

## ارزیابی روش‌های مختلف و پیشنهاد یک روش جدید برای بازسازی خلاءهای آماری داده‌های دمای کمینه روزانه در ایستگاه‌های ارتفاعی ایران

جابر رحیمی<sup>۱</sup>، علی خلیلی<sup>۲\*</sup>، جواد بذرافشان<sup>۳</sup>

۱. دانشجوی دکتری، گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج

۲. استاد، گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج

۳. دانشیار، گروه مهندسی آبیاری و آبادانی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج

(تاریخ دریافت: ۱۳۹۵/۴/۸ - تاریخ تصویب: ۱۳۹۵/۵/۲۳)

### چکیده

داده‌های دمای حداقل روزانه مهم‌ترین نیاز پژوهش‌های اقلیمی و همچنین مطالعات تأثیر تغییر اقلیم در زمینه سرمایه‌های زودرس پاییزه، دیررس بهاره، طول مدت سرما، ارزیابی پتانسیل تولید محصولات باغی و زراعی وابسته به سرمازدگی و نهایتاً امنیت غذایی است. از نگاه دیگر، به‌رغم آنکه ایستگاه‌های مناطق مرتفع نقش مهمی در صحت برآورد شیب ارتفاعی آماره‌های دما و میدانی کردن آنها دارند، تعدادشان در ایران بسیار محدود بوده و اغلب آنها از نظر آماری واجد گسستگی‌های زیادی هستند و بازسازی آنها برای مطالعاتی که مبتنی بر دوره‌های آماری همگن طولانی هستند، امری ضروری محسوب می‌شود. هدف از این بررسی، مقایسه روش‌های مختلف بازسازی داده‌های روزانه دمای کمینه در ایستگاه‌های مرتفع و معرفی برترین روش‌ها برای ترمیم و گسترش سری‌های زمانی مربوطه در دوره اقلیمی همزمان با ایستگاه‌های قدیمی ایران است. برای این منظور، ۱۲ ایستگاه که ارتفاعی بیشتر از ۱۹۰۰ متر داشتند، انتخاب و از دوره‌های موجود آماری آنها ۵۰۰ دمای کمینه روزانه به طور تصادفی حذف و همان داده‌ها با استفاده از ۳۱ روش کلاسیک موجود و یک روش پیشنهادی جدید مبتنی بر توزیع احتمالات وقوع دماهای کمینه بازسازی گردید و اعتبار هر یک از آنها با استفاده از آماره مجذور مربعات خطا و همچنین بررسی حدود خطا در سطح اطمینان ۹۰ و ۹۵٪ بطور کمی ارزیابی شد. بر اساس نتایج به دست آمده سه روش تحلیل مؤلفه‌های اصلی، روش پیشنهادی مبتنی بر تابع توزیع تجمعی احتمالات و روش شبکه عصبی مصنوعی با حدود اعتماد ۹۵٪ خطای بازسازی به ترتیب  $\pm 2/0$ ،  $\pm 2/2$  و  $\pm 3/1$  درجه نسبت به سایر روش‌های بازسازی ارجحیت دارند. این بررسی منجر به تکمیل سری آماری دمای کمینه روزانه این ایستگاه‌ها در دوره ۲۰۱۰-۱۹۶۵ گردید که در مطالعات تغییر اقلیم و ریسک سرمایه زودرس و دیررس و همچنین تغییرات تقویم زراعی در اثر پدیده تغییر اقلیم مورد استفاده قرار می‌گیرند.

**واژه‌های کلیدی:** بازسازی دمای روزانه، ایستگاه‌های هواشناسی مرتفع، خطای برآورد دما، ایران.

### مقدمه

تصحیح این داده‌ها همواره مورد توجه محققان باشد. از آنجایی که گسستگی آماری داده‌ها، انجام تحلیل‌های زمانی تغییرات اقلیمی را مختل می‌سازد، لازم به نظر می‌رسد که به‌منظور تهیه پایه زمانی مشترک آماری، سری‌های زمانی موردنظر با روش مناسبی تکمیل گردیده و سپس از داده‌های آماری استفاده شود. به‌طور کلی روش‌های مختلفی برای بازسازی داده‌های هواشناسی و هیدرولوژیکی وجود دارد که عبارتند از روش‌های جبری، رگرسیونی، احتمالاتی، درون‌یابی، هوش مصنوعی و ترکیبی.

طی سال‌های اخیر تحقیقات بسیاری در مورد مقایسه روش‌های بازسازی داده هواشناسی صورت گرفته است که به عنوان نمونه می‌توان به *Ashraf et al.* (1997)، *Khalil et al.* (2001)، *Teegavarapu and Chandramouli* (2005)، *Mileva*

نمود یک پایه زمانی مشترک طولانی آماری در شبکه ایستگاه‌های هواشناسی کشور که خود ناشی از تأسیس این ایستگاه‌ها در سال‌های مختلف است همواره به عنوان یکی از مسائل و معضلات عمده در روند تجزیه و تحلیل‌های سری‌های زمانی هواشناسی و مطالعات تغییر اقلیم مطرح بوده است. علاوه بر این وجود خلاءهای آماری و یا وجود ارقام مشکوک به‌ویژه در مناطقی که تراکم شبکه هواشناسی کافی نبوده و یا برای مناطقی که به علت سختی راه فاقد ایستگاه و یا فاقد داده‌های معتبر هواشناسی هستند باعث می‌گردد که مسئله تکمیل و

و همچنین طول خلأ آماری متفاوت است. در خصوص تحقیقات داخلی نیز، Khorshiddoust *et al.* (2012) مسئله بازسازی داده های دمای بیشینه و کمینه روزانه در ۵ ایستگاه هواشناسی واقع در غرب استان تهران را با استفاده از روش های نزدیک ترین همسایه و شبکه عصبی مصنوعی مورد بررسی قرار داده و نشان دادند که برای کلیه ایستگاه های مورد مطالعه روش شبکه عصبی متوسط خطای مطلق کمتری را نسبت به روش نزدیک ترین همسایه دارد. همچنین دقت هر دو روش در تخمین دمای بیشینه روزانه بیشتر از دمای کمینه روزانه بود.

