

## تحلیل اکتشافی و مطالعه ناهمگنی سری‌های دما و بارش ایستگاه‌های هواشناسی ایران (بازه زمانی ۲۰۱۸-۱۹۸۹)

محسن رحمدل<sup>۱</sup>، سید حسین ثنائی نژاد<sup>۲\*</sup>، زهره جوانشیری<sup>۳</sup> و آزاده محمدیان<sup>۴</sup>

۱. دانشجوی دکتری، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران

۲. استاد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران

۳. استادیار، پژوهشکده اقلیم‌شناسی، مشهد، ایران

۴. کارشناس ارشد، پژوهشکده اقلیم‌شناسی، مشهد، ایران

(دریافت: ۹۹/۱۱/۱۹، پذیرش نهایی: ۱۴۰۰/۳/۴)

### چکیده

داده‌های دیده‌بانی شده در ایستگاه‌های هواشناسی زیربنای گستره وسیعی از برنامه‌ریزی‌ها، مطالعات کاربردی و مدل‌سازی‌ها در زمینه‌ها و علوم مختلف می‌باشند و استفاده از این داده‌ها در مطالعات و برنامه‌ریزی‌ها بدون اطمینان از صحت (Accuracy) و همگن بودن (Homogeneity) داده‌ها می‌تواند منجر به نوعی عدم قطعیت (Uncertainty) در نتایج به‌دست‌آمده شود. بنابراین با توجه به اهمیت پارامترهای دما و بارش، در این مقاله سری‌های زمانی دمای کمینه و بیشینه و بارش روزانه در ایستگاه‌های هواشناسی کشور با رویکرد تحلیل اکتشافی و بررسی ناهمگنی در دوره زمانی ۳۰ ساله (۱۹۸۹-۲۰۱۸) مورد ارزیابی قرار گرفته است. بررسی‌ها نشان داد در سری‌های زمانی ۳۰ ساله از بین ۱۳۴ ایستگاه به‌طور میانگین ایستگاه‌های هواشناسی کشور در مورد دمای بیشینه ۳ درصد، دمای کمینه ۴ درصد، و بارش روزانه دارای ۲ درصد داده گم شده می‌باشند. در این سری زمانی برای پارامتر دمای بیشینه تعداد ۶۳ مورد داده پرت تشخیص داده شد، که ۵۳ مورد آن مربوط به ایستگاه ژئوفیزیک دانشگاه تهران بود. برای دمای کمینه این عدد به ۵۰ مورد رسید که ۱۱ مورد آن متعلق به ایستگاه ژئوفیزیک دانشگاه تهران می‌باشد. برای پارامتر بارش در این دوره تعداد ۱۳ مورد داده پرت تشخیص داده شد، که ۵ مورد آن به ایستگاه ژئوفیزیک دانشگاه تهران مربوط است. برای پارامتر دمای روزانه در این دوره (بدون احتساب ایستگاه ژئوفیزیک)، ۸۹ ایستگاه همگن و ۴۴ ایستگاه دارای یک یا دو نقطه شکستگی بودند و برای پارامتر بارش نیز در همین دوره، ۱۵ ایستگاه ناهمگن شناخته شدند.

واژه‌های کلیدی: همگن‌سازی، الگوریتم کلایماتول، فراداده، داده‌های گم‌شده، داده‌های پرت.

### ۱. مقدمه

کشورهای کمتر توسعه یافته، مناطق کوهستانی و بیابانی با آن مواجه‌اند. عدم رعایت فواصل مناسب در هنگام احداث ایستگاه‌ها، عدم جانمایی مناسب ایستگاه‌ها، خطاهای انسانی در هنگام قرائت و ثبت داده‌ها، خطاهای مربوط به تجهیزات و ادوات اندازه‌گیری، جابه‌جایی ایستگاه‌های هواشناسی، یکسان نبودن روش‌های مشاهداتی و عدم رعایت شرایط و استانداردهای تعریف‌شده، تغییر شرایط استاندارد در ایستگاه‌ها با گذشت زمان، تغییر در نوع ادوات و حس‌گرهای اندازه‌گیری پارامترهای جوی در طی دوره آماری و ... از جمله مشکلاتی است که می‌تواند صحت داده‌ها و همگنی

داده‌ها و پارامترهای هواشناسی حاصل از اندازه‌گیری در انواع ایستگاه‌های مختلف هواشناسی، برای بسیاری از برنامه‌ها و مطالعات کاربردی در علوم مختلف مانند آب‌وهواشناسی، شهرسازی، کشاورزی، مسائل مربوط به هوانوردی، هیدرولوژی، مدل‌سازی در زمینه‌های مختلف و ... کاربردهای فراوان دارد. بنابراین بدون اطمینان از صحت و همگن بودن این داده‌ها نمی‌توان اقدام به استفاده و استخراج نتایج آنها کرد. وجود داده‌های گم‌شده و داده‌های پرت در ایستگاه‌های هواشناسی، همگن نبودن سری زمانی داده‌ها و ... از جمله مشکلاتی است که پژوهشگران در بخش‌های بزرگی از جهان به‌ویژه

سری آنها را تحت تأثیر قرار دهد. در این زمینه سازمان هواشناسی جهانی با همکاری کمیسیون اقلیم‌شناسی (Commission for Climatology) مجموعه‌ای از پروتوکل‌ها و دستورالعمل‌های لازم را برای همگن‌سازی داده‌ها تهیه نموده و در دسترس قرار داده است. به‌عنوان نمونه می‌توان به سند شماره ۱۱۸۶ سازمان جهانی هواشناسی (۲۰۰۳) و سند شماره ۱۲۰۳ سازمان جهانی هواشناسی (۲۰۱۷) که در منابع به آن اشاره شده است.

دما و بارش از مهم‌ترین عناصر جوی و اقلیمی است که می‌تواند وضعیت جوی، هیدرولوژیکی، پوشش گیاهی، کشاورزی، منابع طبیعی و... را در یک منطقه تحت تأثیر قرار دهد و به‌عنوان بخشی از مهم‌ترین داده‌ها در ورودی سامانه‌های جوی و هیدرولوژیکی می‌باشد که مطالعه و مدل‌سازی آنها در سایر علوم نیز لازم است. در این پژوهش با رویکرد تحلیل اکتشافی و آزمون فرض به ارزیابی و همگن‌سازی سری زمانی داده‌های دما و بارش ایستگاه‌های هواشناسی کشور پرداخته شده است.

## ۲. داده‌ها و روش پژوهش

به‌منظور تحلیل داده‌های ایستگاه‌های هواشناسی و با توجه به اهمیت پارامتر دما و بارش و تأثیرگذاری آنها در فرایندهای اقلیمی، داده‌های دمای کمینه و بیشینه روزانه ایستگاه‌ها و همچنین بارش ۲۴ ساعته این ایستگاه‌ها، در سری زمانی ۳۰ ساله (۱۹۸۹-۲۰۱۸) مورد بررسی قرار گرفتند. برای بررسی سری‌های زمانی این داده‌ها، نسبت به تعیین درصد داده‌های گم‌شده، داده‌های پرت، بررسی همگنی سری‌ها به تفکیک ایستگاه بر اساس دستورالعمل‌های سازمان جهانی هواشناسی اقدام شد. سپس دلایل عدم همگنی سری‌ها در برخی از ایستگاه‌هایی که نیاز به همگن‌سازی داشتند با استفاده از فراداده این ایستگاه‌ها مورد بررسی قرار گرفتند.

گام‌های اصلی در همگن‌سازی داده‌ها عبارت‌اند از:

- تحلیل فراداده و کنترل کیفیت داده‌ها

- ساختن سری مرجع

- تشخیص نقاط شکستگی

- اصلاح داده‌ها

ابتدا ایران بر اساس مشخصه‌های آب‌وهوایی به ۵ خوشه تقسیم شد. در این فرایند، داده‌های ایستگاه‌هایی که به لحاظ اقلیمی نسبتاً مشابه بودند در یک خوشه قرار گرفتند تا برای تعیین ایستگاه‌های مرجع وارد نرم‌افزار R شوند. آنگاه با استفاده از الگوریتم کلایماتول (Climatol) در نرم‌افزار R، تعیین سری‌های مرجع برای هر ایستگاه بر اساس تغییرپذیری پارامتر مورد بررسی، در داخل هر خوشه انجام شد. برای خوشه‌بندی ایستگاه‌ها با توجه فعالیت‌ها و مطالعات قبلی که در این زمینه صورت گرفته بود، با استناد به خوشه‌بندی یکی از مقالات، رحیم‌زاده و نساج زواره (۲۰۱۳) که با توجه به نتایج مطالعات گذشته علیجانی (۱۳۹۰)، دین‌پژوه و همکاران (۱۳۸۲)، مسعودیان (۱۳۸۲)، توکلیان و محمدیان (۲۰۱۱)، بدری‌پور (۱۹۹۲) و عمدتاً با استفاده از روش‌های کمی و تجربی شامل مؤلفه‌های اصلی و تحلیل خوشه‌ای اطلاعات اقلیمی ایستگاه‌های هواشناسی صورت گرفته بود، مناطق مختلف کشور در ۵ خوشه اقلیمی تقسیم‌بندی شد.

با توجه به مطالعات علیجانی (۱۳۸۵) مناطق مختلف کشور، به لحاظ آب و هوایی به ۵ منطقه آب و هوایی ذیل تقسیم‌بندی شده است.

۱- ناحیه خزری: از آستارا تا گرگان و از خط الرأس کوه‌های البرز تا ساحل خزر می‌باشد (علیجانی، ۱۳۸۵).  
حواشی دریای خزر به‌ویژه استان‌های گیلان، مازندران و گلستان که دارای هوای گرم و مرطوب و اقلیم نسبتاً متفاوت نسبت به بقیه نقاط کشور می‌باشند. کوه‌های بلند در غرب، جنوب و شرق این منطقه ساحلی کوچک مانند یک مانع عمل می‌کند و از نفوذ آب‌وهوای گرم و مرطوب به مناطق بیرونی آن جلوگیری می‌کند (رحیم‌زاده و نساج زواره، ۲۰۱۳).

۲- ناحیه کوهستانی: مرز بیرونی این ناحیه در همه‌جا تقریباً بر منحنی ترازنمای ۱۵۰۰ متر منطبق است (علیجانی، ۱۳۸۵). شمال و شمال غرب ایران شامل استان‌های

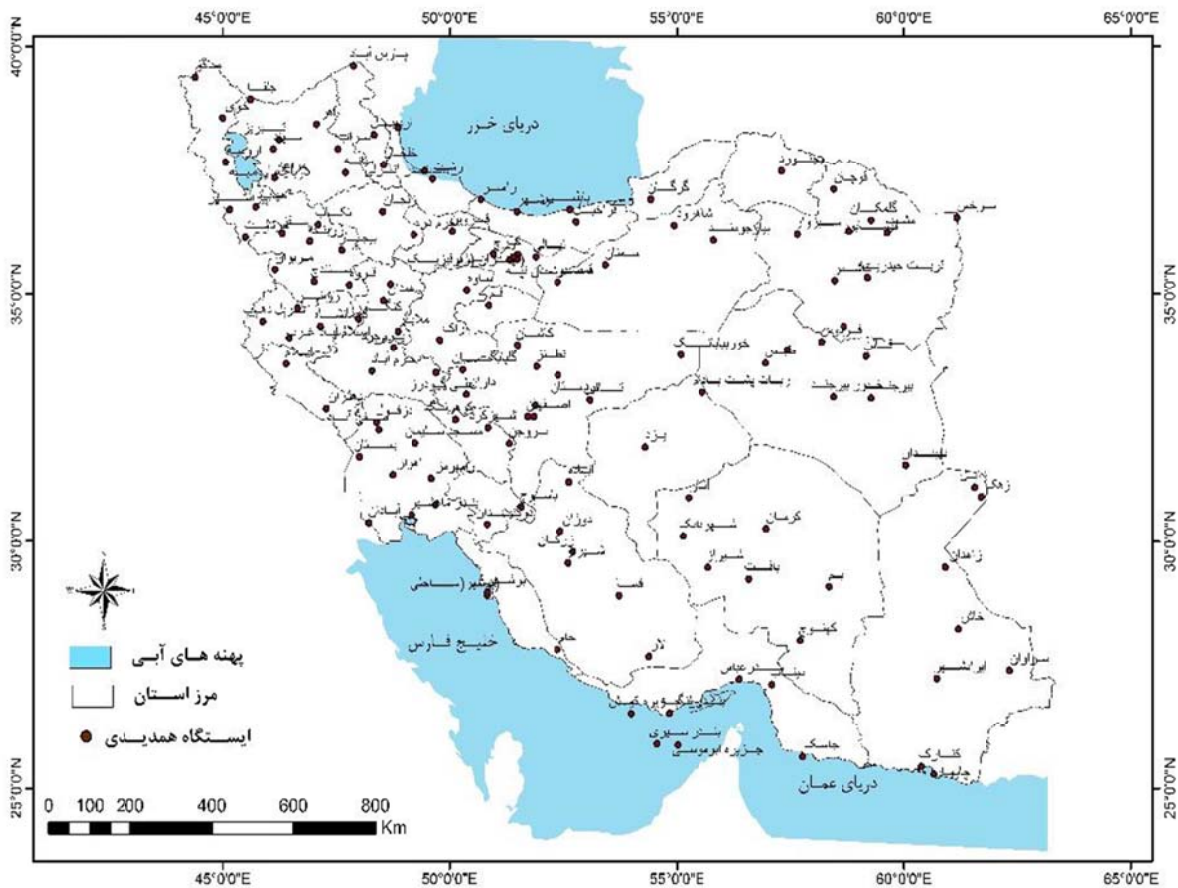
نیز در بر می‌گیرد (علیجانی، ۱۳۸۵).  
 ۵- ناحیه مرکزی (مناطق بیابانی و نیمه‌بیابانی): شامل دشت کویر، دشت لوت و چاله جازموریان است (علیجانی، ۱۳۸۵). مرکز ایران از چندین حوضه بسته تشکیل شده است که در مجموع به‌عنوان فلات مرکزی از آن یاد می‌شود. قسمت شرقی این فلات توسط دو بیابان نمکی دشت کویر و دشت لوت پوشیده شده است. استان‌های قم، اصفهان، یزد، کرمان در منطقه مرکزی قرار دارند، بخش‌هایی از سیستان و بلوچستان، سمنان، خراسان جنوبی و خراسان رضوی نیز جزء مناطق بیابانی و نیمه‌بیابانی محسوب می‌شوند (رحیم‌زاده و نساج‌زواره، ۲۰۱۳).

