نشریه علمی-پژوهشی مهندسی و مدیریت آبخیز
جلد ۱۴، شماره ۱، ۱۴۰۱، صفحات ۲۶-۳۹

Journal of Watershed Engineering and Management
Volume 14, Issue 1, 2022, Pages 26-39
DOI: 10.22092/ijwmse.2021.342929.1793

ارزیابی کارایی روش‌های مرسوم بازسازی نبودهای آماری در مناطق خشک ایران

مرتضی میری*[۱]، محمدرضا کوثری[۲] و مهران زند[۳]

[۱ و ۲] استادیار، پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران و [۳]دانشیار، پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران

تاریخ دریافت: ۱۳۹۹/۰۸/۲۲ تاریخ پذیرش: ۱۳۹۹/۱۲/۲۱

چکیده

یکی از مشکلات شایع و تأثیرگذار در مطالعات اقلیمی، وجود کمبودهای فراوان در سری‌های زمانی داده‌های مختلف اقلیمی و هیدرولوژی است. این پژوهش، با هدف ارزیابی دقت روش‌های بازسازی پارامترهای دمایی در مناطق خشک ایران انجام شده است. برای این منظور، داده‌های روزانه دمای کمینه، متوسط و دمای بیشینه ۷۳ ایستگاه هم‌دیدی با پراکنش مناسب در مناطق خشک ایران از آرشیو سازمان هواشناسی کل کشور دریافت و پردازش شد. روش‌های بازسازی مورد استفاده شامل روش‌های نسبت نرمال، رگرسیون خطی، رگرسیون چندگانه و عکس مجذور فاصله (IDW) می‌شوند. در این تحقیق، کارایی هر یک از روش‌های مذکور در بازسازی نبودهای آماری پارامترهای دمایی در مقیاس‌های زمانی روزانه، ماهانه و سالانه مناطق خشک کشور متناسب با میزان نسبت کمبود داده‌های گم‌شده (از پنج تا ۵۰ درصد کمبود داده‌ها) مورد ارزیابی قرار گرفت. برای ارزیابی آماری روش‌های نام‌برده از آماره‌های ضریب همبستگی R، معیار برآورد خطای RMSE و ضریب Nash استفاده شد. نتایج نشان داد که در حالت کلی، هر کدام از روش‌های نام‌برده شده در سطح خاصی از بازسازی و مقیاس زمانی از عملکرد مناسبی برخوردارند. در مقیاس سالانه و ماهانه روش‌های رگرسیون خطی و نسبت نرمال بیشترین دقت در بازسازی داده‌های دما در منطقه خشک ایران را دارند. مقدار همبستگی بین داده‌های بازسازی و مشاهده‌ای در سطوح مختلف با استفاده از این روش‌ها به بیش از ۰/۹۵ می‌رسد. در مقیاس روزانه تفاوت قابل توجهی بین دقت روش‌های مورد استفاده در بازسازی داده‌های دمایی وجود ندارد و تقریباً هر چهار روش یاد شده از دقت مناسبی برخوردارند؛ چرا که میزان همبستگی بین داده‌های بازسازی‌شده و مشاهده‌شده به بیش از ۹۰ درصد می‌رسد. با وجود این، روش رگرسیون چندگانه با میانگین همبستگی ۰/۹۹ درصد در سطوح مختلف بازسازی از بیشترین دقت در بازسازی داده‌های روزانه برخوردار است.

واژه‌های کلیدی: دما، رگرسیون خطی، ضریب همبستگی، کمبود داده، نسبت نرمال

مقدمه

انجام تحقیقات اقلیمی در درازمدت نیازمند در دسترس بودن داده‌های با کیفت و طول دوره آماری مناسب است. در بسیاری از مناطق جهان به‌ویژه در کشورهای در حال توسعه، نواحی کوهستانی، بیابانی و روستایی، دسترسی به داده‌های جوی و زمینی همچون بارش، دما، رطوبت و دمای خاک همواره با مشکلات متعددی همراه می‌باشد؛ زیرا تراکم پائین

* مسئول مکاتبات: morteza.miri64@gmail.com

ایستگاه‌های هواشناسی در نواحی کوهستانی و بیابانی جهان به‌دلیل محدودیت‌های اقتصادی و یا جغرافیایی، استفاده از داده‌های این ایستگاه‌ها را در مقیاس جهانی و منطقه‌ای با محدودیت مواجه کرده است (Mishra و همکاران ۲۰۱۱؛ Gairola و همکاران ۲۰۱۵؛ Miri و همکاران، ۲۰۱۶). علاوه‌بر دسترس نبودن داده‌ها، وجود داده‌های گم‌شده فراوان در بسیاری از ایستگاه‌های هواشناسی به‌دلیل خطاهای سیستماتیک، خراب شدن دستگاه‌های سنجش، قطع برق، حذف داده‌های نادرست (Aieb و همکاران، ۲۰۱۹) و ... یکی از اساسی‌ترین مشکلاتی است که پژوهشگران کشورهای مختلف با آن روبه‌رو هستند. از این‌رو، در سال‌های اخیر روش‌های متفاوتی برای برآورد و تخمین داده‌های گم‌شده توسط پژوهشگران مختلف ارائه شده است. انتخاب نوع روش به عواملی نظیر طول سری زمانی داده‌های گم‌شده، در دسترس بودن داده‌های آب و هواشناسی از ایستگاه‌های هم‌جوار، فصل، مقادیر گم‌شده، منطقه آب و هوایی موردنظر، دانش و تخصص شخص مسئول تصحیح داده‌ها، طول دوره داده برداری داده‌های موجود، اهمیت کیفیت داده برای مدل موردنظر بستگی دارد (Ress، ۲۰۰۸). روش‌های مختلفی نظیر رگرسیون چندگانه، شبکه‌های عصبی مصنوعی، روش‌های فازی، تحلیل سری‌های زمانی و تکنیک‌های درون‌یابی، رگرسیون خطی، روش نسبت نرمال، روش‌های مبتنی بر تحلیل فضایی داده‌ها مانند روش عکس مجذور فاصله و یا کریجینگ برای این منظور توسعه یافته‌اند. البته، در ایران و در مطالعات منابع طبیعی، از میان روش‌های پرکاربرد، چهار روش رگرسیون خطی ساده و چندگانه، نسبت نرمال و وزن‌دهی عکس فاصله (IDW) بیشترین کاربرد را دارا هستند. در واقع، از این روش‌ها در بسیاری از مطالعات پایه اقلیم‌شناسی خصوصاً در رابطه با مطالعات حوزه‌های آبخیز استفاده می‌شود. با توجه به اهمیت رفع داده‌های مفقودی در سری زمانی اقلیمی، مطالعات زیادی در ارتباط با پیشنهاد روش‌های مختلف بازسازی آماری و ارزیابی کارایی آن‌ها با هدف‌های متفاوت در سطح جهان انجام شده است. از جمله تحقیقات انجام شده می‌توان به Price و همکاران (۲۰۰۰)، Teegavarapu و Chandramouli (۲۰۰۵)، Hofstra و همکاران (۲۰۰۸)، Kim و همکاران (۲۰۱۰) و Wagner و همکاران (۲۰۱۲) اشاره کرد.

Xia و همکاران (۱۹۹۹) به مقایسه کارایی شش روش (میانگین حسابی ساده، معکوس فاصله، نسبت نرمال، کریجینگ عمومی، بهترین برآوردگر منفی و تجزیه و تحلیل رگرسیون چندگانه) بازسازی داده‌های گپ در سری زمانی داده‌های اقلیمی پرداختند. نتایج نشان داد، برای بازسازی سری‌های زمانی دما روش تجزیه و تحلیل رگرسیون چندگانه بهترین نتایج را دارد. Kotsiantis و همکاران (۲۰۰۶) با استفاده از آماره‌های R، RMSE کارایی روش‌های مختلف بازسازی را بررسی و بیان کردند که روش رگرسیون از کارایی مناسبی برای بازسازی داده‌های دمایی برخوردار است. You و همکاران (۲۰۰۸) با مقایسه عملکرد دو روش معکوس فاصله و رگرسیون فضایی در بازسازی حداقل و حداکثر دمای روزانه ایالات‌متحده نشان دادند که روش رگرسیون در مناطق کوهستانی و ساحلی عملکرد بهتری دارد. Hasanpour kashahni و Dinpashoh (۲۰۱۲) ۱۱ روش کلاسیک را برای بازسازی نبودهای آماری داده‌ها مورد ارزیابی قرار دادند. نتایج ایشان نشان می‌دهد که هر چند روش‌های مربوط به هوش مصنوعی با پیچیدگی و محاسبات بیشتری برای پیدا کردن ساختارهای بهینه همراه هستند، اما دقت بیشتری نسبت به روش‌های کلاسیک دارند. Yozgatligil و همکاران (۲۰۱۳) با استفاده از شش روش به بازسازی داده‌های بارش و دمای ماهانه ترکیه پرداختند. نتایج آن‌ها نشان داد، تکنیک جای‌گذاری چندگانه با اعمال برخی اصلاحات نتایج بهتری برای بازسازی نسبت به سایر روش‌ها دارد. Tardivo و Barti (۲۰۱۴) با استفاده از روش‌های مبتنی بر رگرسیون اقدام به بازسازی داده‌های مفقودی پارامترهای دمایی پرداختند. Sattari و همکاران (۲۰۱۶) با استفاده از روش‌های میانگین حسابی، رگرسیون چندگانه و NIPALS (Non-Linear Iterative Partial Least Squares) به بازسازی داده‌های بارش ماهانه در شش ایستگاه استان هرمزگان پرداختند. نتایج آن‌ها نشان داد که روش رگرسیون چندگانه از دقت بیشتری برخوردار است.

