نشریه تحقیقات کاربردی علوم جغرافیایی، سال شانزدهم، شماره ۴۲، پاییز ۹۵

ارزیابی عملکرد روش‌های مختلف در بازسازی داده های بارش ماهانه

دریافت مقاله: ۹۴/۱۰/۱۴ پذیرش نهایی: ۹۵/۵/۱۸

صفحات:۱۷۶-۱۵۵

علی رضازاده جودی: کارشناسی ارشد عمران-آب، باشگاه پژوهشگران جوان و نخبگان، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران[۱].

Email: alijoudi66@gmail.com

محمد تقی ستاری: استادیار، مهندسی منابع آب، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

Email: mtsattar@gmail.com

چکیده

در کلیه مطالعات اقلیمی نیاز به داده‌های صحیح و قابل اعتماد می‌باشد. روش‌های متعددی برای بازسازی داده ها وجود دارند که بسته به نوع داده‌ها و خصوصیات آب و هوائی هر منطقه تعدادی از آن‌ها مورد استفاده قرار می‌گیرد. در این تحقیق بخشی از داده‌های بارش ماهانه ایستگاه سینوپتیک شهر سراب در استان آذربایجان‌شرقی به صورت تصادفی داده های فرضی در نظر گرفته شد. سپس برای بررسی میزان کارایی، روش‌های متعدد بازسازی داده ها، هفت روش کلاسیک آماری و مدل درختی ام۵ بعنوان یکی از روش‌های کارآمد داده‌کاوی و به کمک داده‌های بارش ایستگاه‌های مجاور تخمین گردید. نتایج نشان داد از بین روش‌های کلاسیک آماری بررسی شده به ترتیب روش‌های انتساب چندگانه، نسبت نرمال و رگرسیون خطی چند متغیره دارای نتایج نسبتا دقیق‌تر با خطاهای کمتری می‌باشند. بررسی کلی نتایج نشان داد که مدل درختی ام۵ با توسعه قوانین اگر-آنگاه با ارائه چهار رابطه خطی ساده با آماره‌های (ضریب همبستگی برابر ۰/۹۷۴، ضریب نش-ساتکلیف برابر ۰/۹۴۸، ریشه میانگین مربعات خطا برابر ۵/۱۱ (میلی‌متر) و میانگین خطای مطلق برابر با ۴/۱۸۹ (میلی‌متر) دقیق‌ترین نتیجه را در بین تمامی روش‌های بررسی شده در این مطالعه ارائه می‌دهد. لذا با توجه به سادگی روند مدل‌سازی، کاربردی و قابل فهم بودن و داشتن دقت بالا استفاده از آن برای تخمین مقادیر گم‌شده بارش ماهانه پیشنهاد می‌گردد.

کلیدواژگان: بازسازی داده‌ها، بارش ماهانه، مدل درختی ام۵، روش نسبت نرمال، روش‌های کلاسیک آماری.

---

[۱]. **نویسنده مسئول:** باشگاه پژوهشگران جوان و نخبگان، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

مقدمه

بارش یکی از مهمترین اجزاء چرخه آب بوده و در سنجش خصوصیات اقلیمی هر منطقه نقش بسیار مهمی ایفا می‌کند. قدم اولیه در انجام کلیه مطالعات کلیماتولوژیکی و هیدرولوژیکی دسترسی به داده‌های آماری صحیح و قابل اعتماد می‌باشد. گاهی اوقات به‌دلایل مختلف از جمله عدم ثبت آمار، خرابی دستگاه‌های اندازه‌گیری و حذف آمار غلط، تخمین و برآورد این داده‌ها ضروری می‌باشد. از سویی دیگر، پیش از کسب اطمینان از صحت و کیفیت داده‌های تاریخی و سری‌های زمانی، نمی‌توان اقدام به استفاده از آن‌ها و استخراج نتایج بعدی نمود (قاجارنیا و همکاران، ۱۳۹۳ :۹۱). لذا با توجه به اهمیت دسترسی به داده‌های صحیح و پیوسته جهت انجام مطالعات هیدرولوژیکی دستیابی به یک روش صحیح برآورد بازسازی داده هاو صحت‌سنجی داده‌ها بسیار ضروری است. زیرا از یک طرف موجب کوتاه‌تر شدن مدت مطالعات و از طرف دیگر موجب برآورد دقیق‌تر پارامترهای هدف و کاهش هزینه‌های اجرایی و خسارات بعدی ناشی از اجرای طرح‌های عمرانی می‌گردد (لوک‌زاده، ۱۳۸۳). برای صحت‌سنجی داده‌های تاریخی، روش‌های متعددی از جمله آزمون‌های همگنی و بررسی داده‌های پرت وجود دارد. برای رفع خلأهای داده‌های یک ایستگاه اندازه‌گیری معمولاً از روش‌های آماری و با کمک گرفتن از داده‌های ایستگاه‌های مجاور با تشابه هیدرولوژیکی، کلیماتولوژی و فیزیوگرافی نسبتاً یکسان استفاده می‌گردد (نقدی و همکاران، ۱۳۸۹ :۶۰). با توجه به اهمیت موضوع اشاره شده، همواره محققین بسیاری نسبت به بازسازی و تخمین بازسازی داده هااقدام کرده‌اند. هرکدام از محققین از بین روش‌های مختلف به بررسی و پیشنهاد روشی خاص برای بازسازی بازسازی داده‌ها و صحت‌سنجی آن‌ها پرداخته‌اند. از جمله تحقیقات صورت گرفته در این زمینه می‌توان به تحقیقات دسیلوا و همکاران (۲۰۰۷) اشاره کرد که مقایسه‌ای بین روش‌های مختلف تخمین داده‌های بارش بازسازی شده انجام داد و برتری روش نسبت نرمال را گزارش کردند. همچنین یو و همکاران (۲۰۰۸) با استفاده از دو روش عکس فاصله‌های وزنی و رگرسیون به تخمین داده‌های دما پرداختند. آن‌ها به این نتیجه رسیدند که روش رگرسیون در نواحی ساحلی و کوهستانی بهتر از روش عکس فاصله‌های وزنی می‌باشد. همچنین کیم و پاچسپسکی (۲۰۱۰) به بازسازی داده‌های بارش روزانه با استفاده از روش‌های رگرسیون درختی و شبکه عصبی مصنوعی پرداخته و ادعا کردند که بهترین نتیجه زمانی حاصل می‌شود که از ترکیب این دو روش استفاده گردد. حسنپور کاشانی و دین پژوه (۲۰۱۲) با استفاده از روش‌های آماری کلاسیک شامل یازده روش مختلف به تخمین داده‌های هواشناسی اقدام کردند و نتایج به‌دست آمده را با سه نوع شبکه عصبی مصنوعی مختلف مقایسه و اعلام کردند که با وجود اینکه

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

تشکیل ساختار بهینه شبکه عصبی مصنوعی پیچیده و زمان‌بر است اما نتایج دقیق‌تری نسبت به روش‌های کلاسیک ارائه می‌کند. همچنین آن‌ها از بین روش‌های کلاسیک برتری روش رگرسیون چند متغیره را اعلام نمودند. چوگه و رگولوار (۲۰۱۳) با استفاده از روش شبکه عصبی مصنوعی به تخمین داده‌های بارش پرداخته و بر کاربرد شبکه عصبی مصنوعی در این موضوع تاکید کردند. قانی و همکاران (۲۰۱۴) با استفاده از الگوریتم بیان ژن به تخمین داده‌های بازسازی شده بارش پرداختند و توانایی بالای این مدل را اعلام کردند. از جمله تحقیقاتی که در ایران در این زمینه صورت گرفته است می‌توان به تحقیقات لوک‌زاده (۱۳۸۳) در زمینه مقایسه روش‌های مختلف بازسازی خلأهای آماری بارندگی در مقاطع مختلف زمانی در منطقه البرز مرکزی با روش‌های رگرسیون خطی، نسبت نرمال، محور مختصات و روش زمین آمار اشاره کرد. نتایج مطالعه ایشان نشان‌دهنده برتری روش نسبت نرمال نسبت به سایر روش‌ها بود. انصاری و داوری (۱۳۸۹) با استفاده از تکنیک فازی به تخمین و بازسازی داده‌های بارندگی استان خراسان پرداختند و توانایی بالای این روش را در مقایسه با روش عکس فاصله و میانگین‌گیری گزارش کردند. همچنین ساداتی‌نژاد و همکاران (۱۳۸۹) کارایی روش رگرسیون خطی فازی در برآورد داده‌های ناقص دبی سالانه ایستگاه‌های هیدرومتری را در مقایسه با روش‌های رگرسیون خطی ساده، رگرسیون چند متغیره، محور مختصات و نسبت نرمال بررسی کردند. نتایج نشان‌دهنده برتری روش رگرسیون خطی ساده و بعد از آن رگرسیون فازی بود. نقدی و همکاران (۱۳۸۹) به مقایسه روش‌های مختلف تخمین داده‌های گمشده دبی ماهانه حوضه آبخیز کارون بزرگ پرداخته و برتری روش شبکه عصبی مصنوعی را در مقایسه با سایر روش‌های بررسی شده اعلام کردند. خورشیددوست و همکاران (۱۳۹۱) به بازسازی سری‌های زمانی دمای حداکثر و حداقل روزانه با استفاده از روش نزدیک‌ترین همسایه و شبکه عصبی مصنوعی در غرب تهران پرداختند و برتری روش شبکه عصبی مصنوعی را اعلام کردند. امیدوار و همکاران (۱۳۹۳) با استفاده از درخت تصمیم به پیش‌بینی بارش در کرمانشاه پرداختند و کارایی قابل قبول آنرا گزارش کردند.

