مجله علمی – پژوهشی
عمران مدرس
دوره چهاردهم، شماره چهار، زمستان ۱۳۹۳

تحلیل فراوانی چندمتغیره سیلاب با استفاده از تابع مفصل و توزیع‌های حاشیه‌ای پارامتری و ناپارامتری

محمدصادق عباسیان[۱]، سهیل جلالی[۲*]، سیدسعید موسوی‌ندوشنی[۳]

۱- دانشجوی کارشناسی ارشد مهندسی عمران– آب، دانشگاه صنعتی شریف

۲- دانشجوی کارشناسی ارشد مهندسی عمران– رودخانه، دانشگاه شهید بهشتی– پردیس فنی مهندسی شهید عباسپور

۳- استادیار دانشگاه شهید بهشتی– پردیس فنی مهندسی شهید عباسپور

*Jalali_Soheil@ace.sbu.ac.ir

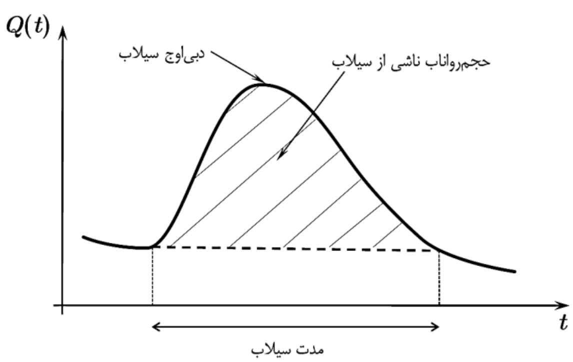
تاریخ دریافت: [۱۳۹۲/٤/۱۳] تاریخ پذیرش: [۱۳۹۳/۲/۱۱]

چکیده– از آن‌جایی که سیلاب پدیده‌ای چندمتغیره است، اتخاذ رویکردهای چند متغیره در تحلیل آن ضروری است. روش سنتی انجام تحلیل‌های چندمتغیره، استفاده از توزیع‌های چندمتغیره کلاسیک با توابع حاشیه‌ای پارامتری است. این در حالی است که هم استفاده از توزیع‌های چندمتغیره کلاسیک و هم استفاده از توزیع‌های پارامتری با محدودیت‌های جدی مواجه است. از مهم‌ترین محدودیت‌های توزیع‌های چندمتغیره کلاسیک می‌توان به لزوم مشخص بودن توابع توزیع حاشیه‌ای و پارامترهای آنها و یکسان بودن توابع توزیع حاشیه‌ای اشاره کرد. همچنین در استفاده از توزیع‌های پارامتری برای متغیرهای حاشیه‌ای از یک توزیع پیش‌فرض برای تفسیر توزیع داده‌ها استفاده می‌شود، در حالی که ممکن است توزیع فرض شده به‌خوبی توزیع واقعی داده‌ها را توصیف نکند. در مقاله حاضر تحلیل‌های توأم متغیرهای سیلاب با استفاده از توابع مفصل که محدودیت توزیع‌های چندمتغیره کلاسیک را ندارند انجام گرفته، به‌گونه‌ای که توزیع‌های حاشیه‌ای از میان توزیع‌های پارامتری و توزیع‌های ناپارامتری که بی‌نیاز از برآورد تعدادی پارامتر است برگزیده شده‌اند.

واژگان کلیدی– تحلیل فراوانی سیلاب، دوره بازگشت توأم، تابع مفصل، توزیع ناپارامتری

۱– مقدمه

سیلاب یکی از مهم‌ترین بلایای طبیعی است که سالانه خسارات مالی و جانی بسیاری را در نقاط مختلف دنیا ایجاد می‌کند که همین موضوع، پیش‌بینی و کنترل آن را ضروری می‌کند. سیلاب پدیده‌ای چند متغیره و پیچیده بوده و در نتیجه دارای ماهیتی تصادفی است. پس تحلیل آن مستلزم استفاده از علم آمار و احتمالات است. همان‌گونه که در شکل (۱) نشان داده شده است، سه متغیر مهم تشکیل دهنده مشخصات سیلاب دبی‌اوج، حجم‌رواناب و مدت سیلاب است.



شکل (۱) شماتیک هیدروگراف سیل و متغیرهای مهم آن

بیشتر در تحلیل‌های سیلاب، تنها متغیر دبی‌اوج در نظر گرفته می‌شود؛ اما بسیاری از پژوهشگران هشدار داده‌اند که رویکردهای تک‌متغیره، یعنی رویکردهایی که در آنها آثار یک متغیر به صورت جداگانه در نظر گرفته می‌شود، منجر به دست بالا گرفتن یا دست پایین گرفتن نتایج می‌شود. یکی از روش‌های سنتی انجام تحلیل‌های چندمتغیره، استفاده از توابع توزیع چندمتغیره کلاسیک است که البته بهره‌گیری از این توابع با محدودیت‌های قابل ملاحظه‌ای مواجه است. از مهم‌ترین این محدودیت‌ها می‌توان به لزوم مشخص بودن توزیع متغیرهای حاشیه‌ای و یکسان بودن آنها اشاره کرد. بنابراین در صورت مواجهه با محدودیت‌های فوق، باید روشی را برای انجام تحلیل‌های چندمتغیره جایگزین کرد که از مناسب‌ترین روش‌های در دسترس که هم بر محدودیت‌های توزیع‌های کلاسیک فایق آمده و هم امکانات جدید را در اختیار می‌دهد، استفاده از دسته خاصی از توابع چندمتغیره به نام تابع مفصل[1] است. در مورد مهم‌ترین مزایای توابع مفصل می‌توان به این موارد اشاره کرد: ۱- برای انتخاب یک تابع مفصل مناسب برای متغیرها، لزومی به مشخص بودن توزیع‌های حاشیه‌ای و پارامترهای آنها نیست؛ ۲- هیچ الزامی در یکسان بودن توزیع‌های حاشیه‌ای وجود ندارد؛ ۳- توزیع‌های حاشیه‌ای می‌توانند پارامتری یا ناپارامتری[2] است؛ ٤- توابع مفصل از ضرایب همبستگی ناپارامتری بهره می‌گیرند؛ ٥- توزیع‌های شرطی به راحتی از روی توابع مفصل قابل محاسبه است. اگر چه محاسبات مربوط به توابع مفصل پیچیده‌تر از محاسبات توابع چندمتغیره کلاسیک است، اما در دسترس بودن نرم‌افزارهای ریاضی برای انجام محاسبات، این مشکل را برطرف کرده است. البته در مورد استفاده از توابع مفصل باید توجه نمود که از این توابع باید هوشمندانه استفاده نمود و نباید برای مدل‌سازی هر نوع همبستگی از آنها استفاده کرد، به این معنی که برای به‌کارگیری توابع مفصل در مسائل جدید، ابتدا باید نتایج تابع مفصل را با نتایج روش‌های معتبر قبلی مقایسه کرده و از صحت عملکرد تابع مفصل در مورد آن مسئله خاص اطمینان حاصل کرد. به عنوان نمونه، Shiau (2006) برای اولین بار توابع مفصل را بر متغیرهای سختی و مدت خشکسالی برازش داد. وی نشان داد که احتمال‌های توأم خشکسالی‌های مشاهده شده و احتمال‌های توأم به دست آمده از توابع مفصل هماهنگی مناسبی با یکدیگر داشته و بنابراین نتیجه گرفت که برای تحلیل خشکسالی‌ها می‌توان از توابع مفصل بهره گرفت.

