

## بررسی و خوشه‌بندی خصوصیات زمانی - مکانی بارش کشور با استفاده از موجک حداکثر هم‌پوشانی و انتروپی چندمقیاسی

سپیده چوبه<sup>۱</sup>، هیراد عبقری<sup>۲\*</sup>، مهدی عرفانیان<sup>۳</sup>

تاریخ دریافت: ۱۴۰۲/۰۳/۰۹

تاریخ پذیرش: ۱۴۰۲/۰۳/۳۱

### چکیده

بارش از عناصر مهم اقلیمی و از عوامل تأثیرگذار در چرخه آب به شمار می‌رود. تغییرات خصوصیات زمانی-مکانی بارش در یک منطقه، نقش مؤثری در مدیریت منابع آب آن دارد. هدف از این تحقیق، ارزیابی ویژگی‌های بارش سالانه ۵۰ ایستگاه سینوپتیک کشور در بازه زمانی ۱۹۸۰-۲۰۲۰ با استفاده از روش‌های تبدیل موجک گسسته حداکثر هم‌پوشانی و خوشه‌بندی است. بدین منظور ابتدا سری زمانی بارش سالانه ایستگاه‌ها با استفاده از روش MODWT و موجک مادر DB4 به چندین زیر سری تجزیه شد، سپس انتروپی زیرسری‌های حاصل از MODWT محاسبه و به‌عنوان ورودی برای منطقه‌بندی بارش استفاده شد. نتایج تجزیه داده‌های بارش نشان داد که در سری زمانی سالانه، زیرسری‌های جزئی کوچک‌تر، فرکانس‌های بزرگ‌تر با تغییرات سریع‌تر را نشان می‌دهند؛ درحالی‌که ضرایب جزئی بزرگ‌تر نشان‌دهنده تغییرات کم می‌باشند. همچنین، A4 (روند تغییرات) کمترین تغییرات را نشان داد. براساس مقادیر معیارهای ارزیابی، تعداد خوشه‌های بهینه برابر با ۴ تعیین شد. مقادیر معیارهای ارزیابی خوشه‌بندی نشان داد که روش K-means با  $CHI=19/53$ ،  $Sci=0/26$  و  $DBi=1/08$  نسبت به روش SOM عملکرد بهتری داشته است. بنابراین ایستگاه‌های سینوپتیک کشور بر مبنای شاخص موجک گسسته حداکثر هم‌پوشانی - انتروپی به ۴ خوشه جدا شد که ایستگاه‌های هر خوشه از نظر عدم قطعیت هیدرولوژیکی و تغییرات، مشابه هم هستند. درنهایت ایستگاه‌های چابهار، شاهرود، آبادان و زنجان به‌عنوان مراکز خوشه انتخاب شدند.

**واژه‌های کلیدی:** انتروپی، ایران، بارش، تبدیل موجک، حداکثر هم‌پوشانی، خوشه‌بندی.

۱. دانشجوی دکتری آبخیزداری - آب، گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران. s.choobeh@urmia.ac.ir

۲. دانشیار، گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران، نویسنده مسئول. h.abghari@urmia.ac.ir

۳. دانشیار، گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران. m.erfanian@urmia.ac.ir

## مقدمه

هوا و اقلیم سیستم پیچیده‌ای بوده و پارامترهای مختلفی را شامل می‌شود که تعامل بین این پارامترها، نوع اقلیم هر منطقه را تعیین می‌کند. دما و بارش از عناصر مهم در شناسایی اقلیم هر منطقه است (امیدوار و خسروی، ۲۰۱۰). تغییرات مکانی و زمانی بارش، پدیده‌های حدی مانند خشکسالی و سیل، افزایش جمعیت، رشد صنعت و اقتصاد منجر به افزایش نیازهای آبی شده و تنش آبی در کشور را به دنبال دارد. با افزایش تنش آبی، مدیریت پایدار آب با مشکلات زیادی مواجه خواهد بود. از طرفی، قرارگیری بخش‌های وسیعی از ایران در کمربند خشک و نیمه‌خشک جهان و همچنین نقش تعیین‌کننده نزولات جوی در تأمین آب کشور، باعث شده که آگاهی بیشتر نسبت به تغییرات بارش کشور از اهمیت زیادی برخوردار باشد. تعیین مناطق همگن بارش و انتخاب ایستگاه‌های شاخص بارش هر منطقه، نقش اساسی در مدیریت منابع آب، تصمیم‌گیری‌ها و سیاست‌گذاری‌های کلان کشور ایفا می‌کند.

در ارزیابی خوشه‌ای تلاش بر این است که ایستگاه‌هایی که متعلق به یک خوشه خاص هستند، از نظر متغیرهای مورد بررسی، بیشترین شباهت را باهم داشته باشند. بدین ترتیب ایستگاه‌های همگن در یک خوشه قرار گرفته و به جای مدل کردن همه ایستگاه‌های موجود در هر خوشه، مدل‌سازی تنها برای مراکز هر خوشه انجام می‌شود. از طرفی، سری‌های زمانی مورد استفاده در مدل‌سازی‌های هیدرولوژیکی، ناپستا و دارای روند و یا نوسانات فصلی هستند. برای لحاظ کردن اثر این عوامل در مدل‌سازی‌ها، پیش‌پردازش زمانی و تعیین عوامل ناپستایی داده‌ها ضروری است (مسعودیان و همکاران، ۲۰۱۱؛ تقوی و همکاران، ۲۰۱۲؛ خسروی و همکاران، ۲۰۱۴؛ رضیعی، ۲۰۱۶؛ مفاخری و همکاران، ۲۰۱۷؛ علیزاده قراملکی، ۲۰۱۸؛ عیسی‌زاده و دین‌پژوه، ۲۰۱۸؛ محمدیاریان و همکاران، ۲۰۱۹؛ رورده و همکاران، ۲۰۱۹؛ روشنگر و همکاران، ۲۰۲۰). همچنین محققان بسیاری در جهان، بارش مناطق مختلف را ناحیه‌بندی نموده‌اند (هسو و لی، ۲۰۱۰؛ سانتوس و همکاران،

۲۰۱۷؛ رضیعی، ۲۰۱۸؛ فاضل و همکاران، ۲۰۱۸؛ گووند و سیواکومار، ۲۰۲۰؛ گانتو و همکاران، ۲۰۲۰؛ البای و همکاران، ۲۰۲۱؛ زروعلی و همکاران، ۲۰۲۲؛ کان چالا و همکاران، ۲۰۲۲؛ کالپس و همکاران، ۲۰۲۲؛ یازاوا و شوجی، ۲۰۲۳؛ اددو و همکاران، ۲۰۲۳).

اولین مطالعات برای تعیین ناحیه‌های بارش ایران توسط دومروس و همکاران (۱۹۹۸) انجام شد که با استفاده از روش‌های تحلیل مؤلفه اصلی و تحلیل خوشه‌ای ۷۱۳ ایستگاه سینوپتیک کشور، ۵ رژیم بارش ماهانه شناسایی نمودند. رورده و همکاران (۲۰۱۹) در مطالعه‌ای، بارش ایران را با استفاده از روش هیبریدی نگاهشت تجزیه و خوشه‌بندی فازی منطقه‌بندی نمودند. بدین منظور داده‌های ۶۳ ایستگاه سینوپتیک کشور (بارش، رطوبت نسبی و دمای نقطه شبنم) در بازه زمانی ۱۹۸۵ تا ۲۰۱۳ را بررسی نمودند و براساس نتایج، بارش کشور را در ۶ دسته ساحلی خزر، سواحل خلیج فارس و دریای عمان، نواحی بسیار خشک مرکزی، ناحیه نیمه‌خشک، ناحیه کوهستانی و نیمه‌کوهستانی خوشه‌بندی نمودند که این خوشه‌ها مطابق با عرض جغرافیایی، توپوگرافی ایران و دوری و نزدیکی از دریا و منابع رطوبتی بوده است. همچنین روشنگر و همکاران (۲۰۲۰)، در تحقیق خود به بررسی خصوصیات زمانی- مکانی بارش پرداختند. در روش پیشنهادی این مطالعه، از تبدیل موجک گسسته ماکزیم همپوشانی به‌همراه خوشه‌بندی K-means برای تجزیه سری زمانی بارش استفاده نمودند. براساس نتایج، تعداد خوشه‌های بهینه برای ایستگاه‌ها برابر با پنج تعیین شد. نتایج نشان داد که با افزایش انرژی خوشه‌ها، مقدار بارش در ایستگاه‌های آن خوشه کاهش می‌یابد و بالعکس. همچنین مقدار ضریب سیلوئت خوشه‌بندی در روش پیشنهادی ۰/۸ محاسبه شد که بیانگر

3Govender & Sivakumar

4Guntu

5Ilbay

6Zerouali

7Canchala

8Claps

9Yazawa & Shoji

10Addou

11Domroes

12Principal component analysis

13Clustering analysis

14Singular value decomposition

15Fuzzy C-Means

1Hsu & Li

2Santos

منطق فازی و رویکرد خوشه‌بندی در شهر بمبئی پرداختند. بدین منظور از رویکردهای خوشه‌بندی C-means و K-means فازی (FCM) برای بررسی الگوی بارندگی استفاده کردند. طبق نتایج، هر دو روش K-means و FCM دو منطقه بارندگی همگن را در بمبئی شناسایی نمودند که دو الگوی بارندگی متمایز را نشان می‌دهند.

آد دو و همکاران (۲۰۲۳) به مطالعه تغییرپذیری بارندگی و منطقه‌بندی داده‌های بارش حوزه آبخیز مولویا در مراکش براساس تجزیه و تحلیل موجک پرداختند. نتایج نشان داد که توان طیفی در پنج محدوده فرکانس متغیر بوده و بیشترین پراکندگی فرکانس منطقه در محدوده ۴ تا ۸ ساله است. در نهایت بر این اساس، حوضه مورد مطالعه خود را به دو خوشه همگن تقسیم نمودند که این خوشه‌ها فقط در تغییرپذیری و رژیم بارندگی متفاوت بودند.

