Original Article

Evaluation of Classical Statistical Methods for Estimating and Reconstructing the Daily Temperature in Iran

Mohammad Khaledi-Alamdari[1] and Ghasem Zarei[2*]

Affiliation

1. *Dep. of Water Engineering,University of Tabriz,Tabriz, Iran*
2. *Dep. of Greenhouse Engineering, Agricultural Engineering Research Institute, Karaj, Iran*

ABSTRACT

Introduction: Effective decision-making and management in the field of sustainable development of natural resources require access to accurate and up-to-date climatic information. This information enables the examination of the role of climate change in various issues and the formulation of effective management strategies accordingly. In this context, temperature as one of the most important climatic indicators plays a key role in environmental analyses and research. Given the fundamental role of temperature in various situations, access to precise and comprehensive temperature data is of high importance. Such data should be sufficiently detailed to provide a clear and complete picture of temperature patterns over time. Unfortunately, climatic data often faces problems such as statistical discontinuities and measurement errors, which can lead to incorrect decisions and inefficient planning. In this research, statistical methods have been employed to analyze existing temperature data and their statistical discontinuities. These methods include geographical coordinates (graphical), normal ratio, weighted correlation coefficient, and arithmetic mean, which are well-established and widely used in completing climatic data. Selecting the most appropriate method from among these can enhance the accuracy of temperature data estimation and play a pivotal role in decisions based on more comprehensive and reliable data. The objective of this research is to identify the optimal methodology for estimating data and addressing statistical discontinuities. This will assist researchers, managers, and policymakers in the field of sustainable development and in better understanding climatic conditions, enabling them to make more informed and effective decisions.

Material and Methods: In order to address the statistical gap, a number of well-known and popular classical statistical methods were evaluated for estimating Iran temperature data. These included geographical coordinates, normal ratio, weighted correlation coefficient, and arithmetic mean. In order to determine the best method for completing missing information, data from 125 stations with complete information (without any missing data) over 21 years (2000 to 2020) were used. Given the extensive and time-consuming nature of the calculations, a random selection of 10% of the stations with an appropriate spatial distribution was employed to carry out the data filling operations. The information of the selected stations was removed at each stage separately and reconstructed based on their five nearest stations. To evaluate the aforementioned methods, statistical evaluation criteria such as R-squared (R2), root mean square error (RMSE), and mean absolute deviation (MAD) were used.

Results and discussion: The results of the analysis of computational values through the normal ratio method were evaluated against observational values. It was found that all the stations under study exhibited a high correlation, indicating the acceptability of the normal ratio method to data estimation. The average values obtained from the evaluation of results indicate that the methods of normal ratio, weighted correlation coefficient, geographical coordinates, and arithmetic mean are prioritized in order, with RMSE values of 3.05, 3.28, 3.30, and 3.51 degrees Celsius, respectively. Consequently, the normal ratio method is the most suitable among the other studied methods and can be employed to address issues such as a lack of information, existing data errors, and also the expansion of the study period.

Conclusion: Among the methods reviewed, the normal ratio method is generally more acceptable and of higher quality than the other methods and is recommended for use in future research within similar study ranges. In subsequent ranks, the methods of geographical coordinates, weighted correlation, and arithmetic mean are placed, respectively. It is notable that, although the other methods are considered of secondary importance, they nevertheless demonstrate satisfactory efficacy in certain locations. Consequently, under varying circumstances, a range of methods may be employed to address data deficiencies, and the optimal approach should be selected and utilized in accordance with the specific study area.

Keywords: Daily temperature, geographic coordinates, normal ratio, statistical discontinuity.

Citation: *Khaledi-Alamdari, M. and Zarei, G., Evaluation of Classical Statistical Methods for Estimating and Reconstructing the Daily Temperature in Iran, Iran J Remote Sens GIS. 16(1):103-112.*

* Corresponding Author: gh.zarei@areeo.ac.ir

DOI: https://doi.org/10.48308/gisj.2023.103725

Received: 2023.01.22

Accepted: 2023.07.15

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license https://creativecommons.org/licenses/by/4.0/).

مقاله پژوهشی

ارزیابی روش‌های آمار کلاسیک در تخمین و بازسازی دمای روزانهٔ کشور ایران

محمد خالدی علمداری[1] و قاسم زارعی[2*]

سمت

۱. گروه علوم و مهندسی آب، دانشگاه تبریز، تبریز، ایران

۲. مؤسسهٔ تحقیقات فنی و مهندسی کشاورزی، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران

چکیده

سابقه و هدف: تصمیم‌گیری و مدیریت مؤثر در زمینهٔ توسعهٔ پایدار منابع طبیعی نیازمند دسترسی به اطلاعات دقیق و به‌روز اقلیمی است. این اطلاعات امکان بررسی نقش تغییرات اقلیمی را در موضوعات گوناگون مهیا می‌کنند و براساس آن، می‌توان راهکارهای مدیریتی مؤثری را تدوین کرد. در این راستا، پارامتر دما یکی از مهم‌ترین شاخص‌های اقلیمی است که نقش محوری را در تحلیل‌ها و پژوهش‌های محیطی ایفا می‌کند. با توجه به نقش اساسی دما در موضوعات گوناگون، دسترسی به داده‌های دمایی دقیق و جامع اهمیت بسیاری دارد. این داده‌ها باید به‌گونه‌ای باشند که بتوانند تصویری واضح و کامل از الگوهای دمایی، در طول زمان، ارائه دهند. اما متأسفانه، داده‌های اقلیمی اغلب با مشکلاتی مانند انقطاع آماری و خطاهای اندازه‌گیری مواجه‌اند. این مشکلات ممکن است به تصمیم‌گیری‌های نادرست و برنامه‌ریزی‌های ناکارآمد منجر شوند. در این پژوهش، با استفاده از روش‌های آماری، سعی شده است داده‌های دمایی موجود و انقطاع‌های آماری آنها با استفاده از روش‌هایی ازجمله مختصات جغرافیایی (گرافیکی)، نسبت نرمال، ضریب همبستگی وزنی و میانگین حسابی که در تکمیل داده‌های اقلیمی شناخته‌شده و پرکاربردند، تحلیل و ارزیابی شود. انتخاب روش مناسب از میان این روش‌ها می‌تواند دقت تخمین داده‌های دمایی را افزایش دهد و در تصمیم‌گیری‌های مبتنی‌بر داده‌های جامع‌تر و معتبرتر، نقش اساسی داشته باشد. درنَهایت، هدف از این پژوهش معرفی بهترین روش برای تخمین اطلاعات و رفع انقطاع آماری است که پژوهشگران، مدیران و سیاست‌گذاران را در زمینهٔ توسعهٔ پایدار و درک بهتر شرایط اقلیمی و اتخاذ تصمیماتی هوشمندانه‌تر و مؤثرتر، یاری خواهد کرد.

