

تحلیل فراوانی منطقه‌ای بارش‌های حداکثر سالانه یک و دو روزه با خوشه‌بندی و نظریه گشتاورهای خطی، مطالعه موردی: استان خراسان رضوی

مهری شاهی^۱ - سید حسین ثنائی نژاد^{۲*} - بیژن قهرمان^۳

تاریخ دریافت: ۱۳۹۰/۱۱/۲۹

تاریخ پذیرش: ۱۳۹۱/۵/۱۵

چکیده

هدف این پژوهش، تحلیل فراوانی منطقه‌ای بارش‌های حداکثر سالانه یک روزه (AM1R) و دو روزه (AM2R) با استفاده از دیدگاه گشتاورهای خطی در استان خراسان رضوی است. بدین منظور ابتدا آزمون‌های پایه آماری شامل: همگنی، استقلال و داده پرت بر روی داده‌های AM1R و AM2R در ۱۲۳ ایستگاه باران‌سنجی استان انجام شد. این استان با روش تحلیل خوشه‌ای و رعایت الگوهای توپوگرافی و بارش متوسط سالانه، به ۴ ناحیه تقسیم، سپس همگنی هیدرولوژیکی این نواحی با آزمون ناهمگنی کنترل و تایید شد. پارامترهای توابع توزیع مقادیر حدی تعمیم‌یافته (GEV)، لوجستیک تعمیم‌یافته (GLO)، پیرسون نوع III (PE3) و لوگ نرمال سه پارامتری (LN3) برای داده‌های AM1R و AM2R، در هر ناحیه برآورد گردید. برای انتخاب تابع توزیع مناسب، نیکویی برازش آن‌ها با نمودار نسبت گشتاورهای خطی و آزمون کلموگروف-اسمیرنوف بررسی و تابع توزیع GEV انتخاب شد. سپس چندک‌های ناحیه‌ای توزیع GEV برای داده‌های AM1R و AM2R به ازای دوره بازگشت‌های مختلف محاسبه شدند. به طور کلی روابط به دست آمده برای برآورد AM1R و AM2R در دوره بازگشت‌های کمتر از ۲۰۰ سال، در تمامی نواحی همگن از نزدیکی نسبتاً خوبی با آمار واقعی برخوردار هستند. میانگین خطای مطلق بین مقادیر بی‌بعد توزیع نظری و مشاهداتی ایستگاه‌های هر ناحیه مطلوب است (بزرگترین MAE معادل ۰/۰۳۸۶ میلی‌متر) که نشان‌دهنده دقت بالای برآوردها است.

واژه‌های کلیدی: تحلیل فراوانی منطقه‌ای، گشتاورهای خطی، بارش‌های حداکثر سالانه یک و دو روزه، روش باران‌نما، تحلیل خوشه‌ای

مقدمه

عامل وقوع سیلاب فقط رگبارهای شدید یا وقوع یک بارش با تداوم مشخص نیست، بلکه بارش‌های منقطع در فاصله زمانی یک یا دو روز نیز می‌تواند سیلاب‌های مخربی تولید کند. زیرا درصد رطوبت خاک قبل از بارندگی، نقش مهمی را در حجم سیلاب ناشی از باران ایفا می‌کند. لذا در این پژوهش، کل بارش‌های به وقوع پیوسته در بازه‌های زمانی یک و دو روزه، مورد نظر بوده است. از طرفی تراکم اندک ایستگاه‌های باران‌سنجی در مناطق خشک و نیمه خشک، کوتاه بودن و عدم یکنواختی طول دوره آماری در این ایستگاه‌ها و وجود داده‌های پرت و غیر قابل اطمینان، مهمترین مشکلات موجود برای تحلیل فراوانی بارش‌های حدی به شمار می‌روند. تحلیل منطقه‌ای برای رفع مشکلات مذکور، استفاده می‌شود. این روش به جای استفاده از آمار یک ایستگاه، از آمار و ویژگی‌های مجموعه‌ای از ایستگاه‌های هم‌رفتار استفاده می‌کند. مطالعات مختلف با استفاده از تکنیک‌های منطقه‌ای بر روی بارش‌های حدی نشان می‌دهند که این روش‌ها تردیدها را در زمینه تخمین ناشی از دیدگاه ایستگاهی به‌طور بالقوه‌ای

تحلیل فراوانی منطقه‌ای بارش‌های حداکثر سالانه یک و چند روزه، می‌تواند در چگونگی مدیریت منابع آب و طراحی کاربری اراضی اطلاعات لازم در اختیار برنامه‌ریزان و متخصصان کشور قرار دهد. بارش‌های حداکثر سالانه که در دوره بازگشت‌های پایین از نظر مقدار قابل توجه هستند، می‌تواند به عنوان پتانسیل منابع آبی شناخته شوند. یک کاربرد آن مکان‌یابی این‌گونه بارش‌ها و اجرای پروژه‌های تغذیه مصنوعی است که می‌تواند از سیلاب‌های مخرب و هدر رفت ناشی از تبخیر جلوگیری کند. همچنین بارش‌های شدید سبب فرسایش خاک شده و مدیریت کاربری اراضی را به دلیل تغییر شرایط رشد گیاه و عملیات کشاورزی، دچار مشکل می‌کند.

۱، ۲ و ۳- به ترتیب دانشجوی کارشناسی ارشد، دانشیار و استاد گروه مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد

(Email: sanaein@gmail.com)

(*- نویسنده مسئول)

ناحیه تقسیم و آزمون همگنی و استخراج توابع توزیع آماری را با استفاده از روش گشتاورهای خطی موفق ارزیابی کردند. بر این اساس مهمترین اهداف این پژوهش به شرح زیر است: ۱- تعیین نواحی همگن از نظر بارش‌های حداکثر سالانه یک و دو روزه در استان خراسان رضوی با تحلیل خوشه‌ای و روش گشتاورهای خطی ۲- برآورد پارامترهای توابع توزیع مختلف با روش گشتاورهای خطی انتخاب تابع توزیع مناسب بارش‌های حداکثر یک روزه و دو روزه سالانه هر ناحیه.

مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

در این پژوهش کل بارش‌های به وقوع پیوسته در بازه‌های زمانی یک و دو روزه مورد نظر بوده است. بدین منظور آمار ۱۲۳ ایستگاه در استان خراسان رضوی انتخاب و بارش‌های حداکثر یک‌روزه (AMIR) و بارش‌های حداکثر دو روزه (AM2R) در هر سال آماری استخراج شده است. این ایستگاه‌ها شامل ۷۸ ایستگاه باران‌سنجی و ۳۴ ایستگاه تبخیرسنجی مربوط به شرکت آب منطقه‌ای خراسان رضوی و ۱۰ ایستگاه سینوپتیک و یک ایستگاه کلیماتولوژی از سازمان هواشناسی می‌باشد (۲ و ۳). همچنین آزمون‌های گردش، والد-ولفویتز، من-ویتنی و گروبز-بک به ترتیب برای تشخیص تصادفی بودن، وجود روند، جهش و همگنی و وجود داده پرت در داده‌های AM1R و AM2R هر ایستگاه استفاده شده است (۱). از میان ایستگاه‌های مورد استفاده، ایستگاه بار-اریه با ۵۹ سال آمار بیشترین و ایستگاه حصار درگز با ۱۵ سال آمار کمترین طول دوره آماری را دارا هستند.

