Research Paper

An Innovative Approach for Estimating Missing Monthly Rainfall Data in the Southern Baluchestan Basin, Iran

Khorshid Karim[1], Hamid Nazaripour[2*] , Mahmood Khosravi[3]Seyed Mahdi Amir Jahanshahi[4]

[1] Master's student, Department of physical Geography, Faculty of Geography and Environmental Planning, University of Sistan and Baluchestan, Zahedan, Iran.
[2] Assistant Professor, Department of physical Geography, Faculty of Geography and Environmental Planning, University of Sistan and Baluchestan, Zahedan, Iran.
[3] Professor, Department of physical Geography, Faculty of Geography and Environmental Planning, University of Sistan and Baluchestan, Zahedan, Iran.
[4] Assistant Professor, Department of Statistics, Faculty of Mathematics, Statistics and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran.

10.22125/IWE.2023.422204.1761

Received:
October 24, 2023
Accepted:
March 2, 2024
Available online:
May 5, 2024

Keywords:
missing rainfall data, spatial interpolation, inverse distance, normal ratio, similarity index

Abstract

Precipitation is a parameter related to climate and research on it is facing problems due to the lack of continuous data. Accurate planning and management of water resources depends on the availability of continuous and accurate precipitation data at meteorological stations. The presence of missing values can lead to biased results and may create obstacles in the analysis. Hence, estimation of missing rainfall data is important in order to obtain more reliable results. Various imputation methods have been proposed and developed by researchers to estimate missing values in daily and monthly rainfall data. In most situations, spatial interpolation techniques such as normal ratio and inverse distance methods are used for estimating missing rainfall values at a particular target station based on the available rainfall values recorded at the neighboring stations. Moreover, these two methods are found to be very useful in the case where the neighboring stations are very close and highly correlated with the target stations. In this study, several modifications and improvements have been proposed to these methods in order to estimate the missing rainfall values at the target station using the information from the nearby stations. The methods have been tested with different percentages of missing rainfall values and also with a radius range of 75 km to 150 km in the catchment area of southern Balochestan. The result indicate that the performance of these modified methods improved the estimation of missing rainfall values at the target station based on the similarity index (S-index), mean absolute error (MAE) and coefficient of correlation (R).

1. Introduction

The availability of complete rainfall data without missing values is important in meteorological, climatological, and hydrological analyses. The problem of missing data is often caused by various reasons. The loss of rainfall data may be due to the relocation of stations, incorrect operation of the measuring device for a certain period of time, and errors in the techniques used in rainfall measurement. The continuity of rainfall records can be affected by factors such as faulty instruments, station displacement, and human negligence during rainfall measurement. These problems affect the estimated

* **Corresponding Author:** Hamid Nazaripour
Address: Department of physical Geography, University of Sistan and Baluchestan, Iran
Email: h.nazaripour@gep.usb.ac.ir
Tel: +989159498907

output and subsequently create incorrect results. To overcome this problem, a number of interpolation techniques have been developed over the decades with the aim of estimating missing observations in climate time series, mainly at monthly, seasonal, and daily time scales. An innovative and reliable approach to selecting estimation methods in order to impute missing rainfall data in Southeast Iran (South Baluchistan catchment area) is presented by this research.

2. Materials and Methods

The data of this research includes monthly rainfall recorded from rain gauge stations in the southern Baluchistan catchment. This research covers the statistical period of 1990 to 2022. Modified spatial weighting methods, including the inverse distance method (IDW), normal ratio method (NR), and correlation coefficient method (CCW), have been selected to impute missing data in the study area. Finally, the performance of the estimation methods has been compared using the similarity index (S-index), mean absolute error (MAE) and coefficient of correlation (R).

3. Results

Both existing and modified methods were used to estimate the missing data from the target station at different spatial intervals. The optimal distance is a distance that has the highest values of similarity index and correlation coefficient, and the lowest average absolute error. A significant difference can be observed between different distances based on error measurement indices. Missing data estimation methods are sensitive to the distance between the target station and neighboring stations. Also, the results of methods for estimating missing data have been evaluated considering different percentages of missing data. The percentage of missing data has a significant impact on missing data estimation methods. When the percentage of missing data for the target stations goes above 20%, the estimation methods perform significantly less.

4. Discussion and Conclusion

In this research, spatial weighting methods including inverse distance weighting methods, normal ratio, and correlation coefficient have been investigated and modified. The results showed that the modified methods are sensitive to the distance as well as the percentage of missing data. The performance of modified spatial weighting methods is always superior to those of existing methods. The most favorable situation for estimating missing data is with a missing percentage of less than 20% and a distance of less than 75 km. The performance improvement percentage of the modified methods has been significant. Therefore, the modified methods are effective in estimating missing data in the study area.

