



تشخیص و تحلیل داده‌های پرت زمانی - مکانی مربوط به پارامترهای سرعت باد و فشار ایستگاه‌های هواشناسی کشور به منظور مطالعات تغییرات آب و هوایی

زهره جوانشیری^{۱*}، محسن رحمدل^۲، آزاده محمدیان^۳

^۱استادیار، گروه اقلیم‌شناسی کاربردی، پژوهشکده اقلیم‌شناسی، پژوهشگاه هواشناسی و علوم جو، مشهد

^۲دانش آموخته دکتری هواشناسی کشاورزی، دانشگاه فردوسی مشهد

^۳کارشناس ارشد پژوهشی، گروه اقلیم‌شناسی کاربردی، پژوهشکده اقلیم‌شناسی، پژوهشگاه هواشناسی و علوم جو، مشهد

تاریخ دریافت: ۱۴۰۰/۸/۲۰؛ تاریخ پذیرش: ۱۴۰۰/۹/۱۳

چکیده

تشخیص و تحلیل داده‌های پرت یکی از گام‌های ابتدایی مهم در تحلیل داده‌هاست. یک داده پرت در سری‌های هواشناسی می‌تواند نشان دهنده‌ی یک رخداد فرین باشد یا خطای ناشی از اندازه‌گیری، مشاهده و ثبت است. اگر داده‌های پرتی که حاصل خطای انسانی یا نقص تجهیزات اندازه‌گیری هستند شناسایی نشده و حذف نگردند، به عنوان داده‌ی فرین ثبت خواهند شد و موجب اریبی در نتایج مطالعات اقلیمی می‌شوند. در این مقاله، داده‌های پرت سری‌های زمانی سرعت باد و فشار برای ۱۴۳ ایستگاه هواشناسی کشور در دوره نرمال اقلیمی ۱۹۹۱-۲۰۲۰ مورد تحلیل قرار گرفت. برای این منظور ابتدا داده‌های پرت مکانی با استفاده از الگوریتم کلایماتول مشخص شد و در مرحله دوم با تحلیل‌های زمانی و هواشناختی، خطاها شناسایی شدند. در مرحله اول برای پارامترهای فشار ایستگاه، فشار تبدیل شده ایستگاه به سطح دریا، فشار بخار، سرعت باد و سرعت باد ماکزیمم به ترتیب ۴۰، ۴۲، ۹۳، ۵۲ و ۴۱ داده پرت شناسایی شد و در مرحله دوم به ترتیب ۲۰، ۱۰، ۵۶، ۲۰ و ۲۷ تعداد از این داده‌ها خطا تشخیص داده شدند. این نتایج به تفکیک ایستگاه و تاریخ گزارش شده‌اند، تا مورد استفاده محققان در سایر مطالعات، بویژه مطالعات تغییر اقلیم، قرار بگیرند.

واژه‌های کلیدی: تشخیص داده پرت، خطا، سرعت باد، فشار

مقدمه

روزانه حجم زیادی از داده‌های هواشناسی توسط تجهیزات و دستگاه‌های اندازه‌گیری مختلف در ایستگاه‌های هواشناسی کشور تولید می‌شود. این داده‌ها یکی از نیازهای اساسی محققین برای درک شرایط آب و هوایی و آشکارسازی تغییرات اقلیمی می‌باشند. یکی از مسائل مهم در ارتباط با داده‌های هواشناسی، تشخیص و تحلیل داده‌های پرت می‌باشد. داده پرت داده‌ای است که از بقیه‌ی داده‌ها فاصله زیادی دارد (بارنت^۲، ۱۹۹۴). اما برای داده‌های مکانی -

زمانی، همچون داده‌های هواشناسی، داده پرت مکانی - زمانی، داده‌ای است که از سایر داده‌ها در همسایگی مکانی و زمانی خود به‌طور معنی داری متفاوت است (پرازانتی^۳، ۲۰۱۶). در سری‌های زمانی هواشناسی، داده پرت می‌تواند خطا یا یک رخداد فرین باشد. به منظور بالا بردن کیفیت داده‌ها، خطاها باید شناسایی و اصلاح شوند، در غیر اینصورت نتایج ناصحیحی را در مطالعات تغییر اقلیم شاهد خواهیم بود. لذا تشخیص داده‌های پرت و ماهیت آن‌ها اهمیت ویژه‌ای دارد.

*نویسنده مسئول: javanshiri@cri.ac.ir

مبتنی بر همسایگی است، برای تعیین نقاط پرت و بررسی همگنی سری‌ها ارائه کرد. در این روش، ابتدا ایستگاه‌های مرجع، که تغییر پذیری مشابهی با ایستگاه مورد نظر دارند، شناسایی می‌شوند و سری می‌شوند. سپس با استفاده از سری آنومالی (تفاضل مقادیر مشاهده شده و برآورد شده) استاندارد شده، داده‌های پرت تعیین می‌شوند. روش کلایماتول، روشی کارا و ساده برای تعیین داده‌های پرت ناشی از خطا می‌باشد و در بسیاری از مطالعات مورد استفاده قرار گرفته است (کل^{۱۳} و همکاران، ۲۰۲۰؛ آروفو^{۱۴} و کارلو^{۱۵}، ۲۰۱۹؛ اسکرینک^{۱۶} و همکاران، ۲۰۲۰).

سرعت باد و فشار، از پارامترهای مورد اهمیت در مطالعات اقلیمی هستند. تقاضای روزافزون برای منابع تجدیدپذیر انرژی در دهه‌های آینده (IEA^{۱۷}، ۲۰۱۳) مستلزم آن است که ما درک روشنی از منابع باد قابل دسترس و حساسیت این منابع در برابر تغییرات اقلیمی داشته باشیم (شفر^{۱۸} و همکاران، ۲۰۱۲). همینطور تغییر در فشار هوا تأثیر بزرگی بر تغییرات آب و هوایی دارد. فشار، گردش هوا را کنترل می‌کند، و بنابراین بر نحوه حرکت رطوبت تأثیر می‌گذارد. تغییرات در گردش هوا می‌تواند بارندگی، دما، باد و طوفان را تغییر دهد.

در این مقاله، به تشخیص و تحلیل داده‌های پرت سری‌های زمانی متوسط سرعت باد، بیشینه سرعت باد، فشار ایستگاه، فشار تبدیل شده ایستگاه به سطح دریا و فشار بخار، ۱۴۳ ایستگاه هواشناسی کشور برای دوره نرمال اقلیمی اخیر (۱۹۹۱-۲۰۲۰) پرداخته‌ایم. از آنجا که هدف اصلی ما از این مطالعه، تعیین داده‌های پرت ناشی از خطا و بالا بردن کیفیت داده‌ها می‌باشد، ابتدا با استفاده از روش کلایماتول، داده‌های پرت را تعیین کردیم و سپس برای اطمینان از این که داده‌های پرت تشخیص داده شده خطا هستند،