تحلیل فراوانی منطقه‌ای با استفاده از روش باران نمایه^۱

تحلیل فراوانی به برآورد چگونگی وقوع یک پیشامد ویژه می‌پردازد. برآورد فراوانی حوادث حدی اغلب از اهمیت زیادی برخوردار است؛ زیرا منابع متعددی از عدم قطعیت درباره فرآیندهای فیزیکی مولد وقایع مشاهده‌ای، وجود دارد. به همین دلیل، دیدگاه آماری برای تحلیل داده‌های حدی مطلوب خواهد بود (۷).

در روش باران‌نمایه فرض می‌شود که توزیع فراوانی بارش‌های حدی همه ایستگاه‌ها در یک ناحیه همگن یکسان است؛ به جز در عامل مقیاس یا باران‌نمایه که انعکاس دهنده خواص متفاوت بارندگی در هر ایستگاه می‌باشد. اگر N ایستگاه واقع در یک منطقه موجود باشد، $Q_i(F)$ چندک احتمال عدم تجاوز F در ایستگاه i ام با طول دوره آماری n_i مطابق رابطه ۱ برآورد می‌شود:

کاهش می‌دهند (۱۱). دارانز و کربای تابع توزیع GEV را برای الگوبندی بارش‌های حدی ساعتی و روزانه برگزیدند. آن‌ها با تحلیل خوشه‌ای و برآوردهای بارش حدی اقدام به منطقه‌ای کردن این‌گونه بارش‌ها برای اطلس باران آلاباما کردند. آن‌ها روش منطقه‌ای کردن را برای تعمیم نتایج در مناطق فاقد ایستگاه مناسب ارزیابی نمودند (۵). اما از مشکلات عمده تعمیم نتایج تحلیل فراوانی از یک ایستگاه به یک منطقه، عدم همگنی مناطق مورد بررسی از نظر هیدرولوژیک می‌باشد. با وجود کاربردی بودن روش‌های تحلیل خوشه‌ای در گروه‌بندی ویژگی‌های هیدرولوژیک، همگن بودن گروه‌های حاصل صد درصد حاصل نمی‌شود. به این منظور پیشنهاد می‌شود نتایج تحلیل خوشه‌ای با روش‌های دیگر مورد آزمون قرار گیرد (۹). یکی از روش‌های مناسب برای بررسی همگنی خوشه‌ها استفاده از گشتاورهای خطی است.

گشتاورهای خطی توسط هاسکینگ (۱۹۹۰) معرفی شد و در تحلیل‌های فراوانی منطقه‌ای کاربردهای بسیاری پیدا کرد. مهم‌ترین کاربرد گشتاورهای خطی، تعیین ایستگاه‌های ناهمگن، تشخیص نواحی همگن، تخمین پارامترهای توابع توزیع و تعیین تابع توزیع مناسب می‌باشد (۷). گشتاورهای خطی برآوردهای نا اریب هستند و نسبت به حجم نمونه حساسیت کمتری دارند. اگر میانگین داده‌ها (گشتاور خطی مرتبه اول) متناهی باشد، آن‌گاه سایر گشتاورهای خطی نیز وجود دارند، در حالی که گشتاورهای معمولی فاقد این خاصیت هستند (۷).

مطالعات زیادی با روش گشتاورهای خطی به تحلیل فراوانی بارش‌های حدی پرداخته‌اند و توابع توزیع مناسب هر منطقه را استخراج نموده‌اند. به عنوان مثال سون و سونگ این روش را برای استخراج بارش حداکثر سالانه ۳۸ ایستگاه کره جنوبی به کار بردند و با استفاده از آزمون نیکویی برازش کلموگروف-اسمیرنوف و نمودار نسبت‌های گشتاور خطی، دو توزیع GLO و GEV را برای آن منطقه مناسب تشخیص دادند (۱۲). یانگ و همکاران (۱۵) حوضه رودخانه پیرل چین را به ۶ خوشه همگن تقسیم کرده و تحلیل فراوانی منطقه‌ای را برای بارش‌های حداکثر سالانه ۱، ۳، ۵ و ۷ روزه با نظریه گشتاورهای خطی انجام دادند. نتایج آزمون‌های نیکویی برازش نشان داد که توابع توزیع GLO، GEV، PE3 به خوبی در بیشتر حوضه‌ها در نواحی همگن برازش داده می‌شوند. به طور مشابه نگونگوندو و همکاران (۸) به تحلیل فراوانی بارش‌های حداکثر سالانه ۱، ۳، ۵ و ۷ روزه برای ۲۳ ایستگاه مالووی پرداختند. آن‌ها نیز در تمام مراحل تحلیل فراوانی از دیدگاه گشتاورهای استفاده کرده و با بررسی خطای برآوردها با آمار مشاهده‌ای، دقت نتایج را خوب ارزیابی نمودند. ابوالوردی و خلیلی (۴) قابلیت گشتاورهای خطی را در تقسیم یک منطقه بزرگ به نواحی همگن بر روی داده‌های حداکثر بارش‌های ۲۴ ساعته سالانه مورد تحقیق قرار دادند. آن‌ها جنوب غرب ایران را به ۴

مطالعات تحلیل منطقه‌ای، از میانگین‌های وزنی گشتاورها و نسبت‌های آن‌ها، با در نظر گرفتن طول دوره آماری هر ایستگاه استفاده می‌شود.