بررسی منابع نشان می‌دهد، محققین مختلف از بین روش‌های موجود جهت بازسازی داده ها صرفاً به بررسی روش‌های خاص و محدودی بسنده کرده‌اند. از طرف دیگر علی‌رغم کاربرد مدل درختی ام۵ در زمینه‌های گوناگونی چون، پیش‌بینی میزان بارش ماهانه (ستاری و همکاران، ۱۳۹۳)، جریان روزانه رودخانه (عبدا.. پور آزاد و همکاران، ۱۳۹۵)، تلفات تبخیر از مخزن سد (ستاری و همکاران، ۱۳۹۳)، تبخیر-تعرق مرجع (اسمعیل‌زاده و ستاری، ۲۰۱۵)، تخمین سیلاب رودخانه (ستاری و همکاران، ۱۳۹۵) و مطالعات هیدرولیکی (صمدیان فرد و ستاری،

[Downloaded from system.khu.ac.ir on 2024-10-12]

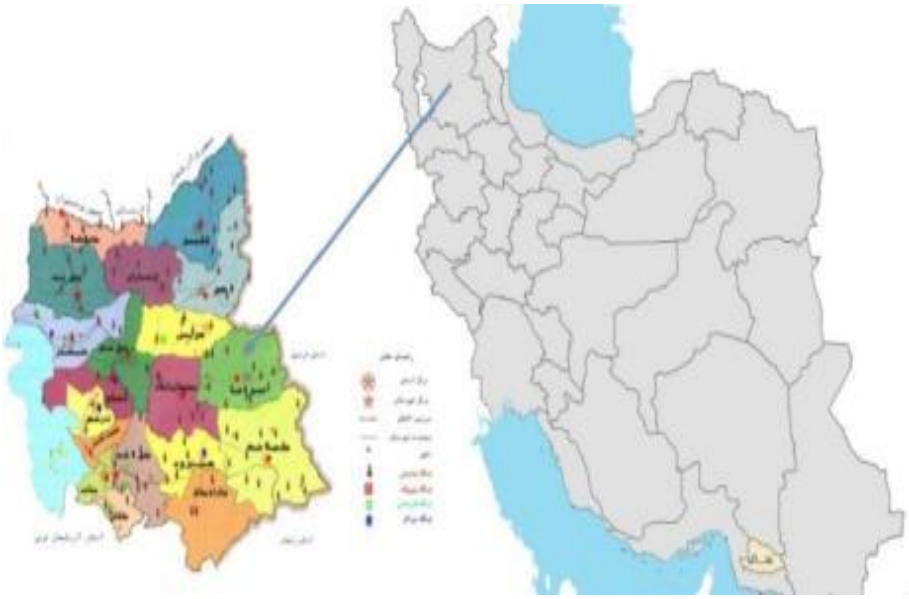
[DOR: 20.1001.1.22287736.1395.16.42.8.3]

۱۳۹۵؛ نهرین و همکاران، ۱۳۹۲؛ مهتابی و ستاری، ۲۰۱۶) مطالعه‌ای در زمینه امکان‌سنجی کاربرد مدل درختی ام۵ در زمینه بازسازی داده‌های بارش ماهانه صورت نپذیرفته است. هدف از مطالعه حاضر بررسی عملکرد و کارایی مدل درختی ام۵ در تخمین بازسازی داده ها بارش ماهانه شهر سراب در استان آذربایجان شرقی، در مقایسه با عملکرد هفت روش آماری کلاسیک شامل روش رگرسیون خطی ساده، رگرسیون خطی چند متغیره، میانگین‌گیری ساده، نسبت نرمال، بهترین تخمین‌گر منفرد، روش نسبت‌ها و روش انتساب چندگانه می‌باشد.

مواد و روش‌ها

منطقه مورد مطالعه و داده‌های استفاده شده

مقادیر بارش ماهانه ایستگاه‌های سینوپتیک استان آذربایجان‌شرقی از قبیل شهرهای تبریز، اهر، مراغه، بناب، سهند، سراب و مرند بین سال‌های ۱۹۵۱ و ۲۰۱۴ موجود می‌باشد. طول دوره آماری با توجه به پایه زمانی مشترک کلیه ایستگاه‌ها، بین سال‌های ۲۰۰۰ تا ۲۰۱۴ در نظر گرفته شده است. شکل (۱) محدوده جغرافیایی منطقه مورد مطالعه و جدول (۱) مختصات جغرافیایی محل قرارگیری ایستگاه‌های سینوپتیک و ارتفاع آنها را نشان می‌دهد.



شکل (۱). موقعیت و محل قرارگیری ایستگاه‌های سینوپتیک مورد استفاده

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

جدول(۱). مختصات جغرافیایی ایستگاه‌های سینوپتیک مورد مطالعه

شهرستان	تبریز	اهر	سراب	مراغه	مرند	سهند	بناب
طول جغرافیایی	۴۶^0 و ۱۷′	۴۷^0 و ۰۴′	۴۷^0 و ۲۳′	۴۶^0 و ۱۰′	۴۵^0 و ۴۶′	۴۶^0 و ۱۲′	۴۶^0 و ۲′
عرض جغرافیایی	۳۸^0 و ۰۵′	۳۸^0 و ۲۶′	۳۷^0 و ۵۶′	۳۷^0 و ۰۱′	۳۸^0 و ۲۶′	۳۷^0 و ۲۰′	۳۷^0 و ۲۰′
ارتفاع از سطح دریا (متر)	۱۳۶۴	۱۳۹۱	۱۶۸۲	۱۳۴۴	۱۵۵۰	۳۶۹۵	۱۲۹۰

در جدول (۲) مشخصات آماری داده‌های استفاده شده در این مطالعه ارائه گردیده است. روش کار در این تحقیق بدین صورت بوده است که، به صورت فرضی تعدادی از داده‌های موجود بارش ماهانه در ایستگاه سراب به صورت تصادفی در دسترس نمی‌باشد و سعی گردیده است کارایی روش‌های کلاسیک آماری و مدل درختی ام۵ در بازسازی داده‌ها مورد ارزیابی قرار گیرد.

