تحقیقات آب و خاک ایران، دورهٔ ۴۸، شمارهٔ ۲، تابستان ۱۳۹۶ (ص ۲۳۹-۲۳۱)

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

جابر رحیمی[۱]، علی خلیلی[۲*]، جواد بذرافشان[۳]

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 سرماهای 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 معدود 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 تجمعی احتمالات وقوع دماهای کمینه بازسازی گردید و اعتبار هریک 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 ۲/۰±، ۲/۲± و ۳/۱± 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 سرمای 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

نبود یک پایه زمانی مشترک طولانی آماری در شبکه ایستگاه-های هواشناسی کشور که خود ناشی از تأسیس این ایستگاه‌ها در سال‌های مختلف است همواره به عنوان یکی از مسائل و معضلات عمده در روند تجزیه‌وتحلیل‌های 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 تصحیح این داده‌ها همواره موردتوجه محققان باشد. ازآنجایی‌که 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. به‌طورکلی 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 صورت‌گرفته است که به عنوان نمونه می‌توان به Ashraf *et al.* (1997)، Khalil *et al.* (2001)، Teegavarapu and Chandramouli (2005)، Mileva

* نویسنده مسئول : akhalili@ut.ac.ir

Archive of SID
www.SID.ir

Kim and Pachepsky ،(2007) Boshkoska and Stankovski (2010)، Di Piazza *et al.* (2011)، Mwale *et al.* (2012) و Wagner *et al.* (2012) اشاره کرد. در میان پژوهش‌های به‌عمل‌آمده در خصوص بازسازی دما، Carrega (1995) به‌منظور بازسازی داده‌های دما در مناطق ارتفاعی، یک روش رگرسیون بر پایه داده‌های 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 روش‌های غیرقطعی را می‌توان در پژوهش Demyanov *et al.* (1998) مشاهده کرد. به‌طوری‌که در این پژوهش باقیمانده‌های شبکه عصبی مصنوعی با استفاده از روش کریجینگ معمولی مورد ارزیابی قرار گرفتند. نتایج به‌دست‌آمده حاکی از بازدهی بالای این روش در نمایش مشخصات آماری دوره‌ای، محلی و متأثر از الگوهای بزرگ‌مقیاس در سری زمانی داده‌ها بود. Xia *et al.* (1999) شش روش بازسازی داده‌های مفقود سری زمانی اقلیمی (شامل روش میانگین حسابی ساده، روش معکوس فاصله، روش نسبت نرمال، بهترین برآوردگر منفرد، تجزیه‌وتحلیل رگرسیون چندگانه و کریجینگ عمومی) را مورد مقایسه قرار داده و نتیجه گرفتند که روش تجزیه‌وتحلیل رگرسیون چندگانه بهترین نتایج را برای سری‌های زمانی دما به دنبال دارد. Coulibaly and Evora (2007) از ترکیب شش شبکه عصبی مصنوعی به‌منظور بازسازی داده‌های کمینه و بیشینه دما استفاده کردند. نتایج نشان داد که شبکه‌های پرسپترون چندلایه، خطای پایین‌تری در بازسازی خلاءهای دمای کمینه و بیشینه دارند. You *et al.* (2008) دو روش معکوس فاصله و رگرسیون فضایی را برای بازسازی داده‌های دما استفاده کرده و به این نتیجه رسیدند که روش رگرسیون در نواحی کوهستانی مناسب‌تر 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)) به‌منظور بازسازی خلاءهای 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 بدنبال خواهد داشت. Henn *et al.* (2013) به مقایسه روش‌های همبستگی مکانی- زمانی با استفاده از توابع متعامد تجربی (Empirical Orthogonal Functions (EOFs))، روش درون‌یابی روزانه سری‌های زمانی و سه روش بر پایه افتاهنگ دما (Rate–Based Filling Lapse) در بازسازی خلأ داده‌های دمای ساعتی پرداخته و نتیجه گرفتند که کارایی این روش‌ها بسته به تعداد ایستگاه‌های موجود در منطقه 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 روش‌های نزدیک‌ترین همسایه و شبکه عصبی مصنوعی مورد بررسی قرار داده و نشان دادند که برای کلیه ایستگاه‌های موردمطالعه روش شبکه عصبی متوسط خطای مطلق کمتری را نسبت به روش نزدیک‌ترین همسایه دارد. همچنین دقت هر دو روش در تخمین دمای بیشینه روزانه بیشتر از دمای کمینه روزانه بود.