تاکنون مطالعات زیادی در زمینه تعیین داده‌های پرت انجام شده است. آدام^۱ و همکاران (۲۰۰۴) از روش مبتنی بر فاصله، که یک روش مکانی است، برای تشخیص نقاط پرت استفاده کردند. چنگ^۲ و همکاران (۲۰۰۴) یک راه بسیار ساده برای تعریف داده پرت مکانی-زمانی ارائه دادند و یک روش چهار مرحله‌ای روشن برای تشخیص آن‌ها را پیشنهاد دادند. لو^۳ و لیانگ^۴ (۲۰۰۴) یک روش کلاس‌بندی فازی موجک^۵ را برای تشخیص داده‌های پرت هواشناسی ارائه کردند. یوکیانگ و همکاران (۲۰۰۵) روش فرمول‌بندی شده‌ای را برای تعریف داده‌های پرت در داده‌های شبکه ای مکانی-زمانی معرفی کردند. هدف آن‌ها تشخیص داده‌های پرت معنی‌دار در مجموعه داده‌های هواشناسی بود. وو^۶ و همکاران (۲۰۱۰)، الگوریتمی را جهت شناسایی داده‌های پرت مکانی-زمانی بارش ارایه کردند. سپس با استفاده از این الگوریتم دنباله‌ی داده‌های پرت را برای داده‌های بارش امریکای جنوبی تعیین کرده و با مقایسه آن‌ها با فاز انسو (ENSO^۷)، به تحلیل آن‌ها پرداختند. در مطالعات ذکر شده هدف اصلی از تعیین داده‌های پرت، مشخص کردن رخداد‌های فرین بوده است. اما مطالعاتی برای تشخیص داده‌های پرت ناشی از خطا و تصحیح آن‌ها، با هدف بالا بردن کیفیت داده‌ها، نیز انجام شده است. ما^۸ و همکاران (۲۰۱۷)، برای شناسایی داده‌های پرت دما الگوریتمی را بر اساس پیش‌بینی پنجره‌ای کشویی^۹ پیشنهاد دادند. رانجان^{۱۰} و همکاران (۲۰۲۰)، روش پیش‌بینی پنجره‌ای کشویی اصلاح شده را برای تشخیص و اصلاح داده‌های پرت ارائه کردند. گویجارو^{۱۱} (۲۰۱۸)، الگوریتم کلایماتول^{۱۲} را که روشی

1. Adam
2. Cheng
3. Lu
4. Liang
5. Wavelet fuzzy classification
6. Wu
7. El Niño Southern Oscillation
8. Ma
9. Sliding Window Prediction
10. Ranjan
11. Guijarro
12. CIIMATOL

13. Coll
14. Aruffo
15. Carlo
16. Skrynyk
17. International Energy Agency
18. Schaeffer

بررسی‌های زمانی و هواشناختی با توجه به ماهیت پارامتر انجام شد.

مواد و روش‌ها

منطقه مورد مطالعه: منطقه مورد مطالعه در این تحقیق کشور ایران است و تعداد ایستگاه‌های همدیدی که دوره ۱۹۹۱-۲۰۲۰ را پوشش می‌دهند،

۱۴۳ ایستگاه می‌باشد که موقعیت این ایستگاه‌ها در شکل (۱) نشان داده شده است. پارامترهای مورد بررسی در این تحقیق، فشار سطح ایستگاه، فشار تبدیل شده ایستگاه به سطح دریا، فشار بخار، متوسط سرعت باد و سرعت باد ماکزیمم می‌باشند که داده‌های مربوط به آن‌ها از سازمان هواشناسی کشور دریافت شدند.



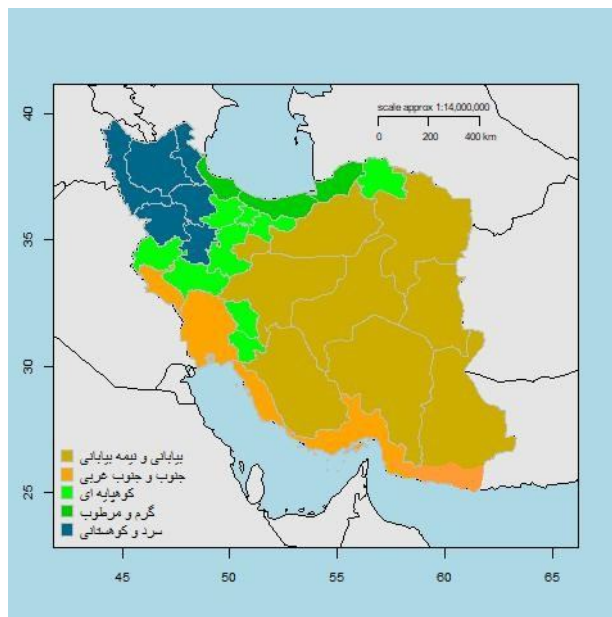
شکل ۱: موقعیت ایستگاه‌های مورد مطالعه

روش کار

ابتدا کشور براساس مشخصه‌های آب و هوایی خوشه‌بندی شد. در مطالعاتی که در این زمینه صورت گرفته است، عمدتاً از روش‌های تحلیل خوشه‌ای و مؤلفه‌های اصلی استفاده شده‌اند (توکلی و محمدیان، ۲۰۱۱؛ بدری پور، ۱۹۹۲؛ علیجانی، ۱۳۸۵؛ دین‌پژوه، ۱۳۸۲؛ مسعودیان، ۱۳۸۲). بر اساس نتایج مطالعات فوق، می‌توان ایران را به ۵ ناحیه تقسیم‌بندی کرد (رحیم‌زاده و نساج زواره، ۲۰۱۴). همانطور که در شکل (۲) نشان داده شده است، این خوشه‌بندی‌ها عبارتند از:

- خراسان جنوبی و رضوی.
- نواحی جنوب و جنوب غربی، شامل استان‌های خوزستان، ایلام، بوشهر، هرمزگان و جنوب سیستان و بلوچستان.
- نواحی کوهپایه‌ای، شامل استان‌های تهران، مرکزی، قزوین، چهارمحال و بختیاری، کرمانشاه و خراسان شمالی.
- نواحی گرم و مرطوب، شامل استان‌های گیلان، مازندران و گلستان.
- نواحی سرد و کوهستانی، شامل استان‌های آذربایجان غربی و شرقی، اردبیل، استان‌های روی رشته کوه زاگرس، کردستان، همدان و زنجان.

- نواحی بیابانی و نیمه‌بیابانی شامل استان‌های قم، اصفهان، یزد، کرمان، سیستان و بلوچستان،



شکل ۲- خوشه‌بندی ایران براساس مشخصه‌های آب و هوایی

پس از تعیین داده‌های پرت با استفاده از الگوریتم کلایماتول، با بررسی‌های هواشناختی با توجه به ماهیت پارامتر مورد مطالعه، خطا بودن داده‌های پرت را بررسی می‌کنیم. بررسی‌هایی که انجام شد، به شرح زیر می‌باشد:

- داده مورد نظر در دامنه کمینه و بیشینه پارامتر مورد نظر قرار دارد یا خیر. اگر در این دامنه قرار ندارد، خطا است.
- داده مورد نظر با داده‌های ساعتی ثبت شده در همان تاریخ مقایسه می‌شود، در صورتی که با داده‌های ساعتی همخوانی نداشت به‌عنوان خطا در نظر گرفته می‌شود.
- اگر در داده‌های ساعتی اختلاف زیادی وجود داشت، وقوع پدیده‌های خاص در آن روز بررسی می‌شود. در صورتی که پدیده خاصی رخ نداده باشد، داده مورد نظر به‌عنوان خطا ثبت می‌شود.
- اگر داده ساعتی موجود یا کافی نبود با داده‌های روزهای قبل و بعد مقایسه می‌شود و در صورت عدم همخوانی به عنوان خطا ثبت می‌شود.
- برای پارامتر فشار بخار از روابطی که بین فشار بخار با رطوبت نسبی و دمای خشک وجود دارد، نیز استفاده گردید. در صورتی که داده‌های ساعتی این پارامتر با هم همخوانی نداشتند، به بررسی

هدف از خوشه‌بندی اولیه، این است که ایستگاه‌های مرجع از بین ایستگاه‌هایی که در شرایط نسبتاً مشابه اقلیمی با ایستگاه مورد نظر قرار دارند، انتخاب شوند. حال الگوریتم کلایماتول با استفاده از نرم‌افزار R، برای هر خوشه به‌طور جداگانه اجرا می‌شود. مراحل کار الگوریتم کلایماتول به‌صورت زیر است:

گام اول: برای هر ایستگاه، ایستگاه‌های مرجع بر اساس تغییرپذیری پارامتر مورد نظر تعیین می‌شوند.

