

ارائه روابط منطقه‌ای برای حداکثر دبی‌های روزانه زیرحوضه‌های غرب ایران با استفاده از مدل‌های رگرسیونی خطی تعمیم یافته (GLMs)

چکیده

روابط منطقه‌ای قابل اطمینان که بتواند با استفاده از خصوصیات حوضه‌های آبخیز بزرگی دبی سیلاب‌ها را به ازای دوره بازگشت‌های مختلف پیش‌بینی کند از اهمیت ویژه‌ای برای مدیریت سیل و طراحی سازه‌های هیدرولیکی در حوضه‌های فاقد ایستگاه هیدرومتری برخوردار است. بنابراین در پژوهش حاضر سعی شد تا با استفاده از مدل‌های رگرسیونی خطی تعمیم یافته (GLMs) مدل‌های منطقه‌ای مناسبی برای برآورد حداکثر دبی‌های روزانه به ازای دوره بازگشت‌های ۲، ۱۰، ۵۰ و ۱۰۰ سال برای ۶۲ زیرحوضه از حوضه‌های کارون بزرگ و کرخه ارائه شود. بر اساس نتایج، حوضه‌های مورد مطالعه توسط برخی از خصوصیات فیزیوگرافی و اقلیمی به چهار زیرگروه گروه‌بندی شدند. بطورکلی نتایج نشان داد که بجز زیرمنطقه شماره دو که شامل زیرحوضه‌های بسیار بزرگ ($A \approx 17300 \text{ km}^2$) می‌باشد، در تمامی زیرمناطق تعیین شده فرآیند مدل‌سازی موفق است. بطوریکه مقادیر ضرایب تبیین اصلاح شده مناسب‌ترین مدل‌ها برای حداکثر دبی‌های ۱۰۰ ساله در زیرمناطق اول، سوم و چهارم به ترتیب ۸۲/۴، ۹۱/۳ و ۹۰/۶ درصد و درصد نسبی خطای آنها (RRMSE) نیز به ترتیب ۹/۵، ۹/۲۳ و ۶/۸ درصد برآورد گردید. همچنین نتایج نشان داد که برای سیلاب‌ها با دوره بازگشت‌های کوتاه مدت مانند ۲ و ۱۰ سال متغیرهایی همچون طول، محیط و مساحت حوضه‌ها تأثیر زیادی بر روی حداکثر دبی‌های روزانه خواهند داشت، در حالی که برای دوره بازگشت‌های بلند مدت مانند ۵۰ و ۱۰۰ سال اغلب خصوصیات آبراهه‌ها و شبکه زهکشی حوضه‌ها مانند طول آبراهه اصلی، مجموع طول سیستم رودخانه‌ها و شیب آبراهه اصلی حداکثر دبی‌های روزانه را تحت تأثیر قرار می‌دهند. با توجه به پژوهش صورت گرفته می‌توان اظهار داشت که رفتار حداکثر دبی‌های روزانه در منطقه مورد مطالعه تا حد زیادی از خصوصیات مختلف اقلیمی و فیزیوگرافی حوضه‌ها تأثیر می‌پذیرد و لذا با مدل‌سازی صحیح آن می‌توان با دقت نسبتاً مناسبی ویژگی‌های این متغیر را در حوضه‌های فاقد ایستگاه هیدرومتری برآورد نمود.

کلمات کلیدی: سیلاب، حوضه آبخیز، مدل‌سازی، روابط غیرخطی، روابط تجربی سیل.

۱- مقدمه

پدیده سیلاب یکی از مخاطرات طبیعی است که کشور ما در سال‌های اخیر بیش از پیش از آن رنج برده است (۲۱). از این حیث این پدیده توجه بسیاری از محققین و تصمیم‌گیران را در چند سال اخیر به خود معطوف ساخته است. از سوی دیگر انجام هر گونه مطالعه و ارائه طرح‌های مدیریتی و سازه‌ای به منظور کاهش سیل (Flood mitigation) در درجه اول نیازمند اطلاعات مناسب از حوضه‌ها و آمار بلند مدت از دبی‌های حداکثر در رودخانه‌ها است (۹). با این وجود بسیاری از حوضه‌ها و رودخانه‌های کشور فاقد ایستگاه‌های هیدرومتری هستند یا دارای ایستگاه‌هایی با طول دوره آماری کوتاه مدت می‌باشند. از این جهت انواع روابط تجربی مانند رابطه‌های کریگر (Creager)، فولر (Fuller) و دیکن (Dicken) با توجه به انواع خصوصیات حوضه‌های آبخیز و به ویژه مساحت حوضه‌ها برای برآورد دبی‌های اوج و حداکثر رودخانه‌ها ارائه شده‌اند (۳). در اکثر روش‌های تجربی سعی می‌شود که

روابط منطقی بین دبی‌های حداکثر و خصوصیات حوضه‌های دارای آمار تعیین شود و سپس این روابط به حوضه‌های فاقد آمار تعمیم داده شوند. از این جهت اخیراً استفاده از تکنیک‌های مختلف برای ارائه مدل‌های منطقه‌ای سیل مورد توجه قرار گرفته‌اند.

عزیز و همکاران (۲) روش‌های مدل‌سازی شبکه عصبی (Artificial neural network; ANN) را با روش‌های مدل‌سازی رگرسیون چندکی بر روی نتایج حاصل از تحلیل فراوانی منطقه‌ای ۴۵۲ حوضه در استرالیا مورد مقایسه قرار دادند. نتایج پژوهش آنها حاکی از عملکرد خوب هر دو روش برای برآورد بزرگی سیلاب‌ها در حوضه‌های فاقد آمار بود. با این وجود نتایج آنها نشان داد که روش ANN برای مناطق بزرگ با حجم زیادی داده بهتر عمل خواهد کرد. هایلپ جرجیس و آلفردسون (۸) رویکرد مدل‌سازی تحلیل فراوانی منطقه‌ای و مدل‌سازی شاخص سیل (Flood index) را بجای مدل‌سازی چندک‌های سیل در حوضه‌های نروژ پیشنهاد کردند. نتایج آنها نشان داد که در صورت منطقه‌بندی مناسب حوضه‌ها، تغییرات شاخص سیل را می‌تواند تا حدود ۹۰٪ (R^2) توسط مساحت حوضه‌ها به تنهایی توجیه کرد. رحمان و همکاران (۱۸) اظهار داشتند که ممکن است خصوصیات حوضه‌های آبخیز با دبی‌های حداکثر سیلاب‌ها دارای روابط غیرخطی پیچیده‌ای باشند که با تغییر شکل لگاریتمی داده‌ها به سادگی قابل مدل‌سازی نیست. از این رو ایشان استفاده از مدل‌های رگرسیونی جمعی تعمیم یافته (Generalized additive models; GAM) را بجای مدل‌های رگرسیونی لگاریتمی برای مدل‌سازی چندک‌های سیل در استرالیا آزمون و پیشنهاد کردند. رحمان و رحمان (۱۹) نیز با بررسی اثر متغیرهای محیطی بر دبی سیلاب‌ها در ۸۸ حوضه در استرالیا اظهار داشتند که خوشه‌بندی حوضه‌ها با مؤلفه‌های اصلی دقت روابط منطقه‌ای سیل را در زیرمناطق تعیین شده به شدت افزایش خواهد داد. ایشان همچنین تکنیک رگرسیون چندکی را برای مدل‌سازی و ارائه روابط منطقه‌ای در استرالیا پیشنهاد کردند. کرد رستمی و همکاران (۱۰) اثر متغیرهای اقلیمی و فیزیوگرافی حوضه‌های آبخیز را بر روی دبی برخی از رودخانه‌های استرالیا از طریق مدل‌های شبکه عصبی مصنوعی مورد بررسی قرار دادند. ایشان اظهار داشتند که افزایش متغیرهای پیشگو در مدل‌ها الزاماً منجر به افزایش کارایی مدل‌ها نخواهد شد و با توجه به نتایج بدست آمده استفاده از مدل‌های شبکه عصبی مصنوعی را برای سایر مناطق استرالیا پیشنهاد کردند. امسیلینی و همکاران (۱۳) سه تکنیک مدل‌سازی رگرسیون لگاریتمی، رگرسیونی جمعی تعمیم یافته و مدل رگرسیونی چند متغیره تطبیقی اسپیلاین (Multivariate adaptive regression spline; MARS) را برای کمی‌سازی و مدل‌سازی نقش متغیرهای محیطی درون حوضه‌ها بر روی دبی سیل در کبک کانادا مورد بررسی قرار دادند. نتایج ایشان نیز نشان داد که استفاده از مدل‌های رگرسیونی مارس و جمعی تعمیم یافته قابلیت بیشتری نسبت به مدل‌های رگرسیونی سنتی لگاریتمی برای ارائه روابط منطقه‌ای جهت برآورد دبی سیل در حوضه‌های فاقد آمار دارند.

