Archive of SID

مجله پژوهش‌های حفاظت آب و خاک
جلد هفدهم، شماره چهارم، ۱۳۸۹
www.gau.ac.ir/journals

کاربرد روش رگرسیون خطی فازی در برآورد داده‌های ناقص دبی سالانه ایستگاه‌های هیدرومتری و مقایسه آن با سایر روش‌های متداول

*سیدجواد ساداتی‌نژاد[۱]، رحیم نقدی[۲] و محمد شایان‌نژاد[۳]

[۱]استادیار گروه مرتع و آبخیزداری، دانشگاه شهرکرد، [۲]دانش‌آموخته کارشناسی‌ارشد گروه مهندسی آب، دانشگاه شهرکرد، [۳]استادیار گروه مهندسی آب، دانشگاه شهرکرد

تاریخ دریافت: ۸۸/۵/۱۶ ؛ تاریخ پذیرش: ۸۹/۷/۱٤

چکیده

داده‌های ناقص دبی یکی از مشکلات پیش‌روی متخصصان و طراحان پروژه‌های منابع آب است و باعث بروز خطا در نتایج مطالعات طرح‌ها شده و اجرای پروژه‌ها را دچار مشکل می‌نماید. در مناطقی که ایستگاه‌های هیدرومتری از نظر تعداد محدود می‌باشند این مسأله حادتر می‌باشد. بنابراین لازم است این نواقص آماری به طریقی برطرف گردد. روش‌های متعددی برای تخمین داده‌های گمشده وجود دارد. در این پژوهش کارایی روش رگرسیون خطی فازی در مقایسه با روش‌های رگرسیون خطی ساده، رگرسیون چندمتغیره و روش محور مختصاتی و نسبت نرمال در مورد بازسازی دبی سالانه ایستگاه‌های هیدرومتری حوضه آبریز کارون بزرگ مورد ارزیابی قرار گرفته است. در این پژوهش با استفاده از روش حذف اعتباری در ۲۷ ایستگاه هیدرومتری حوضه کارون بزرگ در قالب ۱۱ گروه بازسازی پس از حذف عمدی داده‌های مشاهده‌ای دبی ایستگاه‌ها، مقادیر آن‌ها از طریق روش‌های ذکر شده برآورد گردید و آماره جذر میانگین مجذور مربعات خطا به‌ترتیب برای روش‌های ذکر شده برابر با ۱۷/٤۱، ۱٦/۰۵، ۲۱/۰، ۱۱۹/٦ و ۱۷/٦ به‌دست آمد. بنابراین روش رگرسیون خطی ساده به‌عنوان روش برتر در بازسازی داده‌های دبی سالانه تشخیص داده شد. همچنین روش رگرسیون خطی فازی به‌عنوان اولویت دوم جهت بازسازی داده‌های دبی در این حوضه شناخته شد.

واژه‌های کلیدی: رگرسیون خطی فازی، داده‌های گمشده، بازسازی داده‌ها، دبی سالانه، حوضه آبریز کارون

* مسئول مکاتبه: jsadatinejad@kashanu.ac.ir

www.SID.ir

Archive of SID

مقدمه

اساس و پایه مطالعات هیدرولوژی، داده‌های آماری مورد قبول می‌باشد با توجه به خلأهای گسسته و پیوسته در اغلب داده‌های هیدرولوژی مانند دبی رودخانه‌ها که به دلایلی مانند ثبت نکردن آمار، حذف آمار غلط و خرابی یا از بین رفتن دستگاه‌های اندازه‌گیری اتفاق می‌افتد، تخمین و برآورد این داده‌ها لازم و ضروری است. دسترسی به داده‌های کافی و دقیق از یک‌سو موجب کوتاه‌تر شدن مدت مطالعات و از سوی دیگر موجب برآورد دقیق‌تر پارامترهای هدف و کاهش هزینه‌های اجرایی و خسارات بعدی ناشی از اجرای طرح‌ها می‌گردد (لوک‌زاده، ۲۰۰٤).

برای بازسازی و تکمیل داده‌های مختلف از جمله هواشناسی و هیدرولوژیکی از روش‌های مختلفی استفاده می‌شود که بعضی از این روش‌های کلاسیک بوده اما به تازگی از روش‌های جدید و نوینی هم استفاده شده است. از جمله روش‌های کلاسیک می‌توان به روش ایستگاه معرف، نسبت نرمال، رگرسیون خطی و محور مختصات و از روش‌های جدید می‌توان به روش شبکه عصبی مصنوعی، منطق فازی و رگرسیون فازی اشاره نمود.

با بررسی منابع مختلف مشاهده گردید که پژوهش‌های نه چندان زیادی در خصوص بازسازی و یا تطویل آمار هیدرولوژیکی صورت گرفته است و پژوهش‌های صورت گرفته بیش‌تر در زمینه آمار هواشناسی می‌باشد که در اینجا به بخش‌هایی از این پژوهش‌ها اشاره می‌شود.

علیزاده و صلوی‌تبار (۱۹۸۹) با استفاده از آمار یک دوره ۳۰ ساله ایستگاه شیرگاه واقع بر رودخانه تالار نسبت به تولید آمار آب‌دهی رودخانه یاد شده اقدام کردند. در این مطالعه از مدل‌های استوکاستیک (آماری) استفاده شده است. از بین مدل‌های اتورگرسیو، ترکیب اتورگرسیو و میانگین متحرک، زنجیره مارکف و...، روش اتورگرسیون به‌دلیل کاربرد وسیع‌تر در هیدرولوژی و سادگی آن مورد استفاده قرار گرفت. در نهایت با تست معنی‌دار بودن مدل دوسری آمار ۵۰ ساله و ۱۰۰ ساله برای رودخانه یاد شده تولید گردیدکه با بررسی انحراف معیار اختلاف معنی‌داری مشاهده نشد.

ساداتی‌نژاد (۱۹۹۷) در مقایسه آماری روش‌های مختلف بازسازی داده‌های بارش سالانه در استان اصفهان به‌منظور تعیین بهترین روش در هر اقلیم استان یاد شده، ۷ روش بازسازی شامل ایستگاه معرف، نسبت نرمال، محور مختصاتی، رگرسیون خطی ساده، رگرسیون خطی چندمتغیره و اتورگرسیو را مورد بررسی قرار داد که در این مقایسه روش نسبت نرمال را برای اقلیم‌های خشک و مدیترانه‌ای و روش رگرسیون خطی چندمتغیره را برای اقلیم نیمه‌خشک به‌عنوان روش‌های برتر معرفی نمود.

www.SID.ir

Archive of SID

خلیلی (۱۹۹۸) در طرح جامع آب کشور به‌منظور بازسازی خلأهای آماری بارندگی روش تک‌صفحه‌ای و صفحات مکرر را پیشنهاد داد. به این ترتیب که هر ایستگاه را به منزله نقطه‌ای در فضای سه‌بعدی در نظر گرفت که دارای طول، عرض، و بارندگی است. او ٤ ایستگاه را در اطراف ایستگاه بدون آمار انتخاب کرد. در روش تک‌صفحه‌ای بر هر ٤ ایستگاه یک صفحه با رعایت اصل کم‌ترین مربعات عبور می‌کند، و با توجه به طول و عرض ایستگاه بدون آمار مقداری برآورد می‌گردد. در روش صفحات مکرر بر هر سه ایستگاه یک صفحه عبور داده شده و در حالات مختلف مقادیری متفاوت برای ایستگاه فاقد آمار برآورد می‌شود. او این روش‌ها را با سایر روش‌های بازسازی شامل روش نسبت نرمال، همبستگی با بالاترین ضریب، همبستگی‌های مکرر و همبستگی‌های چندمتغیره مقایسه کرد. در نهایت با توجه به فاصله زیاد ایستگاه‌ها روش همبستگی با دو یا سه متغیر را توصیه نمود.