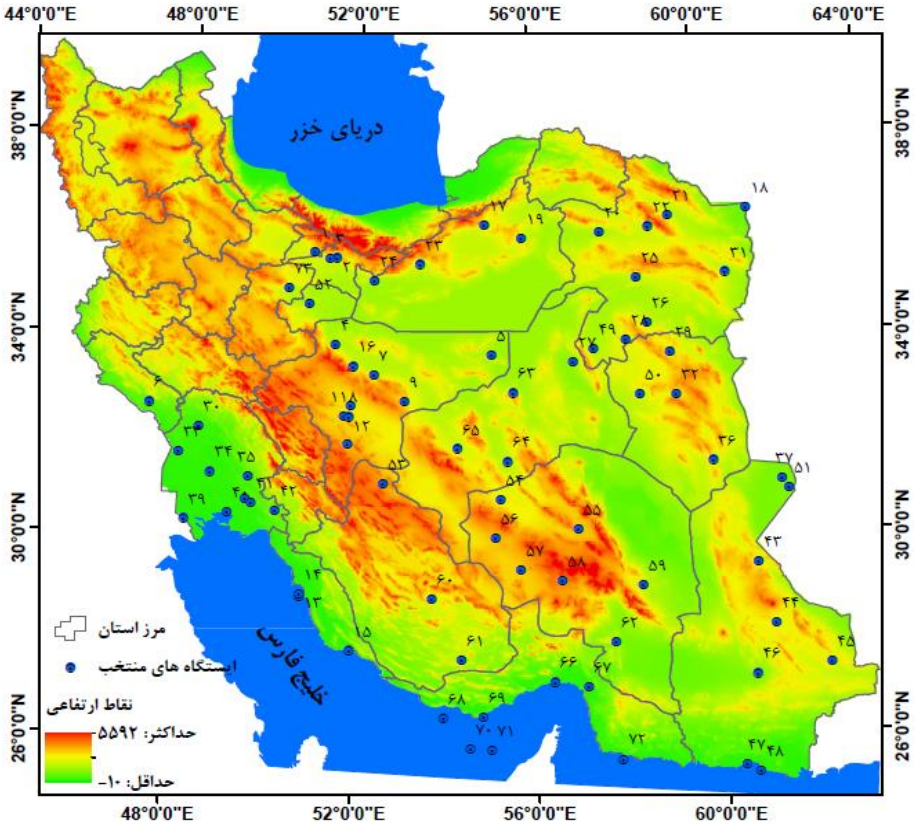
Shabalala و همکاران (۲۰۱۹) به ارزیابی پنج روش بازسازی (میانگین حسابی، روش معکوس فاصله، نسبت‌ها، نسبت نرمال و رگرسیون چندگانه) برای سری‌های زمانی روزانه حداقل و حداکثر دما پرداختند. نتایج نشان داد، رگرسیون چندگانه و نسبت‌ها از عملکرد بهتری نسبت به سایر روش‌ها در بازسازی کمینه و بیشینه دما در استان لیمپوپو آفریقای جنوبی برخوردارند.

Sadatinejad و Mahdavi (1997) روش‌های ایستگاه معرف، نسبت نرمال، محور مختصاتی، رگرسیون خطی، رگرسیون چندگانه و سری زمانی را به‌منظور بازسازی داده‌های بارش در اقلیم‌های متفاوت استان اصفهان مورد بررسی قرار دادند. نتایج نشان‌دهنده آن است که در اقلیم خشک استان اصفهان روش نسبت نرمال، در اقلیم نیمه‌خشک روش رگرسیون چندگانه و در اقلیم مدیترانه‌ای روش نسبت نرمال بهترین بازسازی را برای داده‌های بارش سالانه ارائه نموده است. Khorshiddoost و همکاران (۲۰۱۲) با استفاده از روش‌های نزدیک‌ترین همسایه و شبکه عصبی مصنوعی بازسازی داده‌های دمای بیشینه و کمینه روزانه پنج ایستگاه هواشناسی غرب تهران را بررسی و بیان کردند که روش شبکه عصبی خطای مطلق کمتری نسبت به روش نزدیک‌ترین همسایه دارد. Rahimi و همکاران (۲۰۱۷) با استفاده روش-های مختلف اقدام به بازسازی داده‌های دمای کمینه طی دوره ۲۰۱۰-۱۹۶۵ در ایستگاه‌های ارتفاعی ایران کردند. نتایج آن‌ها نشان داد که روش تحلیل مولفه‌های اصلی، روش توزیع تجمعی احتمالات و روش شبکه عصبی مصنوعی نسبت به سایر روش‌ها از کارایی بیشتری برای بازسازی برخوردارند.

بررسی مطالعات انجام شده نشان می‌دهد، با توجه به اهمیت رفع کمبود داده‌ها در سری‌های زمانی اقلیمی، روش‌های متنوعی برای بازسازی توسعه یافته است و همچنین، عملکرد آن‌ها نیز مورد بررسی قرار گرفته است. با وجود این، هر چند استفاده از داده‌های تولیدشده به‌وسیله روش‌های مختلف بازسازی یک منبع جایگزین مناسب برای ایستگاه‌های همراه با داده گم شده می‌باشد، در عین‌حال این داده‌ها نیز با چالش‌هایی مواجه هستند. برآورد غیرمستقیم شاخص‌ترین کمبود تکنیک‌های بازسازی داده‌های اقلیمی است که به‌ویژه در مناطق خشک و نیمه‌خشک جهان مانند ایران که از تیپ‌های اقلیمی متفاوتی برخوردار است، از اهمیت زیادی برخوردار می‌باشد. بنابراین، قبل از اینکه داده‌ها و نتایج این روش‌ها به‌عنوان ابزار تصمیم‌گیری در یک منطقه به‌کار رود، باید داده‌های حاصل از روش‌های مختلف مورد ارزیابی قرار گیرد. از این‌رو، در این تحقیق سعی بر این است که ضمن معرفی روش‌های مختلف برآورد داده‌های گم‌شده، با استفاده از قابلیت‌های برنامه‌نویسی، حجم قابل توجهی از داده‌های کمینه، متوسط و بیشینه دما مورد تحلیل قرار گیرد. به‌طوری‌که از پنج تا ۵۰ درصد سری‌های زمانی روزانه، ماهانه و سالانه دما به‌صورت ساختگی از سری داده‌ها حذف شد و سپس، عملکرد روش‌های مذکور در بازسازی این داده‌ها مورد ارزیابی قرار گرفته است. نتایج حاصل از این مطالعه می‌تواند به‌عنوان یک راهنمای پایه، در اختیار محققان و کارشناسان مرتبط با داده‌های هواشناسی به‌خصوص اقلیم‌شناسان و هیدرولوژیست‌ها قرار گیرد و به درک بهتر شرایط آب و هوایی در درازمدت منجر شود.

مواد و روش‌ها

داده‌ها: در این پژوهش، از داده‌های روزانه متوسط بارش، کمینه دما، متوسط و بیشینه دمای ۴۰۰ ایستگاه سینوپتیک کشور استفاده شده است. در ابتدا، با مراجعه به سازمان هواشناسی کل کشور داده‌های دما و بارش ۴۰۰ ایستگاه دریافت و بر اساس این داده‌ها، ضریب دومارتن هر یک از ایستگاه‌ها محاسبه شد. در ادامه، ایستگاه‌های دارای ضریب دومارتن کمتر از ۱۰ به‌عنوان مناطق دارای اقلیم خشک انتخاب شدند. از بین ۴۰۰ ایستگاه، تعداد ۷۳ ایستگاهی که از سال ۱۳۶۵ تا ۱۳۹۶ دارای آمار کافی و از پراکنش مناسب در پهنه خشک ایران‌زمین (شکل ۱) برخوردار بودند، برای بررسی کارایی روش‌های بازسازی انتخاب شدند.