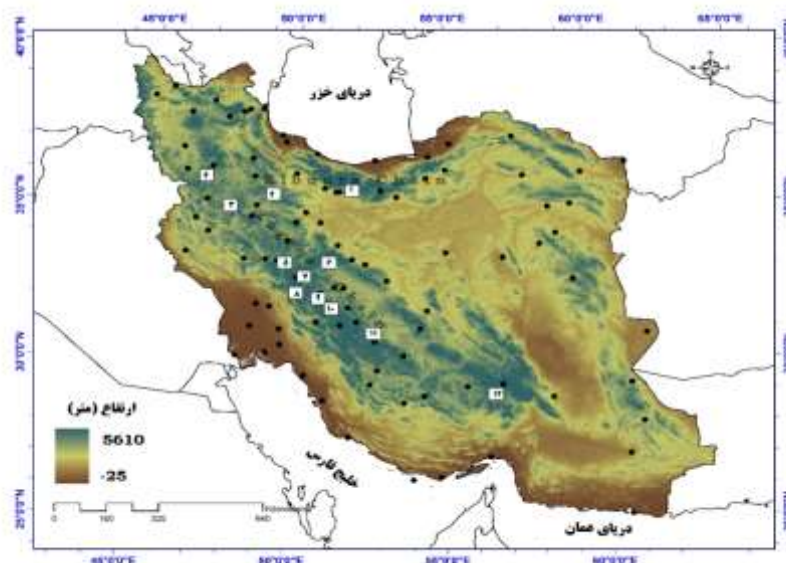
در مجموع اگرچه طی سالیان اخیر تحقیقات مختلفی برای بازسازی آماری پارامترهای هواشناسی در مقیاس زمانی ماهانه صورت گرفته است، اما همچنان جای خالی تحقیقات در زمینه بازسازی پارامترهای هواشناسی در مقیاس زمانی روزانه در کشورمان احساس می شود. لذا هدف از این بررسی بازسازی خلأهای دمای کمینه روزانه و ایجاد سری زمانی پیوسته این عامل در دوره ۲۰۱۰-۱۹۶۵ به منظور بهره گیری بعدی از آنها در مطالعات تغییر اقلیم و اقلیم شناسی رویدادهای سرمای زودرس پاییزه، دیررس بهاره، طول مدت سرما و تعیین تاریخ این رویدادها در فرآیند تغییر اقلیم است که نهایتاً در مسائل پیش نگرانی امنیت غذایی کاربرد دارند.

### مواد و روش ها

در این مطالعه از آمار دمای کمینه روزانه مناسب ترین ایستگاه های مرجع مربوط به ۱۳۰ ایستگاه سینوپتیک کشور برای بازسازی خلأهای آماری تک تک ۱۲ ایستگاه ارتفاعی منتخب استفاده شد. شکل (۱) موقعیت جغرافیایی ایستگاه های مرجع و تحت بازسازی را نشان می دهد.

برای تعیین مناسب ترین ایستگاه های مرجع برای بازسازی خلأهای آماری ایستگاه های منتخب از روش ماتریس فاصله و ماتریس ضرایب همبستگی دو به دو (Khalili, 1991) بین هریک از ۱۲ ایستگاه ارتفاعی با بقیه ایستگاه ها استفاده شد. به این مفهوم که از بین ایستگاه های مرجع، داده های گروه ایستگاه هایی که واجد همبستگی معنی دار حداقل در سطح ۵ درصد با داده های ایستگاه تحت بازسازی بودند با اولویت نزدیکی فاصله، به عنوان ورودی برای بازسازی استفاده گردید. همچنین در ارزیابی همبستگی دو بدو، میزان همبستگی با داده های یک و دو روز قبل به عنوان پیش بینی کننده محتمل نیز آزمون شد. جدول (۱) روش های بازسازی و تکمیل داده های اقلیمی استفاده شده در این بررسی را نشان می دهد.