در شکل ۱ توزیع پراکندگی ایستگاه‌های مورد مطالعه در دوره سی‌ساله (۲۰۱۸-۱۹۸۹) نشان داده شده است.

آذربایجان غربی و شرقی، اردبیل و استان‌های واقع شده در منطقه کوهستانی زاگرس شامل استان‌های کردستان، همدان و زنجان تجربه دماهای پایین و زمستان‌های سرد و خشن‌تری را نسبت به بقیه نقاط کشور دارا هستند (رحیم‌زاده و نساج‌زواره، ۲۰۱۳).

۳- ناحیه کوهپایه‌ای: شامل کوهپایه‌های جنوبی البرز و ارتفاعات خراسان و کوهپایه‌های زاگرس زیر ارتفاعات ۱۵۰۰ متر است (علیجانی، ۱۳۸۵). اقلیم بیشتر این استان‌ها گرم و خشک است اما در مجاورت کوهستان سرد و نیمه‌مرطوب و در مناطق با ارتفاع بیشتر، سرد و دارای زمستان‌های طولانی می‌باشد (رحیم‌زاده و نساج‌زواره، ۲۰۱۳).

۴- ناحیه جنوب: شامل سواحل دریای عمان، خلیج فارس، جلگه خوزستان می‌باشد و مناطقی از جنوب شرق ایران را



شکل ۱. توزیع پراکندگی ایستگاه‌های مورد مطالعه در سری زمانی ۳۰ ساله (۲۰۱۸-۱۹۸۹).

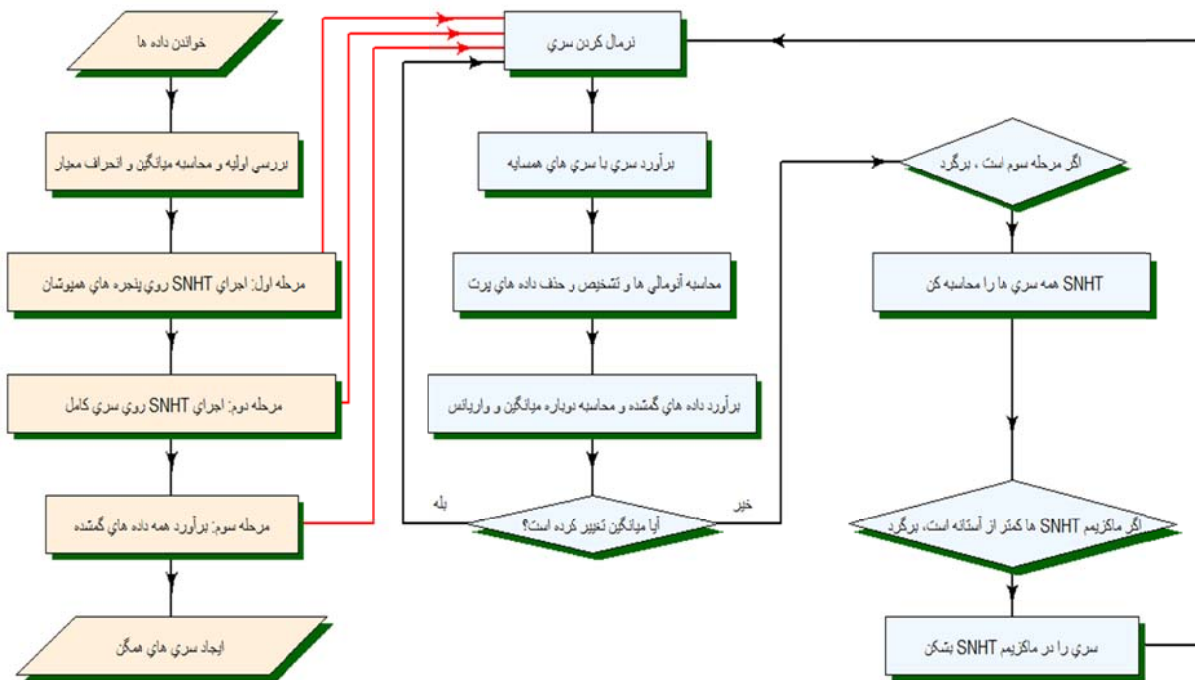
پس از خوشه‌بندی، داده‌های دمای بیشینه، کمینه و بارش روزانه با استفاده از نرم‌افزار SPSS مورد تحلیل آماری قرار گرفتند و درصد داده‌های گم‌شده به تفکیک هر ایستگاه مشخص شد. سپس برای بررسی داده‌های پرت، نقاط ناهمگنی و همگن‌سازی داده‌های کمینه و بیشینه دما و بارش روزانه از پکیج کلایماتول در نرم‌افزار R استفاده شد. به این ترتیب برای کنترل کیفیت و همگن‌سازی داده‌های مربوط به هر ایستگاه، در هر خوشه تعدادی از ایستگاه‌های متعلق به آن خوشه (به‌عنوان زیرخوشه) که دارای تغییرپذیری یکسان برای پارامتر موردنظر با ایستگاه مورد بررسی هستند، به‌عنوان سری‌های مرجع در نظر گرفته می‌شوند. بر اساس این الگوریتم، ابتدا سری موردنظر به‌وسیله سری‌های مرجع به روش رگرسیون نوع (II) برآورد شده و استانداردسازی می‌شود. بعد از برآورد سری‌ها، سری آنومالی استاندارد شده که تفاضل مقادیر مشاهده شده و برآورد شده است، محاسبه می‌شود. سپس داده‌های پرت با استفاده از سری‌های آنومالی برآورد می‌شوند. به این ترتیب که آنومالی‌هایی که از آستانه‌های تعیین‌شده، تجاوز کنند به‌عنوان داده پرت در نظر گرفته می‌شوند. این آستانه‌ها با استفاده از هیستوگرام آنومالی‌ها تعیین شدند. در ادامه پژوهش، برای اطمینان از تشخیص صحیح داده‌های پرت، بررسی وضعیت جوی در روزهای قبل و بعد و همچنین روز موردنظر ایستگاه، از جمله بارش شش ساعته، همخوانی گزارش‌های شش ساعته بارش با بارش گزارش شده ۲۴ ساعته، وضعیت ابرناکی، پدیده‌های هوای حاضر و گذشته و بررسی ارتباطات منطقی بارش با وضعیت جوی گزارش شده و همچنین بررسی داده‌های ساعتی دما و روندهای دمایی و همخوانی با دمای کمینه و بیشینه گزارش شده در تاریخ‌های موردنظر با استفاده از چند پایگاه داده هواشناسی نظیر سامانه درخواست داده‌های هواشناسی متعلق به سازمان هواشناسی کشور

(MESSIR System)، پایگاه OIGMET و METEOMANZ مورد بررسی قرار گرفتند و داده‌های پرت تأییدشده به‌عنوان داده پرت نهایی گزارش شدند. سپس آزمون همگنی نرمال استاندارد (SNHT) بر روی سری‌ها در دو مرحله اجرا شد. از آنجا که آزمون SNHT برای پیدا کردن یک نقطه شکست در سری طراحی شده است، وجود نقاط شکست بیشتر، ممکن است باعث ارزیابی در نتایج شود (جوانشیری و همکاران، ۱۳۹۹). از این‌رو این آزمون در مرحله اول روی پنجره‌هایی همپوشان اجرا می‌شود. به‌طور مثال برای داده‌های روزانه روی ۷۳۰ (دو برابر ۳۶۵) جمله اول سری اجرا شده و در هر مرحله ۳۶۵ جمله به جلو حرکت می‌کند و اگر داده‌ها ماهانه باشند این مقدار از ۳۶۵ به ۶۰ کاهش می‌یابد؛ در مرحله دوم، آزمون روی سری کامل اجرا می‌شود. زمانی که آماره آزمون SNHT بیشتر از آستانه از پیش تعیین‌شده بود، سری در همان نقطه‌ای که ماکزیمم SNHT، مربوط به آن بود شکسته 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 شوند. بعد از تعیین نقاط شکستگی در مرحله دوم، جهت تفکیک جهش‌های اقلیمی و غیراقلیمی، از بانک فراداده و مقایسه با سری‌های ایستگاه‌های مجاور استفاده شد. در مرحله سوم همه داده‌های گم‌شده، داده‌های پرت یا داده‌هایی که به سری جدیدی منتقل شده‌اند برآورد شدند.

۳. مبانی نظری مورد استفاده در مدل‌ها

۳-۱. الگوریتم کلایماتول

در شکل ۲ نمودار روند عملیاتی در الگوریتم کلایماتول نشان داده شده است.



شکل ۲. نمودار روندهای عملیاتی در الگوریتم کلایماتول (گویچارو، ۲۰۱۸).

۲-۳. رگرسیون متعامد نوع (II)

در رگرسیون نوع II در شکل ۳- الف فاصله عمودی نقاط تا خط رگرسیون کمینه می‌شود. به‌طور آماری، رابطه  $\hat{y}_i = x_i$  یک مدل رگرسیون خطی با عنوان رگرسیون متعامد (Orthogonal regression) یا محور اصلی کاهش‌یافته (Redused major axis) نامیده می‌شود که در آن فاصله عمودی نقاط تا خط رگرسیون کمینه می‌شود. در حالی که در رگرسیون نوع I در شکل ۳- ب انحراف تا خط رگرسیون در راستای محور  $y$ ها کمینه می‌شود و فرمول آن با سری‌های استاندارد شده از طریق رابطه  $\hat{y}_i = r x_i$  به‌دست می‌آید. که در آن  $r$  ضریب همبستگی بین متغیر مستقل و وابسته است (گویچارو، ۲۰۱۸). روش رگرسیون متعامد اولین بار توسط پالوس و کوهلر در سال ۱۹۵۲ برای برآورد داده‌های گم‌شده بارش ارائه شد (جوانشیری و همکاران، ۱۳۹۹).

اساس کار کلایماتول، برآورد سری موردنظر به‌وسیله سری‌های مرجع به روش رگرسیون نوع (II) است. به‌این منظور ابتدا سری‌ها با استفاده از رابطه (۱) استاندارد می‌شود.

$$x = \frac{X - m_x}{S_x} \tag{1}$$

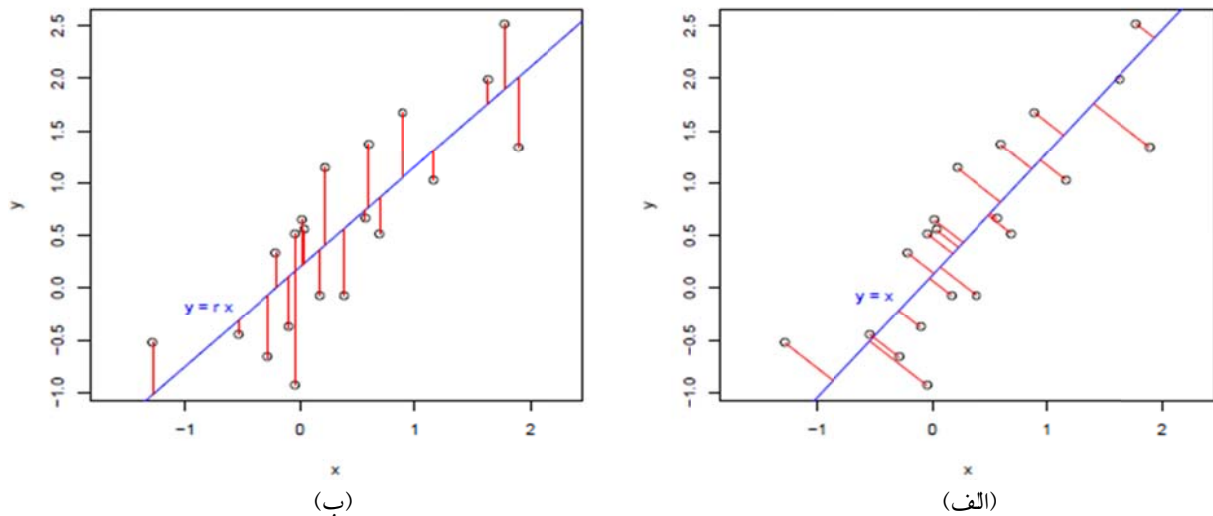
$m_x$  و  $S_x$  به‌ترتیب میانگین و انحراف معیار سری  $X$  هستند. سپس هر سری با استفاده از سری مرجع آن به صورت رابطه (۲) محاسبه می‌شود.

