیر و همکاران (۱۳۹۷) که درباره سری‌های زمانی میانگین سالانه دمای کمینه و بیشینه و بارش ۳۶ ایستگاه همدید کشور در دوره زمانی ۱۹۶۶-۲۰۱۸ با استفاده از آزمون‌های نرمال استاندارد (Standard Normal Homogeneity Test)، پتیت، انحراف تجمعی (Cumulation Deviation Test) و راستی‌آزمایی ورسلی (Worsley's Liklyhood)، انجام شد، ۹۱/۵ درصد سری‌های زمانی دمای کمینه و ۸۸/۵ درصد سری‌های زمانی دمای بیشینه در طبقه ناهمگن قرار گرفتند. در پژوهش کریمی و همکاران (۲۰۲۰)، روند متغیرهای

کمینه و بیشینه روزانه ایستگاه‌های هواشناسی کشور و همچنین بارش ۲۴ ساعته این ایستگاه‌ها، در دوره زمانی ۶۰ ساله (۱۹۵۹-۲۰۱۸) بررسی شدند. پس از بررسی سری‌های زمانی این داده‌ها، نسبت به تعیین درصد داده‌های گم‌شده، داده‌های پرت و بررسی همگنی آنها به تفکیک ایستگاه و بر اساس شیوه‌نامه‌های سازمان جهانی هواشناسی اقدام شد. سپس دلایل ناهمگنی در برخی از ایستگاه‌هایی که نیاز به همگن‌سازی داشتند، با استفاده از فراداده این ایستگاه‌ها بررسی شد. موقعیت ایستگاه‌های بررسی شده در دوره زمانی ۶۰ ساله در شکل ۱ نشان داده شده است.

۲۰۱۸ انجام شده است. تفاوت این مطالعه با سایر مطالعاتی که در این زمینه درباره ایران انجام شده است، دوره آماری مورد مطالعه، روش کار و مقیاس زمانی است. در این مطالعه، دوره طولانی‌تری بررسی شده است، از الگوریتم کلاسیک، که ابزار کارتری نسبت به روش‌های کلاسیک است، برای اولین بار استفاده شده است و داده‌ها برای اولین بار در مقیاس روزانه همگن شده‌اند.

## ۲ داده‌ها و روش تحقیق

به منظور تحلیل داده‌های ایستگاه‌های هواشناسی و با توجه به اهمیت پارامتر دما و بارش و تأثیرگذاری آنها در فرایندهای اقلیمی و زیست‌محیطی، داده‌های دمای



شکل ۱. پراکنندگی ایستگاه‌های مورد مطالعه در سری زمانی ۶۰ ساله (۱۹۵۹-۲۰۱۸).

ایستگاه‌هایی که به لحاظ اقلیمی به نسبت مشابه بودند، در یک خوشه قرار گرفتند تا برای تعیین سری‌های مرجع وارد نرم‌افزار R شوند. آنگاه با استفاده از الگوریتم کلاسیک (Climatol) در نرم‌افزار R، سری‌های مرجع برای هر ایستگاه بر اساس تغییرپذیری پارامتری که بررسی می‌شود، در هر خوشه تعیین شدند. برای خوشه‌بندی اولیه

گام‌های اصلی در همگن‌سازی داده‌ها عبارت‌اند از: تحلیل فراداده و کنترل کیفیت داده‌ها؛ ساختن سری مرجع؛ تشخیص نقاط شکستگی؛ اصلاح داده‌ها (جوانشیری و همکاران، ۱۳۹۹). برای این منظور، ابتدا ایران بر اساس مشخصه‌های آب‌وهوایی به پنج خوشه اقلیمی تقسیم‌بندی شد. در این فرایند، داده‌های

خوشه (به‌عنوان زیرخوشه) که تغییرپذیری یکسانی برای پارامتر مورد نظر با ایستگاه مورد بررسی داشتند، به‌عنوان سری‌های مرجع در نظر گرفته شدند. بر اساس این الگوریتم، ابتدا سری مورد نظر با سری‌های مرجع به روش رگرسیون نوع II برآورد و استانداردسازی شد. بعد از برآورد سری‌ها، سری بی‌هنجاری استاندارد شده محاسبه شد که تفاضل مقادیر مشاهده شده و برآورد شده است. سپس داده‌های پرت با استفاده از سری‌های بی‌هنجاری استاندارد شده برآورد شدند. به این ترتیب که بی‌هنجاری‌هایی که از آستانه‌های تعیین شده تجاوز کنند، داده پرت فرض شدند. این آستانه‌ها با استفاده از بافت‌نگار بی‌هنجاری‌ها تعیین شدند. در ادامه پژوهش، برای اطمینان از تشخیص صحیح داده‌های پرت، وضعیت جوئی در روز مورد نظر ایستگاه و روزهای قبل و بعد از آن، از جمله بارش شش ساعته، همخوانی گزارش‌های شش ساعته بارش با بارش گزارش شده ۲۴ ساعته، وضعیت ابرناکی، پدیده‌های هوای حاضر و گذشته و ارتباطات منطقی بارش با وضعیت جوئی گزارش شده بررسی شد. علاوه بر این، داده‌های ساعتی دما و روندهای دمایی و همخوانی با دمای کمینه و بیشینه گزارش شده در تاریخ‌های مورد نظر با استفاده از چند پایگاه داده هواشناسی نظیر سامانه درخواست داده‌های هواشناسی (<https://data.irimo.ir>) متعلق به سازمان هواشناسی کشور، پایگاه OGIMET (<https://www.ogimet.com/synops.phtml.en>) و METEOMANZ (<http://www.meteomanz.com/index?l=1>) بررسی شدند و داده‌های پرت تأیید شده به‌عنوان داده پرت نهایی گزارش شدند. سپس آزمون همگنی نرمال استاندارد (SNHT) برای سری‌ها در دو مرحله اجرا شد. از آنجا که این آزمون برای پیدا کردن یک نقطه شکست در سری طراحی شده است، وجود نقاط شکست بیشتر،

ایستگاه‌ها با توجه به فعالیت‌ها و مطالعات قبلی که در این زمینه صورت گرفته بود، با استناد به خوشه‌بندی مقاله رحیم‌زاده و نساجی‌زواره (۲۰۱۴) که با توجه به نتایج مطالعات گذشته و بیشتر با استفاده از روش‌های کمی و تجربی شامل تحلیل مؤلفه‌های اصلی و تحلیل خوشه‌ای صورت گرفته بود، مناطق مختلف کشور در پنج خوشه اقلیمی تقسیم‌بندی شد (شکل ۲). علیجانی (۱۳۸۵) در تقسیم‌بندی به نسبت مشابه دیگری، مناطق مختلف کشور را به لحاظ آب‌وهوایی به پنج منطقه آب‌وهوایی تقسیم کرده بود. هدف از این خوشه‌بندی اولیه، تعیین خوشه‌هایی بود که در آن خوشه‌ها، تغییرات آب‌وهوایی تقریباً یکنواخت باشد؛ بنابراین نمودار همبستگی بین ایستگاه‌ها برای پارامترهای مورد بررسی، در هر خوشه به‌طور مجزا ترسیم شد. نمودار ضرایب همبستگی به ارزیابی یکنواختی تغییرات آب‌وهوایی در منطقه مورد نظر کمک می‌کند. به عبارت دیگر، این نمودار وجود عواملی را شناسایی می‌کند که ممکن است باعث تغییرات شدید آب‌وهوایی در منطقه شوند (مانند سلسله‌کوه‌ها و...). بررسی‌ها نشان دادند خوشه‌بندی اولیه برای هدف این پژوهش مناسب بوده است. خوشه‌بندی اولیه برای این است که سری‌های مرجع هر ایستگاه از همان خوشه انتخاب شوند. برای انتخاب سری‌های مرجع دوباره داخل هر خوشه، خوشه‌بندی مجدد بر اساس تغییرپذیری پارامتر مورد بررسی انجام شد. پس از خوشه‌بندی، تحلیل آماری داده‌های دمای بیشینه و کمینه و بارش روزانه با استفاده از نرم‌افزار SPSS انجام و درصد داده‌های گم‌شده به تفکیک هر ایستگاه مشخص شد. سپس برای بررسی داده‌های پرت، نقاط ناهمگنی و همگن‌سازی داده‌های کمینه و بیشینه دما و بارش روزانه از بسته نرم‌افزاری کلایماتول در نرم‌افزار R استفاده شد. به این ترتیب برای کنترل کیفیت و همگن‌سازی داده‌های مربوط به هر ایستگاه، در هر خوشه تعدادی از ایستگاه‌های متعلق به آن

### ۳ مبانی نظری مورد استفاده

#### ۳-۱ الگوریتم کلایماتول

اساس کار کلایماتول، برآورد سری مورد نظر با سری‌های مرجع به روش رگرسیون نوع II است. برای این منظور ابتدا سری‌ها با استفاده از رابطه (۱) استاندارد می‌شوند:

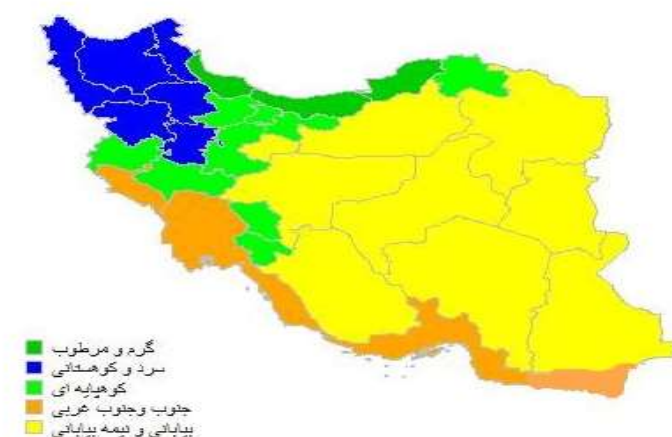
$$x = \frac{X - m_x}{s_x} \quad (1)$$

که  $m_x$  و  $s_x$  به ترتیب میانگین و انحراف معیار سری  $X$  هستند. سپس هر سری با استفاده از سری مرجعش با رابطه (۲) محاسبه می‌شود:

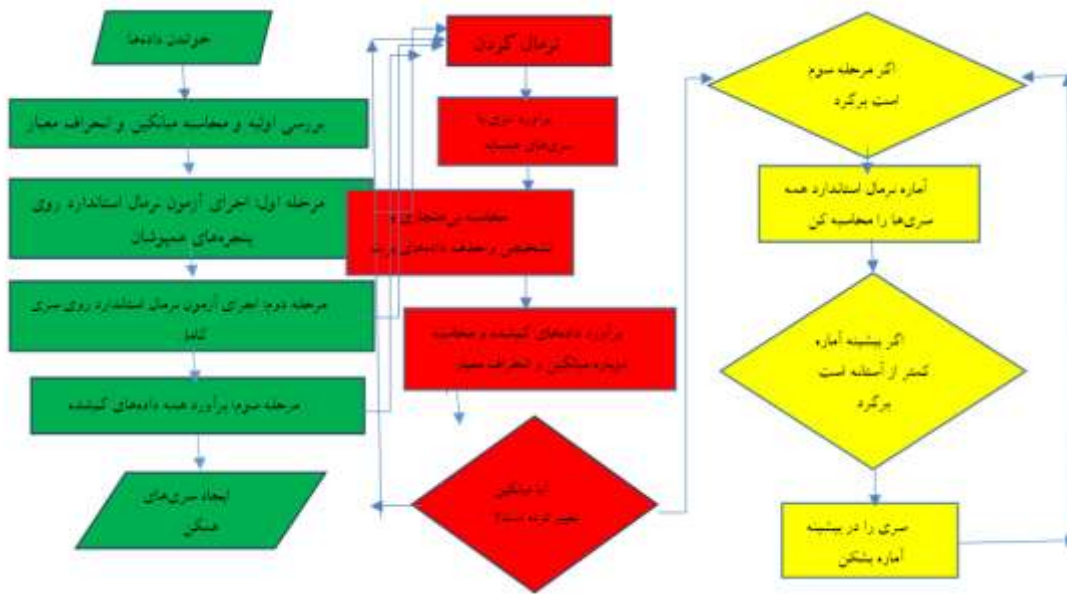
$$\hat{y} = \frac{\sum_{j=1}^{j=N} w_j x_j}{\sum_{j=1}^{j=N} w_j} \quad (2)$$

ن: داده‌ای است که از نزدیک‌ترین تعداد (N) داده مرجع موجود در هر گام زمانی تخمین زده می‌شود.  $x_j$  داده مرجع و  $w_j$  وزنی است که به داده مرجع اختصاص داده شده است و تابعی معکوس از فاصله آنها تا ایستگاه مورد نظر است (گویچارو، ۲۰۱۸). در شکل ۳ نمودار روند عملیاتی در الگوریتم کلایماتول نشان داده شده است.

ممکن است باعث اریبی در نتایج شود (جوانشیری و همکاران، ۱۳۹۹)؛ از این رو این آزمون در مرحله اول برای پنجره‌های همپوشان اجرا می‌شود. به طور مثال برای داده‌های روزانه، آزمون برای ۷۳۰ (دو برابر ۳۶۵) جمله اول سری اجرا شد که در هر مرحله، ۳۶۵ جمله به جلو حرکت می‌کرد. اگر داده‌ها ماهانه باشند، این مقدار از ۳۶۵ به ۶۰ کاهش می‌یابد. در مرحله دوم، آزمون روی سری کامل اجرا شد. هنگامی که آماره آزمون SNHT بیشتر از آستانه از پیش تعیین شده بود، سری در همان نقطه‌ای که پیشینه SNHT مربوط به آن بود، شکسته شد و داده‌های پیش از نقطه شکست به سری جدیدی (با همان مختصات جغرافیایی) انتقال داده شدند و از سری اصلی حذف شدند. این فرایند تکرار شد تا زمانی که تمام سری‌ها همگن شدند. بعد از تعیین نقاط شکستگی در مرحله دوم، جهت تفکیک جهش‌های اقلیمی و غیراقلیمی، از بانک فراداده استفاده شد و سری‌های ایستگاه‌های مجاور بررسی شدند. در مرحله سوم، همه داده‌های گم‌شده، داده‌های پرت یا داده‌هایی که به سری جدیدی منتقل شده‌اند، برآورد شدند.



شکل ۲. خوشه‌بندی اولیه کشور ایران بر اساس مشخصه‌های آب‌وهوایی با استناد به مقاله رحیم‌زاده و نساجی‌زواره (۲۰۱۳).

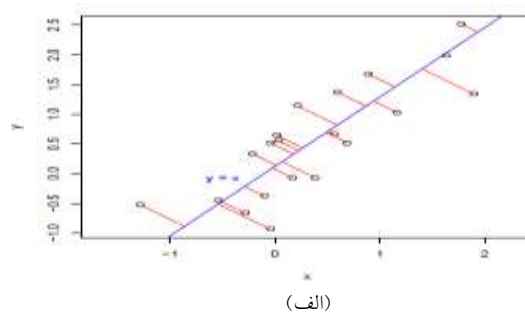
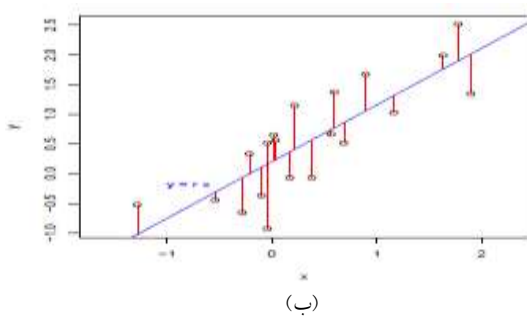


شکل ۳. نمودار روندهای عملیاتی در الگوریتم کلاماتول (گویجارو، ۲۰۱۸).

در رگرسیون نوع I (شکل ۴-ب) انحراف تا خط رگرسیون در راستای محور  $y$  کمینه می‌شود و فرمول آن با سری‌های استاندارد شده با رابطه  $\hat{y}_i = \tau x_i$  به دست می‌آید که  $\tau$  ضریب همبستگی بین متغیر مستقل و وابسته است (گویجارو، ۲۰۱۸). روش رگرسیون عمودی را اولین بار پائولوس و کوهلر (۱۹۵۲) برای برآورد داده‌های گم شده بارش ارائه کردند (جوانشیری و همکاران، ۱۳۹۹).

### ۲-۳ رگرسیون عمودی نوع II

در رگرسیون نوع II (شکل ۴-الف) فاصله عمودی نقاط تا خط رگرسیون کمینه می‌شود. در این نوع رگرسیون، رابطه  $\hat{y}_i = x_i$  یک مدل رگرسیون خطی با نام رگرسیون عمودی (orthogonal regression) یا محور اصلی کاهش یافته (reduced major axis) نامیده می‌شود که در آن فاصله عمودی نقاط تا خط رگرسیون کمینه می‌شود.



شکل ۴. به کمینه رساندن انحراف (خطوط قرمز) از خط رگرسیونی (آبی) با روش (الف) رگرسیون متعامد (رگرسیون نوع II) (ب) حداقل مربعات معمولی (رگرسیون نوع I) (گویجارو، ۲۰۱۸).

### ۳-۳ آزمون نرمال استاندارد (SNHT)

فرض صفر این آزمون، همگنی سری را همراه با توزیع نرمال با میانگین صفر و واریانس یک در مقابل ناهمگنی سری و وجود دست کم یک نقطه مانند  $M$  ارائه می کند که میانگین داده های قبل از آن،  $\mu_1$  و میانگین داده های بعد از آن  $\mu_2$  است.

برای انجام دادن این آزمون، سری استاندارد می شود. اگر  $z_i$  ( $i=1, \dots, n$ ) سری استاندارد شده باشد، آماره آزمون نرمال استاندارد از رابطه (۳) به دست می آید:

$$T = \max\{a\bar{z}_1^2 + (n-a)\bar{z}_2^2\}; 1 \leq a \leq n-1 \quad (3)$$

که  $z_1$  میانگین  $a$  داده اول ( $a=1, \dots, n-1$ ) و  $z_2$  میانگین  $n$  داده باقی است [جوانشیری و همکاران (۱۳۹۹) به نقل از الکساندرسون و مویرگ (۱۹۹۷)].

### ۴ یافته های تحقیق

با توجه به وجود داده های گم شده و مشکوک زیاد در سال ۱۹۵۹ در سری زمانی ۶۰ ساله، شروع دوره آماری از سال ۱۹۶۰ در نظر گرفته شد. ایستگاه قم که در سال ۱۹۵۲ میلادی تأسیس شده است، از سال ۱۹۶۷ تا ۱۹۸۵ داده ای ندارد؛ بنابراین در سری ایستگاه های ۶۰ ساله بررسی نشد. ایستگاه طبس نیز که در سال ۱۹۶۰ تأسیس شده بود و به دلیل زلزله طی سال های ۱۹۷۸ تا ۱۹۸۴ فعال نبود، از فرایند همگن سازی در دوره زمانی ۶۰ ساله کنار گذاشته شد. تاریخ تأسیس ایستگاه های سقز، زابل و چابهار به ترتیب یک، دو و سه سال بعد از سال ۱۹۶۰ بود که این دوره ها با استفاده از سری های مرجع در خوشه برآورد شدند؛ بنابراین تعداد ایستگاه ها در سری زمانی ۶۰ ساله ۳۶ ایستگاه در نظر گرفته شد. در بررسی ها مشخص شد که در بعضی از تاریخ ها، تمامی ایستگاه ها در یک خوشه فاقد داده هستند که احتمالاً ناشی از اشکال در سامانه مسیر (Messir) سازمان هواشناسی بود که داده ها از آن دریافت می شود؛ برای مثال در تاریخ ۲۴/۰۵/۲۰۱۴ در

کلیه خوشه ها هیچ کدام از ایستگاه ها دمای بیشینه ای ندارند. در چنین مواردی برای برآورد دمای بیشینه در تاریخ های مورد نظر از میانگین دمای بیشینه روزهای قبل و بعد خود ایستگاه ها استفاده شد. برای دمای بیشینه در ایستگاه های خوشه چهارم، ۱۸ تاریخ و برای دمای کمینه نیز ۱۳ تاریخ در تمامی ایستگاه های خوشه پنجم در سری زمانی ۶۰ ساله فاقد داده بودند.