با بررسی مطالعات گذشته ملاحظه می‌شود که پارامترهایی نظیر موقعیت جغرافیایی ایستگاه‌ها و مقدار بارش ماهانه و سالانه، مبنای خوشه‌بندی بوده‌اند. اما مطالعات محدودی در زمینه خوشه‌بندی بارش براساس روند تغییرات انجام شده است؛ لذا در این مطالعه مقادیر انترپوی، مبنای خوشه‌بندی بارش سالانه کشور قرار گرفته است.

### مواد و روش‌ها

#### منطقه مورد مطالعه

کشور ایران با مساحتی حدود ۱۶۴۸۱۹۵ کیلومتر مربع بین  $19^{\circ} 02' 44''$  تا  $29^{\circ} 50' 20''$  طول شرقی و  $25^{\circ} 04' 14''$  تا  $39^{\circ} 47' 45''$  عرض شمالی واقع شده است. آب‌وهوای غالب ایران معمولاً از نوع خشک و نیمه‌خشک بوده و مناطق خیلی خشک با حدود ۳۵/۵ درصد، بیشترین مساحت کشور را به خود اختصاص داده است. در این رساله متوسط بارش سالانه در دوره آماری ۱۹۸۰-۲۰۲۰ تقریباً ۲۸۱ میلی‌متر برآورد شده است؛ در صورتی که متوسط بارش سالانه کشور در منابع مختلف ۲۴۳ میلی‌متر ارائه شده است (مرکز پژوهش‌های مجلس شورای اسلامی، ۲۰۱۷). براساس داده‌های این پژوهش، مقدار بارش از صفر در جنوب، شرق و کویرهای مرکزی تا ۱۷۴۴ میلی‌متر در نواحی ساحلی دریای خزر (ایستگاه سینوپتیک بندر انزلی) متغیر است. نقشه متوسط بارندگی سالانه کشور در دوره آماری مذکور و موقعیت ایستگاه‌های

عملکرد قابل قبول روش پیشنهادی در خوشه‌بندی ایستگاه‌های مورد مطالعه است. در مطالعه‌های دیگر، نواحی بارشی کشور بر مبنای تداوم بارش شناسایی شد. نتایج فرایند خوشه‌بندی روی تداوم‌های بارش ایران در دوره زمانی ۱۹۸۷-۲۰۱۶ نشان داد که هفت پهنه بارشی تقریباً همگن در ایران وجود دارد. موقعیت جغرافیایی نواحی تداوم بارش ایران، هم‌بستگی مقدار بارش به ناهمواری‌ها، مسیر سامانه‌های بارش‌زا، نزدیکی به منابع رطوبت و اثر دریا را به وضوح نشان می‌دهد. همچنین ازلحاظ موقعیت قرار گرفتن پهنه‌ها می‌توان بیان نمود که قرار گرفتن چهار پهنه در نیمه غربی ایران با وجود وسعت کم این منطقه در مقابل نیمه شرقی، دلیل بر ناهمگنی بارش در غرب کشور است (تقفی و همکاران، ۲۰۲۴). پرویز و بیرامی (۲۰۲۳) در مطالعه‌ای، به ارزیابی کارایی تبدیل موجک گسسته حداکثر همپوشانی در روندیابی متغیرهای هواشناسی در ایستگاه‌های سینوپتیک منتخب شامل رامسر، کرمانشاه، همدان، سمنان، مشهد، شیراز، اردبیل، ارومیه، زنجان، یزد، تبریز، مراغه و میانه پرداختند. در روش MODWT، روند برخی از زیرسری‌ها با نتایج روند آزمون من-کنندال سری زمانی اصلی داده‌ها هم‌خوانی داشت؛ میزان این تطابق متأثر از مقدار عددی آماره من-کنندال سری زمانی اصلی و علامت آن بوده است. سطح معنی‌داری آزمون من-ویتنی در مقایسه سری زمانی اصلی داده‌های مشاهداتی سالانه با زیرسری‌های حاصل از روش MODWT، بیشتر از ۰/۰۵ بود که بیانگر عملکرد قابل قبول این روش است.

البای و همکاران (۲۰۲۱) در تحقیقی با استفاده از داده‌های بارش ماهانه، ۲۱۵ ایستگاه باران‌سنجی در طی دوره زمانی ۱۹۶۸-۲۰۱۴، به منطقه‌بندی بارش کشور اکوادور پرداختند. بدین منظور از متغیرهای بارش ماهانه و سالانه و موقعیت جغرافیایی ایستگاه‌ها (طول جغرافیایی، عرض جغرافیایی و ارتفاع) برای خوشه‌بندی ایستگاه‌ها استفاده نمودند. در نهایت ۲۲ منطقه همگن شناسایی نمودند. همچنین نتایج نشان دادند که تأثیر دمای سطح دریای اقیانوس آرام، در مناطق مختلف کشور متفاوت است؛ زیرا اکوادور تحت تأثیر حالت‌های تغییرپذیری بارش در کلمبیا و پرو قرار دارد.

نی کو و همکاران (۲۰۲۲) در پژوهشی به بررسی روند بارش و همچنین شناسایی مناطق همگن بارش با استفاده از

سینوپتیک مورد مطالعه در شکل (۱) ارائه شده است.

در این تحقیق به منظور بررسی تغییرات زمانی- مکانی بارش و همچنین تعیین نواحی همگن، از داده‌های سالانه بارش ۵۰ ایستگاه در طی ۴۱ سال (۲۰۲۰-۱۹۸۰) استفاده شده است (شکل ۱). بدین جهت در ابتدا، سری زمانی بارش ایستگاه‌ها بر اساس سطح تجزیه و موجک مادر مناسب به زیرسری‌هایی تجزیه شد و سپس مقادیر انتروپی هر زیرسری محاسبه گردید. در ادامه این مقادیر به عنوان ورودی مدل‌های خوشه‌بندی استفاده شده و در نهایت مناطق همگن بارش از نظر مقادیر انتروپی و روند تغییرات، تعیین شدند.

### تبدیل موجک حداکثر همپوشانی (MODWT)

موجک، تابع مشخص فرضی با میانگین صفر است و بسط بر حسب انتقال و مقیاس‌های آن صورت می‌پذیرد، موجک‌ها در فضا به صورت موضعی بررسی شده و ارتباط نزدیک‌تری بین بعضی توابع و ضرایب آن‌ها وجود دارد. بنابراین پایداری عددی بیشتری را در بازسازی و محاسبات فراهم می‌کنند. تبدیل موجک قادر به پوشش نقاط ضعف تبدیل فوریه است و در ارتباط با سیگنال‌های غیرایستا مانند بارش عملکرد مناسبی دارند؛ زیرا تبدیل فوریه به دنبال یک جزء فرکانسی خاص در سیگنال است که این اطلاعات مستقل از زمان وقوع این جزء است، ولی در تبدیل موجک تحلیل تغییرات سیگنال در بستر زمان می‌باشد. آنالیز موجک شکستن یک سیگنال بارش به نسخه‌های انتقال یافته و مقیاس‌بندی شده از موجک مادر است. اصطلاح موجک، اشاره به این واقعیت دارد که تابع پنجره طول متناهی داشته و نوسانی است؛ در حالی که عرض موجک به عنوان تبدیل تغییر کرده و برای هر جزء طیفی منفرد محاسبه می‌گردد. در حقیقت، تجزیه و تحلیل موجک ارائه یک مقیاس زمان سیگنال بارش را فراهم می‌کند. بدین ترتیب، موجک می‌تواند آنالیزهای موضعی را انجام داده و اطلاعاتی از جمله روند، تغییرات ناگهانی، نقاط شکست و ناپیوستگی را آشکار کند (علیزاده قراملکی، ۲۰۱۸؛ عبقری، ۲۰۰۸).

در تبدیل موجک، سیگنال بارش بر اساس مشخصات فرکانسی به بخش‌های تقریب و جزئیات تقسیم می‌شود. تقریب

نتیجه حذف فرکانس‌های مزاحم است که مانع تخمین درست روند سری‌های زمانی بارش می‌شوند. در تبدیل موجک با تجزیه سری زمانی بارش به چندین زیرسری در هر مرحله تجزیه، می‌توان نویزهایی را که موجب بروز خطا در تخمین روند اصلی سیگنال‌های بارش می‌شوند، حذف نمود. بنابراین آنالیز اصلی بر روی تقریب نهایی سری زمانی بارش انجام می‌گردد. تجزیه سیگنال بارش تا زمانی ادامه می‌یابد که بخش تقریب شکل غالب سیگنال بارش را حفظ کند و جزئیات آن به کمترین مقدار برسد. مزیت تبدیل موجک به تبدیل فوریه این است که موجک‌های پایه، دارای دو خاصیت اساسی یعنی مقیاس و موقعیت هستند. در این تبدیل، مقیاس‌های پایین بیانگر جزئیات بیشتری هستند؛ در صورتی که مقیاس‌های بالا دید کلی از سیگنال بارش را ارائه می‌دهند و فاقد جزئیات هستند. برعکس فرکانس‌های بالا بیانگر اطلاعات جزئی سیگنال بارش هستند که اغلب به یک زمان کوتاه مربوط‌اند؛ اما فرکانس‌های پایین اطلاعات کلی سیگنال بارش را نشان می‌دهند (آرمش، ۲۰۱۷).