در ایران نیز مصطفایی (۱۲) در حوضه‌های استان قزوین اقدام به تحلیل فراوانی منطقه‌ای و مدل‌سازی سیلاب با استفاده از مدل‌های رگرسیونی لگاریتمی و شبکه عصبی مصنوعی کرد. نتایج وی نشان داد که وقایع سیل در قزوین از توزیع‌های لجستیک تعمیم یافته (Generalized logistic; GLO) و پارتو تعمیم یافته (Generalized Pareto; GPA) پیروی می‌کنند. نتایج این مطالعه نشان داد که مدل‌های شبکه عصبی مصنوعی نسبت به رگرسیون‌های لگاریتمی دقت بیشتری دارند. شریفی گرم‌دره و همکاران (۲۰) تکنیک‌های مدل‌سازی رگرسیون بردار پشتیبان (Support vector regression; SVR)، شبکه عصبی مصنوعی (ANN) و مدل‌های رگرسیونی غیرخطی (Nonlinear regression; NLR) را بر روی نتایج تحلیل فراوانی منطقه‌ای حوضه‌های نمکی ایران مرکزی جهت مدل‌سازی دبی پیک سیلاب با دوره بازگشت‌های مختلف بکار بردند. نتایج آنها نیز حاکی از عملکرد مناسب مدل‌های SVR نسبت به مدل‌های ANN و NLR بود. قادری و همکاران (۵) با برازش توزیع‌های آماری مختلف به داده‌های دبی‌های اوج سیلاب ۴۶ ایستگاه هیدرومتری در منطقه کارون و کرخه، توزیع لوگ-پیرسون نوع سوم (Log-Pearson type III; LP-III) را به عنوان توزیع منطقه‌ای مناسب معرفی کردند. ایشان با استفاده از توزیع احتمال LP-III دبی اوج سیلاب‌های ۵۰ ساله را برای ایستگاه‌های مورد مطالعه‌شان برآورد و اقدام به مدل‌سازی آنها به روش‌های درخت تصمیم، شبکه عصبی مصنوعی و شبکه بیزی کردند. نتایج آنها نشان داد که متغیرهای محیط، ضریب شکل، طول آبراهه اصلی و طول حوضه‌ها مؤثرترین خصوصیات فیزیوگرافیکی حوضه‌ها برای بزرگی سیلاب‌های ۵۰ ساله هستند. الله‌بخشیان فارسانی و همکاران (۱) نیز به صورت مشابه در محدوده حوضه‌های کارون بزرگ و کرخه توزیع نرمال تعمیم یافته (Generalized normal; GNO) را به عنوان توزیع منطقه‌ای مناسب معرفی کردند. پس از آن حداکثر دبی‌های روزانه را به ازای دوره بازگشت‌های مختلف توسط توزیع GNO برآورد و از تکنیک‌های مدل‌سازی یادگیری ماشین برای مدل‌سازی سیلاب‌ها با دوره بازگشت‌های مختلف استفاده کردند. نتایج آنها حاکی از عملکرد مناسب‌تر مدل‌های یادگیری ماشین نسبت به مدل‌های غیرخطی رگرسیونی بود. همچنین نتایج ایشان نشان داد که متغیرهای کاربری، حداکثر بارش ۲۴ ساعته، میانگین شیب حوضه‌ها، ضریب فشردگی، میانگین و حداکثر ارتفاع حوضه‌ها جزو مؤثرترین متغیرها بر روی سیلاب‌های منطقه می‌باشند.

یکی از مباحث بنیادی در زمینه مطالعات سیلاب برآورد دبی‌های حداکثر روزانه و لحظه‌ای حوضه‌های آبخیز به ازای دوره بازگشت‌های مشخص می‌باشد. این امر عمدتاً از طریق انجام فرآیند تحلیل فراوانی ایستگاهی و یا منطقه‌ای بر روی داده‌های ایستگاه‌های هیدرومتری میسر می‌شود. با این وجود بسیاری از حوضه‌ها و رودخانه‌های کشور فاقد ایستگاه‌های هیدرومتری هستند یا دارای ایستگاه‌هایی با طول دوره آماری کوتاه مدت می‌باشند؛ بنابراین به عنوان یک راه حل جایگزین در حوضه‌های فاقد آمار یا دارای آمار کوتاه مدت برای داده‌های دبی از انواع روش‌های تجربی و مدل‌سازی دبی‌های حداکثر روزانه و اوج برای برآورد این

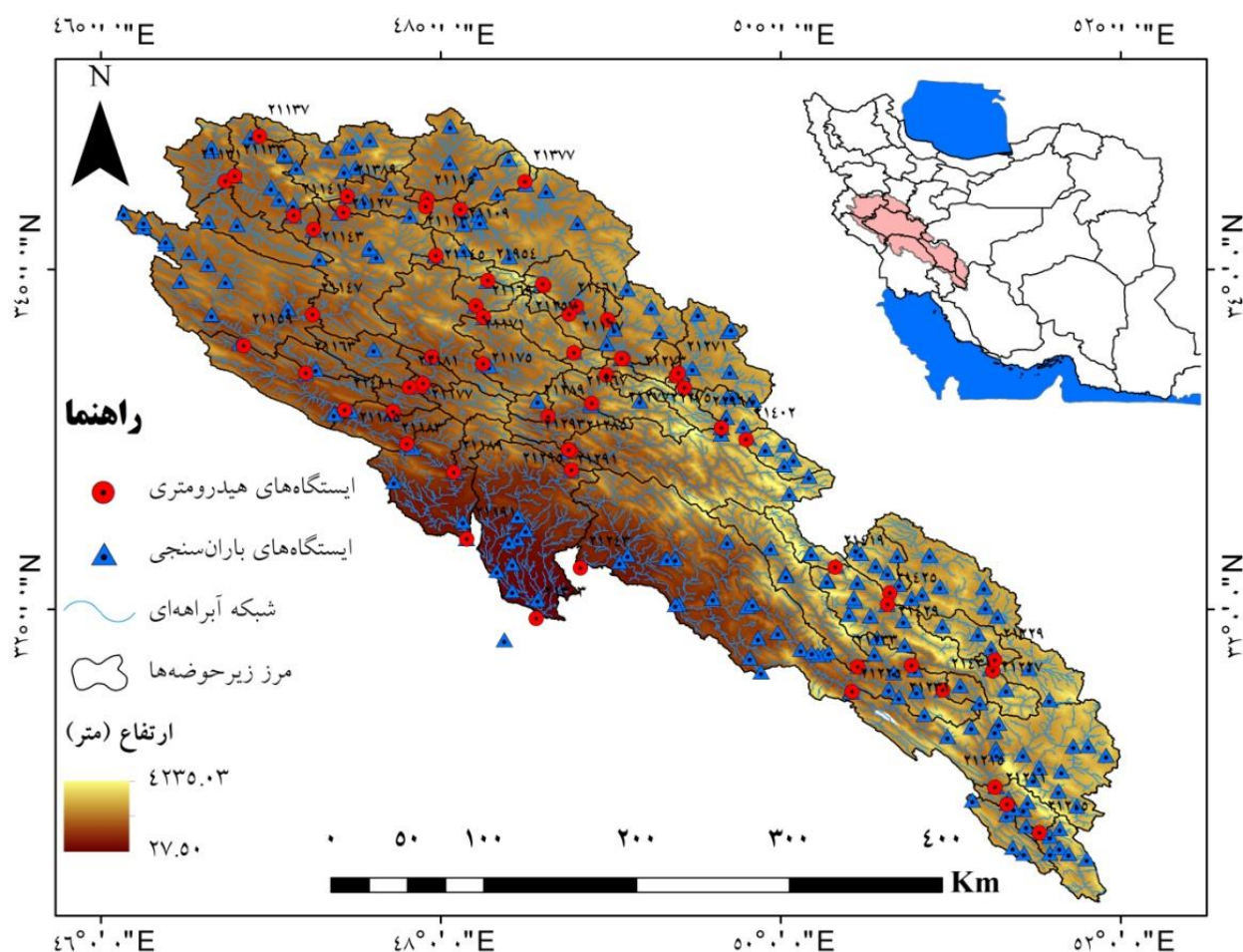
متغیرها به شرط انواع خصوصیات فیزیوگرافی و اقلیمی حوضه‌ها استفاده می‌شود. از این حیث در پژوهش حاضر نیز سعی شد تا ضمن واسنجی کردن ضرایب مدل‌های تجربی با استفاده از مدل‌های خطی تعمیم یافته (GLMs) روابط رگرسیونی منطقه‌ای صریحی به منظور برآورد دبی‌های حداکثر روزانه با دوره بازگشت‌های ۲، ۱۰، ۵۰ و ۱۰۰ سال (متغیرهای وابسته) برای حوضه‌های فاقد ایستگاه هیدرومتری در غرب ایران ارائه شود که برای تمامی کاربران به سادگی قابل استفاده می‌باشند.

۲- مواد و روش‌ها

۲-۱ منطقه مورد مطالعه و داده

پژوهش حاضر از نظر هیدرولوژیکی در محدوده حوضه‌های کارون بزرگ و کرخه در غرب زاگرس انجام شد. از نظر سیاسی منطقه مورد مطالعه بخش‌هایی از استان‌های همدان، لرستان، ایلام، خوزستان، کرمانشاه، کردستان، چهارمحال و بختیاری، کهگیلویه و بویراحمد، اصفهان و مرکزی را شامل می‌شود. منطقه مورد مطالعه از نظر توپوگرافی از محدوده پست جلگه خوزستان با ارتفاع حدود ۲۷ متر از سطح دریا تا ارتفاعات رشته کوه‌های زاگرس با ارتفاعات بیش از ۴۰۰۰ متر را شامل می‌شود (شکل ۱). متوسط بارش در این محدوده از حدود ۱۵۰ تا ۷۵۰ میلی‌متر از مناطق پست تا مناطق مرتفع متغیر است (۵). این محدوده به سبب وجود رودخانه‌های بزرگ کارون و کرخه و سرشاخه‌های آنها از لحاظ هیدرولوژیکی حائز اهمیت است (۱۱). اقلیم غالب در مناطق غربی و جنوب غربی منطقه مورد مطالعه عمدتاً خشک با تابستان‌های گرم است و در مناطق شمالی اقلیم نیمه خشک حاکم است. این درحالی است که در ارتفاعات شرقی و جنوب شرقی منطقه مورد مطالعه اقلیم نیمه مرطوب حاکم است (۵). از نظر کاربری نیز عمده مناطق شرقی و جنوب شرقی منطقه مورد مطالعه شامل کاربری‌های جنگل و مرتع است در حالی که در مناطق غربی و جنوب غربی کاربری کشاورزی متعدداً به چشم می‌خورد.

در راستای نیل به اهداف پژوهش حاضر یازده متغیر فیزیوگرافی که به راحتی برای حوضه‌ها قابل محاسبه هستند به همراه متوسط بارندگی سالانه حوضه‌ها به عنوان متغیرهای مستقل برای مدل‌سازی روابط منطقه‌ای مورد استفاده قرار گرفتند (جدول ۱). در جدول ۱ به خصوصیات مورد مطالعه در این پژوهش اشاره شده است. متوسط بارندگی سالانه حوضه‌ها بر اساس میانگین وزنی بارش سالانه ایستگاه‌های باران‌سنجی (با وزن مساحت پلی‌گون‌های تیسن هر ایستگاه) وزارت نیرو (wts.wtm.ir) در سطح حوضه‌های مورد مطالعه از بدو تأسیس تا سال ۱۳۹۶ محاسبه شد.