### ۴-۱ داده های گم شده

در این مرحله، تحلیل آماری داده های دمای کمینه و بیشینه و بارش روزانه ایستگاه های هواشناسی کشور در دوره زمانی ۶۰ ساله (۲۰۱۸-۱۹۵۹) با نرم افزار SPSS انجام و درصد داده های گم شده پارامترهای مورد نظر به تفکیک ایستگاه ها مشخص شد. این بررسی ها نشان داد در دوره زمانی ۶۰ ساله، ایستگاه ها به طور متوسط برای پارامتر دمای بیشینه، ۵٪؛ دمای کمینه، ۷٪ و بارش، ۲٪ داده گم شده دارند. در دوره زمانی ۶۰ ساله پس از حذف داده های سال ۱۹۵۹، درصد داده های گم شده دمای بیشینه و کمینه و بارش به طور متوسط حدود ۰/۳ درصد کاهش یافت.

### ۴-۲ تحلیل اکتشافی و تشخیص نقاط ناهمگنی

(دوره ۲۰۱۸-۱۹۶۰)

#### ۴-۲-۱ خوشه اول

نواحی مرکزی ایران، شامل دشت کویر، دشت لوت و چاله جازموریان است (علیچانی، ۱۳۸۵). مرکز ایران از چندین حوضه بسته تشکیل شده است که در مجموع با نام فلات مرکزی از آن یاد می شود. قسمت شرقی این فلات با دو بیابان نمکی دشت کویر و دشت لوت پوشیده شده است. استان های قم، اصفهان، یزد و کرمان در منطقه مرکزی قرار دارند. بخش هایی از سیستان و بلوچستان، سمنان، خراسان جنوبی و خراسان رضوی نیز جزء مناطق بیابانی و نیمه بیابانی محسوب می شوند (رحیم زاده و



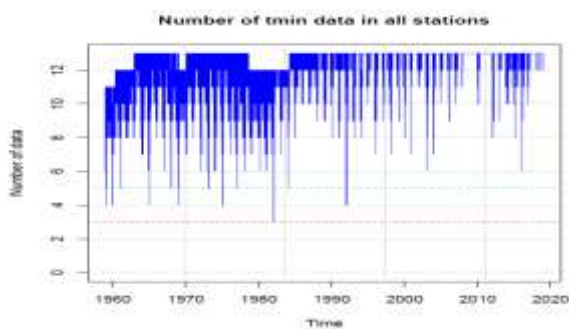
نساجی‌زواره، ۲۰۱۳). ایستگاه‌های واقع در این ناحیه در خوشه اول بررسی شدند.

#### ۴-۲-۱-۱ دمای بیشینه و دمای کمینه

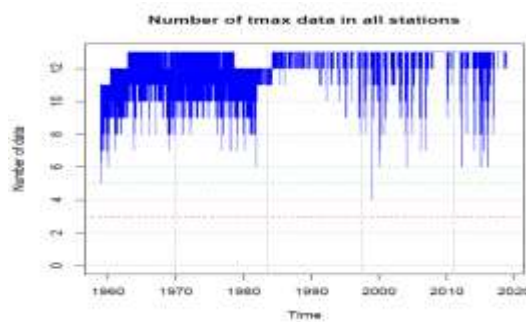
شکل ۵ داده‌های موجود را برحسب ایستگاه نشان می‌دهد. در این شکل خط نقطه‌چین سبز نشان می‌دهد برای تحلیل، در هر گام زمانی، تعداد پنج ایستگاه یا بیشتر، مطلوب و دست‌کم به سه ایستگاه برای هر گام زمانی (خط نقطه‌چین قرمز) نیاز است؛ بنابراین تعداد ایستگاه‌ها برای دمای بیشینه (شکل ۴-الف) و دمای کمینه (شکل ۴-ب) در بیشتر گام‌های زمانی جهت تحلیل‌های آماری، مطلوب و در برخی گام‌های زمانی، شرایط حداقل لازم را داراست. در شکل ۶ خوشه‌بندی ایستگاه‌ها در گروه‌هایی با تغییرپذیری یکسان برای دمای بیشینه و دمای کمینه نشان داده شده است. این شکل نشان می‌دهد ایستگاه‌های خوشه اول، بر اساس تغییرپذیری دمای بیشینه (شکل ۵-الف) و دمای کمینه (شکل ۵-ب) به سه زیرخوشه با تغییرپذیری یکسان تقسیم می‌شود. در شکل ۷ نمودار ضرایب همبستگی بین سری‌ها برای دمای بیشینه (شکل ۶-الف) و دمای کمینه (شکل ۶-ب) نشان داده شده است. نمودار ضرایب همبستگی به ارزیابی یکنواختی تغییرات آب‌وهوایی در منطقه مورد نظر کمک می‌کند. به عبارت دیگر، با این نمودار می‌توان وجود عواملی را شناسایی کرد که ممکن است باعث تغییرات شدید آب‌وهوایی در منطقه شوند (مانند سلسله کوه‌ها و...). همان‌گونه که مشاهده می‌شود، با افزایش فاصله مکانی، همبستگی بین دما در ایستگاه‌ها کاهش می‌یابد. این موضوع نشان می‌دهد منطقه انتخاب‌شده از نظر تغییرات آب‌وهوایی تقریباً یکنواخت بوده و برای هدف موردنظرمان مطلوب است. شکل ۸ نمودار دسته‌بندی ایستگاه‌ها را در خوشه و میزان شباهت و تجانس آنها را با یکدیگر با استفاده از

شاخص عدم تشابه (dissimilarity index) برای دمای بیشینه (شکل ۷-الف) و دمای کمینه (شکل ۷-ب) به‌عنوان یک معیار تفکیک نشان می‌دهد. همان‌گونه که مشاهده می‌شود، تمامی ایستگاه‌هایی که در نقشه به‌عنوان سری مرجع برای یکدیگر در خوشه استفاده می‌شوند، آستانه‌ای پایین‌تر از آستانه شاخص تعیین‌شده (خط قرمز در شکل) دارند. هرچه شاخص عدم تشابه پایین‌تر باشد، به این معنی است که داده‌ها از توزیع یکنواخت‌تری برخوردارند. برای بررسی داده‌های پرت، از بافت‌نگار بی‌هنجاری (تفاوت مقادیر مشاهده‌شده و برآوردشده) استاندارد شده استفاده شده است. در این روش مقادیری به‌عنوان داده پرت شناسایی می‌شوند که تفاوت زیادی با مقادیر همسایگی خود داشته باشند؛ بنابراین پس از بررسی بافت‌نگار بی‌هنجاری استاندارد شده برای دمای بیشینه، بی‌هنجاری‌های استاندارد شده بیشتر از ۸ و کمتر از ۸- و برای دمای کمینه، بی‌هنجاری‌های استاندارد شده بیشتر از ۷ و کمتر از ۷- داده پرت فرض شدند. شکل ۹ بافت‌نگار بی‌هنجاری‌های استاندارد شده را برای دمای بیشینه (شکل ۸-الف) و دمای کمینه (شکل ۸-ب) نشان می‌دهد.

در مرحله بعد، بررسی‌های دقیق هواشناسی برای روزهای همراه با داده‌های دمای پرت که با الگوریتم تشخیص داده شده بودند، در تاریخ مورد نظر و همچنین روزهای قبل و بعد از آن انجام شد. این بررسی‌ها، شامل بررسی دماهای ساعتی گزارش‌شده و همخوانی آنها با دمای کمینه و بیشینه و روندهای دمایی، پدیده‌های جوئی گزارش‌شده، شرایط ابرناکی و مقایسه با دمای بیشینه ایستگاه‌های مجاور بود. داده‌های پرت دمای بیشینه و دمای کمینه ایستگاه‌هایی که پس از بررسی‌های هواشناسی، داده پرت تشخیص داده شدند به ترتیب در جدول‌های ۱ و ۲ نشان داده شده است.

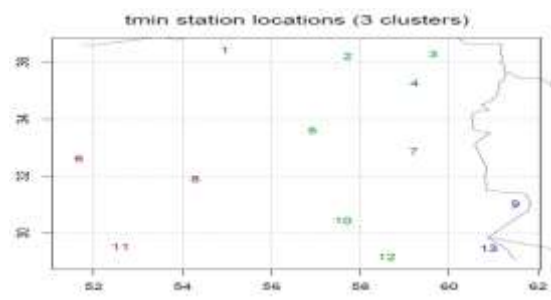


(ب)

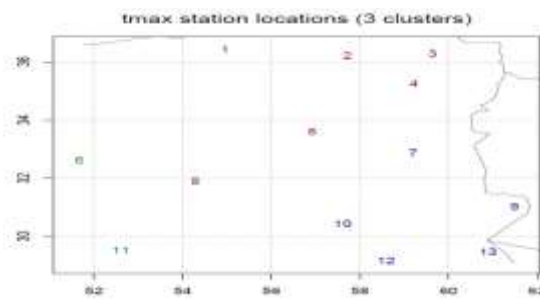


(الف)

شکل ۵. (الف) تعداد داده‌های دمای بیشینه و (ب) دمای کمینه ایستگاه‌های خوشه اول در گام‌های زمانی مختلف در دوره ۱۹۶۰-۲۰۱۸

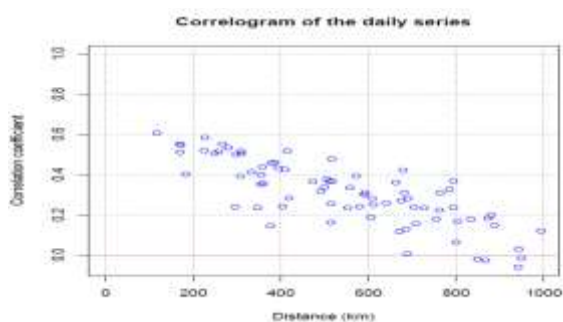


(ب)

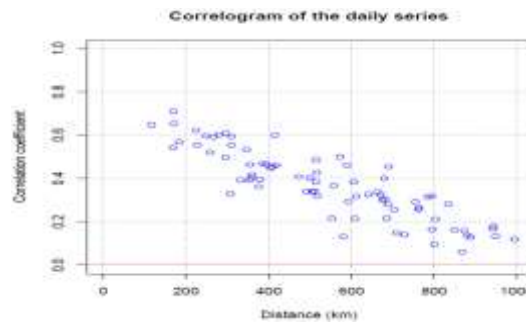


(الف)

شکل ۶. خوشه‌بندی ایستگاه‌های خوشه اول در گروه‌هایی با تغییرپذیری یکسان برای (الف) دمای بیشینه (ب) دمای کمینه در دوره ۱۹۶۰-۲۰۱۸.

