

ارزیابی عملکرد الگوریتم کلونی زنبور عسل در تحلیل تناوب سیلاب

ستار چاوشی* و کورش شیرانی^۱

(تاریخ دریافت: ۱۳۹۸/۷/۷؛ تاریخ پذیرش: ۱۳۹۹/۲/۱۴)

چکیده

انتخاب تابع توزیع مناسب و تخمین پارامترهای آن دو گام اساسی در برآورد دقیق تناوب سیلاب است. در این مقاله برای بهبود تخمین پارامترهای توابع توزیع، از مفهوم بهینه‌سازی الگوریتم فرا ابتکاری استفاده شده است. هدف از این مقاله استفاده از الگوریتم کلونی زنبور عسل (ABC) در راستای برآورد پارامترهای توابع توزیع رایج در تخمین سیلاب است. در این راستا از سه آماره برازندگی، ضریب همبستگی، ضریب تشخیص و مجذور میانگین مربعات خطا، برای مقایسه عملکرد این الگوریتم با روش‌های متعارف تخمین پارامتر، حداکثر درست‌نمایی (ML)، گشتاورها (MOM) و گشتاور وزنی احتمال (PWM)، استفاده شده است. منطقه مورد مطالعه حوضه آبخیز بابل‌رود در حاشیه جنوبی دریای مازندران است که دارای سوابق وقوع سیلاب است. تعداد شش ایستگاه هیدرومتری دارای آمار مشاهداتی دبی حداکثر سیلاب در منطقه شناسایی و داده‌های آنها در تحلیل تناوب سیلاب در تحلیل تناوب سیلاب در گام، پیرسون نوع ۳، مقادیر حد و لجستیک مورد استفاده قرار گرفت. مقایسه عملکرد روش‌های ABC، ML، MOM و PWM در برآورد سه آماره برازندگی نشان می‌دهد که تابع توزیع پیرسون نوع ۳ دارای بیشترین مقدار میانگین ضریب همبستگی (۰/۹۷) با کمترین دامنه تغییرات (۰/۰۴-۰/۰۵) و تابع توزیع گاما دارای بیشترین مقدار ضریب تشخیص (۰/۹۳) با کمترین دامنه تغییرات (۰/۱۰) و همچنین کمترین مقدار مجذور میانگین مربعات خطا (۸/۶۱) با کمترین دامنه تغییرات (۱۴/۳۳) است. در نهایت با جمع‌بندی نتایج به دست آمده است می‌توان نتیجه گرفت که الگوریتم کلونی زنبور عسل عملکرد بهتری (مقادیر بالاتر ضریب همبستگی و ضریب تشخیص و مقدار کمتر مجذور میانگین مربعات خطا) از سایر روش‌های رایج در تخمین پارامترهای توابع توزیع مطالعاتی در منطقه دارد.

واژه‌های کلیدی: تناوب سیلاب، تخمین پارامتر، الگوریتم فرا ابتکاری، الگوریتم کلونی زنبور عسل

۱. بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان اصفهان، سازمان تحقیقات، آموزش و ترویج کشاورزی، اصفهان، ایران

*: مسئول مکاتبات: پست الکترونیکی: chsatar@gmail.com

مقدمه

وقوع سیلاب به لحاظ عواقب آن همچون تلفات جانی و خسارت مالی همواره مورد توجه و حائز اهمیت بوده است. پیش‌بینی وقوع سیلاب و تخمین دبی اوج سیلاب از جمله موضوعات مهم در هیدرولوژی است که در نهایت با هدف کاهش اثرات مخرب سیلاب انجام می‌گیرد. روش‌های متعددی برای برآورد سیلاب ارائه شده است که از جمله می‌توان به روش‌های آماری، روابط تجربی و شبیه‌سازی هیدروگراف جریان اشاره کرد. روش‌های متعارف تناوب سیل نیاز به مجموعه‌ای از داده‌های حد دارد که از لحاظ آماری مستقل و توزیع یکنواختی در طول دوره آماری داشته باشند. بدین ترتیب معلوم می‌شود که داده‌های مشاهداتی از یک جمعیت بوده، محل و روش نمونه‌برداری تغییر معنی‌داری نداشته است. به‌عنوان مثال حوضه بالادست ایستگاه نمونه‌برداری تحت توسعه شهری قرار نگرفته است. انتخاب یک تابع توزیع مناسب برای برآزش با داده‌های مشاهداتی یکی از گام‌های مهم در تحلیل تناوب سیل است. هر تابع توزیع دارای پارامترهایی (موقعیت، شکل و مقیاس) است که توسط روش‌های متعارف نظیر حداکثر درست‌نمایی، روش گشتاورها، روش گشتاور وزنی احتمال و روش گشتاور خطی تعیین می‌شود. روش‌های مبتنی بر گشتاور ساده‌تر و سریع‌تر هستند اما دقت آنها نسبت به روش حداکثر درست‌نمایی کمتر است. در مقابل روش حداکثر درست‌نمایی مبتنی بر محاسبات ریاضی تکراری نظیر مدل‌های نیوتن-رافسون بوده و زمان‌بر است. در سال‌های اخیر در راستای بهبود تخمین پارامترهای توابع توزیع و در نتیجه برآورد دقیق‌تر تناوب سیل از روش‌هایی مبتنی بر الگوریتم‌های فرا ابتکاری استفاده شده است. رهیافت‌های مبتنی بر الگوریتم‌های فرا ابتکاری در راستای حل مسائل مختلف بهینه‌سازی همچون روابط خطی، غیرخطی، احتمالاتی و دینامیک ابداع شده‌اند. از آنجایی که بسیاری از فرایندهای هیدرولوژیکی نظیر سیلاب دارای مکانیسم پیچیده با ماهیت غیرخطی است، استفاده از این الگوریتم‌ها می‌تواند در مطالعه دقیق‌تر رفتار این پدیده‌ها و

پیش‌بینی آن در آینده مفید باشد. در سال‌های اخیر استفاده از الگوریتم‌های فرا ابتکاری در هیدرولوژی بسیار گسترده بوده و وقایع مختلف هیدرولوژیکی همچون سیل و خشکسالی با استفاده از این الگوریتم‌ها مطالعه شده است. در مطالعات اخیر استفاده از این الگوریتم‌ها در تخمین پارامترهای توابع توزیع مختلف به‌عنوان جایگزینی مناسب برای روش‌های رایج تخمین پارامتر شناخته شده است که در زیر به برخی از آنها اشاره می‌شود.

شین و همکاران (۲۲) به‌منظور توصیف خصوصیات آماری داده‌های هیدرواقلمی، از الگوریتم حداکثر محتمل (Expectation Maximization Algorithm, EM) برای تخمین پارامترهای تابع توزیع نرمال مخلوط (Mixture Normal) استفاده کردند. آنها برای تخمین بهینه پارامترهای این توزیع از الگوریتم فرا ابتکاری ژنتیک با انگیزه حداکثر کردن تابع هدف یعنی لگاریتم احتمال استفاده کردند. آنها همچنین در مطالعه‌ای دیگر برای مدل‌سازی وقایع حد هیدرواقلمی، رهیافت جدیدی ابداع کرده و از مفهوم حداکثر درست‌نمایی مبتنی بر الگوریتم فرا ابتکاری (Meta-heuristic Maximum Likelihood) برای تخمین پارامترهای توابع توزیع ناهمگون مخلوط (Heterogeneous Mixture Distributions) استفاده کردند. آنها روش ابداعی را برای مطالعه بارش‌های حداکثر کره جنوبی به‌کار گرفتند و چنین نتیجه گرفتند که این رهیافت برای مطالعه پدیده‌های هیدرواقلمی با منابع مختلف و در نتیجه رفتار آماری مختلف بسیار کارآمد است (۲۳).

از دیگر کاربردهای الگوریتم‌های فرا ابتکاری در مطالعات هیدرواقلمی می‌توان به استفاده از آنها در مطالعه بارش و باد اشاره کرد. یون و همکاران (۲۶) برای مطالعه دقیق رگبار طرح (Design Storm)، شامل انتخاب بهترین تابع توزیع و تخمین پارامترهای آن، از الگوریتم جستجوی هارمونی استفاده کردند. آنها ۷۴ ایستگاه بارندگی در کره جنوبی را مورد مطالعه قرار داده و نتایج این الگوریتم را با نتایج روش‌های متعارف یعنی

مشاهداتی باد، روش گشتاورها مناسب‌ترین روش تخمین و الگوریتم SSO بهترین الگوریتم از لحاظ معیارهای آزمون برازندگی است. یکی از نکات مهم در مطالعه باد انتخاب روش مناسب برای تقلیل خطای برآورد سرعت باد به‌ویژه در تعیین رژیم باد برای پروژه‌های تولید انرژی است. معمولاً از تابع توزیع ۲ پارامتره ویبول برای تعیین تناوب سرعت باد استفاده می‌شود. در تحقیقی پیرامون استفاده بهینه از انرژی باد در ایالت‌های پرنامبوکو (Pernambuco) و ریوگرانده دوسول (Rio Grande do Sul)، به‌عنوان مهم‌ترین مناطق مستعد برای استفاده از منابع تجدید شونده تولید برق در برزیل، داده‌های سرعت باد در ارتفاع ۵۰ متری زمین در سال‌های ۲۰۰۶ تا ۲۰۱۰ جمع‌آوری و تجزیه و تحلیل شد. در این مطالعه ۴ الگوریتم فرا ابتکاری جستجوی هارمونی، الگوریتم بهینه‌سازی کوکو (Cuckoo Search Optimization, CSO)، بهینه‌سازی ازدحام ذرات و کلونی مورچه به‌کار گرفته شد و نتایج حاصل طبق آماره‌های RMSE، MAE و R^2 با روش‌های حداکثر درست‌نمایی، گشتاورها، روابط تجربی و روش انرژی معادل (Equivalent Energy Method) مقایسه شد. نتایج نشان داد که الگوریتم کلونی مورچه بهترین الگوریتم در تخمین پارامترهای تابع توزیع ویبول در مناطق تریونفو (Triunfo) و سائو مارتینهو دسرا (Sao Martinho da Serra) و الگوریتم CSO مناسب‌ترین تابع توزیع در منطقه پترولینا (Petroлина) است (۹).

همچنین الگوریتم‌های فرا ابتکاری در مطالعه جریان رودخانه، رواناب و سیل مورد استفاده قرار گرفته است. لنگات و همکاران (۱۶) روش‌های موجود در تعیین تابع توزیع مناسب برای تخمین داده‌های جریان حداکثر، حداقل و میانگین را مرور کرده و سپس عملکرد شش تابع توزیع را مقایسه و چهار آزمون برازش، ارزیابی و انتخاب را به تصویر کشیدند. روش حداکثر درست‌نمایی برای آزمون برازندگی استفاده و مناسب‌ترین تابع توزیع برای داده‌های جریان حداکثر سالیانه، که پیرسون نوع ۳ و لوگ نرمال است، انتخاب شد. توابع توزیع ویبول، مقادیر حد و گامبل بهترین تابع توزیع برای تخمین جریان حداقل سالانه

حداکثر درست‌نمایی، گشتاورها و گشتاور وزنی احتمال و همچنین الگوریتم ژنتیک مقایسه کردند. نتایج نشان داد که الگوریتم‌های ژنتیک و جستجوی هارمونی نتایج مشابه و در عین حال دقیق‌تری از روش‌های متعارف ارائه می‌دهد.