شکل ۱- پراکنش مکانی ایستگاه‌های مورد استفاده در تحقیق حاضر (مشخصات ایستگاه‌ها بر حسب شماره ردیف در جدول ۱ ارائه شده است)

جدول ۱- مشخصات ایستگاه‌های مورد استفاده در مناطق خشک ایران

ردیف	نام ایستگاه	متوسط بارش	متوسط دما	شاخص خشکی	ردیف	نام ایستگاه	متوسط بارش	متوسط دما	شاخص خشکی
۱	کرج	۲۴۶/۸۸	۱۵/۳۳	۹/۷۴	۳۸	امیدیه (پایگاهی)	۲۲۷/۲۴	۲۵/۷۵	۶/۳۶
۲	دوشان‌تپه	۲۵۰/۱۸	۱۸/۱۲	۸/۹۰	۳۹	آبادان	۱۶۲/۱۴	۲۵/۸۳	۴/۵۲
۳	تهران مهرآباد	۲۳۰/۳۹	۱۷/۷۲	۸/۳۱	۴۰	بندر ماهشهر	۱۹۲/۵۲	۲۵/۹۲	۵/۳۶
۴	کاشان	۱۴۵/۹۵	۱۹/۲۴	۴/۹۹	۴۱	امیدیه (آغاجاری)	۲۵۴/۱۷	۲۶/۰۹	۷/۰۴
۵	خور و بیابانک	۷۶/۵۹	۲۰/۷۰	۲/۴۹	۴۲	بهبهان	۳۲۳/۲۷	۲۴/۹۸	۹/۲۴
۶	دهلران	۲۸۲/۸۵	۲۶/۲۳	۷/۸۱	۴۳	زاهدان	۸۰/۱۹	۱۸/۶۳	۲/۸۰
۷	اردستان	۱۲۸/۹۳	۱۹/۱۷	۴/۴۲	۴۴	خاش	۱۴۲/۵۴	۲۰/۲۶	۴/۷۱
۸	اصفهان (ازن‌سنجی)	۱۲۲/۳۲	۱۶/۵۷	۴/۶۰	۴۵	سراوان	۱۰۶/۰۲	۲۲/۲۸	۳/۲۸
۹	نائین	۹۷/۷۱	۱۶/۸۵	۳/۶۴	۴۶	ایرانشهر	۱۰۷/۵۵	۲۶/۹۵	۲/۹۱
۱۰	اصفهان (فرودگاه)	۱۰۲/۱۹	۱۵/۱۵	۴/۰۶	۴۷	کنارک (فرودگاه)	۱۰۷/۵۸	۲۶/۷۰	۲/۹۳
۱۱	کبوترآباد	۱۱۳/۸۲	۱۵/۶۲	۴/۴۴	۴۸	چابهار	۱۱۶/۴۷	۲۶/۳۴	۳/۲۰
۱۲	شهرضا	۱۳۸/۹۹	۱۴/۷۵	۵/۶۲	۴۹	بشرویه	۸۵/۷۹	۱۹/۴۲	۲/۹۲
۱۳	بوشهر (ساحلی)	۲۵۳/۹۳	۲۵/۲۶	۷/۲۰	۵۰	خور بیرجند	۸۴/۱۹	۱۹/۹۴	۲/۸۱
۱۴	بوشهر (فرودگاه)	۲۳۷/۳۴	۲۴/۹۰	۶/۸۰	۵۱	زهک	۴۸/۴۹	۲۳/۰۵	۱/۴۷

ادامه جدول ۲- مشخصات ایستگاه‌های مورد استفاده در مناطق خشک ایران

ردیف	نام ایستگاه	متوسط بارش	متوسط دما	شاخص خشکی	ردیف	نام ایستگاه	متوسط بارش	متوسط دما	شاخص خشکی
۱۵	بندر دیر	۲۳۱/۵۴	۲۷/۴۴	۶/۱۸	۵۲	قم	۱۳۹/۱۳	۱۸/۴۱	۴/۹۰
۱۶	نطنز	۱۸۴/۸۹	۱۵/۸۵	۷/۱۵	۵۳	آباده	۱۳۴/۰۳	۱۴/۴۶	۵/۴۸
۱۷	شاهرود	۱۶۶/۶۱	۱۴/۹۲	۶/۶۹	۵۴	انار	۷۰/۱۹	۱۸/۴۶	۲/۴۷
۱۸	سرخس	۱۸۸/۱۵	۱۸/۲۸	۶/۶۵	۵۵	کرمان	۱۳۶/۳۲	۱۵/۹۳	۵/۲۶
۱۹	بیارجمند	۱۲۰/۱۱	۱۶/۳۰	۴/۵۷	۵۶	شهر بابک	۱۴۸/۰۱	۱۵/۴۲	۵/۸۲
۲۰	سبزوار	۱۹۶/۳۰	۱۷/۷۹	۷/۰۶	۵۷	سیرجان	۱۳۶/۹۷	۱۷/۴۴	۴/۹۹
۲۱	گلمکان	۲۱۴/۱۶	۱۳/۷۶	۹/۰۲	۵۸	بافت	۲۵۰/۹۲	۱۵/۲۴	۹/۹۴
۲۲	نیشابور	۲۳۴/۷۵	۱۴/۵۴	۹/۵۷	۵۹	بم	۵۸/۹۱	۲۳/۳۸	۱/۷۶
۲۳	سمنان	۱۳۷/۹۵	۱۸/۳۳	۴/۸۷	۶۰	فسا	۲۸۸/۰۶	۱۹/۴۸	۹/۷۷
۲۴	گرمسار	۱۱۲/۰۶	۱۹/۰۹	۳/۸۵	۶۱	لار	۲۰۴/۸۴	۲۳/۸۷	۶/۰۵
۲۵	کاشمر	۱۹۳/۲۹	۱۷/۹۹	۶/۹۱	۶۲	کهنوج	۱۸۲/۸۶	۲۶/۹۴	۴/۹۵
۲۶	گناباد	۱۳۱/۴۸	۱۷/۵۹	۴/۷۶	۶۳	رباط پشت بادام	۱۰۵/۶۴	۱۹/۲۲	۳/۶۱
۲۷	طبس	۷۹/۶۵	۲۲/۲۱	۲/۴۷	۶۴	بافق	۴۶/۸۶	۲۱/۶۵	۱/۴۸
۲۸	فردوس	۱۳۱/۷۰	۱۷/۵۷	۴/۷۸	۶۵	یزد	۶۷/۵۲	۱۹/۵۷	۲/۲۸
۲۹	قائن	۱۵۸/۹۵	۱۴/۷۸	۶/۴۱	۶۶	بندرعباس	۱۶۷/۹۸	۲۷/۰۳	۴/۵۴
۳۰	صفی‌آباد (دزفول)	۳۱۷/۵۹	۲۴/۲۶	۹/۲۷	۶۷	میناب	۱۹۸/۴۵	۲۷/۵۴	۵/۲۹
۳۱	تربت‌جام	۱۶۸/۹۸	۱۵/۸۹	۶/۵۳	۶۸	جزیره کیش	۱۵۳/۸۸	۲۷/۳۱	۴/۱۲
۳۲	بیرجند	۱۷۳/۵۷	۱۶/۵۶	۶/۵۳	۶۹	بندرلنگه	۱۳۲/۷۶	۲۶/۸۷	۳/۶۰
۳۳	بستان	۱۹۵/۱۰	۲۴/۶۶	۵/۶۳	۷۰	جزیره سیری	۱۰۴/۴۴	۲۷/۹۰	۲/۷۶
۳۴	اهواز	۲۱۶/۵۴	۲۵/۷۰	۶/۰۷	۷۱	جزیره ابوموسی	۱۱۸/۵۲	۲۷/۶۳	۳/۱۵
۳۵	رامهرمز	۲۸۵/۱۴	۲۶/۶۷	۷/۷۸	۷۲	جاسک	۱۲۵/۴۸	۲۷/۰۸	۳/۳۸
۳۶	نهبندان	۱۱۲/۷۵	۲۰/۱۳	۳/۷۴	۷۳	ساوه	۱۹۰/۴۵	۱۸/۵۴	۶/۶۷
۳۷	زابل	۵۴/۰۸	۲۲/۳۹	۱/۶۷		میانگین کل	۱۵۹/۳۷	۲۰/۸۲	۵/۲۵

روش تحقیق: روش‌های مرسوم بازسازی مورد استفاده در این پژوهش، شامل روش‌های نسبت نرمال، رگرسیون خطی، رگرسیون چندگانه و عکس مجذور فاصله (IDW) می‌شوند. برای ارزیابی هر یک از روش‌های بازسازی داده، بخشی از داده‌های واقعی به دلخواه از سری داده‌های اصلی کنار گذاشته می‌شوند و سپس، مورد بازسازی واقع می‌شوند. از این‌رو، در این تحقیق آن دسته از سری‌های زمانی مورد بررسی قرار گرفتند که کمینه کمبود آماری را دارا بودند.