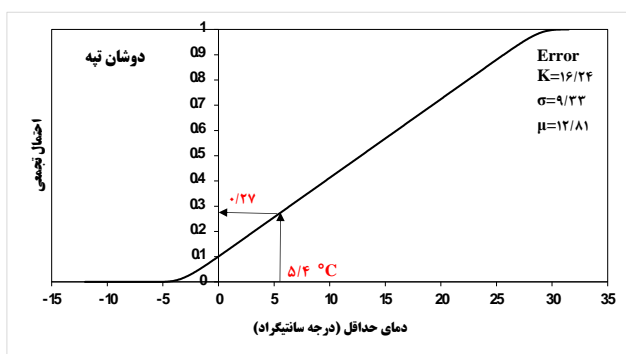
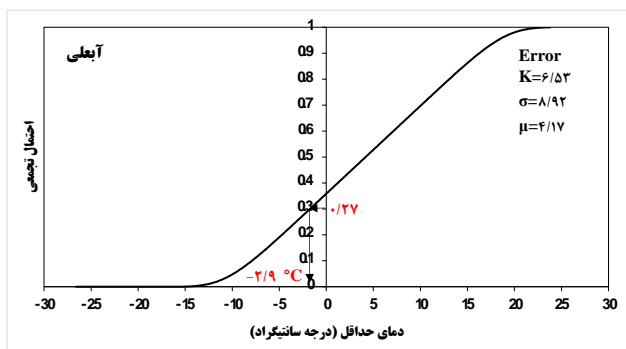
Kim and Pachepsky (2007) Boshkoska and Stankovski (2010) Mwale *et al.* (2011) Di Piazza *et al.* (2012) Wagner *et al.* (2012) اشاره کرد. در میان پژوهش های به عمل آمده در خصوص بازسازی دما، Carrega (1995) به منظور بازسازی داده های دما در مناطق ارتفاعی، یک روش رگرسیون بر پایه داده های رقومی (شیب، جهت، ارتفاع (نسبی و مطلق)، فاصله از دریا) را پیشنهاد و کارایی آن را مورد آزمون قرار داد. همچنین ترکیبی از شبکه عصبی و روش های غیرقطعی را می توان در پژوهش Demyanov *et al.* (1998) مشاهده کرد. به طوری که در این پژوهش باقیمانده های شبکه عصبی مصنوعی با استفاده از روش کریجینگ معمولی مورد ارزیابی قرار گرفتند. نتایج به دست آمده حاکی از بازدهی بالای این روش در نمایش مشخصات آماری دوره ای، محلی و متأثر از الگوهای بزرگ مقیاس در سری زمانی داده ها بود. Xia *et al.* (1999) شش روش بازسازی داده های مفقود سری زمانی اقلیمی (شامل روش میانگین حسابی ساده، روش معکوس فاصله، روش نسبت نرمال، بهترین برآوردگر منفرد، تجزیه و تحلیل رگرسیون چندگانه و کریجینگ عمومی) را مورد مقایسه قرار داده و نتیجه گرفتند که روش تجزیه و تحلیل رگرسیون چندگانه بهترین نتایج را برای سری های زمانی دما به دنبال دارد. Coulibaly and Evora (2007) از ترکیب شش شبکه عصبی مصنوعی به منظور بازسازی داده های کمینه و بیشینه دما استفاده کردند. نتایج نشان داد که شبکه های پرسپترون چندلایه، خطای پایین تری در بازسازی خلأهای دمای کمینه و بیشینه دارند. You *et al.* (2008) دو روش معکوس فاصله و رگرسیون فضایی را برای بازسازی داده های دما استفاده کرده و به این نتیجه رسیدند که روش رگرسیون در نواحی کوهستانی مناسب تر از روش معکوس فاصله است. Yozgatligil *et al.* (2013) از روشی مبتنی بر شبکه عصبی مصنوعی و تکنیک جای گذاری چندگانه (Multiple Imputation Strategy (MI)) به منظور بازسازی خلأهای آماری سری زمانی بارش و دمای ماهانه بهره گرفتند. نتایج نشان داد که بکارگیری تکنیک جای گذاری چندگانه با اعمال برخی اصلاحات نتایج به نسبت بهتری را دنبال خواهد داشت. Henn *et al.* (2013) به مقایسه روش های همبستگی مکانی- زمانی با استفاده از توابع متعامد تجربی (Empirical Orthogonal Functions (EOFs))، روش درونیابی روزانه سری های زمانی و سه روش بر پایه افتاهنگ دما (Rate-Based Filling Lapse) در بازسازی خلأ داده های دمای ساعتی پرداخته و نتیجه گرفتند که کارایی این روش ها بسته به تعداد ایستگاه های موجود در منطقه



شکل ۱. موقعیت جغرافیایی ایستگاه‌های سینوپتیک مرجع (دایره) و ایستگاه‌های ارتفاعی تحت بازسازی (مربع). (۱) آبعلی: ۲۴۶۵ متر، (۲) آوج: ۲۰۳۵ متر، (۳) قروه: ۱۹۰۶ متر، (۴) زرینه اباتو: ۲۱۴۳ متر، (۵) الیگودرز: ۲۰۲۲ متر، (۶) میمه: ۱۹۸۰ متر، (۷) داران: ۲۲۹۰ متر، (۸) کوه‌رنگ: ۲۲۸۵ متر، (۹) شهرکرد: ۲۰۴۹، (۱۰) بروجن: ۲۱۹۷ متر، (۱۱) آباده: ۲۰۳۰ متر، (۱۲) بافت: ۲۲۸۰ متر

جدول ۱. روش‌های بازسازی و تکمیل داده‌های اقلیمی استفاده‌شده در مطالعه

علامت اختصاری	روش	علامت اختصاری	روش
MLRW <sub>r</sub> +1	رگرسیون چندگانه وزن‌دار به اندازه همبستگی به توان ۱	AVNSd	مقدار متوسط نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله
MLRW <sub>r</sub> +1/2	رگرسیون چندگانه وزن‌دار به اندازه همبستگی به توان ۱/۲	AVNSr	مقدار متوسط نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی
MLRW <sub>r</sub> /d+1	رگرسیون چندگانه وزن‌دار به اندازه نسبت همبستگی به فاصله به توان ۱	CMD	متوسط اقلیمی روز مورد نظر
MLRW <sub>r</sub> /d+2	رگرسیون چندگانه وزن‌دار به اندازه نسبت همبستگی به فاصله به توان ۲	NNVd	مقدار نزدیک‌ترین همسایه بر اساس فاصله
ANNNSd	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	NNVr	مقدار نزدیک‌ترین همسایه بر اساس همبستگی
ANNNd	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه به لحاظ فاصله	IDW+1	عکس مجذور فاصله به توان ۱
ANNNSr	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	IDW+2	عکس مجذور فاصله به توان ۲
ANNNr	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه به لحاظ همبستگی	IDW+1/2	عکس مجذور فاصله به توان ۱/۲
ANNNSr/d	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه‌ها بر اساس نسبت همبستگی به فاصله	LRNSd	رگرسیون خطی با نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله
ANNNr/d	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه بر اساس نسبت همبستگی به فاصله	LRNd	رگرسیون خطی با نزدیک‌ترین ایستگاه به لحاظ فاصله
APND	متوسط روزهای ماقبل و مابعد	LRNSr	رگرسیون خطی با نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی
PCANSd	تحلیل مؤلفه‌های اصلی با نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	LRNr	رگرسیون خطی با نزدیک‌ترین ایستگاه به لحاظ همبستگی
PCANSr	تحلیل مؤلفه‌های اصلی با نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	LRNSr/d	رگرسیون خطی با نزدیک‌ترین ایستگاه‌ها بر اساس نسبت همبستگی به فاصله
CDFNd	*روش مبتنی بر توزیع تجمعی داده‌های نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	LRNr/d	رگرسیون خطی با نزدیک‌ترین ایستگاه بر اساس نسبت همبستگی به فاصله
CDFNr	*روش مبتنی بر توزیع تجمعی داده‌های نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	MLRWd-1	۱- رگرسیون چندگانه وزن‌دار به اندازه فاصله به توان ۱
CDFNr/d	*روش مبتنی بر توزیع تجمعی داده‌های نزدیک‌ترین ایستگاه‌ها بر اساس نسبت همبستگی به فاصله	MLRWd-2	۲- رگرسیون چندگانه وزن‌دار به اندازه فاصله به توان ۲