تبدیل موجک حداکثر همپوشانی (MODWT)، اصلاح شده تبدیل موجک گسسته (DWT) است. با توجه به اینکه در هر مرحله از اعمال الگوریتم و جابه‌جایی در محور زمان به مقدار زوج و تقلیل نمونه، بخشی از اطلاعات در حوزه زمان از دست می‌رود، تبدیل موجک گسسته نسبت به انتخاب نقطه شروع سری زمانی حساس است. بنابراین موجک گسسته به MODWT اصلاح شده است (پائول<sup>۳</sup> و همکاران، ۲۰۲۰). در این تبدیل، حرکت در محور زمان، واحد به واحد بوده و فیلتر به صورت گردشی انجام می‌شود و برخلاف روش DWT که پس از هر مرحله از انجام الگوریتم، سری زمانی اصلی به دو سری با طول نصف سری مرحله قبل تقسیم می‌شود، در MODWT در هر مرحله اندازه سری جزئیات و سری تقریب، برابر با سری اصلی بوده و تقلیل نمونه‌ها انجام نشده و اطلاعاتی از بین نمی‌رود و داده‌های اصلی حفظ می‌شود (آداموفسکی<sup>۴</sup> و همکاران، ۲۰۰۹). از تفاوت‌های تبدیل موجک حداکثر همپوشانی و تبدیل موجک گسسته می‌توان به موارد

2 Discrete wavelet transforms

3 Paul

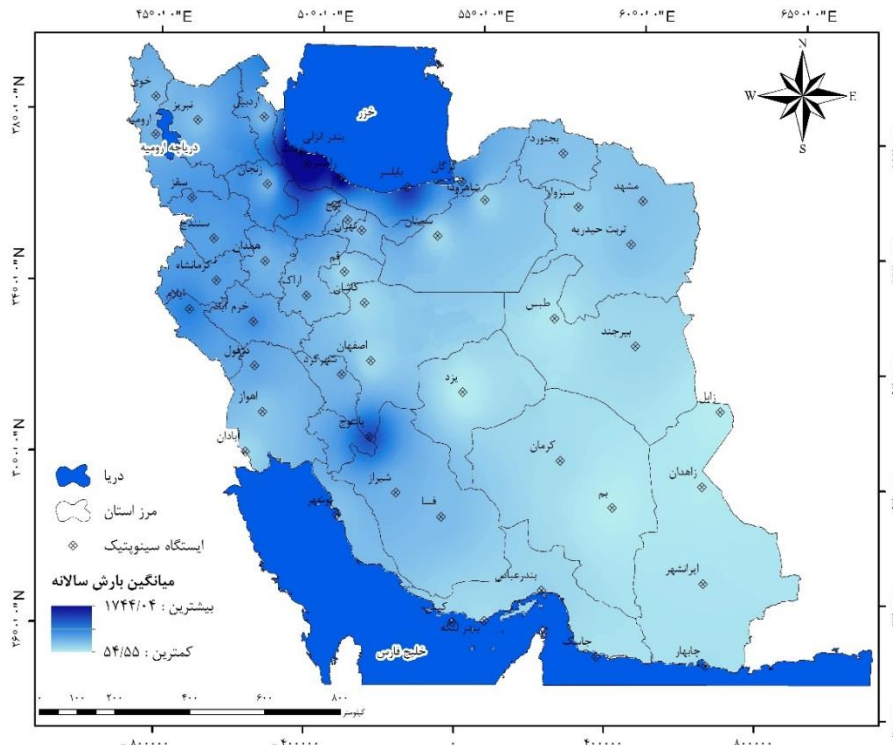
4 Adamowski

1 Maximal overlap discrete wavelet transforms

زیر اشاره کرد:

توابع پایه‌ای غیرصفر در طول یک بازه معین دارند، پشتیبانی کاملی از سیگنال‌های زمانی بارش را ارائه می‌دهند. از تجزیه یک سیگنال توسط تبدیل MODWT، ضرایب موجک و مقیاس گذاری به دست می‌آید. ضرایب مقیاس گذاری ضرایب تبدیل موجک با وضوح بالا را نشان می‌دهند که بیانگر روندهای صاف و هموار را در سری زمانی هستند؛ در مقابل، ضرایب موجک (D) اطلاعات جزئی از روندها را در سری زمانی هیدرولوژیکی بیان می‌کنند. هرکدام از جزءهای D، دوره‌های خاصی از سری زمانی را ارائه می‌دهند. در داده‌های ماهانه، D1، D2، D3 و D4 به ترتیب نشان‌دهنده دوره‌های ۲ ساله، ۴ ساله، ۸ ساله، ۱۶ ساله است. زیرسری تقریب A4 نیز روند تغییرات را نشان می‌دهد.

۱. با استفاده از روش MODWT، سری‌های زمانی با طول‌های متفاوت را می‌توان ارزیابی نمود؛ درحالی‌که روش DWT، محدود به سری‌های زمانی با طول‌های  $2^z$  (z: سطح تجزیه) است.  
 ۲. رخدادها و نوسانات موجود در سری زمانی با تحلیل چند مقیاس تبدیل موجک حداکثر همپوشانی هم‌راستایی بیشتری دارند.  
 ۳. تخمینگر واریانس بر مبنای روش تبدیل موجک حداکثر همپوشانی از تخمینگر روش تبدیل موجک گسسته کارتر است (روشنگر و همکاران، ۲۰۲۳).  
 در این تحقیق از موجک مادر DB4 برای تجزیه سیگنال بارش استفاده شده است. با توجه به اینکه موجک مادر DB



شکل ۱: پراکنش ایستگاه‌های سینوپتیک و متوسط بارش سالانه در ایران در طول دوره آماری ۱۹۸۰-۲۰۲۰

Figure 1: The distribution of synoptic stations and the average annual precipitation in Iran during the statistical period of 1980-2020

پخش شود، توزیع احتمالی برابر با  $\frac{1}{41}$  و انتروپی نیز کمترین مقدار خواهد بود (AE=0). در صورتی‌که تمام بارش تجمعی یک دوره در یک سال متمرکز شود، بیشترین بی‌نظمی و انتروپی بیشینه ( $\log_2 41$ ) اتفاق می‌افتد. بنابراین می‌توان نتیجه گرفت که

### انتروپی تفکیکی<sup>۱</sup> (AE)

انتروپی تفکیکی (AE)، مقدار تغییرات زمانی بارش سالانه در یک دوره زمانی را با استفاده از رابطه (۱) اندازه‌گیری می‌کند. اگر بارش سالانه تجمعی به صورت مساوی در تمام سال‌ها

1 Apportionment Entropy

گام دوم: هر نمونه داده به خوشه‌ای که مرکز آن خوشه کمترین فاصله را تا آن داده دارد، نسبت داده می‌شود.

گام سوم: پس از تعلق تمام داده‌ها به خوشه مربوط، مرکز جدیدی برای هر خوشه محاسبه می‌شود (میانگین نقاط متعلق به هر خوشه).

گام چهارم: مراحل ۲ و ۳ تا زمانی تکرار می‌شوند که تغییری در مراکز خوشه‌ها ایجاد نشود. تابع خطای مربعات نیز به صورت رابطه (۲) محاسبه می‌شود؛ به طوری که در این رابطه  $\|X_i - C_l\|^2$  فاصله داده  $X_{ij}$  از مرکز خوشه و  $n$  تعداد نقاط هستند (هارتیگان و وانگ، ۱۹۷۹).

$$J = \sum_{i=1}^k \sum_{j=1}^n \|X_{ij}^j - C_l\|^2 \quad (2)$$

با توجه به اینکه در روش K-means، محاسبه فاصله بین نقاط براساس تابع فاصله اقلیدسی انجام می‌گردد، از این الگوریتم به صورت استاندارد، فقط برای مقادیر عددی می‌توان استفاده نمود (لی و بیسواز، ۲۰۰۲).

### نقشه‌های خود سازمان‌دهنده (SOM)

این مدل از شبکه‌های عصبی نخستین بار در سال ۱۹۸۱ و توسط کوهنن با الگوبرداری از عصب‌های شبکه چشم ارائه شد. شبکه‌های عصبی خودسازمان‌دهنده، شبکه‌های عصبی بدون نظارتی هستند که قابلیت ارائه خروجی شبکه را در قالب نقشه‌های گرافیکی گویا و قابل فهم دارند. از این رو سرعت درک و تفسیر نتایج آن آسان‌تر خواهد بود. نقشه خود سازمان‌دهنده (SOM) توانایی تبدیل روابط آماری پیچیده و غیرخطی را به روابط ساده هندسی قابل نمایش دارد. در محاسبه مدل توسط یک نقشه خودسازمان‌دهنده، در واقع توصیف مطلوبی از دامنه مشاهدات ارائه می‌شود (لی و همکاران، ۲۰۲۰). در این شبکه، مدل‌های مشابه در محل‌های نزدیک‌تری نسبت به مدل‌های بی‌شبهت به یکدیگر قرار می‌گیرند. بنابراین، SOM را می‌توان یک گراف شباهت و یا یک نمودار خوشه‌بندی نامید (کوهنن، ۲۰۱۳؛ چن و همکاران، ۲۰۱۹).

مقادیر AE بین صفر تا  $\log_2 41$  است (شانون، ۱۹۴۸؛ میشر، ۲۰۰۹).

$$AE = - \sum_{i=1}^{41} \left(\frac{r_i}{R}\right) \log_2 \left(\frac{r_i}{R}\right) = - \sum_{i=1}^{41} (P_i) \log_2 P_i^2 \quad (1)$$

در رابطه فوق AE بیانگر انتروپی تفکیکی، R بارش سالانه تجمعی در طول دوره آماری مورد مطالعه،  $r_i$  بارش سالانه سال مورد نظر و  $P_i$  تابع توزیع احتمالی است.