از طرف دیگر در فرایند تحلیل‌های چندمتغیره سیلاب که نیاز به برون‌یابی برای مقادیر متغیرها داریم، توزیع متغیرهای حاشیه‌ای باید معلوم باشد. به این منظور بیشتر از توزیع‌های پارامتری برای توصیف ساختار هر یک از متغیرها استفاده می‌شود. اما در استفاده از توزیع‌های پارامتری سؤالاتی به این شرح قابل طرح است: در صورتی که نتوان هیچ‌کدام از توابع پارامتری موجود را به خوبی بر داده‌ها برازش داد چه راهی را باید در نظر گرفت؟ آیا روشی وجود دارد که از گزینش تابع پیش‌فرض صرف نظر کرده و در عین حال تفسیر مناسبی از داده‌ها در اختیار قرار دهد؟ سؤالاتی از این دست را می‌توان با ورود به مبحث برآورد ناپارامتری توابع پاسخ داد. به طور کلی روش‌های ناپارامتری برآورد یک تابع سعی دارند با استفاده از داده‌های مشاهده شده، تقریبی موضعی از تابع هدف ارائه نمایند، به طوری که ایده استفاده از توزیع‌های ناپارامتری، دوری کردن از فرضیات محدودکننده در مورد تابع چگالی احتمال است و اینکه دیگر نیازی نباشد تا برای تخمین تابع چگالی احتمال خود را به برآورد تعدادی پارامتر برای داده‌های مورد استفاده محدود کنیم. با این وجود روش ناپارامتری به دلیل اینکه تابع داده‌های مشاهده شده است برای برون‌یابی‌های طولانی محدودیت خاص خود را دارد [۱].

سابقه استفاده از تابع مفصل در هیدرولوژی به یک دهه

1- Copula
2- Nonparametric

اخیر بازمی‌گردد. کاربرد مدل‌سازی با توابع مفصل در هیدرولوژی و محیط‌زیست در ابتدا به وسیله‌ی Salvadori and De Michele (2003) با مدل‌سازی همبستگی متغیرهای شدت و مدت بارش به وسیله‌ی توابع مفصل مطرح شد. سپس استفاده از این توابع در هیدرولوژی توسط Favre et al. (2004) ادامه داده شد و آنها درباره مزیت‌های تحلیل فراوانی چندمتغیره در هیدرولوژی با استفاده از توابع مفصل بحث کردند. از آن پس استفاده از توابع مفصل در هیدرولوژی بیشتر روی مخاطرات طبیعی متمرکز شده است که به طور ویژه می‌توان به تحلیل فراوانی سیلاب (Genest et al., 2007 ؛Zhang and Singh, 2006 ؛Rénard and Lang, 2007 ؛Wang et al., 2009 ؛Karmakar and Simonovic, 2009 ؛Chebana and Ouarda, 2009)، طراحی سیستم‌های کنترل سیلاب، محاسبه خسارت سیلاب و تحلیل فراوانی خشکسالی اشاره کرد. غیر از این موارد، مدل‌سازی همبستگی متغیرهای وابسته به وسیله توابع مفصل به زمینه‌های دیگر هم رسوخ کرده است؛ برای نمونه Bárdossy (2006) قابل استفاده بودن مدل‌سازی به وسیله‌ی توابع مفصل در استخراج مشخصات کیفیت آب زیرزمینی را مورد بحث قرار داد و Chowdhary and Singh (2008) از توابع مفصل برای تحلیل فراوانی بارش استفاده کردند و همچنین Samaniego et al. (2010) از توابع مفصل برای پیش‌بینی جریان در حوضه‌های بدون ایستگاه استفاده کردند.

روش‌های ناپارامتری نیز متعدد در هیدرولوژی و منابع آب مورد استفاده قرار گرفته‌اند. برای اولین بار Wolfowitz (1942) روش ناپارامتری را مطرح کرد. Sharma et al. (1997) با استفاده از برآورد ناپارامتری توابع چگالی احتمال، روشی برای شبیه‌سازی دبی جریان ارائه کردند. Adamowski (1985) با استفاده از برآورد ناپارامتری تابع چگالی احتمال سیلاب‌های سالانه، روشی ناپارامتری برای تحلیل فراوانی سیلاب ارائه کرد.

به عنوان استفاده از مزیت‌های توابع مفصل و توزیع‌های ناپارامتری در تحلیل چندمتغیره سیلاب Karmakar and Simonovic (2009) تحلیل‌های دبی‌اوج–حجم‌رواناب، دبی‌اوج–مدت و حجم‌رواناب–مدت را با استفاده از توابع مفصل و ترکیبی از توزیع‌های پارامتری و ناپارامتری انجام داده و احتمالات توأم و شرطی را به‌دست آوردند.

هدف مقاله حاضر، انجام تحلیل فراوانی توأم سیلاب با استفاده از توابع مفصل با توزیع‌های حاشیه‌ای پارامتری و ناپارامتری است. به این منظور، ابتدا توابع توزیع پارامتری و ناپارامتری بر هر یک از متغیرهای دبی‌اوج، حجم رواناب و مدت سیلاب برازش داده شده و مناسب‌ترین تابع توزیع برای هر یک از متغیرها انتخاب شده است. سپس سه تابع مفصل ارشمیدسی بر زوج‌متغیرهای دبی‌اوج–حجم رواناب، دبی‌اوج–مدت و حجم‌رواناب–مدت برازش داده شده و مناسب‌ترین تابع مفصل برای هر یک از زوج‌متغیرها برگزیده شده است. در پایان پس از محاسبه دوره بازگشت‌های توأم هر یک از زوج متغیرها، نتایج به دست آمده به وسیله‌ی منحنی‌ها و رویه‌های دوره بازگشت‌های توأم به نمایش درآمده است.

۲- مواد و روش‌ها

۲-۱-توزیع‌های پارامتری

تعداد زیادی از توزیع‌های فراوانی با احتمال متفاوت در هیدرولوژی به‌کار برده می‌شوند. توزیع‌های پارامتری استفاده شده در این مقاله عبارتند از هشت توزیع کوشی، نمایی، گاما، گامبل، لوجستیک، لُگ‌نرمال، نرمال و ویبول.