روش نسبت نرمال: در این روش، ایستگاه‌های هم‌جواری که از لحاظ میانگین پارامتر دما به ایستگاه دارای کمبود آماری نزدیک‌تر بودند، وزن بیشتری در بازسازی داده کمبود داشت. محاسبه این روش با رابطه (۱) صورت گرفت.

$$\frac{Px}{Nx}=\frac{1}{N}\sum_{i=1}^{N}\frac{Pi}{Ni} \qquad (۱)$$

که در آن، Px مقدار کمبود آماری دما است که بایستی تخمین زده شود، Nx میانگین دما در ایستگاهی است که دارای کمبود آماری است، یعنی ایستگاه هدف، N تعداد ایستگاه‌های همجوار یا ایستگاه شاهد که در بازسازی کمبود آماری شرکت می‌کنند، Pi مقدار دما در هر یک از ایستگاه‌های همجوار و همزمان با کمبود آماری Px و Ni مقدار میانگین دما در ایستگاه‌های هم‌جوار است.

روش رگرسیون خطی: در این روش، ابتدا همبستگی بین ایستگاهی که دارای کمبود آماری با سایر ایستگاه‌ها بود، از طریق رابطه (۲) تعیین شد.

$$r=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\left[\sum x^2-\frac{(\sum x)^2}{n}\right]\left[\sum y^2-\frac{(\sum y)^2}{n}\right]}} \quad (۲)$$

که در آن، x و y سری‌های زمانی داده‌های دما در ایستگاه دارای کمبود آماری و ایستگاه شاهد هستند، از آنجایی‌که ماتریس همبستگی به‌صورت قرینه است، هر کدام از سری‌های زمانی در ایستگاه شاهد و هدف می‌توانند به‌عنوان x و y معرفی شوند، r میزان همبستگی بین ایستگاه را نشان می‌دهد که مقادیر آن بین ۱- تا ۱ متغیر است، n طول سری زمانی داده‌ها در ایستگاه شاهد یا هدف است. چنانچه ارتباط بین سری زمانی ایستگاه دارای آمار کمبود با برخی ایستگاه‌ها از لحاظ آماری معنی‌دار بود، ایستگاهی به عنوان شاهد انتخاب شد که قدر مطلق مقدار همبستگی آن بیشینه باشد. سپس، از طریق برقراری یک رابطه خطی بین داده‌های ایستگاه شاهد و هدف، مقدار دما در ایستگاه دارای کمبود آماری محاسبه شد (رابطه ۳).

$$y=a+bx \quad (۳)$$

که در آن، y دما در ایستگاه دارای کمبود آماری و x مقدار دما در ایستگاه شاهد و در زمانی است که کمبود آماری در ایستگاه هدف وجود دارد، a و b به‌ترتیب عرض از مبدا و شیب خط در معادله درجه یک هستند.

روش رگرسیون چند متغیره: این روش مشابه روش رگرسیون خطی است. البته، معمولاً بیش از یک ایستگاه شاهد در بازسازی داده‌ها مشارکت می‌کند. در پژوهش حاضر، هر ایستگاهی که ضریب همبستگی آن از لحاظ آماری و در حد پنج درصد با ایستگاه هدف معنی‌دار بود، به‌عنوان ایستگاه شاهد در نظر گرفته شد.

روش عکس مجذور فاصله (IDW): در این روش، میزان نزدیکی بین ایستگاه‌ها، عامل وزن و مشارکت ایستگاه‌های شاهد در بازسازی کمبود داده ایستگاه هدف است. ابتدا، وزن هر یک از ایستگاه‌ها با استفاده از رابطه (۴) محاسبه شده است.

$$W=\frac{1}{X^2+Y^2} \quad (۴)$$

که در آن، X و Y به‌ترتیب فاصله طولی و عرضی بین هر ایستگاه شاهد و ایستگاه هدف است. بدیهی است، هر چه فاصله کمتر باشد، میزان W که وزن آن ایستگاه در بازسازی نبودهای آماری ایستگاه هدف است، بالاتر می‌رود. سپس، بر اساس رابطه (۵)، مقدار دما در ایستگاه هدف P_X و بر اساس وزن هر ایستگاه شاهد و دما آن در تاریخی که کمبود آماری وجود دارد (P_i)، تخمین زده می‌شود.

$$P_X=\frac{\sum W_i.P_i}{\sum W_i} \quad (۵)$$

ارزیابی روش‌ها: از شاخص‌های ضریب همبستگی R (رابطه ۶)، یا معیار برآورد خطای RMSE (رابطه ۷) و ضریب Nash (رابطه ۸) برای ارزیابی دقت داده‌ها استفاده شد.

$$R=\frac{\sum PO-\frac{\sum P\sum O}{n}}{\sqrt{\left[\sum P^2-\frac{(\sum P)^2}{n}\right]\left[\sum O^2-\frac{(\sum O)^2}{n}\right]}} \quad (۶)$$

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}\left(P_i-O_i\right)^2}{n}} \quad (۷)$$

$$Nash=\frac{\sum_{i=1}^{n}\left(P_i-O_i\right)^2}{\sum_{i=1}^{n}\left(O_i-O_{Avr}\right)^2} \quad (۸)$$

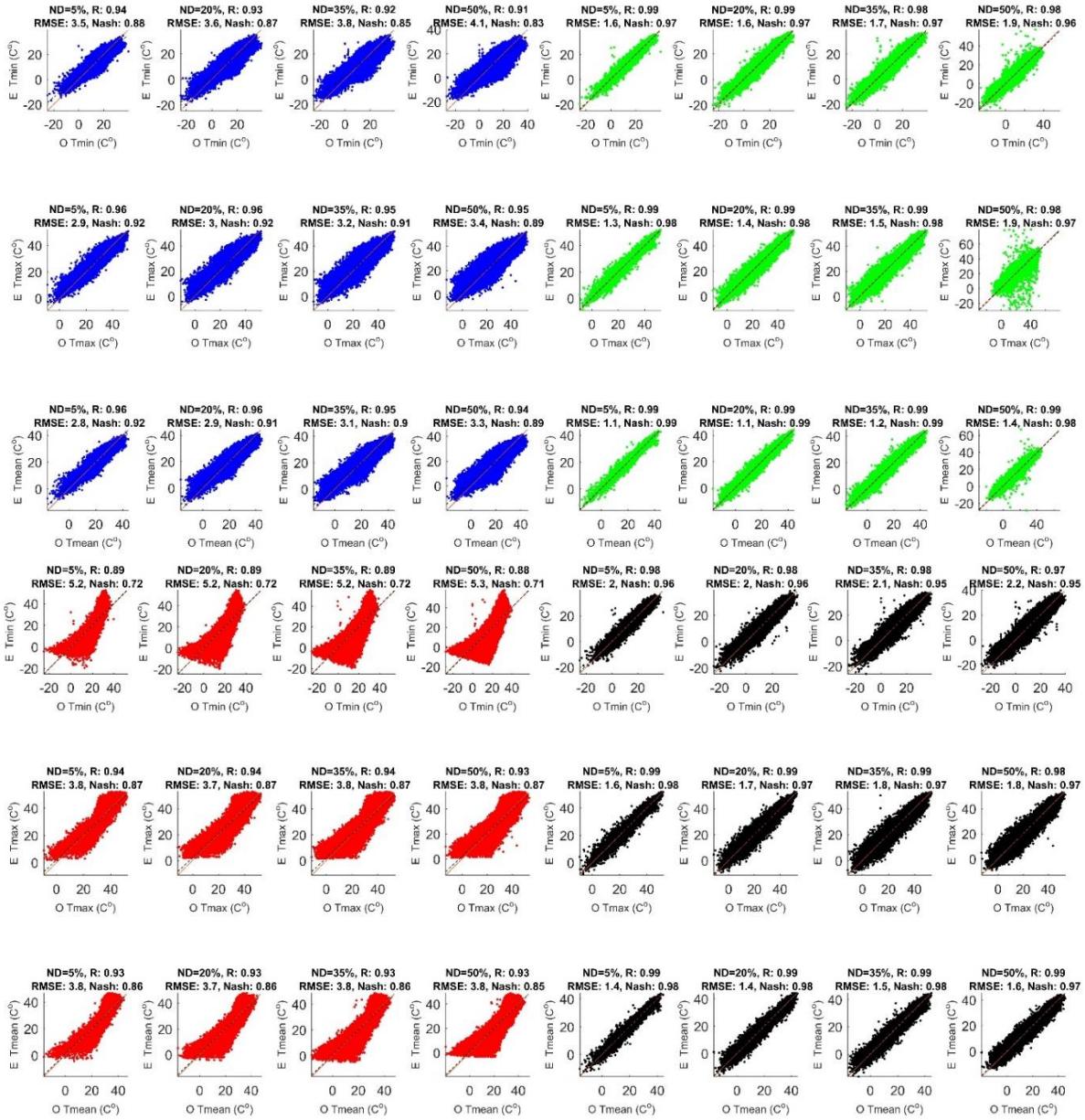
که در آن‌ها، P و O به‌ترتیب مقادیر پیش‌بینی‌شده (Predicted) و واقعی یا مشاهده‌شده (Observed) هستند. O_{Avr} میانگین داده‌های مشاهداتی، n تعداد نمونه‌ها است. همان‌طور که قبلاً نیز بدان اشاره شد، برای ارزیابی روش‌های فوق‌الذکر، به‌صورت تعمدی، بخشی از داده‌های واقعی به داده کمبود (No data) تبدیل شدند. انتخاب آن‌ها به‌صورت کاملاً تصادفی انجام شد. ایجاد داده‌های کمبود به‌گونه‌ای صورت گرفت که طی ۱۰ مرحله، از پنج تا ۵۰ درصد از داده‌های موجود به‌صورت داده کمبود درآمدند تا توان هر یک از روش‌ها در بازسازی داده‌ها متناسب با مقدار داده‌های کمبود نیز تعیین شود. در شرایط واقعی، گاهاً داده‌ها، دارای نسبت قابل‌توجهی از کمبود هستند و لذا، این ارزیابی به کارشناسان و محققان کمک می‌کند

تا کارایی روش را متناسب با حجم داده‌های کمبود نیز ارزیابی کنند. لازم به ذکر است که تمامی محاسبات مربوط به بازسازی نبودهای آماری به سه روش مذکور در این تحقیق، با کمک برنامه‌نویسی و در محیط نرم‌افزار MATLAB انجام شد.

