



دانشگاه گوار، نشریه پژوهش‌های حفاظت آب و خاک

نشریه پژوهش‌های حفاظت آب و خاک

جلد بیستم، شماره ششم، ۱۳۹۲

<http://jwsc.gau.ac.ir>

## تحلیل فراوانی متغیرهای سیلاب با استفاده از روش‌های پارامتری و ناپارامتری

\* میثم سالاری جزی<sup>۱</sup>، علی محمد آخوندعلی<sup>۲</sup>، آرش ادیب<sup>۳</sup> و علیرضا دانشخواه<sup>۴</sup>

<sup>۱</sup> دانشجوی دکتری، گروه هیدرولوژی و منابع آب، دانشگاه شهید چمران اهواز، استاد گروه هیدرولوژی و منابع آب،

دانشگاه شهید چمران اهواز، استادیار گروه مهندسی عمران آب، دانشگاه شهید چمران اهواز،

<sup>۴</sup> استادیار گروه آمار، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۹۱/۳/۷؛ تاریخ پذیرش: ۹۱/۹/۷

### چکیده

شیوه‌های مرسوم در تحلیل فراوانی سیلاب، متغیر دبی پیک سیلاب را مورد بررسی قرار می‌دهد، در حالی که هر رویداد سیلاب دارای سه متغیر تصادفی در طبیعت یعنی دبی پیک، تداوم و حجم سیلاب می‌باشند. فرض دیگر در این‌گونه تحلیل‌ها تبعیت متغیرها از مجموعه‌ای از توابع توزیع پارامتری خاص است. با هدف غلبه بر این محدودیت‌ها مجموعه گسترده‌ای از توابع توزیع پارامتری و روش‌های ناپارامتری مبتنی بر برآورد چگالی کرنل و سری‌های متعامد نرمال برای تعیین توزیع متغیرهای سیلاب در این پژوهش مورد توجه قرار گرفته است. متغیرهای سیلاب مورد بررسی مستخرج از سری زمانی بیشترین سالانه سیلاب در ایستگاه هیدرومتری اهواز است. معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا برای مقایسه برازش توزیع‌های مختلف بر هیستوگرام تجربی داده‌ها به کار گرفته می‌شود و آزمون  $\chi^2$  برای بررسی قابلیت پذیرش توزیع‌های مختلف برای متغیرهای سیلاب مورد استفاده قرار می‌گیرد. نتایج نشان می‌دهد که دبی پیک سیلاب از تابع توزیع پارامتری لوگ پیرسون نوع سه پیروی می‌کند در حالی که روش ناپارامتری سری‌های متعامد نرمال منجر به برآورد مناسب‌تر تابع توزیع تداوم و حجم سیلاب شده است. همچنین مشخص گردید توزیع‌های برآورد شده با روش‌های ناپارامتری سری‌های متعامد نرمال تطابق بهتری با هیستوگرام

\* مسئول مکاتبه: [meysam.salarijazi@gmail.com](mailto:meysam.salarijazi@gmail.com)

تجربی داده‌ها دارد و قادر به شبیه‌سازی حالت دو و سه مدی هیستوگرام تجربی داده‌ها می‌باشد که این حالت توسط هیچ‌کدام از توزیع‌های حدی پارامتری رایج قابل شبیه‌سازی نیست.

**واژه‌های کلیدی:** تحلیل فراوانی، متغیرهای سیلاب، روش‌های پارامتری، برآورد چگالی کرنل، سری‌های متعامد نرمال

### مقدمه

تحلیل فراوانی تک متغیره بر اساس پیک سیلاب به شکل گسترده‌ای در برآورد ریسک سیلاب مد نظر قرار گرفته شده است (گریمالدی و سرینالدی، ۲۰۰۶؛ کایت، ۱۹۷۸؛ کونان، ۱۹۸۷؛ رائو و حامد، ۲۰۰۰). تحلیل و برآورد واقع‌گرایانه‌تر ریسک در زمینه طراحی‌های هیدرولوژیکی و برنامه‌ریزی و مدیریت سیلاب نیازمند به بررسی دبی پیک، تداوم و حجم سیلاب می‌باشد (آدامسون و همکاران، ۱۹۹۹؛ پینیا و همکاران، ۲۰۰۹؛ کارماکار و سیمونوویچ، ۲۰۰۸). تحلیل فراوانی بر مبنای برآورد تابع چگالی احتمال می‌باشد. روش‌های پارامتری برآورد تابع چگالی احتمال فرض می‌کنند که داده‌ها متعلق به تعدادی از خانواده‌های پارامتری شناخته شده و پر کاربرد مانند توزیع مقادیر حدی تعمیم یافته، مقادیر حدی نوع یک و لوگ پیرسون نوع سه و ... می‌باشند. تعداد زیادی از توزیع‌های پارامتری برای تحلیل فراوانی داده‌های سیلاب مورد استفاده قرار گرفته است (کیدسون و ریچارد، ۲۰۰۵؛ کارماکار و سیمونوویچ، ۲۰۰۸). توزیع‌های مقادیر حدی تعمیم یافته و لوگ پیرسون نوع سه به‌عنوان یک روش پایه به‌ترتیب در انگلستان و آمریکا توصیه شده است (هال، ۱۹۸۴؛ آداموسکی، ۱۹۸۹). به‌طور کلی یک تابع توزیع احتمال یا یک فرایند برازش پذیرفته شده عمومی برای متغیرهای هیدرولوژیکی وجود ندارد (آداموسکی، ۱۹۸۵؛ اسماختین، ۲۰۰۱). دوگ (۱۹۸۶) بیان می‌دارد تحلیل‌های متفاوت بر اساس روش‌های مختلف آماری نمی‌تواند بر معایب ناآگاهی از توزیع فراوانی غلبه نماید. در روش‌های پارامتری در ابتدا توزیع خاصی برای جامعه موردنظر فرض می‌شود و در ادامه با شیوه‌های متفاوت تلاش می‌شود پارامترهای توزیع فرض شده برآورد گردد. همچنین قریب به اتفاق توزیع‌های فرض شده در زمینه تحلیل فراوانی تنها قادر به شبیه‌سازی توزیع‌های یک مدی می‌باشند در صورتی‌که در موارد متعددی امکان دارد توزیع تجربی داده‌ها چند مدی باشد. برآورد ناپارامتری تابع چگالی احتمال با هدف غلبه بر برخی از محدودیت‌های روش‌های پارامتری در هیدرولوژی مورد توجه قرار گرفت

(لال، ۱۹۹۵). مزیت روش‌های ناپارامتری عدم نیاز به فرضیات مربوط به توزیع جامعه مورد نظر می‌باشد. روش‌های ناپارامتری برای داده‌های برگرفته از هر جامعه آماری با هر توزیع احتمال معتبر هستند (حقیقت جو، ۲۰۰۸). برآوردهای ناپارامتری تابع، با استفاده از میانگین‌های متحرک وزنی داده‌ها در یک همسایگی کوچک نقطه‌ای که برآورد آن انجام می‌شود، این مزیت را دارد که همیشه ویژگی‌های بیان شده توسط نمونه، از جمله شکل دو یا چند مدی توزیع تجربی نمونه مورد بررسی، را شبیه‌سازی می‌کند (لال، ۱۹۹۵؛ کیم و همکاران، ۲۰۰۳). مطالعات انجام شده در زمینه کاربرد روش‌های ناپارامتری در هیدرولوژی نسبت به روش‌های پارامتری به مراتب محدودتر و اغلب بر اساس کاربرد برآورد چگالی کرنل بوده است که می‌توان به تحقیقات آداموسکی (۱۹۸۵ و ۱۹۸۹) و لال و همکاران (۱۹۹۳) در زمینه تحلیل فراوانی سیلاب، آداموسکی (۱۹۹۶) در زمینه تحلیل فراوانی جریان‌های کم و کیم و همکاران (۲۰۰۳ و ۲۰۰۶) در زمینه تحلیل خشکسالی اشاره کرد. همچنین حقیقت جو و همکاران (۲۰۰۹) از روش ناپارامتری سری فوریه برای پیش‌بینی بارش سالانه و کارماکار و سیمونویچ (۲۰۰۸) از روش ناپارامتری سری‌های متعامد نرمال در تحلیل فراوانی سیلاب استفاده کرده و نشان داده‌اند توابع چگالی ناپارامتری با توزیع تجربی متغیرهای تصادفی در طبیعت می‌توانند به خوبی تطابق نمایند و به‌ویژه برای توزیع‌های چند مدی مناسب می‌باشند. مروری بر منابع آماری نشان می‌دهد که برآوردهای آماری ناپارامتری یک زمینه فعال تحقیقاتی و نیازمند توسعه در زمینه‌های جدید می‌باشند (بومن و آزالینی، ۱۹۹۷؛ سیلورمن، ۱۹۸۶؛ شارما و همکاران، ۱۹۹۷؛ هیگینز، ۲۰۰۴؛ کارماکار و سیمونویچ، ۲۰۰۸). هدف این پژوهش انتخاب تابع چگالی مناسب متغیرهای دبی پیک، حجم و تداوم سیلاب با استفاده از روش‌های پارامتری و ناپارامتری می‌باشد. روش ناپارامتری برمبنای برآورد چگالی کرنل و روش ناپارامتری نوینی برمبنای سری‌های متعامد نرمال برای تعیین توابع توزیع ناپارامتری متغیرهای سیلاب استفاده می‌گردد. استفاده از تحلیل چند متغیره و به‌کارگیری روش‌های ناپارامتری موجب غلبه بر محدودیت‌های روش‌های ممتداول می‌گردد. باتوجه به نتایج پژوهش‌ها در مورد کارایی روش نوین ناپارامتری برمبنای سری‌های متعامد نرمال جزییات این روش تشریح شده است. توسعه ریاضی روش ارائه شده در این پژوهش از مطالعات بومن و آزالینی (۱۹۹۷)، افروموویچ (۱۹۹۹)، گوش و ماژومدار (۲۰۰۷) و کارماکار و سیمونویچ (۲۰۰۸) اقتباس شده است.