مواد و روش‌ها: در این پژوهش، به‌منظور رفع خلأ آماری، روش‌های شناخته‌شده و محبوب آمار کلاسیک شامل روش مختصات جغرافیایی، نسبت نرمال، ضریب همبستگی وزنی و میانگین حسابی، در تخمین داده‌های دمایی کشور ارزیابی شد. به‌منظور بررسی بهترین روش برای تکمیل اطلاعات مفقودی، از اطلاعات ۱۲۵ ایستگاه استفاده شد. این ایستگاه‌ها دارای اطلاعات کامل (بدون هیچ‌گونه مفقودی)، درطول ۲۱ سال (۲۰۲۰-۲۰۰۰ م.) بودند. ازآنجاکه محاسبات گسترده و زمان‌بر بودند، با انتخاب ۱۰٪ این ایستگاه‌ها به‌صورت تصادفی با پراکندگی مکانی مناسب، عملیات پرکردن اطلاعات روی ایستگاه‌های منتخب انجام شد. اطلاعات ایستگاه‌های منتخب، در هر مرحله و به‌صورت جداگانه، حذف و براساس پنج ایستگاه مجاور خود، بازسازی شدند و به‌منظور ارزیابی روش‌های مذکور، از معیارهای ارزیابی آماری ضریب تبیین (R^2)، جذر میانگین مربعات خطا (RMSE) و میانگین انحراف مطلق (MAD) استفاده شد.

نتایج و بحث: با ارزیابی نتایج حاصل از بررسی مقادیر محاسباتی ازطریق روش نسبت نرمال درمقابل مقادیر مشاهداتی، مشخص شد تمامی ایستگاه‌های مورد بررسی همبستگی بالایی دارند؛ این نکته بیانگر مقبولیت روش نسبت نرمال برای تخمین داده‌هاست. با توجه به مقادیر متوسط حاصل از ارزیابی نتایج، روش نسبت نرمال، ضریب همبستگی وزنی، مختصات جغرافیایی و میانگین حسابی به‌ترتیب، با مقدار RMSE معادل ۳/۰۵، ۳/۲۸، ۳/۳۰ و ۳/۵۱ درجهٔ سلسیوس، اولویت‌بندی می‌شوند. بنابراین روش نسبت نرمال در میان سایر روش‌های مورد مطالعه از مقبولیت بیشتری برخوردار است و ازاین‌رو، در رفع مشکلاتی اعم‌از فقدان اطلاعات، خطای موجود در داده‌ها و همچنین گسترش دورهٔ زمانی مطالعاتی، می‌توان از آن بهره برد.

نتیجه‌گیری: در میان روش‌های مورد بررسی، روش نسبت نرمال به‌صورت کلی مقبولیت و کیفیتی بیشتر از دیگر روش‌ها دارد که توصیه می‌شود در پژوهش‌های آتی، در محدودهٔ مطالعاتی مشابه، از این روش استفاده شود. در مراتب بعدی، به‌ترتیب روش مختصات جغرافیایی، همبستگی وزنی و میانگین حسابی قرار دارند. شایان توجه است، با اینکه سایر روش‌ها در مراتب اهمیت بعدی واقع شده‌اند، همچنان در برخی ایستگاه‌ها کارآیی مناسبی نشان می‌دهند؛ بنابراین در شرایط متفاوت، روش‌های متنوعی می‌توانند نیاز به ترمیم داده‌ها را رفع کند و با توجه به محدودهٔ مورد مطالعه، باید بهترین روش انتخاب شود و به‌کار رود.

واژه‌های کلیدی: انقطاع آماری، دمای روزانه، مختصات جغرافیایی، نسبت نرمال.

استناد: *خالدی علمداری، م.، و زارعی، ق.، ارزیابی روش‌های آمار کلاسیک در تخمین و بازسازی دمای روزانهٔ کشور ایران، نشریه سنجش از دور و GIS ایران، سال ۱۶، شماره ۱، بهار ۱۴۰۳: ۱۱۲-۱۰۳.*

* نویسنده عهده‌دار مکاتبات: gh.zarei@areeo.ac.ir

شناسه دیجیتال مقالات: https://doi.org/10.48308/GISJ.2023.103725

دریافت: ۱۴۰۱/۱۱/۰۲

پذیرش: ۱۴۰۲/۰۴/۲۴

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license https://creativecommons.org/licenses/by/4.0/).

۱- مقدمه

بررسی و تصمیم‌گیری دربارهٔ هر موضوعی نیازمند استفاده از داده‌هایی مورد اطمینان، صحیح و تاحد امکان، عاری از خطاست. از دهه‌های گذشته، سیاست و آمادگی دربرابر شرایط حاد آب‌وهوایی مورد توجه ویژهٔ دولت‌ها، سازمان‌های بین‌المللی و منطقه‌ای و سازمان‌های غیردولتی قرار گرفته است (Wilhite, 2000). درپی این نکته، متغیرهای اقلیمی همچون بارش و دما در تعامل با تولیدات کشاورزی، در مقیاس بزرگ و گاه در مقیاس‌های محلی، به‌کار رفته است (Asseng et al., 2011; Bannayan et al., 2003; Palosuo et al., 2011)؛ بنابراین آگاهی از وضعیت دمایی منطقه به کشاورزان این امکان را می‌دهد که تصمیمات بهتری، درمواجهه با مخاطرات ناشی از وضعیت آب‌وهوا، اتخاذ کنند و یا از وضعیت مطلوب اقلیمی، درراستای تولید مضاعف، بهره‌مند شوند زیرا تنوع مکانی، در فراهم‌بودن شرایط مناسب دمایی برای گیاهان مناطق گوناگون، می‌تواند گسترهٔ زمانی بیشتری را برای بهره‌برداری در اختیار کشاورزان قرار دهد (Khormizie et al., 2020).