شرکت مهندسان مشاور جاماب (۱۹۹۹) در طرح جامع آب کشور به‌منظور تکمیل آمار دبی از روابط همبستگی خطی و لگاریتمی بین مقادیر متناظر ماهانه و سالانه ایستگاه‌های مجاور یا آن‌هایی که در محدوده‌های هیدرولوژیکی مشابه قرار دارند استفاده کرد. به این طریق آمار ۵۵ ایستگاه در حوضه کارون و دز کامل گردید و برای تعیین آب‌دهی در محل پروژه‌های موردنظر و ارزیابی پتانسیل منابع آب سطحی به‌کار رفت. از مقایسه نسبت آب‌دهی دوره‌های کوتاه‌مدت به بلندمدت در آمار ایستگاه‌هایی که آمار کامل دارند، برای ارزیابی آب‌دهی در مناطقی که اطلاعات محدود وجود داشته استفاده شد. در مطالعه یاد شده به‌میزان دقت روش بالا جهت رفع خلأهای آمار موجود اشاره‌ای نشده است.

لوک‌زاده (۲۰۰٤) مقایسه‌ای بین روش‌های مختلف بازسازی خلأهای آماری بارندگی در مقاطع مختلف زمانی مختلف در منطقه البرز مرکزی با روش‌های رگرسیون خطی (۱، ۲ و ۳ متغیر مستقل)، نسبت نرمال، عکس مجذور فاصله و روش‌های زمین آماری انجام داد. به این منظور ۱۸ ایستگاه که دارای یک دوره ۲۷ ساله آماری کامل و بدون خلأ بود انتخاب گردید. با حذف ۲۰ درصد آمار هر ایستگاه به‌طور مساوی بین سال‌های خشک، نرمال و تر (ایجاد خلأ مصنوعی) در مقاطع زمانی سالانه، فصلی و ماهانه خلأهای آماری ایجاد شده بازسازی گردید. نتایج ارزیابی با معیار RMSE نشان داد که روش نسبت نرمال در ٦۹/۲ درصد موارد و با معیار ضریب همبستگی در ۵۳/۳ درصد موارد به‌عنوان مناسب‌ترین روش می‌باشد. همچنین بررسی نتایج با آزمون t-student در سطح ۱ درصد نشان داد که برآورد در سال‌های فقط خشک فراتر از مقادیر حقیقی و برای سال‌های فقط تر کم‌تر از مقادیر مشاهده‌ای بود.

www.SID.ir

Archive of SID

مک‌کولوچ و بوس (۱۹۷۰) با استفاده از آمار ۳۰ ساله بارش ایستگاه پرشن حوضه آبخیز انتاریو واقع در کانادا تخمین بارش حوضه را با استفاده از روش رگرسیون مورد بررسی قرار دادند. نتایج این پژوهش نشان داد که داده‌های برآورد شده دارای ٤ درصد خطای سالانه تجمعی و ۱۷ درصد خطای ماهانه با داده‌های واقعی است.

چین (۱۹۹۵) روش آنالیز سری زمانی را در مورد بازسازی آمار بارش و دبی با درصد خطای ناچیز، قابل‌قبول مناسب دانست.

آب و همکاران (۲۰۰۰) برای بازسازی بارندگی‌های مفقوده روش استفاده از منطق فازی را با شبکه‌های عصبی مصنوعی و نسبت نرمال در سه ایستگاه باران‌سنجی واقع در شمال ایتالیا مورد مقایسه قرار دادند. نتایج نشان داد که منطق فازی نسبت به دو روش دیگر، داده‌های گم شده را با خطای کم‌تری برآورد می‌کند.

یو و همکاران (۲۰۰٤) از دو روش عکس مجذور فاصله (گرافیکی) و رگرسیون برای تخمین دما براساس ایستگاه‌های مجاور استفاده کردند و این دو روش را ارزیابی کرده و مقادیر برآوردی را با مقادیر اندازه‌گیری شده واقعی مقایسه نمودند. آن‌ها ضریب کارایی، واریانس، ریشه میانگین مربع خطا و خطاهای سیستماتیک و غیرسیستماتیک را محاسبه کرده و به این نتیجه رسیدند که روش رگرسیون در نواحی کوهستانی مناسب‌تر از روش گرافیکی است. همچنین در مواردی‌که تعداد ایستگاه‌های هواشناسی پراکنش کمی داشته باشند هر دو روش ناتوان می‌باشند.

همان‌گونه که از سوابق پژوهش نمایان است به‌طورکلی مطالعات اندکی در سطح جهان در مورد بازسازی و تکمیل داده‌های هیدرولوژیکی صورت گرفته و بیش‌تر مطالعات موجود در داخل و خارج به موضوع بازسازی و تکمیل داده‌های ناقص هواشناسی از جمله بارندگی، دما و... اختصاص دارد. بنابراین برای مطالعه و بررسی در زمینه پژوهش تلاش گردید از روش‌ها و سوابق پژوهش که در بازسازی پارامترهای اقلیمی استفاده شده و از تجارب کار در بازسازی داده‌های بارش که در سطح وسیع‌تری به‌کار رفته است در زمینه دبی جریان استفاده گردد.

تحلیل رگرسیون یکی از مهم‌ترین ابزارهای مورد استفاده توسط هیدرولوژیست‌ها می‌باشد. اساس تمام آن‌ها ساختن مدل براساس داده‌های مشاهداتی است. در برخی از موارد به‌علت پیچیدگی زیاد ساختار سیستم‌های مورد بررسی و قطعیت نداشتن متغیرهای مورد استفاده، تئوری مجموعه‌های فازی

www.SID.ir

Archive of SID

در تحلیل رگرسیون ابزار مناسبی جهت مدل‌سازی با استفاده از توابع عضویت فازی می‌باشد. روش رگرسیون آماری براساس قوانین احتمالات و رگرسیون فازی هم براساس تئوری مجموعه‌های فازی است. در تحلیل رگرسیون آماری خطای محاسباتی بین داده‌های مشاهداتی و مدل رگرسیون متغیری تصادفی با توزیع نرمال است ولی در رگرسیون فازی خطا به‌صورت میزان فازی بودن ساختار مدل می‌باشد (چنگ و ایوب، ۲۰۰۱). برای اولین بار لطفی‌زاده در سال ۱۹٦۵ مفاهیم مجموعه‌های فازی را بیان کرد و کاربردهایی از داده‌های فازی را در مدل‌های رگرسیون پیشنهاد داد (زاده، ۱۹٦۵) در اوایل رگرسیون فازی با در نظر گرفتن حداقل میزان شاخص فازیت (میزان قطعیت نداشتن) مورد توجه قرار گرفت (تاناکا و همکاران، ۱۹۸۲). همچنین مقایسه‌های بین مدل‌های مختلف رگرسیون فازی انجام و به تفاوت رگرسیون فازی و معمولی نیز اشاره شده است (رادن و ودال، ۱۹۹٤). مقایسه‌های بین رگرسیون فازی و رگرسیون معمولی توسط چنگ و ایوب (۲۰۰۱) انجام گرفت. این مطالعات را سانچز و گومز (۲۰۰۳) گسترش دادند. در سال‌های اخیر مباحث فازی وارد علوم هیدرولوژی گردیده و در زمینه‌های مختلف به‌کار گرفته شده است در این زمینه می‌توان به مطالعات محمدی و طاهری (۲۰۰٦) در زمینه برازش توابع انتقالی خاک با استفاده از رگرسیون خطی و پرویز و همکاران (۲۰۱۰) در مورد پیش‌بینی جریان رودخانه و شایان‌نژاد و همکاران (۲۰۰۸) در مورد برآورد تبخیر و تعرق پتانسیل با استفاده از رگرسیون فازی اشاره نمود.