## مواد و روش‌ها

معرفی منطقه مورد مطالعه و داده‌های مورد استفاده: حوضه آبریز رودخانه‌های کارون و دز (کارون بزرگ) به مساحت تقریبی ۶۷۵۰۰ کیلومتر مربع به مختصات جغرافیایی از ۳۰ درجه تا ۳۴ درجه و ۱۵ دقیقه عرض شمالی و طول ۴۸ درجه تا ۵۲ درجه شرقی محدوده‌ای از جنوب غربی ایران را پوشش می‌دهد. ایستگاه هیدرومتری اهواز بر روی رودخانه کارون بزرگ و در مختصات جغرافیایی ۴۸ درجه و ۴۱ دقیقه طول شرقی و ۳۱ درجه و ۲۰ دقیقه عرض شمالی و ارتفاع ۱۰ متر نسبت به سطح دریا واقع گردیده و مساحت حوضه آبریز آن ۶۰۷۳۷ کیلومتر مربع می‌باشد (سالاری جزی و همکاران، ۲۰۱۲). با توجه به داده‌های موجود، سیلاب‌های سالانه مشخص گردید و سری‌های زمانی متغیرهای پیک، حجم و تداوم سیلاب استخراج گردید. سری زمانی داده‌های سیلاب (با در نظر گرفتن هیدروگراف‌های سیلاب ثبت شده) در دسترس مورد بررسی قرار گرفته و با توجه به عوامل و متغیرهای متفاوت و نیز تحلیل‌های آماری پایه‌ای بازه زمانی حد فاصل سال آبی ۱۳۵۶-۱۳۵۷ تا ۱۳۸۸-۱۳۸۷ برای این پژوهش در نظر گرفته شده است. با توجه به قرار داشتن بسیاری از پروژه‌های عمرانی مانند پل‌ها، مکان‌های مسکونی، جاده‌ها، مناطق صنعتی و ... در امتداد این رودخانه تحلیل منطقی ریسک سیلاب ضروری می‌باشد.

**روش‌های پارامتری:** در طی چندین دهه مطالعه، تعداد زیادی از توابع توزیع پارامتری در هیدرولوژی مورد کاربرد و ارزیابی قرار گرفته است. علاوه بر تعدد توزیع‌های حدی پارامتری روش‌های مختلفی برای تعیین پارامترهای این توزیع‌ها وجود دارد. روش گشتاورها، روش بیشترین درست‌نمایی و روش گشتاورهای خطی از روش‌های رایج برآورد پارامترهای این توزیع‌ها در هیدرولوژی می‌باشند. در این مطالعه تلاش شده است دامنه نسبتاً وسیعی از توزیع‌های پارامتری و روش‌های برآورد پارامترها برای تعیین توزیع مناسب متغیرهای سیلاب مورد بررسی قرار گیرد. براین اساس توزیع‌های پارامتری و برآوردگرهای منتخب در تحلیل فراوانی متغیرهای سیلاب عبارتند از: نمای (بیشترین درست‌نمایی)، پارتو تعمیم یافته (گشتاورها)، پارتو تعمیم یافته (گشتاورهای خطی)، گامبل (گشتاورهای خطی) گامبل (بیشترین درست‌نمایی)، گامبل (گشتاورها)، مقادیر حدی تعمیم یافته (گشتاورها)، مقادیر حدی تعمیم یافته (بیشترین درست‌نمایی)، وایبل (گشتاورها)، وایبل (گشتاورهای خطی)، گاما (گشتاورها)، گاما (گشتاورهای خطی)، لوگ پیرسون نوع سه

(گشتاورها)، لوگ پیرسون نوع سه (گشتاورهای خطی)، لوگ نرمال دو پارامتری (گشتاورها)، لوگ نرمال دو پارامتری (گشتاورهای خطی) و نمایی ریشه مربع (بیشترین درست‌نمایی).  
 روش ناپارامتری کرنل (برآورد چگالی کرنل تک متغیره): بیشتر روش‌های برآورد چگالی ناپارامتری می‌تواند با برآوردگر چگالی کرنل بیان شود که دربرگیرنده میانگین متحرک وزنی توزیع فراوانی تجربی نمونه می‌باشد (اسکات، ۱۹۹۲؛ کیم و همکاران، ۲۰۰۳). تابع کرنل تک متغیره،  $K(x)$ ، ویژگی زیر را دارد.

$$\int_{-\infty}^{+\infty} K(x)dx = 1 \quad (1)$$

بنابراین یک تابع چگالی احتمال می‌تواند به‌عنوان یک تابع کرنل به‌کار رود. در اغلب مطالعات مرتبط با هیدرولوژی و منابع آب توابع کرنل اپانچنیکوف، گوسی یا نرمال، مستطیلی و مثلثی مورد استفاده قرار می‌گیرد (کیم و همکاران، ۲۰۰۳). در این مطالعه تابع کرنل گوسی (نرمال) با میانگین صفر و واریانس یک مورد استفاده قرار گرفته است. برآوردگر چگالی احتمال کرنل تک متغیره،  $\hat{f}(x)$ ، یک تابع چگالی احتمال در نقطه دلخواه  $x$  با رابطه زیر تعریف می‌گردد.

$$\hat{f}(x) = \frac{1}{nh} \sum_{l=1}^n K\{(x - x_l)/h\} \quad (2)$$

دراین رابطه  $n$  تعداد مشاهدات،  $x_l$  مشاهده  $l$  ام و  $h$  پارامتر هموار کننده است که به‌عنوان عرض باند شناخته می‌شود و برای هموارسازی شکل تابع چگالی احتمال برآورد شده استفاده می‌گردد. انتخاب عرض باند یک گام مهم در این روش است و تغییر در عرض باند می‌تواند منجر به تغییرات زیادی در شکل برآورد کرنل گردد (آداموسکی، ۱۹۹۶؛ افروموویچ، ۱۹۹۹). برآوردهای پلاگ این<sup>۱</sup> و ارزیابی مقاطع حداقل مربعات از روش‌های برآورد عرض باند هستند (پولانسکی و بیکر، ۲۰۰۰؛ اسکات، ۱۹۹۲؛ تاروتون و همکاران، ۱۹۹۸). با حداقل کردن انحراف تابع چگالی احتمال برآورد شده و تابع چگالی احتمال واقعی می‌توان عرض باند را برآورد کرد. برای ساده‌سازی ریاضی به‌طور معمول می‌توان فرض کرد که تابع چگالی احتمال واقعی نرمال می‌باشد. برای انتخاب

1- plug-in

عرض باند معیاری به‌عنوان معیار خطا انتخاب می‌گردد و سپس بر اساس این معیار خطا عرض باند بهینه انتخاب می‌شود. معیار خطا به‌صورت میانگین مجموع مربعات خطا ( $MISE$ ) در نظر گرفته می‌شود (کارماکار و سیمونوویچ، ۲۰۰۸).

$$MISE = E[\int \{\hat{f}(x) - f(x)\}^2 dx] \approx \{0.25h^4 \sigma_K^2 \int f''(x)^2 dx\} + (1/nh)\alpha \quad (۳)$$