تغییرات در انتشار گازهای گلخانه‌ای، ناشی از روند تغییرات اقلیمی، به تغییر دما، بارندگی و سایر پارامترهای اقلیمی منجر می‌شود و برنامه‌ریزی‌های مدیریتی را تحت تأثیر فزاینده‌ای قرار می‌دهد. بررسی و تحقیق درزَمینه‌های مرتبط با داده‌های هواشناسی و هیدرولوژیکی و به‌ویژه مطالعات درحوزهٔ توسعهٔ منابع طبیعی پایدار، بهره‌گیری از اطلاعاتی مورد اعتماد، شامل پارامترهای اقلیمی، از مهم‌ترین مراحل به‌شمار می‌رود؛ ازاین‌رو موارد یادشده مستثنا از این نکته نیست و طالب به‌کاربردن داده‌های باکیفیت و عاری از نقص و خطاست. بدین‌منظور داده‌های دریافتی ایستگاه‌های سینوپتیک از سازمان هواشناسی همواره باید بررسی شوند تا، در میان سری‌های زمانی، مشکلاتی همچون نقص دیتا و اطلاعات وجود نداشته باشد. برآورد داده‌های گم‌شده نخستین مرحله در مطالعات هیدرولوژیکی و اقلیم‌شناسی شناخته می‌شود. درواقع داده‌های مشاهداتی ازدست‌رفته، در سری‌های زمانی آب‌وهوایی، اغلب به‌صورت متوالی طی دوره‌های زمانی طولانی مشهودند (Simolo et al., 2010). داده‌های گم‌شده باعث ایجاد اختلال در تخمین پارامترها و شناسایی فرایندها می‌شوند و افزون‌برآن ممکن است از علت‌های تفسیر نادرست، درمورد تغییرات زمانی و مکانی شاخص‌های محیطی باشند (Houari et al., 2016). این داده‌ها به‌دلایلی مانند خرابی موقت دستگاه‌های اندازه‌گیری یا تعویض ادوات، قطعی ارتباط، قرائت‌نشدن ازسوی کارشناس، تغییر محل اندازه‌گیری، تغییر قرائت‌کنندگان و پالایش داده‌ها (حذف داده‌های پرت ازسوی سازمان هواشناسی) ایجاد می‌شوند (Hasanpour & Dinpashoh, 2012). تا کنون روش‌های متفاوتی برای ترمیم داده‌های مفقود پیشنهاد شده‌اند که هریک بر اصول ریاضی خاصی بنا شده است. ازجملهٔ آنها، می‌توان به انواع روش‌های آماری و روش‌های هوشمند اشاره کرد. آرمانوس[1] و همکاران (۲۰۲۰)، با بررسی ۲۱ روش کلاسیک درمورد اطلاعات ۳۴سالهٔ پانزده ایستگاه در منطقهٔ اتیوپی، به این نتیجه رسیدند که روش‌های نسبت نرمال، عکس فاصلهٔ وزن‌دار، رگرسیون خطی چندگانه، ضریب همبستگی وزنی و میانگین حسابی مطمئن‌ترین روش‌ها برای برآورد داده‌های گم‌شدهٔ بارش است و از این بین، نسبت نرمال با بیشترین همبستگی و کمترین خطا، مورد اعتمادتر از دیگر روش‌هاست. اغلب پژوهشگران بر داده‌های گم‌شدهٔ بارش متمرکز شده‌اند. از پیشگامان آن می‌توان به ویلموت[2] و همکاران (۱۹۹۴) اشاره کرد که روش میانگین حسابی داده‌های ایستگاه همسایه را به‌منظور تخمین مقادیر ازدست‌رفتهٔ بارندگی، به‌کار بردند. آلفارو و پاچکو[3] (۲۰۲۰) نیز روش‌های برآورد متفاوتی را برای تخمین داده‌های بارندگی ازدست‌رفته استفاده کردند؛ ازجمله مدل رگرسیون خطی و روش نسبت نرمال. به‌تازگی نیز پژوهشگرانی داده‌های بارندگی در مناطق گوناگون جهان را تخمین زده‌اند (Barrios et al., 2018; Romman et al., 2019). در زمینهٔ دیگر پارامترهای هواشناسی، مونس‌خواه[4] و همکاران (۲۰۲۳) از روش‌های

1. Armanuos
2. Willmott
3. Alfaro & Pacheco
4. Mouneskhah

آمار کلاسیک و روش‌های هوشمند بهره بردند تا داده‌های گم‌شدهٔ ساعت‌های آفتابی در شرق دریاچهٔ ارومیه را تخمین بزنند. درمورد داده‌های دمایی نیز، پژوهش‌هایی به‌منظور رفع مشکلات و دادن راهکارهایی برای ازبین‌بردن نواقص داده‌های سری زمانی هواشناسی در سرتاسر جهان انجام شده است؛ از این میان می‌توان به پژوهش کامپوزانو[1] و همکاران (۲۰۱۴) درمورد داده‌های دما و بارش در حوضهٔ رودخانهٔ پاتو (واقع‌در اکوادور) و پژوهش رافی[2] و همکاران (۲۰۱۹) دربارهٔ داده‌های دمایی در کشور بریتانیا اشاره کرد.

در پژوهش حاضر، چهار روش پرکاربرد و مرسوم شامل مختصات جغرافیایی، نسبت نرمال، ضریب همبستگی وزنی و میانگین حسابی، در ترمیم داده‌های مورد بررسی به‌کار گرفته شدند و بهترین روش در منطقهٔ مورد مطالعه معرفی شد. بر این اساس روشی که برتری مطلوبی بر دیگر روش‌ها داشته معرفی و برای تکمیل و ترمیم داده‌های دردسترس توصیه شده است.