شکل ۱. موقعیت محدوده مورد مطالعه و ایستگاه‌های باران‌سنجی و هیدرومتری مورد بررسی

جدول ۱. متغیرهای مورد استفاده برای مدل‌سازی حداکثر دبی‌های روزانه در زیرحوضه‌های مورد مطالعه

نوع متغیر	نام متغیر	نام اختصاری	حداقل	حداکثر	میانگین	ضریب تغییرات (%)
متغیرهای وابسته (Y_i)	حداکثر دبی‌های روزانه ۲ ساله (m^3/s)	Q02	۳/۸	۱۳۲۴	۲۲۸	۱۴۱
	حداکثر دبی‌های روزانه ۱۰ ساله (m^3/s)	Q10	۹/۳	۲۶۵۷	۴۸۹	۱۳۱
	حداکثر دبی‌های روزانه ۵۰ ساله (m^3/s)	Q50	۱۲/۶	۴۵۸۵	۸۰۲	۱۳۴
	حداکثر دبی‌های روزانه ۱۰۰ ساله (m^3/s)	Q100	۱۳/۱	۵۷۵۷	۹۷۶	۱۳۸
متغیرهای مستقل (X_i)	محیط حوضه (km)	P	۳۱/۱	۱۹۳۴	۵۱۰	۱۰۱
	مساحت حوضه (km^2)	A	۳۱/۴	۴۲۸۰۱	۵۹۱۲	۱۶۵
	طول حوضه (km)	BL	۷/۵۱	۷۳۳	۱۴۷	۱۰۸
	ضریب شکل حوضه	FF	۰/۰۸	۰/۵۶	۰/۲۱	۴۵/۸
	ضریب فشردگی حوضه	CC	۱/۶۳	۳/۲۴	۲/۳۸	۱۶/۴
	مجموع طول شبکه آبراه‌ای حوضه (km)	TLR	۱۶/۵	۳۰۹۳۲	۴۲۰۶	۱۶۶
	طول رودخانه اصلی (km)	MRL	۲/۸۸	۷۲۸	۱۳۹	۱۱۲
	شیب رودخانه اصلی (%)	SMR	۰/۲۹	۱۳/۰	۲/۳۱	۱۳۰
	شیب متوسط حوضه (%)	MSB	۸/۴۳	۵۲/۱	۲۴/۰	۴۲/۳
	ارتفاع متوسط حوضه (m)	E	۱۲۳۶	۲۹۱۲	۲۰۵۰	۱۹/۷
	دامنه ارتفاعی حوضه (m)	R	۸۵۵	۴۳۶۲	۲۲۴۷	۳۴/۹
	معدل بارش سالانه حوضه (mm)	AR	۳۸۷	۱۲۸۹	۶۱۱	۳۳/۹

در شکل ۱ نیز موقعیت ایستگاه‌های باران‌سنجی مورد استفاده در این پژوهش نمایش داده شده است. علاوه بر این دبی‌های حداکثر روزانه متناظر با دوره بازگشت‌های ۲، ۱۰، ۵۰ و ۱۰۰ سال برای ۶۲ حوضه مورد مطالعه بر اساس نتایج تحلیل فراوانی منطقه‌ای مطابق با یافته‌های چیت‌ساز و محیط اصفهانی (۴) مورد استفاده قرار گرفت. شایان ذکر است که بر اساس نتایج چیت‌ساز و محیط اصفهانی (۴) همه ایستگاه‌های مورد مطالعه در این تحقیق یا فاقد سازه هیدرولیکی در بالادست هستند و یا اثر آنها بر ماهیت استوکاستیک داده‌های دبی در ایستگاه‌های پایین دست ناچیز و قابل چشم‌پوشی می‌باشد. اطلاعات تفصیلی در ارتباط با داده‌های دبی حداکثر روزانه حوضه‌های مورد مطالعه در جدول پیوست گزارش شده است.

۲-۲ گروه‌بندی

به منظور ارائه مدل‌های منطقه‌ای دقیق‌تر، فرآیند مدل‌سازی برای دو حالت انجام گرفت. در حالت اول کل زیرحوضه‌های مورد مطالعه در قالب یک گروه در نظر گرفته شد و فرآیند مدل‌سازی برای آنها انجام شد. در حالت دوم زیرحوضه‌های مورد مطالعه در گروه‌های مختلف گروه‌بندی شدند و فرآیند مدل‌سازی در هریک از زیرگروه‌ها به صورت مجزا انجام گرفت. برای این منظور جهت هم مقیاس کردن متغیرهای مستقل به عنوان ویژگی‌های حوضه‌های مورد مطالعه، ابتدا از اعداد لگاریتم طبیعی گرفته شد. سپس با استفاده از تکنیک تجزیه به مؤلفه‌های اصلی (Principal component analysis; PCA) بر روی مقادیر لگاریتمی متغیرها، متغیرهایی که بیشترین سهم را از واریانس سه مؤلفه نخست داشتند (بیشترین بار عاملی در مؤلفه‌های اول تا سوم) انتخاب شد. پس از آن بر اساس لگاریتم طبیعی این متغیرهای مهم، فرآیند خوشه‌بندی تفکیکی به روش kmeans با توجه به تعیین تعداد بهینه خوشه‌ها بر اساس آماره مجموع مربعات درون خوشه‌ای (Total within sum of square; WSS) برای گروه‌بندی زیرحوضه‌های مورد استفاده قرار گرفت (۷ و ۱۶).

۲-۳ مدل‌سازی

در این پژوهش سعی شد تا از مدل‌های خطی تعمیم یافته (GLMs) برای مدل‌سازی و ارائه روابط منطقه‌ای صریح برای حداکثر دبی‌های روزانه با دوره بازگشت‌های ۲، ۱۰، ۵۰ و ۱۰۰ سال استفاده شود. مدل‌های GLM از روش حداکثر درست‌نمایی (Maximum likelihood method; ML) به جای روش حداقل مربعات معمولی (Ordinary least squares; OLS) برای برآورد ضرایب مدل استفاده می‌کنند (۱۴). از این روش در مدل‌های GLM برخلاف مدل‌های رگرسیونی خطی معمولی متغیر وابسته می‌تواند دارای هر توزیع احتمالی از خانواده توزیع‌های نمایی (Exponential distributions) باشد (۱۴ و ۲۲). بطور کلی در مدل‌های GLM بجای استفاده از تابع توزیع احتمال نرمال یا همان گوسی برای متغیر وابسته می‌توان از سایر توابع احتمال مانند گاما، پواسون و توزیع دو جمله‌ای استفاده کرد. در پژوهش حاضر از خانواده‌های گاما و گوسی که مناسب برای داده‌های پیوسته بزرگتر از صفر

هستند، برای مدل‌های GLM استفاده شد. از سوی دیگر مدل‌های GLM این قابلیت را دارند که روابط غیرخطی صریح بین متغیرهای مستقل و وابسته مانند روابط نمایی، درجه‌دو و یا چند جمله‌ای را از طریق یک تابع پیوند (Link function; $g(\cdot)$) برای میانگین شرطی متغیر وابسته در مدل‌سازی لحاظ کنند (۶). فرم کلی مدل‌های GLM به صورت رابطه ۱ است (۱۴ و ۲۲).

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad (1)$$

$$E(Y_i | X_{ik}) \equiv \mu_i = g^{-1}(\mu_i); Y_i \sim EF(\mu_i, \phi)$$

که در آن μ_i معادل مقدار مورد انتظار یا امید ریاضی شرطی متغیر وابسته $(E(Y_i | X_{ik}))$ ، $EF(\mu_i, \phi)$ یکی از توزیع‌های خانواده نمایی با پارامترهای میانگین (μ) و مقیاس (ϕ)، $g(\mu_i)$ تابع پیوند برای امید ریاضی متغیر وابسته، β_k و α نیز ضرایب مدل رگرسیونی و X_{ik} متغیرهای پیشگو یا مستقل می‌باشند. در پژوهش حاضر از توابع پیوند Log Identity و Inverse استفاده شده است (۲۲). مقدار μ_i یا حاصل یک مدل GLM برای هر یک از توابع پیوند مذکور به ترتیب برابر است با $g(\mu_i) = \mu_i$ ، $\exp(g(\mu_i)) = \mu_i$ و $g(\mu_i)^{-1} = \mu_i$. علاوه بر مدل‌های GLM در پژوهش حاضر سعی شد تا فرمت کلی روابط تجربی ارائه شده برای دبی‌های اوج سیلاب که همان مدل‌های رگرسیونی غیرخطی (NLR) بر روی متغیر مساحت می‌باشند نیز مورد بررسی قرار گیرد (رابطه ۲) و ضرایب آن برای مناطق مختلف واسنجی شود.

$$Q_T = cA^n \quad (2)$$

۴-۲ انتخاب پارامتر و ارزیابی مدل

در این پژوهش متغیرهای مناسب برای مدل‌سازی حداکثر دبی‌های روزانه با دوره بازگشت معین در هر منطقه با استفاده از تکنیک رگرسیون گام به گام، استخراج بهترین زیرمدل و استنباط‌های نموداری تعیین شدند. با این وجود در هیچ کدامیک از روش‌های مذکور انواع ارتباطات غیرخطی و یا خطی با حالت نمایی، درجه دو و چند جمله‌ای برای انتخاب متغیرهای پیشگو جهت مدل‌سازی لحاظ نمی‌شود. لذا نمودارهای نقطه‌ای بین هر کدام یک از متغیرهای پیشگو در مقابل متغیرهای وابسته هر منطقه به صورت جداگانه ترسیم شد و بر اساس درک نموداری متغیرهای پیشگویی که دارای روابط غیرخطی با متغیرهای وابسته بودند نیز با انواع فرمت‌های درجه‌دو، لگاریتمی و یا معکوس در فرآیند مدل‌سازی مورد آزمون قرار گرفتند.