جدول (۲). مشخصات آماری مقادیر بارش ماهانه استفاده شده در طول پایه زمانی مشترک ایستگاه‌ها

ایستگاه سینوپتیک	کمینه (میلی‌متر)	بیشینه (میلی‌متر)	میانگین (میلی‌متر)	انحراف استاندارد (میلی‌متر)
اهر	۰	۱۱۹/۶	۲۲/۵۶۶	۲۰/۹۱۹
بناب	۰	۱۲۲/۶	۲۲/۰۷۹	۲۶/۱۸۷
تبریز	۰	۱۱۴/۸	۲۰/۳۰۱	۲۰/۹۴۹
سراب	۰	۱۰۲	۲۰/۶۶۹	۲۰/۰۴۱
سهند	۰	۸۸	۱۷/۷۵۲	۱۹/۳۵۹
مراغه	۰	۱۱۴/۸	۲۱/۱۳۹	۲۴/۸۶۲
مرند	۰	۱۴۳/۷	۳۱/۳۷۱	۲۸/۲۵

روش‌های به‌کار رفته برای کنترل کیفیت آمار

روش آزمون همگنی نرمال استاندارد[1]

روش بررسی همگنی نرمال استاندارد یکی از پرکاربردترین روش‌های آزمون همگنی در تحقیقات روز است. این روش قادر است زمان ایجاد ناپیوستگی و یا بروز ناهمگنی را در سری داده‌ها یافته و گزارش کند. فرض اولیه این آزمون نرمال بودن توزیع آماری داده‌هاست. فرض می‌شود $Y_i(i=1,\ldots,n)$ یک سری زمانی داده بوده و بررسی همگنی آن با توجه به یک سری داده مجاور مورد نظر می‌باشد. در این صورت $Qi(i=1,\ldots,n)$ نیز معرف نسبت (مثلاً داده‌های

[1]- Standard Normal Homogeneity Test

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

بارندگی) موجود بین دو سری داده است. سری Q_i با انجام محاسبات به سری استاندارد Z_I با میانگین صفر و انحراف معیار واحد تبدیل می‌شود:

رابطه(۱) $Z_i=(Q-\bar{Q})/\sigma_Q$

که در آن : $\bar{Q}$ و σQ به ترتیب میانگین و انحراف معیار سری Qi هستند. برای یک تغییر ناگهانی در میانگین سری داده مورد نظر، فرض صفر و فرض یک به صورت زیر تعریف می‌شوند.

رابطه(۲) $H_0: Z_i \approx N(0,1) \quad for\ i = 1, \dots, n \quad and$

رابطه(۳) $H_1: \begin{cases} Z_i \approx N(\mu_1, 1) & for\ i = 1, \dots, a \\ Z_i \approx N(\mu_2, 1) & for\ i = a+1, \dots, n \end{cases}$

که در آن: N معرف توزیع نرمال با پارامترهای مربوطه (میانگین و انحراف معیار) است. الکساندرسون و موبرگ (۱۹۹۷) برای آزمون درستی هر یک از فرض‌های صفر و یک مسأله، یک روش آماری بر اساس نسبت درست نمایی ارائه کردند. بدان معنا که نسبت احتمال درستی H1 به احتمال درستی H_0 برای سری داده مشاهده شده Z_i محاسبه می‌شود. این آزمون عبارت است از:

رابطه(۴) $Z_i = \max\{a\bar{z}_1^2 + (n-a)\bar{z}_2^2\}$

که در آن Z_1 و Z_2 مقادیر میانگین Z_i قبل و بعد از تغییر هستند. مقدار a /محتمل‌ترین زمان بروز تغییر ناگهانی در داده‌ها و یا به بیان دیگر آخرین زمان در سری زمانی جزئی با میانگین Z_1 است. اگر مقدار آماره از مقدار بحرانی آزمون برای سطح بحرانی (مثلاً ۹۵ درصد) بیشتر باشد، آنگاه فرض همگنی صفر در سطح اطمینان مربوطه (مثلاً ۵ درصد) رد می‌شود (قاجارنیا و همکاران، ۱۳۹۳ :۹۴).

آزمون ایستایی داده‌ها توسط آزمون من-کندال[1]

آزمون من-کندال برای بررسی تصادفی بودن و تعیین روند در سری‌ها استفاده می‌شود. در این آزمون تصادفی بودن داده‌ها با عدم وجود روند مشخص می‌شود. در صورت وجود روند داده‌ها غیر تصادفی بوده و برای تعیین تصادفی بودن داده‌ها از آزمون زیر استفاده می‌شود:

رابطه(۵) $T = \dfrac{4P}{N(N-1)}$

در این رابطه T آماره کندال و P مجموع تعداد رتبه‌های بزرگتر از ردیف n_i که بعد از آن قرار می‌گیرند بوده و از رابطه (۶) به‌دست می‌آید.

[1]. Mann-Kendall Trend Test

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

رابطه(۶)

$$P = \sum_{i=1}^{n} ni$$

در این رابطه n تعداد کل سال‌های آماری مورد استفاده می‌باشد. برای سنجش معنی‌دار بودن آماره T از رابطه (۷) استفاده می‌شود:

رابطه(۷)

$$(T)_i = \pm tg\sqrt{\frac{4n+10}{9n(n-1)}}$$

در این رابطه tg برابر با مقدار بحرانی نمره نرمال یا استاندارد (z) با سطح احتمال آزمون است و با سطح احتمال ۹۵ درصد برابر ۱/۹۶ می‌باشد. در صورت اعمال این مقدار، $(T)t$ معادل با ۰/۲۱± می‌شود. با توجه به مقدار بحرانی به‌دست آمده برای $(T)t$ حالات مختلفی بدین شرح مشاهده خواهد شد که، اگر $+(T)t > T > -(T)t$ یا $+0.21 > T > -0.21$ باشد، هیچ‌گونه روند خاصی در سری‌ها مشاهده نمی‌شود و سری‌ها تصادفی هستند. همچنین اگر $T < (T)t$ یا $T < -0.21$ باشد، نشان‌دهنده روند منفی در سری‌ها و در صورتیکه $T > (T)t$ یا $(T)t > +0.21$ باشد، روند مثبت در سری‌ها غالب خواهد بود (عزیزی و روشنی، ۱۳۸۷ :۱۸).

روش‌های استفاده شده برای بازسازی داده های بارش ماهانه

روش رگرسیون خطی ساده[1]

یکی از روش‌های بازسازی داده‌های ناقص روش استفاده از همبستگی بین داده‌های کلیماتولوژی با داده‌های ایستگاه‌های مجاور می‌باشد که از مزایای این روش سادگی نسبی آن در مقایسه با روش‌های دیگر است و معمولاً در بازسازی داده‌های مربوط به ایستگاه‌های کلیماتولوژی واقع در یک منطقه با ویژگی‌های هیدروکلیماتولوژی مشابه نتایج قابل قبولی ارائه می‌دهد (نقدی و همکاران، ۱۳۸۹: ۶۵). در این روش ضریب همبستگی بین داده‌های بارش ماهانه شهر سراب و مقادیر بارش ماهانه سایر ایستگاه‌های واقع در استان محاسبه می‌شود. سپس یک رابطه‌ی خطی بین داده‌های بارش ماهانه ایستگاهی که بالاترین ضریب همبستگی را با ایستگاه شهر سراب داشته باشد با داده‌های بارش ماهانه شهر سراب تشکیل و بازسازی بارش ماهانه بر طبق آن محاسبه می‌گردد.

[1]- Simple Liner Regression

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

روش رگرسیون خطی چند متغیره[1]

تفاوت این روش با روش رگرسیون خطی ساده این‌است که به‌جای استفاده از داده‌های بارش ماهانه یک ایستگاه، از داده‌های بارش ماهانه شامل چندین ایستگاه که بیشترین همبستگی را با مقادیر بارش ماهانه ایستگاه شهر سراب را دارند، استفاده می‌شود و رابطه‌ای خطی بین آن‌ها برازش داده می‌شود.

روش میانگین‌گیری ساده[2]

این روش در واقع همان روش میانگین‌گیری ساده ریاضی است که در هر زمان که ایستگاه مورد نظر دارای داده گم‌شده می‌باشد از اعداد متناظر همان زمان در ایستگاه‌های اطراف ایستگاه مورد نظر میانگین گرفته می‌شود و برابر داده مفقود ایستگاه مورد نظر قرار داده می‌شود. رابطه (۸) جهت محاسبه مقدار مورد نظر توسط این روش ارائه شده است:

رابطه(۸)
$$V_0 = \frac{\sum_{i=1}^{n} V_i}{n}$$

که در این رابطه V_0 داده بازسازی شده ایستگاه مورد نظر، V_i داده iامین ایستگاه نزدیک و n تعداد ایستگاه‌های مجاور می‌باشد.