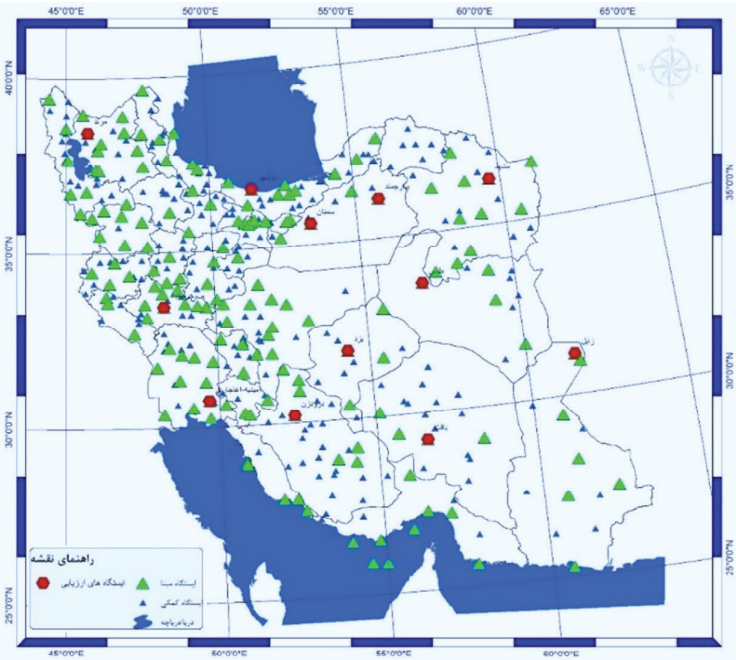
۲- مواد و روش‌ها

این پژوهش بر اطلاعات دمایی کل کشور استوار است و همچون هر پژوهش دیگری، انقطاع و نقص در آمار و اطلاعات آن به‌چشم می‌خورد. به‌منظور اصلاح و تکمیل سری زمانی اطلاعات دردسترس، علاوه‌بر ایستگاه‌های موجود، اطلاعات آماری کل ایستگاه‌های سینوپتیک کشور تهیه شد و ایرادهای اطلاعات ایستگاه‌های منتخب، با استفاده از ایستگاه‌های هم‌جوار و متشابه، رفع شد. در این پژوهش، روش‌های گوناگون اصلاح و تکمیل اطلاعات ناقص مطرح می‌شود؛ این روش‌ها ارزیابی شده‌اند و تمامی اطلاعات ناقص و انقطاع آماری در ایستگاه‌های هدف، براساس روش منتخب، محاسبه و تکمیل شده است. در جدول ۱، مشخصات و ویژگی‌های ایستگاه‌های منتخب، به‌همراه وضعیت اقلیمی آنها براساس شاخص اقلیم نمای دومارتن[3] (۱۹۲۶) آورده شده است. شکل ۱ نیز موقعیت مکانی ایستگاه‌های هواشناسی سینوپتیک در گسترهٔ ایران را نشان می‌دهد.

جدول ۱. مشخصات و موقعیت ایستگاه‌های منتخب

ردیف	نام ایستگاه	استان	طول جغرافیایی	عرض جغرافیایی	ارتفاع از سطح دریا (متر)	دمای بیشینه (درجهٔ سلسیوس)	دمای متوسط (درجهٔ سلسیوس)	دمای کمینه (درجهٔ سلسیوس)	متوسط بارش بلندمدت (میلی‌متر)	اقلیم
۱	امیدیه– آغاجاری	خوزستان	49/69	30/74	27	52/6	26/5	-2/4	253/7	خشک
۲	بیارجمند	سمنان	55/81	36/09	1099/3	43/4	16/43	-17/2	123/7	خشک
۳	خرم‌آباد	لرستان	48/28	33/44	1147/8	46/4	17/13	-14/6	475/7	نیمه‌خشک
۴	درودزن	فارس	52/42	30/21	1642	41/6	18/49	-11/2	449/24	نیمه‌خشک
۵	مرند	آذربایجان شرقی	45/77	38/42	1550	39	13/05	-20	383/92	نیمه‌خشک
۶	نوشهر	مازندران	51/47	36/66	-20/9	36/6	16/48	-5/2	1151/1	بسیار مرطوب
۷	یزد	یزد	54/29	31/9	1230/2	45/6	20/36	-12	51/29	خشک
۸	بافت	کرمان	56/58	29/23	2280	37	15/46	-16/6	249/13	خشک
۹	زابل	سیستان و بلوچستان	61/54	31/09	489/2	49/6	22/53	-11/8	52/32	خشک
۱۰	مشهد	خراسان رضوی	59/63	36/24	999	43/8	15/56	-21	238/48	خشک
۱۱	سمنان	سمنان	53/42	35/59	1127	44/4	18/51	-12/6	121/2	خشک
۱۲	طبس	خراسان شمالی	56/95	33/6	711	49/4	22/95	-6/4	79/48	خشک

1. Campozano
2. Rafii
3. De Martonne



شکل ۱. پراکندگی مکانی ایستگاه‌های هواشناسی مبنا، کمکی دردسترس و منتخب

۱-۲- روش‌های ترمیم داده‌ها

بسیاری از روش‌های کلاسیک در زمینهٔ تکمیل و تخمین اطلاعات معرفی و توصیه شده‌اند که چهار روش پرکاربرد از بین آنها بررسی و ارزیابی شده است. در ادامه، روش‌های مورد استفاده معرفی می‌شود.

۱-۱-۲- مختصات جغرافیایی[1] (روش گرافیکی)

ازجمله روش‌های مورد استفاده در بازسازی داده‌های گم‌شده، روش مختصات جغرافیایی یا روش گرافیکی است. در این روش، پس‌از تعیین موقعیت ایستگاه‌های منطقهٔ مورد مطالعه روی نقشهٔ توپوگرافی با استفاده از مختصات جغرافیایی این ایستگاه‌ها، برای مدل‌سازی، ایستگاه هدف داده به‌منزلهٔ مرکز مختصات جغرافیایی درنظر گرفته و مختصات هریک از ایستگاه‌های اطراف آن به‌نسبت این مختصات جغرافیایی تعیین شد. بدیهی است که ایستگاه‌های نزدیک‌تر به ایستگاه مدنظر سهم بیشتری در بازسازی آن خواهند داشت؛ بنابراین باید ضریب وزنی بزرگ‌تری به آن اختصاص یابد. این ضریب وزنی با استفاده از رابطه (۱) محاسبه می‌شود.

$$W = \frac{1}{x^2+y^2}$$ رابطه (۱)

x و y، به‌ترتیب، طول و عرض مختصاتی ایستگاه‌اند. درنَهایت، داده‌های گم‌شده در ایستگاه هدف با استفاده از رابطه (۲) به‌دست می‌آید.

$$N_x = \frac{\sum_{i=1}^{n} W_i N_i}{\sum_{i=1}^{n} W_i}$$ رابطه (۲)

N_x مقدار برآوردشدهٔ دادهٔ گم‌شده در ایستگاه x، N_i مقدار دادهٔ موجود در ایستگاه i و n معرف تعداد ایستگاه‌هایی است که برای برآورد دادهٔ گم‌شده، از داده‌های آنها استفاده شده است.

