Scientific-Promotional (Professional)Journal of Aquifer

Homepage: https://abkhansj.ut.ac.ir

The Impact of Data Reconstruction Methods on Climate Indicators

Javad Bazrafshan[1], Hamidreza Ghasemi*[2]

Abstract

The importance of choosing appropriate methods for data reconstruction is clearly evident. Given the increasing spread of climate change and its effects on plants and other living organisms, the need for accurate and reliable data is felt more than ever. Therefore, researchers and decision-makers should pay special attention to analyzing and selecting data reconstruction methods so that they can take steps towards optimal resource management and reducing the negative effects of climate change. The aim of this study was to investigate the effect of reconstructing random statistical gaps in the daily minimum and maximum temperature data of the Kohrang station (base station) using the multiple linear regression method and with the help of neighboring stations. For this purpose, 5 to 50 percent (with an interval of 5 percent) of the daily minimum and maximum temperature data at the Kohrang station were randomly considered statistical gaps. Then, data gaps were calculated using the multiple linear regression method and the estimation error was determined for different percentages of statistical gaps. Reconstruction errors were reviewed and evaluated on a monthly and annual basis.

Keywords: *Climate indicators, data gaps, data reconstruction methods.*

[1] Javad Bazrafshan- Associate Professor, University of Tehran.

[2] Hamidreza Ghasemi- Student at University of Tehran (*Corresponding author, email: ghasemi.hamidreza@ut.ac.ir)

نشریه علمی – ترویجی آبخوان (دوره ۱۶، شماره ۲)

Homepage: https://abkhansj.ut.ac.ir

تأثیر روش‌های بازسازی داده‌ها بر روی شاخص‌های اقلیمی

جواد بذرافشان[۱]، حمیدرضا قاسمی[۲*]

چکیده

اهمیت انتخاب روش‌های مناسب برای بازسازی داده‌ها به وضوح مشخص است. با توجه به گسترش روزافزون تغییرات اقلیمی و تأثیرات آن بر گیاهان و دیگر موجودات زنده، نیاز به داده‌های دقیق و قابل اعتماد بیش از پیش احساس می‌شود. بنابراین، پژوهشگران و تصمیم‌گیرندگان باید به تحلیل و انتخاب روش‌های بازسازی داده‌ها توجه ویژه‌ای داشته باشند تا بتوانند در مسیر مدیریت بهینه منابع و کاهش اثرات منفی تغییرات اقلیمی گام بردارند.هدف از اجرای این مطالعه بررسی تاثیر بازسازی خلاءهای آماری تصادفی در داده‌های دمای حداقل و حداکثر روزانه ایستگاه کوهرنگ (ایستگاه مبنا) با استفاده روش رگرسیون خطی چندگانه و به کمک ایستگاههای همجوار بود. برای این منظور، به طور تصادفی، ۵ تا ۵۰ درصد (با فاصله ۵ درصد) از داده‌های دمای حداقل و حداکثر روزانه در ایستگاه کوهرنگ به طور تصادفی خلاء آماری در نظر گرفته شد. سپس، خلاءهای داده با روش رگرسیون خطی چندگانه محاسبه و خطای براورد به ازای درصدهای مختلف خلاء آماری تعیین گردید. خطاهای بازسازی به صورت ماهانه و سالانه مورد بررسی و ارزیابی قرار گرفت.

واژه‌های کلیدی: *شاخص‌های اقلیمی، خلاء داده ،روش‌های بازسازی داده.*

[۱] دانشیار گروه مهندسی آبیاری و ابادانی دانشگاه تهران.

[۲] دانشجوی کارشناسی ارشد گروه مهندسی آبیاری و آبادانی دانشگاه تهران (* نویسنده مسئول، ایمیل: ghasemi.hamidreza@ut.ac.ir)