درمجموع 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 می‌شود. لذا هدف از این بررسی بازسازی خلأهای دمای کمینه روزانه و ایجاد سری زمانی پیوسته این عامل در دوره ۲۰۱۰-۱۹۶۵ به‌منظور بهره‌گیری بعدی از آنها در مطالعات تغییر اقلیم و اقلیم‌شناسی رویدادهای سرمای زودرس پاییزه، دیررس بهاره، طول مدت سرما و تعیین تاریخ این رویدادها در فرآیند تغییر اقلیم است که نهایتاً در مسائل پیش‌نگری امنیت غذایی کاربرد دارند.

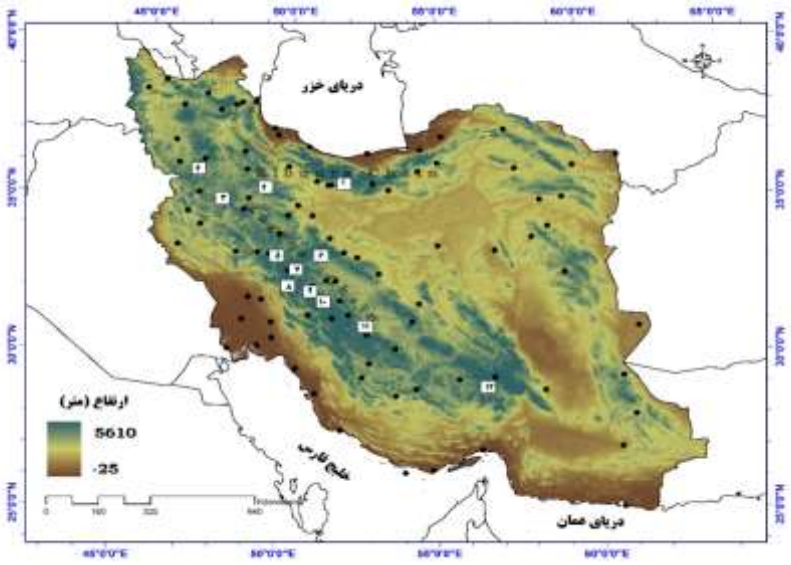
مواد و روش‌ها

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

برای تعیین مناسب‌ترین ایستگاه‌های مرجع برای بازسازی خلأهای آماری ایستگاه‌های منتخب از روش ماتریس فاصله و ماتریس ضرایب همبستگی دو به دو (Khalili, 1991) بین هریک از ۱۲ ایستگاه ارتفاعی با بقیه ایستگاه‌ها استفاده شد. به این مفهوم که از بین ایستگاه‌های مرجع، داده‌های گروه ایستگاه‌هایی که واجد همبستگی معنی‌دار حداقل در سطح ۵ درصد با داده‌های ایستگاه تحت بازسازی بودند با اولویت نزدیکی فاصله، به عنوان ورودی برای بازسازی استفاده گردید. همچنین در ارزیابی همبستگی دو بدو، میزان همبستگی با داده‌های یک و دو روز قبل به عنوان پیش‌بینی کننده محتمل نیز آزمون شد. جدول (۱) روش‌های بازسازی و تکمیل داده‌های اقلیمی استفاده‌شده در این بررسی را نشان می‌دهد.

Archive of SID

www.SID.ir



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 متر، (۱۰) بروجن: ۲۱۹۷ متر، (۱۱) آباده: ۲۰۳۰ متر، (۱۲) بافت: ۲۲۸۰ متر)