1. Geographical Coordinates (GC)

۲-۱-۲- نسبت نرمال[1]

روش نسبت نرمال را ابتدا پولهاس و کوهلر[2] (۱۹۵۲)، برای تخمین داده‌های گم‌شدهٔ بارندگی به‌کار بردند و سپس یانگ[3] (۱۹۹۲) آن را اصلاح کرد. در این روش، اغلب میانگین نسبت داده‌های بین ایستگاه‌های شاهد و ایستگاه هدف نقش مهمی دارد. با استفاده از این شیوه، دما در ایستگاه هدف متناسب است با نسبت میانگین دما در ایستگاه هدف به میانگین ساعت آفتابی در ایستگاه‌های شاهد، ضرب‌در دمای هم‌زمان ایستگاه شاهد که طبق رابطه (۳) محاسبه می‌شود.

$$N_x = \frac{1}{n}\sum_{i=1}^{n}\frac{\overline{N}_x}{\overline{N}_i}N_i \qquad \text{رابطه (۳)}$$

در این معادله، $\overline{N}_x$ میانگین داده‌های دما در ایستگاه هدف، $\overline{N}_i$ میانگین داده‌های دما در ایستگاه شاهد iاُم و N_i داده‌های دما در ایستگاه iاُم است.

۲-۱-۳- ضریب همبستگی وزنی[4]

در این روش، به‌منظور برآورد دادهٔ گم‌شده در ایستگاه هدف، از ضرایب همبستگی ایستگاه‌های شاهد استفاده می‌شود. کارآیی این شیوه به قدرت همبستگی بین ایستگاه هدف و ایستگاه‌های اطراف آن بازمی‌گردد. برای برآورد دادهٔ گم‌شده با استفاده از این روش، از رابطه (۴) استفاده می‌شود (Teegavarapu & Chandramouli, 2005).

$$N_x = \sum_{i=1}^{n}\left(\frac{(r_i)}{\sum_{i=1}^{n}(r_i)}\right)N_i \qquad \text{رابطه (۴)}$$

r_i ضریب همبستگی پیرسون بین داده‌های دمای ایستگاه هدف و ایستگاه شاهد i است.

۲-۱-۴- روش میانگین حسابی

این روش ساده‌ترین و معمول‌ترین روش در بین سایر روش‌های آماری است که در اصلاح یا تکمیل داده‌ها کاربرد دارد. در این شیوه، با استفاده از داده‌های ایستگاه‌های مجاور و بهره‌گیری از میانگین حسابی آنها، دادهٔ مفقود در ایستگاه هدف براساس رابطه (۵) به‌دست می‌آید.

$$N_x = \frac{1}{n}\sum_{i=1}^{n} N_i \qquad \text{رابطه (۵)}$$

N_x مقدار برآوردشدهٔ دادهٔ گم‌شده در ایستگاه x، N_i مقدار دادهٔ موجود در ایستگاه i و n معرف تعداد ایستگاه‌هایی است که برای برآورد دادهٔ گم‌شده، از داده‌های آنها استفاده شده است.

۲-۲- معیارهای ارزیابی نتایج

برای ارزیابی دقت و کارآیی روش‌های هوشمند و روش‌های آماری در مدل‌سازی دما، نتایج به‌دست‌آمده با مقادیر واقعی مقایسه شد. بدین‌منظور از شاخص‌های آماری ضریب تبیین (R^2)، جذر میانگین مربعات خطا (RMSE)، میانگین انحراف مطلق (MAD) برای تعیین میزان ضریب تبیین بین مقادیر دمای واقعی و مقادیر برآوردشده و نیز تعیین مقادیر خطای مدل‌سازی استفاده شد (روابط (۶) تا (۸)).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}\right)^2 \qquad \text{رابطه (۶)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}} \qquad \text{رابطه (۷)}$$

$$MAD = \frac{\sum_{i=1}^{n}|x_i-y_i|}{n} \qquad \text{رابطه (۸)}$$

در این معادلات، x_i و y_i به‌ترتیب iاُمین دادهٔ مشاهداتی و برآور شده، $\bar{x}$ و $\bar{y}$ به‌ترتیب میانگین داده‌های مشاهداتی و برآوردشده و n تعداد ایستگاه‌های همسایه برای مدل‌سازی است.

۳- نتایج و بحث

۳-۱- ارزیابی روش‌های ترمیم اطلاعات در ایستگاه‌های منتخب

در این پژوهش، به‌منظور بررسی بهترین روش برای تکمیل اطلاعات مفقودی، از اطلاعات ۱۲۵ ایستگاه

1. Normal Ratio (NR)
2. Paulhus & Kohler
3. Young
4. Correlation Coefficient Weighted (CCW)

استفاده شد؛ این ایستگاه‌ها دارای اطلاعات کامل (بدون هیچ‌گونه مفقودی) درطول ۲۱ سال (۲۰۲۰-۲۰۰۰ م.) بودند. ازآنجاکه محاسبات گسترده و زمان‌بر بودند، با انتخاب ۱۰٪ این ایستگاه‌ها به‌صورت تصادفی، عملیات پرکردن اطلاعات روی ایستگاه‌های منتخب انجام شد. ایستگاه‌های منتخب دارای پراکندگی مناسبی در سطح کشورند و موقعیت آنها در شکل ۱ آورده شده است. به‌منظور ارزیابی روش‌های یادشده، اطلاعات ایستگاه‌های منتخب در هر مرحله و به‌صورت جداگانه، حذف و براساس پنج ایستگاه مجاور خود، بازسازی شد. نتایج به‌منظور مقایسهٔ کارآیی روش مورد استفاده، در کنار آمار واقعی هر ایستگاه و با معیارهای ارزیابی ذکرشده، بررسی شد.