تعیین نواحی همگن

معمولاً تشخیص مناطق همگن در بین تمامی مراحل تحلیل فراوانی منطقه‌ای، دشوارترین گام است. هدف در این مرحله، گروه‌بندی ایستگاه‌هایی است که به جز عامل مقیاس در نوع تابع توزیع مشترک باشند (۷). این کار در دو گام انجام می‌شود. نخست با انتخاب متغیرهای مختلف خوشه‌بندی صورت می‌گیرد که در این پژوهش از روش تحلیل خوشه‌ای افزایی با الگوریتم k-means برای خوشه‌بندی اولیه استفاده شده است. گام دوم تأیید خوشه است که هم‌نواپی و همگنی ایستگاه‌های هر خوشه با آماره‌های مناسب گشتاور خطی بررسی می‌شود. همچنین جهت تعیین بهترین خوشه‌بندی در گام اول از شاخص عرض سیلپوت، استفاده می‌شود. این شاخص توسط روسیو (۱۰) به صورت رابطه ۴ معرفی شده است.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

در این رابطه متوسط فاصله i امین عضو موجود در یک خوشه با دیگر اعضاء آن خوشه، $b(i)$ ، حداقل متوسط فاصله i امین عضو موجود در یک خوشه با تمامی اعضاء خوشه j ، $(j = 1, \dots, k; j \neq k)$ ، بوده و همواره $-1 \leq s(i) \leq 1$ است. خوشه‌بندی که منجر به تولید حداکثر عرض سیلپوت متوسط شود، خوشه‌بندی بهینه خواهد بود. معمولاً شناخت وضعیت جغرافیایی و اقلیمی منطقه (نگاه کارشناسی) نقش عمده‌ای را در تعیین مناطق همگن ایفا می‌کند، به طوری که بعضی پژوهشگران در تعیین مناطق همگن تلفیق روش‌های تحلیل خوشه‌ای و تجربه را ترجیح داده و خوشه‌های تعیین شده از روش تحلیل خوشه‌ای را مورد بازبینی قرار می‌دهند (۷ و ۱۵).

تعیین ایستگاه‌های ناهم‌نوا و آزمون همگنی خوشه‌ها

ناحیه‌ای شامل N ایستگاه در نظر گرفته می‌شود، به طوری که ایستگاه i ام دارای طول دوره آماری n_i و نسبت گشتاورهای خطی $t_3^{(i)}$ ، $t_4^{(i)}$ و $t_4^{(i)}$ باشد. معیار ناهم‌نواپی D_i برای ایستگاه i ام با استفاده از این نسبت‌ها محاسبه می‌شود. اگر D_i بزرگ باشد، ایستگاه i ام ناهم‌نوا تلقی می‌شود. مقدار بزرگی، به تعداد ایستگاه‌های موجود در گروه بستگی دارد. هنگامی که $N \geq 15$ مقدار بحرانی ۳ برای D_i در نظر گرفته می‌شود. اگر D_i از مقدار بحرانی آن بیشتر باشد، ایستگاه مورد نظر ناهم‌نوا خواهد بود. برای محاسبه اندازه ناهمگنی نیز ابتدا آماره V_1 با استفاده از رابطه ۵ برای

$$Q_i(F) = \mu_i q(F), i = 1, \dots, N. \quad (1)$$

در رابطه ۱ $q(F)$ چندک ناحیه‌ای بدون بعد با احتمال عدم تجاوز F می‌باشد و به آن منحنی رشد منطقه‌ای^۱ نیز می‌گویند. همچنین $\hat{\mu}_i$ ، عامل مقیاس یا باران نمایه در ایستگاه i ام است، که معمولاً پارامترهایی مثل میانگین یا میان در نظر گرفته می‌شود. داده‌های مقیاس شده بدون بعد $q_{ij} = Q_{ij} / \hat{\mu}_i$ ، $j = 1, \dots, n_i$ ، در این $i = 1, \dots, N$ پایه و اساس تخمین $q(F)$ هستند (۷). در این روش پارامترها به طور جداگانه در هر ایستگاه برآورد می‌شوند. k امین پارامتر برآورد شده ایستگاه i ام به شکل $\hat{\theta}_k^{(i)}$ نمایش داده می‌شود. جهت تخمین پارامترهای ناحیه‌ای، می‌بایست پارامترهای هر ایستگاه را با یکدیگر ترکیب کرد. یعنی با استفاده از طول دوره آماری هر ایستگاه به پارامترهای آن ایستگاه وزن داده می‌شود (۷).

معرفی گشتاورهای خطی

گرین‌وود و همکاران (۵) در سال ۱۹۷۹ گشتاورهای وزنی احتمال PWM را که برای تخمین پارامترهای توابع توزیع احتمالی به کار می‌روند به شکل رابطه ۲ معرفی کردند.

$$\beta_r = E\{X[F(X)]^r\} \quad (2)$$

در این رابطه X متغیر تصادفی با تابع توزیع تجمعی F ، و امید ریاضی E است. مقادیر β_r نیز گشتاورهای وزن دار احتمالی و r مرتبه گشتاور می‌باشد. هاسکینگ (۷) گشتاورهای خطی را که ترکیباتی خطی از گشتاورهای وزن دار احتمالی‌اند، مطابق رابطه ۳ معرفی کرد.

$$\lambda_{r+1} = \sum_{k=0}^r (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \beta_k \quad (3)$$

λ_1 (میانگین گشتاور خطی)، معیاری از تمایل به مرکز به عنوان اندازه یا معیار موقعیت، و λ_2 (انحراف معیار گشتاور خطی)، معیاری از پراکنندگی است. نسبت‌های گشتاورهای خطی نیز توسط هاسکینگ به صورت $\tau = \lambda_2 / \lambda_1$ و $\tau_r = \lambda_r / \lambda_2$ (برای $r > 2$) ارائه شده است (۷). آماره τ ، ضریب تغییرات گشتاور خطی^۲، آماره τ_3 ، ضریب چولگی گشتاور خطی^۳، و آماره τ_4 ، به عنوان ضریب کشیدگی گشتاور خطی^۴، نامیده می‌شوند. مقادیر λ_1 ، λ_2 ، τ_3 و τ_4 ، پرکاربردترین آماره‌های گشتاور خطی هستند، به طوری که می‌توان از آن‌ها جهت انتخاب تابع توزیع مناسب نمونه‌های مشاهده شده استفاده کرد (۷). در

- 1- Regional growth curve
- 2- L-coefficient of variation (L-cv)
- 3- L-skewness (L-skew)
- 4- L-kurtosis یا L-kurt

$$MAE = \frac{1}{n} \sum_{k=1}^n |y_k - o_k| \quad (7)$$

(y_k و o_k) در این رابطه، k امین جفت از مقادیر بی‌بعد بارش حدی نظری و تجربی با احتمال مساوی هستند. MAE میانگین حسابی مقادیر مطلق اختلافات بین اعضای هر جفت می‌باشد. اگر مقادیر برآوردی از توزیع نظری با مقادیر مشاهداتی یکسان باشند، MAE صفر و هرچه این اختلاف بزرگتر باشد، MAE افزایش می‌یابد (۱۴). نرم‌افزار مورد استفاده در این پژوهش MATLAB نسخه R2008a است.