نتایج و بحث

شکل ۲، نتایج ارزیابی دقت داده‌های دمای بازسازی‌شده در مقابل داده‌های دمای مشاهده‌ای را در مقیاس روزانه نشان می‌دهد. هر شکل مربوط به یک روش مورد بررسی برای پارامترهای کمینه، متوسط و بیشینه دما است که در بر گیرنده ۱۲ نمودار می‌باشد. نمودارهای ارائه شده در هر شکل متناسب با درصد میزان کمبود داده‌ها (از پنج تا ۵۰ درصد)، پراکنش مقادیر بازسازی‌شده در مقابل مقادیر واقعی را نشان می‌دهند. در عنوان هر یک از نمودارها ND به معنای No Data است و مربوط به همین درصد کمبود شده از داده‌های واقعی است. همچنین، در عنوان هر نمودار، مقدار R و RMSE و ضریب Nash نیز نمایش داده شده است. در محور xها مقادیر واقعی قرار دارند و محور yها مربوط به داده‌های بازسازی‌شده است. بر اساس شکل ۲، در حالت کلی در مقیاس روزانه روش‌های رگرسیون خطی و رگرسیون چندگانه از عملکرد مناسبی در مناطق خشک ایران برخوردارند. مقدار همبستگی بین داده‌های بازسازی‌شده و مشاهده‌ای برای هر سه متغیر دمایی با استفاده از این روش‌ها در سطوح مختلف بازسازی به ۰/۹۹ می‌رسد. همچنین، مقدار ضریب Nash نیز به بیش از ۰/۹۵ می‌رسد که این مقدار از ضریب Nash بیان‌کننده عملکرد مناسب روش‌های بازسازی رگرسیون خطی و رگرسیون چندگانه در بازسازی پارامترهای دمایی مناطق خشک ایران است. از نظر نوع پارامتر بیشترین دقت روش‌های یاد شده برای پارامتر بیشینه دما و کمینه آن برای کمینه دما است. این شرایط به‌دلیل تغییرات بیشتر کمینه دما نسبت به بیشینه دما در مناطق خشک ایران است. در مقیاس زمانی روزانه روش نسبت نرمال نسبت به سایر روش‌ها از دقت کمتری در بازسازی پارامترهای دمایی برخوردار است. هر چند در این روش نیز میزان همبستگی بین داده‌های دمایی بازسازی‌شده با داده‌های مشاهده‌ای برای کمینه دما ۰/۸۸ برای متوسط و بیشینه دما ۰/۹۴ است. با وجود این، بالا بودن مقادیر RMSE و زاویه‌دار شدن پراکنش داده‌ها روی خط برازش نشان‌دهنده بیش برآوردی و کم برآوردی این روش در بازسازی داده‌های دمایی مناطق خشک ایران است. روش IDW نیز از دقت مناسبی برخوردار است و مقدار همبستگی بین مقادیر بازسازی‌شده و داده‌های مشاهده‌ای در سطوح مختلف بازسازی برای کمینه دما بین ۰/۹۱ تا ۰/۹۴ برای متوسط و بیشینه دما ۰/۹۵ تا ۰/۹۶ می‌رسد که این مقدار از ضریب همبستگی بیان‌کننده دقت مناسب این روش در بازسازی داده‌های دمایی می‌باشد.

شکل ۳ نتایج ارزیابی دقت داده‌های دمای بازسازی‌شده در مقابل داده‌های دمای مشاهده‌ای را در مقیاس ماهانه نشان می‌دهد. بر اساس نمودارهای پراکنش نمایش داده شده در شکل ۳، در مقیاس ماهانه روش رگرسیون خطی بیشترین دقت و کارایی در بازسازی داده‌های دمایی در مناطق خشک ایران را دارد. در این روش، مقدار همبستگی بین داده‌های بازسازی‌شده و مشاهده‌ای در بهترین حالت برای بیشینه دما در سطح پنج درصد کمبود داده‌های بیشینه دما به یک و کمترین مقدار آن در سطح ۵۰ درصد کمبود داده‌های متوسط دما ۰/۹۷ است. این مقدار از همبستگی نشان‌دهنده دقت بالای روش رگرسیون خطی در بازسازی پارامترهای دمایی مناطق خشک ایران است. همچنین، پراکنش منظم داده در راستای خط یک و یک نیز بیان‌کننده دقت بالای این روش است. از طرف دیگر، مقادیر آماره RMSE نیز برای هر سه پارامتر دمایی نزدیک به صفر و مقادیر آماره Nash برای کمینه، متوسط و بیشینه دما بیش از ۰/۹۵ است که این مقدار از آماره‌های ذکرشده تأییدکننده دقت و کارایی بالایی روش رگرسیون خطی در بازسازی داده‌های دمایی مناطق خشک در سطوح مختلف کمبود داده‌ها می‌باشد.

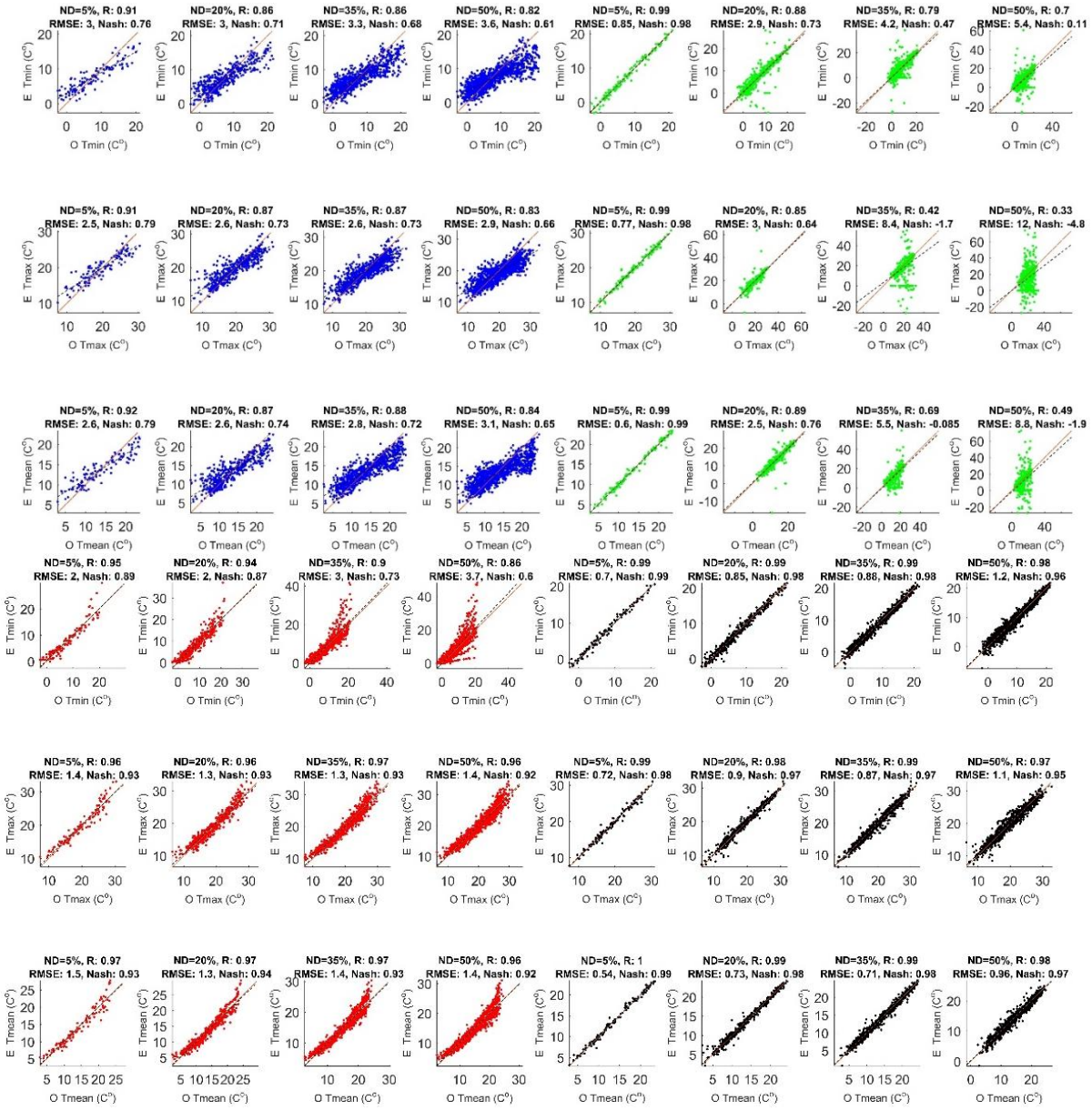


شکل ۲- مقادیر بازسازی‌شده داده‌های دمای روزانه با استفاده از روش‌های رگرسیون چندگانه (سبز)، مجذرو عکس فاصله (آبی)، رگرسیون خطی (مشکی) و نسبت نرمال (قرمز) در مقابل مقادیر واقعی دمای روزانه در حالتی که مقادیر داده‌های گم‌شده به‌صورت ساختگی از پنج تا ۵۰ درصد داده‌ها را شامل می‌شود (در عنوان هر شکل ND به معنای No Data و درصد آن نیز نمایش داده شده است)