5. Six important references

1) Egigu, M. L. (2020). Techniques of Filling Missing Values of Daily and Monthly Rain Fall Data: A Review. SF J Environ Earth Sci. 2020; 3 (1), 1036.
2) Suhaila, J., Sayang, M. D., & Jemain, A. A. (2008). Revised spatial weighting methods for estimation of missing rainfall data. Asia-Pacific Journal of Atmospheric Sciences, 44(2), 93-104.
3) Sattari, M. T., Rezazadeh-Joudi, A., & Kusiak, A. (2017). Assessment of different methods for estimation of missing data in precipitation studies. Hydrology Research, 48(4), 1032-1044.
4) de Carvalho, J. R. P., Almeida Monteiro, J. E. B., Nakai, A. M., & Assad, E. D. (2017). Model for multiple imputation to estimate daily rainfall data and filling of faults. Revista Brasileira de Meteorologia, 32, 575-583.
5) Paulhus, J. L., & Kohler, M. A. (1952). Interpolation of missing precipitation records. Monthly Weather Review, 80(8), 129-133.
6) Kousari, M. R., Esmaeilzadeh Hosseini, M. S., & Miri, M. (2021). Investigation of the efficiency of methods of infilling missing data in relation to the precipitation parameter in arid regions of Iran. Journal of the Earth and Space Physics, 47(2), 315-332.

Conflict of Interest

We have no conflicts of interest to disclose.

Acknowledgments

We are grateful to anonymous reviewers for their generous suggestions on data analysis and interpretations.

نشریه علمی پژوهشی مهندسی آبیاری و آب ایران
سال چهاردهم. شماره پنجاه و پنج .بهار ۱۴۰۳

رویکرد نوآورانه برآورد داده‌های گمشده بارش ماهانه در حوضه آبریز بلوچستان جنوبی

خورشید کریم[۱]، حمید نظری‌پور[۲]*، محمود خسروی[۳] سید مهدی امیرجهانشاهی[۴]

تاریخ ارسال:۱۴۰۲/۰۸/۰۲
تاریخ پذیرش:۱۴۰۲/۱۲/۱۲

مقاله پژوهشی

چکیده

بارش یک پارامتر مرتبط با اقلیم بوده و تحقیقات در مورد آن به دلیل عدم وجود داده‌های مداوم با مشکل مواجه است. برنامه‌ریزی و مدیریت دقیق منابع آب به وجود داده‌های مداوم و دقیق بارش در ایستگاه‌های هواشناسی بستگی دارد. وجود مقادیر از دست‌رفته می‌تواند منجر به نتایج بایاس‌دار شود و ممکن است موانعی در تجزیه و تحلیل ایجاد کند. از این رو، تخمین داده‌های از دست‌رفته بارندگی به منظور به دست آوردن نتایج قابل اعتمادتر مهم است. روش‌های برآورد مختلفی توسط محققان برای تخمین مقادیر گمشده در داده‌های بارندگی روزانه و ماهانه پیشنهاد و توسعه داده شده است. تکنیک-های درون‌یابی فضایی مانند روش‌های نسبت نرمال و عکس فاصله برای تخمین داده‌های گمشده بارندگی در یک ایستگاه خاص بر اساس مقادیر بارندگی موجود ثبت شده در ایستگاه‌های مجاور استفاده می‌شوند. این روش‌ها در مواردی که ایستگاه‌های همسایه بسیار نزدیک و همبستگی بالایی با ایستگاه‌های هدف دارند، بسیار مفید هستند. در این مطالعه، اصلاحات و بهبودهای متعددی بر روی روش‌های عکس فاصله، نسبت نرمال و ضریب همبستگی صورت گرفته تا با استفاده از اطلاعات ایستگاه‌های مجاور، مقادیر بارندگی گمشده در ایستگاه‌های مورد نظر برآورد گردد. این روش‌ها با درصدهای مختلف مقادیر گمشده بارندگی و همچنین با بُردِ فاصله ۷۵ تا ۱۵۰ کیلومتر در حوضه آبریز بلوچستان جنوبی آزمایش شده‌اند. نتایج نشان می‌دهد که عملکرد این روش‌های اصلاح شده در برآورد مقادیر گمشده بارندگی در ایستگاه هدف بر اساس شاخص شباهت (S-index)، میانگین خطای مطلق(MAE) و ضریب همبستگی (R) بهبود یافته است. برآورد داده‌های گمشده به طول داده‌های گمشده و فاصله از ایستگاه‌های همسایه، حساس است. گمشدگی کمتر از ۲۰ درصد و فاصله کمتر از ۷۵ کیلومتری بهترین برآورد را ایجاد می‌کند.

واژه‌های کلیدی: داده‌های گمشده بارش، درون‌یابی فضایی، عکس فاصله، نسبت نرمال، شاخص شباهت، حوضه بلوچستان جنوبی

.