۲-۲- توزیع ناپارامتری

استفاده از توزیع‌های ناپارامتری برای برآورد بر این اصل استوار است که برآورد تابع هدف باید به صورت موضعی انجام شود. این روش سعی در برآورد تابعی احتمالاتی از داده‌های مشاهده شده دارد تا بتوان بر اساس آن به پیش‌بینی متغیر مورد نظر در آینده پرداخت. روش مورد استفاده برای

تحلیل فراوانی در این مقاله بر مبنای برآورد هسته‌ای چگالی[3] است. در این برآورد در صورتی که n داده مشاهده شده از متغیر تصادفی X در دست داشته باشیم، برآورد هسته‌ای چگالی در نقطه x به صورت

$$f(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right) \quad (۱)$$

نوشته می‌شود که در آن x_1، x_2 و ... داده‌های مشاهده‌شده، K تابع هسته‌ای[4] و h نیز پهنای باند[5] محاسبه شده است. به طور کلی اگر $\int K(z)dz = 1$ و $\int zK(z)dz = 0$ باشد، برآورد محاسبه شده به وسیله‌ی رابطه (۱) شرایط تابع چگالی احتمال را داراست [۲۱]. مقدار چگالی احتمال در نقطه x بدین ترتیب محاسبه می‌شود که ابتدا n تابع هسته‌ای مستقل با پهنای باند h و مرکزیت x_i روی هر یک از مشاهدات قرار می‌گیرند. سپس مجموع عرض‌های هسته-های یاد شده در نقطه x محاسبه شده و با تقسیم بر nh، برآورد چگالی در x به دست می‌آید. در واقع تمامی مشاهدات مستقل در محاسبه تابع چگالی احتمال مشارکت خواهند داشت. در مورد انتخاب نوع هسته از نظر تئوری ثابت‌شده است که تابع هسته‌ای انتخاب‌شده، نقش تعیین‌کننده‌ای در عملکرد روش ندارد و انتخاب نوع آن بحرانی نیست و توابع مختلف هسته‌ای در شکل نهایی برآورد، تغییر چندانی نخواهند گذاشت [۱۸].

در این مقاله و برای برآورد توزیع ناپارامتری روی داده‌های مشاهده‌شده، هسته نرمال به عنوان هسته پیش‌فرض تابع چگالی هسته‌ای انتخاب شده است که تابع هسته‌ای آن به صورت زیر است:

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)t^2} \quad (۲)$$

تعیین مؤلفه پهنای باند که در تعیین شکل نهایی برآورد تأثیر به سزایی دارد و در پایان تغییرات چشمگیری در شکل برازش ایجاد خواهد کرد، در برازش ناپارامتری انتخابی کلیدی و بحرانی است. پهنای باند و یا تابع هموارسازی[6] مؤلفه‌ای مثبت است. زمانی که پهنای باند کوچک باشد، تغییرات در نمودار قابل تحلیل نیست و اگر این مؤلفه بیش از حد بزرگ انتخاب شود هم نمی‌توان از آن تحلیلی درست داشت. تنها زمانی که این مؤلفه مقدار بهینه خود را داشته باشد، می‌توان نتایج به دست آمده را بررسی کرد. لازم به ذکر است که این مؤلفه به عنوان یک پارامتر محسوب نمی‌شود. در مورد اندازه این مؤلفه نیز باید به این نکته اشاره کرد که کوچکی و یا بزرگی آن نسبی است و رابطه مستقیمی با اندازه داده‌هایی دارد که در تحلیل فراوانی مورد استفاده قرار می‌گیرند.

به طور کلی روش‌های انتخاب پهنای باند بر مبنای مقایسه چگالی‌های واقعی و برآورد شده است و بر این اساس معیاری از دقت برآورد به دست می‌دهند. در این مقاله محاسبات تعیین پهنای باند بهینه از طریق فرمول

$$A = \min\left\{s, \frac{Q_3 - Q_1}{1.349}\right\}, h = 0.9An^{\frac{-1}{5}} \quad (۳)$$

انجام شده است که در آن Q_1 و Q_3، چارک اول و سوم داده‌ها، n تعداد کل داده و s، انحراف معیار داده‌های دیده شده است [۱۷ و ۲۱].

تابع تجمعی احتمال روش ناپارامتری برآورد چگالی هسته-ای نیز به شرح

$$F(x) = \frac{1}{n}\sum_{j=1}^{n} K_I\left(\frac{x - x_j}{h}\right) \quad (۴)$$

است که در آن

$$K_I(u) = \int_{-\infty}^{u} K(w)dw \quad (۵)$$

است. حال برای محاسبه متغیر با دوره بازگشت مورد نظر (x_T)، ابتدا تابع ناپارامتری بر داده‌ها برازش داده می‌شود و

3- Kernel Density Estimation
4- Kernel Function
5- Bandwidth
6- Smoothing Function

سپس تابع توزیع تجمعی برای داده‌های یاد شده محاسبه و رسم می‌شود.

۲-۳- تابع مفصل

اگر متغیرهای تصادفی X و Y به ترتیب دارای توابع چگالی احتمال $f_X(x)$ و $f_Y(y)$ باشند، آنگاه تابع توزیع تجمعی متغیرها، یعنی $F_X(x)$ و $F_Y(y)$ خود متغیرهای تصادفی، که دارای تابع چگالی احتمال یکنواخت U(0,1) است. حال مطابق با قضیه Sklar[۷]، اگر $F_X(x)$ و $F_Y(y)$ پیوسته بوده و توابع حاشیه‌ای توزیع توأم F باشند، در این صورت تابع مفصل یکتایی وجود دارد که یک تابع توزیع تجمعی بوده و حاشیه‌های آن یکنواخت است؛ یعنی تابع $C:[0,1]^2 \rightarrow [0,1]$ به گونه‌ای که

$$F(x,y) = C(F_X(x), F_Y(y)) \qquad (٦)$$

یک تابع مفصل در حالت دومتغیره است.

خانواده‌های مختلفی از توابع مفصل از جمله خانواده‌های t، بیضوی، Plackett، گاوسی و ارشمیدسی[۸] در اختیار است که از خانواده ارشمیدسی در پژوهش‌های متعددی در مهندسی آب استفاده شده است (برای نمونه De Michele et al. (2006)، Zhang and Singh (2006) و Grimaldi and Serinaldi (2006)). در مقاله حاضر نیز از سه تابع مفصل ارشمیدسی مهم به نام‌های Clayton، Frank و Gumbel به منظور مدل‌سازی همبستگی متغیرها استفاده شده است.