روش نسبت نرمال نیز با اختلاف کمی نسبت به روش رگرسیون خطی در بازسازی داده‌های دمایی مناطق خشک ایران از دقت بالایی برخوردار است. در این روش نیز مقدار بالای ضریب همبستگی و ضریب Nashو مقدار پایین آماره RMSE بیان‌کننده دقت مناسب این روش برای بازسازی داده‌های دمایی در سطوح مختلف کمبود داده‌ها هستند. بررسی مقادیر آماره‌های مورد استفاده بین داده‌های بازسازی‌شده و مشاهده‌ای بر اساس روش معکوس فاصله نشان می‌دهد، هر چند در حالت کلی این روش نیز از دقت قابل قبولی برخوردار است، با وجود این با توجه به مقادیر نسبتاً بالای RMSE و همچنین، ایجاد زاویه بین خط یک و یک نمودار پراکنش داده‌ها با خط وایازش، این روش در بازسازی داده‌های دمایی با بیش برآوردی و کم برآوردی همراه است.

در بین روش‌های یادشده روش رگرسیون چندگانه نسبت به سه روش دیگر از دقت کمتری برخوردار است و در سطوح بالای کمبود داده ضریب همبستگی

بین داده‌های بازسازی‌شده و مشاهده‌ای به کمتر از ۰/۵ می‌رسد. همان‌طور که روی نمودارهای شکل ۳ مشخص است، مقدار آماره همبستگی بین داده‌های بازسازی‌شده و مشاهده‌ای در سطح کمبود پنج درصد همانند روش‌های رگرسیون خطی و نسبت نرمال برای دمای کمینه، متوسط و بیشینه دما بین ۰/۹۸ تا ۰/۹۹ است و در سطح کمبود داده ۱۰ درصد مقدار همبستگی ۰/۸۵ برای متوسط دما تا ۰/۸۹ برای بیشینه دما تغییر می‌کند. این مقدار از ضریب همبستگی در سطوح پنج و ۲۰ درصد کمبود داده بیان‌کننده دقت مناسب روش رگرسیون چند متغیره برای بازسازی داده‌های مفقود به‌ویژه در سطح کمبود پنج درصد می‌باشد.

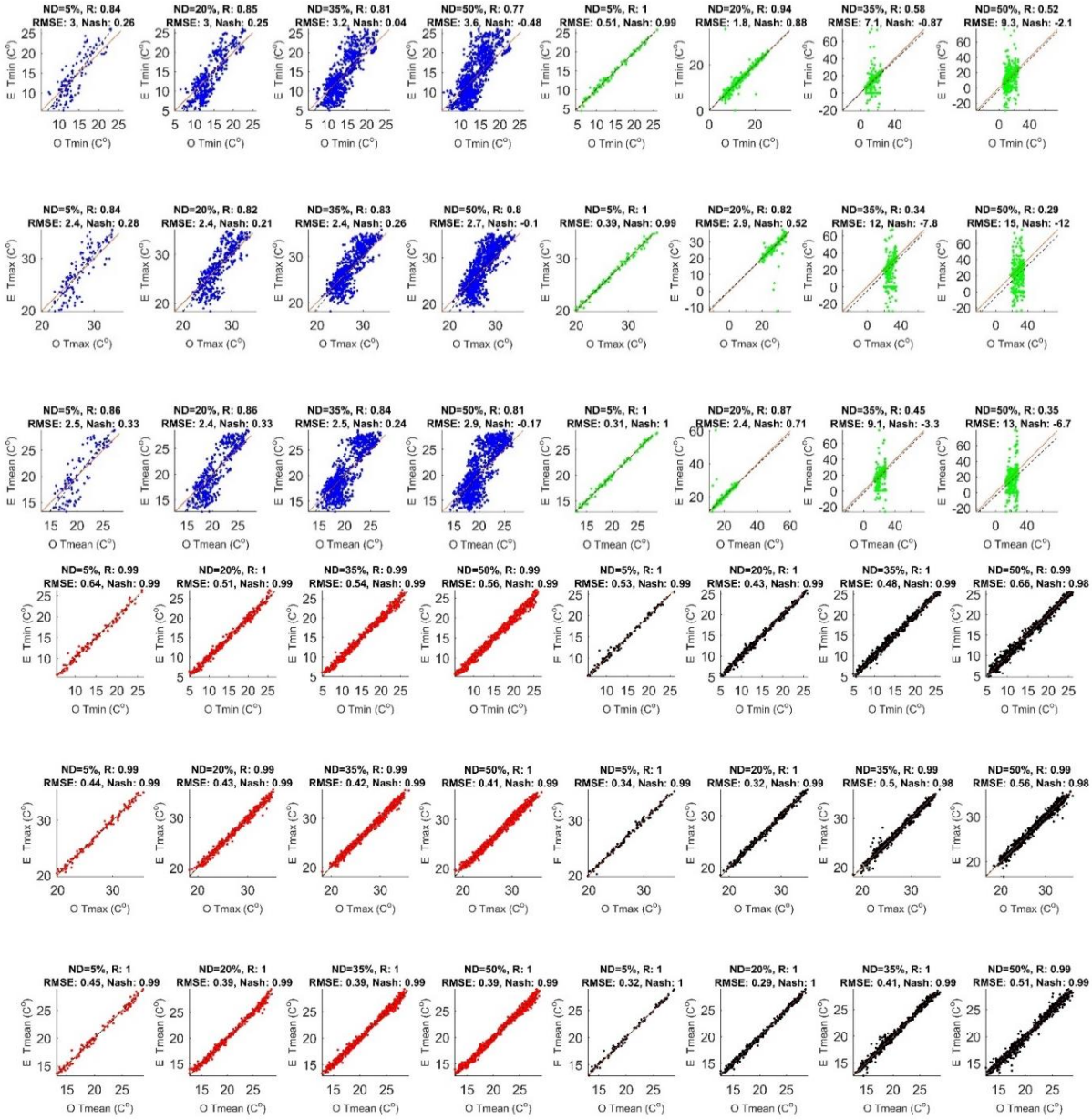


شکل ۳- مقادیر بازسازی‌شده داده‌های دمای ماهانه با استفاده از روش‌های رگرسیون چندگانه (سبز)، مجذرو عکس فاصله (آبی)، رگرسیون خطی (مشکی) و نسبت نرمال (قرمز) در مقابل مقادیر واقعی دمای ماهانه در حالتی که مقادیر داده‌های کمبود به‌صورت ساختگی از پنج تا ۵۰ درصد داده‌ها را شامل می‌شود (در عنوان هر شکل ND به معنای No Data و درصد آن نیز نمایش داده شده است)

علاوه‌بر مقادیر ضریب همبستگی، مقادیر بالای ضریب NASH و مقادیر پایین آماره RMSE نیز تأییدکننده کارایی خوب این روش در سطح کمبود پنج درصد است. در سطوح بالای کمبود داده، کاهش

مقدار آماره همبستگی به کمتر از ۰/۵ به‌ویژه برای متوسط و بیشینه دما و همچنین، مقادیر بالای RMSE و مقادیر پایین و منفی آماره NASH بین داده‌های بازسازی‌شده و مشاهده‌ای بیان‌کننده دقت پایین این روش در بازسازی داده‌های دمایی مناطق خشک ایران است. در این سطوح فاصله بین خط وایازش و خط یک و یک افزایش پیدا می‌کند و زاویه بیشتری ایجاد می‌شود و با توجه به این‌که در محل تمرکز داده‌ها خط وایازش پایین‌تر از خط یک و یک نمودار قرار می‌گیرد، بیان‌کننده کم‌برآوردی روش رگرسیون چندگانه در بازسازی داده‌های دمایی منطقه مورد مطالعه است.