گام دوم: سری مربوط به هر ایستگاه با استفاده از سری‌های مرجع آن ایستگاه برآورد می‌شوند. این برآورد به روش رگرسیون نوع II (رگرسیون عمودی) می‌باشد.

گام سوم: سری آنومالی که عبارت است از تفاضل مقدار مشاهده شده و برآورد شده محاسبه شده و استاندارد می‌شوند.

گام چهارم: نمودار هیستوگرام آنومالی‌های استاندارد شده، رسم می‌شوند.

گام پنجم: با توجه به هیستوگرام رسم شده، آستانه‌ها برای تعیین داده‌های پرت مشخص می‌شوند.

گام ششم: آنومالی‌هایی که کوچکتر یا بزرگتر از آستانه‌های تعیین شده، وجود داده‌ی پرت را نشان می‌دهند.

این نمودارها، آستانه‌های مورد نیاز برای تعیین داده‌های پرت مشخص می‌شود. آستانه‌ها با توجه به شکل هیستوگرام تعیین می‌شوند. مقادیری که در دماهای توزیع قرار دارند و در شکل‌ها با رنگ قرمز مشخص شده‌اند، در واقع نشان‌دهنده مقادیر آنومالی استاندارد شده‌ای از فشار سطح ایستگاه هستند که تفاوت زیادی با ایستگاه‌های مرجع‌شان دارند و به عنوان داده پرت شناخته می‌شوند. به طور مثال در خوشه اول، با توجه به شکل توزیع آستانه‌ها ۲۵- و ۳۰ تعیین شده‌اند و آنومالی‌هایی که بیشتر از ۳۰ و یا کمتر از ۲۵- باشند، نشان‌دهنده داده پرت هستند.

برای آنومالی‌هایی که خارج از آستانه‌های تعیین شده قرار دارند، مشخصات ایستگاه و تاریخ مربوطه به همراه داده مشاهداتی ثبت شده استخراج شد. سپس برای تعیین این‌که، این نقاط خطا هستند یا خیر، به بررسی‌های هواشناسی پرداختیم. برای این‌کار ابتدا بررسی شد که داده مورد نظر در دامنه فشار سطح زمین قرار داشته باشد، به‌طور مثال برای ایستگاه رامسر در تاریخ ۱۶/۴/۲۰۱۶، مقدار ۲۱۸۳،۵۵ ثبت شده است که در دامنه مورد نظر نیست و مسلماً خطا می‌باشد. سپس داده‌های ساعتی آن روز مورد بررسی قرار گرفت، اگر داده مورد نظر با داده‌های ساعتی همخوانی نداشت، خطا محسوب شد. برای ایستگاه‌های شهرضا، انار، فیروزکوه و آستارا این مشکل وجود داشت. در صورتی‌که داده‌های ساعتی در روز مورد نظر کافی نبود، با داده‌های روزهای قبل و بعد مقایسه شد. ۱۳ مورد خطای تشخیص داده شده مربوط به ایستگاه‌های جم و درودزن از این نوع بودند. به طور کلی، از بین ۴۰ داده پرت تشخیص داده شده توسط الگوریتم کلایماتول ۲۰ داده خطا بودند. این خطاها مربوط به ایستگاه‌های شهرضا، انار، درودزن، جم، فیروزکوه، آستارا و رامسر می‌باشند. ایستگاه و تاریخ مربوط به خطاها در جدول (۱) آمده است.

دمای خشک و رطوبت نسبی پرداخته شد. در مواردی که رطوبت نسبی و دمای خشک صحیح به نظر می‌رسید، مجدداً از طریق نرم‌افزار Mettabwin به محاسبه فشاربخار موردنظر پرداخته شد. در صورتی که مقدار محاسبه شده با مقدار ثبت شده تفاوت داشت، به عنوان خطا در نظر گرفته شد.

رگرسیون عمودی یا نوع II: روش رگرسیون عمودی اولین بار توسط پالوس^۱ و کوهرل^۲ (۱۹۵۲) برای برآورد داده‌های گم‌شده بارش ارائه شد. در رگرسیون نوع II، فاصله عمودی نقاط تا خط رگرسیون کمینه می‌شود، در صورتی‌که در رگرسیون نوع I، انحراف نقاط تا خط رگرسیون در راستای محور Y می‌نیم می‌شوند. داگت^۳ (۱۹۷۹)، معادله خط رگرسیون نوع II را به روش کمترین مربعات به دست آورد. جایگزین‌هایی برای این معادله با تقریب‌های بسیار نزدیک وجود دارند، ساده‌ترین آن‌ها RMA^۴ نام دارد و به صورت رابطه (۱) نشان داده می‌شود.

$$\hat{y} = x \quad (1)$$

در رابطه فوق، x و y مقادیر نرمال شده متغیرهای مستقل و وابسته هستند، که به ترتیب در رابطه (۲) و (۳) نشان داده شده است.

$$x = \frac{X - m_X}{s_X} \quad \text{رابطه (۲)}$$

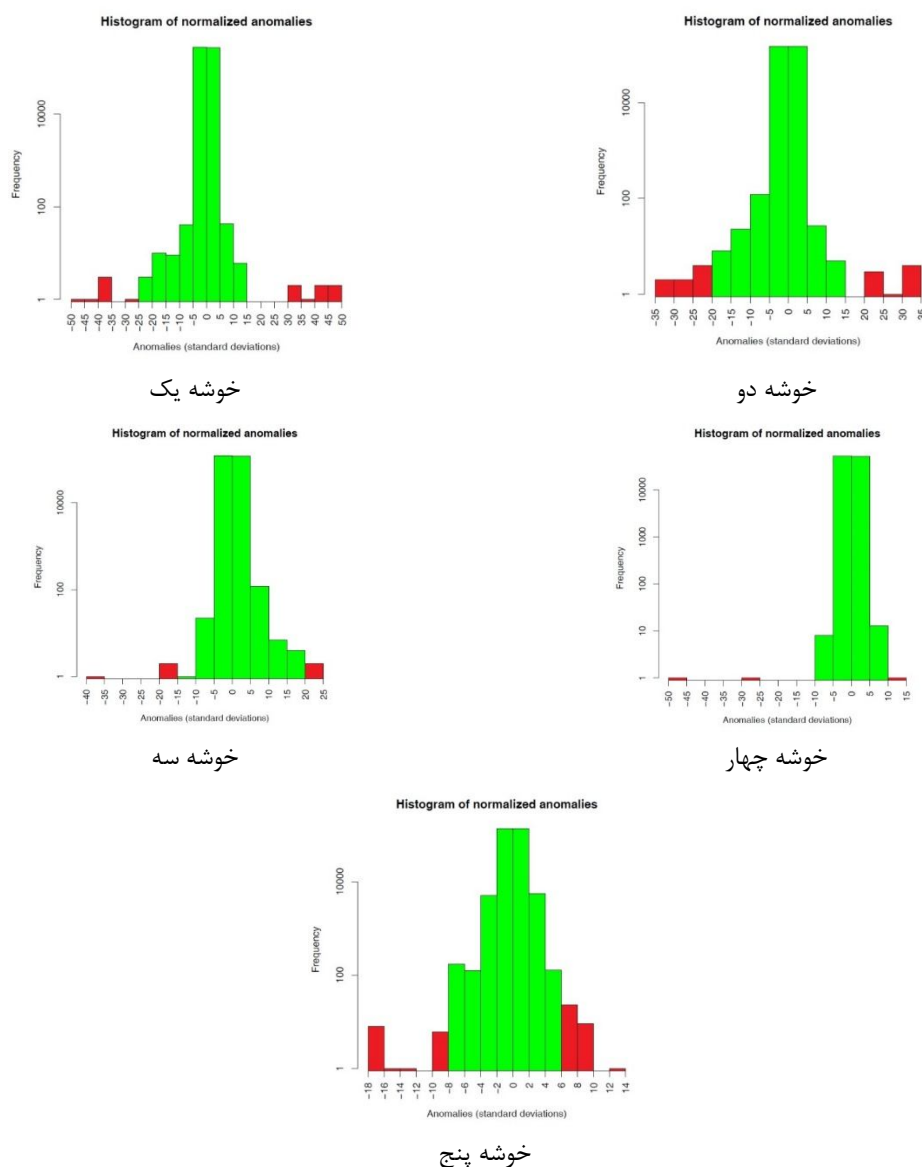
$$y = \frac{Y - m_Y}{s_Y} \quad \text{رابطه (۳)}$$

در روابط فوق، m و s به ترتیب میانگین و انحراف معیار هستند.