۲-۳-۱- توابع مفصل ارشمیدسی

ابتدا φ که به آن مولد[۹] گفته می‌شود را به این صورت تعریف می‌کنیم که تابعی پیوسته و نزولی اکید بوده و $\varphi(0)=1$ به گونه‌ای که:

$$\varphi:[0,+\infty) \rightarrow (0,1]$$

7- Sklar's Theorem
8- Archimedean
9- Generator

حال تابع مفصل ارشمیدسی به صورت

$$C(F_X(x), F_Y(y)) = \varphi(\varphi^{-1}(F_X(x)), \varphi^{-1}(F_Y(y))) \qquad (٧)$$

تعریف می‌شود که در آن φ^{-1} معکوس φ است. به ازای مولدهای متفاوت، توابع مفصل ارشمیدسی مختلفی ساخته می‌شوند. ضمناً هر تابع مفصل ارشمیدسی دارای یک پارامتر می‌باشد که به آن پارامتر همبستگی[۱۰] گفته شده و با θ نشان داده می‌شود. از مزیت‌های مهم توابع مفصل ارشمیدسی می‌توان به این موارد اشاره کرد: ۱- نسبت به خانواده‌های دیگر توابع مفصل به سادگی قابل ساخته شدن هستند؛ ۲- دارای رابطه‌ای صریح است؛ ۳- بسیاری از توابع مفصل در دسترس در این خانواده قرار می‌گیرند. در جدول (۱) ویژگی‌های توابع مفصل ارشمیدسی Clayton، Frank و Gumbel آورده شده است. علت مورد توجه قرار دادن این سه تابع مفصل این است که ۱- تعداد توابع مفصل موجود بسیار زیاد است، به گونه‌ای که فقط بیش از ۲۰ تابع مفصل ارشمیدسی در اختیار است (۲۲ تابع در [۱۳] معرفی شده است)؛ بنابراین بررسی همه توابع مفصل و انتخاب مناسب‌ترین تابع از بین آنها منطقی نبوده و باید چند تابع که در مطالعات پیشین مناسب بوده‌اند را مد نظر قرار داده و از بین آنها مناسب‌ترین تابع را بر اساس آزمون نکویی برازش توابع مفصل انتخاب کرد. از توابع مفصل ارشمیدسی Clayton، Frank و Gumbel در مطالعات متعددی برای تحلیل چندمتغیره پدیده‌های هیدرولوژیکی و به طور خاص تحلیل فراوانی سیلاب استفاده شده است. ۲- با برازش سه تابع مفصل Clayton، Frank و Gumbel بر داده‌ها می‌توان وجود سه نوع همبستگی مختلف را بررسی کرد، به گونه‌ای که تابع Clayton دارای همبستگی دنباله‌ای پایینی بوده و تابع Gumbel دارای همبستگی دنباله‌ای بالایی است و تابع مفصل Frank دارای همبستگی دنباله‌ای پایینی یا بالایی

10- Correlation Parameter

نیست [۱۶]. همبستگی دنباله‌ای پایینی به این معناست که برای مقادیر کم $F_X(x)$ انتظار می‌رود که $F_Y(y)$ نیز مقادیر کمی داشته باشد. همچنین همبستگی دنباله‌ای بالایی به این معناست که برای مقادیر زیاد $F_X(x)$ انتظار می‌رود که $F_Y(y)$ نیز مقادیر زیادی داشته باشد. اگر هیچ‌یک از این رفتارها مشاهده نشود یعنی همبستگی دنباله‌ای وجود ندارد.

۲-۳-۲- برآورد پارامتر و آزمون نکویی برازش توابع مفصل

برای برآورد پارامتر تابع مفصل چهار روش اصلی وجود دارد که از جمله پرکاربردترین آن‌ها روش بیشینه درست‌نمایی است [۱۲]. این روش مانند روش بیشینه درست‌نمایی تک‌متغیره، بر بیشینه کردن لگاریتم تابع درست‌نمایی به‌ازای مقادیر مختلف پارامتر θ، به این شکل

$$l(\theta) = \sum_{i=1}^{n} \log\left[c_\theta \quad F(X_i), G(Y_i)\right] \qquad (۸)$$

بنا شده است. در برآوردگر بیشینه درست‌نمایی، c_θ چگالی تابع مفصل بوده و $F(x_i)$ و $F(y_i)$ توابع توزیع تجمعی حاشیه‌ای است. در مقاله حاضر برای برآورد پارامترها از همین روش استفاده شده است. برای انتخاب بهترین تابع مفصل می‌توان از لگاریتم تابع درست‌نمایی بهره گرفت، به گونه‌ای که هر چه مقدار بیشینه لگاریتم درست‌نمایی تابعی بیشتر باشد، بیشتر مورد پذیرش است. در این مقاله نیز از همین روش بهره گرفته شده است.

۲-۶- دوره بازگشت‌های توأم

برای به دست آوردن دوره بازگشت‌های توأم دو متغیر وابسته با استفاده از تابع مفصل، می‌توان از احتمال تجاوز عطفی یا فصلی استفاده کرد. بر همین اساس، چون در این مقاله، محاسبه دوره بازگشت عطفی مورد توجه قرار گرفته است، از رابطه

$$T_{X,Y} = \frac{\mu_T}{1 - F_X(x) - F_Y(y) + C(F_X(x), F_Y(y))} \qquad (۹)$$

استفاده می‌شود، که در آن μ_T متوسط زمان بین مشاهده‌ی متوالی بر حسب سال و برای داده‌های بیشینه سالانه برابر یک سال است.

۲-۷- محیط آماری R

در کدنویسی برای انجام محاسبات و رسم اشکال این مقاله، از زبان آماری R استفاده شده که یک زبان شی‌گرا و رایگان برای انجام محاسبات آماری و کارهای گرافیکی است.

جدول (۱) توابع مفصل ارشمیدسی مهم و مشخصات آنها

نام تابع	Clayton	Frank	Gumbel
$C_\theta(u,v)$	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$-\theta^{-1}\log\{[1+(e^{-\theta u}-1)(e^{-\theta v}-1)]/[e^{-\theta}-1]\}$	$\exp\left[-((-\log u)^\theta + (-\log v)^\theta)^{1/\theta}\right]$
مولد	$(1+\theta t)^{-1/\theta}$	$-\log(1-(1-e^{-\theta})e^{-t})/\theta$	$\exp(-t^{1/\theta})$
معکوس مولد	$\frac{1}{\theta}(t^{-\theta}-1)$	$-\log((e^{-\theta t}-1)/(e^{-\theta}-1))$	$(-\log t)^\theta$
فضای پارامتر	$0 < \theta < \infty$	$0 < \theta < +\infty$	$1 \le \theta < \infty$

در مقایسه با زبان‌های برنامه‌نویسی مانند Fortran، C++ و این زبان از امکانات مناسب‌تری برخوردار بوده و در ارائه نتایج، به خصوص نتایج گرافیکی، این زبان قابلیت‌های سطح بالاتری را دارد.