روش نسبت نرمال[3]

این روش یکی از روش‌های آماری کلاسیک متداول جهت برآورد بازسازی داده ها و ناقص می‌باشد که اولین بار توسط پال هوس و کوهلر (۱۹۵۲) ارائه داده شد و توسط یانگ (۱۹۹۲) اصلاح گردید که از طریق رابطه‌ی (۹) و (۱۰) قابل محاسبه می‌باشد:

رابطه(۹)
$$V_0 = \frac{\sum_{i=1}^{n} WiVi}{\sum_{i=1}^{n} Wi}$$

(۱۰)
$$W_i = \left[r_i^2 \left(\frac{n_i - 2}{1 - r_i^2} \right) \right]$$

در این رابطه، W_i وزن iامین ایستگاه هواشناسی نزدیک به ایستگاه اصلی، V_i داده مشاهداتی در iامین ایستگاه هواشناسی نزدیک ایستگاه اصلی، r_i ضریب همبستگی بین ایستگاه اصلی و

[1]- Multiple Regression Analysis
[2]- Simple Arithmetic Averaging
[3]- Normal Ratio Method

[Downloaded from system.khu.ac.ir on 2024-10-12]
[DOR: 20.1001.1.22287736.1395.16.42.8.3]

ایستگاه iام و n_i تعداد داده‌های موجود در ایستگاه اصلی (تعداد کل داده‌ها منهای داده‌های گم‌شده) می‌باشند.

بهترین تخمین‌گر منفرد[1]

این روش، روشی بسیار ساده است که اساس کار آن بدین صورت است که، بازسازی داده ها ایستگاه اصلی برابر داده‌های ایستگاه مجاوری خواهد بود که بیشترین ضریب همبستگی را با داده‌های موجود ایستگاه اصلی داشته باشد (حسنپور کاشانی و دین پزوه، ۲۰۱۲ :۶۲).

روش نسبت‌ها[2]

در این روش، میانگین داده‌های موجود در ایستگاه اصلی و ایستگاه مجاور (نزدیک‌ترین ایستگاه و یا میانگین ایستگاه‌های مجاور) محاسبه شده و سپس نسبت میانگین ایستگاه اصلی به ایستگاه مجاور تعیین می‌گردد. در نهایت، نسبت به‌دست آمده در مقدار داده موجود ایستگاه مجاور در دوره نیاز به بازسازی، ضرب گشته و داده‌ی نیاز به بازسازی متناظر به‌دست می‌آید (حسنپور کاشانی و دین پزوه، ۲۰۱۲ :۶۲).

روش انتساب چندگانه[3]

یک روش متداول برای بازسازی داده‌ها، روش انتساب چندگانه است. اساس کار این روش بدین صورت است که، هر مقدار داده ناقص با یک توزیع متناظر انتسابی تخمین زده می‌شود. میانگین مجموعه داده‌های منتسب محاسبه و برای پرکردن داده های ناقص ایستگاه هدف بکار می‌رود. در بسیاری از تحقیقات استفاده از تعداد پنج مجموعه داده منتسب مناسب می‌باشد (رادی و همکاران، ۲۰۱۵ :۴۵). به عنوان مثال (شافر و اولسن، ۱۹۹۸ :۵۶۲) پیشنهاد کردند که تعداد سه الی پنج مجموعه داده منتسب برای دست یافتن به نتایج بهینه مناسب هستند. در این تحقیق برای بازسازی داده‌های بارش ماهانه ایستگاه سینوپتیک سراب توسط روش انتساب چندگانه از نرم‌افزار XLSTAT استفاده گردیده است.

[1]- Single Best Estimator
[2]- UK Traditional Method
[3]- Multiple Imputation Method

[Downloaded from system.khu.ac.ir on 2024-10-12]
[DOR: 20.1001.1.22287736.1395.16.42.8.3]

مدل درختی ام۵[1]

مدل درختی ام۵ یکی از روش‌های داده‌کاوی و از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد که برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد. یعنی درخت تصمیم پیش‌بینی خود را در قالب یکسری قوانین توضیح می‌دهد. درحالی‌که در شبکه‌های عصبی تنها پیش‌بینی بیان می‌شود و چگونگی آن در خود شبکه پنهان باقی می‌ماند. ایده کلی مدل درختی ام۵ از درختان رگرسیونی گرفته شده است با این تفاوت که، درختان رگرسیون مقادیر را در برگ‌ها دارند، در صورتی‌که درختانی که با ام۵ ساخته می‌شوند می‌توانند مدل‌های چند متغیره خطی را در برگ‌ها داشته باشند. مزیت مدل درختی ام۵ نسبت به درختان رگرسیون اینست که مدل درختی معمولا بسیار کوچک‌تری نسبت به درختان رگرسیون دارد و با دقت بیشتری به پیش‌بینی مقادیر می‌پردازد (پال و همکاران، ۲۰۱۲ :۱۰۸۲). در مدل درختی ام۵، T مجموعه‌ای از مثال‌های آموزشی موجود است. هر مثال با مقدار معین مجموعه ویژگی‌های ورودی و یک مقدار خروجی معین مشخص شده است. هدف ساختن یک مدل است که مقدار هدف یک مورد آزمایشی را به مقدار ویژگی ورودی آن مرتبط کند. کیفیت یک مدل به طور کلی با دقت آن در پیش‌بینی مقدار هدف از موارد دیده نشده جدید اندازه گیری می‌شود. مدل‌های مبنی بر درخت به روش تقسیم کن و انتخاب کن ساخته می‌شوند. معیار تقسیم برای مدل درختی ام۵ پایه استفاده از انحراف استاندارد مقادیر طبقه است که به یک گروه به‌عنوان اندازه‌گیری خطا در آن گره می‌رسد و محاسبه کاهش خطای مورد انتظار به‌عنوان یک نتیجه از آزمایش هر ویژگی در آن گره است. فرمول محاسبه کاهش انحراف استاندارد در رابطه (۱۱) ارائه شده است:

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i)$$ رابطه(۱۱)

$$Sd(T) = \sqrt{\frac{1}{N}(\sum_{i=1}^{N} y_i^{\,2} - \frac{1}{N}(\sum_{i=1}^{N} y_i)^2)}$$ رابطه(۱۲)

که در آن T مجموعه ای از مثال‌ها را که به گره می‌رسند نشان می‌دهد، T_i زیر مجموعه‌ای از مثال‌ها را نشان می‌دهد که دارای iامین نتیجه بالقوه آن مجموعه است؛ و s_d انحراف استاندارد را نشان می‌دهد. بعد از بررسی تمام بخش کردن‌های ممکن، مدل درختی ام۵ موردی را که کاهش خطای مورد نظر را حداکثر کند، انتخاب می‌کند. تقسیم‌کردن در این

[1] - M5 Model Tree

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

مدل زمانی متوقف می‌شود که مقادیر طبقه از تمام مواردی که به یک گره رسیده‌اند، فقط کمی متفاوت است یا تنها چند نمونه باقی می‌ماند. این تقسیم‌بندی گاهی یک درخت بزرگ تولید می‌کند که شبیه سازه‌ای است که باید هرس شود، به عنوان مثال به وسیله جایگزینی یک شاخه توسط یک برگ در مرحله نهایی، یک فرایند هموارسازی برای ناپیوستگی‌هایی که ناخواسته بین مدل‌های خطی مجاور در برگ درخت‌های هرس[1] شده ایجاد می‌شود، رخ می‌دهد. به خصوص برای مدل‌هایی که از تعداد کمتری مثال‌های آموزشی ساخته شده‌اند، انجام می‌شود. در هموارسازی[2]، معادلات خطی مجاور به صورتی که خروجی‌های پیش‌بینی شده برای بردارهای ورودی همسایه که مربوط به معادلات مختلف هستند و در حال نزدیک شدن به مقادیر مورد نظر می‌باشند، به‌روز می‌شود (کوئینلن، ۱۹۹۲ :۳۴۰). در این تحقیق جهت بازسازی داده های بارش ماهانه سراب توسط مدل درختی ام۵، از نرم افزار WEKA[3] که در دانشگاه Waikato نیوزلند توسعه داده شده، استفاده گردیده است.