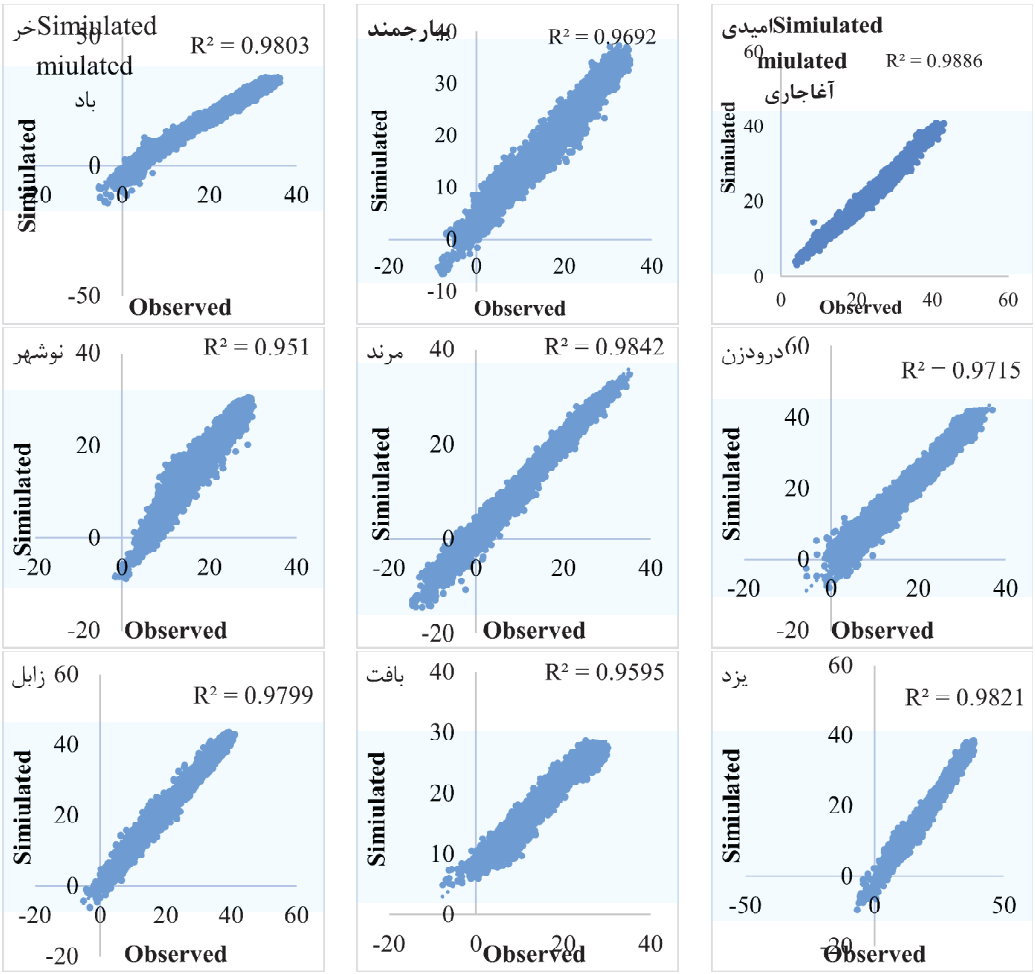
براساس نتایج بازسازی ایستگاه‌های مورد اشاره و با توجه به معیارهای ارزیابی درنظر گرفته‌شده، نتایج طبق جدول ۲ حاصل شد. در این جدول، بهترین نتایج در بین روش‌های متفاوت، به‌صورت توپر نشان داده شده است. شایان ذکر است، چون تغییرات دما به‌صورت هماهنگ در محدوده‌های گوناگون تغییر می‌کند، مقدار ضریب تبیین در این موارد بسیار زیاد و اختلاف این معیار در بین روش‌های متفاوت، اندک است؛ بنابراین مقادیر RMSE و MAD در شناسایی روش منتخب، دارای اهمیت‌اند. با توجه به نتایج، درمجموع مقبولیت روش نسبت نرمال بیشتر از دیگر روش‌های مورد استفاده است. در ادامهٔ پژوهش، به‌منظور تکمیل و رفع نقص اطلاعات ایستگاه‌ها، این روش به‌کار رفته است؛ هرچند روش‌های دیگر را نیز، با توجه به مقدار خطای اندک و تفاوت ناچیز با سایر روش‌ها در اصلاح داده‌ها، می‌توان استفاده کرد. رافی و همکاران (۲۰۱۹) نیز در پژوهشی که در کشور بریتانیا به‌منظور رفع نواقص داده‌های سری زمانی دما انجام دادند، بین روش‌های آماری کلاسیک، روش نسبت نرمال و مختصات جغرافیایی را به‌منزلهٔ روش مورد اعتماد ارزیابی کرده‌اند. با توجه به مقادیر متوسط حاصل از ارزیابی نتایج، روش نسبت نرمال، ضریب همبستگی وزنی، مختصات جغرافیایی و میانگین حسابی، به‌ترتیب با مقدار RMSE

جدول ۲. نتایج معیارهای ارزیابی عملکرد روش‌های مورد استفاده در ترمیم داده‌های مفقودی ایستگاه‌های منتخب