مولینا آگوییلار و همکاران (۱۷) در مطالعه بارش‌های کشور مکزیک به ناهمگون بودن منابع وقوع بارش به‌دلیل جبهه‌های سرد هوا و سیکلون‌های تروپیکال، که پیامد آن وقوع بارش‌های شدید نادر است، اشاره کرده و راهکار مناسب برای مطالعه بارش با منابع وقوع مختلف را استفاده از توابع توزیع مخلوط یک متغیره (Univariate Mixed Probability Distribution Functions) می‌دانند. آنها الگوریتم جستجوی هارمونی (Harmonic Search) را برای تخمین پارامترهای این تابع توزیع استفاده کردند. الرشیدی و همکاران (۴) در مطالعه خود به اهمیت برآورد دقیق سرعت باد برای پروژه‌های استحصال انرژی باد اشاره کرده و از روش‌های تحلیل آماری برای برآورد دقیق سرعت باد ۷ ایستگاه سنجش باد در کشور عربستان استفاده کردند. در این مطالعه توابع توزیع ویبول (Weibull)، رایلی (Rayleigh)، گاما (Gamma) و لوگ نرمال (Lognormal) به‌صورت تک به تک و همچنین تلفیقی، و روش‌های حداقل مربعات (Least Square Method)، گشتاورها (Method of Moment)، و حداکثر درست‌نمایی (Maximum Likelihood Estimation Method) برای برآورد پارامترهای توابع توزیع استفاده شد. آنها در این مطالعه ۴ الگوریتم فرا ابتکاری شامل الگوریتم بهینه‌سازی عنکبوت (Social Spider Optimization, SSO)، الگوریتم ژنتیک، الگوریتم بهینه‌سازی ازدحام ذرات (Particle Swarm Optimization, PSO)، الگوریتم بهینه‌سازی فرایند تخمین پارامترهای توابع توزیع منطبق با سرعت حداکثر باد به‌کار بردند. نتایج نشان داد که تابع توزیع ویبول مناسب‌ترین تابع توزیع قابل برازش با داده‌های

الگوریتم ژنتیک و الگوریتم تکاملی تفاضلی (Differential Evolution) اشاره کرده‌اند.

در منطقه مورد مطالعه مطالعات زیادی در زمینه سیلاب انجام گرفته که عمدتاً از روش‌های کلاسیک استفاده شده است. از جمله می‌توان به مقالات عبدالحی (۱)، آهنی و همکاران (۳)، چاوشی و همکاران (۵ و ۶)، فیاض‌بخش و همکاران (۹)، رستمی کامرود و همکاران (۱۸)، ثقفیان و فرازجو (۱۹)، و شعبانلو و همکاران (۲۰) اشاره کرد.

در این مقاله روش جدیدی برای تخمین پارامترهای توابع توزیع احتمالاتی ارائه شده است که مبتنی بر الگوریتم فرا ابتکاری کلونی زنبور عسل است. هدف از این مقاله بهبود تخمین تناوب سیلاب با استفاده از الگوریتم‌های فرا ابتکاری است.

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه یکی از حوضه‌های آبخیز دریای مازندران انتخاب شده است که سابقه وقوع سالیانه سیلاب و همچنین دارای آمار مشاهداتی درازمدت و قابل اعتماد جریان داشته باشد. حوضه آبخیز بابل‌رود در بخش میانی استان مازندران و در ضلع جنوبی شهرستان بابل واقع است. این حوضه در حدفاصل ۹' - ۳۶° تا ۲۱' - ۳۶° عرض شمالی و ۳۶' - ۵۲° تا ۳۷' - ۵۲° طول شرقی واقع است. این حوضه از شمال به شهر بابل و از جنوب به خط الرأس کوه‌های کتیوک، مهدی‌خیل و کرلین از شرق به کوه‌های پورخانی و از غرب به کوه دزو محدود می‌شود. رودخانه اصلی این حوضه بابل‌رود بوده و رودخانه‌های قرآن‌تالار، سجادرود، و کلارود به رودخانه بابل‌رود می‌ریزند. هر یک از سه رودخانه یاد شده دارای سرشاخه‌های متعددی با نام‌های محلی هستند که مهم‌ترین آنها رود آذر، و اسکلیم است که در محل لفورک به هم متصل و رودخانه قرآن‌تالار را تشکیل می‌دهد. وسعت این حوضه

رودخانه تانا (Tana River) در کشور کنیا تعیین شد، درحالی که توابع توزیع مقادیر حد و لوگ نرمال مناسب‌ترین تابع توزیع برای تخمین جریان میانگین سالانه این رودخانه است.

حسن زاده و همکاران (۱۳) از الگوریتم‌های ژنتیک و کلونی مورچه در تخمین پارامترهای توابع توزیع استفاده کردند. آنها نتایج الگوریتم‌های فوق را با نتایج روش‌های متعارف تخمین پارامتر یعنی حداکثر درست‌نمایی، گشتاورها و گشتاور وزنی احتمال مقایسه کردند و نشان دادند که الگوریتم‌های فرا ابتکاری عملکرد دقیق‌تری از روش‌های متعارف دارد.

طلعت اهری و همکاران (۲۵) نسخه جدیدی از الگوریتم کلونی مورچه را توسعه داده و از آن برای تخمین پارامترهای توابع توزیع رایج در تحلیل فراوانی سیل حوضه منتخب در آذربایجان شرقی با ۱۴ ایستگاه هیدرومتری استفاده کردند. در ادامه و به منظور بررسی عملکرد این الگوریتم، آنها نتایج حاصل را با نتایج روش‌های متعارف ML، MOM، PWM مقایسه کردند. توابع توزیع مورد مطالعه شامل ۷ توزیع، مقادیر حد، پیرسون نوع ۳، لوگ نرمال سه پارامتره، لجستیک تعمیم‌یافته، پارتو تعمیم‌یافته، کاپا ۴ پارامتره و وکبی ۵ پارامتره است. معیارهای سنجش مناسب‌ترین تابع توزیع سه آماره برازندگی، ضریب تعیین (Coefficient of Determination, CD)، ضریب عملکرد (Coefficient of Efficient, CE) و جذر میانگین مربعات خطا (Root Mean Square Error, RMSE) است. نتایج بیانگر عملکرد قابل قبول این الگوریتم در تخمین پارامترهای توابع توزیع مورد مطالعه در منطقه است. مقایسه عملکرد نسخه اصلاح شده الگوریتم کلونی مورچه با نسخه استاندارد این الگوریتم بیانگر قدرت و ثبات (Robustness) نسخه اصلاحی در تخمین پارامترهای توابع توزیع است (۲۵).

سان و همکاران (۲۴) مروری جامع بر استفاده از الگوریتم‌های فرا ابتکاری در تخمین پارامتر در سیستم‌های بیولوژی ارائه و به کاربرد مهم‌ترین الگوریتم‌های فرا ابتکاری همچون الگوریتم شبیه‌سازی تبرید (Simulated Annealing, SA)،

جدول ۱. مشخصات ایستگاه‌های هیدرومتری مورد مطالعه

کد	ایستگاه	رودخانه	طول جغرافیایی	عرض جغرافیایی	ارتفاع (متر)	طول آمار
۱۴۰۰۱	شیرگاه	تالار	۵۲-۵۳-۱۰	۳۶-۱۷-۵۷	۲۲۰	۹۴-۱۳۶۵
۱۴۰۱۳	گلوگاه بندپی	سجادرود	۵۲-۳۷-۳۲	۳۶-۱۸-۰۴	۲۰۰	۷۷-۱۳۶۳
۱۴۰۱۵	دیوا	کلارود	۵۲-۳۴-۰۴	۳۶-۱۹-۰۱	۱۶۰	۷۷-۱۳۶۷
۱۴۰۱۷	کشتارگاه	بابلرود	۵۲-۳۹-۴۹	۳۶-۳۲-۴۳	۱	۹۳-۱۳۵۹
۱۴۰۵۵	تمر	بابک	۵۲-۴۵-۱۶	۳۶-۱۳-۴۲	۳۰۶	۹۳-۱۳۸۱
۱۴۰۷۱	پاشاکلا	بابلرود	۵۲-۴۷-۰۴	۳۶-۱۴-۴۳	۱۳۰	۹۳-۱۳۷۲

آزمون همگنی

یکی از رایج‌ترین آزمون‌های آماری برای بررسی وضعیت تصادفی بودن داده‌ها و امکان تعمیم نتایج نمونه به جامعه اصلی، آزمون ران تست (Run-Test) است. با استفاده از این آزمون، همگن بودن داده‌ها بررسی و مشخص می‌شود تا چه حد دنباله‌ای از اعداد به صورت تصادفی گردآوری شده‌اند (۱۲). در این تحقیق آزمون ران تست در سطح معنی‌دار ۵ درصد با استفاده از نرم‌افزار SPSS ver.16 انجام شد.

آزمون استقلال

این آزمون برای بررسی استقلال داده‌های مشاهداتی و تأثیر احتمالی آنها بر یکدیگر انجام می‌گیرد. در واقع هدف از این آزمون آن است که مطمئن شویم داده‌های مشاهداتی مستقل از یکدیگر بوده و اثری بر یکدیگر نداشته باشند (۱۲). در این تحقیق از آزمون والد-وولفویتز (Wald - Wolfowitz) در نرم‌افزار SPSS ver.16 برای بررسی اختلاف معنی‌دار دو نمونه مشاهداتی در سطح معنی‌دار ۵ درصد استفاده شد.

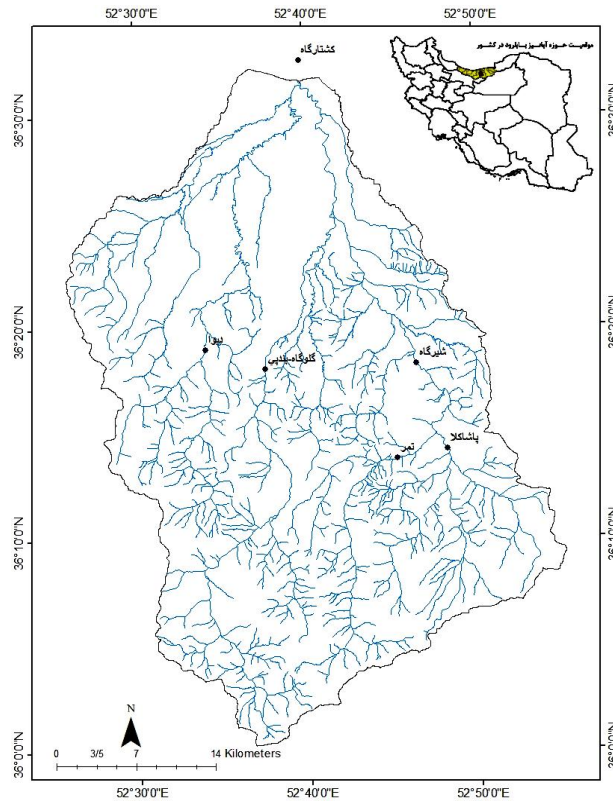
روش تحقیق

در این پژوهش عملکرد الگوریتم کلونی زنبور عسل در تخمین پارامترهای (موقعیت Location، شکل Shape و مقیاس Scale) تابع توزیع مناسب، برای تخمین سیلاب منطقه مورد مطالعه، بررسی شده است. در این راستا توابع توزیع احتمالاتی نظیر نرمال (Normal (NOR)، لجستیک (Logistic (LOG)، پیرسون

۱۴۹۲۹۷ هکتار و شیب متوسط حوضه ۴/۵۷ درصد است. حداقل ارتفاع این حوضه ۸ متر از سطح دریای آزاد در محل خروجی بابلرود از حوضه (پل محمد حسن خان) و بلندترین نقطه این حوضه ۳۷۱۳ در جنوب حوضه (جنوب شرق روستای مهدی خیل) واقع است. حدود سه درصد از سطح حوضه عاری از پوشش گیاهی بوده (ارتفاعات ضلع جنوبی حوضه) و بقیه سطح حوضه را مرتع، جنگل و اراضی زراعی و باغی پوشانده است. طبق مطالعات هواشناسی مبتنی بر آمار ۳۴ ایستگاه هواشناسی موجود در منطقه، متوسط بارندگی سالانه حوضه آبخیز بابلرود ۷۸۵ میلی‌متر، میانگین سالانه دما ۱۳ درجه سانتی‌گراد و تبخیر و تعرق پتانسیل سالانه ۹۹۸ میلی‌متر است. در داخل و پیرامون حوضه ۱۲ ایستگاه هیدرومتری شناسایی و از بین آنها تعداد ۶ ایستگاه که دارای آمار طولانی و قابل اعتماد است مورد مطالعه قرار گرفت (جدول ۱ و شکل ۱).