مقدمه

با پیشرفت فناوری و افزایش نیاز به اطلاعات دقیق در زمینه تغییرات اقلیمی، روش‌های بازسازی داده‌ها به عنوان ابزاری ضروری برای تحلیل و پیش‌بینی وضعیت اقلیمی شناخته می‌شوند. این روش‌ها به ما این امکان را می‌دهند که داده‌های ناقص یا گمشده را تکمیل کنیم و به این ترتیب، تصویر بهتری از وضعیت اقلیمی در دسترس قرار دهیم. به ویژه در مناطقی که جمع‌آوری داده‌ها به دلیل شرایط جغرافیایی یا اجتماعی دشوار است، این روش‌ها می‌توانند به عنوان یک راه‌حل کارآمد عمل کنند. روش‌های مختلف بازسازی داده‌ها شامل تکنیک‌های مختلفی از جمله میانگین‌گیری، رگرسیون و مدل‌های یادگیری ماشین هستند. هر یک از این روش‌ها مزایا و معایب خاص خود را دارند و انتخاب روش مناسب بستگی به نوع داده‌ها و اهداف تحقیق دارد. به عنوان مثال، برخی از روش‌ها ممکن است برای داده‌های زمانی مناسب‌تر باشند، در حالی که دیگران برای داده‌های مکانی کارآمدتر هستند. این انتخاب‌ها تأثیر مستقیمی بر دقت و قابلیت اطمینان داده‌های بازسازی شده خواهند داشت. تأثیر روش‌های بازسازی بر روی شاخص‌های اقلیمی نیز غیرقابل انکار است. شاخص‌هایی مانند دما، بارش و الگوهای جوی به شدت تحت تأثیر کیفیت داده‌ها قرار دارند. با بهبود دقت داده‌های اقلیمی از طریق روش‌های بازسازی، می‌توانیم به تحلیل‌های دقیق‌تری دست یابیم و در نتیجه، پیش‌بینی‌های بهتری درباره تغییرات اقلیمی انجام دهیم. این پیش‌بینی‌ها می‌توانند به برنامه‌ریزی و مدیریت منابع طبیعی و محیط زیست کمک کنند و به ما در مقابله با چالش‌های اقلیمی یاری رسانند. با توجه به این واقعیت که اساس مطالعات پایه در علوم مختلف، به ویژه علوم مرتبط با منابع آب، داده‌های آماری هستند، وجود داده‌های قابل قبول امری ضروری به نظر می‌رسد. با توجه به این که داده‌های موجود به دلایل مختلفی از جمله ثبت نکردن آمار، برداشت غیر دقیق داده‌ها، خرابی وسایل اندازه‌گیری و حذف آمار غلط دچار خلأهای گوناگونی می‌باشند، تخمین و برآورد این داده‌ها امری اجتناب ناپذیر است. دسترسی به داده‌های کافی و دقیق از یک سو موجب کوتاه‌تر شدن مدت مطالعات و از سوی دیگر موجب برآورد دقیق‌تر از پارامترهای هدف و کاهش هزینه‌های اجرایی و خسارات بعدی ناشی از اجرای طرح‌ها می‌گردد (Anderson and Gough, 2018; Barrios et al., 2018).

آمار قدیمی و طولانی‌مدت دمای ماهانه سه شهر و بارش ماهانه پنج شهر ایران توسط سفارت‌های آمریکا و انگلیس در دوره قاجار و قبل از سال ۱۳۳۰ اندازه‌گیری و ثبت شده است. بررسی داده‌های گمشده ماهانه این ایستگاه‌ها نشان می‌دهد که اکثر آن‌ها به دنبال جنگ جهانی اول (۱۹۱۸ و ۱۹۱۹) و جنگ جهانی دوم (۱۹۴۱–۱۹۴۹) رخ داده‌اند. برخی داده‌های گمشده نیز به طور پراکنده در طول دوره آماری وجود دارند. تاکنون ترمیم این داده‌ها در مقیاس ماهانه، به جز دمای مشهد، انجام نشده است.

خلیلی (۱۹۹۱) در طرح جامع آب کشور به منظور بازسازی خلأهای آماری بارندگی، روش تک صفحه‌ای و صفحات مکرر را پیشنهاد داد. ماسیتی (۲۰۱۴) تاثیر بازسازی خلاءهای آماری تصادفی و متوالی را بر محاسبه شاخص‌های دمایی ماهانه با استفاده از سه روش آماری (مدل

نظری، مدل خطی و مدل غیرخطی) مورد بررسی و مقایسه قرار داد. براکا و همکاران (۲۰۲۴) روشی برای افزایش اطمینان به نتایج تجزیه و تحلیل آماری داده های دمای حداقل روزانه بازسازی شده در منطقه مدیترانه (منطقه کالابریا درجنوب ایتالیا) ارائه دادند (Khalili, 1991; Massetti, 2014; Barca et al., 2024).