معیارهای ارزیابی مدل‌ها

در این تحقیق عملکرد روش‌های آماری کلاسیک و مدل درختی ام۵، بر پایه محاسبه ضریب همبستگی[4]، نش-ساتکلیف[5]، جذر میانگین مربعات خطا[6]، و میانگین خطای مطلق[7] مورد ارزیابی قرار گرفت. فرمول های محاسبه آماره های فوق در روابط (۱۳) تا (۱۶) ارائه گردیده است. در این روابط مقادیر X شامل مقادیر مشاهداتی و مقادیر Y شامل مقادیر محاسباتی و N تعداد داده‌ها می‌باشند.

$$r_{pearson} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (۱۳)$$

$$E = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad (۱۴)$$

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n}(y_i - x_i)^2}{N}} \quad (۱۵)$$

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|X_i - Y_i| \quad (۱۶)$$

[1] - Tree Pruning
[2] - Smoothing
[3] - Waikato Environment for Knowledge Analysis
[4] - Correlation coefficient
[5] - Nash-Sutcliffe coefficient (E)
[6] - Root Mean Square Error
[7] - Mean Absolute Error

[Downloaded from system.khu.ac.ir on 2024-10-12]
[DOR: 20.1001.1.22287736.1395.16.42.8.3]

نتایج

نتایج مربوط به آماده‌سازی داده‌ها

با توجه به اهمیت استفاده از داده‌های همگن در مطالعات پیش‌رو در مرحله اول نسبت به انجام آزمون همگنی نرمال استاندارد اقدام گردید. برای انجام این آزمون از نرم‌افزار آماری XLSTAT استفاده گردیده است. در جدول (۳) نتایج مربوط به آزمون همگنی ایستگاه‌های مورد مطالعه به تفکیک ارائه گردیده است.

جدول (۳). نتایج آزمون همگنی نرمال استاندارد برای ایستگاه‌های مورد مطالعه

نام ایستگاه	p-value	α	ریسک رد کردن فرض صفر
اهر	۰/۶۱۴	۰/۰۵	٪۶۱/۳۹
بناب	۰/۶۷۱	۰/۰۵	٪۶۷/۰۷
تبریز	۰/۴۸۲	۰/۰۵	٪۴۸/۲۴
مراغه	۰/۴۹۷	۰/۰۵	٪۴۹/۷۴
مرند	۰/۴۴۳	۰/۰۵	٪۴۴/۳۰
سراب	۰/۷۵۸	۰/۰۵	٪۷۵/۷۹
سهند	۰/۷۶۳	۰/۰۵	٪۷۶/۳۳

در این آزمون فرض صفر همگنی داده‌ها و فرض یک ناهمگن بودن داده‌ها می‌باشد. چنان‌که مقدار p-value از مقدار درجه اطمینان α بزرگ‌تر باشد فرض صفر صحیح است در غیر این صورت فرض یک قابل قبول می‌باشد. نتایج جدول (۳) نشان‌دهنده این مطلب است که داده‌های بارش ماهانه تمامی ایستگاه‌ها همگن بوده و می‌توانند با اطمینان به‌کار برده شوند. همچنین در جدول (۴) نتایج مربوط به آزمون ایستایی من-کندال که توسط نرم‌افزار XLSTAT انجام شده، ارائه گردیده است. در این مطالعه فرض صفر این است که مجموعه داده دارای روند نمی‌باشد و داده‌ها تصادفی هستند و فرض یک این است که مجموعه داده بررسی شده دارای روند می‌باشد.

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

جدول(۴). نتیجه آزمون آماری تعیین روند من-کندال

نام ایستگاه	Kendall's Tau	p-value	α	ریسک رد کردن فرض صفر
اهر	-۰/۰۱۵	۰/۸۰۴	۰/۰۵	٪۸۰/۴۰
بناب	-۰/۰۳۳	۰/۵۸۲	۰/۰۵	٪۵۸/۲۴
تبریز	-۰/۰۳۹	۰/۵۰۷	۰/۰۵	٪۵۰/۶۸
مراغه	-۰/۰۴۲	۰/۴۷۷	۰/۰۵	٪۴۷/۶۵
مرند	۰/۰۲۷	۰/۶۴۰	۰/۰۵	٪۶۴/۰۲
سراب	-۰/۰۰۱	۰/۹۹۲	۰/۰۵	٪۹۹/۲۳
سهند	۰/۰۴۸	۰/۴۱۱	۰/۰۵	٪۴۱/۱۲

همچنان‌که از جدول (۴) استنباط می‌گردد در کلیه ایستگاه‌های مورد مطالعه مقدار p-value بیشتر از مقدار درجه اهمیت α بوده و تمامی داده‌های مورد استفاده تصادفی بوده و دارای روند خاصی نمی‌باشند. با توجه به اینکه میزان همبستگی مقادیر بارش ماهانه در ایستگاه‌های مختلف در مدل‌سازی آن‌ها بسیار مهم و کاربردی می‌باشد، میزان همبستگی بین مقادیر بارش ماهانه ایستگاه‌های بررسی شده در این تحقیق با مقادیر بارش ماهانه ایستگاه سینوپتیک شهر سراب که به‌عنوان ایستگاه هدف در نظر گرفته شده، در جدول (۵) ارائه شده‌اند.

با توجه به جدول (۵)، ایستگاه سینوپتیک شهرسراب بیشترین میزان همبستگی را با ایستگاه تبریز و سپس با ایستگاه‌های اهر و بناب دارا می‌باشد.

جدول(۵). میزان همبستگی مقادیر بارش ماهانه ایستگاه‌های استان با یکدیگر

	بناب	اهر	سهند	تبریز	مراغه	مرند	سراب
بناب	۱						
اهر	۰/۶۴۵	۱					
سهند	۰/۴۰۴	۰/۳۶۷	۱				
تبریز	۰/۸۲۹	۰/۷۶۷	۰/۴۳۳	۱			
مراغه	۰/۹۵۱	۰/۶۳۶	۰/۴۲۰	۰/۸۲۷	۱		
مرند	۰/۷۰۶	۰/۷۰۰	۰/۳۴۲	۰/۷۵۲	۰/۶۹۹	۱	
سراب	۰/۷۷۹	۰/۸۰۵	۰/۴۵۱	۰/۸۴۱	۰/۷۶۱	۰/۶۸۱	۱

نتایج روش‌های کلاسیک آماری و مدل درختی ام۵

در این مرحله نسبت به بازسازی مقادیر بارش ماهانه ایستگاه سراب با استفاده از روش‌های اشاره شده، اقدام شد. در روش‌های کلاسیک بسته به نوع روش از داده‌های ایستگاه‌هایی که

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

بیشترین میزان همبستگی را با داده‌های ایستگاه سراب داشته‌اند استفاده گردید. در بازسازی داده های بارش شهر سراب توسط مدل درختی ام۵، از مدل درختی قواعد ام۵ بهره گرفته شد. در به‌کارگیری این روش ابتدا سناریوهای متعددی شامل ترکیب پارامترهای ورودی شهرهای مختلف طرح گردید و مشاهده شد، بهترین نتیجه زمانی حاصل می‌گردد که از داده‌های بارش ماهانه ایستگاه‌های شهرهای تبریز، اهر و بناب استفاده شود. این روش با ارائه چهار رابطه خطی ساده اگر-آنگاه توانست مقدار بارش ماهانه ایستگاه سراب را تخمین بزند. این روابط به صورت زیر ارائه شده است:

$$Rule\ 1: if\ Tabriz \leq 21.2 \Rightarrow Sarab = (0.017 * Bonab) + (0.2308 * Ahar) + (0.409 * Tabriz) + 4.2236$$ رابطه (۱۷)

$$Rule\ 2: if\ Ahar \leq 52.5 \Rightarrow Sarab = (0.2216 * Bonab) + (0.3418 * Ahar) + (0.0518 * Tabriz) + 9.4086$$ رابطه (۱۸)

$$Rule\ 3: if\ Tabriz > 51.65 \Rightarrow Sarab = (0.2006 * Tabriz) + 60.2181$$ رابطه (۱۹)

$$Rule\ 4: Sarab = 57.35$$ رابطه (۲۰)

به‌عنوان مثال در قانون یک بیان شده است که اگر میزان بارش ماهانه تبریز کوچکتر یا مساوی ۲۱/۲ باشد میزان بارش ماهانه شهر سراب از رابطه (۱۷) قابل محاسبه می‌باشد. همچنین در قانون دو آمده است که اگر میزان بارش ماهانه اهر کوچکتر یا مساوی ۵۲/۵ باشد آنگاه میزان بارش ماهانه شهر سراب از رابطه متناظر (۱۸) محاسبه شده و همچنین قانون سه بیانگر این مطلب است که اگر بارش ماهانه شهر تبریز بزرگتر از ۵۱/۶۵ باشد آنگاه میزان بارش ماهانه شهر سراب از رابطه (۱۹) و طبق قانون چهار در سایر موارد از رابطه (۲۰) قابل محاسبه می‌باشد. نتایج مربوط به آماره‌های ارزیابی به‌دست آمده از روش‌های آماری کلاسیک و مدل درختی قواعد ام۵ در جدول (۶) ارائه گردیده است. در جدول (۷) مقادیر مشاهداتی و محاسباتی بارش ماهانه به‌ازای ماه‌هایی که نیاز به بازسازی دارند، ارائه شده است.