شکل ۲. نمایش بازسازی دمای کمینه در ایستگاه آبعلی با روش توزیع تجمعی داده‌ها به کمک ایستگاه دوشان تپه

به‌طور خلاصه روش کار پژوهش حاضر به این صورت بوده است که در هر مورد پس از حذف تصادفی ۵۰۰ داده مشاهداتی به‌طور تصادفی در ۱۲ ایستگاه ارتفاعی موردبررسی، مقادیر داده‌های حذف‌شده از طریق هر یک از روش‌های فوق برآورد شده و با استفاده از آماره ریشه میانگین مجذور مربعات خطا<sup>۱</sup> و بررسی حدود خطا در سطوح اطمینان ۹۰ و ۹۵ درصد، اولویت هر یک از روش‌ها شناسایی شده است.

## نتایج و بحث

### روش‌های برتر بازسازی در ایستگاه‌های مطالعاتی

جدول (۲) اولویت کارایی روش‌های مختلف بازسازی در ایستگاه‌های موردبررسی را به‌خوبی نمایش می‌دهد. از نکات بارز این جدول می‌توان به کارایی بالاتر روش‌های تحلیل مؤلفه‌های اصلی، روش مبتنی بر توزیع تجمعی داده‌ها و روش شبکه عصبی مصنوعی در مقایسه با روش‌های دیگر اشاره کرد. همچنین با کاهش ارتفاع ایستگاه‌ها بر کارایی روش‌هایی ساده‌تر افزوده می‌شود. در جدول (۳) حدود بالا و پایین حدود اعتماد ۹۰ و ۹۵٪ و همچنین بیشترین خطای مشاهده‌شده در سری بازسازی‌شده مربوط به بهترین روش در هر ایستگاه ارائه شده است. در همین جدول مقدار RMSE مربوطه نیز ارائه شده

چگونگی کاربرد روش‌های رایج مانند مقدار متوسط نزدیک‌ترین ایستگاه‌ها، متوسط اقلیمی روز موردنظر، مقدار نزدیک‌ترین همسایه، عکس مجذور فاصله، رگرسیون خطی، رگرسیون چندگانه وزن‌دار و متوسط روزهای ماقبل و مابعد، شبکه عصبی مصنوعی و روش تحلیل مؤلفه‌های اصلی در منابع مربوطه منعکس است. از آنجایی که در این روش‌ها مسئله استفاده از ایستگاه‌های مجاور چه به لحاظ فاصله و چه به لحاظ همبستگی می‌تواند نتایج مختلفی را بدنبال داشته باشد در این بررسی از حالات مختلف فاصله و ضریب همبستگی پیروان بین ایستگاه‌ها به‌عنوان معیاری برای بازسازی استفاده شده است.

در مطالعه حاضر افزون بر روش‌های پیشین در زمینه بازسازی داده، یک روش جدید بر پایه توزیع تجمعی احتمالات وقوع داده‌های ایستگاه‌های مجاور و یا همبسته نیز پیشنهاد و آزمون شده است. این روش در گام نخست مبتنی بر انتخاب مناسب‌ترین ایستگاه همبسته و مجاور به ایستگاه تحت بازسازی است. سپس برانزده‌ترین تابع توزیع احتمالات وقوع دماها تعیین و احتمال تراکمی دما برای هر ایستگاه مشخص می‌گردد. این تشخیص بر اساس آزمون اندرسون دارلینگ (Anderson-Darling test)

انجام‌پذیر گردید. در گام بعدی برای برآورد دمای یک روز معین در ایستگاه تحت بازسازی، ابتدا احتمال وقوع دمای آن روز از تابع تجمعی احتمالات ایستگاه مرجع تعیین و سپس دمای متناظر با آن احتمال از تابع تجمعی احتمالات ایستگاه تحت بازسازی مشخص می‌گردد. به عبارت دیگر، مقدار متغیر بازسازی‌شده در این روش عبارت است از مقدار متغیری از تابع توزیع تجمعی ایستگاه مورد بازسازی که مقدار احتمال تجمعی آن با احتمال تجمعی متغیر موردنظر در ایستگاه‌های مرجع برابر باشد. برای نمونه در شکل (۲) روند انجام بازسازی با استفاده از روش مبتنی بر توزیع تجمعی داده‌ها در ایستگاه آبعلی بر اساس داده‌های ایستگاه مرجع که دارای بیشترین همبستگی با ایستگاه مذکور است (دوشان تپه،  $R^2 = 0.97$ ) نموده شده و مشخصات تابع چگالی احتمال آنها نیز درج شده است. همان‌طور که از شکل برمی‌آید، احتمال تجمعی دمای ۵/۴ درجه سانتی‌گراد (مربوط به یک روز معین) در ایستگاه دوشان تپه برابر ۰/۲۷ به دست آمده که بر اساس توزیع تجمعی داده‌های دمای کمینه ایستگاه آبعلی مقدار دمای هم احتمال متناظر با آن برابر ۲/۹- درجه سانتی‌گراد به دست می‌آید. بررسی‌های دقیق‌تر به‌منظور بررسی صحت دمای به‌دست‌آمده نشان داد که در طی دوره تاریخی ایستگاه دوشان تپه ۲۳ مرتبه دمای ۵/۴ درجه سانتی‌گراد را تجربه کرده که متوسط دمای کمینه در روزهای فوق برای ایستگاه آبعلی برابر ۳- درجه سانتی‌گراد بوده است.