نام ایستگاه	CCW			AVG5			GC			NR		
	MAD	RMSE	R2	MAD	RMSE	R2	MAD	RMSE	R2	MAD	RMSE	R2
امیدیه–آغاجاری	**0/900**	**1/177**	0/988	0/906	1/185	0/988	1/103	1/449	0/984	1/536	1/798	**0/989**
بیارجمند	**1/794**	**2/198**	**0/977**	1/830	2/245	0/975	2/048	2/456	0/974	2/764	3/156	0/969
خرم‌آباد	2/117	2/465	**0/980**	2/119	2/467	0/980	2/190	2/555	0/979	**1/543**	**1/980**	0/980
درودزن	2/279	2/627	0/972	2/28	2/628	0/971	**1/732**	**2/065**	**0/973**	3/587	4/236	0/971
مرند	**1/130**	**1/437**	0/984	1/131	1/438	0/984	1/239	1/562	**0/985**	1/352	1/642	0/984
نوشهر	2/647	3/130	**0/957**	2/747	3/238	0/955	**2/573**	**3/107**	0/955	2/742	3/410	0/951
یزد	3/187	3/449	**0/984**	3/190	3/453	0/984	3/331	3/594	0/983	**1/475**	**1/891**	0/982
بافت	8/991	9/132	**0/965**	9/052	9/195	0/964	9/897	10/072	0/955	**3/638**	**4/416**	0/959
زابل	2/257	2/625	0/980	2/274	2/644	0/979	3/263	3/711	0/968	**1/979**	**2/378**	**0/980**
مشهد	**1/077**	1/416	0/984	1/078	1/418	0/984	1/258	1/587	**0/985**	1/084	**1/353**	0/984
سمنان	3/859	4/382	0/973	**3/849**	**4/384**	0/972	3/857	4/412	0/971	5/865	7/117	**0/977**
طبس	5/062	5/332	0/979	5/063	5/334	0/979	5/326	5/600	0/978	**2/848**	**3/188**	**0/980**
متوسط مقادیر	2/942	3/281	0/977	2/960	3/302	0/976	3/151	3/514	0/974	2/534	3/047	0/976

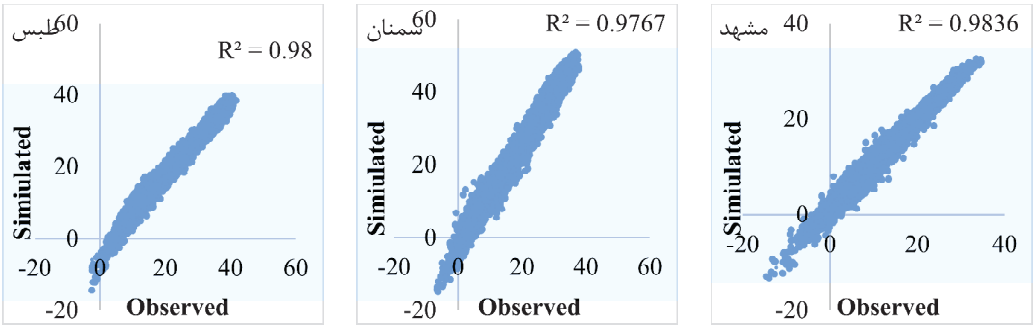
معـادل ۳/۰۵، ۳/۲۸، ۳/۳۰ و ۳/۵۱ درجـهٔ سلسـیوس اولویت‌بندی می‌شوند. در سایر مناطق دنیـا نیـز، نتـایج تاحدودی مشابه در ارزیابی روش‌های یادشده، به‌دسـت آمده و برای نمونه، براساس پژوهش شبالالا[1] و همکاران (۲۰۱۹) که در منطقهٔ لیمپوپو افریقا درمـورد داده‌هـای دمایی انجام شد، روش نسبت نرمال درمقایسه بـا روش میـانگین حسـابی و ضـریب همبسـتگی، دارای برتـری نسبی بوده است.

در شـکل ۲ بـرای ایسـتگاه‌هـای منتخـب، مقـادیر محاسباتی ازطریق روش نسبت نرمال درمقابـل مقـادیر مشاهداتی، ترسیم و ارائه شده است. براساس این شکل، در تمامی ایستگاه‌های مورد بررسـی، مقـادیر مشـاهداتی و بازسازی‌شده دارای همبستگی بالایی است (ضریب تبیـین بالای بین ۰/۹۵۱ تـا ۰/۹۸۹) کـه مقبولیـت روش نسـبت نرمال درمورد تخمین داده‌ها را می‌رساند. شایان ذکر اسـت که هرچـه تـراکم ابـر ایجادشـده از داده‌هـا بیشـتر باشـد، کیفیت تخمین داده‌هـا نیـز بیشـتر اسـت؛ در ایـن شـکل، تـراکم بـالای ابـر ایجادشـده در تمـامی ایسـتگاه‌هـا دیـده می‌شود. این درحالی اسـت کـه بـه‌رغـم بـالابودن مقـادیر همبستگی در سایر روش‌ها، مقادیر جذر میـانگین مربعـات خطا و میانگین انحراف مطلق، به‌طور کلی در روش نسـبت نرمال، مقادیری مناسب‌تر از دیگر روش‌ها دارد.



شکل ۲. مقایسهٔ مقادیر مشاهداتی و محاسباتی نسبت نرمال برای ایستگاه‌های مورد بررسی

1. Shabalala



ادامهٔ شکل ۲.

۴- نتیجه‌گیری کلی

با توجه به اهمیت کیفیت و صحت داده‌ها و همچنین وجودنداشتن داده‌های مفقودی در سری‌های زمانی، به‌منظور بررسی هرچه دقیق‌تر پدیده‌های هواشناسی و نیز تصمیم‌گیری دربارهٔ توسعهٔ پایدار منابع طبیعی و مدیریت آن، تخمین داده‌های گم‌شده یکی از ضروری‌ترین مباحث در آغاز روند تحقیق و مطالعه است زیرا وجود نقص اطلاعاتی، در پایهٔ پژوهش‌ها، به تصمیم‌گیری نادرست دربارهٔ پدیده و هدف مورد نظر منجر خواهد شد. اطلاعات دمایی روزانه ایستگاه‌های سینوپتیک نیز از بااهمیت‌ترین داده‌های اقلیمی‌اند که بنابر نیاز، در بسیاری از زمینه‌های صنعتی، کشاورزی و اجتماعی کاربرد دارند. همچنین شایان ذکر است که تکمیل و رفع خلأ آماری، براساس روش‌هایی که اعتبار آنها ارزیابی نشده است، چه‌بسا به تولید داده‌های نادرست و خارج از واقعیت بینجامد؛ بنابراین در این پژوهش، روش‌های مرسوم آمار کلاسیک، اعم‌از مختصات جغرافیایی (گرافیکی)، نسبت نرمال، ضریب همبستگی وزنی و میانگین حسابی بررسی و ارزیابی شد که هدف از آن اصلاح و ترمیم داده‌های مفقودی متوسط دمای روزانهٔ ایستگاه‌های هواشناسی بود. بدین‌ترتیب مشخص شد که در میان روش‌های مورد بررسی، روش نسبت نرمال با متوسط جذر میانگین مربعات خطای ۳/۰۴۷ مقبولیت و کیفیتی بیشتر از دیگر روش‌ها دارد و بنابراین توصیه می‌شود در محدودهٔ مطالعاتی مشابه در پژوهش‌های آتی، این روش به‌کار رود. در مراتب بعدی، به‌ترتیب روش مختصات جغرافیایی، همبستگی وزنی و میانگین حسابی قرار دارند. شایان توجه است، با اینکه سایر روش‌ها در اولویت‌های بعدی واقع شده‌اند، همچنان در برخی ایستگاه‌ها کارآیی مطلوبی نشان می‌دهند؛ بنابراین در شرایط متفاوت، روش‌های متنوعی می‌توانند نیاز به ترمیم داده‌ها را مرتفع سازد و باید، با توجه به محدودهٔ مورد مطالعه، بهترین روش انتخاب و استفاده شود.