شکل ۴، ارتباط بین مقادیر بازسازی‌شده حاصل از روش‌های بازسازی فوق‌الذکر و مقادیر واقعی متغیرهای دمایی را با استفاده از آماره‌های همبستگی خطی، RMSE، NASH و در مقیاس زمانی سالانه نشان می‌دهد.



شکل ۴- مقادیر بازسازی‌شده داده‌های دمای سالانه با استفاده از روش‌های رگرسیون چندگانه (سبز)، مجذرو عکس فاصله (آبی)، رگرسیون خطی (مشکی) و نسبت نرمال (قرمز) در مقابل مقادیر واقعی دمای سالانه در حالتی که مقادیر داده‌های کمبود به‌صورت ساختگی از پنج تا ۵۰ درصد داده‌ها را شامل می‌شود (در عنوان هر شکل ND به معنای No Data و درصد آن نیز نمایش داده شده است)

همان‌طور که روی نمودارهای شکل ۴ نشان داده شده است، در این مقیاس نیز همانند مقیاس ماهانه روش‌های رگرسیون خطی و نسبت نرمال از دقت بسیار بالایی برای بازسازی داده‌های دمایی در مناطق خشک ایران برخوردارند. مقادیر بالای ضریب همبستگی و ضریب NASH و در مقابل مقادیر بسیار پایین RMSE در سطوح مختلف کمبود داده برای کمینه، متوسط و بیشینه دما تأییدکننده کارایی روش‌های یاد شده در مقیاس سالانه است. در ارتباط با کارایی روش رگرسیون چندگانه، در مقیاس سالانه نیز همانند مقیاس ماهانه این روش در سطح پنج درصد کمبود داده‌های دمایی از دقت مناسبی برای بازسازی داده‌های دمایی برخوردار است، چرا که مقادیر آماره همبستگی برای هر سه پارامتر دما یک و مقادیر ضریب NASH نیز ۰/۹۹ و یک و مقادیر آماره RMSE کمتر از ۰/۵ و نزدیک به صفر است. در سطوح بالای کمبود داده ۳۵ درصد و به‌ویژه ۵۰ درصد همانند مقیاس ماهانه دقت روش رگرسیون چندگانه در مقیاس سالانه کاهش پیدا می‌کند. در سطح ۵۰ درصد مقدار آماره همبستگی بین داده‌های بازسازی‌شده و مشاهده‌ای برای کمینه دما به ۰/۵۲، برای متوسط دما ۰/۲۹ و برای بیشینه دما ۰/۳۵ می‌رسد که این مقدار از همبستگی نشان‌دهنده کارایی نامناسب و ضعیف این روش در مقیاس سالانه است. همچنین، مقادیر منفی ضریب NASH و مقادیر بالای آماره RMSE بین داده‌های بازسازی و مشاهده‌ای تأییدکننده نتایج آماره همبستگی است.

نتیجه‌گیری

در این پژوهش، دقت چهار روش مرسوم رگرسیون خطی، رگرسیون چندگانه، نسبت نرمال و عکس مجذور فاصله برای بازسازی کمینه، متوسط و بیشینه دمای مفقودی در سری زمانی اقلیمی در مقیاس‌های زمانی روزانه، ماهانه و سالانه مناطق خشک ایران با استفاده از آماره‌های R، RMSE و NASH مورد بررسی قرار گرفت. نتایج نشان داد که در حالت کلی، بازسازی دمای بیشینه مناطق خشک ایران با استفاده از روش‌های یاد شده نتایج بهتری نسبت به کمینه و متوسط دما دارد. یکی از دلایل اصلی آن، نوسان کمتر دمای بیشینه نسبت به متوسط و کمینه دما بین ایستگاه‌ها و قسمت‌های مختلف مناطق خشک ایران است. همچنین، مشخص شد که در مجموع از بین چهار روش ذکر شده، روش رگرسیون خطی برای برآورد پارامترهای دمایی مناطق خشک ایران از کارایی و عملکرد بهتری نسبت به سایر روش‌ها برخوردار است. یکی از دلایل اصلی بالا بودن میزان همبستگی در روش رگرسیون، در نظر گرفتن تأثیر تغییرات ارتفاع در افزایش و کاهش مقادیر دمایی در این روش است، در حالی‌که در روش‌های مانند IDW این مهم کمتر مورد توجه قرار می‌گیرد. در بسیاری از مطالعات انجام‌شده (Alijani، ۲۰۰۶؛ Masoudian، ۲۰۱۱) تأثیر مستقیم افزایش و کاهش ارتفاع به‌ترتیب بر کاهش و افزایش دما در مناطق مختلف ایران مورد تأیید قرار گرفته است. بررسی کارایی روش‌های یاد شده در مقیاس‌های زمانی و سطوح مختلف کمبود داده نشان داد که در مقیاس روزانه هرچند بین داده‌های بازی شده بر اساس روش‌های مورد استفاده و داده‌های مشاهده‌ای ارتباط بالایی وجود دارد و عملکرد همه روش‌ها قابل قبول است، اما روش‌های مبتنی بر رگرسیون (رگرسیون خطی و چندگانه) از عملکرد بالایی در بازسازی نبودهای آماری پارامترهای دمایی مناطق خشک ایران برخوردارند، چرا که در روش‌های مبتنی بر رگرسیون علاوه‌بر بالا بودن مقادیر آماره همبستگی، پراکنش داده‌ها در راستای خط یک و یک نمودار انجام شده است و بین خط وایازش و خط یک و یک نیز فاصله و یا زاویه خاصی قابل مشاهده نیست. در حالی‌که در روش نسبت نرمال و به‌ویژه مجذور عکس فاصله با وجود مقادیر همبستگی بالا، مقادیر بالای RMSE و قرارگیری خط وایازش در پایین و بالای خط یک و یک نمودار پراکنش داده‌ها بیانگر عدم دقت بالای این روش‌ها در برآورد داده‌های روزانه مناطق خشک است. در پژوهش انجام‌شده به‌وسیله Kashini و Dinpashoh (۲۰۱۲) نشان داده شد که روش رگرسیون در بازسازی دماهای کمینه و بیشینه ایران عملکرد بهتری نسبت به سایر روش‌ها دارد. همچنین، نتایج تحقیقات Xia و همکاران (۱۹۹۹) و Shabalala و همکاران (۲۰۱۹) بیانگر عملکرد بهتر روش رگرسیون در

بازسازی داده‌های روزانه دما نسبت به سایر روش‌ها است.

نتایج بررسی کارایی روش‌های بازسازی مورد استفاده در مقیاس‌های زمانی ماهانه و سالانه نشان داد که در هر دو مقیاس زمانی به‌ترتیب روش‌های رگرسیون خطی و نسبت نرمال از عملکرد بهتری نسبت به روش‌های دیگر در بازسازی داده‌های دمایی منطقه مورد مطالعه برخوردارند. در این مقیاس‌ها برخلاف مقیاس روزانه، روش رگرسیون چندگانه نسبت به سایر روش‌ها از عملکرد قابل قبولی به‌ویژه در سطوح بالای کمبود داده (۳۵ و ۵۰ درصد) برخوردار نیست. این شرایط می‌تواند به‌دلیل حجم نمونه‌های تعلیمی در اجرای روش رگرسیون چندگانه باشد. چرا که در مقیاس روزانه تعداد داده‌های تعلیمی برای اجرای روش رگرسیون چند متغیره از حجم قابل قبولی برخوردار است. از این‌رو، دقت بازسازی نیز افزایش پیدا می‌کند، ولی در مقیاس سالانه و ماهانه با توجه به این‌که حجم نمونه‌های تعلیمی کاهش پیدا می‌کند، دقت روش نسبت به مقیاس روزانه کمتر می‌شود. دلیل کارایی بهتر روش نسبت نرمال در مقیاس‌های ماهانه و سالانه نسبت به مقایس روزانه به نزدیکی مقادیر میانگین داده‌ها مربوط می‌شود. در روش نسبت نرمال نزدیکی میانگین داده‌ها به یکدیگر نقش مهمی در بازسازی ایفا می‌کند، این در حالی است که در مقیاس روزانه میانگین‌های مورد استفاده ممکن است با یکدیگر نزدیکی قابل توجهی نداشته باشند. در واقع بین ردیف (تاریخ) عدد ناقص و داده‌های موجود هم‌زمان با آن عدد در ایستگاه‌های شاهد و میانگین کلی هر ایستگاه تفاوت وجود داشته باشد. در حالی‌که در مقیاس‌های ماهانه و سالانه این نزدیکی مقادیر افزایش می‌یابد که سبب افزایش کارایی روش نسبت نرمال در بازسازی داده‌های ماهانه و سالانه می‌شود.