به مقدار خطای بالای روش‌های رگرسیون چندگانه وزن‌دار (عکس مجذور فاصله) در ایستگاه آبعلی و روش متوسط روزهای قبل و بعد در ایستگاه زرینه اباتو اشاره کرد که گاهاً تا بیش از ۲۰ درجه اختلاف بین مقادیر واقعی و بازسازی مشاهده شد.

است. با توجه به مقادیر این جدول می‌توان این‌طور استنباط کرد که در بهترین بازسازی‌ها هم ممکن است تا ۳ درجه اختلاف بین مقادیر واقعی و بازسازی دیده شود. به‌طور مشابه در خصوص بدترین روش‌ها در ایستگاه‌های موردبررسی نیز می‌توان

جدول ۲. اولویت کارایی روش‌های مختلف بازسازی دمای حداقل روزانه در ایستگاه‌های موردبررسی با معیار اندازه بازه حدود اطمینان ۹۵٪ خطا

ایستگاه												رتبه
آبعلی	اوج	قره	زرینه اباتو	الیگودرز	میمه	داران	کوه‌رنگ	شهرکرد	بروجن	آباده	بافت	
PCANSr	PCANSr	PCANSd	PCANSr	PCANSr	PCANSd	PCANSd	PCANSr	PCANSr	PCANSr	PCANSr	PCANSr	1
PCANSr	PCANSr	PCANSr	PCANSd	PCANSd	PCANSd	PCANSr	PCANSd	PCANSd	PCANSd	PCANSd	PCANSr	2
AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	AVNSr	3
MLRWtr/d+2	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	MLRWtr/d	4
MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	MLRWd-2	5
CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	CDFNr/d	6
CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	CDFNd	7
MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	MLRWtr/d+1	8
MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	MLRWd-1	9
APND	APND	APND	APND	APND	APND	APND	APND	APND	APND	APND	APND	10
PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	PCANSd	11
ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	ANNNSd	12
ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	ANNNd	13
ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	ANNNSr/d	14
LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	LRNd	15
LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	LRNr/d	16
ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	ANNNr	17
ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	ANNNSr	18
LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	LRNSd	19
LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	LRNSr/d	20
LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	LRNr	21
LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	LRNSr	22
AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	AVNSd	23
IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	IDW+1/2	24
IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	IDW+1	25
NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	NNVr	26
IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	IDW+2	27
NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	NNVd	28
ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	ANNNr/d	29
MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	MLRWtr+1/2	30
MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	MLRWtr+1	31
CMD	CMD	CMD	CMD	CMD	CMD	CMD	CMD	CMD	CMD	CMD	CMD	32

کوچکتر باشد نشانگر کارایی بالاتر روش در فصل موردنظر می‌باشد. به طوری که ملاحظه می‌شود خطای بازسازی در فصول سرد سال بالاتر از سایر فصول بوده و به‌ویژه در فصل زمستان حدود تغییرات خطای بازسازی بیشتر از سایر فصول سال است. همچنین علی‌رغم کارایی بالای روش‌های بازسازی در برآورد مقادیر مفقود در فصل تابستان، در موارد نادر (فراوانی بسیار ناچیز) اختلاف شبیه‌سازی تا ۱۰ درجه سانتی‌گراد هم مشاهده

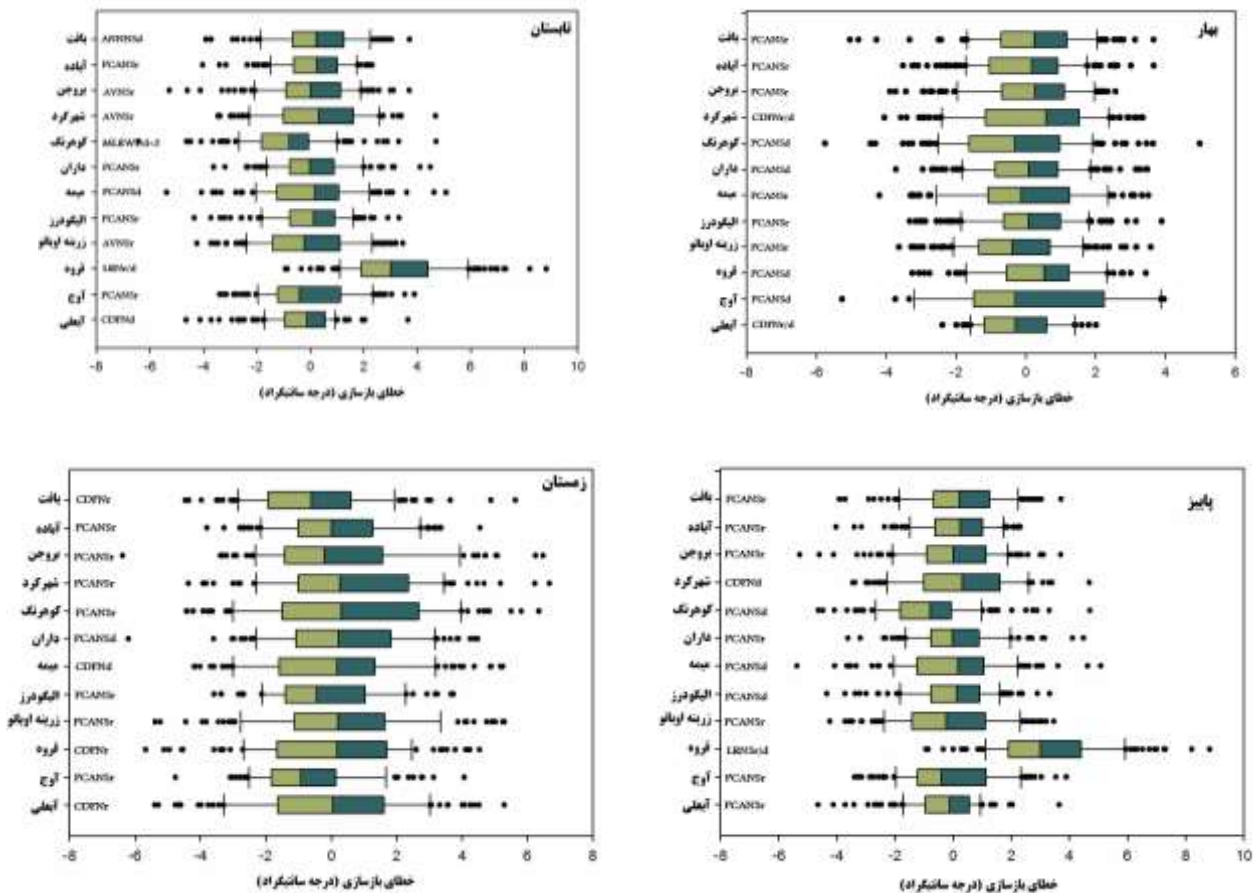
مقایسه کارایی روش‌های بازسازی در فصول مختلف سال شکل (۳) آماره‌های توزیع فراوانی خطای برآورد بهترین روش‌ها در هر یک از ایستگاه‌های مطالعاتی در طی فصول مختلف سال را نمایش می‌دهد. در این شکل مقادیر میانه، صدک‌های ۱۰، ۲۵، ۷۵ و ۹۰ خطای بازسازی در هر یک از این روش‌ها در ایستگاه‌های مورد بحث نیز نموده شده است. طبیعتاً در این شکل هرچه مقدار خطا به صفر نزدیک‌تر و بازه حدود اطمینان