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

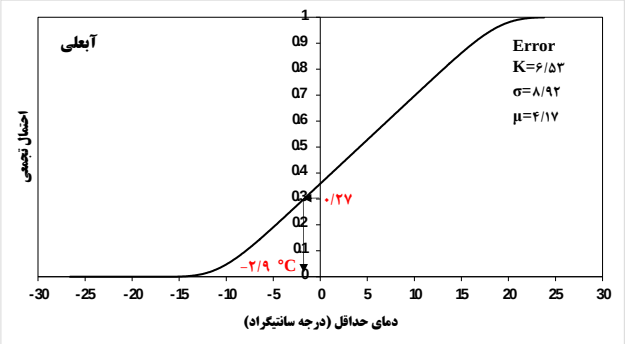
شماره	روش	علامت اختصاری	شماره	روش	علامت اختصاری
۱	مقدار متوسط نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	AVNSd	۱۷	رگرسیون چندگانه وزن‌دار به اندازه همبستگی به توان ۱	MLRWr+1
۲	مقدار متوسط نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	AVNSr	۱۸	رگرسیون چندگانه وزن‌دار به اندازه همبستگی به توان ۱/۲	MLRWr+1/2
۳	متوسط اقلیمی روز مورد نظر	CMD	۱۹	رگرسیون چندگانه وزن‌دار به اندازه نسبت همبستگی به فاصله به توان ۱	MLRWr/d+1
۴	مقدار نزدیک‌ترین همسایه بر اساس فاصله	NNVd	۲۰	رگرسیون چندگانه وزن‌دار به اندازه نسبت همبستگی به فاصله به توان ۲	MLRWr/d+2
۵	مقدار نزدیک‌ترین همسایه بر اساس همبستگی	NNVr	۲۱	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	ANNNSd
۶	عکس مجذور فاصله به توان ۱	IDW+1	۲۲	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه به لحاظ فاصله	ANNNd
۷	عکس مجذور فاصله به توان ۲	IDW+2	۲۳	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	ANNNSr
۸	عکس مجذور فاصله به توان ۱/۲	IDW+1/2	۲۴	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه به لحاظ همبستگی	ANNNr
۹	رگرسیون خطی با نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	LRNSd	۲۵	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه‌ها بر اساس نسبت همبستگی به فاصله	ANNNSr/d
۱۰	رگرسیون خطی با نزدیک‌ترین ایستگاه به لحاظ فاصله	LRNd	۲۶	شبکه عصبی مصنوعی با نزدیک‌ترین ایستگاه بر اساس نسبت همبستگی به فاصله	ANNNr/d
۱۱	رگرسیون خطی با نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	LRNSr	۲۷	متوسط روزهای ماقبل و مابعد	APND
۱۲	رگرسیون خطی با نزدیک‌ترین ایستگاه به لحاظ همبستگی	LRNr	۲۸	تحلیل مؤلفه‌های اصلی با نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	PCANSd
۱۳	رگرسیون خطی با نزدیک‌ترین ایستگاه‌ها بر اساس نسبت همبستگی به فاصله	LRNSr/d	۲۹	تحلیل مؤلفه‌های اصلی با نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	PCANSr
۱۴	رگرسیون خطی با نزدیک‌ترین ایستگاه بر اساس نسبت همبستگی به فاصله	LRNr/d	۳۰	*روش مبتنی بر توزیع تجمعی داده‌های نزدیک‌ترین ایستگاه‌ها به لحاظ فاصله	CDFNd
۱۵	رگرسیون چندگانه وزن‌دار به اندازه فاصله به توان ۱-	MLRWd-1	۳۱	*روش مبتنی بر توزیع تجمعی داده‌های نزدیک‌ترین ایستگاه‌ها به لحاظ همبستگی	CDFNr
۱۶	رگرسیون چندگانه وزن‌دار به اندازه فاصله به توان ۲-	MLRWd-2	۳۲	*روش مبتنی بر توزیع تجمعی داده‌های نزدیک‌ترین ایستگاه‌ها بر اساس نسبت همبستگی به فاصله	CDFNr/d

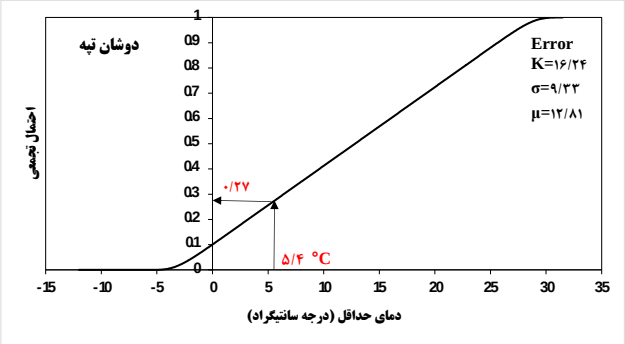
*روش پیشنهادی

Archive of SID
www.SID.ir

چگونگی کاربست 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. ازآنجایی‌که 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 پیرسون 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 برازنده‌ترین 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 =۰/۹۷) 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.





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[1] 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

1. Root Mean Square Error (RMSE)

Archive of SID

www.SID.ir

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 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.