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

جدول (۶). آماره‌های ارزیابی به دست امده از روش‌های کلاسیک آماری و مدل درختی قواعد ام۵

نام روش	شاخص آماری			
	R	N-S	RMSE (mm)	MAE (mm)
رگرسیون خطی ساده	۰/۸۷۰	۰/۷۸۳	۱۰/۵۴۰	۷/۶۰۶
رگرسیون خطی چند متغیره	۰/۹۶۹	۰/۹۱۵	۶/۵۵۷	۴/۲۶۴
میانگین‌گیری ساده	۰/۹۲۸	۰/۸۴۴	۸/۹۱۱	۵/۷۱۹
نسبت نرمال	۰/۹۷۰	۰/۹۱۴	۶/۶۱۲	۴/۹۲۰
بهترین تخمین‌گر منفرد	۰/۸۷۰	۰/۷۵۶	۱۱/۱۷۱	۴/۵۶۶
روش نسبت‌ها	۰/۸۵۲	۰/۷۵۴	۱۱/۲۰۸	۸/۰۰۱
روش انتساب چندگانه	۰/۹۷۴	۰/۹۳۶	۵/۸۰۹	۴/۷۵۳
مدل درختی قواعد ام۵	۰/۹۷۴	۰/۹۴۸	۵/۱۱	۴/۱۸۹

جدول (۷). مقادیر مشاهداتی و محاسباتی بارش ماهانه به‌ازای ماه‌های دارای بازسازی شده

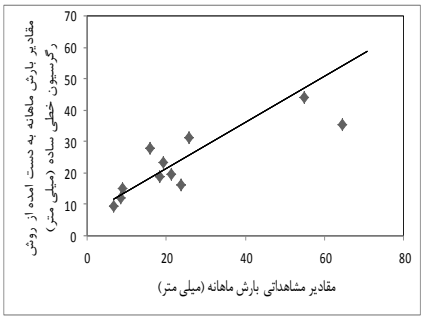
ماه	مشاهداتی	رگرسیون خطی ساده	رگرسیون خطی چند متغیره	میانگین‌گیری ساده	نسبت نرمال	بهترین تخمینگر منفرد	نسبت ها	انتساب چندگانه	مدل درختی ام۵
ژانویه ۲۰۰۳	۱۸/۴	۱۸/۸۳	۱۷/۵۹	۱۳/۸۳	۱۶/۲۷	۱۸/۱	۱۸/۱۷	۱۹/۹۴	۱۵/۲۱
نوامبر ۲۰۰۳	۱۹/۲	۲۳/۵۳	۱۶/۷۲	۲۶/۶۸	۲۵/۶۶	۲۳/۹	۲۳/۹۹	۱۲/۰۷	۱۹/۹۰
مارس ۲۰۰۴	۷۰/۸	۷۸/۱۳	۶۶/۷۲	۵۸/۱۶	۸۳/۱۳	۹۱/۳	۹۱/۶۵	۶۸/۰۸	۷۸/۵۳
مارس ۲۰۰۶	۱۵/۹	۲۸/۰۷	۲۲/۴۳	۲۹/۷	۳۱/۵	۲۹/۵	۲۹/۶۱	۲۲/۳۹	۲۶/۷
فوریه ۲۰۰۷	۲۱/۳	۱۹/۴۸	۲۳/۳۱	۲۸/۲۸	۲۴/۰	۱۸/۹	۱۸/۹۷	۲۰/۸	۲۳/۲۶
اوریل ۲۰۰۸	۸/۵	۱۲/۱۹	۱۴/۳۶	۱۸/۲۷	۱۰/۴۷	۹/۹	۹/۹۴	۱۸/۶۷	۱۱/۶۶
اکتبر ۲۰۰۸	۵۴/۸	۴۳/۹۵	۴۸/۷۴	۵۱/۰۸	۵۷/۹۸	۴۹/۱	۴۹/۲۹	۴۷/۳	۵۷/۳۵
ژانویه ۲۰۰۹	۸/۹	۱۵/۰۳	۱۴/۵۸	۱۶/۶	۱۳/۶۳	۱۳/۴	۱۳/۴۵	۱۶/۹۵	۱/۸۲
مه ۲۰۱۰	۶۴/۵	۳۵/۳۶	۵۳/۳۶	۵۱/۳۶	۶۱/۱۴	۳۸/۵	۳۸/۶۵	۵۵/۸۸	۵۷/۳۵
اکتبر ۲۰۱۱	۲۳/۸	۱۶/۳۲	۲۰/۷۲	۱۴/۲۶	۲۲/۸۷	۱۵	۱۵/۰۶	۲۲/۶۷	۱۸/۸۲
فوریه ۲۰۱۲	۶/۸	۹/۳۶	۹/۷۳	۸/۶۵	۶/۳۵	۶/۴	۶/۴۲	۱۰/۰	۸/۶۲
دسامبر ۲۰۱۴	۲۵/۸	۳۱/۱۵	۲۴/۵۴	۳۲/۲۳	۳۱/۰۲	۳۳/۳	۳۳/۴۳	۲۶/۴۱	۲۴/۵۲

همچنان‌که از نتایج به‌دست آمده از جدول (۶) و (۷) قابل مشاهده است، از بین روش‌های آماری کلاسیک، روش انتساب چندگانه و روش نسبت نرمال دارای نتایج نسبتاً دقیق‌تری می‌باشند و پس از آن‌ها روش رگرسیون خطی چندمتغیره نتایج قابل قبولی ارائه کرده است. همچنین از بین روش‌های بررسی شده در این مطالعه روش نسبت‌ها با ارائه مقادیر کمتر

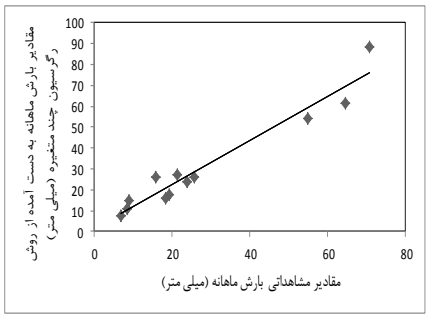
[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

ضرایب همبستگی و بالاتر خطا دارای ضعیف‌ترین نتیجه می‌باشد. اگرچه روش‌های انتساب چندگانه، نسبت نرمال و رگرسیون خطی چند متغیره نتایج دقیقی ارائه کرده‌اند اما با بررسی دقیق‌تر میزان ضریب همبستگی و ضریب نش-ساتکلیف و با در نظر گرفتن آماره‌های مربوط به خطا شامل میانگین خطای مطلق و ریشه میانگین مربعات خطا مشخص می‌شود که در حالت کلی مدل درختی قواعد ام۵ که از روش‌های نوین داده‌کاوی به‌شمار می‌رود و از قوانین اگر-آنگاه بهره می‌گیرد، توانسته است با چهار رابطه خطی ساده و مقادیر آماره‌های (ضریب همبستگی برابر ۰/۹۷۴، ضریب نش-ساتکلیف برابر ۰/۹۴۸، ریشه میانگین مربعات خطا برابر ۵/۱۱ (میلی‌متر) و میانگین خطای مطلق برابر با ۴/۱۸۹(میلی‌متر) دقیق‌ترین نتیجه را از بین روش‌های بررسی شده در این مطالعه ارائه دهد. شکل‌های (۲) تا (۹) نشان‌دهنده نمودارهای پراکنش مقادیر محاسباتی به‌دست آمده از روش‌های مختلف نسبت به مقادیر بازسازی شده بارش ماهانه در ایستگاه سراب می‌باشند.