در رابطه فوق  $\alpha = \int K^2(x)dx$  و  $\sigma_K^2$  واریانس تابع کرنل است. عرض باند باید به گونه‌ای انتخاب شود که معیار خطا  $MISE$  حداقل گردد. برآورد  $MISE$  نیاز به مشتق دوم تابع چگالی دارد که نامعلوم است. با فرض آن‌که توزیع نهایی نرمال باشد براساس روش آزالینی (۱۹۸۱) و سیلورمن (۱۹۸۶) عرض باند می‌تواند تعیین شود. با پذیرش این فرض عرض باند بهینه به‌وسیله رابطه زیر مشخص می‌گردد.

$$h_0 = (1.587)\sigma n^{-1/3} \quad (۴)$$

در رابطه فوق  $\sigma = \min\{S, (IQR/1.349)\}$ ،  $S$  انحراف معیار نمونه و  $IQR$  تفاضل چارک سوم و چارک اول می‌باشد. عرض باند برآورد شده برای برآورد تابع چگالی احتمال داده‌ها با استفاده از رابطه ۲ مورد استفاده قرار می‌گیرد. عرض باند با این فرض برآورد می‌گردد که چگالی واقعی نرمال است که ممکن است این فرض صحیح نباشد. برای غلبه بر مشکل برآورد چگالی کرنل روش سری‌های متعامد نرمال (بومن و آزالینی، ۱۹۹۷؛ افروموویچ، ۱۹۹۹؛ گوش و ماژومدار، ۲۰۰۷؛ کارماکار و سیمونوویچ، ۲۰۰۸) برای برآورد تابع چگالی احتمال ناپارامتری مورد استفاده قرار می‌گیرد. روش ناپارامتری سری‌های متعامد نرمال: تابع چگالی احتمال برای نمونه کوچک با استفاده از روش سری‌های متعامد نرمال (یک سری توابع متعامد به‌دست آمده از نمونه) قابل برآورد است. یک سری متعامد نرمال یک دنباله از توابع متعامد  $\Phi_j(x)$  و  $\Phi_s(x)$  است که در روابط ۵ و ۶ صدق می‌کند.

$$\int \Phi_s(x)\Phi_j(x)dx = 0 \quad \forall s \neq j \quad (۵)$$

$$\int \{\Phi_j(x)\}^2 dx = 1 \quad \forall j \quad (۶)$$

یک تابع چگالی تک متغیره تصادفی  $X$  را می‌توان به‌خوبی با سری متعامد نرمال  $\hat{f}_j(x)$  تقریب زد.

$$\hat{f}_J(x) = \sum_{j=0}^J \theta_j \Phi_j(x) \quad (7)$$

که  $J$  برش نامیده می‌شود.  $\Phi_j(x)$ ها توابع سری متعامد نرمال،  $\theta_j$ ها ضرایب متناظر با هر تابع می‌باشند. در این مطالعه توابع کسینوس از زیر مجموعه سری‌های فوری به‌عنوان سری‌های متعامد نرمال انتخاب شده‌اند.

$$\Phi_0(x) = 1 \quad (8)$$

$$\Phi_j(x) = \sqrt{2} \cos(\pi j x), \quad j = 1, 2, 3, \dots \quad (9)$$

تعیین تابع چگالی احتمال بر مبنای سری‌های متعامد نرمال شامل گام‌های زیر است (گوش و ماژومدار، ۲۰۰۷؛ کارماکار و سیمونویچ، ۲۰۰۸):

گام اول: تعیین تکیه‌گاه و مقیاس‌بندی داده‌ها: روش ارائه شده برای تعیین تابع چگالی احتمال با استفاده از سری‌های متعامد نرمال در حالتی معتبر است که دامنه متغیر تصادفی  $[0-1]$  باشد. متغیر تصادفی مورد بررسی ممکن است دامنه دیگری مانند  $[a-b]$  داشته باشد و بنابراین نیاز است که با تبدیل متغیر  $y$ ، دامنه آن به بازه  $[0-1]$  تغییر کند. متغیر تغییر مقیاس شده  $y$  می‌تواند به صورت  $y = [(x-a)/(b-a)]$  در نظر گرفته شود. با توجه به این که  $a, b$  مقادیر کمترین و بیشترین مجموعه داده می‌باشند تضمینی وجود ندارد که داده‌های غیر مشاهده‌ای در این محدوده باشند بنابراین برای داده‌ها تکیه‌گاه تعریف می‌شود. بر اساس روش ارائه شده توسط (افروموویچ، ۱۹۹۹؛ کارماکار و سیمونویچ، ۲۰۰۸) اگر داده‌های مشاهداتی  $x_1, x_2, \dots, x_n$  مرتب شده باشند  $(x_1 \leq x_2 \leq \dots \leq x_n)$  آنگاه:

$$a = (x_1 - \Psi_1) \quad , \quad b = (x_n + \Psi_2) \quad (10)$$

در رابطه فوق  $\Psi_1 = (x_{1+s} - x_1)/s$  و  $\Psi_2 = (x_n - x_{n-s})/s$  است و با فرض این که چگالی نزدیک مرزهای تکیه‌گاه هموار است  $s$  یک مقدار مثبت و صحیح می‌باشد. مقدار پیش فرض  $s$  برابر یک می‌باشد. (کارماکار و سیمونویچ، ۲۰۰۸).

گام دوم: برآورد ضرایب سری‌های متعامد نرمال: با استفاده از روابط ۸ و ۹ می‌توان توابع مناسب در روش سری‌های متعامد نرمال را به دست آورد. ضرایب  $\theta_j$  ذکر شده در رابطه ۷ را می‌توان با رابطه زیر به دست آورد که در آن  $f(y)$  تابع چگالی احتمال متغیر تصادفی تغییر مقیاس داده شده  $Y$  می‌باشد.

$$\theta_j = \int_{-\infty}^{+\infty} f(y)\Phi_j(y)dy \quad (11)$$

به عبارت دیگر  $\theta_j$  امید ریاضی  $\Phi_j(y)$  می‌باشد، یعنی  $\theta_j = E[\Phi_j(y)]$  که از یک نمونه به اندازه  $n$  (یعنی  $j = 0, 1, 2, \dots, n$ ) به صورت زیر برآورد می‌شود.

$$\theta_j = \frac{1}{n} \sum_{l=1}^n \Phi_j(y_l) \quad (12)$$

گام سوم: برآورد برش  $J$ : تعیین برش  $J$  در رابطه ۷ در این روش ضروری است. افروموویچ (۱۹۹۹)  $J$  را به صورت زیر محاسبه نمود.

$$J = \arg \min_{0 \leq j \leq J_n} \sum_{j=0}^j (2\delta_j n^{-1} - \theta_j^2) \quad (13)$$

در رابطه فوق  $\delta_j = \theta_0 + \theta_{2j} 2^{-1/2} - \theta_j^2$  و  $J_n = [C_{J0} + C_{J1} \ln(n)]$  است. ضرایب فوریه  $\theta_j$  با افزایش  $j$  به طور نسبتاً سریع کاهش می‌یابند بنابراین انتخاب  $\hat{\delta} = \hat{\theta}_0$  به عنوان یک برآورد برای تمامی  $\delta_j$ ‌ها در نظر گرفته می‌شود. مقادیر پیش فرض  $C_{J0}$  و  $C_{J1}$  به ترتیب برابر با ۴ و ۰/۵ پیشنهاد شده است (افروموویچ، ۱۹۹۹). رابطه ۱۳ نشان می‌دهد مقداری از  $\hat{J}$  که به ازای آن عبارت  $\sum_{j=0}^j (2\delta_j n^{-1} - \theta_j^2)$  به کمترین مقدار خود برسد به عنوان برش  $J$  در نظر گرفته می‌شود.