نزدیک‌ترین ایستگاه به لحاظ همبستگی (زیرینه اوباتو و شهرکرد) و روش رگرسیون خطی چندگانه (کوهرنگ) هم جزو بهترین روش‌های بازسازی دسته‌بندی می‌شود.

می‌شود. از سوی دیگر در فصول پاییز، زمستان و بهار همچنان روش تحلیل مؤلفه‌های اصلی و روش مبتنی بر توزیع تجمعی داده‌ها بهترین نتایج را به دنبال داشتند اما در فصل تابستان روش‌های ساده‌تر نظیر استفاده از مقدار دمای کمینه

جدول ۳. حدود اعتماد ۹۰ و ۹۵ درصد، خطای مطلق مشاهده‌شده و مقادیر فصلی RMSE در سری‌های بازسازی‌شده با استفاده از بهترین روش در هر ایستگاه (°C)

ایستگاه	بهترین روش	حدود اعتماد ۹۵%		حدود اعتماد ۹۰%		حداکثر مطلق خطای بازسازی	RMSE		
		حد بالای خطا	حد پایین خطا	حد بالای خطا	حد پایین خطا		تابستان	پاییز	زمستان
آبعلی	CDFNr	2.2	-2.1	1.8	-1.7	5.4	1.5	1.3	2.3
آوج	PCANSr	2.5	-2.5	2.2	-2.2	4.8	2.2	2.3	2.8
قروه	PCANSd	2.7	-2.9	2.2	-2	6.8	1.9	1.9	2.4
زیرینه اباتو	PCANSr	3.1	-3	2.6	-2.4	5.8	2.5	2.9	3.7
الیگودرز	PCANSr	2.1	-2.6	1.8	-2.2	4.3	2	1.9	2.8
میمه	PCANSd	2.9	-3.2	2.3	-2.7	6	2.3	3	2.8
داران	PCANSr	3	-2.3	2.1	-1.8	5.5	1.8	1.8	3.5
کوهرنگ	PCANSd	3.3	-3.5	2.6	-2.8	6.9	1.9	1.9	3.7
شهرکرد	CDFNd	3.3	-3.1	2.7	-2.7	6.6	1.9	1.9	3.8
بروجن	PCANSr	2.5	-2.5	2.2	-2.2	6.5	2	2	3
آباده	PCANSr	2.5	-2.3	2.1	-2	4.5	1.9	1.7	2.9
بافت	PCANSr	2.7	-2.6	2.2	-2.1	5	2.5	2.7	3.2



شکل ۳. توزیع فراوانی خطای بازسازی دمای حداقل به روش‌های برتر در ایستگاه‌های موردبررسی در فصول مختلف

رگرسیون خطی در رتبه‌های بعدی قرار می‌گیرند. در اکثر ایستگاه‌ها روش رگرسیون چندگانه وزن‌دار بدترین نتایج را بدنبال داشته است به طوری که در ایستگاه‌های آبعلی، بروجن و زرینه اوباتو مقادیر بازسازی شده بیش از ۱۰ درجه سانتی‌گراد با مقدار واقعی اختلاف نشان می‌دهند.

در جدول (۴) حدود اعتماد ۹۵ درصد خطای مناسب‌ترین روش‌ها در فصول مختلف سال برای شبکه ایستگاه‌های مطالعاتی منعکس است. با توجه به این جدول می‌توان گفت در مجموع روش مبتنی بر توزیع احتمالات تجمعی و روش تحلیل مؤلفه‌های اصلی نتایج قابل قبول‌تری را در همه فصول بدنبال دارند. پس از این روش، روش شبکه عصبی مصنوعی و

جدول ۴. روش‌های برتر و حدود اعتماد ۹۵٪ خطای بازسازی آن‌ها در ایستگاه‌ها و فصول مختلف سال (°C)