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

$\hat{y}$ : داده‌ای است که از نزدیک‌ترین تعداد ( $N$ ) داده مرجع موجود در هر گام زمانی تخمین زده می‌شود.

$X_j$ : داده مرجع

$W_j$ : وزنی است که به داده مرجع اختصاص داده شده است که تابعی معکوس از فاصله آنها تا ایستگاه مورد نظر است.



شکل ۳. خطوط قرمز به کمینه رساندن انحرافات از خط رگرسیون (آبی) به روش رگرسیون متعامد (رگرسیون نوع II) (الف) و روش حداقل مربعات معمولی (رگرسیون نوع I) (ب) و (گویجارو، ۲۰۱۸).

در سامانه مسیر (MESSIR) سازمان هواشناسی که داده‌ها از طریق آن دریافت می‌شود، مرتبط بود. به‌عنوان مثال در تاریخ ۲۰۱۴/۰۵/۲۴ هیچ‌کدام از ایستگاه‌ها در کلیه خوشه‌ها دمای بیشینه‌ای ندارند. در چنین مواردی برای برآورد دمای بیشینه در تاریخ‌های موردنظر از میانگین دمای روزهای قبل و بعد خود ایستگاه‌ها استفاده کردیم. برای دمای بیشینه در ایستگاه‌های ناحیه خزری، ۱۸ تاریخ در سری زمانی ۳۰ ساله در تمامی ایستگاه‌های خوشه فاقد داده بودند.

#### ۱-۴. داده‌های گم‌شده

در این مرحله داده‌های دمای کمینه، بیشینه و بارش روزانه، ایستگاه‌های هواشناسی کشور، در سری زمانی ۳۰ ساله در بازه زمانی ۲۰۱۸-۱۹۸۹ توسط نرم‌افزار SPSS مورد تحلیل آماری قرار گرفت و درصد داده‌های گم‌شده پارامترهای موردنظر به تفکیک ایستگاه‌ها مشخص شد. این بررسی‌ها نشان داد در سری زمانی ۳۰ ساله، ایستگاه‌های هواشناسی کشور، در مورد دمای بیشینه به‌طور متوسط دارای ۳ درصد، دمای کمینه ۴ درصد و پارامتر بارش دارای ۲ درصد داده گم‌شده می‌باشند.

#### 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 حداقل یک نقطه مانند  $M$  که داده‌های قبل از آن با میانگین  $\mu_1$  و داده‌های بعد از آن با میانگین  $\mu_2$  باشند را ارائه می‌کند.

برای انجام این آزمون، سری استاندارد می‌شود. اگر  $Z_i$ ،  $i=1, \dots, n$  سری استاندارد شده باشد. آماره آزمون نرمال استاندارد از رابطه (۳) به دست می‌آید.

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

$$1 \leq a \leq n-1$$

که در آن برای  $a=1, \dots, n-1$  میانگین  $\bar{z}_1$  داده اول و  $\bar{z}_2$  میانگین  $n-a$  داده باقی است (جوانشیری و همکاران، ۱۳۹۹ به نقل از الکساندر و موبرگ، ۱۹۹۷).

#### ۴. نتایج و بحث

در سری زمانی سی‌ساله (۲۰۱۸-۱۹۸۹) به دلیل خطاهای زیادی که در ایستگاه ژئوفیزیک مشاهده شد، این ایستگاه از فرایند همگن‌سازی کنار گذاشته شد. در بررسی‌ها مشخص شد که در بعضی از تاریخ‌ها تمامی ایستگاه‌ها در یک خوشه فاقد داده می‌باشند؛ این احتمالاً به دلیل اشکال

شکل ۴ نشان داده شده است خط نقطه‌چین سبز نشان می‌دهد مقدار مطلوب برای تحلیل، تعداد ۵ ایستگاه یا بیشتر در هرگام زمانی و حداقل آن ۳ ایستگاه برای هرگام زمانی (خط نقطه‌چین قرمز) مورد نیاز است. بنابراین تعداد ایستگاه‌ها در مورد دمای بیشینه (شکل ۳-الف) و دمای کمینه (شکل ۳-ب) در هر گام زمانی برای تحلیل‌های آماری مطلوب است. شکل ۵ خوشه‌بندی ایستگاه‌ها در گروه‌هایی با تغییرپذیری یکسان برای دمای بیشینه و دمای کمینه را نشان می‌دهد. این شکل نشان می‌دهد ایستگاه‌های نواحی بیابانی و نیمه‌بیابانی بر اساس تغییرپذیری دمای بیشینه (شکل ۵-الف) و دمای کمینه (شکل ۵-ب) به ۸ خوشه با تغییرپذیری یکسان تقسیم می‌شود.

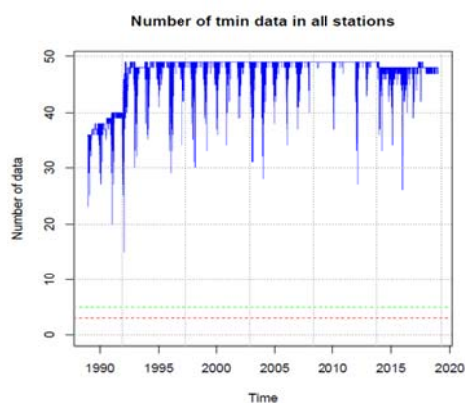
#### ۲-۴. تحلیل اکتشافی و تشخیص نقاط ناهمگنی

با توجه به این که ارائه جزئیات تحلیل برای همه نواحی، حجم مقاله را بسیار طولانی می‌کند، بنابراین در ادامه به‌عنوان نمونه به تحلیل داده‌های دمای کمینه، بیشینه و بارش روزانه نواحی بیابانی و نیمه‌بیابانی در سری زمانی سی‌ساله می‌پردازیم. لازم به‌ذکر است روش و عملکرد مرتبط با سایر نواحی در سری‌های زمانی سی‌ساله نیز مانند نمونه موردبررسی می‌باشد.

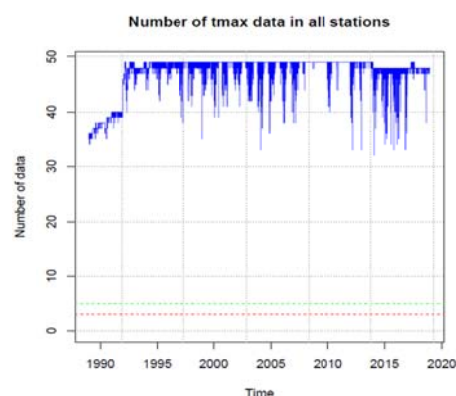
#### ۲-۴-۱. نواحی مرکزی (بیابانی و نیمه‌بیابانی) در سری

زمانی ۱۹۸۹-۲۰۱۸ دمای بیشینه و دمای کمینه

شکل ۴، داده‌های موجود را بر حسب تعداد ایستگاه در گام‌های زمانی مختلف نشان می‌دهد. همان‌گونه که در

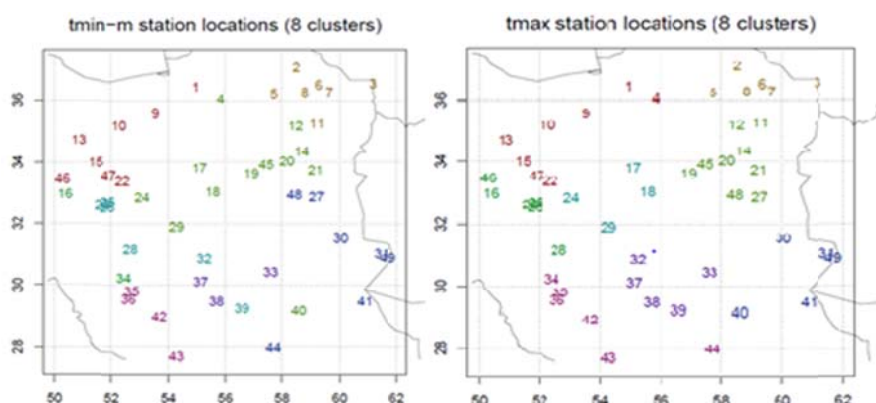


(ب)



(الف)

شکل ۴. تعداد داده‌های دمای بیشینه (الف) و دمای کمینه (ب) ایستگاه‌ها برای نواحی بیابانی و نیمه‌بیابانی در گام‌های زمانی (۱۹۸۹-۲۰۱۸).



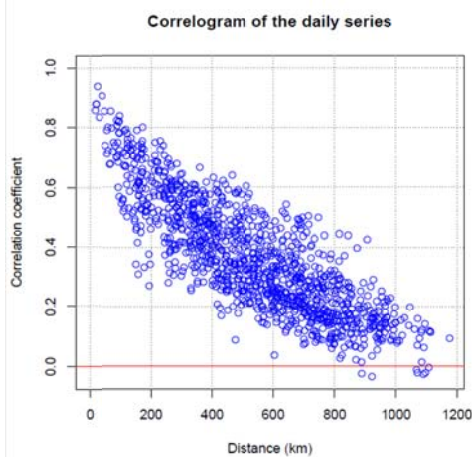
(ب)

(الف)

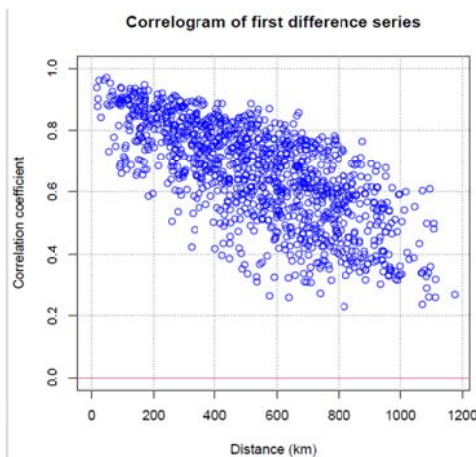
شکل ۵. خوشه‌بندی ایستگاه‌ها در گروه‌هایی با تغییرپذیری یکسان برای دمای بیشینه (الف) و دمای کمینه (ب) برای نواحی بیابانی و نیمه‌بیابانی، دوره زمانی (۱۹۸۹-۲۰۱۸).

بوده و برای هدف مورد نظرمان مطلوب است. شکل ۷ نمودار دسته‌بندی ایستگاه‌ها در خوشه و میزان شباهت و تجانس آنها با یکدیگر را با استفاده از شاخص عدم تشابه (dissimilarity)، برای دمای بیشینه (شکل ۷-۱) و دمای کمینه (شکل ۷-۲) به عنوان یک معیار تفکیک نشان می‌دهد. همان‌گونه که مشاهده می‌شود تمامی ایستگاه‌هایی که در نقشه به عنوان سری مرجع برای یکدیگر در خوشه استفاده می‌شوند، آستانه‌ای پایین‌تر از آستانه شاخص تعیین شده - که با خط قرمز مشخص شده است - دارند. هرچه شاخص عدم تشابه پایین‌تر باشد به این معنی است که داده‌ها از توزیع یکنواخت‌تری برخوردارند.