نتایج و بحث

غریبال داده‌ها و آزمون‌های پایه‌ای

ابتدا بارش‌های حداکثر یک روزه و دو روزه سالانه در هر ایستگاه استخراج و با انجام غریبال اولیه بر روی آن‌ها، نواقص آماری برطرف شد. سپس داده‌های هر ایستگاه از لحاظ عدم وجود روند و جهش، عدم وجود داده پرت، تصادفی بودن و همگنی، با انجام آزمون‌های پایه‌ای در سطح یک درصد مورد ارزیابی قرار گرفته و ایستگاه‌هایی که دارای این معیارها نبودند حذف گردیدند. در مجموع از تعداد ۱۲۳ ایستگاه تعداد ۱۱۵ ایستگاه برای داده‌های AMIR و تعداد ۱۰۳ ایستگاه برای داده‌های AM2R مناسب تشخیص داده شدند.

تشخیص مناطق همگن

ابتدا با انتخاب متغیرهای مختلف و با استفاده از الگوریتم k -means گروه‌بندی ایستگاه‌های مذکور انجام شد. متغیرهای مورد استفاده برای خوشه‌بندی‌های مختلف در هر ایستگاه عبارتند از: عرض جغرافیایی، طول جغرافیایی، ارتفاع ایستگاه، متوسط بارش سالانه و جهت شیب. برای تعیین بهترین تعداد خوشه‌های موجود در مشاهدات، به روش سعی و خطا عمل کرده و با در نظر گرفتن تعداد خوشه‌ها و متغیرهای مختلف مبادرت به تولید متوسط عرض سیلهوت هر یک از آن‌ها شده است. نتایج نشان داد که خوشه‌بندی در حالت‌های ۴ و ۵ خوشه با متغیرهای ارتفاع و متوسط بارش سالانه دارای متوسط عرض سیلهوت بالاتری به ترتیب به میزان ۰/۵۸ و ۰/۶۳ می‌باشد. پس این خوشه‌بندی‌ها به عنوان مبنای نواحی همگن در مرحله اول انتخاب شد. سپس بر اساس شرایط جغرافیایی ایستگاه‌ها از قبیل: الگوی توپوگرافی و جهت شیب دره‌ها اصلاح خوشه‌بندی تعدادی از ایستگاه‌ها انجام گردید و در نهایت ایستگاه‌ها به ۴ خوشه همگن تقسیم شدند. این مرحله با سعی و خطا صورت گرفت و صحت خوشه‌های اصلاح شده با استفاده از آزمون همگنی که در ادامه نتایج آن آمده، مورد تایید قرار گرفت. پراکندگی ایستگاه‌ها بر اساس خوشه‌بندی نهایی در شکل ۱ ارائه شده است.

داده‌های مشاهده‌ای، به دست می‌آید.

$$V_1 = \frac{\sum_{i=1}^N n_i (t^{(i)} - t^{(R)})^2}{\sum_{i=1}^N n_i} \quad (8)$$

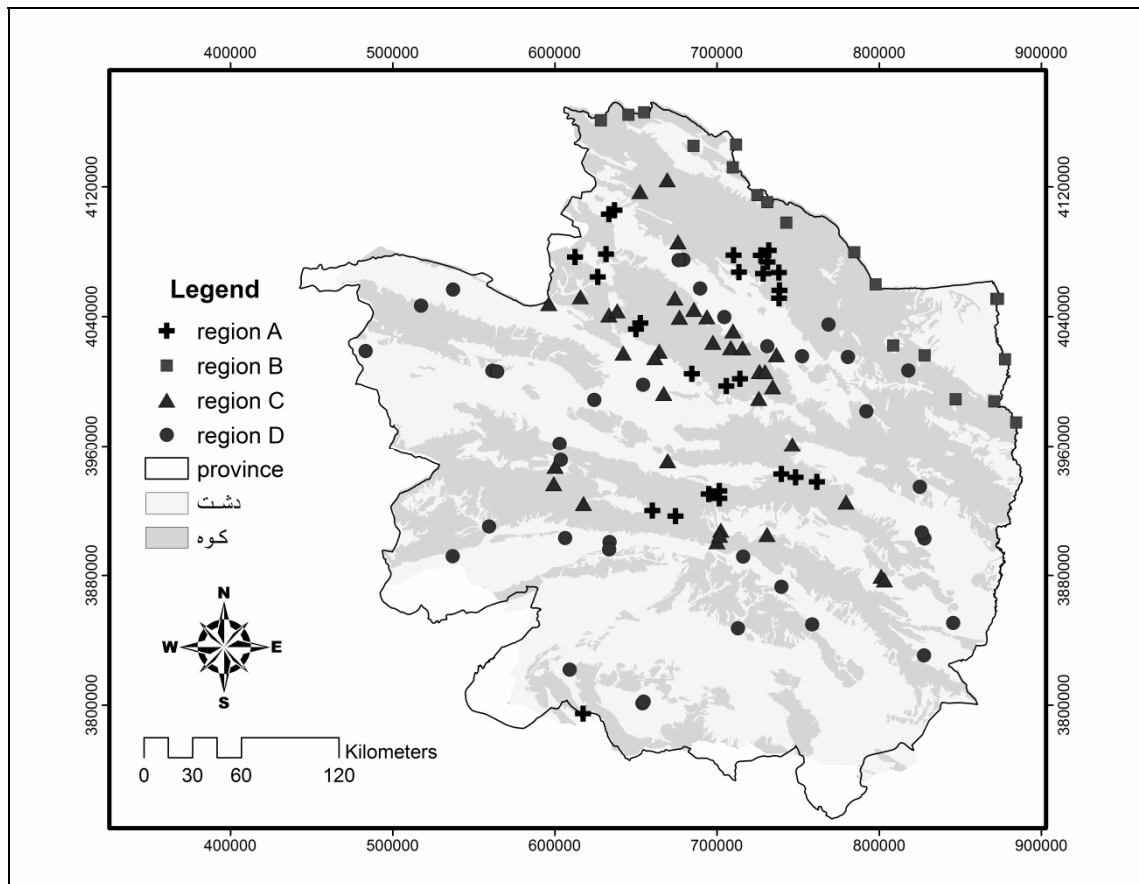
در مرحله بعد برای محاسبه پراکنش نسبت گشتاورهای خطی ناحیه‌ای اقدام به شبیه‌سازی داده‌های ناحیه‌ای معادل می‌شود. برای این منظور می‌بایست تابع توزیع مناسبی را در نظر گرفته و با استفاده از آمار واقعی ناحیه‌ای، پارامترهای تابع انتخابی در هر ناحیه تعیین گردد. برای این منظور تابع چهار پارامتری کاپا به دلیل اینکه دامنه وسیعی از (τ_3, τ_4)ها را شامل می‌شود، به عنوان یک توزیع عمومی که می‌تواند برازش توزیع‌های دو و سه پارامتری را مقایسه کند، مفید خواهد بود (۷). شبیه‌سازی داده‌های منطقه‌ای معادل با محاسبه پارامترهای تابع توزیع کاپا برای هر ناحیه، به روش مونت کارلو انجام می‌شود. فرآیند شبیه‌سازی داده در هر ایستگاه و محاسبه ضرایب گشتاورهای خطی و آماره V_1 مربوط به آن، ۱۰۰۰ بار تکرار شده است. اگر μ_v میانگین و σ_v انحراف معیار مربوط به مقادیر شبیه سازی شده V_1 ها باشد، آنگاه معیار ناهمگنی در هر ناحیه با استفاده از رابطه ۶ قابل محاسبه خواهد بود.