داده‌های مورد استفاده

در این تحقیق از آمار و اطلاعات هواشناسی و هیدرومتری شرکت تحقیقات منابع آب ایران استفاده شده است که شامل آمار دبی حداکثر جریان سالانه و دبی پیک، در طول دوره آماری موجود و همچنین موقعیت مکانی ایستگاه‌های هیدرومتری است. ایستگاه‌های هیدرومتری با طول دوره آماری ۱۰ سال و بیشتر استفاده شده است (جدول ۱) و مراحل آماده‌سازی داده‌ها شامل آزمون‌های همگنی و استقلال داده‌ها به شرح زیر انجام گرفت (۱۲).



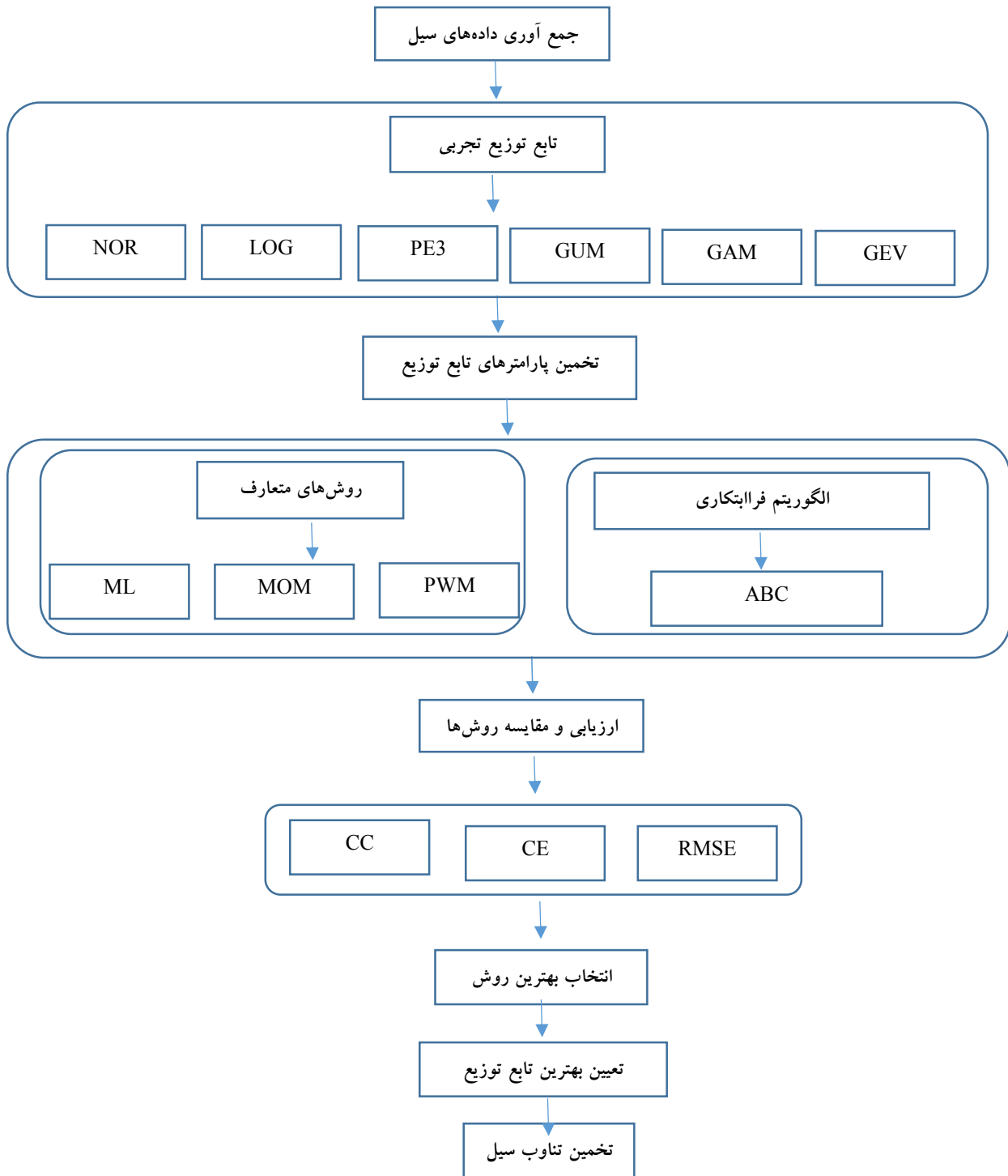
شکل ۱. منطقه مورد مطالعه و پراکنش ایستگاه‌های هیدرومتری

جدول ۲. توابع توزیع احتمالاتی مورد مطالعه

Distribution	Probability Distribution Function	Range
GEV	$f_x(x) = \frac{1}{\alpha} [1 - \beta(\frac{x-u}{\alpha})]^{(\frac{1}{\beta})-1} e^{-\{[1 - \beta(\frac{x-u}{\alpha})]^{1/\beta}\}}$	$-\infty < x < u + \alpha/\beta$
GUM	$f_x(x) = \frac{1}{\alpha} (\frac{x-\gamma}{\alpha}) e^{-\frac{x-\gamma}{\alpha}}$	$-\infty < x < \infty$
PE3	$f_x(x) = \frac{ \beta }{\Gamma_\alpha} [\beta(x-\gamma)]^{\alpha-1} e^{-\beta(x-\gamma)}$	$\gamma < x < \infty$
NOR	$f_x(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$-\infty < x < \infty$
LOG	$f_x(x) = \left(\frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma(1 + e^{-\frac{x-\mu}{\sigma}})^2} \right)$	$-\infty < x < \infty$
GAM	$f_x(x) = \frac{1}{\lambda^K \Gamma(K)} x^{K-1} e^{-\frac{x}{\lambda}}$	$0 < x < \infty$

جدول ۲ آورده شده است. با استفاده از روش‌های تخمین پارامتر تابع توزیع یعنی حداکثر درست‌نمایی (Maximum Likelihood (ML)، گشتاورها (Moments (MOM)

نوع ۳ Pearson Type 3 (PE3)، گمبل (GUMBEL (GUM)، گاما (Gamma (GAM) و مقادیر حد General Extreme Value (GEV) مورد مطالعه قرار گرفته که مشخصات آنها در



شکل ۲. روند انجام تحقیق

عسل برای تخمین پارامترهای توابع توزیع نامبرده استفاده شد. مقادیر چندک‌های حاصل از آزمون هر یک از روش‌های متعارف و الگوریتم فرا ابتکاری مورد استفاده تعیین شد. معادله

و گشتاور وزنی احتمال Probability Weighted Moments (PWM)، مقادیر پارامترهای هر تابع توزیع تعیین شد (۱۱). در ادامه و مطابق با روند نمای (شکل ۲) الگوریتم کلونی زنبور

زیر است:

$$L(\alpha_1, \alpha_2, \dots, \alpha_k) = \prod_{i=1}^n f(x_i, \alpha_1, \alpha_2, \dots, \alpha_k) \quad (1)$$

اگر این نمونه‌ها، هم توزیع باشند به این معنی خواهد بود که تابع احتمال همه یکسان است. در چنین حالتی، نمونه تصادفی را مستقل و هم توزیع می‌نامند (۵ و ۶).

روش گشتاورها

در روش گشتاورها پارامترهای یک تابع توزیع احتمالاتی با بسط رابطه بین گشتاورهای نمونه به گشتاورهای تابع توزیع مربوطه به دست می‌آید. برای تابع توزیعی با تعداد k پارامتر $\alpha_1, \alpha_2, \dots, \alpha_k$ ، اولین گشتاور نمونه به اندازه k معادل با گشتاورهای جمعیت با پارامترهای نامعلوم در نظر گرفته می‌شود. این تعداد k معادله به طور همزمان و برای پارامترهای ناشناخته $\alpha_1, \alpha_2, \dots, \alpha_k$ حل می‌شوند. برای یک تابع توزیع با احتمال تجمعی $f(x)$ PDF گشتاور رتبه r عبارتست از (۱۱):

$$\mu'_r = \int_{-\infty}^{+\infty} x^r f(x) dx \quad (2)$$

گشتاورهای مرکزی مربوطه عبارت است از:

$$\mu_r = \int_{-\infty}^{+\infty} (x - \mu'_1)^r f(x) dx \quad (3)$$

روش گشتاور وزنی احتمال

گشتاور وزنی احتمال برای متغیر تصادفی x و تابع تجمعی مربوطه به شکل زیر بیان می‌شود (۱۱):

$$M_{p,r,s} = E \left\{ x^p [F(x)]^r [1 - F(x)]^s \right\} = \int_0^1 x^p [F(x)]^r [1 - F(x)]^s dF \quad (4)$$

که در آن p, r, s اعداد حقیقی هستند. این معادله در شرایط خاصی که $s=0$ و $p=1$ است به شکل زیر خواهد بود:

$$\beta r = M_{1,r,0} = E \{ x^r [F(x)]^r \} = \int_0^1 x^r [F(x)]^r dF \quad (5)$$

$$r = 0, 1, 2, \dots$$

در این معادله βr تابع خطی از متغیر x برای تخمین پارامترهای یک توزیع است

ویبول (Weibull Plotting Position) برای تعیین دوره‌های بازگشت روش‌های مورد آزمون استفاده شده است (۱۰). در نهایت برای ارزیابی عملکرد این الگوریتم، آزمون برازندگی (Goodness of Fit Test) شامل ضریب همبستگی (Correlation Coefficient (CC)، ضریب تشخیص (Coefficient of Efficiency (CE) و مجذور مربعات میانگین خطا (Root Mean Square Error (RMSE) در مورد هریک از روش‌های فوق انجام و نتایج حاصل با یکدیگر مقایسه شد (۱۲). در این تحقیق از نرم‌افزار متلب Matlab ver. 2016b برای توسعه مدل بهینه‌سازی مبتنی بر الگوریتم کلونی زنبور عسل استفاده شد و سایر پردازش‌های آماری در محیط R-Studio انجام شد.

روش‌های تخمین پارامترهای توزیع احتمالاتی

یکی از مهم‌ترین مسائل در تحلیل فراوانی یک واقعه طبیعی، برآورد پارامترهای تابع توزیع احتمالاتی است که به داده‌های مشاهداتی برازش مناسب دارد. رایج‌ترین روش‌های تخمین پارامترهای توابع توزیع شامل روش حداکثر درست‌نمایی، روش گشتاورها، و روش گشتاور وزنی احتمال است (۵ و ۶).