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

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

مقایسه کارایی روش‌های بازسازی در فصول مختلف سال

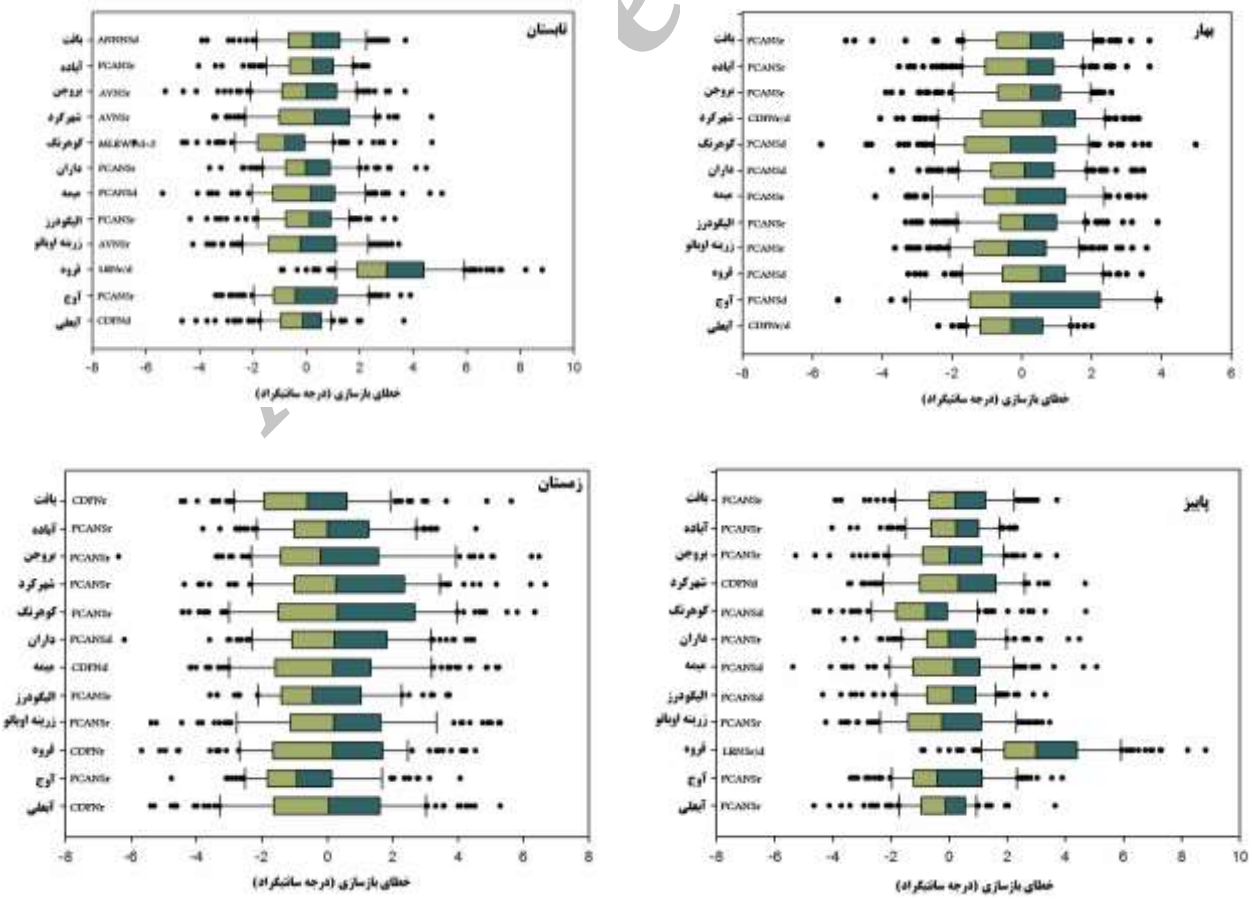
شکل (۳) آماره‌های توزیع فراوانی خطای برآورد بهترین روش‌ها در هر یک از ایستگاه‌های مطالعاتی در طی فصول مختلف سال را نمایش می‌دهد. در این شکل مقادیر میانه، صدک‌های ۱۰، ۲۵، ۷۵ و ۹۰ خطای بازسازی در هر یک از این روش‌ها در ایستگاه‌های مورد بحث نیز نموده شده است. طبیعتاً در این شکل هرچه مقدار خطا به صفر نزدیک‌تر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