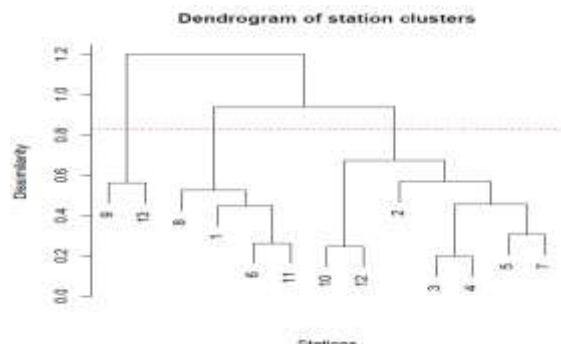


(ب)

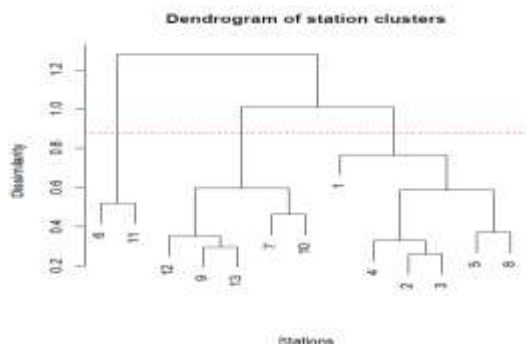


(الف)

شکل ۷. ضرایب همبستگی ایستگاه‌ها در مقابل فاصله مکانی ایستگاه‌های خوشه اول برای (الف) دمای بیشینه (ب) دمای کمینه در دوره ۱۹۶۰-۲۰۱۸.

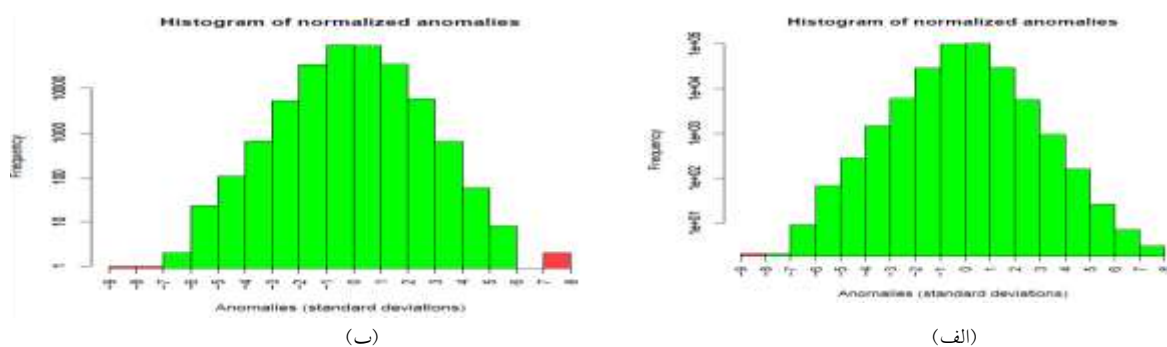


(ب)



(الف)

شکل ۸. دسته‌بندی ایستگاه‌ها در خوشه اول با استفاده از معیار تفکیک عدم تشابه برای (الف) دمای بیشینه (ب) دمای کمینه در دوره ۱۹۸۹-۲۰۱۸.



شکل ۹. بافت‌نگار بی‌هنجاری (تفاوت مقادیر مشاهده‌شده و برآوردشده) استانداردشده برای (الف) دمای بیشینه (ب) دمای کمینه در ایستگاه‌های خوشه اول در سری زمانی ۲۰۱۸-۱۹۶۰.

جدول ۱. داده‌های پرت دمای بیشینه ایستگاه‌های خوشه اول در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده‌شده	مقدار برآوردشده	بی‌هنجاری استانداردشده
۱	زاهدان	۴۰۸۵۶	۱۹۶۵/۰۲/۰۲	۲۱	۸/۵	۷/۴۱

جدول ۲. داده‌های پرت دمای کمینه ایستگاه‌های خوشه اول در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده‌شده	مقدار برآوردشده	بی‌هنجاری استانداردشده
۱	سبزوار	۴۰۷۴۳	۲۰۱۷/۸/۱۹	۲/۵	۲۱/۸	-۸/۶
۲	یزد	۴۰۸۲۱	۱۹۷۹/۵/۷	۲	۱۷/۷	-۷/۶۵

اول، آزمون نرمال استاندارد برای پنجره‌های همپوشان اجرا شد. در مرحله دوم، آزمون نرمال استاندارد برای سری‌های کاملی اجرا شد که از مرحله قبل به‌دست آمده‌بود. در هر نقطه که آماره SNHT بیشتر از حد آستانه بود، سری شکسته شد و همه داده‌های قبل از نقطه شکستگی به سری دیگری با همان طول و عرض جغرافیایی منتقل شدند. این فرایند برای سری‌های جدید اجرا شد تا زمانی که دیگر نقطه ناهمگنی تشخیص داده نشد. پس از تعیین نقاط شکستگی سری‌های زمانی دمای بیشینه و کمینه، با استفاده از بانک فراداده موجود بررسی شد تا معلوم شود این نقاط شکستگی که در مراحل قبل به‌دست آمدند، منشاء اقلیمی دارند یا غیراقلیمی. در بررسی فراداده موجود بسیاری از شهرستان‌ها دلیل مستدلی برای پیدا کردن علت این شکستگی‌ها تشخیص داده نشد. با

#### ۴-۲-۱-۱-۱ بررسی نقاط ناهمگنی و همگن‌سازی

##### سری‌های دمایی در خوشه اول

در این مرحله، سعی شد نقاط ناهمگنی در سری زمانی دمای بیشینه و دمای کمینه مشخص شود. از آنجاکه داده‌ها روزانه بودند و تغییرپذیری داده‌های روزانه زیاد است و این مسئله باعث کاهش کارایی تشخیص ناهمگنی می‌شود، ابتدا نقاط ناهمگنی در داده‌های ماهانه مشخص شد. برای بررسی همگنی داده‌های ماهانه، پس از اجرای تابع dd2m در بسته کلایماتول نرم‌افزار R و تجمیع داده‌های روزانه و به‌دست آوردن داده‌های ماهانه، ابتدا تابع همگن‌سازی با حالت اکتشافی (exploratory mode) برای کشف و تعیین مقادیر مناسب آماره SNHT اجرا شد. پس از مشخص شدن مقادیر مناسب آماره SNHT، دوباره تابع همگن‌سازی با مقادیر تعیین‌شده اجرا شد. در مرحله

ایستگاه‌های نواحی بیابانی و نیمه‌بیابانی نشان داده شده است. برای هر ایستگاه، نمودار بالا، سری اصلی (رنگ سیاه) و سری‌های اصلاح‌شده (سایر رنگ‌ها نظیر سبز و قرمز) و نمودار پایین، میزان اصلاحات انجام‌شده برای هر یک از سری‌ها را نشان می‌دهد. سری‌های همگن‌شده در بانک داده ذخیره می‌شوند و بسته به اینکه در تحقیقات بعدی برای چه منظوری بخواهند بررسی شوند، استفاده خواهند شد.

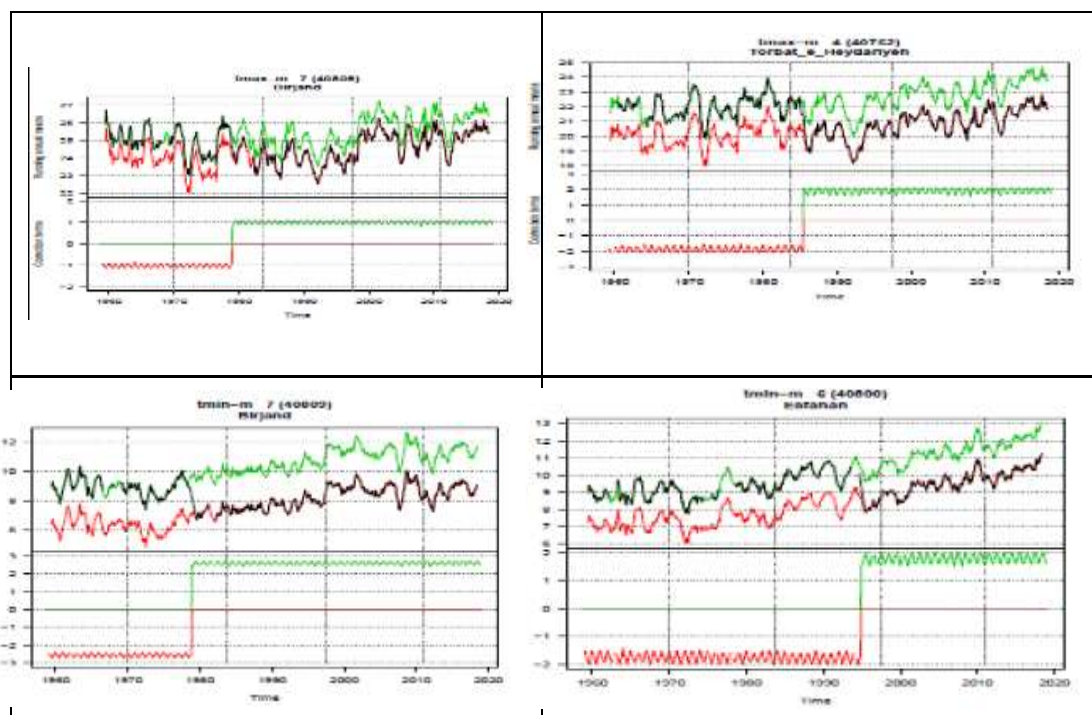
#### ۴-۱-۲-۴ بارش

بافت‌نگار داده‌ها در شکل ۱۱ نشان می‌دهد توزیع داده‌ها نرمال نیست و چولگی به راست دارد و بیشتر داده‌ها صفر هستند که با در نظر داشتن این موضوع که ایستگاه‌ها در ناحیه بیابانی و نیمه‌بیابانی قرار دارند، دور از انتظار نبود. شکل ۱۲ بافت‌نگار بی‌هنجاری استاندارد شده بارش نواحی ایستگاه‌های واقع در خوشه اول را نشان می‌دهد. همان‌طور که از شکل مشخص است، بی‌هنجاری‌های استاندارد شده با مقادیر بیش از ۴۰+ و کمتر از ۲۰-، داده پرت تشخیص داده شدند.