روش حداکثر درست‌نمایی

اصل درست‌نمایی بیان می‌کند (با فرض اینکه مدل مورد نظر درست باشد) همه اطلاعاتی را که داده‌ها درباره پارامترها دربرمی‌گیرند در تابع درست‌نمایی نهفته است. روش حداکثر تابع درست‌نمایی بر این اصل استوار است که پارامترهای توزیع چنان برآورده شوند که به ازای آمار داده شده تابع درست‌نمایی حداکثر باشد. از آنجایی که تابع درست‌نمایی بیشتر به صورت توان است و از طرف دیگر حداکثر کردن لگاریتم یک تابع با حداکثر کردن خود تابع هم‌ارز است لذا برای سهولت محاسبات، حداکثر درست‌نمایی روی لگاریتم تابع انجام می‌گیرد. برای یک توزیع مورد نظر با توزیع تجمعی $f(x)$ PDF و پارامترهای $\alpha_1, \alpha_2, \dots, \alpha_k$ تابع احتمالاتی به شکل

۵. در صورتی که پس از تعدادی تلاش یک منبع غذایی بهتر در مجاورت این منبع غذایی یافت نشود، یک منبع غذایی اولیه به طور تصادفی توسط یک زنبور پیشاهنگ و طبق معادله ۱ تعیین می‌شود.

۶. مراحل ۲ تا ۵، تا وقتی که جواب بهینه به دست آید، تکرار می‌شود.

در این تحقیق الگوریتم کلونی زنبور عسل برای تخمین پارامترهای توابع توزیع استفاده می‌شود. برای دستیابی به نتایج بهینه، تابع هدف زیر استفاده می‌شود که مبتنی بر اختلاف مقادیر مشاهداتی و تخمینی جریان است (۱۳):

$$\text{Minimize}(O_b = \sum_{i=1}^n \frac{(x_i - \bar{X}_i)^2}{x_i}) \quad (10)$$

که در آن O_b تابع هدف، x_i مقدار مشاهداتی نمونه i ام، \bar{X}_i مقدار تخمینی نمونه i ام از تابع توزیع تجربی است.

شاخص‌های ارزیابی عملکرد

به منظور ارزیابی کفایت و دقت روش‌های مورد استفاده در تخمین پارامتر توابع توزیع مورد نظر از سه آماره ضریب همبستگی (CC)، ضریب تشخیص (CE) و مجذور میانگین مربعات خطا (RMSE) به شرح زیر استفاده شده است (۲۵).

$$CC = \frac{\sum_{i=1}^n [(x_i - \bar{x})(X_i - \bar{X})]}{[\sum_{i=1}^n (x_i - \bar{x})^2 (X_i - \bar{X})^2]^{0.5}} \quad (11)$$

$$CE = 1 - \frac{\sum_{i=1}^n (x_i - X_i)^2}{\sum_{i=1}^n (x_i - \bar{X})^2}, -\infty \leq CE \leq 1 \quad (12)$$

(۱۳)

$$RMSE = [\frac{1}{n} \sum_{i=1}^n (x_i - X_i)^2]^{0.5}, 0 \leq RMSE \leq \infty$$

به طوری که X_i ، داده‌های مشاهداتی؛ x_i ، داده‌های تخمینی؛ \bar{x} و \bar{X} به ترتیب میانگین مقادیر مشاهداتی و تخمینی دبی اوج سیلاب و n تعداد داده‌ها است. مقادیر بالای ضریب همبستگی و ضریب تشخیص و مقادیر پایین مجذور میانگین مربعات خطا عملکرد مناسب روش مورد نظر است.

الگوریتم کلونی زنبور عسل (Artificial Bee Colony (ABC)

الگوریتم کلونی زنبور عسل مصنوعی، یک راهکار بهینه‌سازی است که رفتار یک کلونی زنبور عسل را شبیه‌سازی می‌کند و برای اولین بار در سال ۲۰۰۵ توسط کارابوگا (۱۴)، برای مسائل بهینه‌سازی ارائه شد. در این الگوریتم، یک منبع غذایی به عنوان هدفی در فضای جستجو تعریف می‌شود (یک راهکار برای مسئله بهینه‌سازی)، به طوری که تعداد منابع غذایی در ابتدا برابر با تعداد زنبورهای موجود در کندو است. کیفیت منابع غذایی توسط مقدار تابع هدف در آن موقعیت (میزان برازندگی) تعیین می‌شود.

مراحل اجرای الگوریتم به شرح زیر است (۱۴):

۱. تعیین اولیه منابع غذا توسط رابطه معادله ۱

$$x_{ij} = x_j^{\min} + \text{rand}(0.1)(x_j^{\max} - x_j^{\min}) \quad (6)$$

به طوری که $i=1,2,3..SN$ ؛ $j=1,2,3..D$ ؛ SN تعداد منابع غذا؛ D تعداد پارامترها؛ x_j^{\max} و x_j^{\min} به ترتیب حداقل و حداکثر مقدار پارامتر j است.

۲. حرکت زنبورهای کارگر به سمت منابع غذا و تعیین منابع غذایی مجاور طبق معادله ۲:

$$V_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (7)$$

در این معادله j و k به ترتیب پارامترهای منتخب تصادفی و همسایگان تصادفی، ϕ_{ij} یک عدد تصادفی بین -1 و 1 است.

۳. تعیین یک منبع غذایی جدید به نام V_i و حذف x_i در صورتی که مقدار شهد منبع جدید V_i (مقدار برازندگی) بهتر از منبع قبل x_i باشد.

۴. حرکت زنبورهای جستجو به سمت منابع غذا طبق معادلات ۳ و ۴ و تعیین همسایه‌های جدید طبق مراحل ۲ و ۳:

$$p_i = \frac{\text{fitness}_i}{\sum_{j=1}^{SN} \text{fitness}_j} \quad (8)$$

$$\text{fitness}_i = \begin{cases} \frac{1}{1 + \text{fit}_i} & \text{fit}_i \geq 0 \\ 1 + \text{abs}(\text{fit}_i) & \text{fit}_i < 0 \end{cases} \quad (9)$$

به طوری که fit_i ، هزینه (مقدار تابع هدف) برای منبع غذا X_i و fitness_i مقدار شهد (برازش) منبع غذا X_i است.

نتایج و بحث

در این تحقیق نخست داده‌های دبی حداکثر سالانه ایستگاه‌های موجود در منطقه جمع‌آوری و پس از اصلاح و بررسی اولیه صحت و درستی داده‌ها مورد استفاده قرار گرفت. با استفاده از روش‌های رایج تخمین پارامتر یعنی حداکثر درست‌نمایی، گشتاورها و گشتاور وزنی احتمال، مقادیر پارامترهای (موقعیت، شکل و مقیاس) توابع توزیع نرمال، لوجستیک، پیرسون نوع ۳، گمبل، گاما و مقادیر حد در ایستگاه‌های مورد مطالعه تخمین زده شد (جدول ۳).

در ادامه الگوریتم کلونی زنبور عسل برای تخمین پارامترهای توابع توزیع در ایستگاه‌های مطالعاتی به کار گرفته شد. این فرایند با توسعه کدهای بهینه‌سازی مربوط به هر تابع توزیع در محیط متلب انجام گرفت. پارامترهای الگوریتم مورد استفاده در جدول ۴ آورده شده است.

به‌عنوان مثال فرایند بهینه‌سازی الگوریتم کلونی زنبور عسل در ایستگاه کشتارگاه (۱۴۰۱۷) در شکل ۳ نشان داده شده است. همان‌طور که در این شکل دیده می‌شود مقدار خطا در تکرارهای اولیه به سرعت کاهش یافته و سپس روند کاهش به صورت پلکانی ادامه می‌یابد. این روند تا تکرار ۳۷ ادامه یافته و سپس خطا به مقدار ثابت می‌رسد. به عبارت دیگر تعداد تکرار متناسب با روند بهینه‌سازی انتخاب و الگوریتم توانسته است در بازه تعداد تعریف شده تکرار به جواب قطعی برسد.

فرایند بهینه‌سازی توسط این الگوریتم برای کلیه توابع توزیع مطالعاتی در ایستگاه‌های مورد مطالعه انجام گرفت و مقادیر خطای حاصل برای هر تابع توزیع در هر ایستگاه به دست آمد (جدول ۵). به‌عنوان مثال در ایستگاه کشتارگاه (۱۴۰۱۷) توابع توزیع گاما، مقادیر حد و پیرسون نوع سه، کمترین مقدار خطای مدل را نشان می‌دهند.

در مجموع با مقایسه مقادیر خطای حاصل از توابع توزیع مختلف در ایستگاه‌های مطالعاتی، می‌توان نتیجه گرفت که توابع توزیع گاما و مقادیر حد، کمترین خطای مدل را نشان می‌دهند. مقادیر تخمینی پارامترهای (موقعیت، شکل و

مقیاس) توابع توزیع مختلف در ایستگاه‌های مطالعاتی با استفاده از الگوریتم کلونی زنبور عسل در جدول ۳ نشان داده شده است.

نتایج آزمون برازندگی

به‌منظور بررسی عملکرد الگوریتم کلونی زنبور عسل در تخمین بهینه پارامترهای توابع توزیع مختلف در ایستگاه‌های مورد مطالعه در منطقه، نتایج این الگوریتم (ABC) با نتایج حاصل از روش‌های حداکثر درست‌نمایی (ML)، گشتاورها (MOM) و گشتاور وزنی احتمال (PWM) مقایسه شد. با استفاده از معادلات ۱۱ تا ۱۳ مقادیر آماره‌های ضریب همبستگی (CC)، ضریب تشخیص (CE) و مجذور میانگین مربعات خطا (RMSE) برای هر تابع توزیع در هر ایستگاه به‌دست آمد (شکل‌های ۴، ۵ و ۶).

شکل ۴ میانگین و دامنه تغییرات ضریب همبستگی توابع توزیع مختلف در منطقه مطالعاتی با استفاده از روش‌های ABC، ML، MOM و PWM را نشان می‌دهد. به‌عنوان مثال عملکرد الگوریتم ABC نشان می‌دهد که توابع توزیع گاما و پیرسون نوع ۳ دارای بیشترین مقدار میانگین ضریب همبستگی (۰/۹۷) با کمترین دامنه تغییرات (۰/۰۵) است. به عبارت دیگر طبق نتایج این روش و مطابق با عامل ضریب همبستگی، این دو تابع مناسب‌ترین توابع توزیع در منطقه است. طبق روش‌های ML، MOM و PWM، تابع توزیع پیرسون نوع ۳ دارای بیشترین مقدار ضریب همبستگی (۰/۹۷) با کمترین دامنه تغییرات (۰/۰۴) در منطقه است. مقایسه عملکرد چهار روش فوق نشان می‌دهد که تابع توزیع پیرسون نوع ۳ دارای بیشترین مقدار ضریب همبستگی با کمترین دامنه تغییرات در منطقه بوده و مناسب‌ترین تابع توزیع منطقه است. تفاوت معنی‌داری بین چهار روش فوق از لحاظ آماره ضریب همبستگی وجود ندارد.