در نهایت برای حداکثر دبی‌های روزانه متناظر با هر یک از دوره بازگشت‌های ۲، ۱۰، ۵۰ و ۱۰۰ سال در هر زیرمنطقه حداقل دو و حداکثر سه رابطه رگرسیونی به همراه برخی از آماره‌های ارزیابی شایستگی مدل‌ها شامل مجذور میانگین مربعات خطا (RMSE)، ضریب نش-ساتکلیف (NS)، ضریب تبیین اصلاح شده (R^2_{adj}) و آماره اطلاعات آکائیک (Akaike information criterion; AIC) ارائه شد. علاوه بر این به منظور ارزیابی تناسب متغیرهای انتخاب شده برای هر مدل و قابلیت مدل برای پیش‌بینی حداکثر دبی‌های

روزانه در حوضه‌های فاقد ایستگاه نیز از تکنیک جک-نایف استفاده شد. در تکنیک جک-نایف هر بار یکی از مشاهدات از کل فرآیند مدل سازی حذف شده و پارامترهای مدل مربوطه با $n-1$ عضو تعیین خواهد شد. سپس مقدار متغیر وابسته برای حوضه حذف شده با توجه مدل مربوطه پیش‌بینی خواهد شد. بدین ترتیب بر اساس سری‌های مقادیر واقعی و پیش‌بینی شده حداکثر دبی‌های روزانه حوضه‌ها، آماره‌های RMSE و NS مدل‌ها برای حالت جک-نایف نیز محاسبه شد. مقادیر RMSE کمتر و نزدیک به ۰ و NS نزدیک به ۱ در حالت جک-نایف به معنی انتخاب صحیح متغیرهای مستقل و دقت بالای مدل در پیش‌بینی است. علاوه بر این به منظور درک صحیح از تناسب خطاهای مدل‌ها نسبت به میانگین حداکثر دبی‌های روزانه در هر منطقه از آماره مجذور میانگین مربعات خطا نسبی (RRMSE) استفاده شد (رابطه ۳). چنانچه مقادیر این آماره برای یک مدل کمتر از ۱۰٪ باشد مدل عالی، بین ۱۰ تا ۲۰٪ مدل خوب، بین ۲۰ تا ۳۰٪ مدل قابل قبول و بیش از ۳۰٪ مدل ضعیف خواهد بود (۱۵). در رابطه زیر n تعداد مشاهدات، Obs_i و $Pred_i$ به ترتیب مقادیر مشاهده شده (واقعی) و مقادیر پیش‌بینی در i امین حوضه می‌باشند.

$$RRMSE = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (Obs_i - Pred_i)^2}{\sum_{i=1}^n (Pred_i)^2}} \quad (3)$$

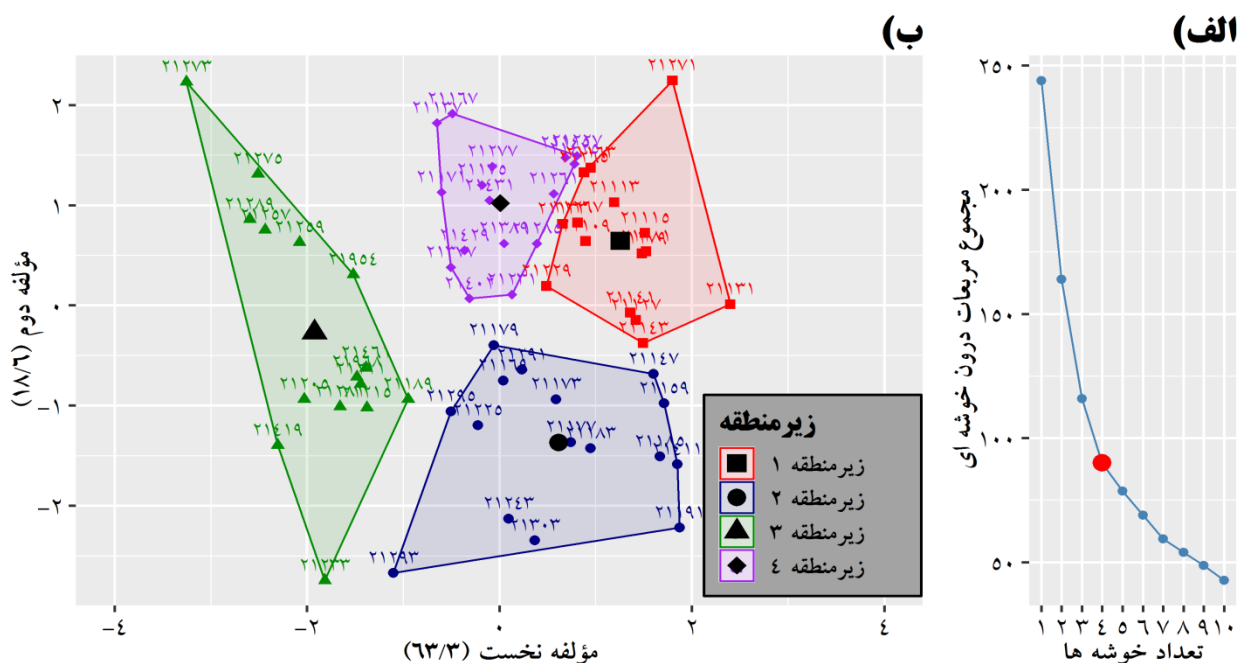
۳- نتایج

۳-۱- گروه‌بندی حوضه‌ها

از آن جایی که در پژوهش حاضر تعداد حوضه‌های مورد مطالعه و تغییرات خصوصیات مورد بررسی در آنها بسیار زیاد است، ابتدا لازم بود تا بر اساس یک روش منطقی متغیرهای مستقلی تعیین شوند که بتوانند نماینده اکثر ویژگی‌های حوضه‌های مورد مطالعه باشند. علاوه بر این گروه‌بندی بر اساس این متغیرهای مهم باید به گونه‌ای باشد که بتوان روابط منطقی و مناسبی بین برخی از ویژگی‌های حوضه‌های آبخیز و حداکثر دبی‌های روزانه در هر زیرمنطقه یافت. از این رو از تکنیک تجزیه به مؤلفه‌های اصلی بر روی لگاریتم طبیعی متغیرهای مستقل استفاده شد. نتایج نشان داد سه مؤلفه اصلی نخست که به ترتیب دارای واریانس ۶۳/۳٪، ۱۸/۶٪ و ۶/۳٪ هستند در مجموع ۸۸/۲٪ از واریانس کل ماتریس لگاریتم طبیعی متغیرهای مستقل را توجیه می‌کنند. به همین ترتیب چهار متغیر طول حوضه (BL) بر روی مؤلفه نخست، شیب متوسط (MSB) و بارش سالانه (AR) حوضه‌ها بر روی مؤلفه دوم، و ضریب شکل حوضه‌ها (FF) بر روی مؤلفه سوم دارای بیشترین بارهای عاملی بودند و به عنوان ویژگی‌های نماینده برای گروه‌بندی حوضه‌ها مورد استفاده قرار گرفتند. بر اساس چهار متغیر مذکور تعداد بهینه گروه‌ها با توجه به آماره WSS برابر چهار گروه در نظر گرفته شد (شکل ۲-الف). در شکل ۲-ب نیز موقعیت گروه‌ها و حوضه‌های مورد مطالعه بر روی دو مؤلفه نخست

ماتریس لگاریتم طبیعی متغیرهای منتخب نمایش داده شده است. همانطور که مشخص است گروه‌ها به خوبی توانسته‌اند توسط متغیرهای منتخب از یکدیگر تفکیک شوند.

بطور کلی نتایج نشان داد که حوضه‌های منطقه دوم شامل حوضه‌های بزرگ با متوسط مساحت حدود ۱۷ هزار کیلومتر مربع و آبراهه‌ی اصلی کم شیب و طول‌تر از ۱۰۰ کیلومتر می‌باشند. در حالی که حوضه‌های منطقه سوم عمدتاً حوضه‌های کوچک با رودخانه‌های کوتاه و پر شیب و با بارش و ارتفاع زیاد را شامل می‌شوند. ویژگی‌های حوضه‌های زیرمناطق اول و دوم حالت حد واسطه زیرمناطق سوم و چهارم شناسایی شد. این حوضه‌ها نیز به خوبی از نظر متغیرهایی مانند بارش سالانه و شیب متوسط حوضه‌ها از گروه‌های دیگر قابل تفکیک می‌باشند.



شکل ۲. تعیین تعداد خوشه‌ها (الف) و موقعیت خوشه‌های تعیین شده و مراکز آنها بر روی دو مؤلفه اصلی نخست (ب)

جدول ۲. میانگین مشخصات حوضه‌های مورد مطالعه به تفکیک زیرمناطق ارائه شده

خصوصیات	زیر منطقه				خصوصیات	زیر منطقه			
	۴	۳	۲	۱		۴	۳	۲	۱
محیط حوضه	۶۱/۲۸	۵۰/۳۸	۳۳۶/۶	۱۰۲/۴	طول رودخانه اصلی	۲۹۹/۸	۱۷۱/۸	۱۱۴۸/۴	۳۹۰/۸
مساحت حوضه	۲/۵۶۱	۵/۰۹	۰/۶	۱/۰۹۲	شیب آبراهه اصلی	۲۱۵۶	۶۴۴/۵	۱۷۳۲۷/۵	۳۰۱۲
طول حوضه	۲۰/۷۶	۳۷/۴۶	۲۳/۶	۱۴/۲۵	شیب متوسط حوضه	۶۷/۷۹	۵۳/۸۸	۳۵۱	۱۰۹/۲
ضریب شکل	۲۱۸۵	۲۳۹۶	۱۸۲۶/۳	۱۸۰۱	ارتفاع متوسط حوضه	۰/۲۸۵	۰/۲۱۸	۰/۱۳	۰/۲۱۴
ضریب فشردگی	۱۸۷۲	۲۱۵۱	۳۱۶۳/۲	۱۷۶۸	دامنه ارتفاعی حوضه	۲/۲۶۱	۲/۱۴۸	۲/۷	۲/۳۴۲
مجموع طول شبکه آبراهه‌ای	۵۳۳/۴	۸۷۸/۲	۵۸۸/۲	۴۵۰/۷	بارش سالانه	۱۴۷۰	۴۲۱/۳	۱۲۳۳۳/۲	۲۲۴۳

وجود اختلافات زیاد برای ویژگی‌های یکسان در گروه‌های مختلف کمک خواهد کرد که پس تعیین روابط منطقه‌ای بتوان به راحتی حوضه‌های فاقد ایستگاه هیدرومتری را به یکی زیرمناطق تعیین شده نسبت داد و با استفاده از روابط منطقه‌ای آن زیرمنطقه دبی‌های حداکثر روزانه با دوره بازگشت‌های معین را برای آن برآورد نمود. از این رو می‌توان با استفاده از تکنیک‌های مختلف اندازه‌گیری فاصله مانند فاصله اقلیدسی، فاصله هر حوضه دلخواه را تا مرکز زیرمناطق ارائه شده (جدول ۲) محاسبه کرد و حوضه مورد نظر را به زیرمنطقه‌ای که کمترین فاصله را با مرکز آن دارد نسبت داد.