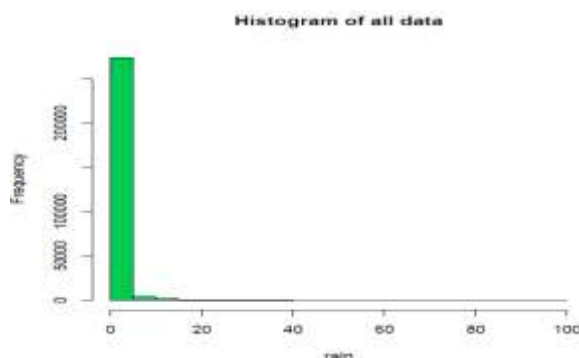
توجه به کامل نبودن بانک فراداده ایستگاه‌های هواشناسی کشور، به‌نظر می‌رسد برای پیدا کردن دلیل قطعی این جهش‌ها در سری زمانی این ایستگاه‌ها، نیاز به بررسی و تهیه فراداده دقیق این ایستگاه‌ها در سازمان هواشناسی وجود دارد؛ بنابراین برای قضاوت درباره منشأ این جهش‌ها به مشاهده و بررسی سری‌های زمانی در ایستگاه‌های مجاور و سری‌های مرجع پرداخته شد. نظر به اینکه در تاریخ‌های مذکور، شکستگی مشهودی در سری‌های زمانی ایستگاه‌های اطراف مشاهده نشد، نتیجه گرفته شد که جهش‌های مذکور در این ایستگاه‌ها نمی‌تواند منشأ اقلیمی داشته باشد و جهش‌های غیراقلیمی در نظر گرفته شدند. در جدول ۳ مشخصات جهش‌های شناسایی شده غیراقلیمی پارامتر دما در ایستگاه‌های خوشه اول در دوره ۲۰۱۸-۱۹۶۰ نشان داده شده است. در مرحله سوم، برای همگن‌سازی داده‌های روزانه، نقاط شکستگی ماهانه در نظر گرفته شد. در این مرحله، داده‌های پرت یا داده‌هایی برآورد شدند که به سری‌های دیگر منتقل شده بودند. در شکل ۱۰ سری‌های کامل بازسازی شده با استفاده از زیردوره‌های همگن مربوط به برخی

جدول ۳. جهش‌های شناسایی شده غیراقلیمی پارامتر دما در ایستگاه‌های خوشه اول در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	سال جهش	دلیل احتمالی جهش
۱	سبزوار	۴۰۷۴۳	۱۹۷۲	جابه‌جایی ایستگاه
			۲۰۰۴	جابه‌جایی ایستگاه
			۲۰۰۸	جابه‌جایی ایستگاه
۲	مشهد	۴۰۷۴۵	۱۹۶۹	-
			۱۹۸۹	جابه‌جایی ایستگاه
۳	تربت حیدریه	۴۰۷۶۲	۱۹۸۵	جابه‌جایی ایستگاه
۴	اصفهان	۴۰۸۰۰	۱۹۹۴	جابه‌جایی ایستگاه
۵	بیرجند	۴۰۸۰۹	۱۹۷۸	جابه‌جایی ایستگاه
۶	زابل	۴۰۸۲۹	۱۹۸۴	-
۷	شیراز	۴۰۸۴۸	۱۹۸۴	جابه‌جایی ایستگاه
			۱۹۹۳	-
			۲۰۰۴	جابه‌جایی ایستگاه



شکل ۱۰. سری‌های کامل بازسازی شده با استفاده از زیردوره‌های همگن مربوط به ایستگاه‌های بیرجند، تربت حیدریه و اصفهان (۲۰۱۸-۱۹۶۰).



شکل ۱۱. بافت‌نگار داده‌های بارش ایستگاه‌های خوشه اول در دوره ۲۰۱۸-۱۹۶۰.

به‌طور دقیق بررسی شدند. داده‌های پرت نهایی تایید شده در جدول ۴ نشان داده شده‌اند.

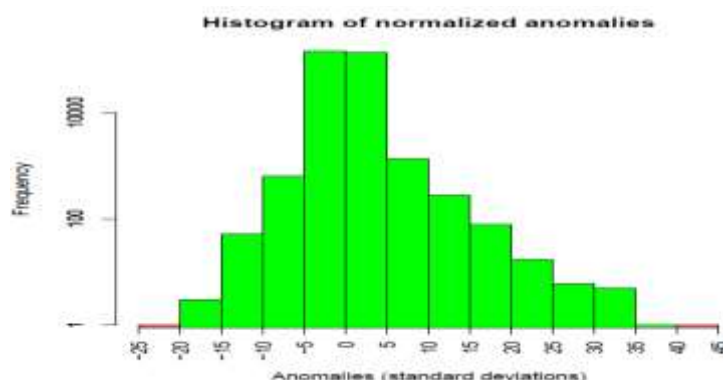
#### ۴-۲-۱-۲-۱ بررسی نقاط ناهمگنی و همگن سازی داده‌های بارش

بررسی همگن بودن سری‌های زمانی داده‌های بارش با توجه به تغییرپذیری خیلی زیادی که دارد، بسیار دشوارتر از بررسی سری‌های زمانی داده‌های دما است. در بررسی

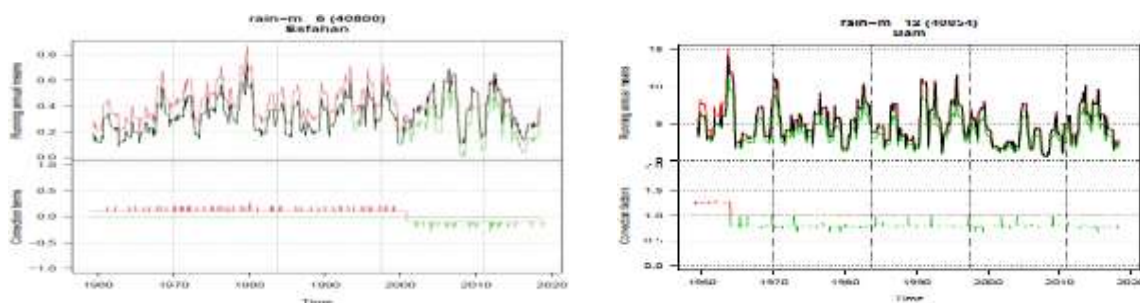
از آنجا که ماهیت رفتار بارش با دما متفاوت است و تغییرپذیری آن نسبت به دما بیشتر است، بررسی وضعیت دقیق جوئی در روزهایی که داده بارش آن پرت تشخیص داده شده است از اهمیت بسزایی برخوردار است؛ بنابراین وضعیت جوئی روزهای مورد نظر از جمله میزان بارش شش ساعته و همخوانی آن با بارش گزارش شده ۲۴ ساعته، وضعیت ابرناکی، پدیده‌ها، میزان رطوبت نسبی و... در روزهای همراه با داده پرت و روزهای قبل و بعد از آن

تشخیص داده نشد. به این ترتیب نقاط شکستگی برای بارش در ایستگاه‌های خوشه اول مشخص شدند که در جدول ۵ نشان داده شده است. در مرحله سوم، همه داده‌های گم‌شده برآورد شدند که شامل داده‌های پرت یا داده‌هایی می‌شود که به سری‌های دیگر منتقل شده بودند. برای برآورد، از همان شکسته‌های خود سری به‌عنوان سری‌های مرجع استفاده می‌شود. سری‌های همگن‌شده در بانک داده ذخیره می‌شوند تا در آینده بر اساس کاربرد آنها انتخاب شوند. برای همگن‌سازی داده‌های روزانه، نقاط شکستگی ماهانه در نظر گرفته شدند. شکل ۱۳ سری‌های کامل بازسازی‌شده بارش را از زیردوره‌های همگن برای ایستگاه‌های فوق نشان می‌دهد.

دما، در خیلی از موارد، همگنی یا ناهمگنی سری با بررسی نمودارهای بی‌هنجاری استاندارد شده تقریباً مشخص می‌شود، اما در بارش وضعیت این گونه نیست. با توجه به تغییرپذیری بسیار زیاد داده‌های روزانه بارش، داده‌های ماهانه بارش بررسی شدند. همانند دما ابتدا آزمون نرمال استاندارد برای پنجره‌های همپوشان اجرا شد. در مرحله دوم، آزمون نرمال استاندارد روی سری‌های کامل اجرا شد و همانند مرحله قبل، در هر نقطه‌ای که بیشینه آماره SNHT بیشتر از حد آستانه بود، سری شکسته شد و همه داده‌ها قبل از نقطه شکستگی به سری دیگری با همان طول و عرض جغرافیایی منتقل شدند. این فرایند برای سری‌های جدید اجرا شد تا زمانی که دیگر نقطه ناهمگنی



شکل ۱۲. بافت‌نگار بی‌هنجاری استاندارد شده بارش ایستگاه‌های خوشه اول در دوره ۱۹۶۰-۲۰۱۸.



شکل ۱۳. سری‌های کامل بازسازی‌شده بارش از زیردوره‌های همگن برای ایستگاه‌های بم و اصفهان دوره ۱۹۶۰-۲۰۱۸.