شکل ۶، نمودار ضرایب همبستگی بین سری‌ها را برای دمای بیشینه (شکل ۶-الف) و دمای کمینه (شکل ۶-ب) نشان می‌دهد. نمودار ضرایب همبستگی به ما کمک می‌کند، یکنواختی تغییرات آب‌وهوایی را در منطقه مورد نظر ارزیابی کنیم یا به عبارت دیگر وجود عواملی را که ممکن است باعث تغییرات شدید آب‌وهوایی در منطقه شوند (مانند سلسه کوه‌ها و مناطق با تیغه توپوگرافی شدید) شناسایی کنیم. همان‌گونه که مشاهده می‌شود با افزایش فاصله مکانی، همبستگی بین دما در ایستگاه‌ها کاهش می‌یابد. این نشان می‌دهد منطقه‌ای که انتخاب کرده‌ایم از نظر تغییرات آب و هوایی تقریباً یکنواخت

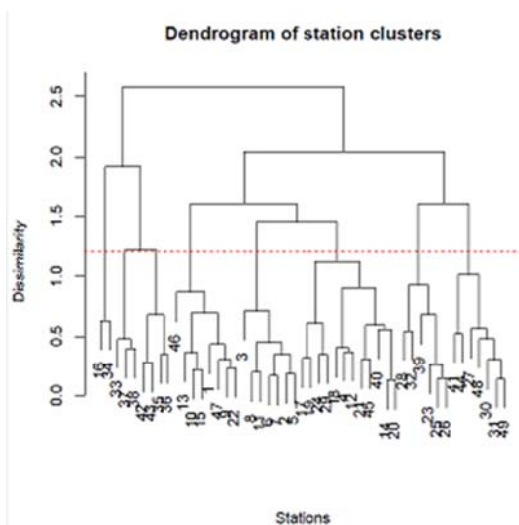


(ب)

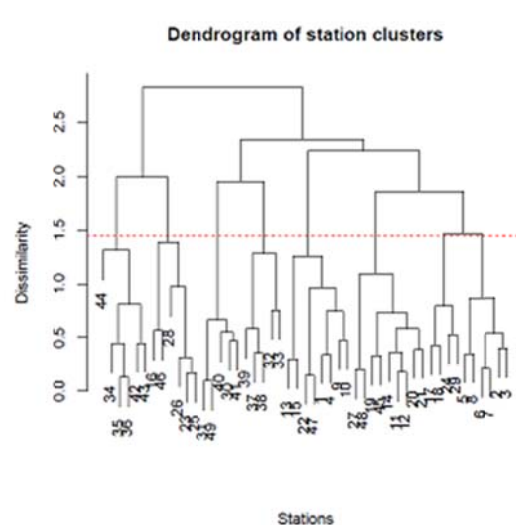


(الف)

شکل ۶. ضرایب همبستگی در مقابل فاصله مکانی برای دمای بیشینه (الف) و دمای کمینه (ب) برای نواحی بیابانی و نیمه‌بیابانی، دوره ۱۹۸۹-۲۰۱۸.



(ب)



(الف)

شکل ۷. دسته‌بندی ایستگاه‌ها در خوشه با استفاده از معیار تفکیک عدم تشابه (dissimilarity) برای دمای بیشینه (الف) و دمای کمینه (ب) برای نواحی بیابانی و

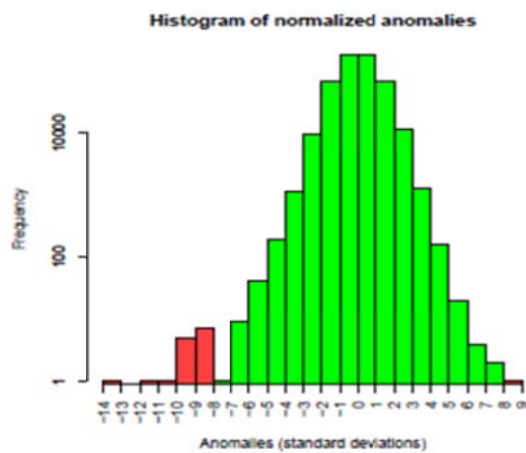
نیمه بیابانی، دوره ۱۹۸۹-۲۰۱۸.



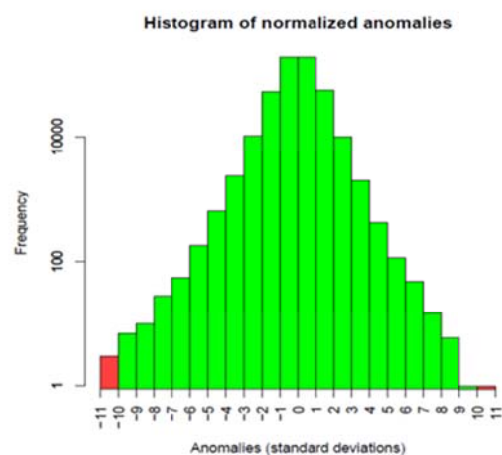
می‌دهد.

در مرحله بعد روزهای همراه با داده‌های دمای پرت که به وسیله الگوریتم تشخیص داده شده بودند، در تاریخ موردنظر و همچنین روزهای قبل و بعد آن مورد بررسی-های دقیق هواشناسی قرار گرفتند. این بررسی‌ها، شامل بررسی دماهای ساعتی گزارش شده و همخوانی آنها با دمای کمینه و بیشینه و روندهای دمایی، پدیده‌های جوی گزارش شده، شرایط ابرناکی و... بود. داده‌های پرتی که پس از بررسی‌های هواشناختی مورد تأیید قرار گرفتند برای دمای بیشینه در جدول ۱ و دمای کمینه در جدول ۲ نشان داده شده است.

برای بررسی داده‌های پرت از هیستوگرام آنومالی (تفاوت مقادیر مشاهده شده و برآورد شده) استاندارد شده استفاده شده است. در این روش مقادیری به عنوان داده پرت شناسایی می‌شوند که تفاوت زیادی با مقادیر همسایگی خود داشته باشند. بنابراین پس از بررسی هیستوگرام آنومالی استاندارد شده برای دمای بیشینه، آنومالی‌های استاندارد شده بیشتر از ۱۰ و کمتر از -۱۰ و برای دمای کمینه آنومالی‌های استاندارد شده بیشتر از ۸ و کمتر از -۸ به عنوان داده پرت در نظر گرفته شدند. شکل ۸ هیستوگرام آنومالی‌های استاندارد شده را برای دمای بیشینه (شکل ۸-الف) و دمای کمینه (شکل ۸-ب) نشان



(ب)



(الف)

شکل ۸. هیستوگرام آنومالی (تفاوت مقادیر مشاهده شده و برآورد شده) استاندارد شده برای دمای بیشینه (الف) و دمای کمینه (ب) برای نواحی بیابانی و نیمه بیابانی در سری زمانی (۲۰۱۸-۱۹۸۹).

جدول ۱. داده‌های پرت نهایی دمای بیشینه ایستگاه‌های نواحی بیابانی و نیمه بیابانی در سری زمانی ۳۰ ساله پس از بررسی‌های هواشناختی.

نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	آنومالی استاندارد شده
گلمکان	۴۰۷۴۴	۲۰۱۶/۱۱/۲۷	۴۳/۸	۸/۹	۱۷/۳۶
اصفهان فرودگاه	۴۰۸۰۲	۲۰۱۴/۱۲/۱۶	۱۹/۸	۱۰/۲	۱۰/۵۸
خور بیرجند	۹۹۴۳۷	۲۰۱۳/۰۶/۰۸	۳۰	۴۳/۴	-۱۰/۱۱
خور بیرجند	۹۹۴۳۷	۲۰۱۳/۰۶/۰۹	۲۷	۴۱/۱	۱۰/۶۱

بالای ۴۰ درجه سلسیوس را ثبت کرده بودند، با توجه به اقلیم منطقه و خصوصیات جغرافیایی ایستگاه از قبیل طول، عرض جغرافیایی و ارتفاع و نیز پرس‌وجو از همکاران هواشناس بومی منطقه، داده بیشینه این ایستگاه در این تاریخ‌ها باید دمای بیش از ۴۰ درجه سلسیوس را نشان می‌داد. حال آن‌که دماهای ثبت شده در ایستگاه خور در تاریخ‌های ۸ و ۹ ژوئیه ۲۰۱۶، به ترتیب ۳۰ و ۲۷ درجه سلسیوس است. بنابراین علی‌رغم همخوانی دماهای ساعتی با دمای بیشینه ایستگاه و همچنین دمای بیشینه روزهای قبل و بعد ایستگاه، ولی به دلیل عدم همخوانی شدید دمای بیشینه با تمامی ایستگاه‌های مجاور، نظیر بیرجند، طبس و نهبندان و همچنین اقلیم منطقه در این فصل، دمای بیشینه ایستگاه و داده‌های ساعتی این ایستگاه پرت تشخیص داده شدند (ایستگاه خور بیرجند در اختیار نیروی هوایی است و فاقد نظارت کافی از سوی سازمان هواشناسی می‌باشد).

دمای بیشینه ایستگاه گل‌مکان-این ایستگاه در فاصله کمتر از ۳۰ کیلومتری شهر مشهد قرار دارد- به دلیل عدم همخوانی با داده‌های دمای ساعتی ایستگاه و از طرفی غیر منطقی بودن مقدار دمای بیشینه ایستگاه در تاریخ ۲۷/۱۱/۲۰۱۶ که مصادف با آذر ماه در اواخر پاییز می‌باشد و رخداد دماهای بالای ۴۰ درجه سلسیوس در این فصل، با توجه به شرایط جغرافیایی و ارتفاع ایستگاه در منطقه موردنظر بی‌سابقه و حتی در فصل تابستان نیز کم سابقه می‌باشد و از طرفی عدم همخوانی دمای بیشینه ایستگاه با دماهای بیشینه ایستگاه‌های مجاور نظیر مشهد و قوچان، در این تاریخ، این داده به‌عنوان داده پرت تشخیص داده شد. دمای بیشینه فرودگاه اصفهان در تاریخ ۱۶/۱۲/۲۰۱۴ به دلیل عدم همخوانی با داده‌های دمای ساعتی ایستگاه به‌عنوان داده پرت تشخیص داده شد. در مورد ایستگاه خور با توجه به این‌که ایستگاه‌های اطراف شامل بیرجند، طبس و نهبندان در این تاریخ‌ها، دماهای

جدول ۲. داده‌های پرت دمای کمینه ایستگاه‌های نواحی بیابانی و نیمه‌بیابانی در سری زمانی ۳۰ ساله پس از بررسی‌های هواشناختی.

نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	آنومالی استاندارد شده
سبزوار	۴۰۷۴۳	۲۰۱۷/۰۸/۱۹	۲/۵	۱۸/۹	-۹/۴۶
نیشابور	۴۰۷۴۶	۲۰۱۵/۰۷/۲۱	۰	۱۴/۸	-۸/۵۹
نیشابور	۴۰۷۴۶	۲۰۱۵/۰۸/۰۲	۰	۱۶	-۹/۳
نیشابور	۴۰۷۴۶	۲۰۱۶/۰۶/۰۲	۰	۱۴/۶	-۸/۴۵
نیشابور	۴۰۷۴۶	۲۰۱۶/۸/۲۳	۰	۱۵/۶	-۹/۰۱
نیشابور	۴۰۷۴۶	۲۰۱۷/۵/۱۳	۰	۱۵/۲	-۸/۸۳
نیشابور	۴۰۷۴۶	۲۰۱۷/۵/۲۴	۰	۱۵/۳	-۸/۸۴
نیشابور	۴۰۷۴۶	۲۰۱۷/۸/۱۹	۰	۱۴/۱	-۸/۳۸
نیشابور	۴۰۷۴۶	۲۰۱۸/۷/۳۰	۰	۱۵/۸	-۹/۱۷
سمنان	۴۰۷۵۷	۲۰۱۷/۶/۲۷	۱۰/۸	۲۴/۴	-۸/۲۶
کاشمر	۴۰۷۶۳	۲۰۱۵/۶/۱۸	۴/۶	۲۳/۵	-۱۱/۴
خور و بیابانک	۴۰۷۸۹	۲۰۱۴/۹/۱۲	۲/۲	۲۱/۱	-۱۳/۸۳
کیوترآباد	۴۰۸۰۳	۲۰۱۶/۲/۲۵	۱۲/۶	۰/۱	۸/۴۳
خور بیرجند	۹۹۴۳۷	۱۹۹۸/۷/۲۲	۲/۶	۲۵/۷	-۱۰/۷۸
خور بیرجند	۹۹۴۳۷	۱۹۹۸/۷/۲۷	۲/۴	۲۲/۳	-۹/۳

آمده بودند، اجرا شد. در این مرحله برای سری زمانی دمای بیشینه نقطه شکستگی تشخیص داده نشد و برای دمای کمینه ۶ نقطه شکستگی مشاهده شد. مشخصات نقاط شکستگی برای دمای بیشینه در مرحله اول در جدول شماره ۳ و برای دمای کمینه در جدول شماره ۴ و ۵ نشان داده شده است.

جدول ۳. نقاط شکستگی برای دمای بیشینه نواحی بیابانی و نیمه بیابانی (مرحله اول) دوره ۲۰۱۸-۱۹۸۹.

نام ایستگاه	کد ایستگاه	تاریخ شکستگی	مقدار آماره SNHT
فردوس	۴۰۷۹۲	۱۹۹۹/۰۷/۰۱	۷۱/۵
سیرجان	۴۰۸۵۱	۲۰۰۵/۱۰/۰۱	۳۰/۴
زاهدان	۴۰۸۵۶	۲۰۰۶/۱۲/۰۱	۳۳/۹
طبس	۴۰۷۹۱	۲۰۱۳/۰۴/۰۱	۴۸/۲

جدول ۴. نقاط شکستگی برای دمای کمینه نواحی بیابانی و نیمه بیابانی (مرحله اول)، ۲۰۱۸-۱۹۸۹.