۲-۳- روابط منطقه‌ای

تحلیل همبستگی بین متغیرهای مورد بررسی برای کل حوضه‌های مورد مطالعه مشخص کرد که بین دبی‌های حداکثر روزانه و ویژگی‌های خطی حوضه‌ها مانند مساحت، محیط، ضریب فشردگی، طول حوضه، مجموع طول آبراهه‌ها و دامنه ارتفاعی حوضه‌ها همبستگی مثبت و یک رابطه خطی قابل توجه وجود دارد. در حالی که ارتباط بین دبی‌های حداکثر روزانه با متغیرهای شیب آبراهه اصلی، ضریب شکل و ارتفاع متوسط حوضه‌ها یک رابطه معکوس خطی است. از سوی دیگر بسیاری از متغیرهای خطی حوضه‌ها با یکدیگر دارای همبستگی زیادی بودند؛ این امر ممکن است منجر به افزایش غیرمنطقی ضریب تبیین مدل‌ها بخاطر اثر همخطی (Colinearity) متغیرهای پیشگو شود. لذا در فرآیند مدل‌سازی سعی شد که با توجه به عامل تورم واریانس (Variance Inflation Factor; VIF) با آستانه عددی بیشتر از ۱۰، حداکثر فقط یکی از متغیرهای خطی حوضه‌ها در هر مدل که دارای بیشترین مقدار معنی‌داری بر اساس آزمون T بود استفاده شود. در جدول ۳ بهترین روابط منطقه‌ای برای حداکثر دبی‌های روزانه متناظر با هر یک از دوره بازگشت‌های ۲، ۱۰، ۵۰ و ۱۰۰ سال در هر یک از مناطق مورد مطالعه ارائه شده است. علت تنوع در مدل‌های منطقه‌ای ارائه شده در این پژوهش بدین خاطر است که بتوان مقادیر حداکثر دبی‌های روزانه را برای انواع حوضه‌های فاقد ایستگاه هیدرومتری و با حداقل اطلاعات برآورد نمود. همچنین شایان ذکر است که برای مدل‌های ارائه شده در جدول ۳ تمامی پیش شرط‌های لازم جهت تأیید مدل رگرسیونی مانند معنی‌داری ضرایب متغیرها، معنی‌داری مدل رگرسیون، عدم وجود الگو باقی‌مانده‌ها در مقابل متغیرهای پیشگو و مقادیر برآورد شده و عدم وجود همخطی بین متغیرها بررسی شده و مورد تأیید می‌باشد. با توجه به نتایج جدول ۳ مشخص شد که در صورت عدم گروه‌بندی حوضه‌ها، مدل‌های GLM به کمک متغیرهایی همچون دامنه ارتفاعی، محیط، مساحت حوضه و شیب آبراهه اصلی حداکثر تا حدود ۸۵٪ از تغییرات حداکثر دبی‌های روزانه حوضه‌های مورد مطالعه را می‌توانند توجیه کنند. مقدار خطا برای این مدل‌ها از حدود ۱۲۹ متر مکعب بر ثانیه برای دوره بازگشت ۲ سال تا حدود ۵۲۱ متر مکعب بر ثانیه برای دوره بازگشت ۱۰۰ سال می‌تواند متغیر باشد.

جدول ۳. روابط منطقه‌ای برای برآورد حداکثر دبی‌های روزانه با دوره بازگشت‌های مختلف و برخی از آماره‌های شایستگی مدل

دوره بازگشت (سال)	توزیع (لینک) مدل GLM	ضابطه ریاضی مدل $Q_{max} =$	آماره‌های ارزیابی مدل بر اساس					
			مدل			جک-نایف		
			NS	RMSE	AIC	NS	RMSE	R ² _{adj}
۲	گاما ("log")	$\exp\{3/467(R)-21/866\}^{**}$	۰/۸	۱۳۶/۲	۶۹۹/۰	۰/۸	۱۴۱/۹	۸۱/۵
	گوسی ("log")	$\exp\{0/474\ln(P)+0/00066(R)+0/825\}$	۰/۸	۱۲۹/۲	۷۸۶/۸	۰/۸	۱۵۳/۸	۸۳/۰
	تجربی	$1/215(A^{0/634})$	۰/۷	۱۸۴/۶	۸۲۹/۰	۰/۶	۱۹۴/۸	۶۵/۴
	گاما ("log")	$\exp\{3/196\ln(R)-18/925\}$	۰/۸	۲۶۸/۰	۸۰۶/۹	۰/۸	۲۷۸/۳	۸۲/۰
	گوسی ("log")	$\exp\{0/066(MSB)-0/733(SMR)+0/0008(P)+4/87\}$	۰/۹	۲۳۸/۱	۸۶۴/۶	۰/۷	۳۵۰/۷	۸۵/۳
	تجربی	$3/553(A^{0/600})$	۰/۷	۳۵۴/۹	۹۰۹/۷	۰/۷	۳۷۳/۲	۶۸/۰
	گوسی ("log")	$\exp\{3/23\ln(R)-18/606\}$	۰/۸	۴۶۱/۸	۹۴۲/۷	۰/۸	۵۰۴/۵	۸۱/۰
	گوسی ("log")	$\exp\{0/078(MSB)-0/938(SMR)+0/00035(A)+5/53\}$	۰/۹	۴۰۵/۱	۹۳۰/۵	۰/۷	۵۸۹/۵	۸۵/۰
	تجربی	$5/068(A^{0/614})$	۰/۷	۶۱۵/۹	۹۷۸/۴	۰/۶	۶۵۳/۵	۶۵/۰
	گوسی ("log")	$\exp\{3/30\ln(R)-18/963\}$	۰/۸	۶۰۵/۸	۹۷۶/۴	۰/۸	۶۵۸/۵	۷۹/۰
۵۰	گوسی ("log")	$\exp\{0/08(MSB)-0/914(SMR)+0/000037(A)+5/61\}$	۰/۹	۵۲۱/۵	۹۶۱/۸	۰/۷	۷۳۸/۶	۸۳/۹
	تجربی	$100/65/550(A^{0/626})$	۰/۷	۷۹۲/۴	۱۰۰/۶	۰/۶	۸۴۲/۸	۶۳/۶
	گاما ("iden.")	$0/019(A)+3/433$	۰/۹	۱۶/۰	۱۱۹/۸	۰/۹	۱۷/۷	۹۰/۴
	گوسی ("log")	$\exp\{0/00002(A)+0/0056(AR)+0/7\}$	۰/۹	۱۴/۵	۱۳۰/۷	۰/۷	۲۷/۸	۹۱/۵
	تجربی	$0/046(A^{0/900})$	۰/۹	۱۵/۲	۱۳۰/۲	۰/۸	۲۱/۰	۹۰/۵
	گاما ("log")	$\exp\{1/473\ln(BL)-1/943\}$	۰/۹	۴۷/۲	۱۶۰/۱	۰/۹	۵۲/۱	۸۶/۷
	گوسی ("Inv")	$\{0/00001(BL^{-1})+0/0008(SMR)-0/003\}^{-1}$	۰/۹	۴۲/۴	۱۶۳/۰	۰/۸	۵۳/۷	۸۸/۳
	تجربی	$0/45(A^{0/746})$	۰/۹	۴۹/۶	۱۶۵/۷	۰/۸	۵۸/۷	۸۴/۰
	گاما ("log")	$\exp\{1/415\ln(BL)-1/06\}$	۰/۸	۱۱۰/۵	۱۸۳/۰	۰/۸	۱۲۵/۲	۸۰/۲
	گاما ("Inv")	$\{0/321(BL^{-1})+0/004(SMR)-0/002\}^{-1}$	۰/۹	۸۹/۱	۱۸۰/۰	۰/۸	۱۱۲/۹	۸۶/۱
۱۰۰	تجربی	$1/09(A^{0/713})$	۰/۸	۱۲۸/۱	۱۹۴/۱	۰/۷	۱۵۲/۴	۷۱/۲
	گاما ("log")	$\exp\{1/397\ln(BL)-0/751\}$	۰/۸	۱۵۳/۴	۱۹۱/۴	۰/۶	۲۰۶/۹	۷۶/۹
	گاما ("Inv")	$\{0/253(BL^{-1})-0/002(SMR^{-1})+0/004\}^{-1}$	۰/۹	۱۲۸/۶	۱۸۹/۵	۰/۶	۲۱۷/۶	۸۲/۴
	تجربی	$1/41(A^{0/709})$	۰/۷	۱۷۹/۳	۲۰۴/۲	۰/۶	۲۱۲/۵	۶۶/۰
	گوسی ("log")	$\exp\{-1/093(R^1)+2/874(R)-6/22\}$	۰/۸	۱۷۲/۴	۲۱۸/۲	۰/۴	۲۸۰/۸	۷۵/۸
	تجربی	$15/014(A^{0/392})$	۰/۴	۳۸۷/۸	۲۳۲/۶	۰/۲	۳۹۰/۷	۳۲/۶
	گاما ("log")	$\exp\{0/00007(R)+125(FF^1)-31/64(FF)+6/75\}$	۰/۸	۳۲۰/۷	۲۴۰/۱	۰/۶	۴۸۱/۰	۷۵/۴
	تجربی	$38/49(A^{0/770})$	۰/۴	۵۴۲/۲	۲۵۲/۹	۰/۳	۶۰۰/۰	۳۵/۰
	گوسی ("log")	$\exp\{0/00007(R)+164(FF^2)-42/34(FF)+7/66\}$	۰/۸	۶۰۰/۰	۲۶۰/۱	۰/۵	۹۲۷/۱	۷۳/۴
	تجربی	$43/714(A^{0/406})$	۰/۴	۱۰۱۰/۸	۲۷۲/۸	۰/۳	۱۱۲۳/۱	۳۰/۳
	گوسی ("log")	$\exp\{0/0008(R)+184(FF^2)-47/89(FF)+8/05\}$	۰/۸	۸۰۰/۴	۲۶۹/۳	۰/۵	۱۲۳۳/۶	۷۱/۷
۱۰۰	تجربی	$43/16(A^{0/428})$	۰/۴	۱۳۳۳/۶	۲۸۱/۷	۰/۲	۱۴۸۵/۳	۲۷/۵