### خوشه‌بندی K-means

اصطلاح K-means نخستین بار توسط جیمز مک‌کویین در سال ۱۹۶۷ استفاده شد. این روش یک روش خوشه‌بندی داده‌ها در داده‌کاو و یافتن مرکز خوشه‌ها در مجموعه‌ای از داده‌هاست. در این روش، داده‌های یک خوشه ویژگی‌های مشترکی دارند. این الگوریتم از ساده‌ترین روش‌های یادگیری خطی و بدون نظارت بوده که برای خوشه‌بندی داده‌های زمانی که در آن تعداد خوشه‌ها از قبل مشخص است، استفاده می‌شود (مک کوئین، ۱۹۶۷؛ سیناگا و یانگ، ۲۰۲۰). تعداد خوشه‌ها بر مبنای ویژگی‌های داده‌هایی که در کالیبراسیون مدل استفاده می‌شوند، تعیین می‌گردد. استفاده از خوشه‌های زیاد در یک مدل پیچیده باعث ایجاد تعداد پارامترهای زیادی خواهد شد و این مستلزم تعداد زیادی از داده‌ها برای به دست آوردن این پارامترها با هر سطح اعتمادی است. بنابراین در این روش ضروری است که تعداد خوشه‌ها با تعداد داده‌ها متناسب باشد. به طور کلی هدف روش خوشه‌بندی K-means، به حداقل رساندن تابع هدف است. مراحل پیاده‌سازی این الگوریتم شامل چهار گام به شرح زیر است (مک کوئین، ۱۹۶۷؛ آگاروال و همکاران، ۲۰۱۶):

گام اول: در ابتدا K نقطه به‌عنوان نماینده خوشه‌ها انتخاب می‌شوند.

7Hartigan & Wong  
8Li & Biswas  
9Self-organizing map  
10Kohonen  
1Chen

1Shannon  
2Mishra  
3Probability Distribution Functions  
4MacQueen  
5Sinaga & Yang  
6Agarwal

## ارزیابی روش خوشه‌بندی

استفاده از رابطه (۴) محاسبه می‌شود:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j=1, \dots, k, i \neq j} \left\{ \frac{\text{diam}(C_i) + \text{diam}(C_j)}{\|C_i - C_j\|} \right\} \quad (4)$$

$$\text{diam}(C_i) = \left( \frac{1}{n_i} \sum_{x \in C_i} \|x - z_i\|^2 \right)^{\frac{1}{2}} \quad (5)$$

در این رابطه،  $k$  تعداد خوشه و  $C$  مراکز خوشه‌ها هستند و  $\text{diam}(C_i)$  حالت قطر خوشه بوده که با استفاده از رابطه (۵) محاسبه می‌شود. با توجه به اینکه هدف، تعیین خوشه با کمترین فاصله درون خوشه‌ای است، کمترین مقدار ممکن شاخص دیویس بولدین مدنظر است.

کالینسکی - هاراباز<sup>۴</sup> (CHI) از ضرایب رایج برای ارزیابی عملکرد روش‌های خوشه‌بندی است و توسط کالینسکی و هاراباز در سال ۱۹۷۴ پیشنهاد شد؛ این شاخص مشابه شاخص سیلووه بوده که شامل دو جزء است: یک جزء نشان‌دهنده درجه نزدیکی بین نمونه‌های اختصاص داده شده به همان خوشه و دیگری نشان‌دهنده درجه پراکندگی نمونه‌های اختصاص داده شده به خوشه‌های مختلف است (لوسیر و همکاران، ۲۰۰۹؛ وانگ و خو، ۲۰۱۹). بنابراین شاخص کالینسکی - هاراباز یک شاخص ارزیابی براساس میزان پراکندگی بین خوشه‌ای و درون خوشه‌ای است و با استفاده از رابطه (۶) محاسبه می‌شود (وانگ و خو، ۲۰۱۹).

$$CH(K) = \frac{B(K)(N - K)}{W(K)(N - K)} \quad (6)$$

در رابطه فوق،  $K$  تعداد مربوط به خوشه‌ها،  $N$  تعداد نمونه‌ها،  $B(K)$  کوواریانس بین خوشه‌ای و  $W(K)$  نیز واگرایی درون خوشه‌ای است و به ترتیب با استفاده از رابطه‌های (۷) و (۸) محاسبه می‌شوند.

$$B(K) = \left( \sum_{k=1}^K a_k \|\bar{X}_k - \bar{X}\|^2 \right) \quad (7)$$

$$W(K) = \left( \sum_{k=1}^K \sum_{c(j)=k} \|\bar{X}_k - \bar{X}\|^2 \right) \quad (8)$$

به منظور تحلیل نتایج خوشه‌بندی، از معیارهای متفاوتی استفاده می‌شود. اگر هیچ اطلاعات اولیه‌ای از ساختار و طبقه‌بندی داده‌ها وجود نداشته باشد، معیار ارزیابی، اندازه‌هایی است که میزان شباهت درون خوشه‌ها و یا عدم شباهت یا فاصله بین خوشه‌ها را تعیین می‌کنند. لذا برای تعیین روش خوشه‌بندی مناسب، می‌توان از میزان شباهت درون خوشه‌ها و شباهت بین خوشه‌ها استفاده کرد. بهترین روش، روشی است که میزان شباهت بین خوشه‌ای کم و شباهت درون خوشه‌ای زیادی داشته باشد. این معیارها به عنوان شاخص‌های ارزیابی درونی بیان می‌شوند.

ضریب سیلووه<sup>۱</sup> (SCI) به منظور ارزیابی عملکرد روش خوشه‌بندی و تعیین تعداد بهینه خوشه‌ها استفاده شد. این ضریب بیشترین کاربرد را در علم هیدرولوژی و بهترین کارایی جهت بررسی عملکرد روش‌های خوشه‌بندی SOM و K-means دارد (هسو و لی، ۲۰۱۰؛ نورانی و همکاران، ۲۰۱۸). مقدار این ضریب بین -۱ تا +۱ متغیر بوده و هرچه به ۱ نزدیک‌تر باشد، نشانگر تفرق بهتر بین خوشه‌هاست. ضریب سیلووه با استفاده از رابطه (۳) محاسبه می‌شود.

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

برای هر داده  $(i)$ ،  $a(i)$  میانگین عدم تشابه داده با دیگر داده‌ها در درون خوشه مورد نظر و  $b(i)$  کمترین عدم تشابه داده با دیگر داده‌ها در درون خوشه مورد نظر است. شاخص سیلووه نشانگر تشابه اعضا به خوشه متعلق به آن در مقایسه با دیگر خوشه‌هاست.

شاخص دیویس بولدین (DBi) از انواع شاخص‌های ارزیابی درون خوشه‌ای است که با توجه به عملکرد مناسب آن در تعیین تعداد خوشه‌های بهینه، در علم هیدرولوژی مورد استفاده قرار می‌گیرد (دیویس و بولدین، ۱۹۷۹؛ کاستوری، ۲۰۰۳).

شاخص دیویس و بولدین تابعی از نسبت مجموع پراکندگی درون خوشه‌ای به جدایی درون خوشه‌ای بوده و با

4Calinski-Harabasz  
5Lucieer  
6Wang & xu

1Silhouette Coefficient index  
2Davies & Bouldin  
3Kasturi

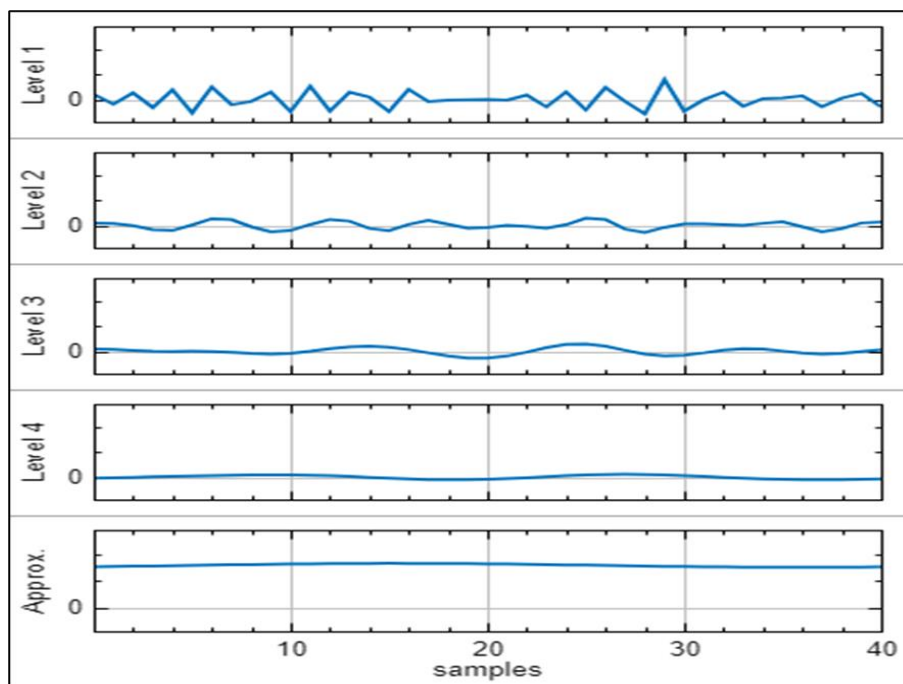
## نتایج

### تجزیه سری‌های زمانی بارش

مقدار بارش سالانه ۵۰ ایستگاه سینوپتیک کشور در طی سال ۴۱ به‌عنوان ورودی مدل MODWT انتخاب شد. تجزیه بارش سالانه ایستگاه شهرکرد در مقیاس سالانه به‌عنوان نمونه در شکل (۲) ارائه شده است. ضرایب جزئی  $D_2$ ،  $D_3$  و  $D_4$  به ترتیب نوسانات ۲، ۴، ۸ و ۱۶ سال را نشان می‌دهد. ضریب تقریب هم روند تغییرات سالانه سری زمانی بارش ایستگاه سینوپتیک شهرکرد را نشان می‌دهد. در سری زمانی سالانه، زیرسری‌های جزئی کوچک‌تر، فرکانس‌های بزرگ‌تر را نشان می‌دهند که تغییرات در آن‌ها سریع‌تر است. اما ضرایب جزئی بزرگ‌تر، فرکانس‌های کم و تغییرات ملایم‌تر را نشان می‌دهند؛ نیز کمترین تغییرات و فرکانس کم را به نمایش می‌گذارد.