در این مطالعه سعی شده از روش رگرسیون فازی که از روش‌های جدیدی است که در علم هیدرولوژی مورد استفاده قرار گرفته در زمینه بازسازی استفاده گردد و دقت آن در بازسازی داده‌های دبی جریان با سایر روش‌های متداول سنجیده شود. بنابراین آن‌چه در این پژوهش مورد بررسی قرار داده شده است ارزیابی توانایی روش رگرسیون فازی در بازسازی داده‌های سالانه دبی ایستگاه‌های هیدرومتری حوضه آبریز کارون بزرگ در مقایسه با روش‌های رایج بازسازی خلأهای آماری دبی جریان مانند روش رگرسیون خطی، چندمتغیره، نسبت نرمال و محور مختصاتی می‌باشد.

مواد و روش‌ها

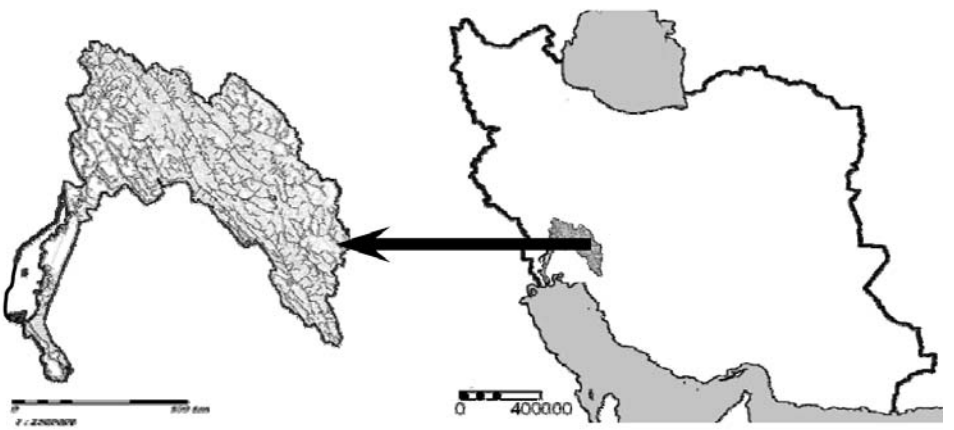
معرفی حوضه آبریز مورد مطالعه: از آنجا که هدف از این پژوهش مقایسه روش‌های مختلف جهت تکمیل داده‌های هیدرومتری است در انتخاب محدوده مورد مطالعه سعی شده تنوع شرایط اقلیمی،

www.SID.ir

Archive of SID

فیزیوگرافی و سنوات آماری موردنظر قرارگیرد. با توجه به این موضوع حوضه آبریز کارون بزرگ که مشتمل بر حوضه‌های کارون و دز است، جهت انجام این مطالعه انتخاب گردید. حوضه‌های آبریز رودخانه‌های دز و کارون در داخل ارتفاعات زاگرس میانی قرار دارند و به مختصات جغرافیایی ٤٨ درجه و ۱۰ دقیقه تا ۵۲ درجه و ۳۰ دقیقه طول شرقی و ۳۰ درجه و ۲۰ دقیقه تا ۳٤ درجه و ۵ دقیقه عرض شمالی محدود می‌باشند (شکل ۱) این دو حوضه در تقسیم‌بندی‌های طرح جامع آب کشور به‌عنوان حوضه‌های سوم و چهارم از منطقه سوم مطالعاتی می‌باشند.

در حوضه رودخانه کارون بزرگ ۸۳ ایستگاه هیدرومتری در شاخه‌های اصلی و فرعی رودخانه‌ها تأسیس شده است. در این ایستگاه‌ها آب‌دهی رودخانه‌ها، غلظت رسوبات و پارامترهای کیفیت آب اندازه‌گیری می‌شود. طولانی‌ترین آمار در این حوضه مربوط به رودخانه کارون در شهر اهواز می‌باشد که از سال ۱۲۷۳، آمار تغییرات سطح آب و از سال ۱۳۲۹ آب‌دهی آن اندازه‌گیری و ثبت شده است. از این ۸۳ ایستگاه تعداد ۲۰ ایستگاه تعطیل شده و ۳ ایستگاه جدید در سال‌های ۱۳۷۰ و ۱۳۷۱ تأسیس گردیده‌اند. حوضه‌های آبریز رودخانه‌های کارون و دز به لحاظ مشخصه‌های پستی و بلندی و هیدرولوژی از ۸ زیرحوضه تشکیل شده‌اند که مشخصات آن‌ها در جدول ۱ آمده است.



شکل ۱- موقعیت حوضه آبریز مورد مطالعه.

www.SID.ir

Archive of SID

جدول ۱- مشخصات زیرحوضه‌های آبریز کارون و دز (کارون بزرگ).

کد زیرحوضه	رودخانه‌های اصلی	مساحت (کیلومترمربع)	ارتفاع حداکثر (متر)	ارتفاع حداقل (متر)
۳-۳-۱	تیره- ماربره	٦٢٧٤	٤٠٥٠	١٥٠٠
۳-۳-۲	سبزه- زاز- سزار	٤٧٢٩	٤٠٨٢	٥٠٠
۳-۳-۳	بختیاری- زالکی	٦٤٤٨	٤٠٨٢	١٥٠٠
۳-۳-٤	دزپایین و شاوور	٥٧٩٩	٢٥٨٥	١٥
۳-٤-۱	بشار- ماربره- خرسان	٨٩٧٦	٤٤٠٩	١٧٩٠
۳-٤-۲	آب ونک- جونقان- کیار- بازفت	١٥٠٣٧	٤٢٢١	١٠٠٠
۳-٤-۳	کارون	٨٥٣١	٣٧٠١	٢٠٠
۳-٤-٤	بهلول- شور	١٢٦٨٧	١٣٩١	٠

روش کار: تکمیل داده‌های گمشده براساس داده‌های ثبت شده در ایستگاه‌های دیگر، مستلزم یافتن شباهتی بین ایستگاه‌های مورد بررسی است. این شباهت می‌تواند شباهت هیدرولوژیکی، فیزیوگرافی و یا کلیماتولوژی باشد. برای بازسازی داده‌های دبی حداقل ٤ ایستگاه مجاور هم که دارای طول دوره آماری مشترک و بدون خلأ آماری باشند مورد نیاز است (خلیلی ۱۹۹۸). بنابراین لازم است که حوضه به چند زیرحوضه تقسیم شود و ایستگاه‌هایی که در هر زیرحوضه قرار می‌گیرند و دارای طول دوره آماری مناسب و بدون نقص هستند انتخاب گردند. برای این کار از تقسیم‌بندی زیر حوضه‌ای طرح جامع آب کشور در حوضه کارون بزرگ استفاده شده است (شرکت مهندسان مشاور جاماب، ۱۹۹۹) با در نظر گرفتن ۱۳ سال به‌عنوان طول دوره آماری مشترک ۲۷ ایستگاه در قالب ۱۱ گروه برای بازسازی آماده شد که مشخصات آن‌ها در جدول ۲ آمده است.