موضوعه‌های منطقه ۱

موضوعه‌های منطقه ۱

موضوعه‌های منطقه ۲

*: Iden Inv و Log به ترتیب توابع پیوند معکوس، مشخص و لگاریتمی هستند. **: ردیف‌های برجسته بیانگر بهترین مدل بر اساس آماره‌های شایستگی

ادامه جدول ۳. روابط منطقه‌ای برای برآورد حداکثر دبی‌های روزانه با دوره بازگشت‌های مختلف و برخی از آماره‌های شایستگی مدل

آماره‌های ارزیابی مدل بر اساس						ضابطه ریاضی مدل	توزیع	دوره بازگشت
جک-نایف						$Q_{max} =$	(لینک) مدل	(سال)
NS	RMSE	R^2_{adj}	NS	RMSE	AIC		GLM	
۰/۹	۴۸/۲	۹۳/۰	۰/۹	۴۰/۴	۱۵۹/۵	$\exp\{0.0061(P)+3.433\}$	گوسی ("log")	
۰/۹	۵۱/۸	۹۲/۳	۰/۹	۴۰/۴	۱۶۱/۵	$\exp\{0.0166(MRL)+0.653(SMF^{-1})+3.279\}$	گوسی ("log")	۲
۰/۴	۱۱۸/۲	۸۴/۵	۰/۹	۵۷/۷	۱۷۰/۲	$0.377(A^{0.909})$	تجربی	
۰/۹	۹۸/۹	۹۲/۴	۰/۹	۸۴/۴	۱۸۱/۶	$\exp\{0.0057(P)+4.312\}$	گوسی ("log")	
۰/۹	۱۰۰/۶	۹۴/۱	۱/۰	۷۱/۱	۱۷۸/۵	$\exp\{0.0051(P)+0.0011(AR)+3.385\}$	گوسی ("log")	۱۰
۰/۵	۲۲۲/۳	۸۶/۰	۰/۹	۱۱۰/۱	۱۸۹/۶	$1/285(A^{0.844})$	تجربی	
۰/۹	۱۵۶/۳	۹۱/۲	۰/۹	۱۴۳/۳	۱۹۷/۵	$\exp\{0.017(BL)+4.804\}$	گوسی ("log")	
۰/۹	۱۴۴/۸	۹۳/۲	۰/۹	۱۲۰/۸	۱۹۴/۴	$\exp\{0.0153(MRL)+0.0012(AR)+3.843\}$	گوسی ("log")	۵۰
۰/۷	۵۰۸/۴	۶۹/۹	۰/۷	۲۵۴/۸	۲۱۴/۸	$3/469(A^{0.765})$	تجربی	
۰/۹	۲۳۶/۵	۸۹/۹	۰/۹	۱۸۸/۶	۲۰۵/۸	$\exp\{0.0181(BL)+4.906\}$	گوسی ("log")	
۰/۸	۲۴۹/۹	۹۱/۳	۰/۹	۱۶۷/۹	۲۰۴/۳	$\exp\{0.0161(MRL)+0.0011(AR)+4.001\}$	گوسی ("log")	۱۰۰
۰/۶	۷۰۶/۸	۶۰/۶	۰/۷	۳۵۶/۷	۲۲۴/۹	$4/99(A^{0.737})$	تجربی	
۱/۰	۱۸/۴	۹۷/۲	۱/۰	۱۷/۲	۱۴۲/۵	$\exp\{0.0158(BL)+3.032\}$	گوسی ("log")	
۰/۹	۲۵/۵	۹۴/۷	۱/۰	۲۳/۹	۱۵۲/۹	$\exp\{0.00025(A)+3.612\}$	گوسی ("log")	۲
۰/۸	۵۰/۶	۹۰/۹	۰/۹	۳۰/۲	۱۶۰/۴	$0.023(A^{1/0.561})$	تجربی	
۰/۹	۷۲/۶	۹۴/۷	۱/۰	۵۲/۷	۱۷۸/۲	$\exp\{0.015(BL)+3.969\}$	گوسی ("log")	
۰/۹	۸۰/۸	۹۲/۷	۰/۹	۶۱/۹	۱۸۳/۴	$\exp\{0.000236(A)+4.523\}$	گوسی ("log")	۱۰
۰/۹	۸۹/۷	۹۰/۰	۰/۹	۶۸/۰	۱۸۶/۴	$0.089(A^{0.997})$	تجربی	
۰/۹	۱۱۹/۸	۹۱/۷	۰/۹	۱۱۲/۰	۲۰۲/۴	$\exp\{0.0152(BL)+4.463\}$	گوسی ("log")	
۰/۹	۱۱۸/۱	۹۲/۱	۰/۹	۱۰۹/۵	۲۰۱/۷	$\exp\{0.0157(MRL)+4.541\}$	گوسی ("log")	۵۰
۰/۷	۲۱۳/۷	۸۴/۰	۰/۹	۱۴۶/۸	۲۱۱/۱	$0.163(A^{0.887})$	تجربی	
۰/۹	۱۷۹/۳	۸۹/۸	۰/۹	۱۵۴/۱	۲۱۲/۶	$\exp\{0.0154(BL)+4.637\}$	گوسی ("log")	
۰/۹	۱۶۵/۲	۹۰/۶	۰/۹	۱۴۷/۸	۲۱۱/۳	$\exp\{0.0159(MRL)+4.715\}$	گوسی ("log")	۱۰۰
۰/۶	۳۱۰/۰	۸۰/۹	۰/۸	۲۰۲/۴	۲۲۱/۳	$0.199(A^{0.987})$	تجربی	

حوضه‌های منطقه ۲

حوضه‌های منطقه ۴

*: Iden Inv و Log به ترتیب توابع پیوند معکوس، مشخص و لگاریتمی هستند. **: ردیف‌های برجسته بیانگر بهترین مدل بر اساس آماره‌های شایستگی

بطورکلی نتایج نشان داد که برای دوره بازگشت ۱۰۰ سال که به عنوان یکی از دوره بازگشت‌های مهم برای طراحی اغلب

سازه‌های هیدرولیکی در نظر گرفته می‌شود، ضریب تبیین اصلاح شده روابط تجربی تنها حدود ۶۳/۶٪ و مقدار خطای آن ۷۹۲ متر

مکعب بر ثانیه است. در حالی که مدل GLM با توزیع گوسی و تابع پیوند لگاریتمی و متغیرهای پیشگو شیب آبراهه اصلی و

مساحت حوضه با عملکرد بهتر می‌تواند تا ۸۴٪ از تغییرات حداکثر دبی‌های روزانه ۱۰۰ ساله را با خطای ۵۲۱ متر مکعب بر ثانیه

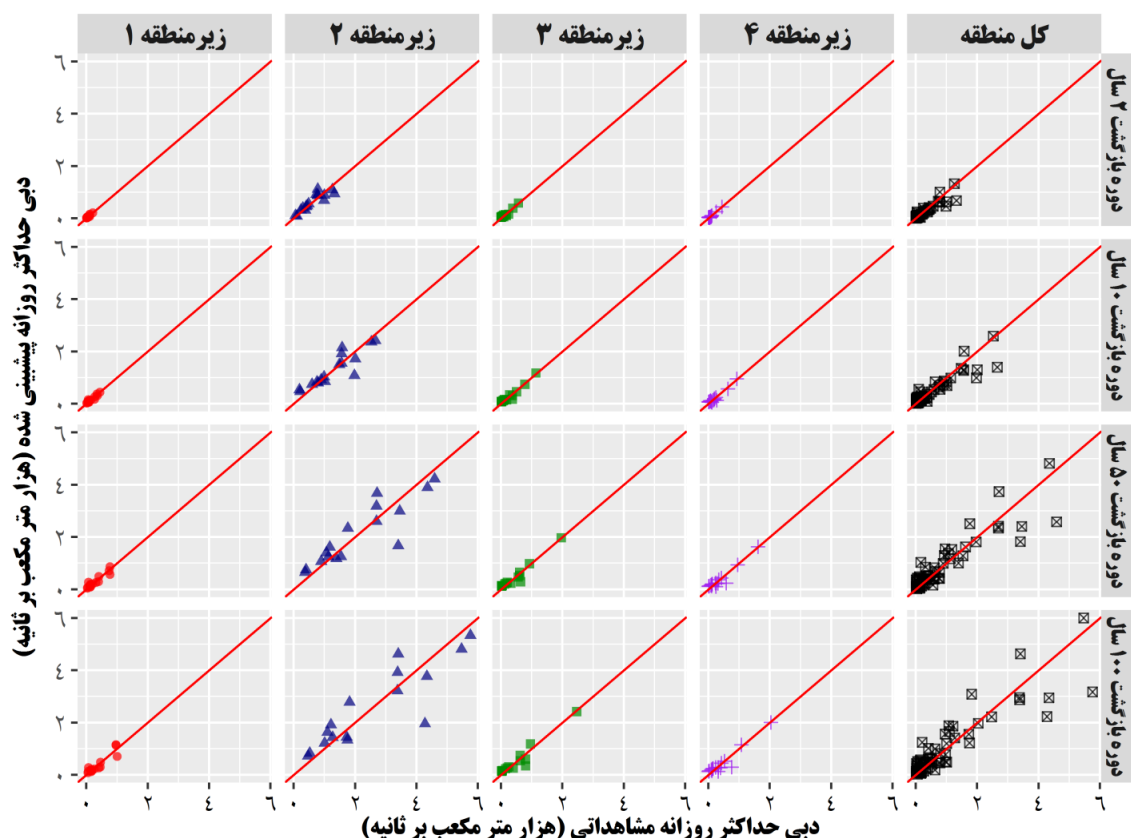
توجیه کند. با این وجود چه مدل‌های GLM و چه مدل‌های تجربی در حالتی که از تمامی حوضه‌ها برای مدل‌سازی استفاده شوند دارای خطای قابل توجهی به ویژه برای حالت پیش‌بینی در روش جک-نایف هستند.

از این رو فرآیند مدل‌سازی برای هر یک از زیرمناطق تعیین شده به صورت مجزا نیز انجام شد (جدول ۳). با توجه به نتایج مشخص شد که به غیر از زیرمنطقه شماره دو که شامل زیرحوضه‌های بسیار بزرگ هستند ($23000 \text{ km}^2 < A < 10000$)، در تمامی زیرمناطق فرآیند مدل‌سازی موفق بوده است. بطوریکه مقادیر ضرایب تبیین اصلاح شده مناسب‌ترین مدل‌ها برای حداکثر دبی‌های ۱۰۰ ساله در زیرمناطق اول، سوم و چهارم به ترتیب $0.82/4$ ، $0.91/3$ و $0.90/6$ و مقادیر خطای آنها به ترتیب $128/58$ ، $167/9$ و $147/8$ متر مکعب بر ثانیه برآورد گردید.