جدول ۴. داده‌های پرت تأییدشده بارش پس از بررسی‌های هواشناختی در ایستگاه‌های خوشه اول در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآوردشده
۱	یزد	۴۰۸۲۱	۱۹۷۹/۱۲/۲۱	۴۸	۳/۶
۲	زابل	۴۰۸۲۹	۱۹۶۳/۲/۱۴	۳۹	۱/۱
۳	زابل	۴۰۸۲۹	۱۹۸۰/۱۰/۲۸	۳۲	۰/۷
۴	زابل	۴۰۸۲۹	۱۹۹۴/۳/۲	۴۵	۱/۳
۵	کرمان	۴۰۸۴۱	۱۹۷۴/۵/۱۱	۶۰	۰/۶

جدول ۵. جهش‌های شناسایی شده غیراقليمی پارامتر بارش در نواحی بیابانی و نیمه‌بیابانی در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	سال جهش	دلیل احتمالی جهش
۱	اصفهان	۴۰۸۰۰	۲۰۰۰	-
۲	بم	۴۰۸۸۵۴	۱۹۶۴	-

۲۱۱ میلیمتر بارش ثبت شده است که با توجه به نبود سایر اطلاعات جوئی به جز داده‌های بارش و همچنین نبود داده‌های سایر ایستگاه‌های خوشه، تصمیم‌گیری درباره پرت بودن داده مورد نظر مشکل است، به‌خصوص با توجه به گزارش شکل‌گیری طوفان جنب حاره‌ای در ۲۰ ژوئن ۱۹۷۹ در منطقه مصیره کشور عمان، احتمال وجود بارش‌های سنگین در دریای عمان و خلیج فارس دور از انتظار نبوده است.

نقاط شکستگی شناسایی شده سری‌های دما در نواحی جنوب در جدول ۸ نشان داده شده است. برای ایستگاه‌های نواحی جنوب، نقطه شکستگی در سری زمانی بارش تشخیص داده نشد.

با توجه به اینکه ارائه جزئیات تحلیل اکتشافی برای همه نواحی، مقاله را بسیار طولانی می‌کند؛ بنابراین در ادامه تنها به ذکر نتایج و خروجی سایر خوشه‌ها پرداخته و از ذکر جزئیات تحلیل خودداری شده است.

#### ۴-۲-۲ خوشه دوم

خوشه دوم شامل سواحل دریای عمان، خلیج فارس و جلگه خوزستان است و مناطقی از جنوب شرق ایران را نیز دربرمی‌گیرد (علیچانی، ۱۳۸۵). داده‌های پرت ناحیه جنوب برای دمای بیشینه و کمینه به ترتیب در جداول ۶ و ۷ نشان داده شده است.

در تاریخ ۱۹۷۹/۱/۲۴ در سامانه مسیر سازمان هواشناسی، برای ایستگاه بندر عباس با کد ۴۰۸۷۵، میزان

جدول ۶. داده‌های پرت دمای بیشینه ایستگاه‌های خوشه دوم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآوردشده	بی‌هنجاری استانداردشده
۱	آبادان	۴۰۸۳۱	۱۹۶۴/۱/۱۹	۱۴	۴/۹	۶/۰۷

جدول ۷. داده‌های پرت دمای کمینه ایستگاه‌های خوشه دوم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآوردشده	بی‌هنجاری استانداردشده
۱	چابهار	۴۰۸۹۸	۲۰۱۷/۲/۲۷	۱/۱	۱۷	-۶/۹۶
۲	چابهار	۴۰۸۹۸	۲۰۱۷/۹/۳	۲/۱۶	۲۸/۷	-۱۱/۳۷

جدول ۸. جهش های شناسایی شده غیراقلیمی پارامتر دما در خوشه دوم در دوره ۲۰۱۸-۱۹۶۰

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی اول	دلیل شکستگی	تاریخ شکستگی دوم	دلیل شکستگی	تاریخ شکستگی سوم	دلیل شکستگی
۱	بوشهر	۴۰۸۵۸	۱۹۶۲	جابه جایی ایستگاه	۱۹۷۲	جابه جایی ایستگاه	۲۰۰۲	-
۲	اهواز	۴۰۸۱۱	۱۹۸۴	جابه جایی ایستگاه	-	-	-	-
۳	چابهار	۴۰۸۷۸	۱۹۸۵	-	-	-	-	-

## ۴-۲-۳ خوشه سوم

این خوشه شامل کوهپایه های جنوبی البرز، ارتفاعات خراسان و کوهپایه های زاگرس است (علیچانی، ۱۳۸۵). اقلیم بیشتر این استان ها گرم و خشک است اما در مجاورت کوهستان های سرد و نیمه مرطوب قرار دارند. مناطق با ارتفاع بیشتر، سرد هستند و زمستان های طولانی دارند (رحیم زاده و نساجی زواره، ۲۰۱۳).

داده های پرت ایستگاه های واقع در این خوشه برای دمای بیشینه و بارش به ترتیب در جداول ۹ و ۱۰ نشان داده شده است. دمای کمینه در این خوشه فاقد داده پرت بود. نقاط شکستگی شناسایی شده سری های زمانی دما در ایستگاه های این خوشه در جدول ۱۱ و برای پارامتر بارش نیز در جدول ۱۲ نشان داده شده است.

جدول ۹. داده های پرت دمای بیشینه ایستگاه های خوشه سوم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی هنجاری استاندارد شده
۱	اراک	۴۷۶۹	۲۰۰۰/۲/۵	-۸/۴	۸/۵	-۸/۷

جدول ۱۰. داده های پرت بارش روزانه ایستگاه های خوشه سوم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی هنجاری استاندارد شده
۱	خرم آباد	۴۰۷۸۲	۲۰۱۵/۱۱/۱	۸۸	۲/۱	۲۱/۵۵

جدول ۱۱. جهش های شناسایی شده غیراقلیمی پارامتر دما در ایستگاه های خوشه سوم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی اول	دلیل شکستگی	تاریخ شکستگی دوم	دلیل شکستگی
۱	قزوین	۴۰۷۳۱	۱۹۷۴	جابه جایی ایستگاه	-	-
۲	خرم آباد	۴۰۷۸۲	۱۹۷۹	جابه جایی ایستگاه	-	-
۳	شهرکرد	۴۰۷۹۸	۱۹۸۴	-	۲۰۰۳	جابه جایی ایستگاه

جدول ۱۲. جهش های شناسایی شده غیراقلیمی پارامتر بارش در ایستگاه های خوشه سوم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی	دلیل شکستگی
۱	کرمانشاه	۴۰۷۶۶	۱۹۷۶	-
۲	اراک	۴۰۷۶۹	۱۹۸۶	-



#### ۴-۲-۴ خوشه چهارم

خوشه چهارم از آستارا تا گرگان و از خط‌الرأس کوه‌های البرز تا ساحل خزر است (علیجانی، ۱۳۸۵). حواشی دریای خزر به‌ویژه استان‌های گیلان، مازندران و گلستان هوای گرم و مرطوب و اقلیم به‌نسبت متفاوتی نسبت به بقیه نقاط کشور دارند. کوه‌های بلند در غرب، جنوب و شرق این منطقه ساحلی کوچک مانند یک مانع عمل می‌کند و از نفوذ آب‌وهوای گرم و مرطوب جلوگیری می‌کند (رحیم‌زاده و نساجی‌زواره، ۲۰۱۳).

داده‌های پرت نواحی خزری برای دمای بیشینه و کمینه و بارش به‌ترتیب در جداول ۱۳ تا ۱۵ نشان داده شده است. در تاریخ ۱۳۰۹/۰۹/۱۷، بارش ۲۴ ساعته ایستگاه هواشناسی رامسر با کد ۴۰۷۳۲ در سامانه، ۲۶۰ میلیمتر ثبت شده است که بیشتر ناشی از بارشی است که در ساعت ۰۳ در بازه زمانی ۱۲ ساعته مقدار ۱۷۰ میلیمتر بارش را ثبت کرده است. به دلیل نبود اطلاعات جوئی ۱۲ ساعت قبل از ساعت ۰۳، قضاوت درباره این میزان بارش ۱۲ ساعته سخت است.

نقاط شکستگی شناسایی شده سری‌های دما در نواحی کوهپایه‌ای در بازه زمانی ۲۰۱۸-۱۹۶۰ در جدول ۱۶ نشان داده شده است. در ایستگاه‌های ناحیه خزری، سری ناهمگن در سری زمانی ۲۰۱۸-۱۹۶۰ برای پارامتر بارش تشخیص داده نشد.

#### ۴-۲-۵ خوشه پنجم

شمال و شمال غرب ایران شامل استان‌های آذربایجان غربی و شرقی، اردبیل و استان‌های واقع در منطقه

کوهستانی زاگرس شامل استان‌های کردستان، همدان و زنجان تجربه دماهای پایین و زمستان‌های سرد و خشن‌تری را نسبت به بقیه نقاط کشور دارند (رحیم‌زاده و نساجی‌زواره، ۲۰۱۳). این ناحیه مناطق گسترده کوهستانی و فلات‌های مرتفع را دربرمی‌گیرد. در غرب کشور، فلات آذربایجان و ارتفاعات کردستان از کوه‌های تالش تا مرز عراق و ترکیه در این ناحیه قرار دارد (علیجانی، ۱۳۸۵). این ناحیه کوهستانی در بیشتر جاها بر منحنی ترازنمای ۱۵۰۰ متر منطبق است ولی باید متذکر شد شهرها و به‌تبع آن، ایستگاه‌های هواشناسی، در نقاط پست این ناحیه قرار دارند (علیجانی، ۱۳۸۵).