یافته‌های تحقیق

۱. **فشار سطح ایستگاه (QFE):** شکل (۳) هیستوگرام آنومالی داده‌های فشار سطح ایستگاه را برای ۵ خوشه به تفکیک نشان می‌دهد. با استفاده از

1. Paulhus
2. Kohler
3. Daget
4. Reduced Major Axis



شکل ۳: هیستوگرام آنومالی استاندارد شده داده‌های فشار سطح ایستگاه (QFE) به تفکیک در هر خوشه برای تشخیص حدود داده‌های پرت

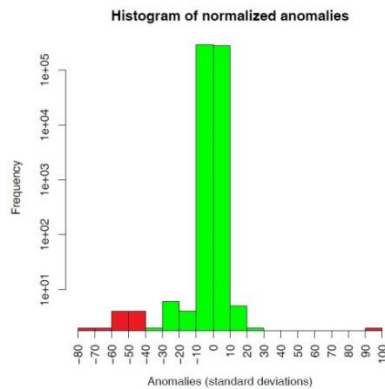
مشخص شده اند. داده‌های پرت مشخص شده با استفاده از این آستانه‌ها تعیین شدند. برای تشخیص خطاها به تحلیل هواشناختی همانند آنچه در بخش قبل گفته شد پرداختیم. برای ایستگاه‌های نیشابور، تربت جام، آباده، لار، خاش، جم، اراک، سیرجان، قراخیل و زرینه خطاهایی مشخص شدند. خطاهای مربوط به مورد ایستگاه‌های سیرجان، قراخیل و زرینه به دلیل عدم همخوانی آن‌ها با داده‌های ساعتی بود اما در مورد سایر ایستگاه‌ها، مقادیری که خطا تشخیص داده شده اند، خارج از دامنه متعارف تغییرات فشار

شکل (۶) فراوانی خطاها و داده‌های پرت فشار ایستگاه را به تفکیک در هر خوشه نشان می‌دهد، همانطور که مشاهده می‌شود تعداد خطاها در خوشه یک و دو بیشتر از سایر مناطق است و بیشتر این خطاها مربوط به دو ایستگاه جم و درودزن می‌باشند.

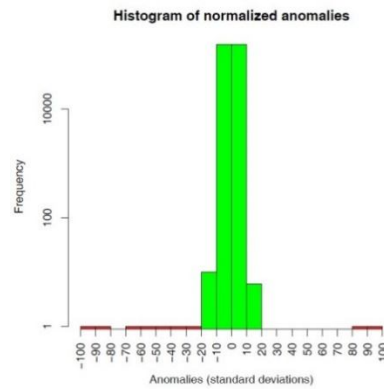
۲. فشار تبدیل شده ایستگاه به سطح متوسط دریا در دمای واقعی (QFF): شکل (۴) هیستوگرام آنومالی داده‌های فشار تبدیل شده ایستگاه به سطح متوسط دریا را برای هر خوشه نشان می‌دهد. آستانه تعیین شده برای تشخیص داده‌های پرت، در شکل‌ها

هر خوشه نشان می‌دهد، همانطور که مشاهده می‌شود تعداد داده‌های پرت تشخیص داده شده و همینطور خطاها در خوشه یک بیشتر از سایر مناطق است و این امر با توجه به اینکه خوشه یک ایستگاه‌های بیشتری را شامل می‌شود قابل انتظار است.

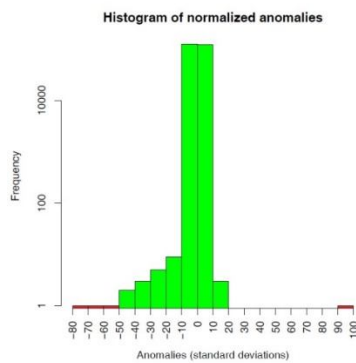
سطح دریا در ایستگاه بودند. به‌طور کلی از بین ۴۲ داده پرت تشخیص داده شده توسط الگوریتم کلایماتول ۱۰ داده خطا بودند. ایستگاه و تاریخ مربوط به خطاها در جدول (۱) آمده است. شکل (۶) فراوانی خطاها و داده‌های پرت فشار (QFF) را به تفکیک در



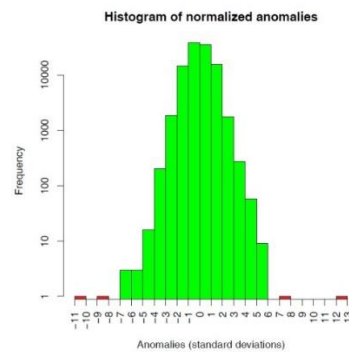
خوشه یک



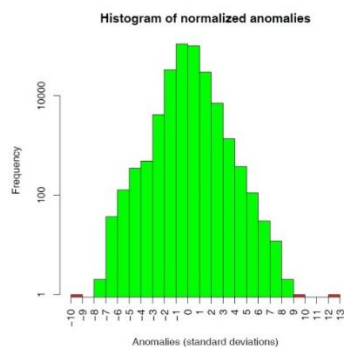
خوشه دو



خوشه سه



خوشه چهار

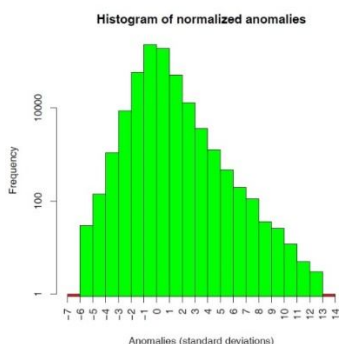


خوشه پنج

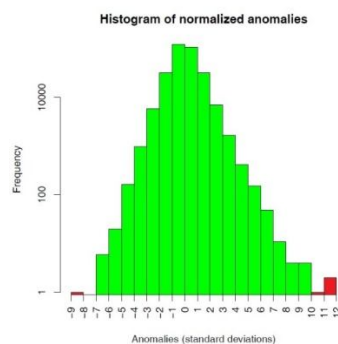
شکل ۴- هیستوگرام آنومالی داده‌های فشار تبدیل شده ایستگاه به سطح متوسط دریا (QFF) به تفکیک در هر خوشه برای تشخیص حدود داده‌های پرت

۳. فشار بخار: شکل (۵) هیستوگرام آنومالی فشار بخار را به تفکیک در ۵ خوشه نشان می‌دهد. نواحی قرمز در شکل نشان‌دهنده آنومالی مربوط به داده‌های پرت است. برای تشخیص خطاها، در گام اول بررسی شد که داده‌ها در دامنه استاندارد قرار دارد یا خیر. اگر داده ثبت شده در دامنه استاندارد فشار بخار قرار نداشت به‌عنوان خطا در نظر گرفته شد. به طور مثال در مورد ایستگاه گل‌مکان در تاریخ ۱/۱۹/۲۰۱۸ مقدار

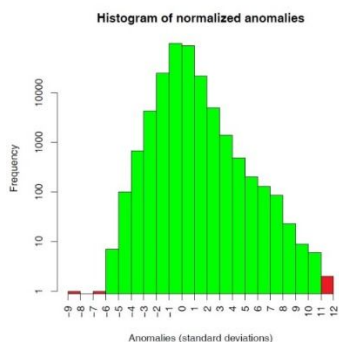
۱،۲- ثبت شده است، یا ایستگاه گلپایگان از تاریخ ۹/۵/۲۰۱۸ تا ۲۳/۶/۲۰۱۸ مقادیر منفی ثبت شده است، که با توجه به اینکه فشار بخار نمی‌تواند منفی باشد، خطا محسوب می‌شوند. در گام دوم داده‌های ساعتی روز مورد نظر بررسی شد، در صورتی که مقدار ثبت شده با داده‌های ساعتی همخوانی نداشت، آن را خطا در نظر گرفتیم.