ایستگاه	بهار		تابستان		پاییز		زمستان	
	بهترین روش	حدود اعتماد	بهترین روش	حدود اعتماد	بهترین روش	حدود اعتماد	بهترین روش	حدود اعتماد
آبعلی	CDFNr/d	1/41 -1/60	CDFNd	1/66 -1/61	PCANSr	1/33 -2/54	CDFNr	3/57 -4/05
آوج	PCANSd	3/87 -3/55	PCANSr	2/28 -2/38	PCANSr	2/66 -2/46	PCANSr	2/49 -2/77
قروه	PCANSd	2/77 -2/16	LRNr/d	2/16 -2/47	LRNSr/d	6/83 0/42	CDFNr	3/52 -3/55
زرینه اوباتو	PCANSr	2/25 -2/63	AVNSr	7/78 1/89	PCANSr	2/70 -2/74	PCANSr	4/62 -3/40
الیگودرز	PCANSr	2/17 -2/45	PCANSr	1/92 -2/50	PCANSd	1/87 -2/54	PCANSr	3/18 -2/72
میمه	PCANSr	3/02 -3/05	PCANSd	2/29 -2/90	PCANSd	2/58 -2/68	CDFNd	3/95 -3/33
داران	PCANSd	2/31 -2/30	PCANSr	2/04 -2/32	PCANSr	2/44 -1/90	PCANSd	3/58 -2/58
کوه‌رنگ	PCANSd	2/25 -3/02	MLRWrd+2	-1/62 -5/55	PCANSd	1/50 -3/34	PCANSr	4/73 -3/67
شهرکرد	CDFNr/d	2/92 -2/81	AVNSr	4/66 0/10	CDFNd	2/72 -2/73	PCANSr	4/18 -2/86
بروجن	PCANSr	2/09 -2/70	AVNSr	4/21 -0/52	PCANSr	2/32 -2/59	PCANSr	4/65 -2/97
آباده	PCANSr	2/42 -2/30	PCANSr	2/98 -2/35	PCANSr	1/87 -2/09	PCANSr	3/00 -2/54
بافت	PCANSr	2/35 -1/82	ANNNSd	1/87 -3/08	PCANSr	2/76 -2/16	CDFNr	2/52 -3/30

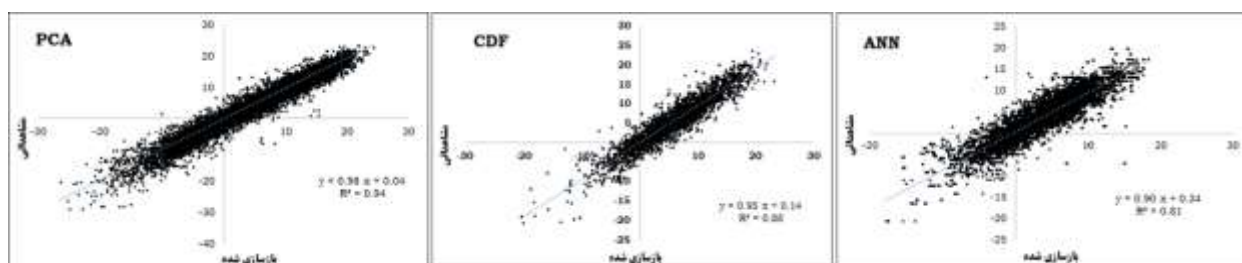
نمایش می‌دهد. بر اساس نتایج بدست آمده، سه روش تحلیل مؤلفه‌های اصلی، روش مبتنی بر توزیع تجمعی داده‌ها و روش شبکه عصبی مصنوعی به‌عنوان بهترین روش‌های بازسازی، به ترتیب با امتیازهای ۲۹/۸، ۲۸/۹ و ۱۹/۹ ارزیابی شدند. روش "متوسط اقلیمی روز موردنظر" با ۲/۷ امتیاز ضعیف‌ترین روش بازسازی محسوب می‌گردد. در شکل (۴) نتیجه یک‌کاسه کردن کلیه مقادیر بازسازی شده دمای حداقل روزانه در کل شبکه به‌صورت نموداری برای روش‌های برگزیده جمع‌بندی و ارائه شده است. به طوری که ملاحظه می‌شود اولویت‌های اصلی روش‌های بازسازی را به ترتیب روش‌های تحلیل مؤلفه‌های اصلی (R<sup>2</sup>=۰/۹۸)، روش مبتنی بر توزیع تجمعی (R<sup>2</sup>=۰/۸۸) و شبکه عصبی مصنوعی (R<sup>2</sup>=۰/۸۱) برای تشکیل می‌دهند.

#### مقایسه روش‌های برتر در بازسازی دمای کمینه روزانه-نگرش یکپارچه

همان‌طور که پیشتر نیز عنوان شد در این مطالعه خطای برآورد هر یک از ۳۲ روش موردنظر در ایستگاه‌های ارتفاعی با استفاده از آماره اندازه بازه اطمینان ۹۵ درصد و همچنین ریشه میانگین مربعات خطا مورد بررسی قرار گرفت و به این ترتیب بهترین روش‌ها در هر یک از این ایستگاه‌ها شناسایی گردید. در گام بعدی به‌منظور ارزیابی کمی اولویت روش‌های بازسازی داده دمای حداقل روزانه اقدام به رتبه‌بندی روش‌ها در هر یک از ایستگاه‌ها شد، به این صورت که به هر یک از روش‌ها حسب ترتیب اولویت امتیازی بین ۳۲ (برترین) تا ۱ (نامناسب‌ترین) داده شد و سپس بین رتبه هر یک از روش‌ها در ایستگاه‌های ۱۲ گانه میانگین‌گیری شد. جدول (۵) نتایج حاصل از این بررسی را

جدول ۵. اولویت امتیازی روش‌های بازسازی دمای حداقل روزانه در ایستگاه‌های مورد بررسی (معیار ۱ تا ۳۲)

اولویت	روش	علامت اختصاری	میانگین امتیاز رتبه‌بندی
1	تحلیل مؤلفه‌های اصلی	PCA	۲۹/۸
2	روش مبتنی بر توزیع تجمعی داده‌ها	CDF	۲۸/۹
3	شبکه عصبی مصنوعی	ANN	۱۹/۹
4	مقدار متوسط نزدیک‌ترین ایستگاه‌ها	AVN	۱۷/۹
5	عکس مجذور فاصله	IDW	۱۷/۵
6	رگرسیون خطی	LR	۱۵/۶
7	مقدار نزدیک‌ترین همسایه	NN	۱۲
8	متوسط روزهای ماقبل و مابعد	APND	۱۱/۳
9	رگرسیون چندگانه وزن‌دار	MLR	۸/۱
10	متوسط اقلیمی روز مورد نظر	CMD	۲/۷