نام ایستگاه	کد ایستگاه	تاریخ شکستگی	مقدار آماره SNHT
قوچان	۴۰۷۴۰	۲۰۰۷/۰۵/۰۱	۷۲/۴
سبزوار	۴۰۷۴۳	۲۰۰۳/۱۲/۰۱	۹۵/۱
سبزوار	۴۰۷۴۳	۲۰۰۸/۱۱/۰۱	۸۲/۱
سمنان	۴۰۷۵۷	۲۰۰۱/۰۷/۰۱	۵۶/۷
گرمسار	۴۰۷۵۸	۲۰۰۶/۰۱/۰۱	۷۹/۵
کاشان	۴۰۷۸۵	۲۰۰۴/۰۳/۰۱	۴۴/۶
فردوس	۴۰۷۹۲	۲۰۰۷/۱۰/۰۱	۴۰/۶
اصفهان	۴۰۸۰۰	۱۹۹۴/۱۲/۰۱	۹۴/۲
فرودگاه اصفهان	۴۰۸۰۲	۲۰۱۰/۱۱/۰۱	۵۳/۱
شیراز	۴۰۸۴۸	۱۹۹۳/۰۸/۰۱	۴۹/۷
شیراز	۴۰۸۴۸	۲۰۰۳/۱۱/۰۱	۷۸/۱

دمای کمینه این ایستگاه‌ها نیز با توجه به عدم همخوانی با دماهای ساعتی گزارش شده، اختلاف با دماهای کمینه ایستگاه‌های مجاور و عدم همخوانی دمای کمینه با شرایط اقلیمی منطقه در تاریخ مورد نظر به عنوان داده پرت نهایی مورد تأیید قرار گرفتند.

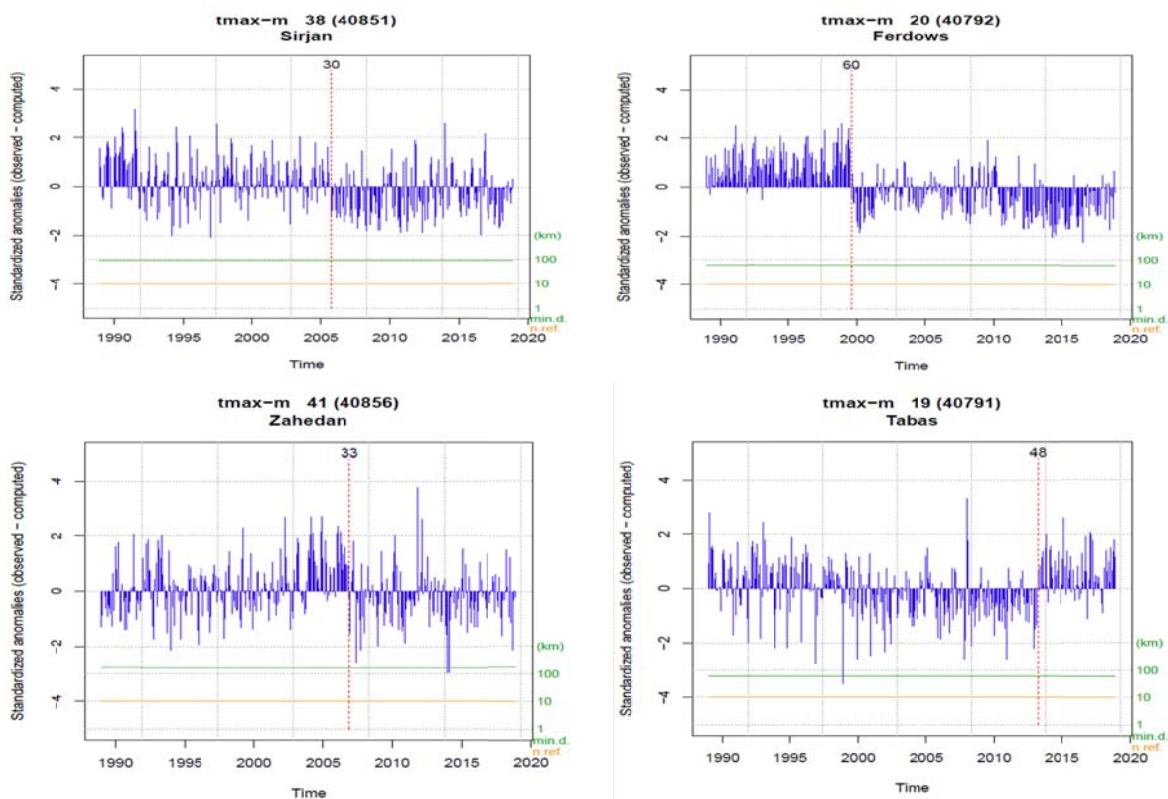
#### ۴-۲-۲. بررسی نقاط ناهمگنی و همگن سازی داده‌های دما

در این مرحله، سعی شد تا نقاط ناهمگنی در سری زمانی دمای بیشینه و دمای کمینه مشخص شود، از آنجا که داده‌ها روزانه بودند و تغییرپذیری داده‌های روزانه زیاد می‌باشد و این مسأله باعث کاهش کارایی تشخیص ناهمگنی می‌شود، ابتدا نقاط ناهمگنی در داده‌های ماهانه مشخص شد. برای بررسی همگنی داده‌های ماهیانه، پس از اجراء تابع dd2m در پکیج کلایماتول در نرم افزار R و جمع داده‌های روزانه و به دست آوردن داده‌های ماهیانه، ابتدا تابع همگن سازی (Homogen) با مد اکتشافی (exploratory) برای کشف و تعیین مقادیر مناسب آماره SNHT مرحله اول بر روی (پنجره‌های همپوشان) و مرحله دوم (سری‌های کامل) اجرا شد. پس از مشخص شدن مقادیر مناسب آماره SNHT برای هر یک از این دو مرحله دوباره تابع همگن سازی با مقادیر تعیین شده اجرا شد. در مرحله اول، آزمون نرمال استاندارد روی پنجره‌های همپوشان، اجرا شد. در این مرحله ۴ نقطه شکستگی برای دمای بیشینه و ۱۱ نقطه شکستگی برای دمای کمینه مشخص شد. در واقع، 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 روی سری‌های جدید اجرا شد تا زمانی که دیگر نقطه ناهمگنی تشخیص داده نشد. در مرحله دوم، در این مرحله آزمون نرمال استاندارد روی سری‌های کامل که از مرحله قبل به دست

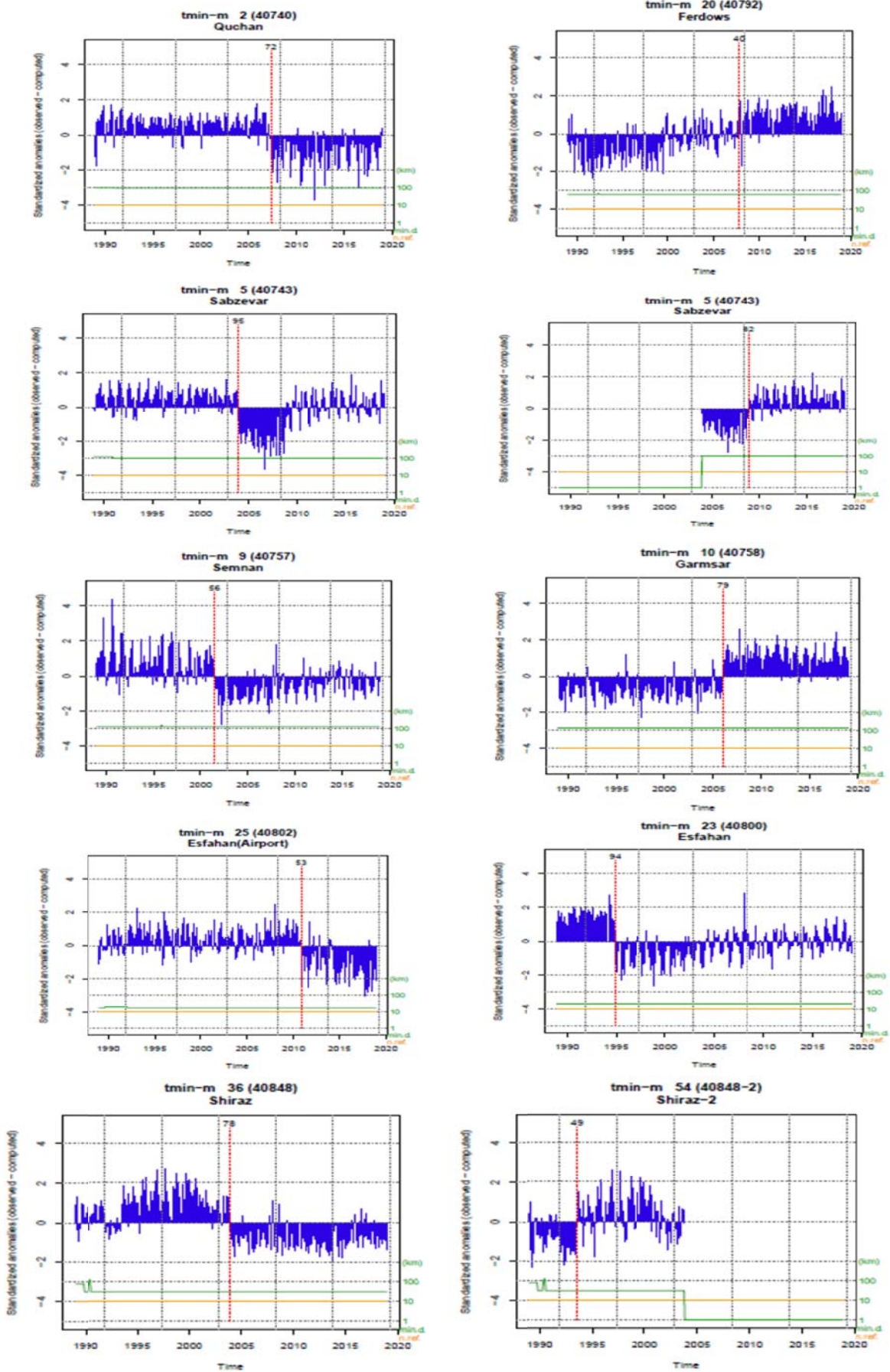
جدول ۵. نقاط شکستگی برای دمای کمینه نواحی بیابانی و نیمه‌بیابانی (مرحله دوم)، دوره ۱۹۸۹-۲۰۱۸.

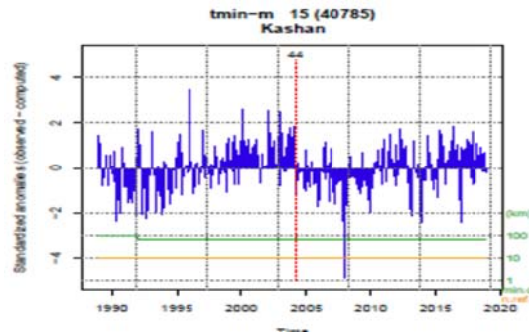
نام ایستگاه	کد ایستگاه	تاریخ شکستگی	مقدار آماره SNHT
۴۰۸۴۴	درودزن	۲۰۰۶/۱۰/۱	۱۵۸/۹
۴۰۷۸۹	خوروبیابانک	۲۰۰۸/۰۲/۱	۹۱/۳
۴۰۸۲۱	یزد	۲۰۰۱/۰۱/۱	۹۰/۷
۴۰۸۵۹	فسا	۲۰۱۰/۰۵/۱	۷۲/۴
۴۰۷۹۲	فردوس	۱۹۹۹/۰۷/۱	۷۱/۵
۴۰۸۴۹	شهربابک	۲۰۰۷/۱۰/۱	۷۰/۷

برای دمای بیشینه و دمای کمینه تشخیص نقاط شکستگی و نمودار آنومالی استاندارد شده در شکل‌های ۹ تا ۱۱ نشان داده شده است.