در جهت بررسی کارایی روش رگرسیون خطی فازی (FLR)[1] در بازسازی داده‌های دبی سالیانه ایستگاه‌های هیدرومتری از روش حذف اعتباری[2] استفاده گردید. در این روش هر یک از داده‌های دبی سالیانه مشاهده‌ای ایستگاه‌های هیدرومتری حوضه آبریز کارون بزرگ به‌صورت عمدی حذف شده و سپس با استفاده از روش‌های رگرسیون فازی، نسبت نرمال (NR)[3]، روش محور مختصات یا گرافیکی (GR)[4]، روش رگرسیون خطی ساده (SLR)[5]، روش رگرسیون خطی چندمتغیره (MLR)[6]، بازسازی گردیده و سپس با آمار مشاهده‌ای مقایسه شده و مورد ارزیابی قرار می‌گیرد.

1- Fuzzy Linear Regression (FLR)
2- Cross Validation
3- Normal Ratio (NR)
4- Graphic (GR)
5- Simple Linear Regression (SLR)
6- Multi Linear Regression (MLR)

www.SID.ir

Archive of SID

جدول ۲- گروه‌های بازسازی بر مبنای ناحیه‌بندی زیرحوضه‌ای کارون بزرگ.

گروه بازسازی	رودخانه	ایستگاه هیدرومتری	کد زیرحوضه	طول دوره آماری مشترک
گروه ۱	گله‌رود تیره ماربره کمندان ماربره آب سبزه	ونایی درود دره تخت کمندان درود چم‌چیت	۱-۳-۳ ۲-۳-۳	۱۳٤۸-٤۹ تا ۱۳٦۲-٦۳ (۱۵ سال)
گروه ۲	بشار کبکیان ماربره ماربره کرم رود خرسان	شاه مختار بطاری دهکده شهید کتا پتاوه بارز	۱-٤-۳	۱۳۵۸-۵۹ تا ۱۳۷۸-۷۹ (۲۱ سال)
گروه ۳	سراب سفید گله رود کمندان ماربره تیره	ونایی ونایی کمندان دره تخت درود	۱-۳-۳	۱۳۵۲-۵۳ تا ۱۳٦٦-٦۷ (۱۵ سال)
گروه ٤	سرخاب زاز آب سبزه ماربره تیره	کشور سپیددشت چم‌چیت درود درود	۲-۳-۳ ۱-۳-۳	۱۳٤۲-٤۳ تا ۱۳٦۲-٦۳ (۲۱ سال)
گروه ۵	شاوور دز دز دز دز	پل شاوور دزفول حرمله بامدژ تله زنگ	٤-۳-۳ ۲-۳-۳	۱۳۵۲-۵۳ تا ۱۳٦۸-٦۹ (۱۷سال)

www.SID.ir

Archive of SID

ادامه جدول ۲-

گروه بازسازی	رودخانه	ایستگاه هیدرومتری	کد زیرحوضه	طول دوره آماری مشترک
گروه ٦	کارون	ملاثانی	٤-٤-٣	١٣٥٢-٥٣ تا ١٣٦٨-٦٩ (١٧ سال)
	کارون	سدگتوند		
	کارون	اهواز		
	دز	حرمله	٤-٣-٣	
	دز	بامدژ		
گروه ٧	سراب سفید	ونایی	١-٣-٣	١٣٥٢-٥٣ تا ١٣٦٦-٦٧ (١٥ سال)
	گله‌رود	ونایی		
	کمندان	کمندان		
	تیره	درود		
گروه ٨	سراب سفید	ونایی	١-٣-٣	١٣٥٤-٥٥ تا ١٣٦٦-٦٧ (١٣ سال)
	گله‌رود	ونایی		
	سیلا خور	رحیم آباد		
	تیره	درود		
گروه ٩	دز	تله‌زنگ	٢-٣-٣	١٣٤٢-٤٣ تا ١٣٦٢-٦٣ (٢١ سال)
	زاز	سپیددشت		
	سرخاب	کشور		
	آب سبزه	چم چیت		
گروه ١٠	بشار	شاه مختار	١-٤-٣	١٣٥٣-٥٤ تا ١٣٦٤-٦٥ (١٢ سال)
	بشار	دارشاهی		
	کبکیان	بطاری		
	بشار	بطاری		
گروه ١١	کارون	ملاثانی	٤-٤-٣	١٣٤٨-٤٩ تا ١٣٦٦-٦٧ (١٩ سال)
	کارون	سدگتوند		
	کارون	اهواز		
	شور	بتوند	٤-٣-٣	
	دز	بامدژ		

www.SID.ir

Archive of SID

قبل از انجام بازسازی همگنی داده‌های ایستگاه‌های هیدرومتری و آزمون گردش[1]، نرمال بودن آن‌ها نیز توسط آزمون کلموگراف و اسمیرنوف[2] مورد ارزیابی قرار گرفت (مهدوی، ۱۹۹۳).

برای تعیین متغیر مستقل در روش رگرسیون خطی ساده برای داده‌های دبی سالانه در هر یک از گروه‌ها همبستگی بین هر یک از ایستگاه‌ها با سایر ایستگاه‌های گروه برقرار گردید و با توجه به ضرایب همبستگی، ایستگاهی که بالاترین ضریب همبستگی را با ایستگاه موردنظر داشته به‌عنوان متغیر مستقل انتخاب شده است.

برای مقایسه روش‌های بازسازی داده‌های دبی سالانه از آماره ریشه مجذور مربعات خطا[3] استفاده شده است. علاوه‌بر آماره یاد شده تطابق آمار بازسازی شده در سه سطح حداکثر اختلاف ۲۰±، ۱۰± و ۵± درصد نسبت به آمار مشاهده شده مورد بررسی قرار گرفت و در هر ایستگاه درصد آماری که در هر سطح اختلاف قرار دارد مشخص شد که از آن به‌عنوان معیاری جهت سنجش بهترین روش بازسازی استفاده گردید.

در این پژوهش از روش‌های رگرسیون ساده، چندمتغیره، نسبت نرمال، گرافیکی و رگرسیون فازی جهت بازسازی داده‌ها استفاده شده است.

با توجه به آشنایی عمومی که با روش رگرسیون خطی ساده و چندمتغیره وجود دارد از توضیح آن‌ها خودداری می‌شود و در مورد روش نسبت نرمال دبی سالانه ایستگاه دارای داده ناقص با استفاده از ایستگاه‌های اطراف براساس رابطه زیر بازسازی می‌شود:

$$Q_X = \frac{1}{n}\left[\left(\frac{Q_X^-}{Q_A}.Q_A\right)+\left(\frac{Q_X^-}{Q_B}.Q_B\right)+...\right] \qquad (۱)$$

که در آن Q_X: دبی سالی که آمار آن مفقود است Q_A و Q_B: دبی ایستگاه‌های هم‌جوار در همان سال می‌باشد و Q_X^-، Q_A^- و Q_B^-: میانگین سالانه دبی در هر سه ایستگاه X، A، B و n تعداد ایستگاه‌های هم‌جوار مورد استفاده است.