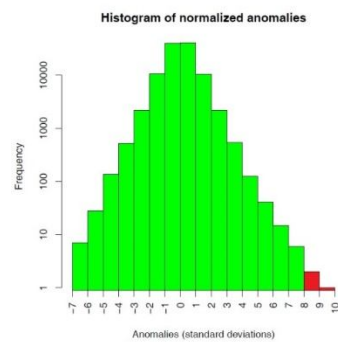
خوشه یک



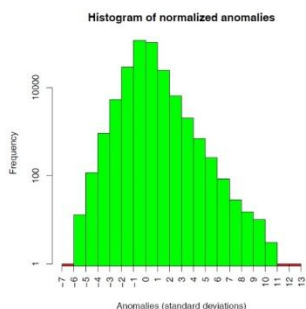
خوشه دو



خوشه سه



خوشه چهار



خوشه پنج

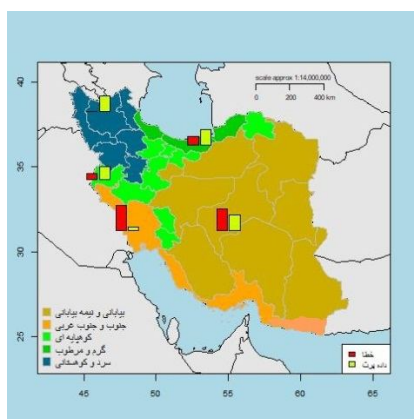
شکل ۵- هیستوگرام آنومالی داده‌های فشار بخار به تفکیک در هر خوشه برای تشخیص حدود داده‌های پرت

۴. در صورتی که داده‌های ساعتی کامل نبودند، به بررسی دمای خشک و رطوبت نسبی پرداخته شد. در

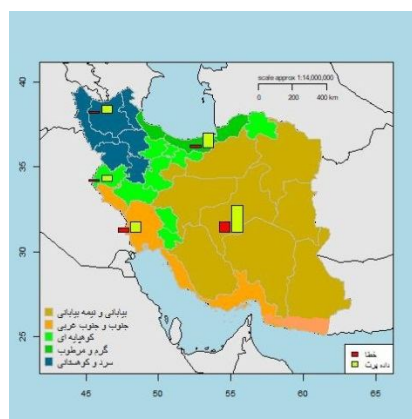
مورد ایستگاه‌های خوروبابانک، بستان، آبعلی، بندر انزلی، نوشهر، سراب و مهاباد مقادیر خطای تشخیص

کلی ۹۳ داده پرت در مرحله اول شناسایی شد، در تحلیل مرحله دوم ۵۶ داده خطا تشخیص داده شدند که از این تعداد ۴۶ داده مربوط به ایستگاه گلپایگان می‌باشد. بعد از بررسی‌های انجام شده مشخص شد که علت خطا در فشار بخار ایستگاه گلپایگان ناشی از خطای ثبت رطوبت می‌باشد که در تاریخ‌های ذکر شده، صفر ثبت شده‌اند. ایستگاه و تاریخ خطاها به تفکیک در جدول (۱) آمده است. شکل (۶) فراوانی خطاها و داده‌های پرت فشار بخار را به تفکیک در هر خوشه نشان می‌دهد، همانطور که مشاهده می‌شود، داده‌های پرت و تعداد خطاها در خوشه یک بسیار بیشتر از سایر خوشه‌هاست و علت آن همان ایستگاه گلپایگان می‌باشد.

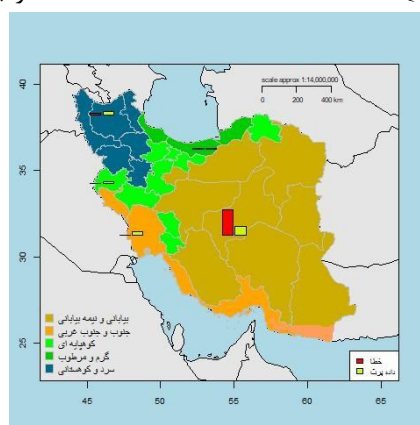
داده‌شده به دلیل عدم همخوانی فشار بخار با رطوبت نسبی ثبت شده بود. در مواردی که رطوبت نسبی و دمای خشک صحیح به نظر می‌رسید، فشار بخار مجدداً از طریق نرم افزار Mettabwin محاسبه شد و در صورت تفاوت با مقدار ثبت شده، خطا در نظر گرفته شد. به طور مثال برای ایستگاه شهر بابک با استفاده از الگوریتم کلایماتول ۵ داده پرت تشخیص داده شد، با توجه به اینکه داده‌های ساعتی برای تاریخ‌های مورد نظر موجود نبودند، دمای خشک و رطوبت نسبی روزهای مورد نظر بررسی شد و در نهایت با استفاده از نرم‌افزار Mettabwin، فشار بخار دوباره محاسبه شد و نتیجه این بررسی نشان داد که این داده‌ها خطا نیستند. اما در مورد ایستگاه بافق و زرینه خطاها به دلیل اشتباه در محاسبه بودند. به طور



فشار ایستگاه



فشار (QFF)



فشار بخار

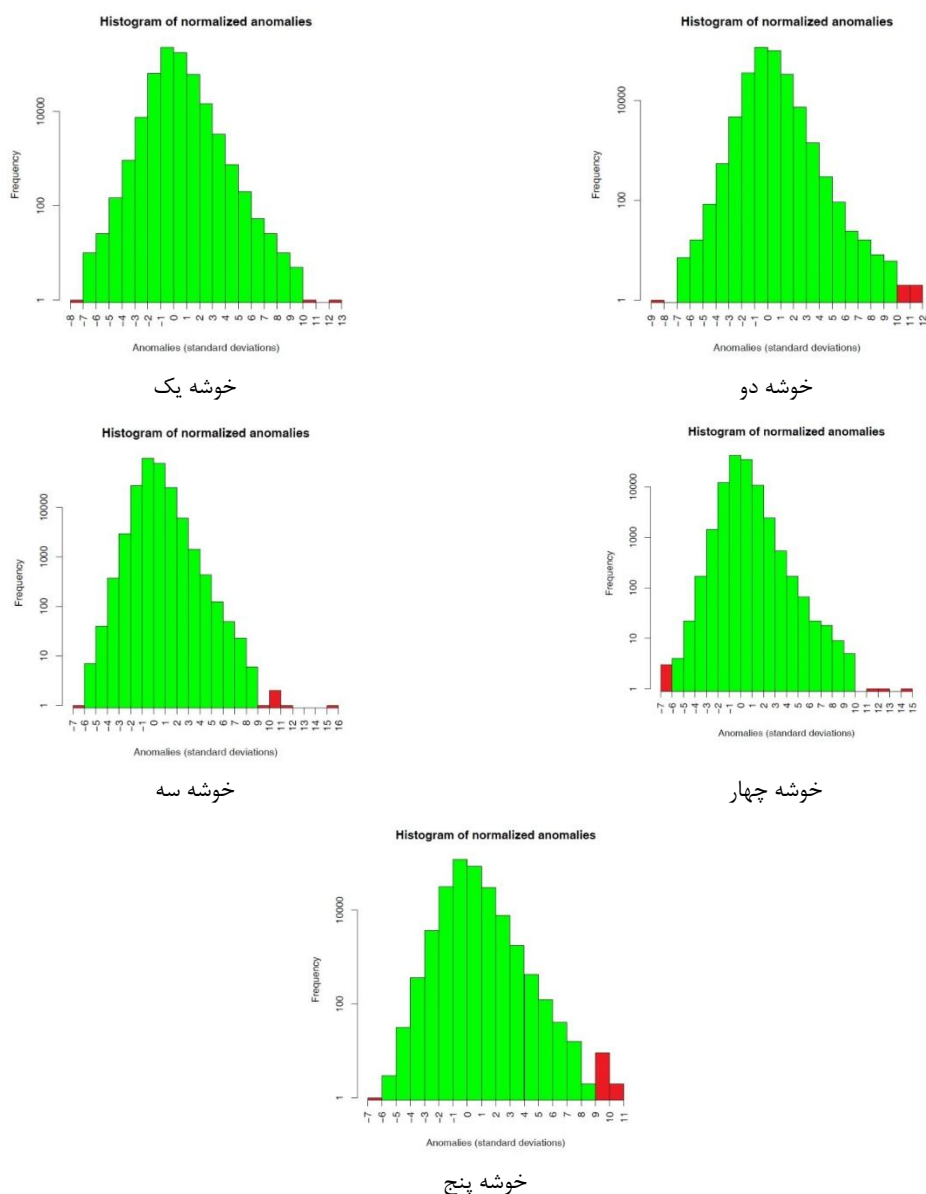
شکل ۶- فراوانی خطاها و داده‌های پرت پارامترهای فشار QEF، QFF و فشار بخار در هر خوشه