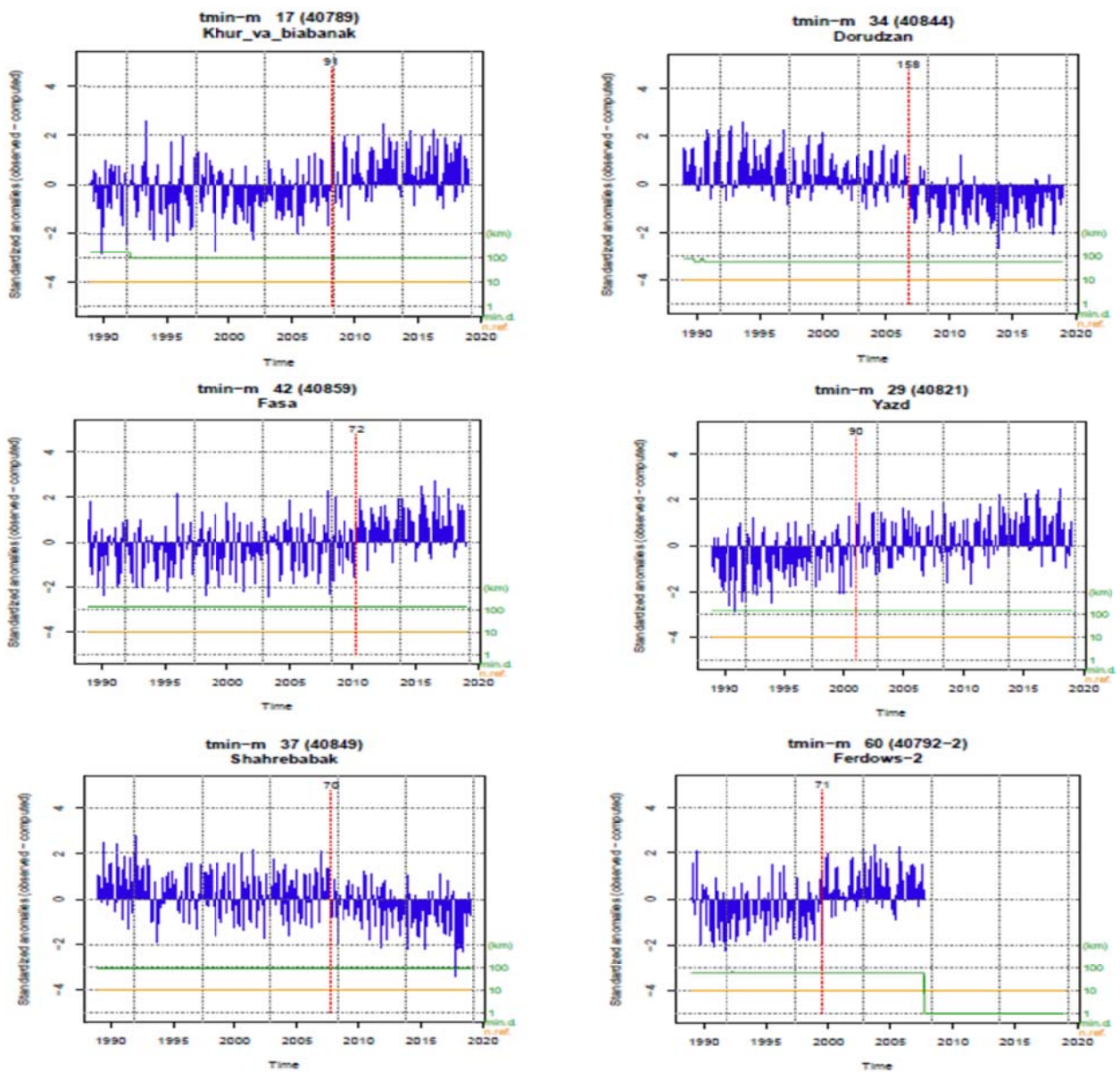


شکل ۹. نمودار آنومالی استاندارد شده و تشخیص نقاط شکستگی در سری زمانی ماهانه دمای بیشینه (مرحله اول) در نواحی بیابانی و نیمه‌بیابانی، دوره (۲۰۱۸-۱۹۸۹).





شکل ۱۰. نمودار آنومالی استاندارد شده و تشخیص نقاط شکستگی در سری زمانی ماهانه دمای کمینه برای نواحی بیابانی و نیمه‌بیابانی (مرحله اول)، دوره ۲۰۱۸-۱۹۸۹.



شکل ۱۱. نمودار آنومالی استاندارد شده و تشخیص نقاط شکستگی در سری زمانی ماهانه دمای کمینه برای نواحی بیابانی و نیمه بیابانی (مرحله دوم)، دوره ۲۰۱۸-۱۹۸۹.

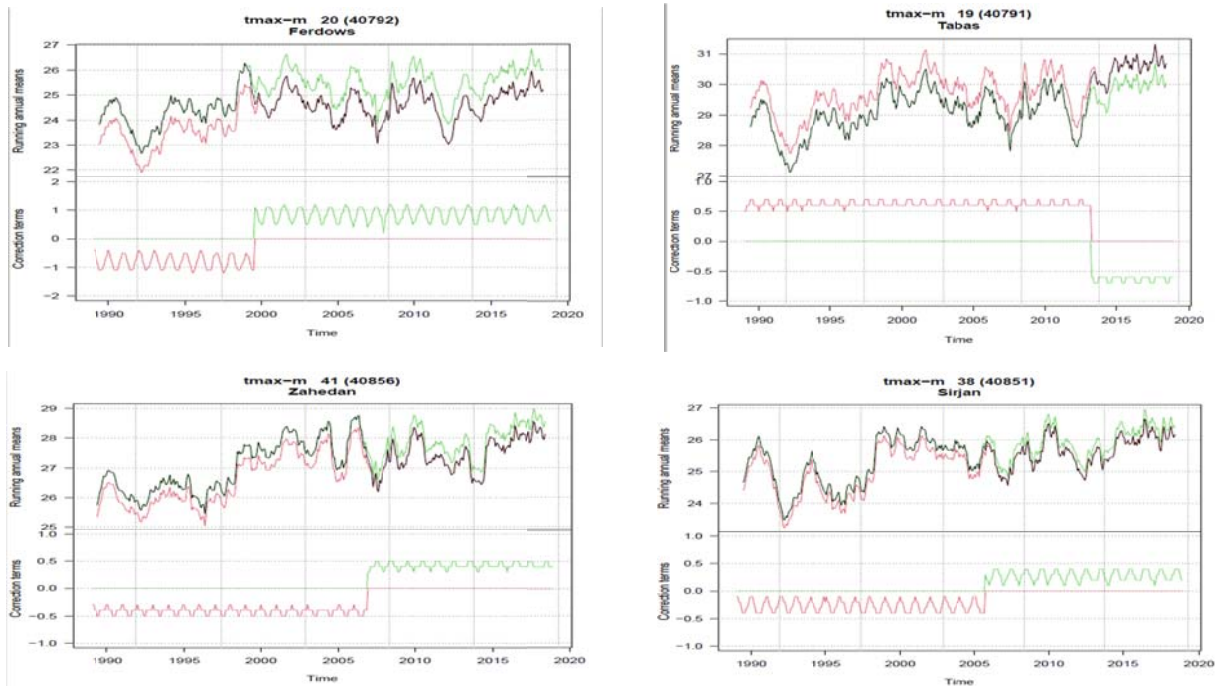
نمی‌تواند منشأ اقلیمی داشته باشد و به‌عنوان جهش‌های غیراقلیمی در نظر گرفته شد. در جدول شماره ۶ دلیل برخی از این جهش‌ها که با بررسی فراداده ایستگاه‌ها مشخص شده، نشان داده شده است.

در مرحله سوم، برای همگن‌سازی داده‌های روزانه، نقاط شکستگی ماهانه در نظر گرفته شدند. در این مرحله داده‌های پرت، یا داده‌هایی که به سری‌های دیگر منتقل شده، برآورد شدند. در شکل ۱۲ سری‌های کامل بازسازی شده با استفاده از زیردوره‌های همگن مربوط به ایستگاه‌های فردوس طیس، سیرجان و زاهدان به‌عنوان نمونه نشان داده شده اند. برای هر ایستگاه، نمودار بالا سری اصلی (رنگ سیاه) و سری‌های اصلاح شده با سایر رنگ‌ها (نظیر سبز و قرمز) و نمودار پایین میزان اصلاحات انجام‌شده روی هر یک از سری‌ها را نشان می‌دهد. سری‌های همگن شده در بانک داده ذخیره شده و بسته به این که در تحقیقات بعدی برای چه منظوری بخواهند مورد بررسی قرار گیرند، استفاده خواهند شد.

پس از تعیین نقاط شکستگی سری‌های زمانی سی‌ساله دمای بیشینه و کمینه، با استفاده از بانک فراداده موجود بررسی شد تا معلوم شود این نقاط شکستگی که در مراحل قبل به دست آمدند، منشأ اقلیمی یا غیراقلیمی دارند. در بررسی فراداده موجود بسیاری از شهرستان‌ها دلیل مستدلی برای پیدا کردن دلیل این شکستگی‌ها تشخیص داده نشد. با توجه به کامل نبودن بانک فراداده ایستگاه‌های هواشناسی کشور، به نظر می‌رسد برای پیدا کردن دلیل قطعی این جهش‌ها در سری زمانی این ایستگاه‌ها، نیاز به بررسی و تهیه فراداده دقیق این ایستگاه‌ها توسط ادارات کل هواشناسی استان‌های مربوطه از طریق بررسی اسناد موجود در بخش‌های مختلف این ادارات وجود دارد؛ بنابراین برای قضاوت در مورد منشأ این جهش‌ها، به مشاهده و بررسی سری‌های زمانی در ایستگاه‌های مجاور و سری‌های مرجع پرداخته شد. نظر به این که در تاریخ‌های گفته‌شده شکستگی مشهودی در سری‌های زمانی ایستگاه‌های اطراف مشاهده نشد به این نتیجه رسیدیم که جهش‌های مذکور در این ایستگاه‌ها

جدول ۶. دلایل شناسایی شده برخی از جهش‌های صورت‌گرفته با بررسی فراداده ایستگاه‌ها در بازه ۲۰۱۸-۱۹۸۹.

نام ایستگاه	کد ایستگاه	تاریخ شکستگی	دلایل شناسایی شده شکست‌های سری
سبزوار	۴۰۷۴۳	۲۰۰۳/۱۲/۰۱	تغییر مکان ایستگاه
سبزوار	۴۰۷۴۳	۲۰۰۸/۱۱/۰۱	تغییر مکان ایستگاه
سمنان	۴۰۷۵۷	۲۰۰۱/۰۷/۰۱	تغییر مکان ایستگاه
اصفهان (ازن‌سنجی)	۴۰۸۰۰	۱۹۹۴/۱۲/۰۱	تغییر مکان ایستگاه
شیراز	۴۰۸۴۸	۲۰۰۳/۱۱/۰۱	تغییر مکان ایستگاه

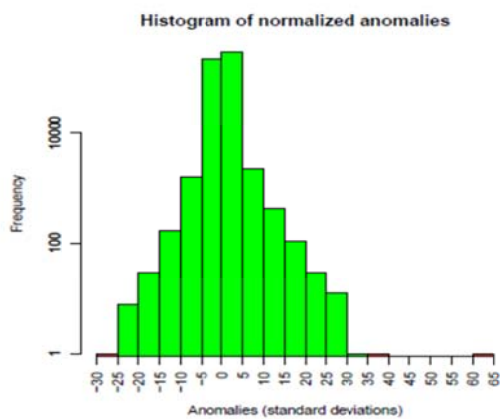


شکل ۱۲. سری‌های کامل بازسازی شده با استفاده از زبردوره‌های همگن مربوط به ایستگاه‌های فردوس طبس، سیرجان و زاهدان (۲۰۱۸-۱۹۸۹).

شدند.

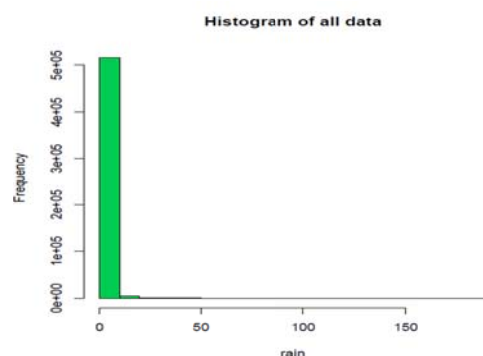
۴-۲-۱-۲. بارش

هیستوگرام داده‌ها (شکل ۱۳) نشان می‌دهد توزیع داده‌ها نرمال نیست و چولگی به‌راست دارد و اکثر داده‌ها صفر هستند که با در نظر داشتن این موضوع که ایستگاه‌ها در ناحیه بیابانی و نیمه‌بیابانی قرار دارند، چیز دور از انتظاری نبود.



شکل ۱۴. هیستوگرام آنومالی استاندارد شده بارش نواحی بیابانی و نیمه‌بیابانی، دوره ۲۰۱۸-۱۹۸۹.

از آنجا که رفتار بارش ماهیتاً با دما متفاوت است و تغییرپذیری آن نسبت به دما بالاتر است، لذا در مورد بارش بررسی وضعیت دقیق جوی در روزهایی که داده آن پرت تشخیص داده شده، از اهمیت به‌سزایی برخوردار است. بنابراین وضعیت جوی روزهای موردنظر از جمله، میزان بارش‌های شش ساعته و همخوانی آن با بارش گزارش شده ۲۴ ساعته، وضعیت ابرناکی، پدیده‌ها، میزان



شکل ۱۳. هیستوگرام داده‌های بارش نواحی بیابانی و نیمه‌بیابانی دوره ۲۰۱۸-۱۹۸۹.

شکل ۱۴ هیستوگرام آنومالی استاندارد شده بارش نواحی بیابانی و نیمه‌بیابانی را نشان می‌دهد. همان‌طور که از شکل مشخص است آنومالی‌های استاندارد شده با مقادیر بیش از ۳۵+ و کمتر از ۲۵- به‌عنوان داده پرت تشخیص داده



در خیلی از موارد همگنی یا عدم‌همگنی سری با بررسی نمودارهای آنومالی استاندارد شده تقریباً مشخص می‌شود، اما در مورد بارش وضعیت این‌گونه نیست. با توجه به تغییرپذیری بسیار زیاد داده‌های روزانه بارش، داده‌های ماهانه بارش را مورد بررسی قرار دادیم. همانند دما ابتدا آزمون نرمال استاندارد روی پنجره‌های همپوشان اجرا شد. در مرحله اول ۴ نقطه شکستگی برای بارش تشخیص داده شده است و ۴۹ سری، به ۵۳ سری همگن تبدیل شده است. تاریخ این شکستگی‌ها در جدول ۸ و نمودار آنومالی‌ها در شکل ۱۵ نشان داده شده است.