گام چهارم: هموارسازی تابع چگالی احتمال برآورد شده: در بسیاری از موارد هموارسازی ضرایب فوریه با ضرب آن‌ها در بعضی مقادیر ثابت که در بازه  $[0-1]$  قرار دارند مورد توجه هستند. پس از هموارسازی، یک تابع چگالی احتمال اصلاح شده به صورت رابطه ۱۴ زیر به دست می‌آید.

$$\hat{f}_J(y) = \sum_{j=0}^J w_j \theta_j \Phi_j(y); \quad 0 \leq y \leq 1, \quad 0 \leq w_j \leq 1 \quad (14)$$



که در آن وزن‌های  $w_j$  برای هموارسازی ضرایب فوریه به صورت زیر تعیین می‌گردند.

$$w_0 = 1 \quad (15)$$

$$w_j = \left[ 1 - \frac{\hat{\delta}}{(n\theta_j^2)} \right]_+, \quad \forall j > 0 \quad (16)$$

تعداد  $j$  تای اول ضرایب فوریه، یک تابع چگالی نیازمند تعداد نسبتاً زیادی از ضرایب برای تخمین مناسب می‌باشد. بنابراین جمله‌هایی با بسامد بالا اضافه خواهند شد. بعد از اضافه کردن جمله‌ها در رابطه ۱۴ برآورد چگالی احتمال به صورت زیر به دست می‌آید.

$$\hat{f}_j(y) = \sum_{j=0}^J w_j \theta_j \Phi_j(y) + \sum_{j=J+1}^{C_{JM} \times J_n} I_{\{\theta_j^2 > C_T \hat{\sigma} \ln(n/n)\}} \theta_j \Phi_j(y); \quad 0 \leq y \leq 1, \quad 0 \leq w_j \leq 1 \quad (17)$$

که  $C_{JM}$  و  $C_T$  پارامترهای بخش‌الحاقی می‌باشند که تعداد بیشترین اجزای شامل شده در برآورد تابع چگالی احتمال را تعریف می‌کند. مقادیر پیش فرض ۶ و ۴ می‌باشند (افرومویچ، ۱۹۹۹).  $I$  یک تابع نشانگر است به طوری که اگر  $A$  درست باشد  $I_{(A)}$  برابر با یک و اگر درست نباشد برابر با صفر می‌باشد.

گام پنجم: تصحیح مساحت زیر تابع چگالی احتمال و مقادیر منفی: پس از برآورد تابع چگالی احتمال این امکان وجود دارد که تابع چگالی احتمال در بعضی نواحی مقداری منفی گردد. برای اصلاح تابع چگالی احتمال از الگوریتم زیر استفاده می‌گردد (گوش و ماژومدار، ۲۰۰۷).

الف- اگر مقادیر منفی در بعضی نقاط وجود دارد بیشترین این مقادیر محاسبه می‌شوند.

ب- حداکثر مقدار منفی یافت شده به  $\hat{f}_j(y)$  اضافه می‌شود. این عمل موجب می‌شود مقادیر تابع در همه دامنه مثبت باشد.

ج- مساحت زیر نمودار تعیین می‌گردد. اگر این مقدار کمتر از واحد باشد مقدار  $c$  به نحوی که رابطه زیر برقرار باشد با استفاده از روش‌های عددی محاسبه شود. مقدار  $\hat{f}_j(y)$  با اضافه کردن مقدار  $c$  به دست آمده از رابطه زیر تصحیح می‌گردد.

$$\int_{-\infty}^{+\infty} [\hat{f}_j(y) + c] dy = 1 \quad (18)$$

د- اگر مساحت زیر نمودار کمتر از واحد باشد مقدار  $c_1$  با استفاده از همین فرآیند مشخص می‌گردد.

$$\int_{-\infty}^{+\infty} [\hat{f}_j(y) + c_1] dy = 1 \quad (19)$$

با کم کردن  $c_1$  از  $\hat{f}_j(y)$  ممکن است تابع چگالی احتمال در بعضی از نقاط دامنه مقادیر منفی به خود بگیرد. در چنین شرایطی فقط بخش مثبت  $\hat{f}_j(y)$  در نظر گرفته می‌شود.

$$\hat{f}_j(y) = [\hat{f}_j(y) - c_1]_+ \quad (20)$$

که  $\hat{f}_j(y) = [\hat{f}_j(y) - c_1]_+ = \max\{0, [\hat{f}_j(y) - c_1]_+\}$  مساحت زیر نمودار دوباره محاسبه می‌شود و اگر برابر با واحد نباشد دوباره باید به گام ج رفته و روند را ادامه داد تا این که مساحت زیر تابع چگالی احتمال واحد گردد.

گام ششم: برآورد تابع چگالی احتمال برای داده‌های مقیاس‌بندی نشده: داده‌های مقیاس‌بندی شده مطابق با چگالی  $f_Y(y)$  توزیع می‌شود که  $y \in [0-1]$  می‌باشد. بنابراین تابع چگالی احتمال برآورد شده متناظر با داده‌های مقیاس‌بندی شده  $y$  در بازه  $[0-1]$  می‌باشد. برآورد  $f_X(x)$  داده‌های متغیر اصلی  $x$  در بازه  $[a-b]$  با توجه به رابطه زیر قابل دستیابی است.

$$f_X(x) = (b-a)^{-1} f_Y(y); \quad x \in [a-b] \quad (21)$$

پس از برآورد تابع چگالی احتمال متغیرهای سیلاب با استفاده از روش ناپارامتری سری‌های متعامد نرمال می‌توان با استفاده از انتگرال عددی تابع توزیع تجمعی هر متغیر را به دست آورد. یکی از مزیت‌های سری‌های متعامد نرمال نسبت به برآورد چگالی کرنل تولید تابع مناسب می‌باشد. هنگامی که  $\theta_j$  محاسبه شد برآورد تابع چگالی احتمال بر روی تمام مقادیر ممکن قابل تعریف می‌باشد.

**نکوی برازش:** توابع توزیع تجمعی منتخب متغیرهای سیلاب برآورد شده توسط روش‌های پارامتری و ناپارامتری با احتمال عدم وقوع تجربی که از روابط ترسیم موقعیت به دست می‌آید مقایسه می‌گردند. در این پژوهش برای دستیابی به احتمال عدم وقوع تجربی از رابطه ترسیم موقعیت گرینگورتن (۱۹۶۳) استفاده شده که به صورت زیر می‌باشد.

$$P_k = \frac{k - 0.44}{N + 0.12} \quad (22)$$

$k$  ردیف مشاهدات در یک ترتیب صعودی،  $N$  تعداد مشاهدات و  $P_k$  فراوانی تجمعی (احتمال این که یک مقدار مشخص از  $k$  امین مشاهده مرتب شده به صورت صعودی کمتر است) می باشد. در این مطالعه ریشه میانگین مربعات خطا ( $RMSE$ )، معیار اطلاعات آکائیک ( $AIC$ ) (آکائیک، ۱۹۷۴) و معیار اطلاعات بیزی ( $BIC$ ) یا معیار شوآرز (شوآرز، ۱۹۷۸) به عنوان معیار نکویی برازش برای توابع توزیع احتمال پارامتری و ناپارامتری متغیرهای سیلاب استفاده شده است که با استفاده از روابط ۲۳-۲۵ تعیین می گردند. در این رابطه  $x_c(i)$  و  $x_o(i)$  مقادیر محاسبه شده و مشاهده شده  $i$ ام می باشند و  $k$  تعداد پارامترهایی می باشد که در برآورد مقادیر محاسباتی استفاده شده است همچنین  $N$  تعداد مشاهدات است (ژانگ و سینگ، ۲۰۰۷).

$$RMSE = (MSE)^{\frac{1}{2}} = \left\{ \frac{1}{N - k} \sum_{i=1}^N [x_c(i) - x_o(i)]^2 \right\}^{\frac{1}{2}} \quad (23)$$