Archive of SID
www.SID.ir

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

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 95%		حدود اعتماد 90%		حداکثر مطلق خطای بازسازی	RMSE			
		حد پایین خطا	حد بالای خطا	حد پایین خطا	حد بالای خطا		بهار	تابستان	پاییز	زمستان
آبعلی	CDFNr	-2.1	2.2	-1.7	1.8	5.4	1.7	1.5	1.3	2.3
آوج	PCANSr	-2.5	2.5	-2.2	2.2	4.8	2.5	2.2	2.3	2.8
قروه	PCANSd	-2.9	2.7	-2	2.2	6.8	1.9	1.9	1.9	2.4
زرینه اباتو	PCANSr	-3	3.1	-2.4	2.6	5.8	2	2.5	2.9	3.7
الیگودرز	PCANSr	-2.6	2.1	-2.2	1.8	4.3	2	2	1.9	2.8
میمه	PCANSd	-3.2	2.9	-2.7	2.3	6	2.3	2.3	3	2.8
داران	PCANSr	-2.3	3	-1.8	2.1	5.5	1.9	1.8	1.8	3.5
کوهرنگ	PCANSd	-3.5	3.3	-2.8	2.6	6.9	2.1	1.9	1.9	3.7
شهرکرد	CDFNd	-3.1	3.3	-2.7	2.7	6.6	3	1.9	1.9	3.8
بروجن	PCANSr	-2.5	2.5	-2.2	2.2	6.5	2.1	2	2	3
آباده	PCANSr	-2.3	2.5	-2	2.1	4.5	2	1.9	1.7	2.9
بافت	PCANSr	-2.6	2.7	-2.1	2.2	5	2.5	2.5	2.7	3.2



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

www.SID.ir

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 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 بازسازی‌شده بیش از ۱۰ درجه سانتی‌گراد با مقدار واقعی اختلاف نشان می‌دهند.

جدول ۴. روش‌های برتر و حدود اعتماد ۹۵٪ خطای بازسازی آن ها 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 اباتو	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	MLRWr/d+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

مقایسه روش‌های برتر در بازسازی دمای کمینه روزانه-نگرش یکپارچه

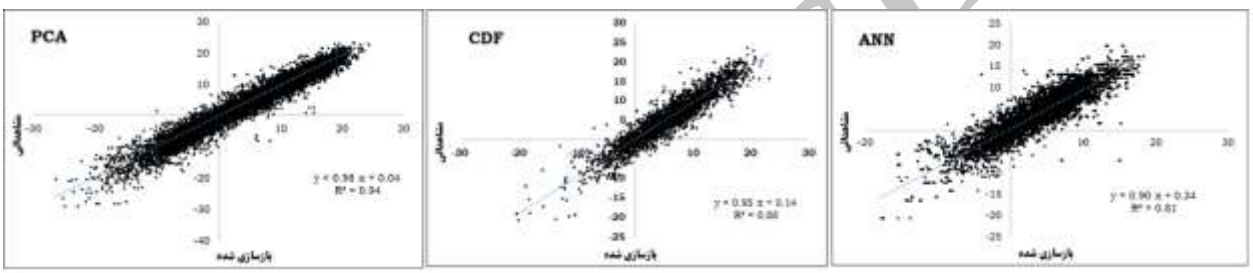
همان‌طور که پیشتر نیز عنوان شد در این مطالعه خطای برآورد هر یک از ۳۲ روش موردنظر در ایستگاه‌های ارتفاعی با استفاده از آماره اندازه بازه اطمینان ۹۵ درصد و همچنین ریشه میانگین مربعات خطا مورد بررسی قرار گرفت و به این ترتیب بهترین روش‌ها در هر یک از این ایستگاه‌ها شناسایی گردید. در گام بعدی به‌منظور ارزیابی 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 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 بازسازی‌شده 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^2=۰/۹۸)، روش مبتنی بر توزیع تجمعی (R^2=۰/۸۸) و شبکه عصبی مصنوعی (R^2=۰/۸۱) برای تشکیل می‌دهند.

Archive of SID

www.SID.ir

جدول ۵. اولویت امتیازی روش‌های بازسازی دمای حداقل روزانه در ایستگاه‌های موردبررسی (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 موردنظر	CMD	۲/۷



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

نتیجه‌گیری کلی

در این تحقیق به‌منظور 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 به‌صورت زیر خلاصه کرد:

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

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.

درمجموع 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.

سپاسگزاری

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

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
μ: continuous location parameter
σ: continuous scale parameter
K: continuous shape parameter

Archive of SID
www.SID.ir

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.

Archive of SID

www.SID.ir