۳- نتایج و بحث

۳-۱- حوضه آبریز مورد مطالعه

حوضه مورد مطالعه حوضه آبریز کسیلیان است. این حوضه به مساحت حدود ۶۸ کیلومتر مربع در استان مازندران در محدوده عرض جغرافیایی ۳۵°۵۸' و ۳۶°۷' شمالی و طول

جغرافیایی '۵۳°۸ و '۵۳°۱۵ شرقی در شهرستان سوادکوه واقع شده است. بیشینه، کمینه و ارتفاع متوسط حوضه به ترتیب ۳۳۴۹، ۱۱۲۰ و ۱۶۷۲ است. متوسط شیب حوضه ۱۵/۸ درصد، متوسط شیب آبراهه اصلی ۱۳ درصد و طول آبراهه اصلی ۱۶/۵ کیلومتر است. این حوضه دارای دو ایستگاه کلیماتولوژی، ده ایستگاه باران‌سنجی و یک ایستگاه هیدرومتری است. آمار مورد نیاز برای انجام مطالعات از ۴۳ سیلاب حوضه یاد شده بین سال‌های ۶۷ و ۷۸ به دست آمده است. در جدول (۲) برخی از آمارهای مه.۸م برای هر یک از متغیرهای سیلاب نمایش داده شده است.

۳-۲- برازش توزیع‌های پارامتری و ناپارامتری

هشت تابع توزیع یاد شده در بخش ۲-۱ بر داده‌های هر یک از متغیرهای دبی‌اوج، حجم‌رواناب و مدت سیلاب برازش داده شده و پارامترهای برآورد شده در جدول (۳) آورده شده است.

حال آزمون نکویی برازش KS برای هر یک از توزیع‌ها انجام شده و برای متغیرهای حجم‌رواناب و مدت سیلاب تمامی توزیع‌ها پذیرفته شده‌اند؛ در حالی که برای متغیر دبی‌اوج، فقط توزیع‌های گاما، لگ‌نرمال و ویبول پذیرفته شده‌اند. سپس توزیع ناپارامتری به روش برازش چگالی هسته‌ای بر سری داده‌های یاد شده برازش داده شده است. با در نظر گرفتن هسته نرمال و بر اساس رابطه (۳) میزان پهنای باند برای متغیر مدت سیلاب برابر h=۰/۲۱، برای متغیر حجم رواناب برابر h=۰/۰۸، و برای متغیر دبی اوج برابر h=۰/۱۷۵، محاسبه شده است. شکل (۲) بیانگر برازش چگالی هسته‌ای بر روی سه متغیر سیلاب است.

نکته قابل توجه در مورد این شکل‌ها این است که این برازش به خوبی توانسته تغییرات چگالی داده‌ها را به نمایش بگذارد.

جدول (۲) برخی از آماره‌های مهم متغیرهای سیلاب

متغیر	میانگین	انحراف معیار	حداقل	حداکثر	چولگی	کشیدگی
دبی‌اوج (MCM/Day)	۰/٤۹۳	۰/٤۱۲	۰/۱۳۳۹	۱/٦٥۸۹	۱/۱٥۹	۳/۰۹٥
حجم‌رواناب (MCM)	۰/۳۲۷	۰/۲۹۱	۰/۰۲۹٥	۱/۳۲٤	۱/۹٤٥	٦/۷۳۷
مدت (Day)	۱/۹۹٤	۰/٦۷۸	۰/۸۷٥	۳/۸۷٥	۰/۷۸٤	۳/۳۹۰

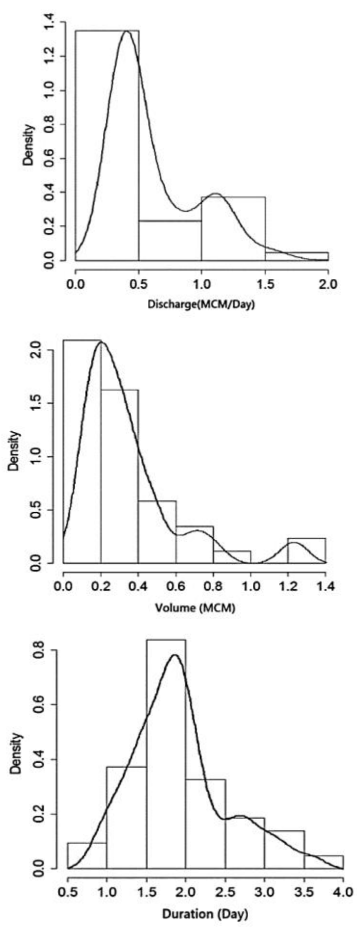
جدول (۳) مقادیر برآورد پارامترهای توزیع‌های پارامتری برازش داده شده بر متغیرهای سیلاب

توزیع	نام پارامتر	مقدار		
		دبی‌اوج	حجم‌رواناب	مدت سیلاب
نمایی	پارامتر نرخ	۲/۰۲٥۹	۳/۰٥۸۲	۰/٥۰۱٤٦
گاما	پارامتر شکل	۱/٤۳۰۹	۱/۲٦۱٦	۸/٦٥۲۳
	پارامتر مقیاس	۰/۳٤٤۹٦	۰/۲٥۹۱۸	۰/۲۳۰٤۸
نرمال	پارامتر مکان (میانگین)	۰/٤۹۳٦۱	۰/۳۲٦۹۹	۱/۹۹٤۲
	پارامتر مقیاس (انحراف معیار)	۰/٤۱۲٦٤	۰/۲۹۱۱۲	۰/٦۷۷۹٥
لگ‌نرمال	پارامتر مکان (لگاریتم میانگین)	-۱/۰۱۱۸	-۱/٤۳۹۳	۰/٦۳٥
	پارامتر مقیاس (لگاریتم انحراف)	۰/۷٦۱۷٦	۰/۸۱۱٥۲	۰/۳۳٤٥۹

توزیع	نام پارامتر	مقدار		
		دبی‌اوج	حجم‌رواناب	مدت سیلاب
گامبل	پارامتر مکان	۰/۳۰۷۹	۰/۱۹۵۹۷	۱/٦٨۹۱
	پارامتر مقیاس	۰/۳۲۱۷٤	۰/۲۲٦۹۸	۰/٥۲۸٦
ویبول	پارامتر شکل	۱/٤۱٥۷	۱/٤٥۲٥	۳/٥۳۱٦
	پارامتر مقیاس	۰/٥۱٥۲٤	۰/۳۳۱۱٥	۲/۱٦٤۳
کوشی	پارامتر مکان	۰/۲٥۱۸٥	۰/۲۱۳۷٦	۱/۸۸۹۷
	پارامتر مقیاس	۰/۱۱۷٤۱	۰/۱۰۷۳٦	۰/۳۲۲٦٦
لوجستیک	پارامتر مکان	۰/۲٥۱۸٥	۰/۲۱۳۷٦	۱/۸۸۹۷
	پارامتر مقیاس	۰/۱۱۷٤۱	۰/۱٦۰٥	۰/۳۷۳۷۸

حال RMSE برای تک‌تک توزیع‌های برازش داده شده بر داده‌های متغیرهای سیلاب محاسبه شده و نتایج آنها در جدول (٤) آورده شده است.