شکل ۴. مقایسه مقادیر بازسازی شده دمای کمینه روزانه به روش‌های برتر با مقادیر حقیقی آنها در شبکه ایستگاه‌های مطالعاتی

## نتیجه‌گیری کلی

در مجموع بر اساس یافته‌های این پژوهش، سه روش تحلیل مؤلفه‌های اصلی با درجه برتری ۲۹/۸، روش پیشنهادی مبتنی بر توزیع تجمعی با امتیاز ۲۸/۹ و شبکه عصبی مصنوعی با اولویت ۱۹/۹ برای بازسازی داده‌های مفقود در ایستگاه‌های ارتفاعی قابل پیشنهاد هستند.

در این تحقیق به منظور آزمون کارایی روش‌های بازسازی داده‌های دمای کمینه روزانه در ایستگاه‌های مرتفع کشور ضمن دسته‌بندی روش‌ها از ۳۱ روش عام و یک روش پیشنهادی استفاده شد. مهم‌ترین نتایج حاصل از این پژوهش را می‌توان به صورت زیر خلاصه کرد:

## سپاسگزاری

این پژوهش با پشتیبانی صندوق حمایت از پژوهشگران و فناوران کشور به انجام رسیده است.

کارایی روش‌های بازسازی با افزایش ارتفاع ایستگاه کاهش می‌یابد، آن چنانکه در ایستگاه‌های ابعلی، داران و کوه‌رنگ نتایج در مقایسه با سایر ایستگاه‌ها ضعیف‌تر ارزیابی شد.

## نمادها

AVNS: Average value nearer stations  
 CMD: Climatological mean of the day  
 NNV: Nearest Neighbor Value  
 IDW: Inverse Distance Weight  
 LR: Linear Regression  
 MLRW: Weighted Multiple Linear Regression  
 ANN: Artificial Neural Network  
 APND: Average of Previous and Next Day  
 PCA: Principal component analysis  
 CDF: Cumulative distribution function  
 $\mu$ : continuous location parameter  
 $\sigma$ : continuous scale parameter  
 K: continuous shape parameter

در همه فصول روش‌های مبتنی بر توزیع تجمعی و تحلیل مؤلفه‌های اصلی برترین روش‌ها محسوب می‌شوند و روش شبکه عصبی مصنوعی در بعضی از ایستگاه‌ها برتر محسوب نمی‌گردد. کارایی روش‌های بازسازی در فصل زمستان پایین‌تر از سایر فصول است و روش‌های تحلیل مؤلفه‌های اصلی و شبکه عصبی مصنوعی و روش‌های مبتنی بر توزیع تجمعی نتایج بهتری را در این فصل عاید می‌سازند.

روش‌های مقدار متوسط نزدیک‌ترین ایستگاه‌ها در فصل بهار و رگرسیون چندگانه وزن‌دار در فصول دیگر از کمترین کارایی در زمینه بازسازی خلأهای آماری داده‌های دمای کمینه روزانه برخوردارند.



## REFERENCES

- Ashraf, M., Loftis, J. C., & Hubbard, K. G. (1997). Application of geostatistics to evaluate partial weather station networks. *Agricultural and forest meteorology*, 84(3), 255-271.
- Carrega, P. (1995). A method for the reconstruction of mountain air temperatures with automatic cartographic applications. *Theoretical and applied climatology*, 52(1-2), 69-84.
- Coulibaly, P., & Evora, N. D. (2007). Comparison of neural network methods for infilling missing daily weather records. *Journal of hydrology*, 341(1), 27-41.
- Demyanov, V., Kanevsky, M., Chernov, S., Savelieva, E., & Timonin, V. (1998). Neural network residual kriging application for climatic data. *Journal of Geographic Information and Decision Analysis*, 2(2), 215-232.
- Di Piazza, A., Conti, F. L., Noto, L. V., Viola, F., & La Loggia, G. (2011). Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation*, 13(3), 396-408.
- Khalil, M., Panu, U. S., & Lennox, W. C. (2001). Groups and neural networks based streamflow data infilling procedures. *Journal of Hydrology*, 241(3), 153-176.
- Khalili A (1991) Integrated Water Plan of Iran, Jamab Consulting Engineering Co., The Ministry of Energy, Tehran, 111-122. (In Farsi)
- Henn, B., Raleigh, M. S., Fisher, A., & Lundquist, J. D. (2013). A comparison of methods for filling gaps in hourly near-surface air temperature data. *Journal of Hydrometeorology*, 14(3), 929-945.
- Khorshiddoust, A. M., Nassaji, Z. M., and Ghermez, C. B. (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 Farsi)
- Kim, J. W., & Pachepsky, Y. A. (2010). Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *Journal of hydrology*, 394(3), 305-314.
- Mileva-Boshkoska, B., & Stankovski, M. (2007). Prediction of missing data for ozone concentrations using support vector machines and radial basis neural networks. *Informatica*, 31(4).
- Mwale, F. D., Adeloye, A. J., & Rustum, R. (2012). Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi—A self organizing map approach. *Physics and Chemistry of the Earth, Parts A/B/C*, 50, 34-43.
- Teegavarapu, R. S., & Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312(1), 191-206.
- Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., & Schneider, K. (2012). Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *Journal of Hydrology*, 464, 388-400.
- Xia, Y., Fabian, P., Stohl, A., & Winterhalter, M. (1999). Forest climatology: estimation of missing values for Bavaria, Germany. *Agricultural and Forest Meteorology*, 96(1), 131-144.
- Yozgatligil, C., Aslan, S., Iyigun, C., & Batmaz, I. (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.
- You, J., Hubbard, K. G., & Goddard, S. (2008). Comparison of methods for spatially estimating station temperatures in a quality control system. *International Journal of Climatology*, 28(6), 777-787.