### محاسبه انتروپی تفکیکی

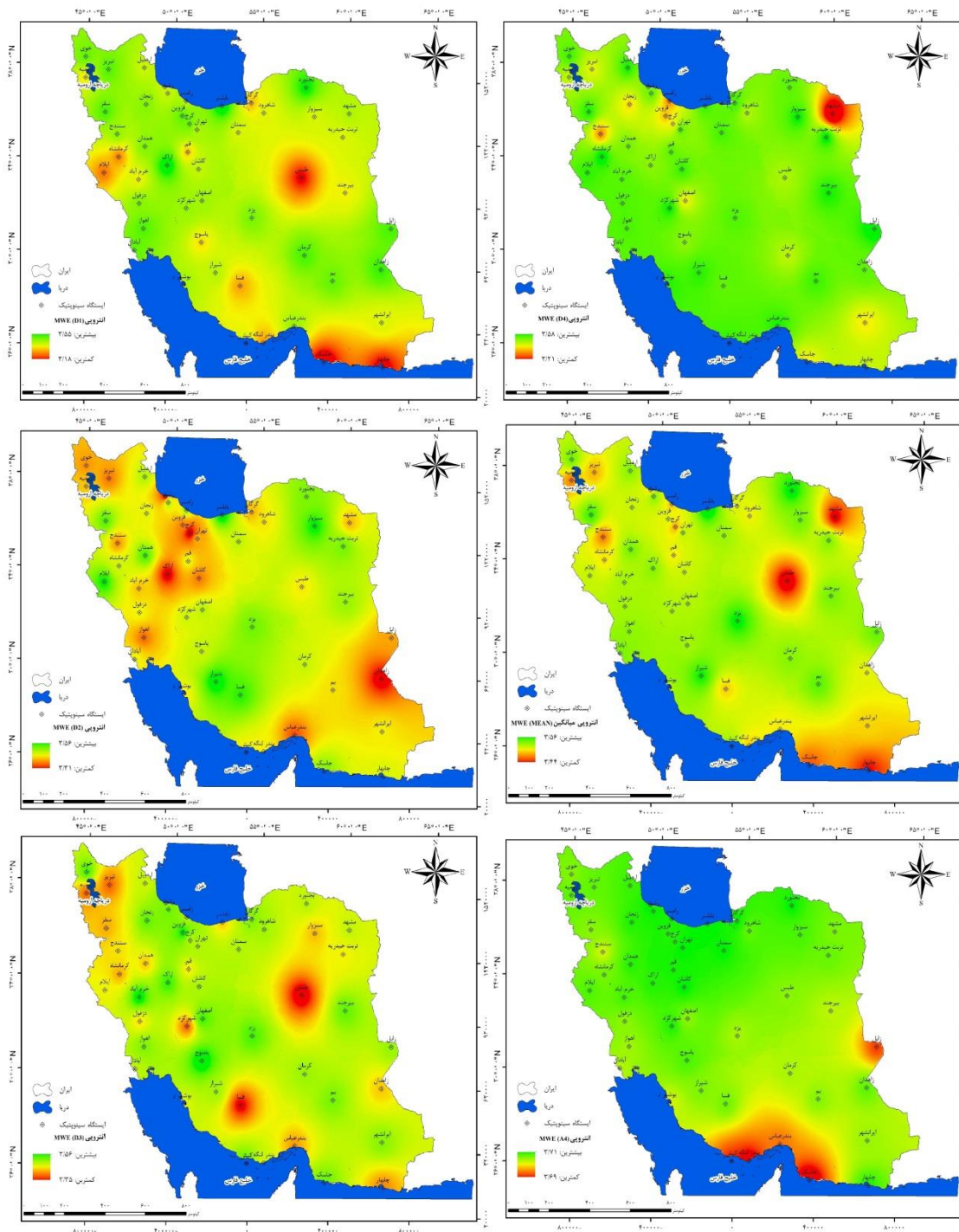
در این مرحله، انتروپی زیرسری‌های بارش حاصل از موجک گسسته حداکثر هم‌پوشانی محاسبه و به‌عنوان ورودی برای منطقه‌بندی بارش مورد استفاده قرار گرفت. توزیع مکانی تغییرات موجک گسسته حداکثر هم‌پوشانی-انتروپی برای چهار زیرسری جزئی و زیرسری تقریبی در مقیاس سالانه (MDE) در شکل (۳) ارائه شده است. با توجه به شکل (۳) مشاهده می‌شود که تغییرات موجک گسسته حداکثر هم‌پوشانی انتروپی در شمال و شمال غرب ایران تغییرات بیشتری دارند. شایان ذکر است که این مناطق اکثراً بارانی و یا سرد و بعضاً نیمه‌خشک می‌باشند. در بخش‌های شرقی و جنوبی که اکثراً مناطق خشک هستند، تغییرات ملایم‌تر است. بنابراین می‌توان نتیجه گرفت که در شمال و شمال غرب کشور تغییرپذیری بارش بیشتر است. همچنین بیشترین تغییرات موجک انتروپی برای زیرسری‌های  $D_3$  (هشت سال) مشاهده شد.



شکل (۲): تجزیه سری زمانی بارش سالانه ایستگاه شهرکرد با استفاده از  $DB(4,4)$

Figure (2): Decomposition of the annual precipitation time series at the Shahr-e Kord station using  $DB(4,4)$





شکل (۳): توزیع مکانی تغییرات موجک گسسته حداکثر همپوشانی-انترویی (MDE) سالانه در ایران

Figure (3): Spatial distribution of annual MDE change in Iran

ارزیابی روش‌های خوشه‌بندی منطقه مورد مطالعه

نتایج معیارهای ارزیابی و تعیین تعداد مناسب خوشه‌ها در مقیاس سالانه در جدول (۱) ارائه شده است. با در نظر گرفتن مقادیر معیارهای ارزیابی و نتایج کیفی منطقه‌بندی ایستگاه‌های سینوپتیک، روش K-means با  $Chi = ۱۴/۵۸$ ،  $Sci = ۰/۴۱$  و  $DBi = ۱/۲۲$ ، بهترین عملکرد را داشته است.

پنج مقدار موجک گسسته حداکثر همپوشانی-انترویی سالانه (D1، D2، D3، D4 و A4) به‌عنوان ورودی‌های مدل‌های خوشه‌بندی به‌منظور منطقه‌بندی بارش کشور انتخاب شدند. برای تعیین تعداد خوشه‌ها نیز معیارهای ارزیابی سیلوه، کالینسکی-هاراباز و دیویس بولدین محاسبه شدند.

جدول (۱): انتخاب بهترین تعداد خوشه‌ها براساس شاخص‌های آماری مختلف در مقیاس سالانه (MDE)

Table (1): Selection of the optimal number of clusters based on various statistical indices in annual scale (MDE)

تعداد خوشه‌ها									شاخص	روش خوشه‌بندی
۱۰	۹	۸	۷	۶	۵	۴	۳	۲		
۱/۴۷	۱/۱۲	۱/۰۸	۱/۱۴	۱/۱۸	۱/۳۸	۱/۱۴	۱/۲۴	۱/۲۲	DBi	SOM
۱۱/۸۶	۱۴/۱۴	۱۵/۲۹	۱۶/۷۲	۱۷/۲۸	۱۵/۲۲	۱۹/۵۳	۱۸/۸۳	۱۶/۶۰	CHi	
۰/۱۹	۰/۳۰	۰/۳۲	۰/۳۹	۰/۳۷	۰/۲۶	۰/۴۱	۰/۳۹	۰/۴۴	SCi	
۱/۳۶	۱/۲۳	۱/۲۶	۰/۹۹	۱/۷۶	۱/۳۲	۱/۲۲	۱/۲۸	۱/۵۸	DBi	k-means
۱۷/۰۸	۱۵/۶۶	۱۹/۱۳	۱۵/۵۱	۱۶/۳۶	۱۷/۹۴	۱۴/۵۸	۱۹/۲۳	۱۵/۱۶	CHi	
۰/۱۹	۰/۳۰	۰/۳۲	۰/۳۹	۰/۳۷	۰/۲۶	۰/۴۱	۰/۳۹	۰/۴۴	SCi	