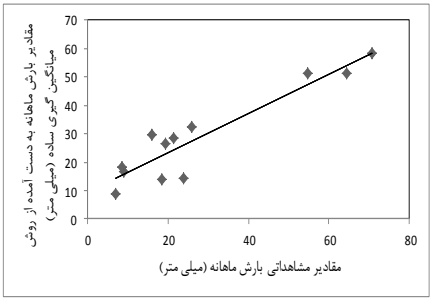
شکل(۲). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش رگرسیون خطی ساده نسبت به مقادیر مشاهداتی



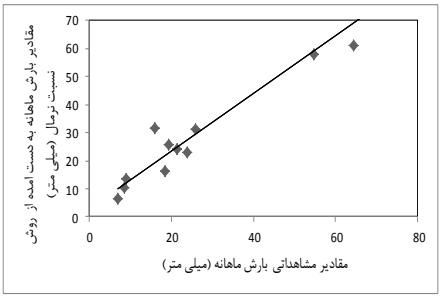
شکل(۳). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش رگرسیون چند متغیره نسبت به مقادیر مشاهداتی

[Downloaded from system.khu.ac.ir on 2024-10-12]

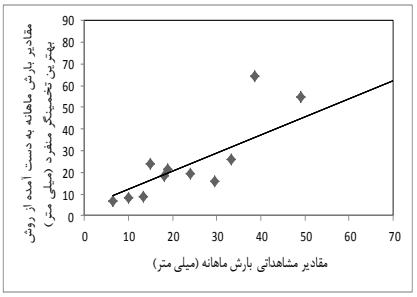
[DOR: 20.1001.1.22287736.1395.16.42.8.3]



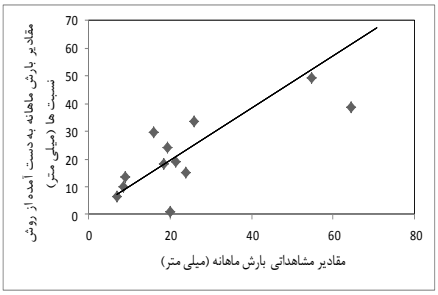
شکل(۴). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش میانگین‌گیری ساده نسبت به مقادیر مشاهداتی



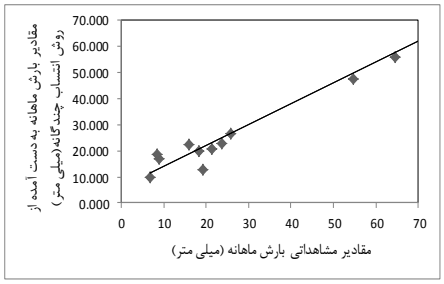
شکل(۵). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش نسبت نرمال نسبت به مقادیر مشاهداتی



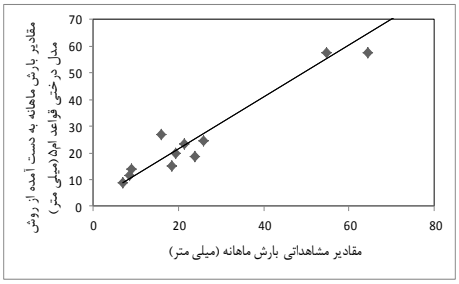
شکل(۶). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش بهترین تخمین گر منفرد نسبت به مقادیر مشاهداتی



شکل(۷). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش نسبت‌ها نسبت به مقادیر مشاهداتی



شکل(۸). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش انتساب چندگانه نسبت به مقادیر مشاهداتی



شکل(۹). مقادیر محاسباتی بارش ماهانه به‌دست آمده از روش مدل درختی قواعد ام۵ نسبت به مقادیر مشاهداتی

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

همچنان‌که از بررسی دقیق‌تر شکل‌های (۲) تا (۹) مشخص می‌گردد، مقادیر به‌دست آمده از مدل درختی قواعد ام۵ و روش انتساب چندگانه برای داده های بازسازی شده ایستگاه سراب از پراکنش و همخوانی بیشتری نسبت به داده‌های مشاهداتی آن ایستگاه، نسبت به سایر روش‌های بررسی شده در این مطالعه دارد که نشان‌دهنده توانایی بالای این روش‌ها در تخمین مقادیر بازسازی شده بارش ماهانه می‌باشد. در شکل (۱۰) سری زمانی بارش ماهانه به دست آمده توسط روش انتساب چندگانه، روش نسبت نرمال و رگرسیون خطی چند متغیره به‌عنوان بهترین روش‌های کلاسیک آماری بررسی شده و مدل درختی قواعد ام۵ نسبت به مقادیر مشاهداتی بارش ماهانه ایستگاه سراب که به صورت بازسازی در نظر گرفته شده‌اند، ارائه شده است که نشان‌دهنده دقت و همخوانی بیشتر نتایج به‌دست آمده از روش مدل درختی قواعد ام۵ نسبت به سایر روش‌ها می‌باشد.

شکل (۱۰). سری زمانی بارش ماهانه به دست آمده از روش‌های انتساب چندگانه، نسبت نرمال، رگرسیون خطی چند متغیره و مدل درختی قواعد ام۵ و مقادیر مشاهداتی بارش ماهانه

نتیجه‌گیری

با توجه به نقش پررنگ دسترسی به آمار و اطلاعات دقیق در مطالعات هیدرولوژیکی و کلیماتولوژیکی، در این تحقیق آمار مقادیر بارش ماهانه ایستگاه‌های سینوپتیک استان آذربایجان‌شرقی از نقطه نظر همگنی توسط آزمون نرمال استاندارد و از نظر ایستایی توسط آزمون آماری من- کندال مورد بررسی قرار گرفتند. نتایج نشان‌دهنده همگن بودن و عدم وجود

[Downloaded from system.khu.ac.ir on 2024-10-12]
[DOR: 20.1001.1.22287736.1395.16.42.8.3]

روند خاص در مجموعه داده‌های مورد بررسی است. سپس با استفاده از روش‌های مختلف شامل روش‌های رگرسیون خطی ساده، رگرسیون خطی چند متغیره، میانگین‌گیری ساده، نسبت نرمال، بهترین تخمین‌گر منفرد، روش نسبت‌ها، روش انتساب چندگانه و مدل درختی قواعد ام۵، نسبت به تخمین مقادیر بارش ماهانه ایستگاه سراب که به‌صورت تصادفی تعدادی از داده‌های آن نیاز به بازسازی داشته، اقدام گردید. نتایج نشان داد از بین روش‌های کلاسیک آماری بررسی شده روش انتساب چندگانه و روش نسبت نرمال دارای نتایج دقیق‌تر و خطاهای کمتری می‌باشند.

در تحقیق مشابهی (لوک زاده، ۱۳۸۳) نیز برتری روش آماری نسبت نرمال را در اکثر موارد بررسی شده اعلام نمود. همچنین در تحقیق دیگری حسنپور کاشانی و دین پژوه (۲۰۱۲) بعد از اعلام برتری روش‌های شبکه عصبی مصنوعی از میان روش‌های کلاسیک آماری بررسی شده برتری روش رگرسیون خطی چند متغیره را اعلام نمودند که این روش در تحقیق حاضر نیز نتایج دقیقی ارائه کرده است. در روش انتساب چندگانه مشخص شد که بهترین نتیجه زمانی حاصل می‌گردد که از تعداد ۵ ایستگاه وابسته استفاده گردد، که این مسأله نیز با تحقیقات رادی و همکاران (۲۰۱۵) همخوانی دارد. بررسی کلی نتایج نشان‌دهنده این موضوع می‌باشد که مدل درختی قواعد ام۵ که از روش‌های کارآمد داده‌کاوی به‌شمار می‌رود و از قوانین اگر- آنگاه بهره می‌گیرد توانسته است با ارائه چهار رابطه خطی اگر- آنگاه ساده و با ارائه مقادیر آماره‌های (ضریب همبستگی برابر ۰/۹۷۴، ضریب نش-ساتکلیف برابر ۰/۹۴۸، ریشه میانگین مربعات خطا برابر ۵/۱۱ (میلی‌متر) و میانگین خطای مطلق برابر با ۴/۱۸۹ (میلی‌متر) دقیق‌ترین نتیجه را در بین کلیه روش‌های بررسی شده در این مطالعه ارائه دهد. لذا با توجه به سادگی روند مدل‌سازی با این مدل و دقت بالای آن استفاده از آن برای تخمین مقادیر گم‌شده بارش ماهانه در موارد مشابه کلیماتولوژیکی پیشنهاد می‌گردد.

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

منابع و ماخذ

۱. امیدوار، کمال.، شفیعی، شهاب.، تقیزاده، زهرا.، علی پور، مهرداد. (۱۳۹۳). **ارزیابی کارایی مدل درخت تصمیم در پیش‌بینی بارش ایستگاه سینوپتیک کرمانشاه**، مجله تحقیقات کاربردی علوم جغرافیایی، ۳۴(۳): ۸۹-۱۱۰.