۵- تشکر و قدردانی

با سپاس از سازمان هواشناسی کشور که اطلاعات دمایی مورد نیاز را به‌منظور پیشبرد مقالهٔ تحقیقاتی حاضر، دردسترس نگارندگان قرار داده است.

۶- منابع

Alfaro, R. & Pacheco, R., 2000, **Aplicación de algunos métodos de relleno a series anuales de lluvia de diferentes regiones de Costa Rica**, Tópicos Meteorológicos y Oceanográficos, 7(1), PP. 1-20.

Armanuos, A.M., Al-Ansari, N. & Yaseen, Z.M., 2020, **Cross Assessment of Twenty-One Different Methods for Missing Precipitation Data Estimation**, Atmosphere, 11(4), PP. 1-34.

Asseng, S., Foster, I. & Turner, N.C., 2011, **The Impact of Temperature Variability on Wheat Yields**, Global Change Biology, 17, PP., 997-1012.

Bannayan, M., Crout, N. & Hoogenboom, G., 2003, **Application of the CERES-Wheat Model for within-Season Prediction of Winter Wheat Yield in the United Kingdom**, Agronomy Journal, 95, PP. 114-125.

Barrios, A., Trincado, G. & Garreaud, R., 2018, **Alternative Approaches for Estimating Missing Climate Data: Application to Monthly Precipitation Records in South-Central Chile**, Forest Ecosystems, 5(1), PP. 1-10.

Campozano, L., Sánchez, E., Avilés, Á. & Samaniego, E., 2014, **Evaluation of Infilling Methods for Time Series of Daily Precipitation and Temperature: The Case of the Ecuadorian Andes**, Maskana, 5(1), PP. 99-115.

De Martonne, E., 1926, **Aerisme, et índices d'aridite**, Comptesrendus de L'Academie des Sciences, 182, PP. 1395-1398.

Hasanpour Kashani, M. & Dinpashoh, Y., 2012, **Evaluation of Efficiency of Different Estimation Methods for Missing Climatological Data**, Stochastic Environmental Research and Risk Assessment, 26(1), PP. 59-71.

Houari, R., Bounceur, A., Kechadi, M.T., Tari, A.K. & Euler, R., 2016, **Dimensionality Reduction in Data Mining: A Copula Approach**, Expert Systems with Applications, 64, PP. 247-260.

Mouneskhah, V., Khaledi, M., Hadi, M. & Samadianfard, S., 2023, **Comparison of the Efficiency of Intelligent and Statistical Methods in the Reconstruction of Sunshine Hours Data (Case Study: East of Urmia Lake Basin)**, Journal of Agricultural Meteorology, 10(2), PP. 28-36.

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C. & Takáč, J., 2011, **Simulation of Winter Wheat Yield and Its Variability in Different Climates of Europe: A Comparison of Eight Crop Growth Models**, European Journal of Agronomy, 35, PP. 103-114.

Paulhus, J.L. & Kohler, M.A., 1952, **Interpolation of Missing Precipitation Records**, Monthly Weather Review, 80(8), PP. 129-133.

Rafii, F. & Kechadi, T., 2019, **Collection of Historical Weather Data: Issues with Missing Values**, Proceedings of the 4th International Conference on Smart City Application.

Romman, Z.A., Al-Bakri, J.T. & Al Kuisi, M.M., 2019, **Estimation of Rainfall Missing Data in an Arid Area Using Spatial and EM Methods**, Journal of Software Engineering & Applications, 9, PP. 76-80.

Shabalala, Z.P., Moeletsi, M.E., Tongwane, M.I. & Mazibuko, S.M., 2019, **Evaluation of Infilling Methods for Time Series of Daily Temperature Data: Case Study of Limpopo Province, South Africa**, Climate, 7(7), P. 86.

Simolo, C., Brunetti, M., Maugeri, M. & Nanni, T., 2010, **Improving Estimation of Missing Values in Daily Precipitation Series Bya Probability Density Function-Preserving Approach**, International Journal of Climatology, 30, PP. 1564-1576.

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(1-4), PP. 191-206.

Wilhite, D.A., 2000, **Drought as a Natural Hazard: Concepts and Definitions**, Drought: A Global Assessment, Springer, London, Routledge.1(1), PP. 3-18.

Willmott, C.J., Robeson, S.M. & Feddema, J.J., 1994, **Estimating Continental and Terrestrial Precipitation Averages from Rain-Gauge Networks**, International Journal of Climatology, 14(4), PP. 403-414.

Young, K.C., 1992, **A Three-Way Model for Interpolating for Monthly Precipitation Values**, Monthly Weather Review, 120(11), PP. 2561-2569.

Zare khormizie, H. & Ghafarian Malamiri, H.R., 2020, **Effect of Height and Temperature on Plant Phenological Processes Using Harmonic Analysis of MODIS NDVI Time Series (Case Study: Shirkouh, Yazd Province)**, Iranian Journal of Remote Sensing & GIS, 12(3), PP. 1-22.