مواد و روش‌ها

الف) روش وزن‌دهی معکوس فاصله (IDW)

ب) روش وزنی معکوس فاصله تعدیل‌شده (IDWm)

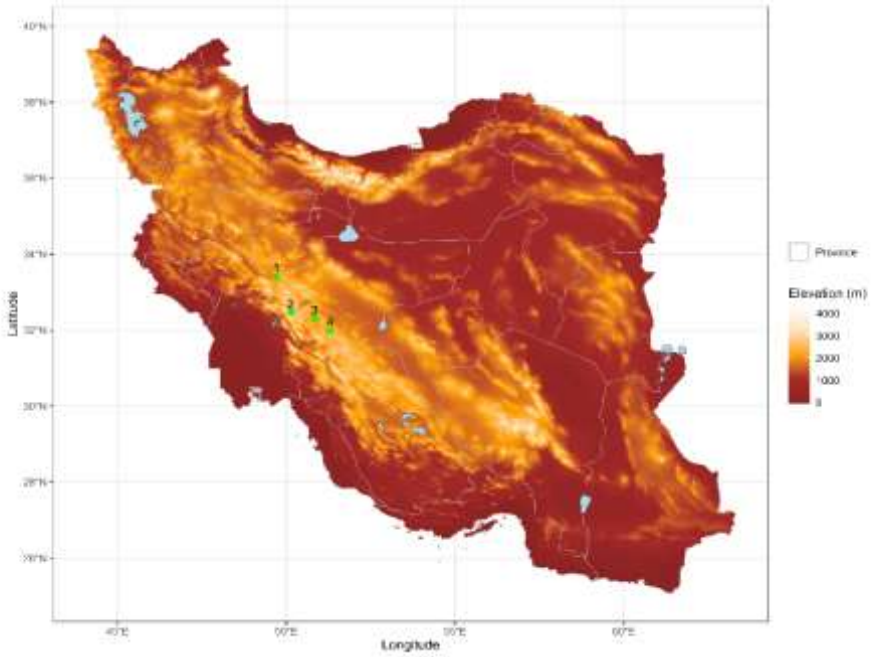
ج) وزن‌دهی ضریب همبستگی (CCW)

د) رگرسیون خطی چندگانه (MLR)

و) شبکه‌های عصبی مصنوعی (ANN)

موقعیت منطقه مطالعاتی در شکل ۱ نشان داده شده است. چهار ایستگاه هواشناسی الیگودرز، کوهرنگ، شهرکرد و بروجن با مشخصات مندرج در جدول ۱ برای این تحقیق انتخاب گردید. هر چهار ایستگاه در ارتفاعات بالاتر از ۲۰۰۰ متر قرار گرفته و کوهرنگ با ارتفاع ۲۳۶۵ متر مرتفع‌ترین ایستگاه مطالعاتی را تشکیل می‌دهد و مقدار بارندگی متوسط سالانه آن ۱۲۳۷ میلی‌متر و میانگین حداقل دما ۹/۵- درجه سلسیوس است. برای انجام این مطالعه از داده‌های دمای حداقل و حداکثر روزانه ۲۰۱۹-۱۹۹۲ استفاده شد. ایستگاه‌ها فاقد خلاء آماری هستند.

هدف از اجرای این مطالعه، بررسی تاثیر بازسازی خلاءهای آماری تصادفی در داده‌های دمای حداقل و حداکثر روزانه ایستگاه کوهرنگ (ایستگاه مبنا) با استفاده روش رگرسیون خطی چندگانه و به کمک ایستگاههای همجوار بود. خطاهای بازسازی به صورت ماهانه و سالانه مورد بررسی و ارزیابی قرار گرفت.



شکل ۱- موقعیت ایستگاه‌های مطالعاتی (شماره ایستگاه‌ها متناظر با نام ایستگاه‌ها در جدول ۱ است).