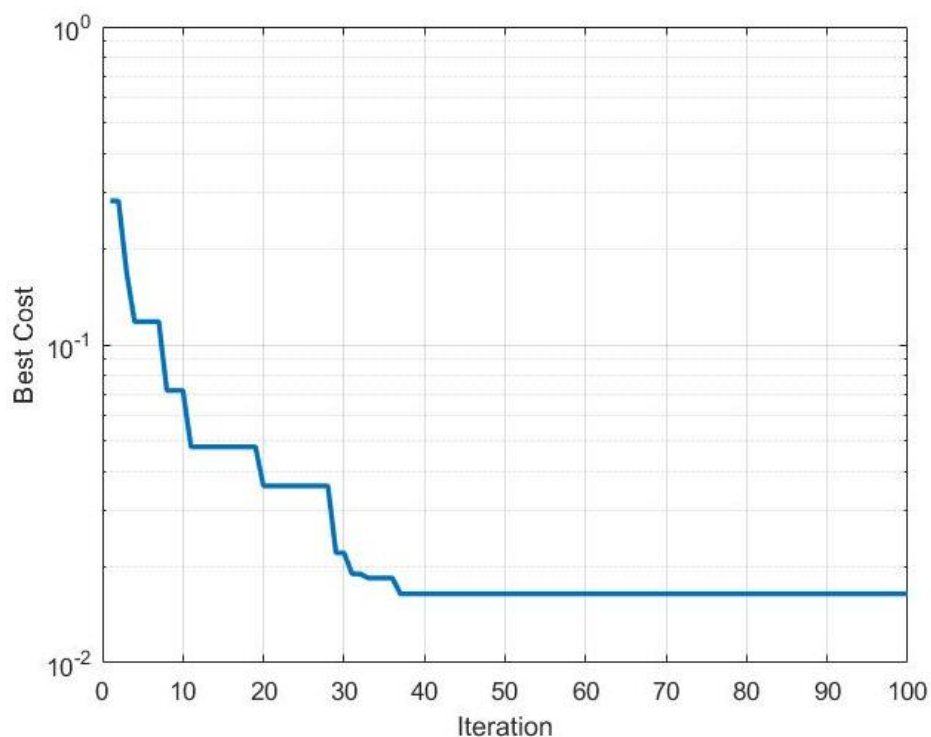
شکل ۵ میانگین و تغییرات آماره ضریب تشخیص در منطقه را نشان می‌دهد. در این شکل طبق الگوریتم ABC تابع توزیع گاما دارای بیشترین مقدار میانگین ضریب تشخیص (۰/۹۳) با

جدول ۳. مقادیر تخمینی پارامترهای توابع استفاده روش‌های مورد مطالعه

PE3			NOR			LOG			GEEV			GAM			GUM				
Scale	Shape	Location	Scale	Shape	Location	Scale	Shape	Location	Scale	Shape	Location	Scale	Shape	Location	Scale	Shape	Location	Site	Method
۴۷/۳۷	۲/۰۱	۶۹/۶۱	۳۴/۸۷	-	۶۹/۳۷	۲۶/۹	-۰/۰۲	۵۲/۳۲	۱۷/۱۷	۲/۰۵	-	۲۸/۰۸	-	۵۲/۲۶	۱۴۰۰۱	-	-	-	-
۳۷/۱۷	۲/۳۴	۳۸/۳	۳۰/۰۱	-	۳۷/۳۷	۱۳/۸۴	-۰/۲۷	۲۰/۷۵	۱۹/۸۴	۱/۸۹	-	۱/۸/۱	-	۲۶/۷۵	۱۴۰۱۳	-	-	-	-
۹/۳۸	۲/۱۵	۱۹/۹۵	۷/۵۵	-	۱۹/۶۳	۲/۸۵	-۰/۲۵	۱۵/۸۱	۲/۳۹	۸/۲۱	-	۲/۹/۸	-	۱۶/۴۴	۱۴۰۱۵	-	-	-	ML
۱۴/۸۲	۱/۴۸	۲۵-۲/۲۸	۱۳۲/۰۲	-	۲۵-۲/۲۶	۷۴/۸۴	-۰/۰۹	۱۸۵/۶۸	۶/۸/۴۴	۳/۶۷	-	۱۰۰/۵۷	-	۱۹-۰/۱۲	۱۴۰۱۷	-	-	-	-
۴۲/۴۴	۲/۸۲	۴۲/۱۴	۳۵/۳۲	-	۴۲/۳۷	۱۷/۵۳	-۰/۵۲	۳۳/۲۹	۲۲/۳۱	۱/۹۲	-	۲۰/۸۱	-	۲۸/۰۵	۱۴۰۲۵	-	-	-	-
۳۶/۸۸	۲/۳۴	۴۲/۱۸	۳۳/۶۵	-	۴۸-۰/۶	۱۷/۵۹	-۰/۲۲	۴۲/۵۴	۲۰/۸۲	۲/۲۲	-	۳۲/۲۴	-	۳۳/۵۹	۱۴۰۷۱	-	-	-	-
۲/۶۶	-۱/۰۷	۱۶/۴۲	۳۵/۴۴	-	۶۹/۴۷	۱۹/۲۲	-	۵۲/۸۱	۱۷/۵۰	۳/۹۷	-	۱۷/۱۹	-	۵۲/۷۸	۱۴۰۰۱	-	-	-	-
۲/۸۳	۱/۴/۸۹	۱۸/۴۷	۳۱/۰/۶	-	۳۷/۳۷	۱۶/۵۵	-۰/۰۱	۳۳/۳۵	۲۴/۱۱	۱/۵۵	-	۳۲/۶۰	-	۳۳/۸۷	۱۴۰۱۳	-	-	-	-
۲/۹۲	۶/۱/۶	۴/۶۱	۷/۸۹	-	۱۹/۶۳	۴/۱/۶	-۰/۰۱	۱۶/۰۸	۲/۹۰	۶/۷/۶	-	۵/۸۷	-	۱۶/۳۳	۱۴۰۱۵	-	-	-	MOM
۵/۹۵	۷/۶۸۲	۵۲/۹۷	۱۳۴/۰۷	-	۲۵-۲/۲۶	۷۲/۸۹	-	۲۵-۲/۲۷	۶۹/۶۵	۳/۵۹	-	۱۰۰/۲/۴	-	۱۹-۰/۸۴	۱۴۰۱۷	-	-	-	-
۲/۱/۶	۱/۱/۶۵	۲۴/۹/۶	۳/۵۵	-	۴۲/۳۸	۱۹/۴۷	-۰/۰۳	۲۵/۶۹	۲۹/۴۴	۱/۴۴	-	۳۷/۵۴	-	۲۶/۴۸	۱۴۰۲۵	-	-	-	-
۲/۵۹	-۱/۲۵	۲۱/۳/۶	۳۴/۳۷	-	۴۸-۰/۶	۱۸/۵۵	-۰/۰۲	۳۳/۴۹	۳۳/۵۶	۲/۰/۴	-	۲۶/۲۴	-	۳۳/۹۲	۱۴۰۷۱	-	-	-	-
۳۵/۴۹	۰/۹۵	۶۹/۴۷	۳۴/۴۹	-	۶۹/۵۱	۱۸/۶۸	-	۵۲/۵۲	۱۸/۲۹	۳/۷۹	-	۲۸/۰۸	-	۵۲/۳۷	۱۴۰۰۱	-	-	-	-
۳۳/۳۰	۲/۴۰	۳۷/۳۷	۲۸/۹۴	-	۳۷/۴۷	۱۲/۳۸	-۰/۳۳	۲۰/۹۷	۲۶/۸۷	۱/۳۹	-	۳۲/۵۶	-	۳۳/۸۷	۱۴۰۱۳	-	-	-	-
۸/۳۲	۲/۰/۶	۱۹/۶۳	۷/۳۲	-	۱۷/۴۳	۳/۳۷	-۰/۲۵	۱۵/۶۱	۲/۸۳	۶/۹۴	-	۵/۹/۶	-	۱۶/۱/۹	۱۴۰۱۵	-	-	-	-
۱۲/۶۹	۱/۲۱	۲۵-۲/۲۶	۱۳۲/۵۰	-	۲۵-۲/۲۶	۶۹/۹۱	-	۲۲/۱۱	۱۰-۲/۰/۹	۳/۳۱	-	۱۰۰/۸/۴	-	۱۸/۸۰/۱	۱۴۰۱۷	-	-	-	-
۳۹/۹۷	۲/۵۹	۴۲/۳۸	۳۲/۸۸	-	۴۲/۳۸	۱۳/۴۱	-	۳۰/۴۱	۳۰/۴۴	۱/۳۹	-	۲۶/۸۷	-	۲۶/۹۳	۱۴۰۲۵	-	-	-	-
۳۵/۳۸	۱/۶۷	۰/۸۴	۳۲/۵۰	-	۴۸-۰/۶	۱۶/۰/۵	-	۳۹/۹۱	۲۵/۰/۷	۱/۹۲	-	۲۶/۲۶	-	۳۳/۸۹	۱۴۰۷۱	-	-	-	-
۳۳/۹/۶	۲/۶۴	۷/۷۰	۳/۶/۸۷	-	۶۹/۵۳	۲۱/۳۷	-۰/۱۵	۲۱/۸/۶	۲۱/۰/۲	۳/۲۱	-	۳۱/۰/۴	-	۵۲/۹۲	۱۴۰۰۱	-	-	-	-
۴۱/۲۲	۷/۰/۰	۱۱/۰/۷	۳۲/۱/۵	-	۳۷/۳۸	۱۸/۸/۶	-	۳۹/۳۷	۴۰/۵۱	۰/۹/۹	-	۳۷/۹۹	-	۳۳/۰/۵	۱۴۰۱۳	-	-	-	-
۱۲/۱۳	۰/۸۷	۱۱/۴/۹	۸/۲۲	-	۱۹/۶۳	۴/۹/۷	-	۱۹/۶۳	۴/۳۳	۴/۶۳	-	۷/۲/۹	-	۱۵/۹/۷	۱۴۰۱۵	-	-	-	-
۹۹/۲۵	۲/۲۵	۳۷/۴۵	۱۴۰/۸۳	-	۲۵-۰/۰/۸	۸۱/۱۰	-	۲۴۹/۹۸	۱۱۰/۲۰	۲/۹/۷	-	۱۱/۶/۸	-	۱۸/۷/۳۱	۱۴۰۱۷	-	-	-	-
۳۵/۳۳	۱/۵۶	-۰/۱/۶	۳۷/۰/۹	-	۴۲/۵۱	۲۱/۸/۶	-	۴۲/۵۶	۶۰/۷۵	۰/۷/۶	-	۳۳/۰/۰	-	۲۵/۸۲	۱۴۰۲۵	-	-	-	-
۳۵/۶۷	۱/۳۰	۱/۸۷	۳۵/۵۲	-	۴۷/۹۹	۲۰/۷/۱	-	۴۷/۹۶	۳۲/۶۹	۱/۵۲	-	۳۰/۱/۷	-	۳۳/۰/۵	۱۴۰۷۱	-	-	-	-