در مورد روش گرافیکی با توجه به مختصات جغرافیایی ایستگاه دارای آمار مفقود و سایر ایستگاه‌های اطراف آن و براساس رابطه زیر بازسازی صورت می‌پذیرد:

$$Q_X = \frac{W_A Q_A + W_B Q_B + ...}{W_A + W_B + ...} \qquad (۲)$$

1- Run Test
2- Kolmogorov-Smironov Test
3- Root Mean Square Error (RMSE)

www.SID.ir

Archive of SID

که در آن $W = \frac{۱}{X^۲ + Y^۲}$ بوده و X و Y: مختصات جغرافیایی ایستگاه‌ها می‌باشد. Q_X: دبی سال موردنظر گمشده Q_A و Q_B: دبی همان سال در ایستگاه‌های هم‌جوار می‌باشد. با توجه به اهمیت روش رگرسیون خطی فازی این روش به‌صورت مفصل‌تر در زیر تشریح می‌گردد.

روش رگرسیون خطی فازی: در رگرسیون خطی کلاسیک به‌ازای هر دسته از متغیرهای ورودی، یک مقدار مشخص برای متغیر خروجی محاسبه می‌گردد. در حالی‌که در رگرسیون فازی بازه‌ای از مقادیر ممکن برای متغیر خروجی تخمین زده می‌شود. توزیع این مقادیر به‌صورت تابع عضویت مشخص می‌شود.

به‌طورکلی برای برازش یک معادله رگرسیون خطی فازی سه دسته مدل وجود دارد:

۱. مدل‌های رگرسیون امکانی فازی[1]

۲. مدل‌های رگرسیون کم‌ترین مربعات[2]

۳. مدل‌های رگرسیون مبتنی بر تحلیل بازه‌ای[3]

مدل‌های رگرسیون امکانی فازی اولین بار توسط تاناکا و همکاران (۱۹۸۲) ارایه گردید. این مدل‌ها بهترین معادله رگرسیون را با حداقل کردن میزان فازی بودن به‌دست می‌دهند. این کار با حداقل کردن میزان فازی بودن به‌دست می‌دهند. این کار با حداقل کردن مجموع کل پهنای توابع عضویت ضرایب فازی معادله رگرسیون اجرا می‌شود. یکی از مدل‌های رگرسیون فازی امکانی، مدلی است که در آن ضرایب، فازی می‌باشد و ورودی و خروجی مشاهده‌ای غیرفازی است. در این پژوهش از این مدل استفاده شده است. این مدل با معادله ۳ بیان می‌شود.

$$\tilde{y} = \tilde{A}_۰ + \tilde{A}_۱ x_۱ + ... + \tilde{A}_n x_n \qquad (۳)$$

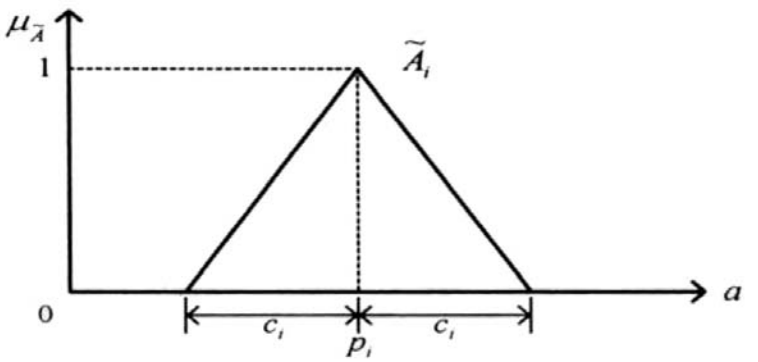
ضرایب معادله بالا یعنی $\tilde{A}_۰$، $\tilde{A}_۱$، ... و $\tilde{A}_n$ اعداد فازی و متغیرهای ورودی مشاهده‌ای؛ یعنی، $x_۱$ و ... و x_n اعداد معمولی هستند برای هر n متغیر یک عدد فازی مانند $\tilde{y}$ به‌عنوان خروجی محاسبه‌ای به‌دست می‌آید.

1- Fuzzy Possibilistic Regression
2- Fuzzy Least Squares Regression
3- Interval Regression

www.SID.ir

Archive of SID

فرض کنید مقدار m سطر داده مشاهده‌ای وجود دارد و در هر سطر n متغیر ورودی (x_{ij}) و یک متغیر خروجی می‌باشد. همچنین فرض کنید عدد فازی به‌صورت مثلثی متقارن به شکل ۱ باشد (در این شکل $C_i =$ پهنای عدد و $P_i =$ مرکز عدد فازی). در این صورت می‌توان تابع عضویت را به‌صورت رابطه ٤ نوشت.



شکل ۲- تابع عضویت ضرایب فازی.

$$\mu_{\tilde{A}}(a_i) = \begin{cases} ۱ - \dfrac{|p_i - a_i|}{c_i} & p_i - c_i \le a_i \le p_i + c \\ ۰ \end{cases} \qquad (٤)$$

$$\tilde{y} = (p_۰, c_۰) + (p_۱, c_۱)x_۱ + \ldots + (p_n, c_n)x_n \qquad (٥)$$

عدد فازی $\tilde{A}$ در شکل ۲ برای نشان دادن مقدار «تقریباً برابر p_i» می‌باشد و c_i نشانگر میزان فازی بودن آن است که این مفهوم را می‌توان به شکل $\tilde{A}_i = (p_i, c_i)$ نمایش داد بنابراین رابطه رگرسیون فازی به‌صورت معادله ٥ می‌باشد. تابع عضویت متغیر فازی خروجی؛ یعنی $\tilde{y}$ به‌صورت رابطه ٦ ارایه می‌شود.

$$\mu_{\tilde{y}}(y) = \begin{cases} \max(\min[\mu_{\tilde{A}i}(a_i)]) & \{a | y = f(x,a) \neq \phi \\ ۰ \end{cases} \qquad (٦)$$

www.SID.ir

Archive of SID

با جای‌گذاری رابطه ۳ در رابطه ۵، رابطه ٦ به‌دست می‌آید.

$$\mu_{\tilde{y}}(y)=\begin{cases}1-\dfrac{y-p_0-\sum_{1}^{n}p_i x_i}{} & x_i\neq 0\\ 1 & x_i=0, y=0\\ 0 & x_i=0, y\neq 0\end{cases} \qquad (۷)$$

برای حل مسأله رگرسیون خطی فازی، الگوریتم‌های مختلفی پیشنهاد شده است که یکی از آن‌ها تبدیل مسأله رگرسیون خطی فازی به یک مسأله برنامه‌ریزی خطی است. در این حالت که داده‌ها، غیرفازی هستند، هدف مدل رگرسیون، تعیین مقادیر بهینه $\tilde{A}$ می‌باشد به گونه‌ای که درجه عضویت متغیر خروجی فازی برای تمام داده‌ها از یک مقدار معنی مانند h که توسط کاربر تعیین می‌شود، بزرگ‌تر باشد. به بیان دیگر برای m سطر داده، $(i=1,2,3,...,m)$ باید نامساوی زیر صادق باشد:

$$\mu_{\tilde{y}j}(y_i)\geq h \qquad (۸)$$

با افزایش مقدار h میزان فازی بودن خروجی‌ها نیز افزایش می‌یابد. رابطه ۸ بیان می‌کند که خروجی فازی باید بین دو مقدار A و B که در شکل ۳ مشخص شده‌اند، قرار بگیرد. با توجه به رابطه ۷ مرکز و پهنای تابع عضویت خروجی به‌ترتیب برابر با $p_0+\sum_{1}^{n}p_i x_i$ و $c_0+\Sigma c_i x_i$ در نظر گرفته می‌شود.