رطوبت نسبی و ... در روزهای همراه با داده پرت و روزهای قبل و بعد از آن مورد بررسی دقیق قرار گرفت. و داده‌هایی که پس از این بررسی‌ها به‌عنوان داده پرت نهایی تأیید شدند، در جدول ۷ نشان داده شده‌اند.

#### ۳-۲-۴. بررسی نقاط ناهمگنی و همگن‌سازی داده‌های بارش

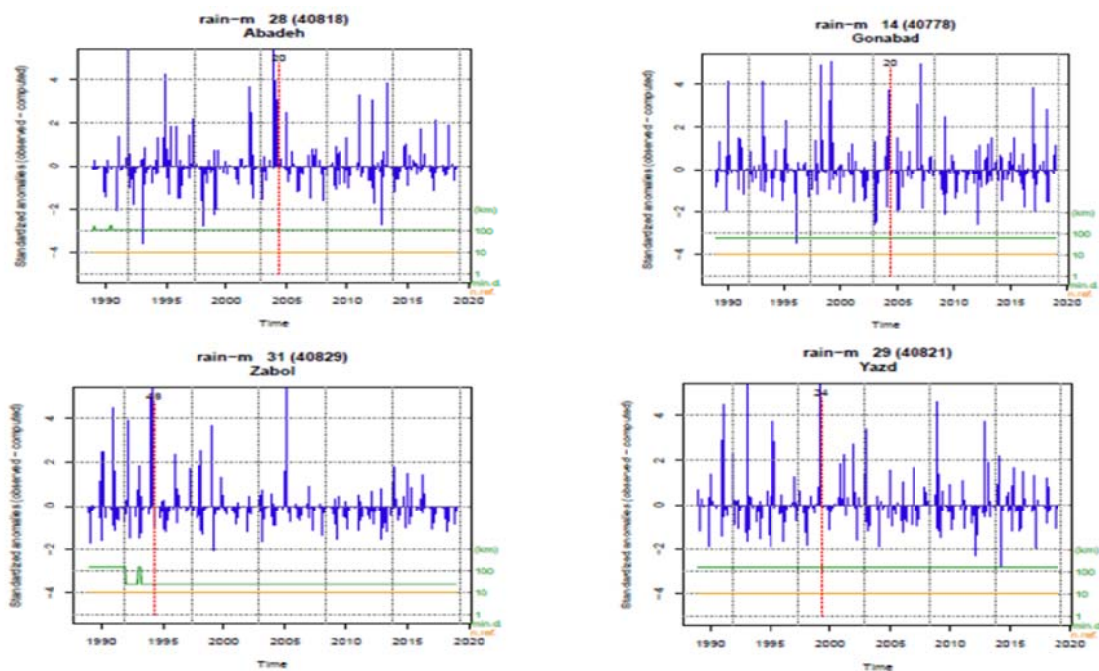
بررسی همگن بودن سری‌های زمانی داده‌های بارش با توجه به تغییرپذیری خیلی زیادی که دارد، بسیار دشوارتر از بررسی سری‌های زمانی داده‌های دما است. در مورد دما

جدول ۷. داده‌های پرت تأیید شده بارش پس از بررسی‌های هواشناختی در نواحی بیابانی و نیمه‌بیابانی ۲۰۱۸-۱۹۸۹.

نام ایستگاه	کد ایستگاه	تاریخ	مقدار مشاهده شده	مقدار برآورد شده	آنومالی استاندارد شده
زابل	۴۰۸۲۹	۱۹۹۱/۰۱/۲۶	۰/۹	۲۸/۳	-۲۹/۶
زابل	۴۰۸۲۹	۱۹۹۴/۰۳/۰۲	۴۵	۴/۳	۳۵/۳۸
گلپایگان	۹۹۴۱۷	۲۰۱۸/۰۷/۱۹	۱۹۰	۰	۶۱/۴۵

جدول ۸. مشخصات نقاط شکستگی برای بارش نواحی کویری و نیمه‌کویری (مرحله اول) در سری زمانی ۲۰۱۸-۱۹۸۹.

نام ایستگاه	کد ایستگاه	تاریخ	مقدار آماره SNHT
گناباد	۴۰۷۷۸	۲۰۰۴/۱/۵	۲۰/۵
آباده	۴۰۸۱۸	۲۰۰۴/۱/۵	۲۰/۷
یزد	۴۰۸۲۱	۱۹۹۹/۱/۴	۲۴/۴
زابل	۴۰۸۲۹	۱۹۹۴/۱/۴	۲۸/۴



شکل ۱۵. تشخیص نقاط شکستگی در داده‌های ماهانه بارش برای نواحی بیابانی و نیمه‌بیابانی (مرحله اول)، دوره ۲۰۱۸-۱۹۸۹.

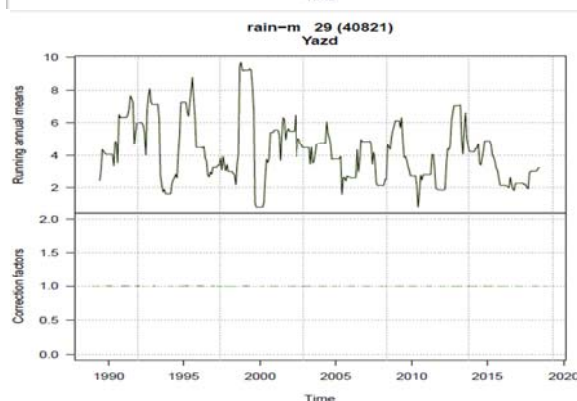
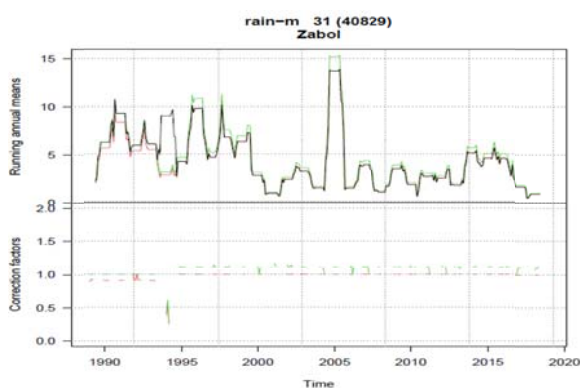
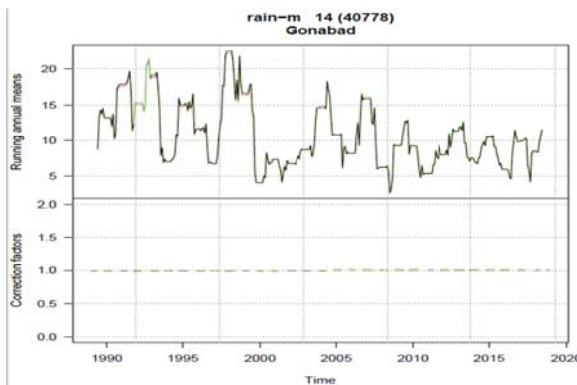
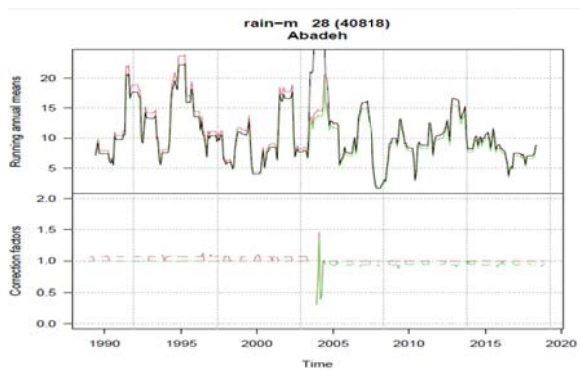
و بر اساس این که برای چه منظوری می‌خواهند مورد استفاده قرار گیرند، انتخاب شدند. برای همگن‌سازی داده‌های روزانه، نقاط شکستگی ماهانه در نظر گرفته شدند. شکل ۱۶ سری‌های کامل بازسازی شده بارش را از زیردوره‌های همگن برای ایستگاه‌ها نشان می‌دهد.

۳-۴. مشخصات سری‌های زمانی ناهمگن دما و بارش ایستگاه‌های هواشناسی کشور در بازه زمانی ۲۰۱۸-۱۹۸۹

در ادامه به‌طور خلاصه مشخصات ایستگاه‌هایی که در آنها عدم همگنی در سری‌های زمانی سی‌ساله مشاهده شد، به‌همراه تاریخ جهش و دلیل احتمالی شناسایی شده آن، به‌ترتیب برای سری‌های زمانی دما در جدول ۹ و برای سری‌های زمانی بارش در جدول ۱۰ نشان داده شده است.

مرحله دوم: در این مرحله آزمون همگنی نرمال استاندارد روی سری‌های کامل ۵۳ سری که از مرحله قبل به‌دست آمده بود اجرا شد و همانند مرحله قبل در هر نقطه‌ای که ماکزیمم آماره SNHT بیشتر از حد آستانه بود، سری شکسته و همه داده‌ها قبل از نقطه شکستگی به سری دیگری با همان طول و عرض جغرافیایی منتقل شدند. این فرایند روی سری‌های جدید اجرا شد تا زمانی که دیگر نقطه ناهمگنی تشخیص داده نشد. در این مرحله هیچ نقطه شکستگی یافت نشد.

مرحله سوم: در این مرحله، همه داده‌های گم‌شده شامل داده‌های پرت یا داده‌هایی که به سری‌های دیگر منتقل شده بودند برآورد شدند. برای برآورد از همان شکسته‌های خود سری به‌عنوان سری‌های مرجع استفاده می‌شود. سری‌های همگن شده در بانک داده ذخیره شده



شکل ۱۶. سری‌های کامل بازسازی شده بارش از زیردوره‌های همگن برای ایستگاه‌های آباد، گناباد، زابل، یزد دوره ۲۰۱۸-۱۹۸۹.

جدول ۹. خلاصه مشخصات ایستگاه‌های دارای شکستگی در سری زمانی سی‌ساله دما (۲۰۱۸-۱۹۸۹).

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی اول	دلیل شکستگی	تاریخ شکستگی دوم	دلیل شکستگی
۱	فردوس	۴۰۷۹۲	۱۹۹۹	-	۲۰۰۷	-
۲	سیرجان	۴۰۸۵۱	۲۰۰۵	-	-	-
۳	زاهدان	۴۰۸۵۶	۲۰۰۶	-	-	-
۴	طیس	۴۰۷۹۱	۲۰۱۳	-	-	-
۵	قوچان	۴۰۷۴۰	۲۰۰۷	-	-	-
۶	سبزوار	۴۰۷۴۳	۲۰۰۳	جابه‌جایی ایستگاه	۲۰۰۸	جابه‌جایی ایستگاه
۷	سمنان	۴۰۷۵۷	۲۰۰۱	جابه‌جایی ایستگاه	-	-
۸	گرمسار	۴۰۷۵۸	۲۰۰۶	-	-	-
۹	کاشان	۴۰۷۸۵	۲۰۰۴	-	-	-
۱۰	اصفهان	۴۰۸۰۰	۱۹۹۴	جابه‌جایی ایستگاه	-	-
۱۱	فرودگاه اصفهان	۴۰۸۰۲	۲۰۱۰	-	-	-
۱۲	شیراز	۴۰۸۴۸	۱۹۹۳	-	۲۰۰۳	جابه‌جایی ایستگاه
۱۳	درودزن	۴۰۸۴۴	۲۰۰۶	-	-	-
۱۴	خورویبابانک	۴۰۷۸۹	۲۰۰۸	-	-	-
۱۵	یزد	۴۰۸۲۱	۲۰۰۱	-	-	-
۱۶	فسا	۴۰۸۵۹	۲۰۱۰	-	-	-
۱۷	شهربابک	۴۰۸۴۹	۲۰۰۷	-	-	-
۱۸	رامهرمز	۴۰۸۱۳	۱۹۹۲	-	-	-
۱۹	بوشهر	۴۰۸۵۸	۲۰۰۲	-	-	-
۲۰	ایلام	۴۰۷۸۰	۲۰۰۳	-	-	-
۲۱	جزیره کیش	۴۰۸۸۲	۱۹۹۶	-	-	-
۲۲	شهرکرد	۴۰۷۹۸	۲۰۰۳	جابه‌جایی ایستگاه	-	-
۲۳	یاسوج	۴۰۸۳۶	۲۰۰۲	-	-	-
۲۴	ساوه	۹۹۳۷۲	۲۰۰۳	-	-	-
۲۵	تهران (شمیران)	۴۰۷۵۱	۲۰۰۴	-	-	-
۲۶	دوشان تپه	۴۰۷۵۳	۲۰۰۷	-	-	-
۲۷	تهران مهرآباد	۴۰۷۵۴	۲۰۰۹	-	-	-
۲۸	رشت	۴۰۷۱۹	۱۹۹۶	جابه‌جایی ایستگاه	۲۰۱۴	-
۲۹	نوشهر	۴۰۷۳۴	۲۰۱۲	-	-	-
۳۰	گرگان	۴۰۷۳۸	۲۰۰۷	جابه‌جایی ایستگاه	-	-
۳۱	آستارا	۴۰۷۰۹	۲۰۱۴	-	-	-
۳۲	بندر انزلی	۴۰۷۱۸	۲۰۱۰	-	-	-
۳۳	بابلسر	۴۰۷۳۶	۲۰۱۲	-	-	-
۳۴	تبریز	۴۰۷۰۶	۲۰۰۶	-	-	-
۳۵	سهند	۴۰۷۰۷	۱۹۹۶	-	-	-
۳۶	اردبیل	۴۰۷۰۸	۱۹۹۷	-	-	-
۳۷	ارومیه	۴۰۷۱۲	۲۰۰۳	جابه‌جایی ایستگاه	-	-
۳۸	مراغه	۴۰۷۱۳	۲۰۰۲	-	-	-
۳۹	پیرانشهر	۴۰۷۲۴	۱۹۹۷	-	۲۰۰۷	-
۴۰	سنندج	۴۰۷۴۷	۱۹۸۸	جابه‌جایی ایستگاه	-	-
۴۱	سردشت	۴۰۷۲۵	۱۹۹۸	-	-	-
۴۲	زنجان	۴۰۷۲۹	۱۹۹۶	-	۲۰۱۴	-
۴۳	(نوژه) همدان	۴۰۷۶۷	۱۹۹۵	-	-	-
۴۴	ملایر	۴۰۷۷۵	۱۹۹۹	-	-	-