جدول ۱- خطاهای تشخیص داده شده برای داده‌های QEE, QFE و فشار بخار

خوشه		۱		۲		۳	۴	۵
فشار ایستگاه (QFE)	نام ایستگاه تاریخ مشاهداتی	فشار (QFF)	نام ایستگاه تاریخ مقدار مشاهداتی	فشار بخار	نام ایستگاه تاریخ مشاهداتی	فیروزکوه ۱۹۹۳/۱۰/۱۱ ۷۴۰ ۱۹۹۳/۱۴/۱۱	آستارا ۲۰۱۹/۱۴/۰۴ ۹۶۱,۸ ۲۰۱۹/۱۵/۰۴ ۹۸۱,۱	رامسر ۰۱/۰۴/۲۰۱۶ ۲۱۸۳,۵۵
	شهرضا ۲۰۱۸/۲۲/۰۵ انار ۲۰۱۹/۰۵/۰۵ دروزن ۱۹۹۴/۱۵/۰۷ ۱۰۱۸,۴ ۱۰۱۵,۸ ۱۰۰۸ ۱۰۰۷,۴ ۱۰۰۰	نیشابور ۲۱۳۲,۶۶/۴/۲۰۱۴ تربت جام ۰۶/۱۲/۲۰۱۴ آباد ۰۷/۰۷/۲۰۲۰ سیرجان ۰۲/۰۳/۲۰۱۴ لار ۲/۳/۲۰۱۴ ۲۱۳۷,۹	گل‌مکان ۱۱/۹/۲۰۱۸ خور ۱۹۹۳/۲۷/۵ بیابانک ۸,۲ ۲۰۱۸/۳۰/۱۰ ۲۷,۵ کلیپگان از ۹/۵/۲۰۱۸ تا مقادیر منفی ۲۳/۶/۲۰۱۸	بستان ۱۹۹۶/۲۰/۱۶ ۳۹,۶	بستان ۱۹۹۶/۲۳/۰۶ ۱۰۲۱,۸ ۱۰۱۹,۴ ۱۰۲۷,۴ ۱۰۲۲,۴ ۱۰۲۰ ۱۰۱۷,۶ ۱۰۱۸,۶ ۱۰۱۷,۲	۷۷۰,۷ ۱۹۹۳/۱۰/۱۱ ۷۴۰ ۱۹۹۳/۱۴/۱۱	۹۶۱,۸ ۲۰۱۹/۱۴/۰۴ ۹۸۱,۱ ۲۰۱۹/۱۵/۰۴	۲۱۸۳,۵۵ ۰۱/۰۴/۲۰۱۶
						۲۲۹۹,۳ ۳/۱۷/۲۰۱۵	۱۰۱۳,۲ ۳/۲۵/۲۰۱۶	۹۹۸,۵ ۲/۴/۲۰۱۹
						۴۱۹۹۶/۲۵/۷ آبعلی	۳۵ ۲۰۰۲/۱۸/۷ نوشهر ۲۵,۹۲۰ ۲۰۱۳/۳/۱۰	۱۹,۴ ۱۹۹۹/۳/۱۶ مهاباد

۵. **سرعت باد:** شکل (۷) هیستوگرام آنومالی داده‌های سرعت باد را برای ۵ خوشه نشان می‌دهد. با توجه به توزیع آنومالی استاندارد شده، آستانه‌ها برای تعیین داده‌های پرت تعیین شده‌اند و مقادیری که خارج از آستانه‌های تعیین شده قرار دارند با رنگ قرمز مشخص شده‌اند. داده‌های پرت مشخص شده در این مرحله به تفکیک ایستگاه ثبت شدند. برای تحلیل مرحله دوم، داده‌های ساعتی همان روز و داده‌های

روزهای قبل و بعد آن مورد بررسی قرار گرفت. همین‌طور با رفتار پارامتر در دوره آماری ماه مورد نظر مورد مقایسه قرار گرفت. در مرحله اول ۵۲ داده پرت تشخیص داده شده است، که از این تعداد ۲۰ داده خطا بودند. خطاها در جدول (۲) آمده‌اند، همان‌طور که مشاهده می‌شود بیشتر خطاها مربوط به ایستگاه‌های بندرعباس، سردشت و آبعلی می‌باشد.



شکل ۷- هیستوگرام آنومالی داده‌های سرعت باد به تفکیک در هر خوشه برای تشخیص حدود داده‌های پرت

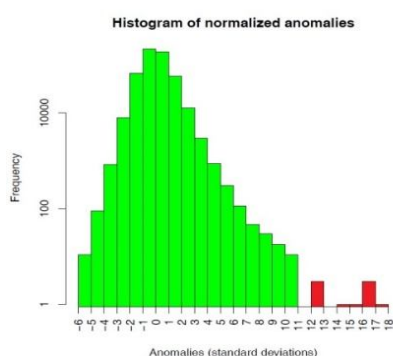
نواحی قرمز رنگ در شکل داده‌های پرت را نشان می‌دهد. برای تحلیل مرحله دوم، همخوانی داده مورد نظر با داده‌های ساعتی همان روز و داده‌های روزهای قبل و بعد مورد بررسی قرار گرفتند. علاوه بر این، برای داده‌های پرت بزرگ (بادهایی با سرعت بسیار زیاد)، وقوع پدیده خاص در آن روز بررسی شد، همینطور وقوع بادهایی با سرعت مشابه در ایستگاه مورد نظر در طول دوره آماری مورد مطالعه قرار گرفت و در صورتی که سابقه ثبت چنین بادهایی وجود نداشت و همینطور پدیده خاصی رخ نداده بود، به

شکل (۹) فراوانی خطاها و داده‌های پرت سرعت باد را به تفکیک در هر خوشه نشان می‌دهد. همانطور که مشاهده می‌شود، تعداد داده‌های پرت و خطاهای خوشه یک نسبت به سایر خوشه‌ها کمتر است و این مسأله با توجه به اینکه این خوشه ایستگاه‌های زیادی را شامل می‌شود قابل تأمل است.