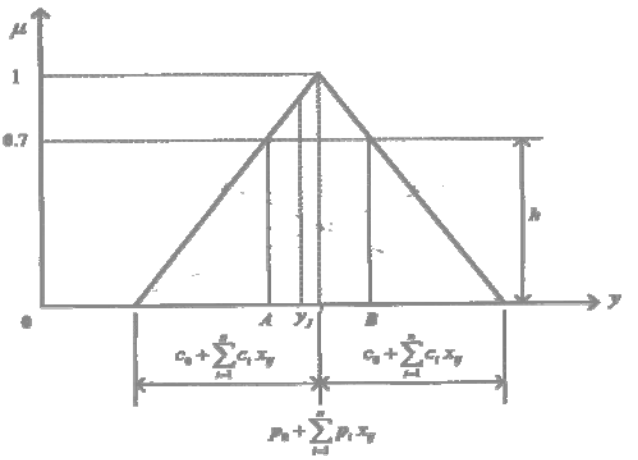
در روش رگرسیون فازی امکانی ضرایب به گونه‌ای تعیین می‌شود که پهنای خروجی فازی برای تمام مجموعه‌های داده‌ها حداقل گردد. قیدهای مسأله با جای‌گذاری رابطه ۷ در نامعادله ۸ به‌دست آمده است.

بنابراین برای حل یک مسأله رگرسیون خطی با ضرایب فازی و داده‌های غیرفازی، کافی است که یک مدل برنامه‌ریزی خطی حل گردد. تعداد متغیرهای ورودی عبارت است از داده‌های دبی در همان سال موردنظر در سایر ایستگاه‌های هیدرومتری اطراف که دارای آمار کامل می‌باشند. برای این کار از نرم‌افزار هیدروژنراتور[1] (ساداتی‌نژاد و همکاران، ۲۰۰٦) استفاده شده است.

1- Hydrogenerator

www.SID.ir

Archive of SID



شکل ۳- تابع عضویت خروجی فازی.

بحث و نتیجه‌گیری

داده‌های ایستگاه‌های هیدرومتری حوضه کارون به‌وسیله روش آزمون گردش (مهدوی، ۱۹۹۳) ارزیابی شده و همگنی آن مورد تأیید قرار گرفت و همچنین با توجه به غیرنرمال بودن داده‌ها براساس آزمون کلموگراف و اسمیرنوف (مهدوی، ۱۹۹۳) با استفاده از تبدیل LNX، داده‌ها نرمال شده و مورد استفاده قرار گرفت.

پس از بازسازی داده‌های هر یک از ایستگاه‌های هیدرومتری با روش‌های پنج‌گانه ذکر شده در بخش قبل، با استفاده از معیار ریشه مجذور مربعات خطا داده‌ای مشاهده‌ای با داده‌های بازسازی شده مقایسه گردیده است که جدول ٤ مقادیر آماره ریشه مجذور مربعات خطا، در هر یک از ایستگاه‌های هیدرومتری گروه ۱ را به‌عنوان نمونه نشان می‌دهد.

علاوه‌بر معیار ریشه مجذور مربعات خطا از معیار درصد فراوانی داده‌های بازسازی شده در سه سطح اختلاف ٥±، ۱۰± و ۲۰± درصد داده‌های مشاهده‌ای نیز برای تعیین بهترین روش بازسازی در هر ایستگاه استفاده شده است. در این روش داده بازسازی شده هر سال هر ایستگاه را با داده مشاهده‌ای همان سال ایستگاه مقایسه کرده و براساس میزان اختلاف برآورد مشخص می‌گردد که در کدام یک از ۳ دامنه مشخص شده بالا نسبت به داده مشاهده‌ای قرار می‌گیرد. بدیهی است هرچه میزان درصد فراوانی داده‌های بازسازی در سطح اختلاف کم‌تری از داده‌های مشاهده‌ای قرار گیرد موفقیت روش بیش‌تر خواهد بود. جدول ٥ مقادیر درصد فراوانی داده‌های بازسازی شده را در سطوح مختلف در ایستگاه‌های گروه ۱ نشان می‌دهد.

www.SID.ir

Archive of SID

جدول ٤- مقادیر ریشه مجذور مربعات خطا در روش‌های مختلف بازسازی آمار دبی سالانه ایستگاه‌های هیدرومتری گروه یک.

گروه بازسازی	رودخانه- ایستگاه	روش بازسازی				
		MLR	SLR	GR	NR	FLR
گروه ۱	گله‌رود- ونایی	۰/۹۵	۰/٦٤	۸/۷۰	۱/۰۷	۰/٦۲
	تیره- درود	۳/۱٤	٤/۰۰	۷/۰٤	۲/۳٤	۳/٤۹
	ماربره- دره تخت	۱/۷٦	۱/۳۳	٥/۰٤	۲/۷۰	۱/۳٦
	کمندان- کمندان	۰/۲٤	۰/۱۷	۷/۹۳	۰/۳۹	۰/۲۷
	ماربره- درود	۲/۸۱	۱/۹۹	٦/۸۳	٤/۳٦	۲/۳۸
	آب سبزه- چم‌جیت	۱/۳۲	۱/۰۸	۷/۲۰	۱/٤۰	۱/٤٥

FLR: روش رگرسیون خطی فازی، NR: روش نسبت نرمال، GR: روش گرافیکی، SLR: روش رگرسیون خطی ساده، MLR: روش رگرسیون چندمتغیره.

جدول ٥- درصد فراوانی داده‌های بازسازی شده دبی سالانه ایستگاه‌های گروه ۱ در سطوح اختلاف داده‌های مشاهده‌ای.