$$H_1 = \frac{(V_1 - \mu_v)}{\sigma_v} \quad (9)$$

اگر H_1 به اندازه کافی بزرگ باشد، ناحیه ناهمگن خواهد بود. هاسکینگ و والیس در سال ۱۹۹۷ پیشنهاد کردند که یک ناحیه به طور قابل قبولی همگن است اگر $H_1 < 1$ ، و ممکن است ناحیه ناهمگن باشد اگر $1 \leq H_1 < 2$ و به عنوان ناحیه مشخصاً ناهمگن خواهد بود اگر $H_1 > 2$ باشد. در بسیاری از پژوهش‌ها وقتی مقادیر $0 < H_1 < 1$ برآورد می‌شوند، ناحیه همگن در نظر گرفته می‌شود (۴ و ۱۵). هاسکینگ و والیس همچنین توصیه کردند که اعتبار H_1 بستگی به انتخاب درست ایستگاه‌ها در خوشه‌ها بر مبنای مشخصات فیزیکی و منطقه جغرافیایی ایستگاه دارد (۷).

آزمون نیکویی برازش و ارزیابی صحت خوشه‌بندی

در این پژوهش از نمودار نسبت گشتاورهای خطی و آزمون نیکویی برازش کلموگروف-اسمیرنوف (k-s) برای انتخاب تابع توزیع مناسب استفاده شده است. لازم به ذکر است که نمودار نسبت گشتاورهای خطی به عنوان یک ابزار مفید برای تعیین تابع توزیع مناسب بر داده‌های منطقه‌ای پیشنهاد شده است (۷) و بسیاری از پژوهشگران این نمودار را به عنوان قسمتی از فرآیند انتخاب توزیع آماری در داده‌های منطقه‌ای به کار برده‌اند. همچنین میانگین خطای مطلق (MAE) برای ارزیابی دقت برآورد تابع توزیع منتخب در پیش‌بینی مقادیر بارش حدی، استفاده شده است (رابطه ۷).



شکل ۱- خوشه‌بندی اصلاح شده ایستگاه‌ها در استان خراسان رضوی با تعداد ۴ خوشه

در جدول ۱ قابل مشاهده است. بر اساس این جدول مقادیر H_1 صرف نظر از علامت آن کمتر از یک است، که نشان از همگنی ایستگاه‌های هر ناحیه دارد.

برازش توابع توزیع مختلف و آزمون نیکویی برازش

پس از بررسی همگنی نواحی مورد مطالعه می‌بایست توابع توزیعی که بهترین برازش را با داده‌های هر یک از نواحی مذکور دارند، تعیین گردد. ابتدا پارامترهای توابع توزیع سه پارامتری لوجستیک تعمیم‌یافته (GLO)، مقادیر حدی تعمیم‌یافته (GEV)، پیرسون نوع سه (PE3)، لوگ نرمال سه پارامتری (LN3) در نواحی چهارگانه برای داده‌های AM1R و AM2R محاسبه شده است. با داشتن این پارامترها و به کمک روابط توزیع تجمعی توابع احتمالاتی، مقادیر احتمال تجمعی آمار بی‌بعد شده هر یک از نواحی مورد مطالعه محاسبه شد. همچنین با استفاده از رابطه تجربی مربوط به هر توزیع، مقادیر احتمال تجمعی آمار بی‌بعد به روش تجربی نیز برآورد شده و مقدار D_{max} (آماره آزمون k-S) محاسبه گردیده است. نتایج حاصل از انجام این آزمون و مقادیر بحرانی D_{α} برای آزمون نیکویی برازش k-S در جدول ۲ ارائه شده است.

آزمون ناهم‌نوایی و همگنی خوشه‌ها

به منظور تعیین ایستگاه‌های ناهم‌نوا با سایر ایستگاه‌ها، ابتدا ضرایب گشتاور خطی (T_1 ، T_2 و T_3) ایستگاه‌های واقع در هر ناحیه محاسبه شده است. جدول ۱ میانگین وزنی ضرایب گشتاورهای خطی ایستگاه‌های هر ناحیه را برای داده‌های AM1R و AM2R نشان می‌دهد. پس از محاسبه ضرایب گشتاور خطی در هر ایستگاه، آماره ناهم‌نوایی D_i برای ایستگاه‌های واقع در هر ناحیه محاسبه شده است. بر اساس این آماره در داده‌های AM1R هیچ ایستگاهی حذف نمی‌شود. اما تعداد ۳ ایستگاه برای داده‌های AM2R ناهم‌نوا تشخیص داده شده که از محاسبات حذف شدند.

برای اجرای آزمون ناهمگنی، ابتدا گشتاورهای وزن دار احتمالی آمار بی‌بعد هر یک از نواحی محاسبه و سپس اقدام به برآورد مقادیر چهار پارامتر توزیع کاپا در نواحی ۴ گانه شده است. با داشتن تابع توزیع کاپا و پارامترهای آن، سری داده‌ی ناحیه‌ای شبیه‌سازی شده تولید و پارامترهای آزمون ناهمگنی شامل σ_v ، V_{obs} و μ_v محاسبه شده‌اند. با در اختیار داشتن این پارامترها و با استفاده از رابطه ۶ آماره H_1 آزمون ناهمگنی محاسبه شده که مقادیر این آماره

جدول ۱- نتایج آزمون‌های ناهمنوایی و ناهمگنی در خوشه‌بندی نهایی

متغیر	ناحیه مورد بررسی برای آزمون همگنی	تعداد ایستگاه‌ها	تعداد ایستگاه‌های حذف شده در آزمون ناهمنوایی	τ	τ_3	τ_4	اندازه ناهمگنی H_1
AM1R	A	۲۶	ندارد	۰/۱۷	۰/۱۳	۰/۱۴	-۰/۸۶
	B	۱۹	ندارد	۰/۱۹	۰/۱۶	۰/۱۶	-۰/۷۹
	C	۳۵	ندارد	۰/۱۹	۰/۱۵	۰/۱۶	-۰/۰۶
	D	۳۵	ندارد	۰/۲	۰/۱۸	۰/۱۵	-۰/۷۸
AM2R	A	۲۵	۱	۰/۲۰	۰/۱۷	۰/۱۳	-۰/۹۱
	B	۱۷	ندارد	۰/۱۹	۰/۱۸	۰/۱۴	-۰/۸۷
	C	۳۲	۲	۰/۲	۰/۱۷	۰/۱۵	-۰/۶۷
	D	۲۹	ندارد	۰/۲	۰/۱۶	۰/۱۴	-۰/۹۷