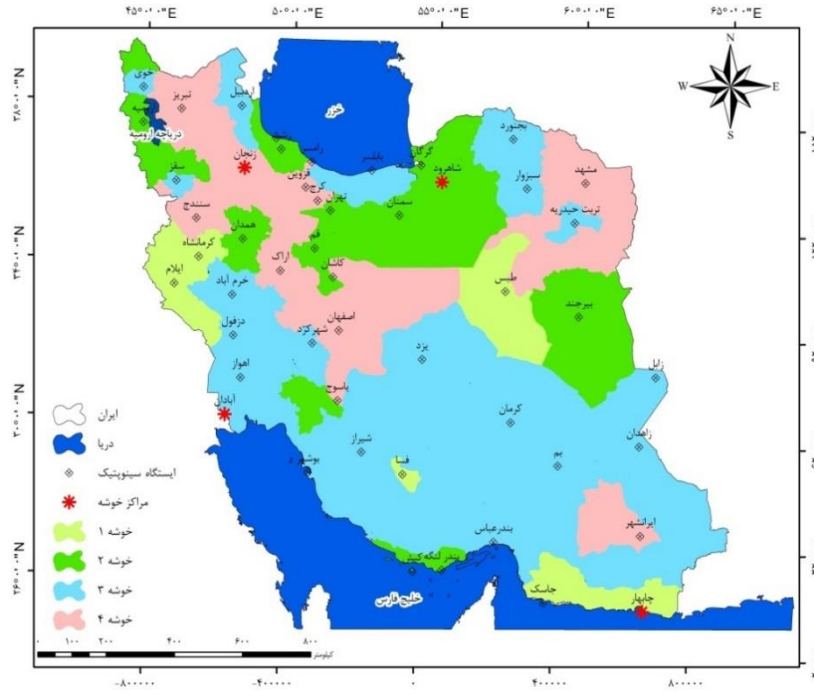
نتایج ارزیابی با استفاده از شاخص کالینسکی-هاراباز نشان داد که تعداد خوشه برابر با ۴، عملکرد بهتری در تعیین نواحی همگن بارش از لحاظ عدم قطعیت‌های چندمقیاسی انرژی-اساس داشته است؛ درحالی که ارزیابی براساس شاخص سیلوه و دیویس بولدین، تعداد خوشه برابر با ۵، ۶ یا ۸ می‌تواند عملکرد بهتری در تعیین نواحی همگن بارش داشته باشد. در تعداد خوشه‌های برابر با ۵، ۶ و ۸، ایستگاه‌های سینوپتیک مورد مطالعه از نظر کیفی پراکنش مناسبی ندارند؛ به طوری که یک خوشه حدود ۷۰ درصد ایستگاه‌ها و یک خوشه تنها دو ایستگاه را شامل می‌شود. زمانی که تعداد خوشه‌های بهینه برابر با ۴ انتخاب شود، توزیع مکانی ایستگاه‌های سینوپتیک به لحاظ آماری و تحلیلی منطقی به نظر می‌رسد.

بر این اساس، ایستگاه‌های چابهار، فسا، ایلام، جاسک، کرمانشاه و طبرستان در خوشه ۱، ایستگاه‌های بیرجند، بندر انزلی، بندرلنگه، گرگان، کاشان، خرم‌آباد، کیش، قم، رشت، سمنان، شاهرود، تهران، یاسوج و ارومیه در خوشه ۲، ایستگاه‌های آبادان، اهواز، اردبیل، بابل، بام، بندرعباس، بجنورد، بوشهر، دزفول، همدان، کرمان، خوی، سبزوار، سقز، شهرکرد، شیراز، تربت حیدریه، یزد، زابل و زاهدان در خوشه ۳ و ایستگاه‌های اراک، ایرانشهر، اصفهان، کرج، مشهد، قزوین، رامسر، سنندج، تبریز و زنجان نیز در خوشه ۴ قرار گرفتند.

#### خوشه‌بندی منطقه مورد مطالعه

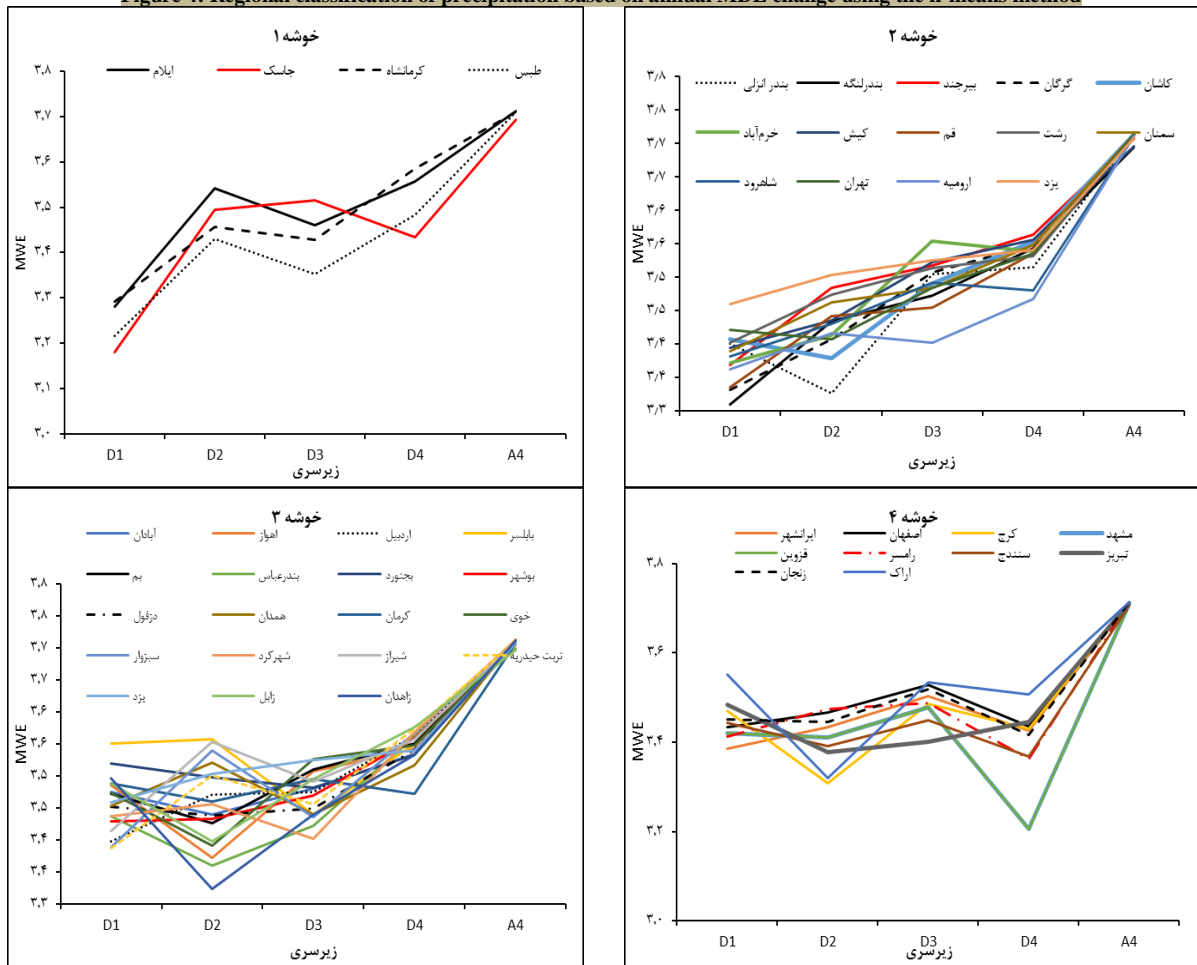
براساس نتایج معیارهای ارزیابی خوشه‌بندی، ۵۰ ایستگاه سینوپتیک کشور بر مبنای شاخص MDE به ۴ خوشه تقسیم شدند و ایستگاه‌های چابهار، شاهرود، آبادان و زنجان به عنوان مراکز خوشه انتخاب شدند (شکل ۴). مطابق نتایج ارائه شده در شکل (۴)، مشاهده می‌شود که در روش‌های آنتروپی چندمقیاسی، ایستگاه‌های سینوپتیک مربوط به یک خوشه مشخص در مناطق جغرافیایی متفاوتی پراکنش دارند و قرابت جغرافیایی ایستگاه، مبنای خوشه‌بندی نبوده است. ایستگاه‌های

روند تغییرات موجک آنتروپی چندمقیاسی (MDE) برای هر خوشه در شکل (۵) ارائه شده است. با توجه به شکل مشاهده می‌شود که مقادیر آنتروپی چندمقیاسی تغییرات زیادی دارند؛ در صورتی که از لحاظ درون خوشه‌ای همگن هستند و بیشترین شباهت را دارند. مقادیر پیک هر خوشه بیانگر بیشترین تغییرپذیری در مقیاس سالانه است.



شکل (۴): منطقه‌بندی بارش براساس مقادیر موجک گسسته حداکثر هم‌پوشانی و انترپولی سالانه با استفاده از روش kmeans

Figure 4: Regional classification of precipitation based on annual MDE change using the k-means method



شکل (۵): روند تغییرات مقادیر MWE سالانه برای خوشه‌های مختلف

Figure (5): Changes in yearly MWE values for different clusters

### بحث و نتیجه گیری

در این مطالعه، به منظور بررسی خصوصیات زمانی-مکانی بارش سالانه کشور و منطقه بندی آن، از یک روش مبتنی بر تحلیل موجک و روش K-means استفاده شد. در این روش، MODWT برای استخراج ویژگی های غیرایستا و چندمقیاسی سری زمانی بارش سالانه به کار گرفته شد و پس از محاسبه انتروپی زیرسری های حاصل، این مقادیر به عنوان ورودی روش خوشه بندی K-means استفاده شدند. نتایج نشان داد زمانی که تعداد خوشه ها برابر با ۴ در نظر گرفته شود، نتایج از نظر کیفی (پراکنش ایستگاه ها در هر خوشه) و کمی (شاخص های ارزیابی درونی) قابل توجیه است. در این روش، عدم قطعیت هیدرولوژیکی و تغییرات تناوبی زمانی- مکانی بارش قابل تشخیص است؛ به طوری که ایستگاه های هر خوشه از نظر عدم قطعیت هیدرولوژیکی و تغییرات، مشابه هم بوده و ویژگی های منحصر به فردی دارند که با خوشه های دیگر متفاوت است. منطقه بندی ایستگاه های بارش در ایران توسط محققان دیگر (رضیئی و همکاران، ۲۰۰۸؛ مدرس، ۲۰۰۶؛ دومروس و همکاران، ۱۹۹۸؛ دارند و منصور، ۲۰۱۴؛ رضیئی، ۲۰۱۸) نیز انجام شده است که نزدیکی جغرافیایی ایستگاه ها و مقدار بارش مبنای خوشه بندی بوده است؛ اما در این تحقیق خوشه بندی انجام شده نواحی را تعیین کرد که در آن هر خوشه از لحاظ عدم قطعیت و پیچیدگی مشابه هم هستند که با نتایج مطالعات روشنگر و علیزاده (۲۰۱۹) و روشنگر و همکاران (۱۳۹۹) مطابق می باشد.