روش بازسازی	MLR			SLR			GR			NR			FLR		
ایستگاه	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
۱	۰/۰۰	٦/٦۷	۲۰/۰۰	٦/٦۷	٦/٦۷	٥۳/۳۳	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۲٦/٦۷	۰/۰۰	٦/٦۷	۳۳/۳۳
۲	۲۰/۰۰	٤۰/۰۰	۸۰/۰۰	۱۳/۳۳	٤٦/٦۷	۷۳/۳۳	٦/٦۷	٦/٦۷	۱۳/۳۳	۲٦/٦۷	٤٦/٦۷	۷۳/۳۳	٦/٦۷	٤۰/۰۰	٦٦/٦۷
۳	۲٦/٦۷	۳۳/۳۳	۷۳/۳۳	٥۳/۳۳	۷۳/۳۳	۹۳/۳۳	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	٦۰/۰۰	۸٦/٦۷	۲۰/۰۰	٤۰/۰۰	٦٦/٦۷
٤	۲۰/۰۰	۱۰/۰۰	۸۰/۰۰	۲٦/٦۷	٦٦/٦۷	۸٦/٦۷	۰/۰۰	۰/۰۰	۱۳/۳۳	۲٦/٦۷	٦۰/۰۰	٦٦/٦۷	٤۰/۰۰	٥۳/۳۳	٦٦/٦۷
٥	۲۰/۰۰	٤٦/٦۷	٦۰/۰۰	۱۳/۳۳	٦٦/٦۷	۸۰/۰۰	۰/۰۰	٦/٦۷	۱۳/۳۳	۲۰/۰۰	٦/٦۷	٦۰/۰۰	٦/٦۷	۲۰/۰۰	٥۳/۳۳
٦	٦/٦۷	۲٦/٦۷	٦٦/٦۷	۱۳/۳۳	٥۳/۳۳	۸۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۲۰/۰۰	٤٦/٦۷	۷۳/۳۳	۲٦/٦۷	٤٦/٦۷	۷۳/۳۳

۱. گله‌رود- ونایی، ۲. تیره- درود، ۳. ماربره- دره تخت، ٤. کمندان- کمندان، ٥. ماربره- درود، ٦. آب سبزه- چم چیت.
a: سطح اختلاف ٥±، b: سطح اختلاف ۱۰± و c: سطح اختلاف ۲۰±. FLR: روش رگرسیون خطی فازی، NR: روش نسبت نرمال، GR: روش گرافیکی، SLR: روش رگرسیون خطی ساده، MLR: روش رگرسیون چندمتغیره.

هر گاه مقادیر آماره ریشه مجذور مربعات خطا در دو روش بازسازی اعمال شده در یک ایستگاه برابر شود با کمک درصد فراوانی این سطوح اختلاف روش برتر تشخیص داده شده است.

با استفاده از نتایج به‌دست آمده از محاسبه معیار ریشه مجذور مربعات خطا و نیز درصد فراوانی داده‌های بازسازی شده در سطوح اختلاف داده‌های مشاهده‌ای می‌توان اولویت هر یک از روش‌های

www.SID.ir

Archive of SID

بازسازی را در ایستگاه‌های هیدرومتری گروه‌های مختلف شناسایی نمود. به این منظور هر روشی که مقدار آماره ریشه مجذور مربعات خطای کم‌تری دارد روش مطلوب‌تری محسوب می‌شود. جدول ٦ اولویت بازسازی داده‌های دبی سالانه را در ایستگاه‌های گروه ۱ به‌عنوان نمونه نشان می‌دهد.

جدول ٦- اولویت روش‌های بازسازی دبی سالانه در ایستگاه‌های هیدرومتری گروه ۱.

گروه بازسازی	رودخانه- ایستگاه	اولویت				
		۱	۲	۳	٤	٥
گروه ۱	گله‌رود- ونایی	GR	NR	MLR	SLR	FR
	تیره- درود	GR	SLR	FR	MLR	NR
	ماربره- دره‌تخت	GR	NR	MLR	FR	SLR
	کمندان- کمندان	GR	NR	FR	MLR	SLR
	ماربره- درود	GR	NR	MLR	FR	SLR
	آب سبزه- چم چیت	GR	FR	NR	MLR	SLR

با مشخص شدن اولویت بازسازی هر یک از ایستگاه‌ها با استفاده از روش متوسط‌گیری آماره ریشه مجذور مربعات خطا در ایستگاه‌های هر حوضه روش برتر برای بازسازی داده‌های دبی سالانه هر زیر حوضه مشخص می‌گردد و در نهایت با میانگین‌گیری ریشه مجذور مربعات خطا در زیرحوضه‌ها، روش برتر جهت بازسازی داده‌های دبی در حوضه کارون بزرگ به‌دست می‌آید (جدول ۷).

همان‌گونه که ملاحظه می‌شود در بین روش‌های بازسازی دبی سالانه، رگرسیون خطی ساده در سه زیرحوضه ۲-۳-۳، ٤-۳-۳ و ٤-٤-۳، رگرسیون فازی در زیرحوضه ۱-٤-۳ و رگرسیون خطی چندمتغیره در زیرحوضه ۱-۳-۳ حداقل خطای ممکن را در بازسازی داده‌ها داشته‌اند. بنابراین بهترین روش بازسازی در این زیرحوضه‌ها شناخته شدند.

با توجه به میانگین ریشه مجذور مربعات خطا در زیرحوضه‌های کارون بزرگ که در جدول ۷ آمده است در حوضه ۱-۳-۳ و ۱-٤-۳ روش رگرسیون فازی و در زیرحوضه‌های ۲-۳-۳، ٤-۳-۳ و ٤-٤-۳ روش رگرسیون خطی ساده روش برتر انتخاب گردیده‌اند. براساس متوسط ریشه مجذور مربعات خطا زیرحوضه‌های پنج‌گانه کارون بزرگ روش رگرسیون ساده خطی به‌عنوان روش برتر برای بازسازی داده‌های سالانه دبی تشخیص داده می‌شود (جدول ۸).

www.SID.ir

Archive of SID

جدول ۷- مقادیر ریشه مجذور مربعات خطا روش‌های بازسازی دبی سالانه در زیرحوضه‌های کارون بزرگ.

کد زیرحوضه	روش بازسازی					روش منتخب
	NR	GR	SLR	MLR	FR	
۳-۳-۱	۱/۷۲	۵/۷	۱/۵٤	۱/٤٥	۱/٥۲	FLR
۳-۳-۲	۸/۱٥	٥۸	۷/۱٤	۱۱/۸۸	۱۰/٦۹	SIR
۳-۳-٤	۳۰/٥۹	۱۹۳/۰۸	۲۹/۹٥	۳۲/۹٤	۳۰/٦۳	SLR
۳-٤-۱	۱۳/۹۷	۲۸/۳۳	۱۲/۹۷	۱٤/۲	۱۲/۳۳	FLR
۳-٤-٤	۳٤	۳۱۳/۰٦	۲۸/٦٥	٤٦/٦۷	۳۱/۹۱	SLR
میانگین حوضه	۱۷/٦	۱۱۹/٦	۱٦/۰٥	۲۱/٤	۱۷/٤۱	SLR

جدول ۸- اولویت روش‌های بازسازی دبی سالانه در ایستگاه‌های هیدرومتری حوضه آبریز کارون بزرگ.

اولویت بازسازی	٥	٤	۳	۲	۱
روش بازسازی	GR	MLR	NR	FLR	SLR

SLR: روش رگرسیون ساده، FLR: روش رگرسیون خطی فازی، NR: روش نسبت نرمال، GR: روش گرافیکی، MLR: روش رگرسیون چندمتغیره.

با توجه به جدول‌های ۷ و ۸ روش رگرسیون ساده و فازی و نسبت نرمال با مقدار ریشه مجذور مربعات خطا ۱٦/۰٥، ۱۷/٤۱ و ۱۷/٦ به‌عنوان اولویت اول تا سوم در بازسازی داده‌های دبی سالانه ایستگاه‌های هیدرومتری حوضه کارون بزرگ معرفی می‌گردد. به‌طوری‌که مشاهده می‌شود روش گرافیکی با ریشه مجذور مربعات خطا ۱۱۹/٦ در بازسازی دبی‌های سالانه در پایین‌ترین جایگاه ممکن قرار دارد و دارای بیش‌ترین خطا در برآورد داده‌های دبی است. درصد فراوانی داده‌های بازسازی شده در سطوح مختلف اختلاف نیز اولویت‌بندی انجام شده با آماره ریشه مجذور مربعات خطا را تأیید می‌نماید.