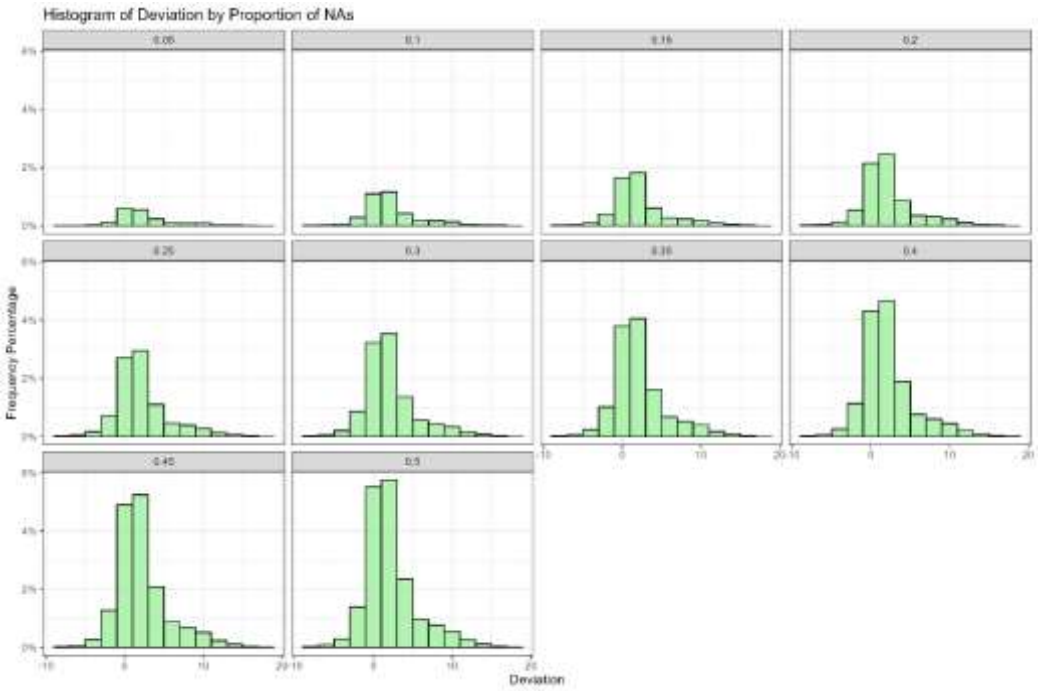
جدول ۱- مشخصات ایستگاههای منتخب

Number	Station	Lat	Long	Altitude	P	T	Tmin	D.C.	Humdity Class	Thermal Class
1	Aligudarz	33.41	49.7	2022.1	391.8	12.8	-5.7	17.2	Semi-arid	Cold
2	Kuhrang	32.46	50.13	2365	1237.2	9.9	-9.5	62.1	Per-humid B	Very Cold
3	Shahrekord	32.29	50.84	2048.9	299	11.5	-7.7	13.9	Semi-arid	Very Cold
4	Borujen	31.98	51.3	2260	247.6	11.2	-6.8	11.7	Semi-arid	Cold