در مقیاس ماهانه بر اساس روش رگرسیون خطی، خطای محاسبات به‌طور میانگین در کمترین حالت، برای بیشینه دما در سطح پنج درصد کمبود داده ۰/۵ سانتی‌گراد و در بیشترین حالت، برای کمینه دما و در سطح ۵۰ درصد کمبود داده ۱/۲ درجه سانتی‌گراد است. در مقیاس سالانه نیز همانند مقیاس ماهانه کمترین خطای محاسباتی برای بیشینه دما و در سطح کمبود پنج درصد داده‌ها به‌طور متوسط ۰/۳۲ درجه سانتی‌گراد و بیشترین خطای محاسباتی برای کمینه دما و در سطح ۵۰ درصد کمبود داده‌ها به‌طور متوسط ۰/۶۶ درجه سانتی‌گراد است. شایان ذکر است که این مقدار متوسط منطقه مورد مطالعه هست و به‌صورت نقطه‌ای با توجه به شرایط محیط و توپوگرافی این مقدار می‌تواند با افزایش و کاهش همراه شود. همان‌طور که نتایج فوق نشان داد، زمانی که بازسازی داده‌ها در مقیاس‌های زمانی ماهانه و به‌ویژه سالانه انجام می‌شود، دقت و کارایی روش‌های فوق افزایش بیشتری پیدا می‌کند. افزایش دقت روش‌ها در مقیاس ماهانه و سالانه از یک طرف، به‌دلیل نزدیک شدن میانگین دماهای ماهانه و سالانه ایستگاه‌های مشابه در سری‌های زمانی ماهانه و سالانه است و از طرف دیگر، به‌دلیل کاهش تأثیر رخدادهای حدی، کاهش تضاریس داده‌ها و هموارتر شدن سری‌های زمانی سالانه و ماهانه پارامترهای دمایی نسبت به مقیاس روزانه است.

در مجموع، بر اساس نتایج این پژوهش و تحقیقات مشابه انجام‌شده می‌توان گفت که استفاده از روش‌های بازسازی مناسب در مناطق متفاوت آب و هوایی به شرایط مختلفی از جمله نوع سری زمانی و نوع پارامتر مورد نظر، فصل مقادیر گم‌شده، میزان مقادیر گم‌شده در داده‌ها، منطقه آب و هوایی، دانش و تخصص شخص مسئول تصحیح داده‌ها، امکانات در دسترس خصوصاً از لحاظ سخت‌افزاری و نرم‌افزاری، توانایی و دقت روش، طول دوره داده‌برداری داده‌های موجود و ... بستگی دارد. از این‌رو، با توجه به پویا بودن فرایند بازسازی داده‌ها، روش مناسب برای بازسازی باید با توجه به شرایط موجود در هر منطقه انتخاب و حتماً عملکرد آن نیز مورد ارزیابی قرار گیرد. لذا، لازم است تا فرایند بازسازی داده‌ها خصوصاً در مطالعات علوم محیطی به‌صورت یک نرم‌افزار واحد ارائه شود تا هم فرایندهای بازسازی داده یکپارچه شود و هم علاوه‌بر کوتاه‌تر شدن این عملیات وقت‌گیر از لحاظ حجم محاسبات، مورد ارزیابی نیز قرار گرفته، در نهایت، روشی انتخاب شود که متناسب با شرایط، بهترین عملکرد را نسبت به سایر روش‌ها داراست. علاوه‌بر آن، می‌توان از این نرم‌افزار، همانند بسته

نرم‌افزاری که به‌وسیله Serano و همکاران (۲۰۱۷) و یا Hu و Huang (۲۰۲۰) در محیط نرم‌افزار R برای تولید داده‌های بارش روزانه ایجاد شده است، برای تولید نقشه‌های پارامترهای هواشناسی در نقاط فاقد داده نیز استفاده کرد. به هر حال، در قالب یک بسته نرم‌افزاری بهتر است، دقت سایر روش‌های بازسازی نبودهای آماری خصوصاً روش‌های مبتنی بر هوش مصنوعی نیز مورد بررسی قرار گیرند که در تحقیقات آتی به آن پرداخته خواهد شد.

منابع مورد استفاده

1. Aieb, A., K. Madani, M.L. Scarpa and B. Bonaccorso. 2019. A new approach for processing climate missing databases applied to daily rainfall data in Soummam Watershed, Algeria. Heliyon, 5(2):1247.
2. Alijani, B. 2006. Climate of Iran. Payam-e-Noor University Press, 125 pages (in Persian).
3. Gairola, R.M., S. Prakash and P.K. Pal. 2015. Improved rainfall estimation over the Indian monsoon region by synergistic use of Kalpana-1 and rain gauge data. Atmósfera, 28: 51-61.
4. Hasanpour Kashani, M. and Y. Dinpashoh. 2012. Evaluation of efficiency of different estimation methods for missing climatological data. Stochastic Environmental Research and Risk Assessment, 1(26): 59-71.
5. Hofstra, N., M. Haylock, M. New, P. Jones and C. Frei. 2008. Comparison of six methods for the interpolation of daily, European climate data. Journal of Geophysical Research: Atmospheres, doi.org/10.1029/2008JD010100.
6. Hu, M. and Y. Huang. 2020. Atakrig: an R package for multivariate area-to-area and area-to-point kriging predictions. Computers and Geosciences, 139: 104471.doi.org/10.1016/j.cageo.2020. 104471.
7. Kashani, M.H. and Y. Dinpashoh. 2012. Evaluation of eciency of di erent estimation methods for missing climatological data. Stochastic Environmental Research and Risk Assessment, 26: 59–71.
8. Khorshiddoust, A.M., Z.M. Nassaji and B. Ghermezcheshmeh. 2012. Time series reconstruction of daily maximum and minimum temperature using nearest neighborhood and artificial neural network techniques, case study: west of Tehran Province. Geographical Space, 12(38): 197-214 (in Persian).
9. Kim, J.W. and Y.A. Pachepsky. 2010. Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. Journal of Hydrology, 394(3): 305-314.
10. Kotsiantis, S., A. Kostoulas, S. Lykoudis, A. Argiriou and K. Menagias. 2006. Filling missing temperature values in weather data banks. 2nd IEE International Conference on Intelligent Environments, 5-6 July, 2006, Athens, Greece, 1: 327-334.
11. Masoudian, S.A. 2011. Climate of Iran, Isfahan. Isfahan University Press, 125 pages (in Persian).
12. Miri, M., T. Raziei and M. Rahimi. 2016. Evaluation and statistically comparison of TRMM and GPCC datasets with observed precipitation in Iran. Journal of the Earth and Space Physics, 42: 657–672 (in Persian).
13. Mishra, A.K., R.M. Gairola, A.K. Varma and V.K. Agarwal. 2011. Improved rainfall estimation over the Indian region using satellite infrared technique. Advances in Space Research, 48: 49–55.
14. Price, D.T., D.W. McKenney, I.A. Nalder, M.F. Hutchinson and J.L. Kesteven. 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. Agricultural and Forest Meteorology, 101(2-3): 81-94.
15. Rahimi, J., A. Khalili and J. Bazr Afshan. 2017. Evaluation of different missing data reconstruction methods for daily minimum temperature in elevated stations of Iran: comparison with new proposed approach. Iranian Journal of Soil and Water Research, 48(2): 231-239 (in Persian).
16. Rees, G. 2008. Hydrological data. In: Gustard, Alan; Demuth, Siegfried, (eds.) Manual on Low-flow Estimation and Prediction. Geneva, Switzerland. World Meteorological Organization, 22-35, 136 pages.
17. Sadatinezhad, S.J. and M. Mahdavi. 1997. Statistical comparison and different methods of rainfall data reconstruction in Isfahan Province, in natural resources. Tarbiat Modares University, 165 pages (in Persian).
18. Sattari, M.T., A. Rezazadeh-Joudi and A. Kusiak. 2016. Assessment of different methods for estimation of missing data in precipitation studies. Hydrology Research, 4(48): 1032-1044.
19. Serrano-Notivoli, R., M. de Luis and S. Beguería. 2017. An R package for daily precipitation climate series reconstruction. Environmental Modelling and Software, 89: 190-195.

20. Shabalala, Z.P., M. Moeletsi, M. Tongwane and S. Mazibuko. 2019. Evaluation of infilling methods for time series of daily temperature data: case study of Limpopo Province, South Africa. Climate, 7(7): 86-102.
21. Tardivo, G. and A. Berti. 2014. The selection of predictors in a regression-based method for gap filling in daily temperature datasets. International Journal of Climatology, 34: 1311–1317.
22. Teegavarapu, R.S. and V. Chandramouli. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Journal of Hydrology, 312(1): 191-206.
23. Wagner, P.D., P. Fiener, F. Wilken, S. Kumar and K. Schneider. 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology, 464: 388-400.
24. Xia, Y., P. Fabian, A. Stohl and M. Winterhalter. 1999. Forest climatology: estimation of missing values for Bavaria, Germany. Agricultural and Forest Meteorology, 96(1): 131-144.
25. You, J., K.G. Hubbard and S. Goddard. 2008. Comparison of methods for spatially estimating station temperatures in a quality control system. International Journal of Climatology, 28(6): 777-787.
26. Yozgatligil, C., S. Aslan, C. Iyigun and I. Batmaz. 2013. Comparison of missing value imputation methods in time series: the case of Turkish meteorological data. Theoretical and Applied Climatology, 112(1-2): 143-167.