جدول ۴. مقادیر تعریف شده پارامترهای الگوریتم کلونی زنبور عسل

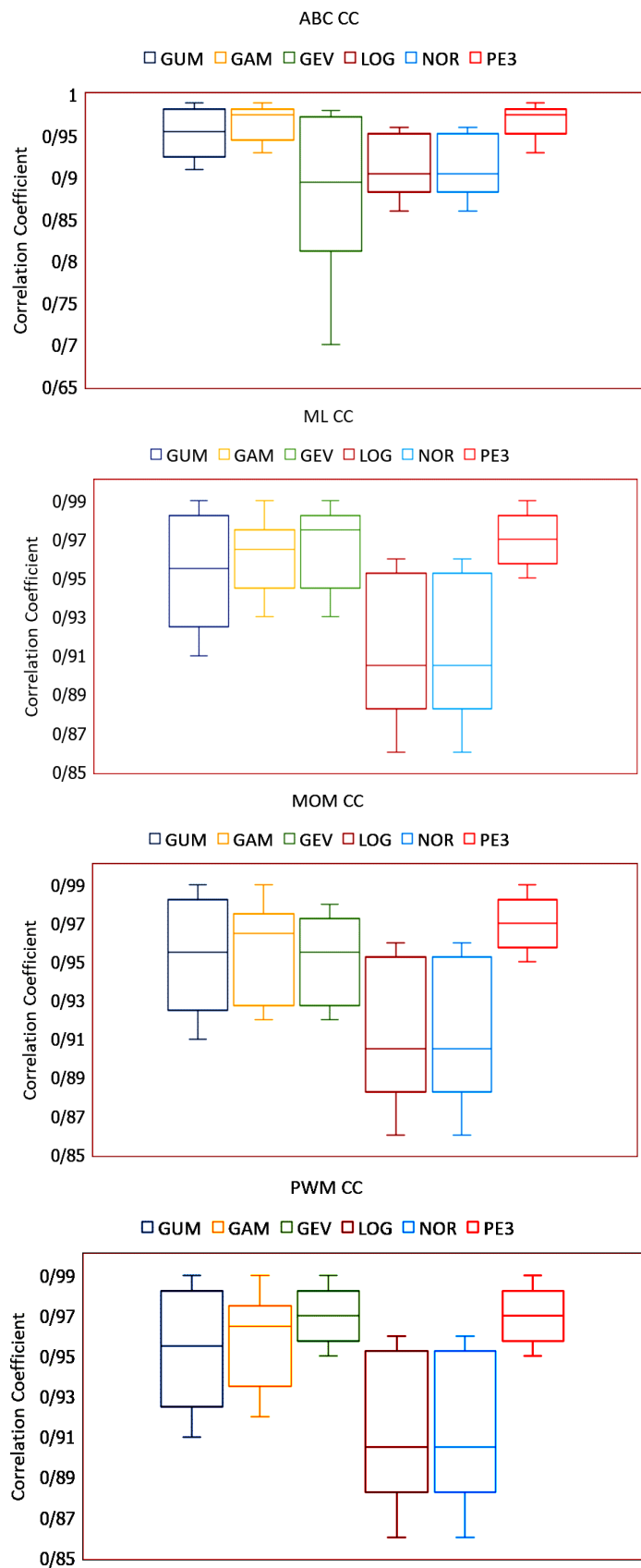
مقدار	توصیف	پارامتر
۱	Acceleration Coefficient Upper Bound	a
۳۰	Abandonment Limit Parameter	L
۱۰	Colony Size	n
۱۰۰	Maximum Number of Iteration	m
۱۰	Number of Onlooker Bees	o



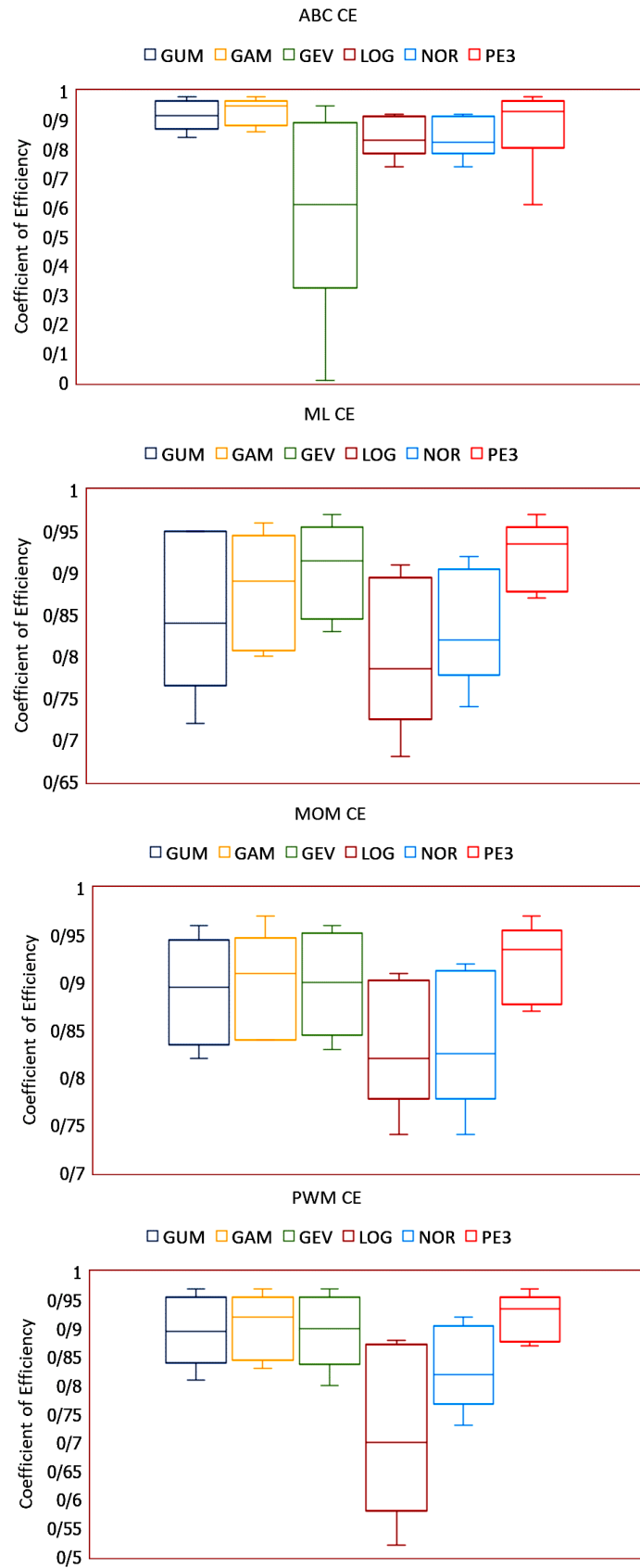
شکل ۳. تغییرات خطای الگوریتم کلونی زنبور عسل در تخمین بهینه پارامترهای تابع توزیع گاما در ایستگاه کشتارگاه

جدول ۵. مقادیر خطای الگوریتم ACO در تخمین پارامترهای توابع توزیع در ایستگاه‌های مطالعاتی

PE3	NOR	LOG	GEV	GAM	GUM	ایستگاه
۰,۰۳۵	۰,۰۸۰	۰,۰۸۴	۰,۰۲۷	۰,۰۳۶	۰,۰۳	۱۴۰۰۱
۰,۰۵۳	۰,۱۹۰	۰,۱۹۲	۰,۰۴۸	۰,۰۴۸	۰,۱۰	۱۴۰۱۳
۰,۰۶۲	۰,۱۸۸	۰,۱۸۰	۰,۰۵۰	۰,۱۱۸	۰,۱۰۷	۱۴۰۱۵
۰,۰۱۶	۰,۰۶۹	۰,۰۶۷	۰,۰۱۶	۰,۰۱۶	۰,۰۱۸	۱۴۰۱۷
۰,۱۱۰	۰,۲۵۵	۰,۲۵۱	۰,۰۱۳	۰,۰۷۲	۰,۱۵۰	۱۴۰۵۵
۰,۰۳۶	۰,۱۳۷	۰,۱۲۸	۰,۰۲۵	۰,۰۳۵	۰,۰۵۴	۱۴۰۷۱



شکل ۴. مقایسه تغییرات ضریب همبستگی روش‌های مورد مطالعه در تخمین پارامترهای توابع توزیع



شکل ۵. مقایسه تغییرات ضریب تشخیص روش‌های مورد مطالعه در تخمین پارامترهای توابع توزیع

مورد مطالعه توزیع گاما است. نتایج این جدول و نمودارها نشان می‌دهد که این تابع توزیع کمترین خطای تخمین پارامتر توسط الگوریتم کلونی زنبور عسل، بیشترین مقدار پارامترهای ضریب همبستگی و ضریب تشخیص و کمترین مجذور میانگین مربعات خطا را در ایستگاه‌های مورد مطالعه دارد. مطالعات گذشته در این منطقه نتایج متفاوتی را نشان می‌دهد که می‌تواند به علت روش تخمین پارامتر و همچنین طول دوره آماری متفاوت باشد. به‌عنوان مثال شیخ و همکاران (۲۱) با استفاده از تحلیل منطقه‌ای به روش گشتاورهای خطی، تابع توزیع لجستیک تعمیم‌یافته را به‌عنوان مناسب‌ترین تابع توزیع منطقه معرفی کردند. فرسادنیا و همکاران (۷) با استفاده از روش نروفازی حوضه‌های آبخیز مازندران را به سه ناحیه همگن تقسیم کردند. نتایج تحقیق آنها نشان داد که تابع توزیع لجستیک تعمیم‌یافته مناسب‌ترین تابع توزیع در این نواحی است. کاظمی‌کیا و همکاران (۱۵) از مقادیر سری جزئی و حداکثر برای تحلیل سیلاب چهار ایستگاه هیدرومتری موجود در حوضه‌های آبخیز بابل‌رود و تالار در استان مازندران استفاده کردند. آنها نتیجه گرفتند که توابع توزیع پیرسون نوع ۳ و گاما مناسب‌ترین توابع توزیع به‌ترتیب در ایستگاه‌های شیرگاه-تالار و بابل-کشتارگاه و تابع توزیع لوگ نرمال سه پارامتره بهترین تابع توزیع در ایستگاه‌های کیاکلا و قرآن‌تالار است.

تخمین تناوب سیل در منطقه

با انتخاب تابع توزیع گاما به‌عنوان مناسب‌ترین تابع توزیع و پارامترهای تعیین شده توسط الگوریتم کلونی زنبور عسل (جدول ۶)، مقادیر سیلاب با دوره بازگشت‌های مختلف در هر یک از ایستگاه‌های مورد مطالعه تخمین زده شد (جدول ۷).