جدول ۲- مقادیر D_α (سطح معنی‌دار بودن $\alpha=0.05$) و D_{max} برآورد شده آزمون k-s در نواحی ۴ گانه

متغیر	ناحیه	حجم نمونه	D_α	لوگ نرمال سه پارامتری	پیرسون نوع III	مقادیر حدی تعمیم یافته	لوجستیک تعمیم یافته
AM1R	A	۶۸۴	۰/۰۳۴	۰/۰۲۲	۰/۰۲۶	۰/۰۲۱	۰/۰۳۳
	B	۵۴۸	۰/۰۳۸	۰/۰۳۳	۰/۰۳۹	۰/۰۳۰	۰/۰۲۸
	C	۱۰۰۳	۰/۰۲۸	۰/۰۱۹	۰/۰۲۴	۰/۰۲	۰/۰۲
	D	۱۰۰۵	۰/۰۲۸	۰/۰۲۳	۰/۰۲۵	۰/۰۲۲	۰/۰۲۶
AM2R	A	۶۹۶	۰/۰۳۴	۰/۰۲۴	۰/۰۲۴	۰/۰۲۴	۰/۰۳۳
	B	۵۰۲	۰/۰۴۰	۰/۰۲۰	۰/۰۲۶	۰/۰۱۸	۰/۰۳۳
	C	۸۲۶	۰/۰۳۱	۰/۰۲۸	۰/۰۳۵	۰/۰۲۵	۰/۰۳۳
	D	۸۱۵	۰/۰۳۱	۰/۰۱۵	۰/۰۲۱	۰/۰۱۶	۰/۰۲۴

وقوع است که بر اساس رابطه $F = 1 - 1/T$ برای دوره بازگشت T قابل محاسبه می‌باشد.

بر پایه نتایج به دست آمده، مقادیر بی‌بعد AM1R و AM2R در دوره بازگشت‌های پایین در نواحی مختلف نزدیک به هم هستند و با افزایش دوره بازگشت روند صعودی دارند. روند صعودی داده‌های AM1R با افزایش دوره بازگشت در نواحی چهارگانه به یک میزان نبوده، به طوریکه روند افزایشی چندک بی‌بعد AM1R با افزایش دوره بازگشت در ناحیه D بیشترین مقدار و در ناحیه A کمترین مقدار را به خود اختصاص داده است. مقایسه ضریب تغییرات گشتاور خطی (τ) در این نواحی که در جدول ۱ آمده نیز موید این مطلب است. بر اساس این جدول ضریب تغییرات نواحی D و A به ترتیب بزرگترین و کوچکترین مقادیر را در نواحی چهارگانه داده‌های AM1R به خود اختصاص داده‌اند. اما چندک بی‌بعد AM2R در هر ۴ ناحیه با افزایش دوره بازگشت تقریباً به یک میزان افزایش می‌یابد. شکل ۲ روند افزایش مقادیر بی‌بعد AM1R و AM2R را با افزایش دوره بازگشت برای نواحی چهارگانه نشان می‌دهد.

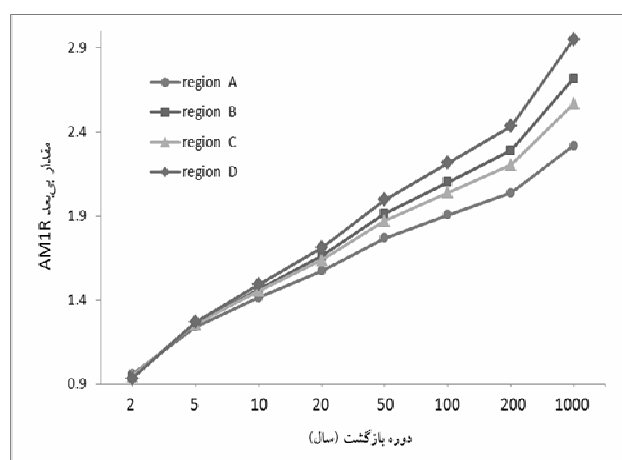
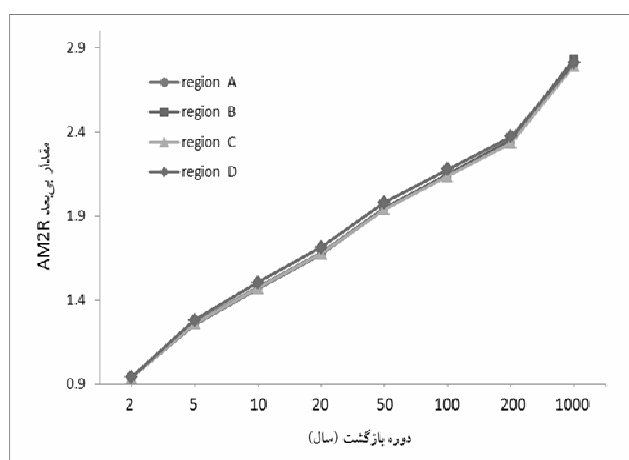
با مقایسه D_{max} برآورد شده برای توزیع‌های مختلف با مقدار بحرانی D_α مشخص می‌شود که در نواحی چهارگانه داده‌های AM1R و AM2R غیر از ۳ مورد، D_{max} تابع توزیع GEV در بین توابع توزیع کوچکترین مقدار را دارد. و در این ۳ مورد هم D_{max} تابع توزیع GEV دارای اختلاف بسیار اندکی با کوچکترین D_{max} است. از آنجایی که در مطالعات هیدرولوژیکی برای جلوگیری از توابع توزیع متعدد در برآوردها استفاده از یک نوع تابع توزیع برای تمامی نواحی مورد مطالعه، ترجیح دارد، در نتیجه می‌توان تابع توزیع GEV را به عنوان تابع منتخب در تمام نواحی برای آمار AM1R و AM2R در نظر گرفت. البته صحت این انتخاب بر اساس نمودار نسبت‌های گشتاور خطی نیز تایید شده است.