$$AIC = N \log(MSE) + 2k \quad (24)$$

$$BIC = N \log(MSE) + [k \times \log(N)] \quad (25)$$

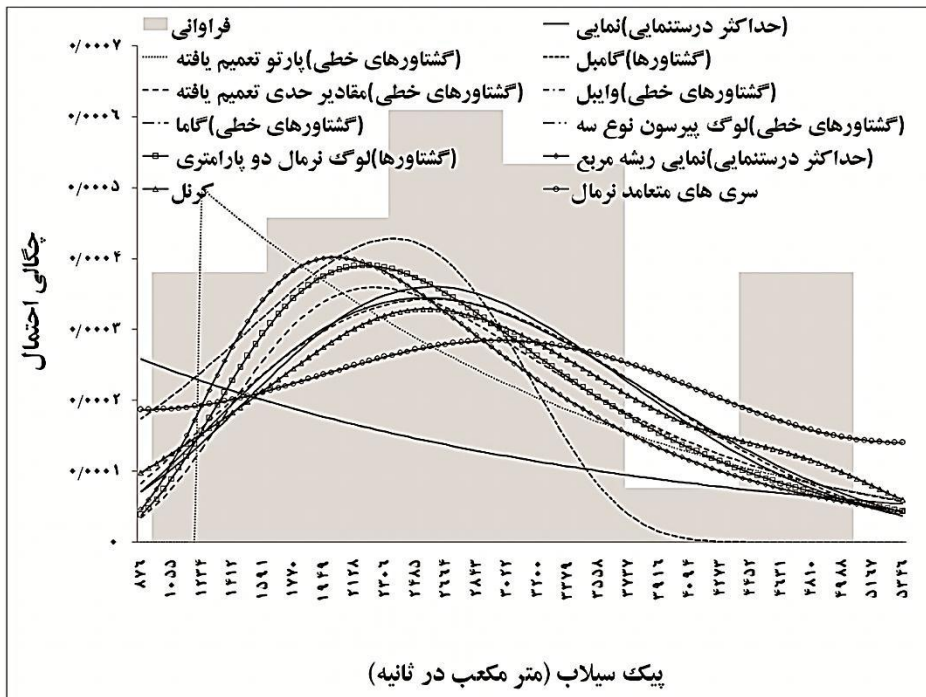
بهترین تابع توزیع احتمال توزیعی است که حداقل معیار اطلاعات آکائیک و حداقل معیار اطلاعات بیزی را دارد. برای بررسی این که آیا داده های نمونه از یک جمعیت یا یک توزیع خاصی می باشند از آزمون  $\chi^2$  (اسندکور و کوچران، ۱۹۸۹) استفاده گردید. یکی از ویژگی های جالب آزمون نکویی برازش  $\chi^2$  این است که می تواند برای هر توزیع تک متغیره که تابع توزیع تجمعی آن قابل محاسبه باشد اعمال گردد. این آزمون بر داده های متغیر مورد بررسی که در دسته ها طبقه بندی می گردند اعمال می شود. مقدار آماره آزمون  $\chi^2$  وابسته به این است که داده های به چه صورت در دسته ها طبقه بندی شده اند. یکی از معایب آزمون  $\chi^2$  این است که نیازمند به اندازه نمونه کافی و مناسب است تا تقریب  $\chi^2$  معتبر باشد. آماره آزمون به صورت رابطه زیر تعریف می شود:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (۲۶)$$

در این رابطه  $O_i$  فراوانی مشاهده شده و  $E_i$  فراوانی موردانتظار برای دسته  $i$  می‌باشد. فراوانی موردانتظار به صورت  $E_i = N[F(Y_u) - F(Y_l)]$  محاسبه می‌شود.  $E$  تابع توزیع تجمعی توزیعی است که آزمون می‌گردد و  $Y_u$  حد بالایی دسته  $i$  و  $Y_l$  حد پایینی دسته  $i$  و  $N$  اندازه نمونه می‌باشد. آماره آزمون تقریباً از توزیع  $\chi^2$  با درجه آزادی  $v = k - p - 1$  (که  $k$  تعداد دسته‌های غیرتهی و  $p$  تعداد پارامترهای برآورد شده است) پیروی می‌کند. اگر  $\chi^2 > \chi^2_{(\alpha, v)}$  سطح معنی‌داری و  $v$  درجه آزادی) باشد فرض این‌که توزیع مورد بررسی بر داده‌ها مناسب است رد می‌شود.

### نتایج و بحث

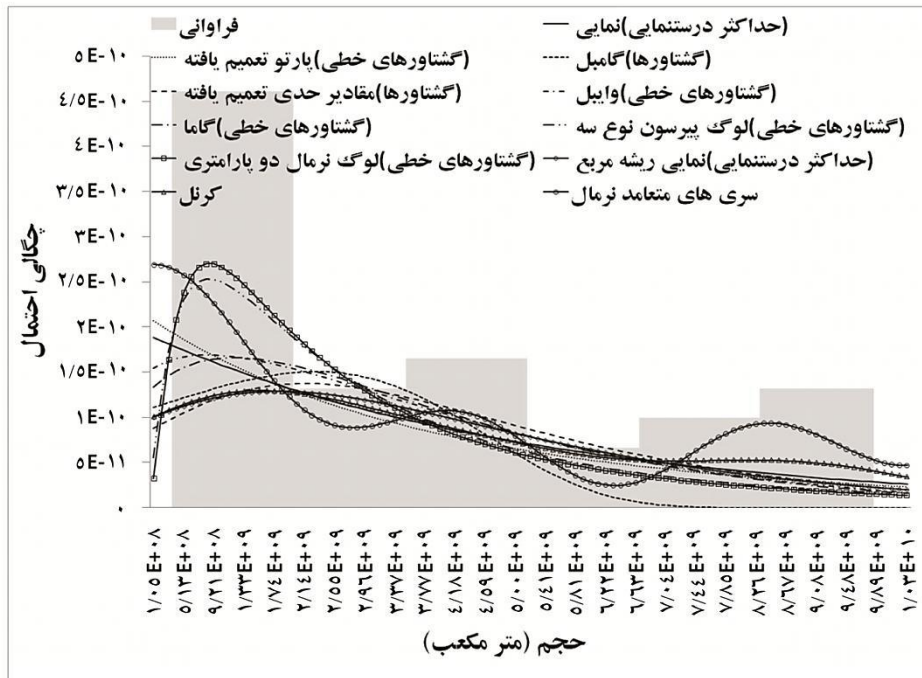
روش‌های مرسوم پارامتری و روش‌های ناپارامتری برآورد چگالی کرنل و سری‌های متعامد نرمال برای برآورد توابع چگالی احتمال متغیرهای سیلاب مورد استفاده قرار گرفت و توابع چگالی احتمال برآورد شده در شکل‌های ۱، ۲ و ۳ نشان داده شده است (با در نظر گرفتن این نکته که ترسیم همه توابع چگالی احتمال برآورد شده در تحلیل این شکل‌ها مشکل ایجاد می‌کند، برای هر توزیع پارامتری فقط بهترین تابع چگالی احتمال برآورد شده رسم شده است). بررسی شکل‌های ۱، ۲ و ۳ نشان می‌دهد که متغیرهای مورد بررسی دو و سه مدی هستند. بررسی مشاهده‌ای برازش توابع چگالی احتمال روشی مرسوم است ولی مقایسه بین توزیع‌هایی که برازش نسبتاً مشابهی دارند دشوار و همراه با خطا می‌باشد. بنابراین برای مقایسه بین توزیع‌های مختلف معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا مدنظر قرار گرفت. جدول ۱ مقادیر این معیارها را در توابع توزیع مختلف به صورت عددی نشان می‌دهد.



شکل ۱- هیستوگرام تجربی داده‌ها و توابع چگالی احتمال پارامتری و ناپارامتری برای متغیر پیک سیلاب.

جدول ۱- مقادیر معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا در روش‌های پارامتری و ناپارامتری برای متغیرهای سیلاب.