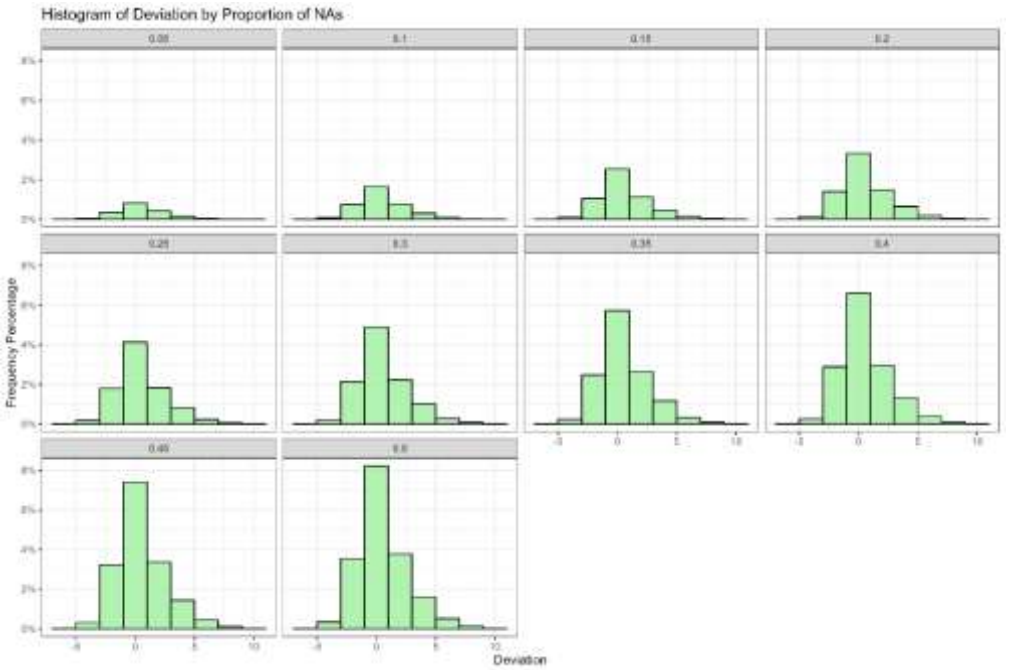
نتیجه‌گیری

الف) تاثیر بازسازی داده‌ها بر توزیع فراوانی خطا

شکل‌های ۲ و ۳ هیستوگرام فراوانی خطای براورد دمای حداقل و دمای حداکثر را در ایستگاه کوهرنگ نشان می‌دهد. با توجه به شکل‌ها مشاهده می‌شود که با افزایش درصد خلاءها، فراوانی خطاها نیز افزایش می‌شود. با این حال، بیشترین خطای براورد حول و حوش صفر است که از این نظر ایده آل محسوب می‌شود. در درصدهای بالاتر خلاء، انحرافات بزرگی تا ۲۰ درجه سلسیوس (در دمای حداقل) و تا ۱۰ درجه سلسیوس (در دمای حداکثر) پیرامون میانگین دیده می‌شود.



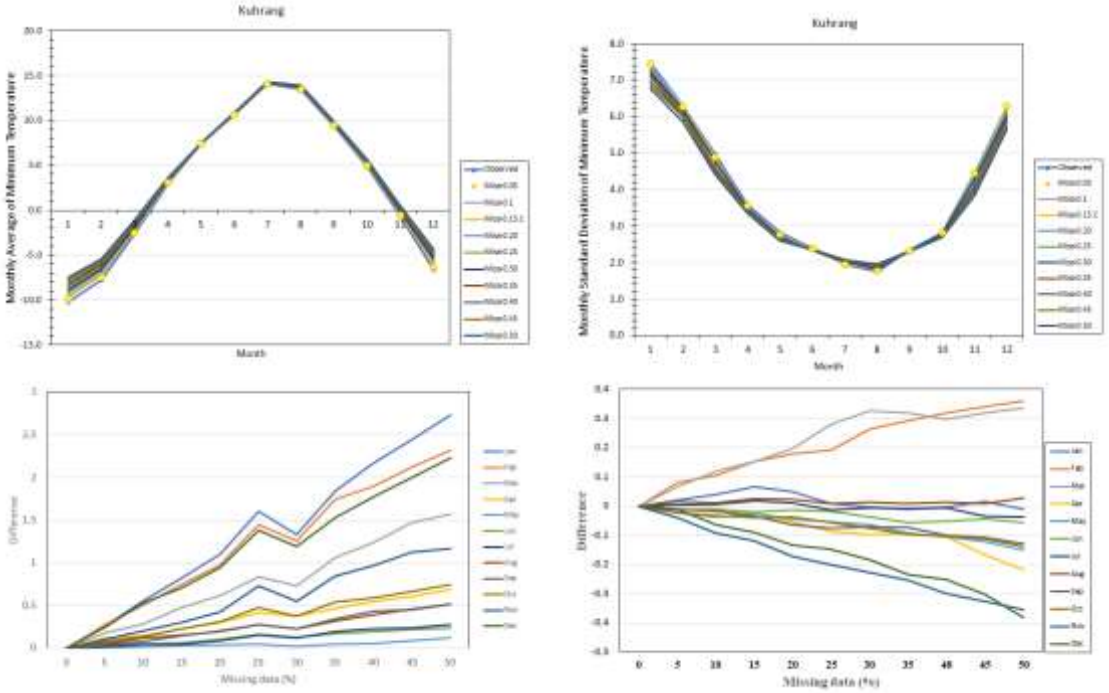
شکل ۲- هیستوگرام فراوانی خطاهای براورد داده مفقود دمای حداقل در ایستگاه کوهرنگ.