برآورد AM1R و AM2R در دوره بازگشت‌های مختلف

جدول ۳ تابع چندک بدون بعد GEV را برای نواحی چهارگانه داده‌های AM1R و AM2R نشان می‌دهد. F در این توابع، احتمال

جدول ۳- تابع چندک بدون بعد GEV برای نواحی چهارگانه آمار AM1R و AM2R

متغیر	ناحیه	تابع چندک بدون بعد ناحیه‌ای	MAE (mm)
AM1R	A	$x(F) = 0.86 + 3.9 [1 - (-\ln F)^{0.068}]$	۰/۰۳۸۶
	B	$x(F) = 0.84 + 32 [1 - (-\ln F)^{0.0087}]$	۰/۰۱۳۷
	C	$x(F) = 0.85 + 8 [1 - (-\ln F)^{0.0349}]$	۰/۰۲۳۷
	D	$x(F) = 0.83 - 17.6 [1 - (-\ln F)^{-0.016}]$	۰/۰۱۲۹
AM2R	A	$x(F) = 0.84 + 118 [1 - (-\ln F)^{0.024}]$	۰/۰۰۹۹
	B	$x(F) = 0.84 - 22 [1 - (-\ln F)^{-0.012}]$	۰/۰۱۰۵
	C	$x(F) = 0.84 - 195 [1 - (-\ln F)^{-0.0014}]$	۰/۰۰۹۹
	D	$x(F) = 0.83 - 15.7 [1 - (-\ln F)^{0.0196}]$	۰/۰۱۶۵



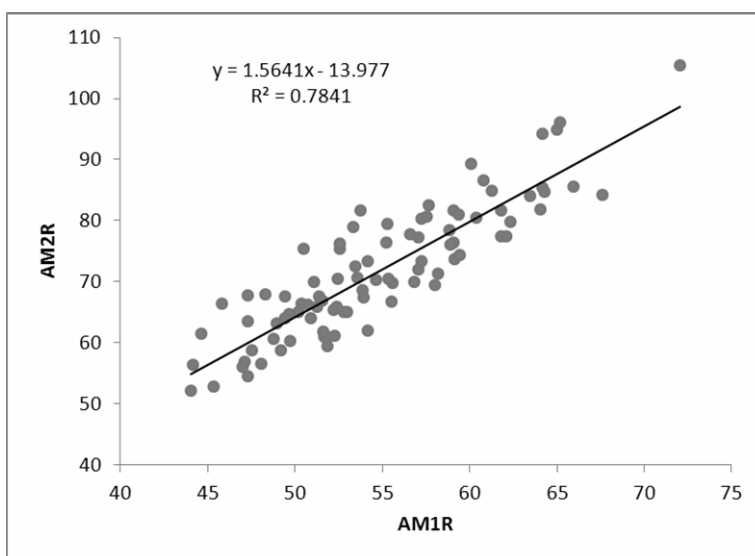
شکل ۲- روند مقادیر بی بعد AM2R و AM1R به ازای دوره بازگشت‌های مختلف در نواحی چهارگانه

قبلاً نیز اشاره شد، مطابق رابطه ۱ با ضرب متوسط AM1R و AM2R هر ایستگاه در مقادیر بی‌بعد ناحیه‌ای AM1R و AM2R، می‌توان مقادیر AM1R و AM2R را با دوره بازگشت‌های مختلف در محل مورد نظر محاسبه کرد.

همچنین با داشتن مقادیر AM1R و AM2R در دوره بازگشت‌های مختلف می‌توان در هر ایستگاه روابط همبستگی بین AM1R و AM2R را به ازای دوره بازگشت‌های مختلف استخراج نمود. برای نمونه رابطه همبستگی خطی مقادیر AM1R و AM2R برای ۹۰ ایستگاه خراسان رضوی در دوره بازگشت ۱۰۰ ساله در شکل ۳ ارائه شده است. تحلیل نتایج این رگرسیون در محیط SPSS صورت گرفته است. جدول ۴، جدول آنوا برای رابطه مذکور و جدول ۵، معنی‌داری ضرایب رابطه همبستگی را نشان می‌دهند. بر اساس این جداول، رابطه رگرسیونی و ضرایب آن در سطح ۵ درصد معنی‌دار هستند.

بر اساس این شکل به خوبی دیده می‌شود که مقدار بی‌بعد AM2R در هر یک از دوره بازگشت‌ها برای هر ۴ ناحیه بر هم منطبق هستند. یعنی می‌توان گفت مقدار بی‌بعد AM2R وابسته به نواحی همگن نیست و در همه نواحی تقریباً یک مقدار در نظر گرفته می‌شود. برای بررسی صحت این فرضیه آزمون ناهمگنی برای داده‌های AM2R در کل ایستگاه‌های استان اجرا گردید. نتایج نشان داد که کل ایستگاه‌های استان از نظر داده‌های AM2R همگن هستند ($H_1 = 0/98$).

می‌توان با در اختیار داشتن روابط چندک بی‌بعد ناحیه‌ای، به مقایسه مقادیر AM1R و AM2R بی‌بعد واقعی و برآورد شده در هر یک از نواحی پرداخت. بر این اساس مقدار MAE برای دوره بازگشت‌های کمتر از ۲۰۰ سال در هر ناحیه در جدول ۱ ارائه شده است. مقدار MAE در همه نواحی همگن بسیار کوچک است و نشان می‌دهد که روابط به دست آمده به منظور برآورد AM1R و AM2R، از نزدیکی نسبتاً خوبی با مقادیر واقعی برخوردار هستند. همانطور که



شکل ۳- رابطه همبستگی مقادیر AM1R و AM2R در دوره بازگشت ۱۰۰ ساله برای ۹۰ ایستگاه استان

جدول ۴- جدول آنوا برای رابطه همبستگی AM1R و AM2R در دوره بازگشت ۱۰۰ ساله برای ۹۰ ایستگاه استان

ANOVA ^b	میانگین مربعات	درجه آزادی	میانگین مربعات	مقدار F	سطح معنی داری
رگرسیون	۷۶۳۰	۱	۷۶۳۰	۳۱۷	۰/۰۰۰ ^a
باقیمانده	۲۱۲۱	۸۸	۲۴		
کل	۹۷۵۱	۸۹			

a. Predictors: (Constant), AM1R

b. Dependent Variable: AM2R

جدول ۵- جدول ضرایب رابطه همبستگی AM1R و AM2R در دوره بازگشت ۱۰۰ ساله برای ۹۰ ایستگاه استان

Coefficients ^a	ضرایب غیر استاندارد شده		t	سطح معنی داری
	B	Std. Error		
(Constant)	-۱۳/۷۸۲	۴/۸۳۲	-۲/۸۵۲	۰/۰۰۵
AM1R	۱/۵۶۱	۰/۰۸۸	۱۷/۷۹۳	۰/۰۰۰

a. Dependent Variable: AM2R

نتیجه گیری

رشد منطقه‌ای برای داده‌های AM2R، نشان دهنده عدم وابستگی این داده‌ها به نواحی همگن بوده و آزمون ناهمگنی نیز این فرضیه را تایید می‌کند. علت این امر را باید در الگوهای سینوپتیکی به وجود آورنده بارش‌های حداکثر دو روزه جست. معمولاً بارش‌هایی که در بازه زمانی دو روزه رخ می‌دهند ناشی از نفوذ توده‌های هوا و بارش‌های جبهه‌ای هستند. چون گستره این نوع بارش‌ها وسیع است و معمولاً سطح وسیعی از استان را می‌پوشانند، لذا شرایط افزایش دهنده AM2R در هر ناحیه یکسان است. ولی این بدان معنی نیست که مقدار واقعی AM2R به ازای یک دوره بازگشت معین در این نواحی یکسان است. بلکه از آنجایی که با ضرب مقادیر بی‌بعد AM2R در متوسط AM2R هر ایستگاه مقدار AM2R در یک دوره بازگشت معین