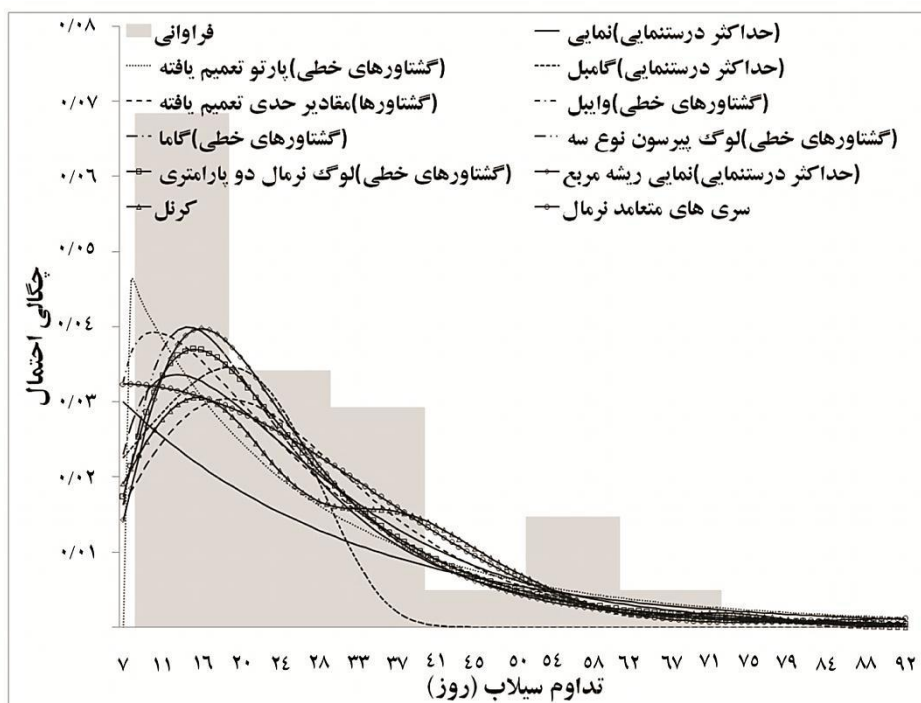
|   | پیک سیلاب |         |        | حجم سیلاب |         |        | تداوم سیلاب |         |        |
|---|-----------|---------|--------|-----------|---------|--------|-------------|---------|--------|
|   | AIC       | BIC     | RMSE   | AIC       | BIC     | RMSE   | AIC         | BIC     | RMSE   |
| نمایی (بیشترین درست‌نمایی)                  | -۱۴۶/۴۶   | -۱۴۶/۴۵ | ۰/۱۸۴۱ | -۲۱۴۰/۰۲  | -۲۱۴/۰۱ | ۰/۰۸۵۲ | -۲۲۶/۲۰     | -۲۲۶/۱۹ | ۰/۰۷۴۲ |
| پارتو تعمیم یافته (گشتاورها)                | -۱۲۳/۵۶   | -۱۲۳/۵۵ | ۰/۲۲۸۳ | -۱۵۹/۶۳   | -۱۵۹/۶۲ | ۰/۱۵۱۴ | -۲۰۶/۱۵     | -۲۰۶/۱۴ | ۰/۰۸۹۱ |
| پارتو تعمیم یافته (گشتاورهای خطی)           | -۲۳۷/۴۳   | -۲۳۷/۴۲ | ۰/۰۶۴۲ | -۲۲۱/۰۷   | -۲۲۱/۰۶ | ۰/۰۷۵۱ | -۲۴۱/۱۵     | -۲۴۱/۱۳ | ۰/۰۵۹۸ |
| گامبل (گشتاورها)                            | -۱۳۱/۵۸   | -۱۳۱/۵۷ | ۰/۲۱۳۲ | -۱۳۱/۷۴   | -۱۳۱/۷۳ | ۰/۲۱۲۸ | -۱۵۸/۲۲     | -۱۵۸/۲۱ | ۰/۱۵۷۴ |
| گامبل (گشتاورهای خطی)                       | -۱۲۸/۳۷   | -۱۲۸/۳۶ | ۰/۲۲۱۲ | -۱۳۰/۸۸   | -۱۳۰/۸۷ | ۰/۲۱۴۹ | -۱۵۷/۸۸     | -۱۵۷/۸۷ | ۰/۱۵۸۰ |
| گامبل (بیشترین درست‌نمایی)                  | -۱۲۸/۴۴   | -۱۲۸/۴۳ | ۰/۲۲۱۰ | -۱۳۰/۸۶   | -۱۳۰/۸۵ | ۰/۲۱۵۰ | -۱۵۸/۹۴     | -۱۵۸/۹۴ | ۰/۱۵۶۱ |
| مقادیر حدی تعمیم یافته (گشتاورها)           | -۲۵۲/۴۹   | -۲۵۲/۴۷ | ۰/۰۵۲۵ | -۲۲۹/۵۰   | -۲۲۹/۴۹ | ۰/۰۶۸۳ | -۲۷۲/۰۵     | -۲۷۲/۰۴ | ۰/۰۴۲۰ |
| مقادیر حدی تعمیم یافته (گشتاورهای خطی)      | -۲۶۶/۴۸   | -۲۶۶/۴۶ | ۰/۰۴۴۸ | -۱۹۴/۲۵   | -۱۹۴/۲۴ | ۰/۱۰۲۰ | -۲۲۹/۷۲     | -۲۲۹/۷۰ | ۰/۰۶۸۱ |
| مقادیر حدی تعمیم یافته (بیشترین درست‌نمایی) | -۲۳۵/۷۶   | -۲۳۵/۷۵ | ۰/۰۵۱۸ | -۱۱۸/۸۳   | -۱۱۸/۸۲ | ۰/۲۴۱۰ | -۱۸۱/۲۰     | -۱۸۱/۱۸ | ۰/۱۱۸۴ |
| وایبل (گشتاورها)                            | -۳۰۲/۸۹   | -۳۰۲/۸۸ | ۰/۰۲۹۶ | -۲۳۰/۰۳   | -۲۳۰/۰۲ | ۰/۰۶۷۸ | -۲۸۳/۸۶     | -۲۸۳/۸۵ | ۰/۰۳۶۷ |
| وایبل (گشتاورهای خطی)                       | -۳۰۶/۷۶   | -۳۰۶/۷۴ | ۰/۰۲۸۳ | -۲۳۹/۶۳   | -۲۳۹/۶۲ | ۰/۰۶۰۸ | -۲۸۸/۰۰     | -۲۸۷/۹۸ | ۰/۰۳۵۰ |
| گاما (گشتاورها)                             | -۳۰۲/۱۵   | -۳۰۲/۱۴ | ۰/۰۳۰۲ | -۲۲۰/۵۰   | -۲۲۰/۴۸ | ۰/۰۷۲۳ | -۲۸۶/۸۵     | -۲۸۶/۸۴ | ۰/۰۳۵۵ |
| گاما (گشتاورهای خطی)                        | -۳۰۷/۷۶   | -۳۰۷/۷۴ | ۰/۰۲۸۰ | -۲۳۶/۲۴   | -۲۳۶/۲۲ | ۰/۰۶۳۲ | -۲۸۹/۸۲     | -۲۸۹/۸۱ | ۰/۰۳۴۳ |
| لوگ پیرسون نوع سه (گشتاورها)                | -۳۰۹/۸۳   | -۳۰۹/۸۱ | ۰/۰۲۷۳ | -۲۲۸/۷۳   | -۲۲۸/۷۱ | ۰/۰۶۸۹ | -۲۸۲/۰۶     | -۲۸۲/۰۴ | ۰/۰۳۷۵ |
| لوگ پیرسون نوع سه (گشتاورهای خطی)           | -۳۱۰/۰۵   | -۳۱۰/۰۳ | ۰/۰۲۷۲ | -۲۳۳/۶۲   | -۲۳۳/۶۱ | ۰/۰۶۵۱ | -۲۸۷/۸۹     | -۲۳۳/۸۸ | ۰/۰۳۵۱ |
| لوگ نرمال دو پارامتری (گشتاورها)            | -۲۹۲/۷۴   | -۲۹۲/۷۳ | ۰/۰۳۴۰ | -۲۲۸/۳۷   | -۲۲۸/۳۶ | ۰/۰۷۰۷ | -۲۸۱/۶۶     | -۲۸۱/۶۵ | ۰/۰۳۸۵ |
| لوگ نرمال دو پارامتری (گشتاورهای خطی)       | -۲۹۲/۳۶   | -۲۹۲/۳۵ | ۰/۰۳۴۱ | -۲۳۳/۲۶   | -۲۳۳/۲۵ | ۰/۰۶۶۹ | -۲۸۸/۴۱     | -۲۸۸/۴۰ | ۰/۰۳۵۷ |
| نمایی ریشه مربع (بیشترین درست‌نمایی)        | -۲۵۸/۰۱   | -۲۵۸/۰۱ | ۰/۰۵۰۵ | -۲۴۴/۴۱   | -۲۴۴/۴۱ | ۰/۰۵۸۹ | -۲۷۱/۳۲     | -۲۷۱/۳۱ | ۰/۰۴۳۴ |
| کرنل  | -۲۷۳/۷۰   | -۲۷۳/۷۰ | ۰/۰۴۴۲ | -۱۶۷/۵۵   | -۱۶۷/۵۵ | ۰/۱۴۸۱ | -۲۰۸/۷۶     | -۲۰۸/۷۶ | ۰/۰۹۲۶ |
| سری‌های متعامد نرمال                        | -۲۶۸/۸۲   | -۲۶۸/۸۲ | ۰/۰۴۹۴ | -۲۹۵/۳۵   | -۲۹۵/۳۵ | ۰/۰۳۴۵ | -۲۹۱/۹۰     | -۲۹۱/۹۰ | ۰/۰۳۴۰ |



شکل ۲- هیستوگرام تجربی داده‌ها و توابع چگالی احتمال پارامتری و ناپارامتری برای متغیر حجم سیلاب.