با این وجود برای حوضه‌های زیرمنطقه دوم که بالاترین مقادیر حداکثر دبی‌های روزانه -بطور متوسط 4936 متر مکعب بر ثانیه برای دوره بازگشت ۱۰۰ سال- و سیستم زهکشی بسیار پیچیده‌ای دارند مدل GLM با خانواده گوسی و تابع پیوند لگاریتمی توانسته است با متغیرهای پیشگو ضریب شکل و دامنه ارتفاعی حوضه‌ها تغییرات حداکثر دبی‌های روزانه ۱۰۰ ساله را تا $0.71/7$ با خطای 800 متر مکعب بر ثانیه توجیه کند. همچنین نتایج نشان داد که در زیرمنطقه اول شیب آبراهه اصلی به همراه مساحت و طول حوضه‌ها از جمله خصوصیات فیزیوگرافیکی مؤثر بر حداکثر دبی‌های روزانه حوضه‌های این منطقه هستند. به همین ترتیب در زیرمنطقه دوم متغیرهای دامنه ارتفاعی و ضریب شکل حوضه‌ها و در زیرمناطق سوم و چهارم متغیرهایی همچون طول حوضه، طول آبراهه اصلی و بارش سالانه از جمله عوامل کنترل کننده حداکثر دبی‌های روزانه در حوضه‌های مورد مطالعه می‌باشند.

علاوه بر این بر اساس مدل‌های برتر جدول ۳ حداکثر دبی‌های روزانه در هر یک از حوضه‌های مورد مطالعه به روش جک-نایف نیز پیش‌بینی و در مقابل مقادیر مشاهداتی آنها (۴) در شکل ۳ نمایش داده شد. همانطور که مشخص است برای دوره بازگشت‌های کوتاه مدت ۲ و ۱۰ سال تقریباً تمامی مدل‌ها به خوبی توانسته‌اند حداکثر دبی‌های روزانه را پیش‌بینی کنند، بطوریکه فاصله نقاط با خط یک به یک بسیار ناچیز است. این امر برای حداکثر دبی‌های روزانه با دوره بازگشت‌های ۵۰ و ۱۰۰ سال در زیرمناطق اول، سوم و چهارم نیز صادق است. این در حالی است که با افزایش دوره بازگشت‌ها (۵۰ و ۱۰۰ سال)، مدل‌های زیرمنطقه دوم و کل منطقه برای پیش‌بینی حداکثر دبی‌های روزانه حوضه‌ها با دبی بیش از 2000 متر مکعب بر ثانیه اختلاف زیادی را حاصل می‌شوند. برای نمونه حوضه‌های 21185 ، 21243 و 21295 به ترتیب با مساحت $39/45$ ، $32/75$ و $16/12$ هزار کیلومتر مربع جزو زیرحوضه‌های زیرمنطقه دوم هستند و سه نمونه بارز از مواردی می‌باشند که مقدار اختلاف بین مقادیر واقعی این حوضه‌ها از تحلیل فراوانی داده‌های ایستگاه‌های هیدرومتری- و مقادیر پیش‌بینی شده حداکثر دبی‌های روزانه آنها توسط مدل‌های منطقه‌ای

بسیار زیاد است. البته باید خاطر نشان کرد که این امر برای حوضه‌های بسیار بزرگ مانند موارد ذکر شده می‌تواند کاملاً طبیعی باشد؛ چرا که در حوضه‌های بسیار بزرگ عوامل متعددی فرآیندهای هیدرولوژیکی حوضه‌ها را تحت تأثیر قرار خواهند داد و توجیه چنین فرآیندهای پیچیده‌ای توسط یک یا چند متغیر پیشگو همواره همراه با خطا و عدم قطعیت‌های بسیاری خواهد بود. با این وجود نتایج پژوهش حاضر نشان داد که در حوضه‌های کوچکتر مدل‌های رگرسیونی GLM به خوبی قادر به پیش‌بینی حداکثر دبی‌های روزانه حتی برای دوره بازگشت‌های بلند مدت نظیر ۵۰ و ۱۰۰ سال می‌باشند.



شکل ۳. دبی‌های حداکثر روزانه مشاهداتی در مقابل دبی‌های پیش‌بینی شده توسط مدل‌های منطقه‌ای به روش جک-نایف برای مناطق و دوره بازگشت‌های مختلف. خط قرمز رنگ خط ۱:۱ است.

از سوی دیگر به منظور استاندارد کردن خطای مدل‌های ارائه شده و مقایسه مدل‌های زیرمناطق مختلف با یکدیگر در جدول ۴ آماره مجذور میانگین مربعات خطا نسبی (RRMSE) برای هر یک از زیرمناطق و دوره بازگشت‌های مورد مطالعه نمایش داده شده است. باتوجه به مدل‌های برتر در جدول ۳ (ردیف‌های برجسته)، مقادیر RRMSE به جز مدل‌های منطقه دوم که در طبقه خوب ($10\% < RRMSE < 20\%$) هستند سایر مدل‌های ارائه شده در پژوهش حاضر در طبقه عالی ($RRMSE < 10\%$) قرار دارند. از این حیث حتی برای دوره بازگشت‌های بلند مدت، مدل‌ها از عملکرد خوبی برخوردار هستند. بطورکلی نتایج نشان می‌دهد که مناسب‌ترین مدل‌ها به ترتیب متعلق به مدل‌های منطقه چهارم، سوم، اول و سپس دوم هستند. نکته شایان ذکر دیگر مقادیر RRMSE کم برای مدل‌های کل منطقه است، در این مدل‌ها بخاطر اینکه اندازه نمونه زیاد است (۶۲ حوضه) در نهایت مقادیر پیش

بینی شده با دقت زیاد (خطای کم) در حوضه‌های کوچک (با تعداد زیاد) میانگین خطا مدل‌ها را به ویژه برای حوضه‌های بزرگ (با تعداد کم) کاهش خواهند داد. لذا برای ارزیابی و مقایسه مدل‌ها با یکدیگر لازم است تا چندین آماره شایستگی (جدول ۳) بررسی و در نهایت مناسب‌ترین مدل‌ها به عنوان روابط منطقه‌ای سیلاب مورد استفاده قرار گیرند.

جدول ۴. مقادیر RRMSE برای مدل‌های منطقه‌ای به همراه کلاس دقت آنها برای پیش‌بینی

منطقه	مقادیر RRMSE برای بازگشت‌های			
	۲-سال	۱۰-سال	۵۰-سال	۱۰۰-سال
زیرمنطقه ۱	۵/۵۱ (عالی)	۶/۴۹ (عالی)	۷/۴۰ (عالی)	۹/۴۹ (عالی)
زیرمنطقه ۲	۱۰/۳۴ (خوب)	۸/۷۷ (عالی)	۱۰/۴۳ (خوب)	۱۱/۳۴ (خوب)
زیرمنطقه ۳	۵/۹۹ (عالی)	۶/۰۰ (عالی)	۵/۷۳ (عالی)	۹/۲۳ (عالی)
زیرمنطقه ۴	۳/۳۵ (عالی)	۵/۴۱ (عالی)	۵/۶۱ (عالی)	۶/۸۱ (عالی)
کل منطقه	۵/۵۱ (عالی)	۵/۲۴ (عالی)	۴/۹۶ (عالی)	۵/۲۷ (عالی)

۴- بحث و جمع‌بندی

در پژوهش حاضر سعی شد تا با استفاده از اطلاعات پژوهش (۴) روابط منطقی و منطقه‌ای ارائه شود که بتوان با استفاده از آنها حداکثر دبی‌های روزانه را برای حوضه‌های فاقد ایستگاه هیدرومتری نیز با دقت مناسبی برآورد نمود. این امر از یک سو با توجه به فراوانی کم ایستگاه‌های هیدرومتری و از یک سو اهمیت وقایع سیلابی که اخیراً در غرب کشور اتفاق افتاده است از لحاظ اقدامات مدیریتی و کاهش سیل حائز اهمیت است (۲۱). هدف ما از پژوهش حاضر برآورد حداکثر دبی‌های روزانه با استفاده از متغیرهای پیشگو ساده و با ثبات در طول زمان بود. با این حال در صورتی که هدف از مدل‌سازی شناسایی عوامل کنترل‌کننده سیلاب‌ها برای مدیریت و کاهش اثرات آن باشد، لازم است تا متغیرهای قابل کنترل بشر و موثر بر سیلاب‌ها نظیر کاربری‌ها و شرایط پوشش گیاهی نیز در مدل‌سازی لحاظ شوند. برای نمونه الله‌بخشیان فارسانی و همکاران (۱) در مطالعه مشابه خود در محدوده حوضه‌های کارون و کرخه نشان دادند که نوع کاربری به ویژه اراضی آیش می‌تواند تأثیرات جدی بر روی خصوصیات دبی سیلاب‌ها داشته باشد. با این وجود وسعت اراضی آیش می‌تواند از حوضه‌ای به حوضه دیگر و تحت مدیریت‌های مختلف و یا حتی در طول زمان با تغییر شرایط ترسالی‌ها و خشکسالی‌ها متغیر باشد. لذا ارائه مدل‌های منطقه‌ای بر اساس چنین خصوصیتی که در طول زمان می‌توانند تغییرپذیر باشند نیازمند به بروز رسانی و واسنجی کردن متداوم مدل دارد. از این رو چنین مدل‌هایی ممکن است در برخی از دوره‌های زمانی چندان معتبر نباشند و از سوی دیگر تهیه متغیرهایی مانند کاربری نیازمند وقت و هزینه زیاد است که ممکن در بسیاری از موارد مقرون به صرفه نباشد. بنابراین ارائه مدل‌های منطقه‌ای با متغیرهای با ثبات نظیر خصوصیات فیزیوگرافی در

حوضه‌ها نسبت به مدل‌های منطقه‌ای با متغیرهای تغییرپذیر در زمان و مکان نظیر کاربری و پوشش گیاهی می‌تواند در بسیاری از موارد ارجح باشد.