نتایج این پژوهش نشان می‌دهد که می‌توان ایستگاه‌های موجود در استان خراسان رضوی را به چهار ناحیه همگن از نظر بارش‌های حداکثر یک روزه تقسیم نمود. همچنین با انجام تحلیل فراوانی منطقه‌ای و آزمون نیکویی برازش، مشخص شد که تابع توزیع GEV مناسب‌ترین تابع جهت برازش مقادیر AM1R و AM2R می‌باشد. سایر پژوهشگران نیز با انجام روش‌های مشابه برای بارش‌های حداکثر سالانه این تابع توزیع را پیشنهاد نموده‌اند (۱۲ و ۱۵). روابط به دست آمده به منظور برآورد AM1R و AM2R، در تمامی نواحی همگن دارای دقت بالایی است. به طوری که بزرگترین مقدار MAE معادل ۰/۰۳۸۶ میلی‌متر، موید این مطلب می‌باشد. همچنین منحنی

مقادیر AM1R و AM2R مربوط به دشت‌های سبزوار، نیشابور، مشهد، درگز و دشت‌های جنوبی استان می‌باشد. با توجه به اینکه در این پژوهش از سری حداکثر سالانه بارش استفاده شده است، پیشنهاد می‌شود که در مطالعات بعدی از سری‌های مقادیر جزئی استفاده شده تا فراوانی وقوع بارش‌های بزرگتر از یک مقدار مشخص نیز بدست آید.

تعیین می‌شود، می‌توان گفت موقعیت مکانی و شرایط ویژه جغرافیایی هر ایستگاه موجب دریافت‌های متفاوت از بارش‌های با گستره وسیع می‌شود. از نقطه نظر مکانی نیز بزرگترین مقادیر AM1R و AM2R با دوره بازگشت ۱۰۰ ساله در ارتفاعات مرکزی استان به‌خصوص در ارتفاعات تربت حیدریه، جنوب شرقی رشته‌کوه‌های بینالود و ناحیه کوچکی در شرق و شمال استان دیده می‌شود. همچنین کمترین

منابع

- ۱- رضایی پژند ح. ۱۳۸۰. کاربرد آمار و احتمال در منابع آب. چاپ اول. انتشارات سخن گستر، مشهد.
- ۲- سازمان هواشناسی کشور. ۱۳۸۹. داده‌های بارش حداکثر سالانه ایستگاه‌های سینوپتیک و کليماتولوژی.
- ۳- شرکت آب منطقه‌ای خراسان رضوی. ۱۳۹۰. مطالعات پایه منابع آب، داده‌های ایستگاه‌های باران‌سنجی و تبخیرسنجی. <http://www.khrw.ir/Default.aspx?alias=www.khrw.ir/motaleat>
- 3- Abolverdi J., and Khalili D. 2010. Development of regional rainfall annual maxima for southwestern Iran by L-moments. *Water Resour Manage*, 24:2501-2526.
- 4- Durrans S.R., and Kirby J.T. 2004. Regionalization of extreme precipitation estimates for the Alabama rainfall atlas. *Journal of Hydrology*, 295: 101-107.
- 5- Greenwood J.A., Landwehr J.M., Matalas N.C., and Wallis J.R. 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15 (5):1049-1054.
- 6- Hosking J.R., and Wallis J.R. 1997. *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press, Cambridge, UK.
- 7- Ngongondo C.S., Xu C.-Yu., Tallaksen L.M., Alemaw B., and Chirwa T. 2011. Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches. *Journal of Stoch Environ Res Risk Assess*, 25:939-955.
- 8- Rao A.R., and Srinivas V.V. 2006. Regionalization of watersheds by Hybrid-Cluster analysis. *Journal of Hydrology*. 318: 37-56.
- 9- Rousseeuw P.J. 1987. Silhouette: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, 20: 53-65.
- 10- Schaefer M.G. 1990. Regional analysis of precipitation annual maximum in Washington State. *Water Resources Research*, 26:119-131.
- 11- Soon H.L., and Sung J.M. 2003. Frequency analysis of extreme rainfall using L-moment. *Journal of Irrigation And Drainage*, 52: 219-230.
- 12- Wallis J.R., Matalas N.C., and Slack J.R. 1974. Just a moment!, *Water Resources Research*, 10(2): 211-219.
- 13- Wilks D.S. 2006. *Statistical methods in the atmospheric sciences*. 2nd ed. Academic Press is an imprint of Elsevier.
- 14- Yang T., Shao Q., Hao Z.C., Chen X., Zhang Z., Xu C.Y., and Sun L. 2010. Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. *Journal of Hydrology*, 380:386-405.

Regional Frequency Analysis of Annual Maximum 1-day And 2-day Rainfalls Using Clustering And L-moments, Case study: Khorasan Razavi Province

M. Shahedi¹- S.H. Sanaeinejad^{2*} - B. Ghahreman³

Received:18-02-2012

Accepted:05-08-2012

Abstract

The purpose of this study is regional frequency analysis of Annual Maximum 1-day Rainfall (AM1R) and Annual Maximum 2-day Rainfall (AM2R) using L-moments theory in Khorasan Razavi Province. For this purpose, the basic statistical tests include: homogeneity, independency and outlier data for AM1R and AM2R were surveyed in 123 rainfall stations. The province was divided into four regions based on cluster analysis, topography and spatial pattern of precipitation. Hydrology homogeneity was also controlled by running heterogeneity test for each region. generalized extreme value (GEV), generalized logistic (GLO), Pearson type III (PE3) and Log Normal type III (LN3) probability distributions were estimated for every region. To select the appropriate distribution of AM1R and AM2R data, the fitness was judged using an L-moment ratio diagram and the Kolmogorov-Smirnov test and GEV distribution select. The regionally quantile estimateions for GEV distribution were also calculated for AM1R and AM2R data. In all of the Homogeneous regions, the estimated values of AM1R and AM2R from the obtained relations are close enough to the real data of return periods less than 200 years (The largest MAE was 0.0386). The average absolute error between the estimated and the observation values in each region is favorable, showing a high accuracy of the estimation.

Keywords: Regional frequency analysis, L-moments, Annual Maximum 1-day And 2-day Rainfalls, Index rainfall method, Cluster analysis

1,2,3- MSc Student, Associate Professor and Professor, Department of Water Engeenring, Faculty of Agriculture, Ferdowsi University of Mashhad, Respectively
(*- Corresponding Author Email: sanaein@gmail.com)