شکل ۷ منحنی توابع توزیع تجمعی تجربی و گاما و شکل ۸ منحنی Q-Q plot (روش گرافیکی برای نمایش مقادیر چندک‌های تابع توزیع تجربی نمونه در مقابل مقادیر چندک‌های تابع توزیع رفرنس) دبی اوج سیلاب در ایستگاه کشتارگاه را

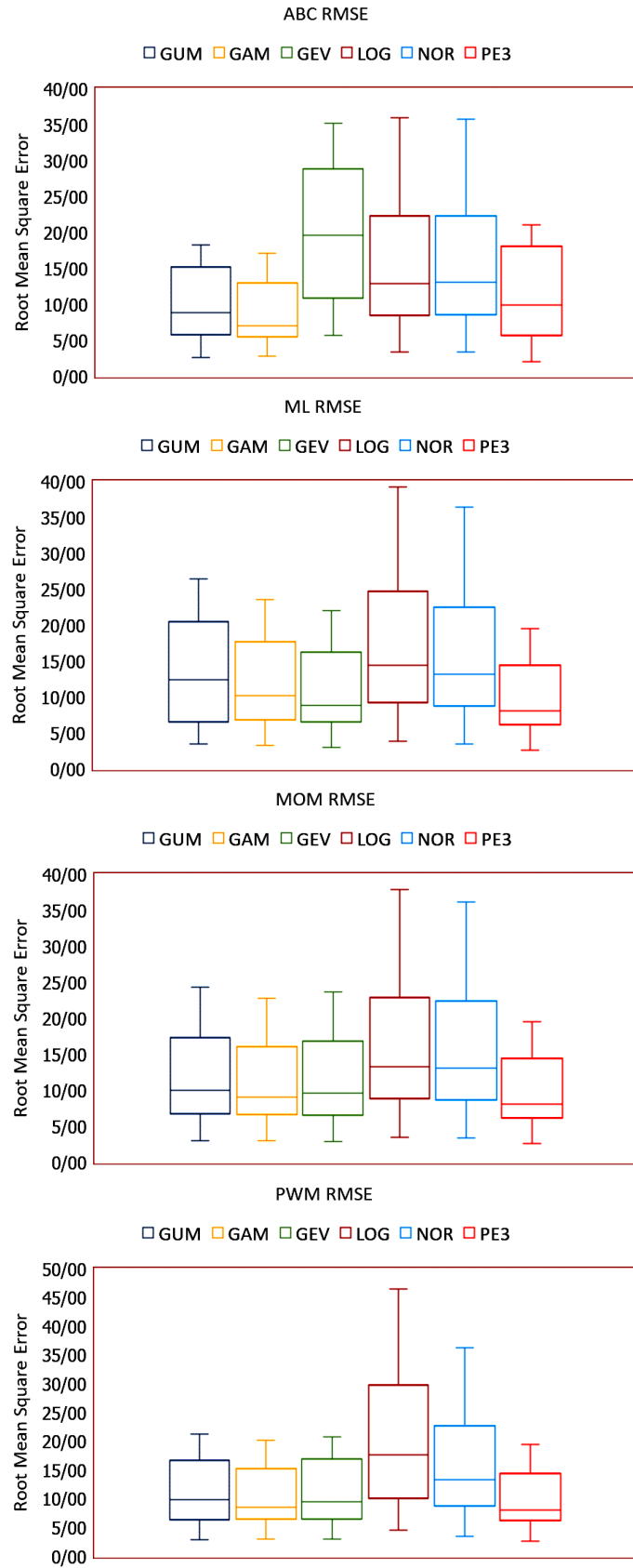
کمترین دامنه تغییرات (۰/۱۰) است درحالی که طبق عملکرد سایر روش‌ها، تابع توزیع پیرسون نوع ۳ دارای بیشترین مقدار میانگین ضریب تشخیص (۰/۹۷) با کمترین دامنه تغییرات (۰/۲۴) است. مقایسه عملکرد چهار روش فوق نشان می‌دهد که تابع توزیع گاما دارای بیشترین مقدار ضریب تشخیص با کمترین دامنه تغییرات در منطقه است. لذا از لحاظ آماره ضریب تشخیص، الگوریتم ABC عملکرد بهتری از سایر روش‌ها داشته و تابع توزیع گاما بهترین تابع توزیع در منطقه است.

شکل ۶ میانگین و تغییرات آماره مجذور میانگین مربعات خطا در منطقه را نشان می‌دهد. در این شکل طبق الگوریتم ABC تابع توزیع گاما دارای کمترین مقدار مجذور میانگین مربعات خطا (۸/۶۱) با کمترین دامنه تغییرات (۱۴/۳۳) است درحالی‌که طبق روش‌های ML، MOM و PWM کمترین مقدار مجذور میانگین مربعات خطا (۹/۷۱) با کمترین دامنه تغییرات در منطقه (۱۶/۸۹) متعلق به تابع توزیع پیرسون نوع ۳ است. مقایسه عملکرد چهار روش فوق نشان می‌دهد که تابع توزیع گاما دارای کمترین مقدار مجذور میانگین مربعات خطا با کمترین دامنه تغییرات در منطقه است. لذا از لحاظ آماره مجذور میانگین مربعات خطا، الگوریتم ABC عملکرد بهتری از سایر روش‌ها داشته و تابع توزیع گاما بهترین تابع توزیع در منطقه است.

درنهایت با جمع‌بندی نتایج حاصل از هر سه آماره برازندگی می‌توان نتیجه گرفت که الگوریتم کلونی زنبور عسل عملکرد بهتری (مقادیر بالاتر ضریب همبستگی و ضریب تشخیص و مقدار کمتر مجذور میانگین مربعات خطا) از سایر روش‌های رایج در تخمین پارامترهای توابع توزیع مطالعاتی در منطقه را دارد.

انتخاب تابع توزیع مناسب در منطقه

با استفاده از نتایج جدول ۵ و نمودارهای ۳ تا ۵ مناسب‌ترین تابع توزیع برای برازش با دبی حداکثر سیلاب سالیانه در منطقه



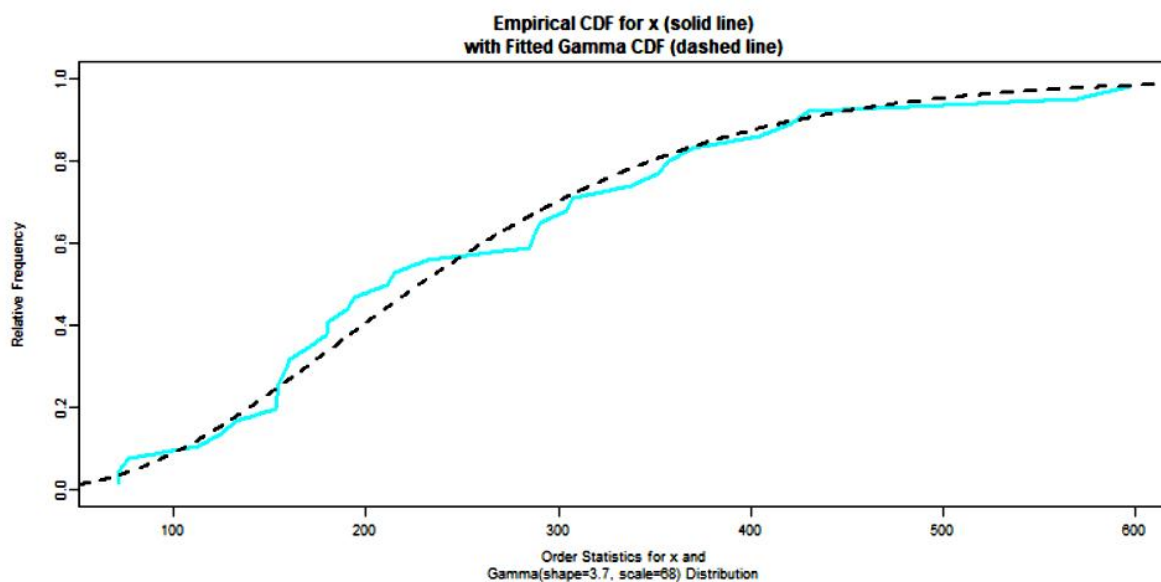
شکل ۶. مقایسه تغییرات مجذور میانگین مربعات خطای روش‌های مورد مطالعه در تخمین پارامترهای توابع توزیع

جدول ۶. مقادیر پارامترهای تخمینی تابع توزیع گاما توسط الگوریتم کلونی زنبور عسل در ایستگاه‌های مطالعاتی

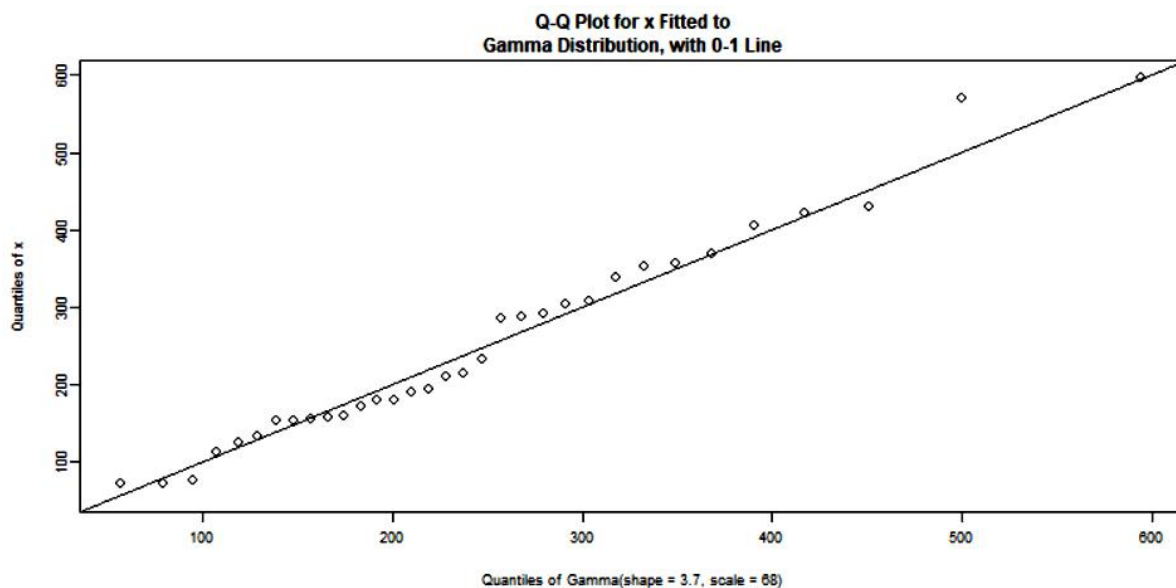
ایستگاه	Shape	Scale
۱۴۰۰۱	۳/۲۰۹	۲۲/۰۲۳
۱۴۰۱۳	۰/۹۹	۴۰/۵۱
۱۴۰۱۵	۴/۶۳	۴/۳۲
۱۴۰۱۷	۲/۹۶۵	۸۵/۸۹
۱۴۰۵۵	۰/۷۶	۶۰/۷۵
۱۴۰۷۱	۱/۵۲	۳۲/۶۹

جدول ۷. مقادیر تخمینی دبی سیلاب (مترمکعب بر ثانیه) با دوره بازگشت‌های مختلف در ایستگاه‌های مطالعاتی

ایستگاه/سال	۲	۵	۱۰	۲۰	۵۰	۱۰۰
۱۴۰۰۱	۱۱/۲۳	۱۴/۴۱	۲۰/۴۱	۲۷/۱۲	۳۷/۲۴	۶۳/۴۸
۱۴۰۱۳	۰/۳۸	۰/۷۸	۲	۴/۱۵	۸/۸۴	۲۷/۶۹
۱۴۰۱۵	۴/۷۷	۵/۷۶	۷/۵۲	۹/۳۹	۱۲/۰۷	۱۸/۵۸
۱۴۰۱۷	۳۶/۴۲	۴۷/۵۰	۶۸/۷۰	۹۲/۸۲	۱۲۹/۶۲	۲۲۶/۶۸
۱۴۰۵۵	۰/۱۳	۰/۳۱	۱/۰۶	۲/۶۸	۶/۹۵	۲۸/۰۵
۱۴۰۷۱	۱/۹۸	۳/۱۷	۵/۹۸	۹/۸۶	۱۶/۸۷	۳۹/۳۹



شکل ۷. منحنی توابع توزیع تجمعی تجربی و گامای دبی اوج سیلاب در ایستگاه کشتارگاه (۱۴۰۰۱)



شکل ۸. منحنی Q-Q plot دبی اوج سیلاب در ایستگاه کشتارگاه (۱۴۰۰۱)