بر اساس نتایج RMSE محاسبه شده برای هر یک از سه متغیر سیلاب، خطای استاندارد توزیع ناپارامتری در هر سه متغیر از خطای استاندارد توزیع‌های پارامتری کمتر شده و این توزیع به عنوان توزیع برتر از میان توزیع‌های پارامتری و توزیع ناپارامتری انتخاب شده است.



شکل (۲) برازش توزیع ناپارامتری بر متغیرهای دبی‌اوج، حجم‌رواناب و مدت سیلاب

جدول (٤) محاسبات خطای استاندارد برای توزیع‌های پارامتری و توزیع ناپارامتری

توزیع	حجم‌رواناب	مدت‌سیلاب	دبی‌اوج
کوشی	۰/۲٦۰۷	۰/۷۲۸	رد شده توسط آزمون KS
نمایی	۰/۰٥۷۹	۱/۱۰۲۸	رد شده توسط آزمون KS
گاما	۰/۰۷۰۲	۰/۱۱۳٦	۰/۱۰۹۹
گامبل	۰/۰۹٤۳	۰/۱۰٦۲	رد شده توسط آزمون KS
لوجستیک	۰/۱۳۷۷	۱/٥٦۰٥	رد شده توسط آزمون KS
لگ‌نرمال	۰/۰٦۱٥	۰/۱۰۲٤	۰/۱۳۰۳
نرمال	۰/۱٤۲٤	۰/۱٦٤۷	رد شده توسط آزمون KS
ویبول	۰/۱۲۷۳	۰/۱۹۲۳	۰/۱٤۹۳
ناپارامتری	۰/۰۳۱	۰/۰٦۱٥	۰/۰۸۷۲

۳-۳- همبستگی بین متغیرهای سیلاب

در جدول (۵) ضرایب همبستگی پیرسون، کندال و اسپیرمن محاسبه شده بین متغیرها آورده شده است. وجود همبستگی مثبت بین زوج‌متغیرها نشان دهنده این است که انجام تحلیل سه‌متغیره دبی‌اوج-حجم‌رواناب-مدت سیلاب مناسب است. اما همان گونه که ذکر شد، هدف این مقاله انجام تحلیل‌های دومتغیره است.

جدول (۵) مقادیر ضریب‌های همبستگی بین متغیرهای سیلاب

متغیرها	دبی اوج-حجم رواناب	دبی اوج-مدت	حجم رواناب-مدت
ضریب همبستگی	۰/۸۲۳	۰/۳۶۲	۰/۷۰۸
ضریب همبستگی	۰/٦۸۷	۰/۲۸۷	۰/۵۳۵
ضریب همبستگی	۰/۸۵۱	۰/۳۹۱	۰/٦۹۹

۳-٤- برآورد پارامتر و انتخاب مناسب‌ترین تابع مفصل

در جدول (٦) پارامترهای برآورد شده از طریق روش بیشینه درست‌نمایی و همچنین مقدار بیشینه تابع لگاریتم درست‌نمایی آورده شده است. همان‌گونه که از این جدول به دست می‌آید، برای زوج‌متغیر دبی‌اوج حجم‌رواناب تابع مفصل Frank با پارامتر θ = ۱۰/۷۰، برای زوج‌متغیر دبی‌اوج- مدت تابع مفصل Frank با پارامتر θ = ۲/۸۳، و برای زوج‌متغیر حجم‌رواناب- مدت تابع مفصل Gumbel با پارامتر θ = ۲/۰۸، دارای بیشترین مقدار بیشینه درست‌نمایی بوده و به عنوان بهترین تابع مفصل از میان سه گزینه مورد بررسی برگزیده می‌شوند. با مشخص شدن تابع مفصل مناسب و پارامتر آن، در واقع ساختار همبستگی زوج‌متغیرها به دست آمده است.

جدول (٦) مقادیر پارامتر توابع مفصل (به روش حداکثر درست‌نمایی) و حداکثر درست‌نمایی

متغیرها	تابع مفصل	مقدار پارامتر	مقدار حداکثر تابع لگاریتم درست‌نمایی
دبی‌اوج-حجم‌رواناب	Clayton	۱/۹٤	۱٤/۹۸
	Frank	۱۰/۷۰	۲۷/٤۲
	Gumbel	۲/۹۳	۲٦/۵٦
دبی‌اوج-مدت	Clayton	۰/۵۱	۱/۵۱
	Frank	۲/۸۳	۳/۷۲
	Gumbel	۱/۳٤	۲/۸۷
حجم‌رواناب-مدت	Clayton	۱/٦٤	۱۲/٤۰
	Frank	٦/۲۲	۱٤/۵۵
	Gumbel	۲/۰۸	۱۵/۲۰

۳-۵- دوره بازگشت‌های توأم عطفی

برای استفاده از رابطه (۹) به منظور محاسبه دوره بازگشت‌های توأم عطفی، نخست باید مقدار μ_T مشخص باشد. به این منظور متوسط فاصله زمانی بین هیدروگراف‌های متوالی محاسبه شده و برابر ۰/۳ سال به دست آمده است. باید توجه کرد که در رابطه (۹) $F_X(x)$ و $F_Y(y)$ که ورودی‌های تابع مفصل است، خود از طریق رابطه (٤) به‌دست آمده‌اند. در شکل (۳) (سمت راست) رویه‌ها و در شکل (۳) (سمت چپ) خطوط تراز دوره بازگشت‌های توأم عطفی با بیشینه دوره بازگشت ۱۰۰۰ سال رسم شده است.

٤- نتیجه‌گیری

در این پژوهش به منظور انجام تحلیل‌های توأم دبی‌اوج-حجم‌رواناب، دبی‌اوج- مدت و حجم‌رواناب- مدت سیلاب و به دست آوردن دوره بازگشت‌های توأم زوج‌متغیرهای یاد شده، دسته قدرتمندی از توابع چندمتغیره با نام مفصل با توابع توزیع حاشیه‌ای ناپارامتری استفاده شده است. نکته حائز اهمیت در مطالعه انجام گرفته این گونه است:

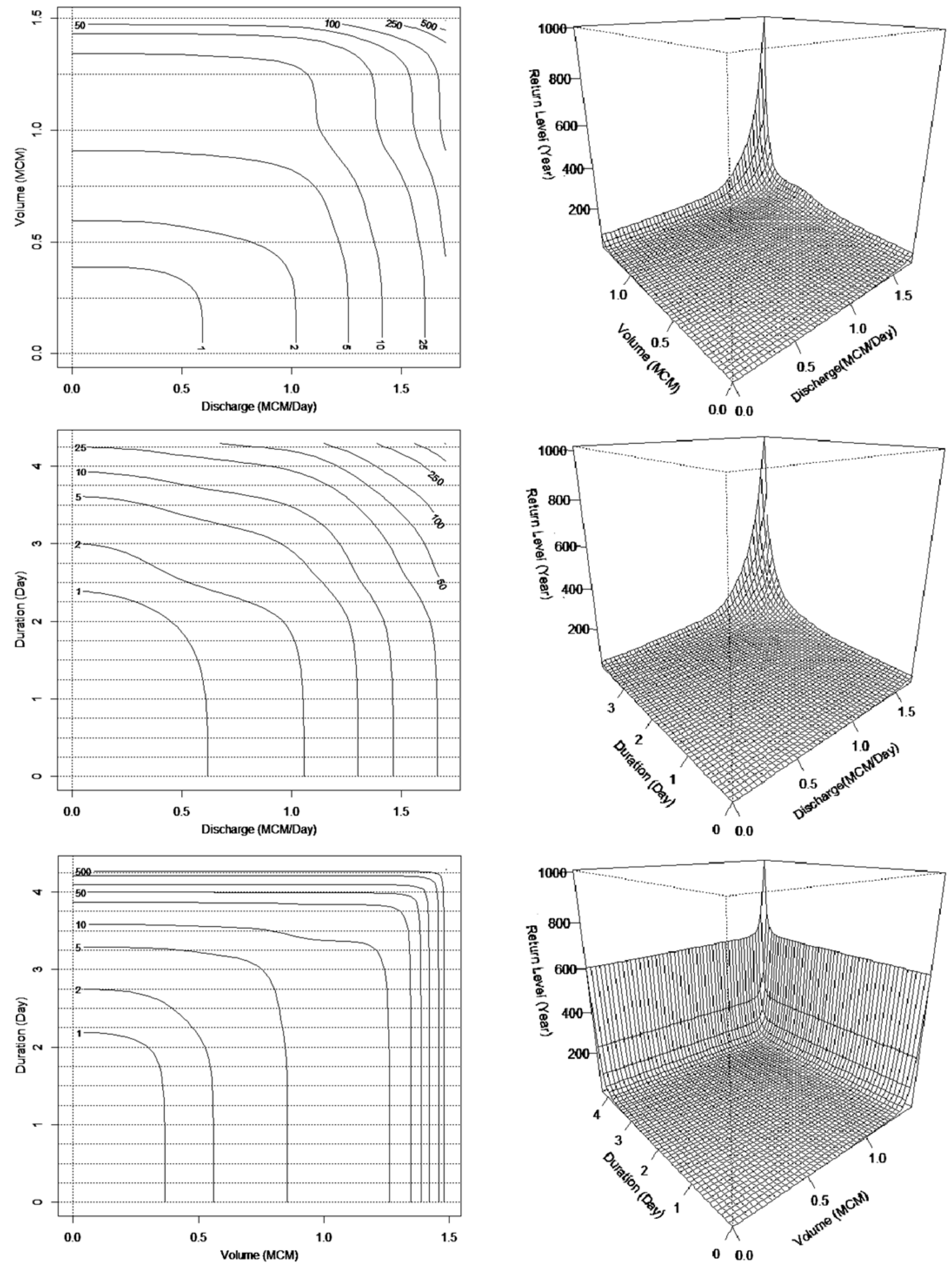
۱- هدف از به کارگیری تابع مفصل برطرف کردن محدودیت‌های قابل توجه در استفاده از توزیع‌های چندمتغیره کلاسیک (مانند نرمال، لگ‌نرمال، گامبل و ...) در مدل‌سازی همبستگی متغیرهای وابسته بوده که مهم‌ترین این محدودیت‌ها لزوم مشخص بودن توابع توزیع حاشیه‌ای و یکسان بودن آنها است. در به خدمت گرفتن تابع مفصل الزامی در مشخص کردن توابع توزیع حاشیه‌ای و برآورد پارامتر آنها وجود ندارد، به گونه‌ای که برای برآورد پارامتر توابع مفصل می‌توان از توزیع‌های تجربی متغیرهای حاشیه‌ای استفاده کرد.

۲- با توجه به نیاز به برون‌یابی در هر متغیر برای یافتن مقادیر با احتمالات بزرگ به منظور انجام تحلیل فراوانی، برازش یک تابع احتمال مناسب به هر متغیر ضروری است. نظر به قابلیت مهم تابع مفصل مبنی بر اینکه توابع توزیع حاشیه‌ای آن می‌توانند از بین توابع پارامتری یا ناپارامتری انتخاب شده به شکلی که حتی برخی از آنها پارامتری و برخی دیگر ناپارامتری باشند، در این پژوهش انتخاب بهترین توابع توزیع حاشیه‌ای از بین توابع پارامتری و ناپارامتری بر مبنای RMSE صورت گرفته که در انتها برای هر یک از متغیرهای دبی‌اوج، حجم‌رواناب و مدت سیلاب توزیع ناپارامتری برگزیده شده است.

۳- چگونگی برازش توابع ناپارامتری بر داده‌های متغیرها که در شکل (۲) به‌نمایش درآمده است نشان می‌دهد که بر خلاف توزیع‌های پارامتری که پس از مشخص شده پارامترها شکل تابع مشخص شده و از تغییرات موضعی توزیع واقعی داده‌های مشاهده شده پیروی نمی‌کنند، توزیع‌های ناپارامتری نسبت به تغییرات موضعی توزیع واقعی داده‌ها حساس بوده و سعی در هماهنگی خود با توزیع واقعی دارند. این موضوع را می‌توان از مقایسه هیستوگرام متغیرها با منحنی‌های تابع توزیع ناپارامتری آنها دریافت.

۴- اگرچه توزیع‌های ناپارامتری می‌توانند گزینه بسیار مناسبی برای درون‌یابی باشند، چون به میزان قابل توجهی متأثر از داده‌های مشاهده شده است، برای برون‌یابی با احتمالات بزرگ مناسب نبوده و بنابراین در این پژوهش به محاسبه دوره بازگشت‌های توأم با بیشینه دوره بازگشت ۱۰۰۰ سال بسنده شده است. در مقابل برای توزیع‌های پارامتری از حیث نظری حد توقفی وجود ندارد و برای کنترل توزیع مورد نظر باید فاصله اطمینان آن را رسم نمود تا بتوان در مورد توزیع برازش یافته اظهار نظر کرد.

۵- پس از مشخص کردن بهترین تابع مفصل از بین سه تابع مفصل ارشمیدسی Clayton، Frank و Gumbel، امکان به‌دست آوردن دوره بازگشت‌های توأم زوج‌متغیرهای دبی‌اوج- حجم‌رواناب، دبی‌اوج- مدت و حجم‌رواناب- مدت فراهم آمده و رویه‌ها و خطوط تراز دوره بازگشت‌های توأم رسم شده است. از شکل ۳ (سمت راست) چگونگی تغییرات دوره بازگشت عطفی در برابر متغیرهای حاشیه‌ای قابل ملاحظه است. وقتی مقادیر متغیرها از یک آستانه عبور نماید، رویه‌های دوره بازگشت توأم به شدت اوج می‌گیرند. علت این موضوع به ماهیت توزیع‌های ناپارامتری برمی‌گردد که تحت اثر داده‌های مشاهده شده است. در صورت به کارگیری توابع توزیع حاشیه‌ای پارامتری، تغییرات رویه‌های دوره بازگشت‌های توأم ملایم‌تر خواهد بود.