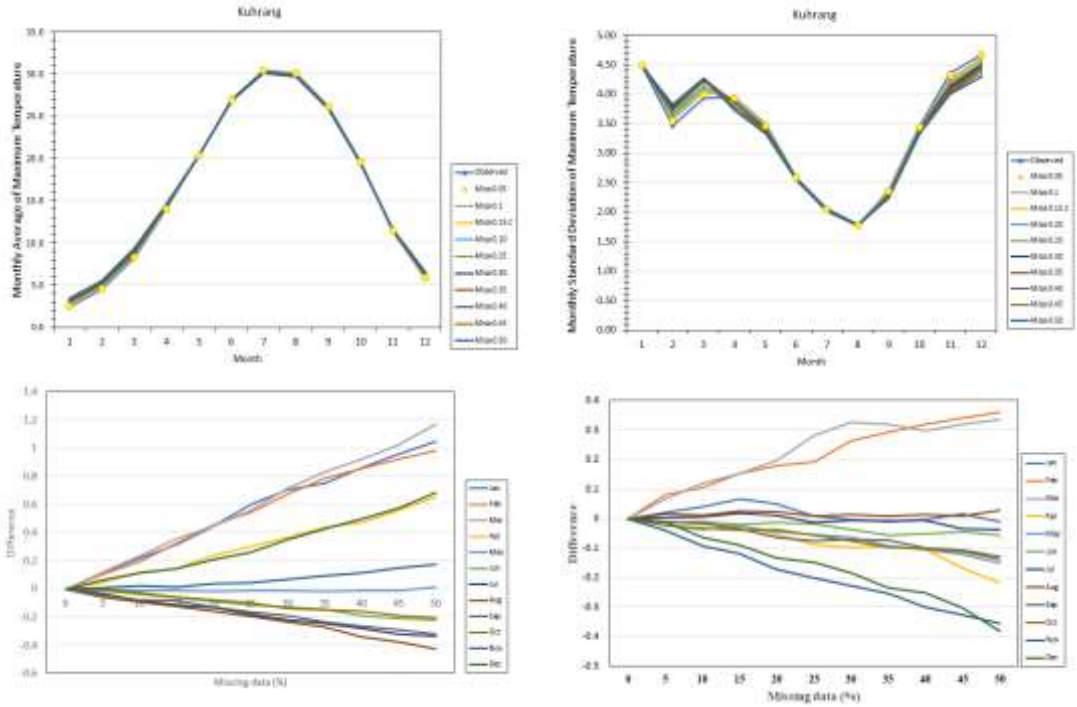
شکل ۳- هیستوگرام فراوانی خطاهای براورد داده مفقود دمای حداکثر در ایستگاه کوهرنگ.

ب) تاثیر بازسازی داده‌ها بر میانگین و انحراف معیار ماهانه

شکل‌های ۴ و ۵ میانگین (بالا سمت چپ)، انحراف معیار (بالا سمت راست)، اختلاف میانگین‌ها (پایین سمت چپ) و اختلاف انحراف معیارهای مقادیر بازسازی شده و مشاهده شده برای دمای حداقل و حداکثر را نشان می‌دهند. مشاهده می‌شود که اختلاف میانگین‌ها و انحراف معیارها در ماه‌های دسامبر تا مارس بیشتر از دیگر ماه‌های سال است. این اختلافات در ماه ژانویه در درصدهای مختلف بازسازی داده از سایر ماه‌ها بیشتر است و در درصد بازسازی ۵۰ درصد تا ۲/۵ درجه سلسیوس در دمای حداقل و تا ۱/۲ درجه سلسیوس در دمای حداکثر می‌رسد که بسیار قابل توجه است. تا درصد بازسازی ۲۰ درصد، اختلافات در میانگین دمای حداقل، کمتر از ۱ درجه سلسیوس و در دمای حداکثر کمتر از ۰/۵ درجه سلسیوس است.