با بررسی جدول ۱ مشخص می‌گردد که تابع توزیع‌نمایی ریشه مربع (برآوردگر گشتاورهای خطی) در بین روش‌های پارامتری دارای کمترین معیارهای خطا می‌باشد. برای این توزیع مقادیر معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا به ترتیب  $۲۴۴/۴۱$ ،  $۲۴۴/۴۱$  - و  $۰/۰۵۸۹$  محاسبه شده است. توابع توزیع وایل (برآوردگر گشتاورهای خطی) و گاما (برآوردگر گشتاورهای خطی) رتبه‌های بعدی مناسب‌ترین برآورد تابع چگالی احتمال در روش‌های پارامتری می‌باشند. در بین روش‌های ناپارامتری روش سری‌های متعامد نرمال منجر به برآورد بهتری نسبت به روش تابع چگالی کرنل شده است. با توجه به شکل ۲ هیستوگرام تجربی داده‌ها سه مدی می‌باشد که روش سری‌های متعامد نرمال تابع چگالی احتمال را به صورت سه مدی و روش تابع چگالی کرنل به صورت دو مدی برآورد نموده است. روش تابع چگالی کرنل برای این متغیر تابع چگالی احتمال را با دقت مناسب برآورد نکرده است. در مجموع در بین روش‌های پارامتری و ناپارامتری روش ناپارامتری سری‌های متعامد نرمال دارای حداقل معیارهای خطا است. در این روش مقادیر معیار

اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا به ترتیب ۲۹۵/۳۵-، ۲۹۵/۳۵- و ۰/۰۳۴۵ می‌باشد.



شکل ۳- هیستوگرام تجربی داده‌ها و توابع چگالی احتمال پارامتری و ناپارامتری برای متغیر تداوم سیلاب.

باتوجه به جدول ۱ مشخص می‌گردد که تابع توزیع احتمال گاما (برآوردگر گشتاورهای خطی) نسبت به سایر روش‌های پارامتری تابع چگالی احتمال داده‌ها را بهتر برآورد کرده است. مقادیر معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا برای این روش به ترتیب ۲۸۹/۸۲-، ۲۸۹/۸۱- و ۰/۰۳۴۳ بوده است. همچنین توابع توزیع لوگ نرمال دو پارامتری (برآوردگر گشتاورهای خطی) و وایبل (برآوردگر گشتاورهای خطی) نسبت به دیگر توابع توزیع پارامتری مناسب هستند. مقایسه بین روش‌های ناپارامتری نشان می‌دهد که روش سری‌های متعامد نرمال نسبت به روش برآورد چگالی کرنل به مراتب مناسب‌تر است. این برتری با توجه به مقادیر معیارهای خطا به‌طور کامل



مشهود است. مقادیر معیارهای خطا یعنی معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا به ترتیب ۲۹۱/۹۰، ۲۹۱/۹۰- و ۰/۰۳۴۰ می باشد، بنابراین روش سری های متعامد نرمال در مجموع بهترین برآورد تابع چگالی احتمال متغیر تداوم سیلاب را دارا می باشد. با توجه به شکل ۳ داده های تداوم سیلاب دو مدی می باشد و توابع چگالی احتمال برآورد شده با روش های ناپارامتری نیز دو مدی می باشند. در گام بعدی برای بررسی نکویی برازش آزمون  $\chi^2$  مورد استفاده قرار می گیرد. برای محاسبه نکویی برازش بر مبنای آزمون  $\chi^2$  داده ها به  $k$  دسته تقسیم می شوند. در این مطالعه سری زمانی ۳۲ سال متغیرهای سیلاب در ۶ دسته طبقه بندی شده و مورد بررسی قرار می گیرند. جدول ۲ نتایج آزمون  $\chi^2$  برای توابع توزیع احتمال منتخب متغیرهای سیلاب را نشان می دهد. نتایج این جدول بیانگر این نکته است که توزیع های منتخب یعنی توزیع لوگ پیرسون نوع سه برای متغیر پیک سیلاب در سطح معنی داری ۹۷/۵ درصد و تابع توزیع احتمال منتج شده از سری های متعامد نرمال برای متغیرهای تداوم و حجم سیلاب در سطح معنی داری ۹۹/۵ درصد پذیرفته می شوند.

جدول ۲- نتایج آزمون  $\chi^2$  برای سری های زمانی متغیرهای سیلاب.

| نتیجه | مقدار حدی توزیع $\chi^2_{(a,v)}$ در جدول | مقدار $\chi^2$ | تابع توزیع منتخب     | متغیر       |
|-------|--|----------------|----------------------|-------------|
| پذیرش | ۰/۰۵۱ (۲ و ۰/۹۷۵)                        | ۰/۰۲۷          | لوگ پیرسون نوع سه    | پیک سیلاب   |
| پذیرش | ۰/۴۱۲ (۵ و ۰/۹۹۵)                        | ۰/۰۱۶          | سری های متعامد نرمال | حجم سیلاب   |
| پذیرش | ۰/۴۱۲ (۵ و ۰/۹۹۵)                        | ۰/۰۲۸          | سری های متعامد نرمال | تداوم سیلاب |

### نتیجه گیری

روش های مرسوم در تحلیل فراوانی سیلاب در قریب به اتفاق موارد به تحلیل یک متغیره پیک دبی سیلاب می پردازد و دیگر متغیرهای تصادفی مرتبط یعنی حجم و تداوم سیلاب را مدنظر قرار نمی دهد. دیگر محدودیت عمده این روش ها فرض تبعیت متغیر مورد بررسی از خانواده های خاصی از توابع توزیع پارامتری می باشد با در نظر داشتن دو محدودیت عمده نام برده علاوه بر طیف نسبتاً وسیعی از توزیع های پارامتری و روش های تعیین پارامتر، روش های ناپارامتری برآورد چگالی کرنل و سری های تعامد نیز برای تعیین تابع چگالی احتمال متغیرهای سیلاب مدنظر قرار گرفت. معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا به عنوان معیارهای تعیین توزیع

منتخب متغیرهای سیلاب در نظر گرفته شده و آزمون  $\chi^2$  برای بررسی قابلیت پذیرش توابع چگالی احتمال منتخب به کار گرفته شد. براساس معیارهای عددی نکویی برازش، توزیع پارامتری لوگ پیرسون نوع سه بهترین برازش را برای متغیر دبی پیک سیلاب داشته است. تابع چگالی احتمال برآورد شده براساس روش ناپارامتری سری‌های متعامد نرمال نیز مناسب‌ترین برازش را برای متغیرهای حجم و تداوم سیلاب نشان داده است. نتیجه‌گیری فوق براین اساس است که مناسب‌ترین توزیع آن است که بیشترین معیار اطلاعات آکائیک، معیار اطلاعات بیزی و ریشه میانگین مربعات خطا را داشته باشد. نتایج آزمون  $\chi^2$  نیز پذیرش توابع چگالی احتمال منتخب را برای متغیرهای سیلاب تأیید می‌کند. مقایسه دو روش متفاوت ناپارامتری به کار گرفته شده در این پژوهش بر مبنای معیارهای مشاهده‌ای و عددی نشان می‌دهد روش ناپارامتری سری‌های متعامد نرمال نسبت به برآورد چگالی کرنل منجر به نتایج بهتری در برآورد تابع چگالی احتمال متغیرهای سیلاب شده است زیرا این روش می‌تواند تابع چگالی احتمال را بر روی تمامی دامنه مقادیر ممکن برآورد کند (بومن و آزالینی، ۱۹۹۷). بررسی شکل‌های ۲ و ۳ نشان می‌دهد که روش‌های ناپارامتری به خوبی توانسته است تابع چگالی احتمال متغیرهای حجم و تداوم سیلاب را براساس هیستوگرام تجربی داده‌ها به صورت دو و سه مدی برآورد کند در حالی که هیچ یک از توابع توزیع حدی پارامتری رایج توانایی برآورد تابع چگالی احتمال دو یا چند مدی را ندارند. این موضوع با نتایج گوش و ماژومدار (۲۰۰۷)، کارماکار و سیمونویچ (۲۰۰۸) و حقیقت‌جو و همکاران (۲۰۰۹) تطابق دارد. بررسی شکل ۱ نیز نشان می‌دهد اگرچه روش‌های ناپارامتری تابع چگالی احتمال را به صورت یک مدی برآورد کرده‌اند اما این برآورد با هیستوگرام تجربی داده‌ها تطابق خوبی دارد زیرا در حقیقت روش‌های ناپارامتری از میانگین‌های ریاضی برای هموارسازی هیستوگرام تجربی داده‌ها استفاده می‌کند (لال، ۱۹۹۵). جمع‌بندی نتایج گویای این مطلب است که روش ناپارامتری سری‌های متعامد نرمال کارایی مناسبی در برآورد تابع چگالی احتمال متغیرهای سیلاب داشته است که این نتیجه با نتیجه مطالعه کارماکار و سیمونویچ (۲۰۰۸) همخوانی دارد. با توجه به کارایی روش ناپارامتری سری‌های متعامد نرمال در برآورد تابع چگالی احتمال متغیرهای سیلاب در این پژوهش پیشنهاد می‌گردد کارایی این روش برای تحلیل فراوانی متغیرهای سیلاب در رودخانه‌های مختلف دیگر مناطق کشور و با در نظر قرار دادن شرایط متنوع اقلیمی و نیز اندازه‌های متفاوت سری‌های زمانی مورد بررسی قرار بگیرد. همچنین بررسی روش‌های ناپارامتری در

تحلیل فراوانی دیگر متغیرهای هیدرولوژیکی که دارای توزیع‌های تجربی دو یا چند مدی هستند نیز می‌تواند یک زمینه مناسب پژوهشی باشد.