الگوریتم کلونی زنبور عسل در قیاس با روش‌های متعارف تخمین پارامتر نشان داده شد. بهترین عملکرد این الگوریتم در تخمین پارامتر مربوط به توابع توزیع گاما و مقادیر حد بوده که کمترین خطای مدل را نشان می‌دهند. مقایسه عملکرد روش‌های ABC، ML، MOM و PWM در برآورد سه آماره برازندگی نشان می‌دهد که تابع توزیع پیرسون نوع ۳ دارای بیشترین مقدار میانگین ضریب همبستگی با کمترین دامنه تغییرات و تابع توزیع گاما دارای بیشترین مقدار ضریب تشخیص با کمترین دامنه تغییرات و همچنین کمترین مقدار مجذور میانگین مربعات خطا با کمترین دامنه تغییرات است. مناسب‌ترین تابع توزیع برای برازش با دبی حداکثر سیلاب سالیانه در منطقه مورد مطالعه توزیع گاما است. این تابع توزیع کمترین خطای تخمین پارامتر توسط الگوریتم کلونی زنبور عسل، بیشترین مقدار پارامترهای ضریب همبستگی و ضریب تشخیص و کمترین مقدار مجذور میانگین مربعات خطا را در ایستگاه‌های مورد مطالعه دارد. در نهایت با جمع‌بندی نتایج به‌دست آمده می‌توان نتیجه گرفت که الگوریتم کلونی زنبور عسل عملکرد بهتری (مقادیر بالاتر ضریب همبستگی و ضریب تشخیص و مقدار کمتر مجذور میانگین مربعات خطا)

نشان می‌دهد. طبق تعریف، در نمودار Q-Q plot، هر چقدر نقاط نزدیک‌تر به خط فرضی پراکنده باشند احتمال پذیرفتن فرض صفر بیشتر است. به عبارت دیگر داده‌های مورد مطالعه از تابع توزیع مورد نظر پیروی می‌کند.

نتیجه‌گیری

برآورد دقیق سیلاب با دوره بازگشت‌های مختلف خطر شکست سازه‌های کنترل جریان (به‌واسطه برآورد کم جریان) و هزینه‌های ساخت آن را (به‌واسطه برآورد بیشینه جریان) کاهش می‌دهد. به‌واسطه طول آماری کوتاه‌مدت سیلاب مشاهداتی و مکانیسم غیرخطی آن، تخمین سیلاب با دوره بازگشت‌های بزرگ بسیار مشکل است. لذا انتخاب روش مناسب که بتواند با تکیه بر طول کوتاه‌مدت آمار برآورد دقیق‌تری از سیلاب‌های بزرگ ارائه دهد بسیار مهم است. الگوریتم‌های فرا ابتکاری به‌واسطه ماهیت پیچیده ریاضی در حل مسائل بهینه‌سازی مختلف استفاده می‌شود. آزمون الگوریتم‌های مختلف در تخمین پارامترهای توابع توزیع و مقایسه عملکرد آنها می‌تواند در انتخاب روش مناسب تخمین سیلاب مفید واقع شود. در این تحقیق عملکرد مناسب

الگوریتم ژنتیک، ازدحام ذرات و کلونی مورچه نیز مورد مطالعه قرار گیرد.

از سایر روش‌های رایج در تخمین پارامترهای توابع توزیع مطالعاتی در منطقه را دارد. با این حال پیشنهاد می‌شود تا برای بهبود نتایج این تحقیق، الگوریتم‌های دیگر بهینه‌سازی نظیر

منابع مورد استفاده

1. Abdolhay, A. 2008. Comparison of hydrological homogenization methods for development of flood regional models. Master thesis, Universiti Putra Malaysia, Serdang, Malaysia.
2. Adamowski, K. 1981. Plotting formula for flood frequency. *Journal of the American Water Resources Association* 17(2): 197-202.
3. Ahani, A., S. Emamgholizadeh, M. N. Mousavi Nadoushani and K. Azhdari. 2016. Regional flood frequency analysis by hybrid cluster analysis and L-moments. *Journal of Watershed Management Research* 6(12) 11-20. (In Farsi).
4. Alrashidi, M., S. Rahman and M. Pipattanasomporn. 2020. Metaheuristic optimization algorithms to estimate statistical distribution parameters for characterizing wind speeds. *Renewable Energy* 149: 664-681.
5. Chavoshi, S., W. N. Azmin Sulaiman, B. Saghafian, M. D. Sulaiman and L. A. Manaf. 2012. Soft and hard clustering methods for delineation of hydrological homogeneous regions in the southern strip of the Caspian Sea watershed. *Journal of Flood Risk Management* 5 (4): 282-294.
6. Chavoshi, S., W. N. Azmin Sulaiman, B. Saghafian, M. D. Sulaiman and L. A. Manaf. 2013. Flood prediction in southern strip of Caspian Sea watershed. *Journal of Water Resources* 40(6): 593-605. DOI: 10.1134/S0097807813060122.
7. Farsadnia, F., P. Haghghat Jou, Z. Shamohamadi Heidari and A. R. Moghaddam Nia. 2014. Flood regionalization of Mazandaran Province's watersheds by using Fuzzy Clustering Algorithm. *Journal of Iran-Watershed Management Science and Engineering* 8(24): 61-65. (In Farsi).
8. Fayaz Bakhsh, H., M. Kaviani Poor and M. Hedaiati Zadeh. 2015. The comparison of artificial neural network and linear regression in regional flood modeling. Proceeding of the 10th International Congress on Civil Engineering, Civil Engineering Faculty, Tabriz University.
9. Freitas de Andrade, C., L. Ferreira dos Santos, M. V. Silveira Macedo, P. A. Costa Rocha and F. Ferreira Gomes. 2019. Four heuristic optimization algorithms applied to wind energy: determination of Weibull curve parameters for three Brazilian sites. *International Journal of Energy Environment Engineering* 10: 1-12.
10. Gelder, P. H. A. J. M. 2003. Statistical estimation methods in hydrological engineering. In: Proceedings of 2003 International Scientific Seminar, Publishing House of the Institute of Geography, Delft University of Technology, Netherlands.
11. Greenwood, J. A., J. M. Landwehr, N. C. Matalas and J. R. Wallis. 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* 15(5): 1049-1054.
12. Grimaldi S, S. -C. Kao, A. Castellarin, S. -M. Papalexioiu, A. Viglione, F. Laio, H. Aksoy and A. Gedikli. 2011. Statistical hydrology. PP. 479-517. In: Peter Wilderer (Eds.), Treatise on Water Science, Oxford: Academic Press.
13. Hassanzadeh, A., A. Abdi, S. Talatahari and V. P. Singh. 2011. Metaheuristic algorithms for hydrologic frequency analysis. *Water Resources Management* 25(7): 1855-1879.
14. Karaboga, D. 2005. An idea based on honey bee swarm for numerical optimization. Technical Report-TR06, Department of Computer Engineering, Engineering Faculty, Erciyes University.
15. Kazemikia, S., T. Besharati, M. Zolfaghari and M. R. Ghanbarpour. 2016. Comparative study of maximum and partial duration series in flood frequency analysis (Case study in Talar and Babolrud watersheds in Mazandaran Province). *Journal of Iran-Watershed Management Science and Engineering* 10(3): 113-125. (In Farsi).
16. Langat, P. K., L. Kumar, R. Koech and M. K. Ghosh. 2019. Hydro-morphological characteristics using flow duration curve, Historical data and remote sensing: Effects of land use and climate. *Water* 11: 309-329.
17. Molina-Aguilar, J. P., A. Gutierrez-Lopez, J. A. Raynal-Villaseñor and L. G. Garcia-Valenzuela. 2019. Optimization of parameters in the generalized extreme-value distribution type 1 for three populations using harmonic search. *Atmosphere* 10: 257.
18. Rostamin Kamrood, M., Z. Shah Mohamad Heidari, P. Haghghatjoo and A. Moghadamnia. 2011. Regional flood frequency analysis using L-moments methods in Gilan Province' catchments. *Journal of Iran Water Resources Engineering* 4: 39-50. (In Farsi).

19. Saghafian, B. and H. Farazjoo. 2007. Flood generation delineating and flooding priority of hydrologic units in Golstan Dam basin. *Journal of Iran-Watershed Management Science and Engineering* 1(1): 1-11. (In Farsi).
20. Shabanloo, S., H. Sedghi, B. Saghafian and H. Moosavi Jahromi. 2008. Flood mapping in Golestan river network using GIS. *Journal of Iran Water Research* 2 (3): 11-22. (In Farsi).
21. Sheikh, Z., A. Dehvar and F. Farsadnia. 2014. Comparison of canonical kriging and linear moments methods for regional flood frequency analysis in Mazandaran province. *Journal of Iran-Watershed Management Science and Engineering* 8(25): 25-38. (In Farsi).
22. Shin, J., J. Heo, C. Jeong and T. Lee. 2014. Meta-heuristic maximum likelihood parameter estimation of the mixture normal distribution for hydro-meteorological variables. *Stochastic Environmental Research and Risk Assessment* 28: 347-358.
23. Shin, J. -Y., T. Lee and T. B. M. Ouarda. 2015. Heterogeneous mixture distributions for modeling multisource extreme rainfalls. *Journal of Hydrometeorology* 16: 2639-2657.
24. Sun, J., J. M. Garibaldi and C. Hodgman. 2012. Parameter estimation using metaheuristics in systems biology: a comprehensive review. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* 9(1): 185-202.
25. Yang, X., A. Gandomi, S. Talatahari and A. H. Alavi. 2012. Metaheuristics in Water, Geotechnical and Transport Engineering. Elsevier, Hardcover ISBN 0123983177, 9780123983176.
26. Yoon, Y., Ch. Jeong and T. Lee. 2013. Application of harmony search to design storm estimation from probability distribution models *Journal of Applied Mathematics* 1-11. Article ID 932943, 11 pages. Available online at: <http://dx.doi.org/10.1155/2013/932943>.

Evaluating the Performance of the Artificial Bee Colony Algorithm in Flood Frequency Analysis

S. Chavoshi* and K. Shirani¹

(Received: September 29-2019; Accepted: May 3-2020)

Abstract

Selection of the appropriate distribution function and estimation of its parameters are two fundamental steps in the accurate estimation of flood magnitude. This study relied on the concept of optimization by meta heuristic algorithms to improve the results obtained from the conventional methods of parameter estimation, such as maximum likelihood (ML), moments (MOM) and probability weighted moments (PWM) methods. More specifically, this study aimed to improve flood frequency analysis using the Artificial Bee Colony algorithm (ABC). The overall performance of this algorithm was compared to the conventional methods by employing goodness of fit statistics, correlation coefficient (CC), coefficient of efficiency (CE) and root mean square error (RMSE). The study area, Babolrood catchment located in southern bank of Caspian Sea, has been subjected to annual flooding events. A total of 6 hydrometry stations in the study area were delineated and their data were used in the analysis of 6 distribution functions of Normal, Gumbel, Gamma, Pearson Type 3, General Extreme Value and General Logistic. This analysis indicated that Gamma and Pearson Type 3 were the most appropriate distribution functions for flood appraisal in the study area, according to the ABC and conventional methods, respectively. Also, the results showed that ABC outperformed ML, MOM and PWM; so, Gamma could be recommended as the most reliable distribution function for flood frequency analysis in the study area.

Keywords: Flood frequency, Parameter estimation, Metaheuristic algorithm, Artificial bee colony.

1. Soil Conservation and Watershed Management Research Department, Isfahan Agricultural and Natural Resources Research and Education Center, AREEO, Isfahan, Iran.

*: Corresponding author, Email: chsatar@gmail.com