شکل ۴- میانگین (بالا سمت چپ)، انحراف معیار (بالا سمت راست)، اختلاف میانگین‌ها (پایین سمت چپ) و اختلاف انحراف معیارهای مقادیر بازسازی شده و مشاهده شده مربوط به دمای حداقل روزانه در ایستگاه کوهرنگ



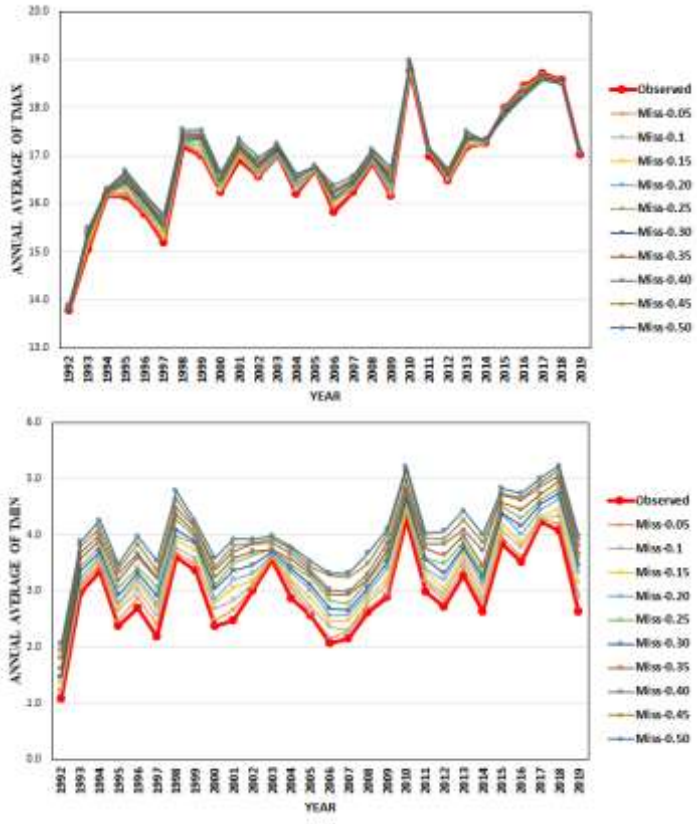
شکل ۵- میانگین (بالا سمت چپ)، انحراف معیار (بالا سمت راست)، اختلاف میانگین‌ها (پایین سمت چپ) و اختلاف انحراف معیارهای مقادیر بازسازی شده و مشاهده شده مربوط به دمای حداکثر روزانه در ایستگاه کوهرنگ

ج) تاثیر بازسازی داده‌ها بر میانگین سالانه دما

سری‌های متوسط سالانه دمای حداقل (نمودار پایینی) و حداکثر (نمودار بالایی) در شکل ۶ نشان داده شده است.

د) نتیجه‌گیری مطالعه موردی

در این مطالعه، تأثیر بازسازی خلاءهای آماری در داده‌های دمای حداقل و حداکثر ایستگاه کوهرنگ با استفاده از رگرسیون خطی چندگانه بررسی شد. نتایج نشان داد اختلاف میانگین‌ها و انحراف معیارها در ماه‌های سرد سال (دسامبر تا مارس) بیشتر از سایر ماه‌ها بوده و در ماه ژانویه این اختلافات به حداکثر خود می‌رسد. به‌طور کلی، دمای حداقل سالانه نسبت به درصد خلاءهای آماری حساسیت بیشتری دارد در حالی که دمای حداکثر سالانه حساسیت کمتری نشان می‌دهد.



شکل ۶- میانگین سالانه دمای حداکثر (بالا) و حداقل (پایین) مشاهده شده و بازسازی شده با درصدهای ۵ تا ۵۰ درصد خلاء آماری در ایستگاه کوهرنگ

REFERENCES

Anderson, C. I., & Gough, W. A. (2018). Accounting for missing data in monthly temperature series: Testing rule-of-thumb omission of months with missing values. *International Journal of Climatology,* 38, 4990–5002.

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, 28.

Khalili, A. (1991). Integrated Water Plan of Iran. *Jamab Consulting Engineering*

Co., The Ministry of Energy, Tehran, 111-122. (In Persian).

Massetti, L. (2014). Analysis and estimation of the effects of missing values on the calculation of monthly temperature indices. *Theoretical and Applied Climatology,* 117, 511–519.

Barca, E., Guagliardi, I., & Caloiero, T. (2024). A methodological approach for filling the gap in extreme daily temperature data: An application in the Calabria region (Southern Italy). *Theoretical and Applied Climatology,* 155, 7447–7461.