### منابع

1. Adamowski, K. 1985. Nonparametric kernel estimation of flood frequencies. *Water Resour Res.* 21: 11. 1885–1890.
2. Adamowski, K. 1989. A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies. *J. Hydrol.* 108: 295–308.
3. Adamowski, K. 1996. Nonparametric estimation of low-flow frequencies. *J. Hyd Eng.* 122: 1. 46–49.
4. Adamson, P.T., Metcalfe, A.V., and Parmentier, B., 1999. Bivariate extreme value distributions: an application of the Gibbs sampler to the analysis of floods. *Water Resour Res.* 35: 9. 2825–2832.
5. Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans Autom Control.* 19: 6. 716–722.
6. Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika.* 68: 326–328.
7. Bowman, A., and Azzalini, A., 1997. *Applied smoothing techniques for data analysis: the Kernel approach with S-plus illustrations.* Oxford University Press, New York. 198p.
8. Cunnane, C., 1987. Review of statistical models for flood frequency estimation. P49-95, In: V.P. Singh, (ed), *Hydrologic frequency modeling.* Springer, Dordrecht, the Netherlands.
9. Dooge, J.C.E., 1986. Looking for hydrologic laws. *Water Resour Res.* 22: 9. 465-485.
10. Efromovich, S., 1999. *Nonparametric curve estimation: methods, theory and applications.* Springer-Verlag, New York. 411p.
11. Ghosh, S., and Mujumdar, P.P., 2007. Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment. *Water Resour Res.* 43, W07405. doi: 10.1029/2006WR005351.
12. Grimaldi, S., and Serinaldi, F., 2006. Asymmetric copula in multivariate flood frequency analysis. *Adv Water Resour.* 29: 8. 1155–1167.
13. Hall, M.J., 1984. *Urban hydrology.* Elsevier, Barking, UK. 299p.
14. Higgins, J.J., 2004. *Introduction to modern nonparametric statistics.* Thomson Learning, Belmont, CA: Brooks/Cole. 366p.
15. Haghghat Jou, P.H., 2008. A comparative evaluation of parametric and nonparametric statistics in precipitation frequency analysis. PhD. Thesis In Shahid Chamran University, 115p. (In Persian)

16. Haghghat Jou, P.H., Akhoond-Ali, A.M., Behnia, A., and Chinipardaz, R., 2009. A comparison of parametric and nonparametric density functions for estimating annual precipitation in Iran. *Research J. Environmental Sciences*. 3: 1. 62-70.
17. Karmakar, S., and Simonovic, S.P., 2008. Bivariate flood frequency analysis: Part 1. Determination of marginals by parametric and nonparametric techniques. *J. Flood Risk Management*. 1: 190-200.
18. Kidson, R., and Richards, K.S., 2005. Flood frequency analysis: assumptions and alternatives. *Prog Phys Geogr*. 29: 3. 392-410.
19. Kim, T.W., Vald'es, J.B., and Yoo, C., 2003. Nonparametric approach for estimating return periods of droughts in arid regions. *J. Hydrol Eng, ASCE*. 8: 5. 237-246.
20. Kim, T.W., Vald'es, J.B., and Yoo, C. 2006. Nonparametric approach for bivariate drought characterization using Palmer drought index. *J. Hydrol Eng, ASCE*. 11: 2. 134-143.
21. Kite, G.W., 1978. *Frequency and risk analysis in hydrology*. Water resource publications, Fort Collins, CO. 257p.
22. Lall, U., 1995. Nonparametric function estimation: Recent hydrologic contributions. *Reviews of Geophysics, Contributions in Hydrology, U.S. National Report to the IUGG 1991-1994*. 1093-1099.
23. Lall, U., Moon, Y.I., and Bosworth, K., 1993. Kernel flood frequency estimators: bandwidth selection and kernel choice. *Water Resour Res*. 29: 4. 1003-1015.
24. Pinya, M.A.S., Madsen, H., Rosbjerg, D., 2009. Assessment of the risk of inland flooding in a tidal sluice regulated catchment using multi-variate statistical techniques. *Phys Chem Earth*, 34: 10-12. 662-669.
25. Polansky, A.M., and Baker, E.R., 2000. Multistage plug-in bandwidth selection for kernel distribution function estimates. *J. Stat Comput Simul*. 65: 63-80.
26. Rao, A.R., and Hamed, K.H., 2000. *Flood frequency analysis*. CRC, Boca Raton, FL. 376p.
27. Salarijazi, M., Akhond-Ali, A-M., Adib, A., and Daneshkhah, A., 2012. Trend and change-point detection for the annual stream-flow series of the Karun River at the Ahvaz hydrometric station. *African Journal of Agricultural Research*. 7: 32. 4540-4552.
28. Schwarz, G., 1978. Estimating the dimension of a model. *Ann Stat*. 6: 2. 461-464.
29. Scott, D.W. 1992. *Multivariate density estimation, theory, practice, and visualization*. John Wiley, New York. 376p.
30. Sharma, A., Tarboton, D.G., and Lall, U., 1997. Streamflow simulation: a nonparametric approach. *Water Resour Res*. 33: 2. 291-308.

31. Silverman, B.W., 1986. Density estimation for statistics and data analysis, 1st edition. Chapman and Hall, London. 176p.
32. Smakhtin, V.U., 2001. Low flow hydrology: a review. *J. Hydrol.* 240: 147–186.
33. Snedecor, G.W., and Cochran, W.G., 1989. Statistical methods, 8th edition. Iowa State University Press. 503p.
34. Tarboton, D.G., Sharma, A., and Lall U., 1998. Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resour Res.* 34: 1. 107–119.
35. Zhang, L., and Singh, V.P., 2007. Trivariate flood frequency analysis using the Gumbel–Hougaard copula. *J. Hydrol Eng, ASCE.* 12: 4. 431–439.

Archive of SID



Gorgan University of Agricultural  
Sciences and Natural Resources

*J. of Water and Soil Conservation*, Vol. 20(6), 2013

<http://jwsc.gau.ac.ir>

## **Flood variables frequency analysis using parametric and non-parametric methods**

**\*M. Salarijazi<sup>1</sup>, A.M. Akhound-Ali<sup>2</sup>, A. Adib<sup>3</sup> and A.R. Daneshkhah<sup>4</sup>**

<sup>1</sup>Ph.D. Student, Dept., of Hydrology and Water Resources, Shahid Chamran University of Ahvaz, Khuzestan, Iran, <sup>2</sup>Professor, Dept., of Hydrology and Water Resources, Shahid Chamran University of Ahvaz, Khuzestan, Iran, <sup>3</sup>Associate Professor, Dept., of Civil Engineering, Shahid Chamran University of Ahvaz, Khuzestan, Iran, <sup>4</sup>Assistant Professor, Dept., of Statistics, Shahid Chamran University of Ahvaz

Received: 05/27/2012 ; Accepted: 11/27/2012

### **Abstract**

The common methods in the flood frequency analysis are used to examine the flood peak variable while each flood event consists of three naturally random variables: flood's peak, volume and duration. Furthermore, it was essential to assume the variable of interest follows a specific parametric distribution chosen from an appropriate distributions' family. With the aim to overcome these limitations a wide range of the well-known parametric distributions functions and non-parametric methods based on kernel density estimation and orthonormal series approximation is considered to estimate the distribution of flood variables. The investigated variables is extracted from annual maximum flood series at Ahvaz hydrometric station. In order to compare between the fitted parametric and non-parametric distributions to the data, we used the statistical criterions such as Akaike Information Criteria, Bayesian Information Criteria and the Root mean square error. We then used the chi-square goodness of fit test to examine that the acceptability of the chosen distribution in the previous stage. The results showed that the flood peak follows the parametric log Pearson type III distribution function while the non-parametric orthogonal series approximation were the best fit for the flood volume and flood duration. The estimated distributions based on the orthogonal series expansion were able to capture the graphical features of the data and the corresponding fitted densities reproduced the same unimodality or multimodality that the original histogram of data illustrates. This is not the case when one uses the traditional methods to analyze the frequency of flood.

**Keywords:** Frequency analysis, Flood variables, Parametric methods, Kernel density estimation, Orthonormal series.

---

\* Corresponding author; Email: meysam.salarijazi@gmail.com