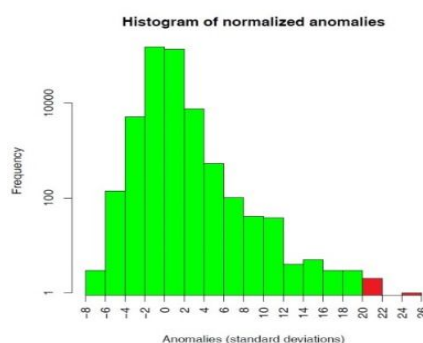
۶. **بیشینه سرعت باد:** شکل (۸) هیستوگرام آنومالی داده‌های سرعت باد بیشینه را به تفکیک در هر خوشه نشان می‌دهد. با توجه به توزیع آنومالی‌ها، آستانه‌ها برای تعیین نقاط پرت مشخص شده‌اند.

عنوان خطا تشخیص داده شدند. جدول (۲) نتایج را به تفکیک ایستگاه و تاریخ نشان می‌دهد. به طور کلی برای پارامتر بیشینه سرعت باد در مرحله اول ۴۱ داده پرت مشخص شد که از این تعداد ۲۷ داده خطا بودند. در بیشتر موارد خطاهای تشخیص داده شده به دلیل عدم همخوانی با داده‌های ساعتی یا داده‌های روزهای قبل و بعد بود. در موارد نادری همچون ایستگاه کهنوج به دلیل نامتعارف بودن مقدار، خطا شناخته

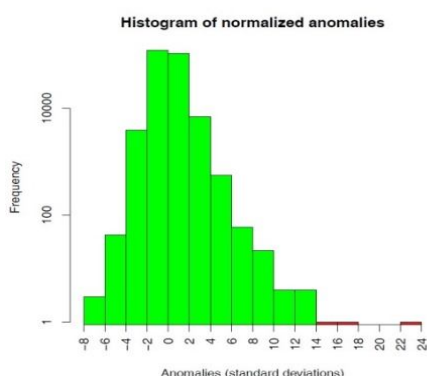
شد. شکل (۹) فراوانی خطاها و داده‌های پرت پارامتر سرعت باد را به تفکیک در هر خوشه نشان می‌دهد، همانطور که مشاهده می‌شود تعداد داده‌های پرت تشخیص داده شده و همینطور خطاها در خوشه یک بیشتر از سایر مناطق است و این امر با توجه به اینکه خوشه یک ایستگاه‌های بیشتری را شامل می‌شود قابل انتظار است.



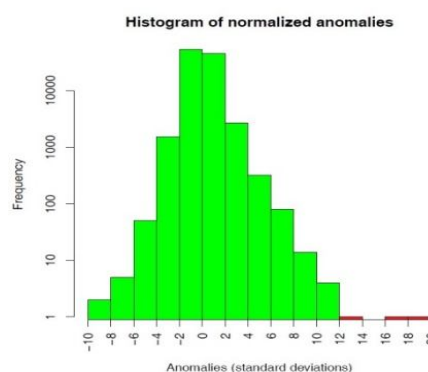
خوشه یک



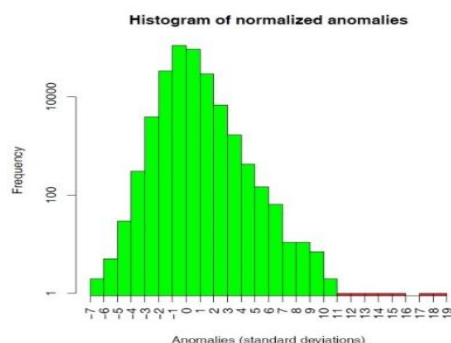
خوشه دو



خوشه سه

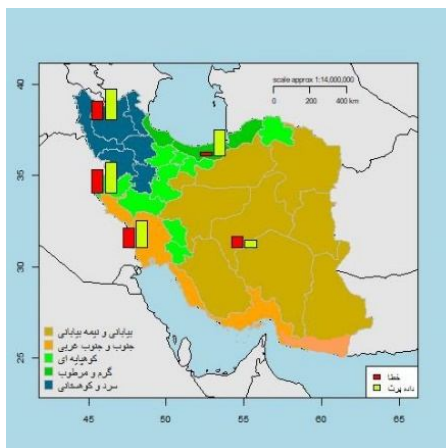


خوشه چهار

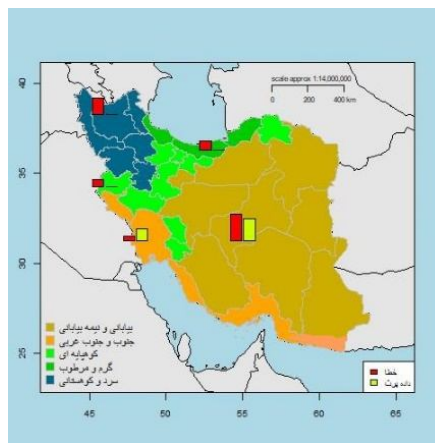


خوشه پنج

شکل ۸- هیستوگرام آنومالی داده‌های سرعت باد بیشینه به تفکیک در هر خوشه برای تشخیص حدود داده‌های پرت



سرعت باد



سرعت باد بیشینه

شکل ۹- فراوانی خطاها و داده‌های پرت به تفکیک در هر خوشه

جدول ۱- خطاهای تشخیص داده شده برای داده‌های سرعت باد و سرعت باد بیشینه

ردیف	سرعت باد			بیشینه سرعت باد		
	ایستگاه	تاریخ	مقدار مشاهداتی	ایستگاه	تاریخ	مقدار مشاهداتی
۱	زرقان	۱۲/۱۴/۱۹۹۱	۱۲,۴	گناباد	۲۰۱۷/۲۵/۱۰	۴۰
	جیرفت			طیس	۲۰۱۶/۹/۵	۴۱
	(میانده)	۰۶/۳۰/۲۰۰۰	۱۵,۶		۲۰۱۷/۱۲/۱۱	۵۰
	نطنز	۰۲/۰۹/۲۰۰۶	۲۰,۸	اردستان	۱/۷/۲۰۱۴	۲۳۰
				سیرجان	۲۰۱۴/۷/۶	۵۰
					۵۰ ۲۰۱۵/۱۲/۱	
					۴۰ ۲۰۱۶/۳۰/۱۲	
					۵۰ ۲۰۱۷/۲۹/۶	
					۳۸ ۲۰۱۸/۲۴/۴	
				بم	۲۰۱۸/۱/۷	۴۴
۲	بندرعباس	۱۲/۱۱/۱۹۹۵	۷	بندر ماهشهر	۲۰۱۵/۳۱/۲	۵۰
		۰۱/۱۱/۱۹۹۶	۶,۶	سراوان	۲۰۱۷/۲۴/۸	۸۰
		۰۲/۱۴/۱۹۹۶	۳,۶			
		۱۱/۱۹/۱۹۹۶	۸			
	میناب	۱۱/۱۸/۱۹۹۶	۱۴,۹			
۳	آبعلی	۲۹/۰۱/۲۰۰۸	۲۹,۰	تهران (شمیران)	۲۰۱۹/۲۲/۹	۵۰
		۰۱/۲۲/۲۰۰۹	۱۷,۶	کرمانشاه	۲۰۱۷/۲۷/۷	۵۰
		۱۲/۰۷/۲۰۱۵	۲۱,۸	شهرکرد	۲۰۱۹/۱۹/۲	۴۸
	سرپل ذهاب	۱۲/۲۰/۲۰۰۲	۱۸			
		۱۱/۱۰/۲۰۰۳	۲۰,۱			
	ساوه	۲۲/۲۰/۱۵/۱۹۹۳	۲۲,۲			