شکل (۳) رویه‌های دوره بازگشت توأم (سمت راست) و خطوط تراز دوره بازگشت توأم (سمت چپ)

زوج متغیرهای دبی‌اوج– حجم‌رواناب، دبی‌اوج– مدت و حجم‌رواناب– مدت

٦- از شکل (۳) (سمت چپ) امکان قرائت مقادیر متغیرهای حاشیه‌ای به ازای دوره بازگشت‌های مختلف فراهم شده است. اثر اوج گرفتن ناگهانی رویه‌ها به صورت نزدیک شدن خطوط دوره بازگشت در دوره بازگشت‌های بالا نمایان شده و به علاوه در خطوط تراز اعوجاج‌هایی دیده می‌شود (به گونه‌ای خاص در دوره بازگشت توأم ۱۰ ساله برای زوج‌متغیرهای متغیرهای دبی‌اوج و حجم‌رواناب) که این موضوع ناشی از نامنظم بودن شکل توابع ناپارامتری است. اگر توزیع‌های حاشیه‌ای از توزیع‌های پارامتری باشند، شکل خطوط تراز بسیار هموارتر خواهد بود.

۷- تحلیل‌های توأم برخلاف تحلیل‌های تک‌متغیره که تنها یک عدد را به ازای هر دوره بازگشت در اختیار می‌دهند،

بازه‌ای از مقادیر را برای هر متغیر به ازای هر دوره بازگشت در اختیار قرار داده‌اند که این نشان دهنده در نظر گرفتن اثر متقابل متغیرهای وابسته سیلاب در تحلیل این پدیده است. بنابراین در تحلیل ریسک که به احتمال وقوع یا دوره بازگشت سیلاب نیاز است و در طرح‌های سازه‌ای یا غیرسازه‌ای مهندسی رودخانه، می‌توان از نتایج تحلیل‌های چندمتغیره برای تصمیم‌گیری درباره مقادیر دبی‌اوج طرح، حجم‌رواناب طرح و مدت طرح استفاده نمود. به عنوان نمونه، از آن‌جایی که ریسک احتمال شکست است، اگر بدانیم مقادیر بحرانی دبی‌اوج، حجم‌رواناب و مدت‌سیلاب بحرانی که منجر به شکست پروژه می‌شوند چقدر است، می‌توان ریسک را از روی احتمال‌ها یا دوره بازگشت‌های توأم به‌دست آورد.

۵- مراجع

۱- اشرف‌زاده؛ ا.؛ خلقی؛ م.؛ "توسعه و کاربرد یک مدل غیرپارامتری برای شبیه‌سازی دبی جریان رودخانه"؛ نشریه علوم کشاورزی ایران؛ ۳۶ (٤)، ۱۳۸۳، صفحات ۹۹۱-۹۹۹.

2- Adamowski K., “Nonparametric Kernel Estimation of Flood Frequencies”, Water Resources Research, 21 (11), 1585-1590, 1985.

3- Bárdossy, A., “Copula-based geostatistical models for groundwater quality parameters”, Water Resources Research, 42(11), W11416, 2006.

4- Chebana F., Ouarda T. B. M. J., “Index flood-based multivariate regional frequency analysis”, Water Resources Research, 45(10), 2009.

5- Chowdhary, H., Singh, V. P., “Gains from copula-based multivariate Distribution Functions for Rainfall Processes”, World Environmental and Water Resources Congress, Chapter 569, 1-10, 2008.

6- De Michele C., Salvadori G., Canossi M., Petaccia A., Rosso, R., “Bivariate statistical approach to check adequacy of dam spillway”, Journal of Hydrologic Engineering, 10(1), 50-57, 2005.

7- De Michele C., Salvadori G., “A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas”, Geophysical Research Atoms, 108(D2), 2003.

8- Favre A. C., El Adlouni S., Perreault L., Thié monge N., Bobée B., “Multivariate hydrological frequency analysis using copulas”, Water Resources Research, 40 (1), 2004.

9- Genest, C., Favre, A-C., Béliveau, J., Jacques, C., “Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data”, Water Resources Research, 43(9), W09401, 2007.

10- Grimaldi S., Serinaldi F., “Asymmetric copula in multivariate flood frequency analysis”, Advances in Water Resources, 29(8), 1115–1167, 2006.

11- Karmakar S., Simonovic S. P. “Bivariate flood frequency analysis, Part 2: a copula-based approach with mixed marginal distributions”, Journal of Flood Risk Management, 2(1), 32-44, 2009.

12- Nazemi A. Elshorbagy, A. A., “Application of Copula Modelling to the Performance Assessment of Reconstructed Watersheds”, Stochastic Environmental Research & Risk Assessment, 26(2), 189-205, 2012.

13- Nelsen, R.B., “An Introduction to Copulas”, In: Lecture Notes in Statistics, vol. 139, New York: Springer, 2006.

14- Rénard B., Lang M., “Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology”, Advances in Water Resources, 30(4), 897-912, 2007.

15- Samaniego, L., Bárdossy, A., and Kumar, R., “Streamflow prediction in ungauged catchments using copula-based dissimilarity measures”, Water Resources Research, 46(2), W02506, 2010.

16- Schmidt, T., “Coping with Copulas”, Risk Books: Copulas – From Theory to Applications in Finance, London: Incisive Financial Publishing, 2006.

17- Scott D. W., “Multivariate Density Estimation. Theory, Practice, and Visualization”, New York: Wiley, 1992.

18- Shabri A., “Nonparametric Kernel Estimation of Annual Maximum Stream Flow Quantiles”, Matematika, 18 (2), 99-107, 2002.

19- Sharma A., Lall U., Tarboton D.G., “Stream flow simulation: A nonparametric approach”, Water Resources Research , 33(2), 291-308, 1997.

20- Shiau, J. T., “Fitting Drought Duration and Severity with Two-Dimensional Copulas”, Water Resources Management, 20(5), 795–815, 2006.

21- Silverman B. W., “Density Estimation for Statistics and Data Analysis”, London: Chapman and Hall, 1986.

22- Wang C., Ni-Bin Chang N. B., Yeh, G. T., “Copula-based flood frequency (COFF) analysis at the confluences of river systems”, Hydrological Processes, 23(10), 1471-1486, 2009.

23- Wolfowitz J., “Additive Partition Functions and a Class of Statistical Hypotheses”, Annals of Statistics, 13 (3), 247-279, 1942.

24- Zhang L., Singh V. P., “Bivariate flood frequency analysis using the copula method”, Journal of Hydrological Engineering, 11(2), 150–16, 2006