نتایج

با توجه به مطالب ارایه شده و تحلیل‌های صورت گرفته در بخش قبلی و در راستای ارایه یک روش مطلوب برای بازسازی داده‌های دبی سالانه به مهندسان و شرکت‌های مشاور می‌توان پیشنهاد نمود در حوضه کارون بزرگ دو روش رگرسیون خطی ساده و رگرسیون خطی فازی برآورد مناسبی را در بازسازی داده‌های گمشده دبی سالانه دارند و با توجه به سادگی و عمومیت روش رگرسیون

www.SID.ir

Archive of SID

خطی ساده این روش برای کاربرد توصیه می‌گردد. براساس نتایج جدول‌های ۳ و ٤ بیش‌ترین خطا با مقدار ریشه مجذور مربعات خطا (۱۱۹/٦) در بازسازی داده‌ها را روش محور مختصات دارد که لازم است از به‌کارگیری آن احتراز گردد. همچنین پیشنهاد می‌شود این پژوهش در سایر حوضه‌های آبریز کشور نیز مدنظر قرار گرفته و برای داده‌های ماهانه و روزانه نیز تکرار گردد. بدیهی است نتایج حاضر برای کاربرد در حوضه کارون بزرگ توصیه می‌شود و کاربرد آن در سایر حوضه‌ها همراه با احتیاط و تطابق شرایط امکان‌پذیر است.

سپاسگزاری

بدین‌وسیله از حمایت‌های مالی معاونت پژوهش و مطالعات پایه شرکت مدیریت منابع آب ایران (دفتر امور پژوهشی و پشتیبانی علمی) سپاسگزاری می‌نمائیم.

منابع

1.Abeb, X.J., Solomatine, D.P. and Vennerker, R.G.W. 2000. Application of adaptiv fuzzy rule-based methods for reconstruction of missing precipitation events. Hydro. Sci. J. 45: 3. 425-436.
2.Alizadeh, M. and Sovlitabar, A. 1989. Reconstruction of discharge river, P 519-537. Proceeding of the First Conference of Hydrology, Tehran University, Tehran, Iran. (In Persian)
3.Chang Y. and Ayyub, B. 2001. Fuzzy regression methods-a comparative assessment. Fuzzy Sets and Systems, 119: 187-203.
4.Chin, D.A. 1995. A scale model of multivariate rainfull time series. J. Hydro. 168: 1. 4-15.
5.Jamab Engineering Consulting Company. 1999. National water Resources master plan: Karoo and Dez river basin, Iran. (In Persian)
6.Khalili, A. 1998. Water Comprehensive Plan of Iran. Minestry of Energy, 370p. (In Persian)
7.Lookzadeh, S. 2004. Evaluation of several different construction method for precipitation data in cenral Alborz Region. M.Sc. Thesis, Tehran University Iran, 183p. (In Persian)
8.Macculloch, J.A.W. and Booth, M. 1970. Estimation of basin precipitation by regression equation .Water Resources Research, 16: 6. 1753-1758.
9.Mahdavi, M. 1993. Applaid hydrology. Tehran University Press, 330p. (In Persian)
10.Mohammadi, J. and Taheri, S.M. 2006. Pedo transfer function fitting with fuzzy regression., J. Sci. and Tech. Agric. and Natur. Resour. 9: 2. 51-60. (In Persian)

www.SID.ir

Archive of SID

11.Parviz, L., Kholghi, M. and Fakhorifard, A. 2010. Forecasting annual streamflow using autoregressive integrated moving average model and fuzzy regression, J. Soil and Water Sci. 19: 1. 66.82. (In Persian)
12.Radden D.T. and Woodall W.H. 1994. Preperties f certain fuzzy linear regression methods. Fuzzy Sets and Systems, 104: 361-375.
13.Sadatinejad, S.J. 1997. Statistical comparison of different methods for data generation of precipitation in Isfahan Province, M.Sc. Thesis, Watershed Manesement Dept. Tarbiat Modarres University, 145p. (In Persian)
14.Sadatinejad, S.J., Abdollahi, Kh. and Shayannejad, M. 2006. Zoning the region of influence of hedrometric station in Karoon river basin based on the suitable method of data reconstruction, Iranian Water Resources Manegement Organization, 198p. (In Persian)
15.Sanchez, J.D.A. and Gomez, A.T. 2003. Applications of fuzzy regression in actuarial analysis, J. Risk and Ins. 70: 4. 797-802.
16.Shayannejad, M., Sadatinejad, S.J. and Fahmi, H. 2008. Determination of potential evapotranspiration using the fuzzy regression method. J. Water Resour. Res. 3: 3. 76-86. (In Persian)
17.Tanaka, H., Uejima, S. and Asai, K. 1982. Linear regression analysis with fuzzy model, IEEE Transactions on Systems Man. Cybern. 12: 6. 903-907.
18.You, J., Hubbard, K.G. and Goddard, S. 2004. Comparison of air temperature estimates from spatial regression and inverse distance method. J. Atmos. Oceaanic Tech. Submitted.
19.Zadeh, L.A. 1965. Fuzzy sets. Information and Control, 8: 338-353.

www.SID.ir

Archive of SID

Gorgan University of Agricultural Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 17(4), 2011
www.gau.ac.ir/journals

Application of Fuzzy Linear Regression for predicting annual discharge missing data in hydrometric station compared with other conventional methods

***S.J. Sadati Nejad[1], R. Naghdi[2] and M. Shayan Nejad[3]**

[1]Assistant Prof., Dept. of Watershed and Management, University of Shahrekord,
[2]M.Sc. Graduated, Dept. of Water Engineering, University of Shahrekord,
[3]Assistant Prof., Dept. of Water Engineering, University of Shahrekord
Received: 2009/08/07; Accepted: 2010/10/06

Abstract

Discharge missing data, is an important challenge facing water resource project designers and hydrologists, as it leads to error in design studies and implementations with enormous financial and safety considerations. When the number of hydrometric stations is limited in an area of concern, this challenge becomes more serious. Therefore, this data needs to be completed statistically for practical purposes. Several methods have been proposed for predicting such missing data. In this article, fuzzy linear regression is compared with simple linear regression, multi-linear regression, graphical method and normal ratio method, for reconstruction of annual discharge data in Great Karoon river basin in south-west Iran. For this purpose, observational discharge data from 27 stations in a set of 11 groups were analyzed. By deliberate exclusion of real annual data for each station, the excluded data was reconstructed using the above methods and the error in calculation was measured using RMSE index which was found to be 17.41, 16.05, 21.00, 119.60 and 17.60, respectively. Hence, simple linear regression was found to be the superior method for reconstruction of annual discharge data for this basin. Fuzzy linear regression was also shown to be the second most appropriate method in this regard.

Keywords: Fuzzy linear regression, Missing data, Reconstruction data, Annual discharge, Karoon river basin

* Corresponding Author; Email: jsadatinejad@kashanu.ac.ir

www.SID.ir