داده‌های پرت ایستگاه‌های واقع در این نواحی برای دمای بیشینه و کمینه به‌ترتیب در جداول ۱۷ و ۱۸ نشان داده شده است. درباره داده‌های پرت بارش روزانه، در ایستگاه‌های این خوشه برای ایستگاه‌های تبریز در تاریخ‌های ۱۹۶۳/۱۱/۲۴، ۱۹۷۷/۳/۳ و ۱۹۸۱/۴/۱۷ به‌ترتیب ۳۷، ۶۳ و ۵۳ میلیمتر و برای ایستگاه ارومیه نیز در تاریخ‌های ۱۹۶۷/۴/۱۳، ۱۹۹۳/۵/۱۲ و ۲۰۰۶/۱۰/۲۹ به‌ترتیب ۵۴، ۶۱ و ۵۵ میلیمتر بارش ثبت شده است. نرم‌افزار، این داده‌ها را پرت تشخیص داده بود، اما به دلیل نبود اطلاعات جوئی در تاریخ‌های مورد نظر در سامانه مسیر سازمان هواشناسی، امکان قضاوت قطعی درباره پرت بودن داده‌های مورد نظر وجود نداشت.

نقاط شکستگی شناسایی شده سری‌های زمانی دما و بارش در این نواحی به‌ترتیب در جداول ۱۹ و ۲۰ نشان داده شده است.

جدول ۱۳. داده‌های پرت دمای بیشینه ایستگاه‌های خوشه چهارم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی‌هنجاری استاندارد شده
۱	بندر انزلی	۴۰۷۱۸	۱۹۷۰/۰۲/۱۴	۲۷	۱۱/۵	۸/۵۱
۲	رامسر	۴۰۷۳۲	۱۹۷۷/۴/۲۴	۳۰	۱۷/۶	۸/۸۳

جدول ۱۴. داده‌های پرت دمای کمینه ایستگاه‌های خوشه چهارم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی‌هنجاری استاندارد شده
۱	رشت	۴۰۷۱۹	۱۹۷۸/۱۲/۲	-۷	۸/۴	-۸/۶۷
۲	رشت	۴۰۷۱۹	۱۹۷۸/۱۲/۱۰	-۹	۸/۱	-۹/۵۸

جدول ۱۵. داده‌های پرت بارش روزانه ایستگاه‌های خوشه چهارم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی‌هنجاری استاندارد شده
۱	انزلی	۴۰۷۱۸	۱۹۶۰/۱۱/۲۰	۲۴۸	۵۸/۳	۲۱/۹
۲	رشت	۴۰۷۱۹	۱۹۸۵/۱۰/۱۰	۱۵۰/۹	۱۹/۴	۲۰/۳۱

جدول ۱۶. جهش‌های شناسایی شده غیراقليمی پارامتر دما در ایستگاه‌های خوشه چهارم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی اول	دلیل شکستگی	تاریخ شکستگی دوم	دلیل شکستگی	تاریخ شکستگی سوم	دلیل شکستگی
۱	رشت	۴۰۷۱۹	۱۹۶۹	-	۱۹۷۳	-	۱۹۹۵	جابه‌جایی ایستگاه
۲	گرگان	۴۰۷۳۸	۱۹۸۴	جابه‌جایی ایستگاه	۲۰۰۷	جابه‌جایی ایستگاه	-	-
۳	بندر انزلی	۴۰۷۱۸	۱۹۸۶	جابه‌جایی ایستگاه	۲۰۱۰	جابه‌جایی ایستگاه	-	-

جدول ۱۷. داده‌های پرت دمای بیشینه ایستگاه‌های خوشه پنجم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی‌هنجاری استاندارد شده
۱	ارومیه	۴۰۷۱۲	۱۹۷۲/۲/۱	۱۱	-۶/۶	۹/۸۹
۲	سنندج	۴۰۷۴۷	۱۹۶۳/۱۲/۲۵	۱۳	-۰/۴	۷/۰۶

جدول ۱۸. داده‌های پرت دمای کمینه ایستگاه‌های خوشه پنجم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	بی‌هنجاری استاندارد شده
۱	خوی	۴۰۷۰۳	۱۹۶۲/۱/۸	-۲۰	-۱/۲	-۷/۲۱

جدول ۱۹. جهش‌های شناسایی شده غیراقليمی پارامتر دما در خوشه پنجم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی اول	دلیل شکستگی	تاریخ شکستگی دوم	دلیل شکستگی
۱	ارومیه	۴۰۷۱۲	۲۰۰۳	جابه‌جایی ایستگاه	-	-
۲	سقز	۴۰۷۲۷	۱۹۸۷	-	-	-
۳	زنجان	۴۰۷۲۹	۱۹۸۳	جابه‌جایی ایستگاه	۲۰۱۴	-
۴	سنندج	۴۰۷۴۷	۱۹۸۸	جابه‌جایی ایستگاه	-	-

جدول ۲۰. جهش‌های شناسایی شده غیراقليمی پارامتر بارش در خوشه پنجم در دوره ۲۰۱۸-۱۹۶۰.

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی	دلیل شکستگی
۱	همدان (نورژه)	۴۰۷۶۷	۱۹۶۴	-

## ۵ نتیجه‌گیری

با بررسی‌های انجام‌شده مشخص شد ایستگاه‌های هواشناسی کشور، در دوره زمانی ۶۰ ساله برای دمای بیشینه روزانه، به‌طور متوسط ۰.۵٪؛ دمای کمینه، بیش از ۰.۷٪ و برای بارش روزانه حدود ۰.۲٪ داده گم‌شده دارند. با توجه به وجود خطاهای زیاد در داده‌های سال ۱۹۵۹، پس از حذف داده‌های این سال در این سری‌های زمانی، درصد داده‌های گم‌شده دمای بیشینه و کمینه و بارش روزانه به‌طور متوسط در حدود ۰/۳ درصد کاهش یافت. برای پارامتر دمای بیشینه تعداد هفت نمونه و برای دمای کمینه نیز هفت نمونه داده پرت ناشی از خطا تشخیص داده شد. برای پارامتر بارش، در مرحله اول شانزده داده پرت تشخیص داده شد که در بررسی ثانویه، هشت مورد از آنها خطا شناخته شدند.

در این دوره زمانی بدون در نظر گرفتن داده‌های سال ۱۹۵۹ و بدون احتساب ایستگاه طبس، برای پارامتر دمای روزانه، از بین ۳۶ ایستگاه، ۱۶ ایستگاه همگن تشخیص داده شد و در ۲۰ ایستگاه، یک یا دو و در برخی ایستگاه‌ها تا سه نقطه شکستگی در سری زمانی مشاهده شد. با بررسی فراداده مشخص شد در ۱۷ ایستگاه جابه‌جایی صورت گرفته است. برای پارامتر بارش نیز در همین دوره ۵ ایستگاه ناهمگن شناخته شدند.

نتیجه بررسی سایر منابع این بود که فراداده ایستگاه شامل ثبت ساخت‌وساز در ایستگاه‌ها، تغییرات محل ایستگاه، تغییر تجهیزات و ادوات ایستگاه، تغییرات در روش‌های مشاهداتی و... به‌طور میانگین تنها برای ۱۰ تا ۳۰٪ ایستگاه‌ها تهیه شده است (رحم‌دل و همکاران، ۱۴۰۰)؛ بنابراین با توجه به نبود بانک جامع فراداده، در برخی ایستگاه‌ها دلیل قاطعی برای بسیاری از شکستگی‌ها و ناهمگنی در سری‌های زمانی پیدا نشد. گفتنی است درباره تلفیق، نتیجه‌گیری و لزوم تهیه و به‌روزرسانی بانک فراداده و شناسنامه اقلیمی انواع ایستگاه‌های هواشناسی

کشور، پس از بررسی و پیگیری‌های فراوان، اصلاح و گردآوری فراداده این ایستگاه‌ها، پس از یک وقفه به‌نسبت طولانی، از تیرماه سال ۱۳۹۹ دوباره در دستور کار سازمان هواشناسی قرار گرفت و هم‌اکنون فراداده به‌نسبت کاملی از وضعیت حال حاضر ایستگاه‌های هواشناسی کشور تهیه شده است. البته درباره وضعیت گذشته ایستگاه‌ها و همچنین به‌روزرسانی شناسنامه در سطح فعالیت‌های روزمره ایستگاه‌ها همچنان مسائل اساسی وجود دارد (رحم‌دل، ۱۴۰۰).

## ۶ پیشنهادها

الف) با توجه به تکمیل نبودن بانک فراداده ایستگاه‌های هواشناسی کشور، در زمینه تهیه و تکمیل بانک جامع فراداده ایستگاه‌ها، به‌خصوص درباره گذشته ایستگاه‌ها و همچنین به‌روزرسانی نحوه فعالیت‌های روزمره در شناسنامه ایستگاه‌ها، ادارات کل هواشناسی استان‌های مربوطه بررسی‌های دقیقی را با همکاری و نظارت سازمان هواشناسی انجام دهند.

ب) با توجه به نقص‌های فراوانی که در ارزیابی‌های صورت گرفته برای سری‌های دما و بارش ایستگاه‌ها مشاهده شد، برای هر مطالعه‌ای درباره داده‌های مشاهداتی ایستگاه‌ها، از داده‌هایی استفاده شود که همگن شده‌اند و کنترل کیفیت آماری برای آنها انجام شده است.

ج) در کارهای مطالعاتی آینده جهت تأیید، تکمیل و بررسی جهش‌های غیراقلیمی و همچنین تفکیک جهش‌های اقلیمی و غیراقلیمی، از سایر روش‌های آماری نظیر آزمون مقایسه زوجی (Pairwise detection test) و نرم‌افزارهایی مانند هومر که محصول سازمان همکاری‌های اروپا است، برای تشخیص ناهمگنی و همگن‌سازی داده‌های سالانه و ماهانه استفاده شود.