۴	رشت ۱۳/۴ ۱۲/۲۸/۱۹۹۵	آستارا ۳۵ ۲۰۰۱/۲۳/۲ ۳۵ ۲۰۰۱/۲۴/۲ بابلسر ۴۵ ۱۹۹۳/۱۳/۱۱ ۴۷ ۱۹۹۳/۱۴/۱۱
۵	سردشت ۲۶/۷ ۲/۱۰/۱۹۹۲	خوی ۵۰ ۲۰۱۹/۳/۸ تبریز ۳۸ ۲۰۱۴/۲/۴ ارومیه ۴۰ ۲۰۱۳/۲۵/۵ ۳۴ ۲۰۱۷/۷/۹ خرمدره ۵۰ ۲۰۱۷/۲۳/۱۰ مریوان ۴۰ ۲۰۱۴/۱۶/۷ ۵۰ ۲۰۱۵/۸/۷

جمع‌بندی

در این مقاله، به تشخیص داده‌های پرت زمانی- مکانی پارامترهای فشار ایستگاه، فشار تبدیل شده ایستگاه به سطح دریا، فشار بخار، سرعت باد، سرعت باد بیشینه در ۱۴۳ ایستگاه هواشناسی کشور برای دوره نرمال اقلیمی ۱۹۹۱-۲۰۲۰ پرداخته شد. هدف اصلی از این مطالعه تعیین داده‌های پرت ناشی از خطا بود و از این رو در دو مرحله کار انجام گرفت. ابتدا با استفاده از الگوریتم کلایماتول، داده‌های پرت مکانی مشخص شدند و در مرحله دوم با تحلیل‌های زمانی و هواشناختی به بررسی اینکه آیا این داده‌های پرت، خطا هستند پرداختیم. به طور کلی در مرحله اول برای پارامترهای فشار ایستگاه، فشار ایستگاه تبدیل شده به سطح دریا، فشار بخار، سرعت باد و سرعت باد بیشینه به ترتیب ۴۰، ۴۲، ۹۳، ۵۲ و ۴۱ داده پرت

مشخص شد و در مرحله دوم به ترتیب ۲۰، ۱۰، ۵۶، ۲۰ و ۲۷ تعداد از این داده‌ها خطا در نظر گرفته شدند. بررسی توزیع خطاها و داده‌های پرت در خوشه‌های مختلف نشان داد که ارتباطی بین میزان خطا و منطقه مورد بررسی وجود ندارد. در مواردی که در یک منطقه شاهد خطاهای بسیار بیشتری نسبت به سایر مناطق هستیم غالباً به علت اشکال در ثبت داده در یک ایستگاه خاص می‌باشد. خطاها به تفکیک ایستگاه و تاریخ گزارش شده اند. امید است که این نتایج بتوانند در مطالعات تغییر اقلیم مفید واقع شوند. محققان قبل از انجام هر گونه مطالعه‌ای بر روی داده‌ها می‌توانند خطاهای مشخص شده در داده‌ها را حذف یا اصلاح کرده و نتایج قابل اعتمادتری را ارائه کنند.

منابع

۱. دین‌پژوه، یعقوب، فاخری، احمد، مقدم، محمد، میرنیا، میرکمال، و جهانبخش اصل، سعید. (۱۳۸۲). پهنه‌بندی اقلیمی ایران با استفاده از تحلیل‌های چند متغیره برای استفاده در مطالعات کشاورزی. م. دانش کشاورزی، ۱۳(۱)، ۹۰-۷۱.
۲. علیجانی، بهلول. (۱۳۹۰). اقلیم‌شناسی سینوپتیک، چاپ چهارم، انتشارات سمت، تهران- ایران.
۳. مسعودیان، سید. ابوالفضل. (۱۳۸۲). نواحی اقلیمی ایران. م. جغرافیا و توسعه. ۱(۲)، ۱۸۴-۱۷۱.
4. Barnett, V. and Lewis, T. (1994). Outliers in statistical data. John Wiley & Sons, 3rd edition.
5. 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>).
6. Barnett, V. and Lewis, T. (1994). Outliers in statistical data. John Wiley & Sons, 3rd edition.
7. Coll, J., Domonkos, P., Guijarro, J., Curley, M., Elke Rustemeier, E., Aguilar,

16. Prasanthi, M.L., Krishna Chaitanya, A., Sambasiva Rao, Dr. N. (2016). A Survey On Outlier Detection Methods In Spatio-Temporal Datasets, IJAERS, Vol-3, Issue-11, Nov-, 168-172.
17. Rahimzadeh, F., Nassaji Zavareh, M., (2014). Effects of adjustment for non-climatic discontinuities in determination of temperature trends and variability over Iran. *Int. J. Climatol.* 34:2079–2096.
18. Ranjan, K., Tripathy, D.S., Prusty, B.R., and Jena, D. (2020). An improved sliding window prediction-based outlier detection and correction for volatile time-series, *Int J Numer Model.* 34:e2816. <https://doi.org/10.1002/jnm.2816>.
19. Skrynyk, O., Aguilar, E., Guijarro, J., Yannick, L., Randriamarolaza, A., and Bubin, S. (2020). Uncertainty evaluation of Climatol's adjustment algorithm applied to daily air temperature time series. *Int. J. Climatol.* 41 (S1), E2395-E2419.
20. Schaeffer R. et al. (2012). Energy sector vulnerability to climate change: A review. *Energy*, 38, 1-12.
21. Tavakoli, M., and Mohmoudian, H.A., (2011). GIS based considerations for development in different Iranian climatic regions. *J. American. Science.* 7(4), 182–187.
22. Cheng, T. and Li, Z. (2004). A Hybrid Approach to Detect Spatial-temporal Outliers. In *Proc. GeoInformaticas*, p 173-178.
23. Yuxiang, S., Kunqing, X., Xiujun, M., Xingxing, J., Wen, P., and Xiaoping, G., (2005). Detecting spatio-temporal outliers in climate dataset: a method study, 2005 IEEE International Geoscience and Remote Sensing Symposium. IGARSS'05.
24. Wu, E., Liu, W., and Chawla, S. (2008). Spatio-temporal Outlier Detection in Precipitation Data. *Knowledge Discovery from Sensor Data.* 115-133.
- E., Walsh, S., and Sweeney, J. (2020). Application of homogenization methods for Ireland's monthly precipitation records: Comparison of break detection results. *Int. J. Climatol.* 40 (14), 6169- 6188.
8. Lu, C. and Liang, L.R. 2004. Wavelet Fuzzy Classification for Detecting and Tracking Region Outliers in Meteorological Data. *GIS'04*, pages 258-265.
9. Daget, J. (1979). *Les modèles mathématiques en écologie.* Collection d'Écologie 8, 172, Masson, Paris.
10. Eleonora Aruffo, E., Piero Di Carlo, P.D. (2019). Homogenization of instrumental time series of air temperature in Central Italy (1930–2015), *Clim. Res.* Vol. 77: 193–204, 193-203.
11. Guijarro, J.A. (2018). Homogenization of climatic series with Climatol. Available at: http://www.Climatol.eu/homog_Climatol-en.pdf [Accessed 28th March 2020].
12. International Energy Agency (IEA) (2013). *Wind Power Technology Roadmap 2013 Edition.*
13. Ma, L., Gu, X. and Wang, B. 2017. Correction of Outliers in Temperature Time Series Based on Sliding Window Prediction in Meteorological Sensor Network, *Information* 2017, 8(2), 60, <https://doi.org/10.3390/info8020060>.
14. N.R. Adam, V.P. Janeja, and V. Atluri. Neighborhood based detection of anomalies in high dimensional spatio-temporal Sensor Datasets. *SAC'04*, pages 576-583, 2004.
15. Paulhus, J.L.H., Kohler, M.A. (1952). Interpolation of missing precipitation records. *Monthly Weather Review* 80(8), 129-133.