عدم قطعیت بارش ممکن است ناشی از تغییرات طبیعی و اقلیمی از جمله تغییرات فصلی باد، فعالیت های انسانی و افزایش دمای زمین و... باشد. این عدم قطعیت در بارش به ویژه در مناطق خشک می تواند اثرات شدیدی بر منابع آب و مدیریت آن ها داشته باشد. در مناطق خشک، بارش محدود و ناپایدار است و در اغلب موارد در طول فصل بارندگی، همراه با وقفه های طولانی بین دوره های بارشی رخ می دهد. خوشه بندی

### منابع

1. Abghari, H., 2008. Investigation of Wavelet Neural Networks and Auto-regressive Models to Intelligent Prediction of Monthly River Flow. Ph.D. Thesis in Hydrology, Faculty of Natural Resources, University of Tehran.
2. Adamowski, K., Prokoph, A., & Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation.

ایستگاه ها می تواند تا حدی با پوشش منطقه ای، عدم قطعیت مربوط به بارش را کاهش دهد. در مناطق خشک که بارش کم است و پراکنش نامنظمی در طول سال دارد، خوشه بندی ایستگاه ها می تواند برای بهبود مدیریت منابع آب، از جمله تخصیص منابع و برنامه ریزی بلندمدت بسیار مفید باشد. از طرفی، بارندگی اولین منبع تأمین آب در سطح زمین است؛ از این رو مطالعه وضعیت و تغییرات مکانی- زمانی آن، نقش بسیار مهمی در تعیین وضعیت فعلی و آتی بیابان آبی هر منطقه و در نهایت مدیریت صحیح آب بخش های مختلف کشاورزی، شرب و صنعت دارد. این موضوع در مناطق خشک و نیمه خشک از جمله ایران، به دلیل بیابان منفی منابع آبی در اغلب بخش های آن از یک سو و همچنین وجود شرایط اقلیمی متفاوت در بخش های مختلف آن از سوی دیگر بسیار حائز اهمیت است.

خوشه بندی یکی از روش های پر کاربرد در منطقه بندی خصوصیات مختلف بارندگی است که با تعیین مناطق همگن، مطالعه خصوصیات بارندگی در مناطق همگن را بسیار دقیق تر و راحت تر می سازد. این مسئله موجب می گردد تا راهکارها و تصمیمات مدیریتی مناسبی با شرایط هر منطقه همگن اتخاذ گردد.

شناخت مناطق هیدرولوژیکی همگن و شناسایی ویژگی های بارش مرتبط، کارایی مدیریت منابع آب را در سازگاری با تغییرات اقلیمی، جلوگیری از تخریب محیط زیست و منابع آبی و کاهش اثرات بلایای ناشی از تغییر اقلیم بهبود می بخشد. روش پیشنهادی می تواند برای خوشه بندی مکانی ایستگاه های سینوپتیک و در نهایت مدیریت منابع آب و سایر اهداف مرتبط با بارش و متغیرهای مربوط به آن (مانند رواناب، رطوبت خاک و...) مفید باشد. با این حال، این مطالعه با محدودیت تعداد ایستگاه های سینوپتیک و طول دوره آماری داده ها مواجه بوده است؛ لذا کاربرد مدل برای مناطق با ایستگاه های بیشتر و با مقیاس های زمانی مختلف پیشنهاد می شود.

- Hydrological Processes 23(18), 2686-2696.
3. Addou, R., Hanchane, M., Krakauer, N. Y., Kessabi, R., Obda, K., Souab, M., & Achir, I. E., 2023. Wavelet Analysis for Studying Rainfall Variability and Regionalizing Data: An Applied Study of the Moulouya Watershed in Morocco. *Applied Sciences* 13(6), 3841.
  4. Agarwal, A., Maheswaran, R., Sehgal, V., Khos, R., Sivakumar, B., & Bernhofer, C., 2016. Hydrologic regionalization using wavelet-based multiscale entropy method. *Journal of Hydrology* 538, 22-32.
  5. Alizadeh Gharamaleki, F. 2018. Investigating changeability of precipitation and related variables using Empirical Mode Decomposition algorithm and spatial clustering. Ph.D. Thesis in civil engineering Hydraulic structures, Aras International Campus, University of Tabriz.
  6. Armesh, M., 2017. Analysis of the Variability in Indian Summer Monsoon precipitations Influences on the Southeast of Iran Regarding to Teleconnection and Climatic Indices. Ph.D. Thesis in Climatology, Faculty of Geography and Environmental Planning, University of Sistan and Baluchestan.
  7. Canchala, T., Ocampo-Marulanda, C., Alfonso-Morales, W., Carvajal-Escobar, Y., Ceron, W. L., & Caicedo-Bravo, E., 2022. Techniques for monthly rainfall regionalization in southwestern Colombia. *Anais da Academia Brasileira de Ciências* 94.
  8. Chen, N., Chen, L., Ma, Y., & Chen, A., 2019. Regional disaster risk assessment of China based on self-organizing map: clustering, visualization and ranking. *International Journal of Disaster Risk Reduction* 33, 196-206.
  9. Claps, P., Ganora, D., & Mazzoglio, P., 2022. Rainfall regionalization techniques. In *Rainfall*. Elsevier. Amsterdam, Nederland (pp. 327-350).
  10. Darand, M., Mansouri Daneshvar, M.R., 2014. Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Environmental Processes* 1, 517-532.
  11. Davies, D.L., & Bouldin, D.W., 1979. A cluster separation measure. *IEEE Trans Pattern*
  12. *Anal Mach Intell* 1(2):224-227.
  13. Domroes, M., Kaviani, M., & Schaefer, D., 1998. An analysis of regional and intra-annual precipitation variability over Iran using multivariate statistical methods. *Theoretical and Applied Climatology* 61:151-159.
  14. Isazade, M., & Dinpashoh, Y., 2018. Spatio-Temporal delineation of Iran's Precipitation Climate and Selection of Indicator Stations Using the Multivariate Statistical Methods. *Water and Soil Science* 28(3), 169-181.
  15. Fazel, N., Berndtsson, R., Uvo, C. B., Madani, K., & Kløve, B., 2018. Regionalization of precipitation characteristics in Iran's Lake Urmia basin. *Theoretical and Applied Climatology* 132, 363-373.
  16. Govender, P., & Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980-2019). *Atmospheric pollution research* 11(1), 40-56.
  17. Guntu, R.K., Maheswaran, R., Agarwal, A., & Singh, V.P. 2020. Accounting for temporal variability for improved precipitation regionalization based on self-organizing map coupled with information theory. *Journal of Hydrology* 590, 125236.
  18. Hartigan, J. A., & Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Applied Statistics* 28(1), 100-108.
  19. Hsu, K.C., & Li, S. T., 2010. Clustering spatial-temporal precipitation data using wavelet transform and self-organizing map neural network. *Advances in Water Resources* 33(2), 190-200.
  20. Ilbay-Yupa, M., Lavado-Casimiro, W., Rau, P., Zubieta, R., & Castellón, F., 2021. Updating regionalization of precipitation in Ecuador. *Theoretical and Applied Climatology* 143, 1513-1528.
  21. Isazadeh, M., & Dinpashoh, Y. (2018). Spatio-Temporal delineation of Iran's Precipitation Climate and Selection of Indicator Stations Using the Multivariate Statistical Methods. *Water and Soil Science*, 28(3), 169-181.
  22. Khosravi, M., Doostkamian, M., Mirmoosavi, S.H., Bayat, A., & Beyg Rezaei, E., 2014. Classification of temperature and precipitation in Iran using geo statistics and cluster analysis methods. *Journal of Regional Planning* 4(13), 121-132.
  23. Kasturi, J., Acharya, J., & Ramanathan, M., 2003. An information theoretic approach for analyzing temporal patterns of gene expression. *Bioinformatics* 19(4):449-458.
  24. Kohonen, T., 2013. Essentials of the self-organizing map. *Neural networks* 37, 52-65.
  25. Li, M., Jiang, Z., Zhou, P., Le Treut, H., & Li, L., 2020. Projection and possible causes of summer precipitation in eastern China using self-organizing map. *Climate Dynamics* 54(5-6), 2815-2830.
  26. Li, C., & Biswas, G., 2002. A Bayesian approach for structural learning with hidden Markov models. *Scientific Programming* 10(3), 201-219.
  27. Lucieer, V., & Lucieer, A., 2009. Fuzzy clustering for seafloor classification. *Marine Geology* 264(3-4), 230-241.
  28. MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proceeding of Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 281-297.