د) علاوه بر پارامترهای دما و بارش، بررسی همگنی و تحلیل‌های اکتشافی برای سری داده‌های سایر پارامترهای

- رابطه با پارامتر بارش در مناطق خشک ایران: مجله فیزیک زمین و فضا، ۲(۴۷)، ۳۱۵-۳۳۲.
- مسعودیان، س. ا.، ۱۳۸۲، نواحی اقلیمی ایران: مجله جغرافیا و توسعه، ۱(۲)، ۱۷۱-۱۸۴.
- Alexanderson, H., and Mobeg, A., 1997, Homogenization of Swedish temperature data part I: Homogeneity test for linear trends: *International Journal of Climatology*, **17**, 25-34.
- Guijarro, J. A., 2014, User's guide to Climatol. State Meteorological Agency (AEMET), Balearic Islands Office, Spain.
- Guijarro, J. A., 2018, Homogenization of climatic series with Climatol: [http://www.climatol.eu/homog\\_climatol-en.pdf](http://www.climatol.eu/homog_climatol-en.pdf), Accessed 27 February 2021.
- Karimi, M., Vicente-Serrano, S. M., Reig, F., et al., 2020, Recent trends in atmospheric evaporative demand in Southwest Iran: implications for change in drought severity: *Theoretical and Applied Climatology*, **142**, 945-958, <https://doi.org/10.1007/s00704-020-03349-3>.
- Lin, N. J., Abd Aziz, S., Feng, H. Y., Wayayok, A., and Rowshon Kamal, M., 2015, Homogeneity analysis of rainfall in Kelantan, Malaysia: *Jurnal Teknologi*, **76**(15), 1-6.
- Mestre, O., Domonkos, P., Picard, F., and Auer, I., 2013, HOMER: Homogenization software - methods and applications: *Időjárás, Quarterly Journal of the Hungarian Meteorological Service*, **117**(1), 47-67.
- Ming, K., and Fadhilah, Y., 2012, Homogeneity tests on daily rainfall series in peninsular Malaysia: *International Journal of Contemporary Mathematical Sciences*, **7**(1), 9-22.
- Omar, M. A., Mahmood Agha, S., Çağatay Bağçacı, S., and Sarlak, N., 2017, Homogeneity analysis of precipitation series in North Iraq: *Journal of Applied Geology and Geophysics*, **3**(5), 57-63.
- Paulhus, J. L. H., and Kohler, M. A., 1952, Interpolation of missing precipitation records: *Monthly Weather Review*, **80**(8), 129-133.
- Rahimzadeh, F., and Nassaj, Z. M., 2013, Effects of adjustment for non-climatic discontinuities on determination of temperature trends and variability over Iran: *International Journal of Climatology*, **34**, 2079-2096.
- Tavakoli, M., and Mahmoudian, A. H., 2011, GIS based considerations for development in تأثیرگذار بر مسائل زیست محیطی نظیر سمت و سرعت باد، رطوبت نسبی و تابش نیز انجام شود.
- منابع**
- بازگیر، س.، عباسی، ف.، اسعدی اسکویی، ا.، حقیقت، م.، رضازاده، پ.، ۱۳۹۷، تحلیل همگنی داده‌های دما و بارش در ایران با رویکرد اقلیمی: نشریه تحلیل فضایی مخاطرات محیطی، ۶(۱)، ۵۱-۷۰.
- جوانشیری، ز.، محمدیان، آ.، پاکدامن، م.، فلامرزی، ی.، کوهی، م.، ۱۳۹۹، کنترل کیفیت و همگن سازی داده‌های روزانه پارامترهای دما و بارش ایستگاه‌های هواشناسی کل کشور: طرح پژوهشی، سازمان هواشناسی کشور.
- دین پژوه، ی.، فاخری، ا.، مقدم، م.، میرنیا، م.، جهانبخش اصل، س.، ۱۳۸۲، پهنه‌بندی اقلیمی ایران با استفاده از تحلیل‌های چندمتغیره برای استفاده در مطالعات کشاورزی: مجله دانش کشاورزی، ۱۳(۱)، ۷۱-۹۰.
- رحم‌دل، م.، ثنائی‌نژاد، ح.، جوانشیری، ز.، ۱۴۰۰، بررسی و مستندسازی اشکالات سامانه پایش داده‌های جوّی در سازمان هواشناسی ایران با رویکرد آسیب شناسی: نشریه پژوهش‌های اقلیم‌شناسی، ۱۲(۴۵)، ۲۵-۴۲.
- رحم‌دل، م.، ۱۴۰۰، مطالعه تفحصی و مستندسازی اشکالات سامانه پایش و تفسیر داده‌های جوّی در ایران: رساله دکتری، دانشگاه فردوسی مشهد.
- رضئی، ط.، ۱۳۹۵، شناسایی مناطق همگن بارشی ایران با استفاده از روش تحلیل مؤلفه‌های اصلی: مجله ژئوفیزیک ایران، ۳(۱۰)، ۱۲۸-۱۴۴.
- علیجانی، ب.، ۱۳۸۵، آب‌وهوای ایران، چاپ هفتم: انتشارات سمت، تهران-ایران.
- علیجانی، ب.، ۱۳۹۰، اقلیم‌شناسی سینوپتیک، چاپ چهارم: انتشارات سمت، تهران-ایران.
- کوثری، م.، اسماعیل‌زاده حسینی، م.، میری، م.، ۱۴۰۰، بررسی کارایی روش‌های بازسازی نواقص آماری در

Homogenization: WMO/TD No. 1186.  
World Meteorological Organization, 2017, WMO  
Guidelines on the Calculation of Climate  
Normals: WMO-No. 1203.

different Iranian climatic regions: Journal of  
American Science, 7(4), 182-187.  
World Meteorological Organization, 2003,  
Guidelines on Climate Metadata and

## Exploratory Analysis and Detection of In-homogeneities in Temperature and Precipitation Series of Meteorological Stations in Iran (Period 1959-2018)

Mohsen Rahmdel<sup>1\*</sup>, Zohreh Javanshiri<sup>2</sup> and Hossein Sanaeinejad<sup>3</sup>

<sup>1</sup> Ph.D. of Agricultural Meteorology, Ferdowsi University, Mashhad, Iran

<sup>2</sup> Assistant Professor, Climatological Research Institute, Mashhad, Iran

<sup>3</sup> Professor, Department of Water Science and Engineering, Faculty of Agriculture, Ferdowsi University, Mashhad, Iran

(Received: 07 October 2021, Accepted: 25 January 2022)

### Summary

In-situ observations underlie a wide range of planning, applied studies and modeling in various fields and sciences. Using this data in studies and planning without ensuring the accuracy and homogeneity of them can lead to uncertainty in the results. The major problems that researchers face are the poor data quality, missing data, outliers and in-homogeneity in time series. Therefore, in this paper, the minimum and maximum daily temperature series and daily rainfall series were analyzed at 37 weather stations in Iran for outliers and homogeneity over the period 1959-2018. In this regard, the World Meteorological Organization in cooperation with the Climatology Commission has provided instructions for data homogenization (e.g. WMO/TD Document No. 1186, Guidelines on climate metadata and homogenization; WMO Document No. 1203, WMO Guidelines on the Calculation of Climate Normal).

The main steps in data homogenization are:

- 1- Metadata analysis and data quality control;
- 2- Creating a reference series;
- 3- Detection of break points;
- 4- Data correction.

To do this, in the initial clustering, according to the previous activities and studies in this field which have mostly used empirical and quantitative methods, including principal components and cluster analysis, Iran was divided into 5 clusters based on the climatic characteristics. After initial clustering, the daily maximum and minimum temperatures and daily rainfall series were statistically analyzed using SPSS software and the percentage of missing data was determined for each station. Then, Climatol package in R software was used to study outliers, in-homogeneity and homogenization. In each cluster, the series are re-clustered based on the variability of desired parameter, and for each station, the other stations with similar variability belonging to that cluster are considered as reference stations.

Based on this algorithm, first the desired series is estimated and standardized by reference series using type (II) regression method. After estimating the series, the standardized anomaly series are calculated, in which the difference between the observed and estimated values is calculated. For detecting outliers, two steps were followed. Original data corresponding to the standardized anomalies greater than the prescribed thresholds were detected as outliers. In the second step, in order to confirm the outliers, the detected outliers in the first step were compared with the values of the days before and after for temperature series. If they differed significantly, they would be accepted as outliers and deleted. For the precipitation series, the atmospheric condition of the desired dates would be checked. For detection of in-homogeneity, the standard normal homogeneity test (SNHT) was performed on the monthly series. If the SNHT test statistic was greater than the prescribed threshold, the series was split at the point of the maximum SNHT and all the data before the break point were transferred to a new series with the same geographic coordinates. This process was repeated until all series were homogeneous. If break points were confirmed by metadata, they would then be accepted as non-climatic breaks. Finally, all the missing data in every homogenous series are estimated using same estimation procedure. The only difference is that the fragments of series are used as references.

\*Corresponding author:

mohsen.rahmdel@mail.um.ac.ir

Given the large number of missing and suspicious data in 1959, we considered the beginning of the statistical period from 1960. Investigations have shown that on some dates, all stations in a cluster lack data, possibly due to glitches in the MESSIR-CLIM system of Meteorological Organization through which data is received. In such cases, the average data of the days before and after the mentioned stations was used to estimate the data on that dates. MESSIR-CLIM is the database of IRIMO including climatic database management system that is based on PostgreSQL. The main functions of the system receive and store all kinds of weather and climatic data. The system is able to collect and process massive amounts of information and provide meteorological products (such as charts, maps, tables, and reports).

For the maximum temperature in the stations of the Caspian region (Cluster 4), 18 dates and for the minimum temperature, 13 dates in the mountainous areas (Cluster 5), in all the cluster stations were missing data.

The maximum and minimum temperature and daily precipitation series for 37 weather stations of Iran have an average of 5%, 7% and 2% missing values, respectively.

In the 60-year time series (1959-2018) after deleting the 1959 data, the percentage of missing data at maximum and minimum daily temperature and daily precipitation decreased by an average of about 0.3%.

In this time series, excluding 1959 data, 7 outliers were detected for the maximum temperature parameter. For the minimum temperature, this number reached 7 and for the precipitation parameter, 8 outliers were identified. In 8 cases, due to the lack of atmospheric data on the desired dates, it was not possible to make a definitive judgment about the accuracy of precipitation data outliers.

In terms of daily temperature series, with the exception of Tabas station, out of 36 stations, 16 stations were homogeneous and 20 stations had one or two or three breakpoints. For the precipitation parameter, 5 in-homogeneous stations were identified.

Unfortunately, due to the lack of a comprehensive metadata bank, there was no definitive reason for many of these fractures at some stations.

**Keywords:** Precipitation, temperature, in-homogeneity, metadata, Climatol package