۲. انصاری، حسین؛ داوری، کامران (۱۳۸۹). **تخمین و بازسازی داده‌های بارندگی با تکنیک فازی**، مجله تحقیقات منابع آب ایران، شماره ۶(۱): ۳۹-۴۷.

۳. ساداتی‌نژاد، سید جواد؛ نقدی، رحیم؛ شایان‌نژاد، محمد (۱۳۸۹). **کاربرد روش رگرسیون خطی فازی در برآورد داده‌های ناقص دبی سالانه ایستگاه‌های هیدرومتری و مقایسه آن با سایر روش‌های متداول**، مجله پژوهش‌های حفاظت آب و خاک، شماره ۱۷(۴): ۶۷-۸۶.

۴. ستاری، محمدتقی؛ احمدی‌فر، وحدت و پاشاپور خلف انصار، رسول (۱۳۹۳). **مدلسازی تلفات تبخیر از مخزن سد علویان با استفاده از مدل درختی M5 و مقایسه آن با روش‌های تجربی**، مهندسی آبیاری و آب، ۵(۱۷): ۱۱۰-۱۲۲.

۵. ستاری، محمدتقی؛ رضازاده جودی، علی و نهرین، فرناز (۱۳۹۳). **پیش‌بینی مقادیر بارش ماهانه با استفاده از شبکه عصبی مصنوعی و مدل درختی M5 (مطالعه موردی: ایستگاه اهر)**، پژوهش‌های جغرافیای طبیعی، ۴۶(۲): ۴۷-۲۶۰.

۶. ستاری، محمدتقی؛ عبدا.. پور آزاد، محمدرضا و میرعباسی نجف آبادی، رسول (۱۳۹۵). **پیش‌بینی سیلاب‌های ساعتی رودخانه اهرچای با استفاده از روش‌های یادگیری ماشین**، نشریه مهندسی و مدیریت آبخیز، ۸(۱): ۱۱۵-۱۲۷.

۷. صمدیان فرد، سعید؛ ستاری، محمدتقی (۱۳۹۵). **مدل‌سازی ضریب اصطکاک جریان در لوله‌های آبیاری با استفاده از روش‌های یادگیری ماشینی و مقایسه عملکرد آنها با روابط تجربی**، نشریه دانش آب و خاک، ۲۶(۱/۲): ۴۵-۵۷.

۸. عبدا.. پور آزاد، محمدرضا؛ ستاری، محمدتقی و میرعباسی نجف آبادی، رسول (۱۳۹۵). **پیش‌بینی جریان روزانه رودخانه اهرچای با استفاده از مدل قوانین M5 و مقایسه آن با شبکه‌های عصبی مصنوعی المانی (ENN)**، ۱۰(۳۳): ۱۱-۱۸.

۹. عزیزی، قاسم؛ روشنی، محمود (۱۳۸۷). **مطالعه تغییر اقلیم در سواحل جنوبی دریای خزر به روش من-کندال**، مجله پژوهش‌های جغرافیایی، شماره ۶۴(۲): ۱۳-۲۸.

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

۱۰. قاجارنیا، نوید؛ لیاقت، عبدالمجید و دانش‌کار آراسته، پیمان (۱۳۹۳). **صحت‌سنجی داده‌های بارندگی ایستگاه‌های غیر ثبات سازمان هواشناسی و تماب در حوزه آبریز دریاچه اورمیه،** نشریه حفاظن منابع آب و خاک، شماره ۴(۱): ۹۱-۱۰۹.

۱۱. لوک‌زاده، صمد (۱۳۸۳). **ارزیابی چند روش مختلف بازسازی خلأهای آماری بارندگی در مقاطع زمانی مختلف در منطقه البرز مرکزی،** پایان‌نامه کارشناسی ارشد، هواشناسی کشاورزی، دانشکده کشاورزی، دانشگاه تهران.

۱۲. محمدخورشیددوست، علی؛ نساجی زواره، مجتبی؛ قرمزچشمه، باقر (۱۳۹۱). **بازسازی سری‌های زمانی دمای حداکثر و حداقل روزانه با استفاده از روش نزدیک‌ترین همسایه و شبکه عصبی مصنوعی (مطالعه موردی: غرب استان تهران)** ، فصلنامه علمی-پژوهشی فضای جغرافیایی، شماره ۳۸(۲): ۱۹۷-۲۱۴.

۱۳. نقدی، رحیم؛ شایان‌نژاد، محمد و ساداتی‌نژاد، سید جواد (۱۳۸۹). **مقایسه روش‌های مختلف تخمین بازسازی داده‌های دبی ماهانه حوضه آبخیز کارون بزرگ،** پژوهشنامه مدیریت حوزه آبخیز، شماره ۱(۱): ۵۹-۷۳.

۱۴. نهرین، فرناز؛ ستاری، محمدتقی و سلماسی، فرزین (۱۳۹۲). **پیش‌بینی اتلاف کارمایه در سرریس توری سنگی پله‌ای با استفاده از شبیه درختی M5**، مهندسی منابع آب، ۶: ۷۵-۸۶.

Alexandersson H, Moberg A. (1997). ***Homogenization of Swedish temperature data***. Part 1: homogeneity test for linear trends. International Journal of Climatology, 17: pp. 25–34.

Che Ghani, Nor Zaimah, Abuhasan, Zorkefle, and Tze Liang, Lau, (2014). ***Estimation of Missing Rainfall Data Using GEP: Case Study of Raja River, Alor Setar, Kedah, Advances in Artificial Intelligence,*** http://dx.doi.org/10.1155/2014/716398, pp. 1-5.

De silva, R.P, Dayawansa, N.D.K, Ratnasiri, M.D, (2007). **A *Comparison of methods used in Estimating missing rainfall data***, the journal of agricultural sciences, 3(2): pp. 101-108.

Esmaeilzadeh B, Sattari MT. (2015). ***Monthly Evapotranspiration Modeling using Intelligent Systems in Tabriz, Iran***, Agriculture Science Developments, 4(3): PP. 35-40.

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]

Hasanpur Kashani, Mahsa, and Dinpashoh, Yagub, (2012). ***Evaluation of efficiency of different estimation methods for missing climatological data***, journal of stochastic environment research risk assessment, 26: pp. 59-71.

K. Choge, Harshand, and Regulwar, D.G, (2013). ***Artificial Neural Network Method for Estimation of Missing Data,*** International Journal of Advanced Technology in Civil Engineering, 2(1): pp. 1-4.

Kim, Jung-Woo, and A. Pachepsky, Yakov, (2010). ***Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation***, Journal of Hydrology 394: pp. 305–314.

Mahesh Pal; N. K. Singh; N. K. Tiwari, (2012). ***M5 Model Tree for Pier Scour Prediction Using Field Dataset***, KSCE Journal of Civil Engineering, 16(6): pp. 1079-1084.

Mahtabi G, Sattari MT. (2016). ***Investigation of Hydraulic Jump Characteristics in Rough Beds Using M5 Model Tree***, Jordan Journal of Agricultural Sciences, 12(2): pp. 631-647.

Paulhus JLH, Kohler MA, (1952). ***Interpolation of missing precipitation records***. Mon Weather Rev 80: pp.129–133.

Quinlan J.R, (1992). ***Learning with continuous classes***, In proceedings AI, 92 (Adams & Sterling, Eds), Singapore: World Scientific, pp. 343-348.

Radi N, Zakaria R, Azman M. (2015). ***Estimation of Missing Rainfall Data Using Spatial Interpolation and Imputation Methods***. AIP conference proceedings, 1643(1), 42-48. DOI: 10.1063/1.4907423.

Schafer, J. L., and Olsen, M. K., (1998). ***Multivariate behavioral research***, 33(4), pp. 545-571.

You, Jinsheng, Hubbard, Kenneth G, Goddard, Steve, (2008). ***Comparison of methods for spatially estimating station temperatures in a quality control system,*** International Journal of Climatology, 28: pp. 777-787.

Young KC, (1992). ***A three-way model for interpolating monthly precipitation values***. Mon Weather Rev, 120: pp. 2561–2569.

[Downloaded from system.khu.ac.ir on 2024-10-12]

[DOR: 20.1001.1.22287736.1395.16.42.8.3]