29. Mafakheri, O., Saligheh, M., Alijani, B., & Akbary, M., 2017. Zonation of temporal changes and uniformity of rainfall in Iran. *Physical Geography Quarterly* 49(2), 191-205.
30. Masoudian, S.A., Darand, M., & Karsaz, S., 2011. Rainfall zoning in west and north-west of Iran by cluster analysis method. *Physical Geography Quarterly* 4(11), 35-44.
31. Mishra, A.K., Özger, M., & Singh, V.P., 2009. An entropy-based investigation into the variability of precipitation. *Journal of Hydrology* 370(1-4), 139-154.
32. Modarres, R., 2006. Regional precipitation climates of Iran. *Journal of Hydrology*: 45 (1), 13-27.
33. Mohamadyariyan M, Tavousi T, khosravi M, Hamidiyanpour M. Zoning of Iranian Heavy Precipitation Regime. *GeoRes* 2019; 34 (2) :183-192.
34. Naikoo, M.W., Talukdar, S., Das, T., & Rahman, A., 2022. Identification of homogenous rainfall regions with trend analysis using fuzzy logic and clustering approach coupled with advanced trend analysis techniques in Mumbai city. *Urban Climate* 46, 101306.
35. Nourani, V., Roushangar, K., & Andalib, G., 2018. An inverse method for watershed change detection using hybrid conceptual and artificial intelligence approaches. *Journal of Hydrology* 562, 371-384.
36. Omidvar, K., Omidvar, K., & Khosravi, Y. (2010). Investigation of Change of Some Climatic Elements in North Coast of Persian Gulf Using Kendal Test. *Geography and Environmental Planning*, 21(2), 33-46.
37. Paul, R.K., Paul, A.K., & Bhar, L.M., 2020. Wavelet-based combination approach for modeling sub-divisional rainfall in India. *Theoretical and Applied Climatology* 139, 949-963.
38. Parviz, L., & Beyrami, A., 2023. Evaluation of empirical mode decomposition and maximal overlap discrete wavelet transform approaches in trend analysis of meteorological variables. *Journal of Agricultural Meteorology*. Articles in Press, Available Online from 15 May 2023
39. Razinei, T., 2016. Identification of homogeneous precipitation sub-regions for Iran using principal component analysis. *Iranian Journal of Geophysics* 10(3), 128-144.
40. Razinei, T., 2018. A precipitation regionalization and regime for Iran based on multivariate analysis. *Theoretical and applied climatology* 131(3-4), 1429-1448.
41. Razinei, T., Bordi, I., & Pereira, L.S., 2008. A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrology and Earth System Sciences* 12, 1309-1321.
42. Roradeh, H., Ghasemi, J., Yousefi, Y., & Ghasemi, Z., 2019. Clustering the rainfall of Iran with using new approach based on Singular Value Decomposition Mapping and Fuzzy C-Means Clustering. *Geographical Planning of Space Quarterly Journal* 9(31), 113-124.
43. Roushangar, K., & Alizadeh, F., 2019. Using multi-temporal analysis to classify monthly precipitation based on maximal overlap discrete wavelet transform. *Journal of Hydroinformatics* 21(4), pp.541-557.
44. Roushangar, K., Dolatshahi, M., & Alizadeh, F., 2023. MODWT and wavelet coherence-based analysis of groundwater levels changes detection. *Paddy and Water Environment* 21, 1, 59-83.
45. Roushangar, K., Moghaddas, M., & Ghasempour, R. (2020). Investigation of Temporal-Spatial Characteristics of Precipitation Using Discrete Maximal Overlap Wavelet Transform (MODWT) and Spatial Clustering Tools. *Journal of Climate Research*, 1399(43), 143-158.
- Saghafi, M., Barati, G.R., Alijani, B., & Moradi, M., 2024. Zoning and analysis of pervasive rainfall in rainy areas of Iran In the statistical period of 30 years (1987-1987). *Journal of Applied researches in Geographical Sciences* 23(71), 103-121.
46. Santos, M., Fragoso, M., & Santos, J.A., 2017. Regionalization and susceptibility assessment to daily precipitation extremes in mainland Portugal. *Applied Geography* 86, 128-138.
47. Shannon, C.E., 1948. A mathematical theory of communication. *Bell system technical journal* 27(3), 379-423.
48. Sinaga, K. P., & Yang, M.S., 2020. Unsupervised K-means clustering algorithm. *IEEE access* 8, 80716-80727.
49. Taghavi, F., Nasserli, M., Bayat, B., Motevallian, S. S., & Azadifard, D., 2012. The Identification of Climatic Patterns of Iran Based on Spectral Analysis and Clustering of Precipitation and Temperature Extreme Values. *Physical Geography Research Quarterly* 43(77), 109-124.
50. Wang, X., & Xu, Y., 2019. "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index". 2nd International Conference on Advanced Materials, Intelligent Manufacturing and Automation. China
51. Yazawa, T., & Shoji, A., 2023. Spatial analysis of historical extreme rainfall characteristics using regionalization in the Lake Biwa and Yodo River Basin, Japan. *Journal of Water and Climate Change* 14(3), 916-936.
52. Zerouali, B., Chettih, M., Abda, Z., Mesbah, M., Santos, C.A.G., & Brasil Neto, R. M., 2022. A new regionalization of rainfall patterns based on wavelet transform information and hierarchical cluster analysis in northeastern Algeria. *Theoretical and Applied Climatology* 147(3-4), 1489-1510.

## Investigating and Classifying Temporal-Spatial Characteristics of Iran's Annual Precipitation Using Maximal Overlap Discrete Wavelet Transform and Multiscale Entropy

Sepideh Choobeh<sup>۱</sup>, Hiran Abghari<sup>۲\*</sup>, Mahdi Erfanian<sup>۳</sup>

Received: 30/05/2023

Accepted: 17/06/2023

### Extended Abstract

**Introduction:** Assessing precipitation alterations in a large area like Iran is required for the identification of those areas that are more vulnerable to changes in precipitation patterns, considering the fact that such changes may significantly influence water availability, agriculture, and other sectors that are dependent on water resources. On the other hand, understanding the spatial variability of precipitation patterns can help develop purposive strategies, including drought or flood management in specific regions. Moreover, as severe weather events such as floods and droughts can devastate communities and their infrastructure, such an understanding can inform decisions made concerning disaster risk reduction efforts. Therefore, assessing precipitation variations is essential for the effective management of water resources and the reduction of disaster risks.

**Materials and methods:** This study suggests a new method for analyzing precipitation properties in Iran, using a mixture of Maximal Overlap Discrete Wavelet Transform (MODWT) and Multiscale Entropy (MDE) techniques. This approach allows for a more detailed and accurate assessment of the spatial and temporal characteristics of precipitation properties, preparing the ground for the development of appropriate strategies for different regions in Iran. To this end, annual precipitation data collected from fifty Iranian synoptic stations for 1980-2020 were analyzed. Then, after classifying the precipitation data into different subseries, the concept of entropy was used to measure precipitation variability. Moreover, MDE values were used as input data for clustering purposes, followed by the calculation of internal evaluation criteria to be used for the determination of the optimal number of clusters and the most suitable clustering method calculated.

**Results:** The variations and trends of the precipitation data can be identified through the analysis of partial coefficients D1-D4 and the approximation coefficient. Accordingly, while the smaller partial sub-categories indicate more rapid variations at higher frequencies, the greater partial coefficients show more moderate variations at lower frequencies. Moreover, the approximation coefficient reveals the slightest variations at low frequencies in annual time series. The study's results suggested that northern and northwestern Iranian regions that are primarily characterized by rainy, cold, and in some cases semi-arid climates experienced the greatest variations in terms of annual precipitation. On the other hand, the eastern and southern parts of Iran, which are mostly dry areas, experienced more moderate variations in annual precipitation rates. Therefore, according to the results found in this study, it could be argued that northern and northwestern Iran enjoy more precipitation variability than other parts of the country. Furthermore, the D3 sub-category (eight years) was found to have the

1. PhD Student of Watershed Management, Faculty of Natural Resources, Urmia University, Urmia, Iran; s.choobeh@urmia.ac.ir

2. Corresponding author: Associate Professor, Department of Range and Watershed Management, Faculty of Natural Resources, Urmia University, Urmia, Iran; h.abghari@urmia.ac.ir

3. Associate Professor, Department of Range and Watershed Management, Faculty of Natural Resources, Urmia University, Urmia, Iran; m.erfanian@urmia.ac.ir

doi: 10.22052/deej.2023.253023.1015

greatest variations in terms of MWE. On the other hand, based on the values of Sci, DBi, and CH index, the k-means clustering method performed better than SOM (Sci=0.41, DBi= 1.22, and CH= 14.58). Finally, fifty Iranian synoptic stations were categorized into four clusters based on the MWE index, with Chabahar, Shahrud, Abadan, and Zanjan being selected as core clusters.

**Discussion and Conclusion:** The current study proposed a methodology for analyzing and zoning Iran's annual precipitation based on the multiscale entropy method, considering the fact that the period and trend of the annual time series could be identified via the analysis of the precipitation series. Following the analysis of the collected data, this study used the multiscale entropy method to record precipitation variability in each synoptic station. On the other hand, the determination of SOM and k-means input data based on MWE values helped reduce the input data, leading to an increase in the accuracy of the zoning method and the selection of the homogenous clusters based on the proposed methodology. The results of the study indicated that generally, the K-means method offered more homogeneous areas than the SOM method. Moreover, the homogeneous distribution of annual precipitation variations obtained based on the K-means-derived clusters confirmed the positive performance of the methodology proposed in this study, which involves the identification of hydrologic uncertainty and temporal-spatial variations of precipitation in those stations related to a specific cluster. The methodology seeks to create distinct clusters of stations that share common features. This approach can provide insights into the spatial variability of hydrological processes and help improve water resource management by identifying areas susceptible to hydrological extremes such as floods or droughts. Some studies have already been conducted on the spatial clustering of precipitation stations in Iran using different methods, including the application of geographical proximity or precipitation rate as criteria for clustering. However, the method used in the current study involved clustering based on similarity in hydrological uncertainty and temporal-spatial complexity, which is consistent with the findings of Roshangar and Alizadeh (2019), and Roshangar et al. (2019). Identifying hydrologically homogeneous regions and their associated precipitation characteristics can significantly enhance the effective management of water resources in terms of adapting to climate change, preventing damage to water environments, and mitigating the impact of climate-related disasters. Therefore, the proposed methodology for spatial clustering of synoptic stations could be useful in managing water resources and all precipitation-related sectors and variables such as runoff and soil moisture. However, this study faced some limitations such as the small number of synoptic stations and the short length of the statistical period. Therefore, it is recommended that the methodology be applied to more stations at more varied time scales.

**Keywords:** Clustering, Discrete Wavelet Transform, Entropy, Iran, Maximal Overlap, Precipitation.