جدول ۱۰. خلاصه مشخصات ایستگاه‌های دارای شکستگی در سری زمانی سی‌ساله بارش (۲۰۱۸-۱۹۸۹).

ردیف	نام ایستگاه	کد ایستگاه	تاریخ شکستگی	دلیل شکستگی
۱	گناباد	۴۰۷۷۸	۲۰۰۴	-
۲	آباده	۴۰۸۱۸	۲۰۰۴	-
۳	یزد	۴۰۸۲۱	۱۹۹۹	جابه‌جایی ایستگاه
۴	زابل	۴۰۸۲۹	۱۹۹۴	-
۵	اصفهان	۴۰۸۰۰	۲۰۰۰	-
۶	بندرماهشهر	۴۰۸۳۲	۱۹۹۹	-
۷	امیدیه	۴۰۸۳۳	۱۹۹۴	-
۸	جم	۸۸۱۸۷	۱۹۹۳	-
۹	کرج	۴۰۷۵۲	۲۰۰۵	-
۱۰	تهران(مهرآباد)	۴۰۷۵۴	۲۰۱۲	-
۱۱	رشت	۴۰۷۱۹	۱۹۹۶	جابه‌جایی ایستگاه
۱۲	بابلسر	۴۰۷۳۶	۲۰۰۲	-
۱۳	تبریز	۴۰۷۰۶	۱۹۹۶	جابه‌جایی ایستگاه
۱۴	مراغه	۴۰۷۱۳	۱۹۹۶	-
۱۵	مهاباد	۴۰۷۲۶	۲۰۰۸	-

##### ۵. نتیجه‌گیری

در دوره سی‌ساله ۱۹۸۹-۲۰۱۸ برای دمای بیشینه تعداد ۶۳ مورد داده پرت تشخیص داده شد که ۵۳ مورد آن مربوط به ایستگاه ژئوفیزیک دانشگاه تهران است. در این دوره برای دمای کمینه تعداد ۵۰ مورد داده پرت تشخیص داده شد، که ۱۱ مورد آن مربوط به ایستگاه ژئوفیزیک دانشگاه تهران است. برای بارش در این سری زمانی تعداد ۱۳ مورد داده پرت تشخیص داده شد، که ۵ مورد آن مربوط به ایستگاه ژئوفیزیک دانشگاه تهران است. در مورد بارش، یک مورد هم امکان قضاوت قطعی در مورد داده مورد بررسی با توجه به نقص در اطلاعات جوی تاریخ موردنظر، وجود نداشت.

برای سری‌های زمانی دمای روزانه در دوره ۱۹۸۹-۲۰۱۸ از بین ۱۳۳ ایستگاه (بدون احتساب ایستگاه ژئوفیزیک)، ۸۹ ایستگاه همگن و ۴۴ ایستگاه دارای یک نقطه یا دو نقطه شکستگی بودند. با بررسی فراداده موجود تنها دلیل مورد از جهش‌ها در سری‌های زمانی دما، به‌دلیل جابه‌جایی ایستگاه‌ها تشخیص داده شدند. برای پارامتر بارش نیز در دوره ۳۰ ساله، ۱۵ ایستگاه ناهمگن شناخته

ارزیابی سری زمانی سی‌ساله پارامترهای موردبررسی در بازه زمانی ۱۹۸۹-۲۰۱۸ نشان داد، دمای بیشینه روزانه ایستگاه‌های هواشناسی کشور به‌طور میانگین دارای بیش از ۳ درصد داده گم‌شده می‌باشند. در این میان ایستگاه فرودگاه دزفول (کد ۴۰۷۹۵) با بیش از ۲۴ درصد داده گم‌شده، بیشترین داده گم‌شده دمای بیشینه را در سطح ایستگاه‌های هواشناسی کشور، به خود اختصاص داده است. در مورد دمای کمینه بررسی‌ها نشان داد، به‌طور متوسط ایستگاه‌های هواشناسی دارای بیش از ۴ درصد داده گم‌شده دمای کمینه، در دوره سی‌ساله می‌باشند، که ایستگاه فرودگاه دزفول با حدود ۲۵ درصد داده گم‌شده، بیشترین فقدان دمای کمینه را در بین ایستگاه‌های کشور داراست. در مورد بارش روزانه نیز آمار نشان داد که ایستگاه‌های هواشناسی کشور دارای ۲ درصد داده گم‌شده در این سری زمانی می‌باشند. باز هم ایستگاه فرودگاه دزفول با حدود ۲۳ درصد داده گم‌شده بارش در سری زمانی سی‌ساله رکورددار است.

هواشناسی کل کشور. طرح پژوهشی، کارفرما: سازمان هواشناسی کشور.

دین‌پژوه، ی.، فاخری، ا.، مقدم، م.، میرنیا، م.، و جهانبخش اصل، س. ۱۳۸۲. پهنه‌بندی اقلیمی ایران با استفاده از تحلیل‌های چند متغیره برای استفاده در مطالعات کشاورزی. م. دانش کشاورزی، ۱۳، ۱، ۹۰-۷۱. علیجانی، ب.، ۱۳۸۵، آب و هوای ایران، انتشارات سمت، تهران-ایران.

علیجانی، ب.، ۱۳۹۰، اقلیم‌شناسی سینوپتیک، انتشارات سمت، تهران-ایران.

مسعودیان، س. ا.، ۱۳۸۲، نواحی اقلیمی ایران. م. جغرافیا و توسعه. ۱، ۲، ۱۸۴-۱۷۱.

Alexanderson, H., Mobeg, A., 1997, Homogenization of Swedish temprature data partI: Homogeneity test for linear trends. Int. J. Climatol, 17, 25-34.

Badripour. H., 1992, Islamic Republic of Iran (FAO, Report on the round table on pastoralism. FAO Technical Cooperation Programme, Project TCP/IRA/2255, Rome. FAOSTAT. 2004 ([http://www.fao.org/waicent/portal/statistics\\_en.asp](http://www.fao.org/waicent/portal/statistics_en.asp)).

Guijarro, J. A., 2018, Homogenization of climatic series with Climatol, <https://CRAN.R-project.org/package=climatol>.

<https://data.irimo.ir/>

<https://www.ogimet.com/synops.phtml.en>

<http://www.meteomanz.com/index?l=1>

Paulhus, J. L. H., Kohler, M. A., 1952,

شدند. با بررسی فراداده موجود تنها دلیل ۳ مورد از جهش‌ها در ایستگاه‌های یزد، تبریز و رشت به علت جابه‌جایی این ایستگاه‌ها شناسایی شدند.

به‌دلیل نبود اطلاعات مستند در مورد تغییراتی که در گذشته بر روی ایستگاه‌ها انجام شده، برای بسیاری از ایستگاه‌ها امکان قضاوت در مورد علت شکستگی‌ها در سری‌های زمانی مورد بررسی مقدور نشد.

## مراجع

جوانشیری، ز.، محمدیان، آ.، پاکدامن، م.، فلامرزی، ی.، و کوهی، م. ۱۳۹۹. کنترل کیفیت و همگن‌سازی داده‌های روزانه پارامترهای دما و بارش ایستگاه‌های

Interpolation of missing precipitation records. Monthly Weather Review. 80(8), 129-133.

Rahimzadeh, F. and Nassaji Zavareh, M. 2013, Effects of adjustment for non-climatic discontinuities on determination of temperature trends and variability over Iran. Int. J. Climatol. 2014. 34: 2079-2096 .

Tavakoli, M. and Mohamadian, A. H., 2011, GIS based considerations for development in different Iranian climatic regions. J. American. Science. 7(4). 182-187.

World Meteorological Organization, 2017, WMO Guidelines on the Calculation of Climate Normals. WMO-No.1203.

World Meteorological Organization, 2003, Guidelines on Climate Metadata and Homogenization. WMO/TD No. 1186.

## Exploratory analysis and in-homogeneity study of temperature and rainfall series of meteorological stations in Iran (period 1989-2018)

Rahmdel, M.<sup>1</sup>, Sanaeinejad, S. H.<sup>2\*</sup>, Javanshiri, Z.<sup>3</sup> and Mohamadian, A.<sup>4</sup>

1. Ph.D. Student, Department of Water Science and Engineering, Faculty of Agriculture, Ferdowsi University Mashhad, Mashhad, Iran

2. Professor, Department of Water Science and Engineering, Faculty of Agriculture, Ferdowsi University Mashhad, Mashhad, Iran

3. Assistant Professor, Climatological Research Institute, Mashhad, Iran

4. M.Sc., Climatological Research Institute, Mashhad, Iran

(Received: 7 Feb 2021, Accepted: 25 May 2021)

### Summary

In-situ observations underlies a wide range of planning, applied studies and modeling in various fields and sciences, and using this data in studies and planning without ensuring the accuracy and homogeneity of them, can lead to uncertainty in the results. Major problems that researchers face are the poor quality data, missing data, outliers and in-homogeneity in time series. Inappropriate co-locating of stations, human errors in reading and recording data, errors in measuring equipment, changes in measurement tools, different methods of observation, non maintenance and calibration of equipment, constructions around the stations, changes in the type of instruments and sensors for atmospheric parameters measurements and station relocation during the statistical period are problems that affect the accuracy and homogeneity of the meteorological data. Therefore, in this paper, the minimum and maximum daily temperature series and daily rainfall series at 134 weather stations in Iran were analyzed for outliers and homogeneity over the period 1989-2018. First, Iran was divided into 5 clusters based on climatic characteristics. After clustering, the daily maximum and minimum temperatures and daily rainfall data were statistically analyzed using SPSS software and the percentage of missing data was determined separately 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 needed parameter, and for each station, the other stations belonging to that cluster are considered as reference stations. Based on this algorithm, first the desired series is estimated and standardized by reference series by type (II) regression method. After estimating the series, the standardized anomaly series is calculated, in which the difference between the observed and estimated values is calculated. For detecting outliers, two steps were followed. Original data corresponding to standardized anomalies that were greater than the prescribed thresholds, were detected as outliers. In the second step, in order to ensure the correct detection of the outliers, for temperature series the detected outliers in the first step were compared with the values of the days before and after. If they differed significantly, they would be accepted as outliers and deleted. For the precipitation series, the atmospheric condition of the desired dates were 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 maximum SNHT and all the data before the break 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 all homogeneous series and subseries infilled with the same data estimation procedure using only the reference of their own other fragments.

The maximum, minimum temperatures and precipitation series for 134 weather stations of Iran have an average 3%, 4% and 2% missing values, respectively. In this time series, 63 outliers were detected for the maximum temperature parameter that 53 of them were related to the Geophysics station of the University of Tehran. For the minimum temperature, this number reached 50 that 11 of them belong to the Geophysics station and for the precipitation parameter 13 outliers were identified that 5 of them are related to the Geophysics station. For the daily temperature series (excluding Geophysics station), 89 stations were homogeneous and 44 stations had one or two break points, and for the precipitation series, 15 stations were identified as in-homogeneous.

**Keywords:** Homogenization, Climatol algorithm, Metadata, Outliers, Missing data.

\* Corresponding author:

sanaein@gmail.com