همچنین یافته‌های پژوهش حاضر نشان داد که مساحت حوضه‌ها به تنهایی قابلیت توجیه حداکثر دبی‌های روزانه را ندارد. به عبارت دیگر تغییرات حداکثر دبی‌های روزانه در زیرحوضه‌های منطقه مورد مطالعه تحت تأثیر چندین متغیر متفاوت است. پیش از این رحیمی و همکاران (۱۷) نیز بر اساس نتایج مبتنی بر تکنیک مقیاس مکانی بر روی حوضه‌های کرخه به وضوح بر این موضوع تأکید داشتند که از متغیر مساحت حوضه‌ها به تنهایی نمی‌توان برای مدل‌سازی و ارائه روابط منطقه‌ای سیل در محدوده حوضه کرخه استفاده کرد. از سوی دیگر مقایسه نتایج پژوهش حاضر با مطالعات مشابه که به مدل‌سازی منطقه‌ای سیل با استفاده از روش‌های یادگیری ماشین در حوضه‌های کارون و کرخه پرداخته بودند (۱ و ۵)، نشان داد که اولاً مهمترین متغیرهای فیزیوگرافی تأثیرگذار بر سیلاب‌های حوضه‌های کارون بزرگ و کرخه عبارت هستند از محیط، ضریب شکل، طول آبراهه اصلی و طول حوضه‌ها و دوم اینکه مدل‌های GLM با توزیع احتمال و تابع پیوند مناسب می‌توانند با روابط ساده‌تر و قابل کاربرد برای همه کاربران دقتی برابر و یا بالاتری از بسیاری از روش‌های پیچیده ماشین یادگیری و یا فرم رگرسیون لگاریتمی ساده داشته باشند. بطورکلی نتایج پژوهش حاضر نشان داد که متغیرهای خطی یک بعدی مانند طول آبراهه اصلی، محیط و طول حوضه‌ها نسب به متغیرهای خطی چند بعدی مانند مساحت حوضه‌ها دارای همبستگی بیشتری با حداکثر دبی‌های روزانه برای حوضه‌های مورد مطالعه بودند. همچنین مشخص شد که برای سیلاب‌ها با دوره بازگشت‌های کوتاه مدت ۲ و ۱۰ سال متغیرهایی همچون طول، محیط و مساحت حوضه‌ها تأثیر زیادی بر روی بزرگی حداکثر دبی‌های روزانه خواهند داشت. در حالی که برای دوره بازگشت‌های بلند مدت ۵۰ و ۱۰۰ سال اغلب خصوصیات آبراهه‌ها و شبکه زهکشی حوضه‌ها مانند طول آبراهه اصلی، مجموع طول سیستم رودخانه‌ها و شیب آبراهه اصلی حداکثر دبی‌های روزانه را تحت تأثیر قرار خواهند داد (جدول ۳).

همچنین نتایج نشان داد که برخی از متغیرها مانند شیب آبراهه اصلی، ضریب شکل و دامنه ارتفاعی حوضه‌ها دارای روابط غیرخطی با متغیر حداکثر دبی‌های روزانه در حوضه‌های مورد مطالعه بودند. بطوریکه برای توجیه تغییرات حداکثر دبی‌های روزانه توسط این متغیرها در حالت‌های کل منطقه و حوضه‌های منطقه دوم توابع پیوند لگاریتمی و معکوس برای مدل‌ها بکار گرفته شد و حتی خود این متغیرها نیز به صورت چند جمله‌ای و یا لگاریتمی و معکوس در سمت راست مدل‌های منطقه‌ای قرار گرفتند. این امر حاکی از ماهیت غیرخطی بسیاری از فرایندهای هیدرولوژیکی است که در اغلب مطالعات مدل‌سازی سیلاب به صورت خطی فرض شده‌اند (۱۸).

در نهایت باید اظهار داشت که مدل‌های ارائه شده در پژوهش حاضر نه تنها برای طراحی مناسب سازه‌های هیدرولیکی و پروژه‌های کاهش خطر سیل در حوضه‌های فاقد ایستگاه هیدرومتری قابل کاربرد هستند، بلکه حتی برای تدقیق نتایج تحلیل فراوانی در حوضه‌هایی که دارای آمار مناسب می‌باشند هم نیز قابل کاربرد هستند. بطوریکه می‌توان از نتایج آنها برای انتخاب توزیع‌های آماری مناسب و برآورد صحیح‌تر پارامترهای توزیع‌های آماری نیز استفاده نمود.

منابع

1. Allahbakhshian-Farsani, P., M. Vafakhah, H. Khosravi-Farsani and E. Hertig. 2020. Regional flood frequency analysis through some machine learning models in semi-arid regions. *Water Resources Management* 34(9): 2887-2909.
2. Aziz, K., A. Rahman, G. Fang and S. Shrestha. 2014. Application of artificial neural networks in regional flood frequency analysis: a case study for Australia. *Stochastic environmental research and risk assessment* 28(3): 541-554.
3. Chandramohan, T., MK. Jose, BK. Purandara and B. Venkatesh. 2018. Revision of Empirical Coefficients of Commonly Used Flood Formulae Using Flow Data from Karnataka Rivers. In: Hydrologic modeling: Select Proceedings of International Conference on Water, Environment, Energy and Society (ICWEES-2016), Madhya Pradesh, India, 555-563.
4. Chitsaaz, V., and P. Mohit Isfahani. 2022. Regional Flood Frequency Analysis in West of Iran. *Iran-Water Resources Research* 18(1): 84-103.
5. Ghaderi, K., B. Motamedvaziri, M. Vafakhah and AA. Dehghani. 2019. Comparison of Artificial Neural Network, Decision Tree and Bayesian Network Models in Regional Flood Frequency Analysis using L-moments and Maximum Likelihood Methods in Karkheh and Karun Watersheds. *Journal of Water and Soil Science* 25(31): 313-326.
6. Guisan, A., J.R.T.C. Edwards and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling* 157(2-3): 89-100.
7. Guldemir, H., and A. Sengur. 2006. Comparison of clustering algorithms for analog modulation classification. *Expert Systems with Applications* 30: 642-649.
8. Hailegeorgis T.T., and K. Alfredsen. 2017. Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies* 9: 104-126.
9. Hosking, J.R.M, and J.R. Wallis. 1997. Regional frequency analysis: An approach based on l-moments. Cambridge University Press, Cambridge.
10. Kordrostami, S., M.A. Alim, F. Karim, and A. Rahman. 2020. Regional flood frequency analysis using an artificial neural network model. *Geosciences* 10(4): 127.
11. Mehdi Nasab, M. 2020. Survey of 1 April Flood in Kashkan Catchment in Lorestan Province and Presenting Solutions. *Journal of Environment and Suprasectorial Development* 5(67): 17-30.
12. Mosaffaie, J. 2016. Application of artificial neural network, multiple-regression and index-flood techniques in regional flood frequency estimation. *International Journal of Water* 10(4): 328-342.
13. Msilini, A., C. Charron, T.B. Ouarda, and P. Masselot. 2022. Flood frequency analysis at ungauged catchments with the GAM and MARS approaches in the Montreal region, Canada. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques* 1-11.
14. Nelder, J.A., and RW. Wedderburn. 1972 Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* 135(3): 370-384.
15. Padhma, M. 2020. End-to-End Introduction to Evaluating Regression Models. Available online at: <https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/>. Accessed 1 September 2022.
16. Panseira, W.A., B.M. Gomes, M.A.V. Boas and E.L. de Mello. 2013. Clustering rainfall stations among regional frequency analysis. *Journal of Food, Agriculture and Environment* 11(2): 877-885.
17. Rahimi, M., R. Modarres and S. Soltani. 2021. Comparison of spatial scaling of peak discharge in Caspian sea basins and Karkheh river basin. *Iran-Water Resources Research* 16(4): 241-249.
18. Rahman, A., C. Charron, T.B. Ouarda and F. Chebana. 2018 Development of regional flood frequency analysis techniques using generalized additive models for Australia. *Stochastic environmental research and risk assessment* 32(1): 123-139.
19. Rahman, A.S. and A. Rahman 2020. Application of principal component analysis and cluster analysis in regional flood frequency analysis: a case study in New South Wales, Australia. *Water* 12(3): 781.

20. Sharifi Garmdareh, E., M. Vafakhah and S. Eslamian. 2019. Assessment the Performance of Support Vector Machine and Artificial Neural Network Systems for Regional Flood Frequency Analysis (A Case Study: Namak Lake Watershed). *Journal of Water and Soil Science* 23(1): 351-366.
21. Special Panel on National Floods Report (SPNFR). 2020. Story, Analyze, Experiences and Offers: 314p.
22. Wood, S. 2006. Generalized Additive Models: *An Introduction with R*. CRC Press.

Providing Regional Models for Maximum Daily Discharges of Iran's Western Sub-Basins Using Generalized Linear Models (GLMs)

Abstract

Introducing reliable regional models to predict maximum discharge of floods using characteristics of sub-basins has a special importance in terms of flood management and designing hydraulic structures in basins that have no any hydrometric station. Hence, in the present study, we aimed to provide appropriate regional flood models, using Generalized Linear Models (GLMs), to estimate 2-, 10-, 50- and 100- year maximum daily discharges of 62 sub-basins in Great-Karoon and Karkhe basins. According to the results, the sub-basins were categorized into four sub-regions based on some physiographic and climatic characteristics of study sub-basins. Overall, results showed that regional flood modeling was successful in all sub-regions except the sub-region II, which includes very large basins ($A \approx 17300 \text{ km}^2$). As the adjusted R^2 of the best models in sub-regions I, III and IV were estimated around 82.4, 91.3 and 90.6 percent and these models have a relative error (RRMSE) around 9.5, 9.23 and 6.7 percent, respectively. Also, it was found that more frequent floods with 2- and 10- year return-periods influenced by properties such as basin's length, perimeter and area, while rare floods with 50- and 100- year return-periods are mostly influenced by river system characteristics such as the main river length, total lengths of river system and slope of the main river. According to the research, it can be stated that the behavior of maximum daily discharges in the study area is extremely influenced by the different climatic and physiographic characteristics of the watersheds. Therefore, with appropriate modeling of maximum daily discharges in gauged catchments, it can be estimated accurately at ungauged sites.

Keywords: Flood, Watershed, Modeling, Nonlinear Relations, Empirical Flood Modeling.