[۱] دانشجوی کارشناسی ارشد، گروه جغرافیای طبیعی، دانشکده جغرافیا و برنامه ریزی محیطی، دانشگاه سیستان و بلوچستان، ایران.
[۲] استادیار، گروه جغرافیای طبیعی، دانشکده جغرافیا و برنامه ریزی محیطی، دانشگاه سیستان و بلوچستان، ایران. نویسنده مسئول (h.nazaripour@gep.usb.ac.ir)
[۳] استاد، گروه جغرافیای طبیعی، دانشکده جغرافیا و برنامه ریزی محیطی، دانشگاه سیستان و بلوچستان، ایران.
[۴] استادیار، گروه آمار، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه سیستان و بلوچستان، ایران.

مقدمه

در دسترس بودن داده‌های کامل بارندگی بدون مقادیر گمشده در تحلیل‌های هواشناسی، اقلیم‌شناسی و هیدرولوژیکی مهم است. با این حال، مشکل گمشدگی داده‌ها اغلب به دلایل مختلفی رخ می‌دهد. گمشدگی داده-های بارش ممکن است ناشی از جابه‌جایی ایستگاه‌ها، عملکرد نادرست دستگاه اندازه‌گیری برای یک دوره زمانی خاص و خطا در تکنیک‌های مورد استفاده در اندازه‌گیری بارندگی باشد. سری‌های ناقص از داده‌های بارندگی می‌تواند نتایج مدل‌های هیدرولوژیکی و هواشناسی را تحت تأثیر قرار دهد. بنابراین، برآورد مقادیر گمشده بارندگی به یک وظیفه مهم تبدیل می‌شود و روش‌های مناسب برای جایگزینی مقادیر گمشده همواره مورد توجه مطالعات هیدرولوژیکی و هواشناسی است. تخمین داده‌های گمشده، همچنین به‌عنوان تکمیل (Coulibaly and Evora, 2007)، برآورد (Abebe et al, 2000)، برآورد سری داده‌های هواشناسی (Ramos-Calzado et al., 2008)، یا انتساب داده‌های هواشناسی (Schneider, 2001) نیز شناخته می‌شود و همواره موضوع مورد علاقه در بسیاری از مطالعات مدل‌سازی هیدرولوژیکی و دیگر محیط‌زیست می‌باشد. نتیجه تجزیه و تحلیل داده‌ها به کیفیت و کامل بودن داده ها بستگی دارد. با این حال، برای اکثر کاربران، داده‌های هواشناسی نقش مهمی در تصمیم‌گیری دارد و ضروری است که آن‌ها کامل و قابل اعتماد باشند. داده‌ها بایستی شامل سری‌هایی از داده‌های قوی و پیوسته باشند (de Carvalho et al., 2017). شکست یا وقفه در سری-های تاریخی مطمئناً منجر به افزایش عدم اطمینان برای پیش‌بینی هوا و تعریف سناریوهای اقلیمی آینده می‌گردد (Nascimento et al., 2023). فقدان داده‌های بارندگی با کیفیت خوب، پیامد بدی بر روند تحلیل خواهد داشت و متعاقباً منجر به نتایج مغرضانه تحلیل می‌شود. علاوه بر این، دقت نتایج تحلیل‌های اقلیم‌شناسی به شدت به کیفیت داده‌های بارندگی مورد استفاده بستگی دارد. با این حال، سری‌های زمانی بارندگی اغلب دارای عدم قطعیت (مانند نقاط پرت) به دلیل تغییرات زمانی و مکانی اندازه‌گیری بارندگی است که بر کیفیت داده‌ها تأثیر می‌گذارد. علاوه بر این، عواملی مانند معیوب بودن ابزار، جابه‌جایی ایستگاه‌ها و سهل‌انگاری انسان در حین اندازه‌گیری بارندگی ممکن است بر تداوم رکوردهای بارندگی تأثیر بگذارد (Suhaila et al, 2008). این مشکلات بر خروجی تخمین اثر می-گذارد و متعاقباً نتایج نادرست ایجاد می‌کند.

در مکان‌هایی که امکان ثبت دقیق و پیوسته داده‌های بارش در یک مقطع زمانی خاص وجود نداشته است، استفاده از روش‌هایی برای تخمین داده‌های بارش گمشده و اعمال آن در مدل‌های هیدرولوژیکی ضروری است. برای غلبه بر این مشکل، تعدادی از تکنیک های درون‌یابی در طول دهه‌ها با هدف تخمین مشاهدات گمشده در سری های زمانی اقلیمی، عمدتاً به صورت مقیاس‌های ماهانه، فصلی و زمانی روزانه توسعه یافته‌اند. این تکنیک‌ها مبتنی بر درون‌یابی فضایی می‌باشند؛ به این معنی که مقادیر منتسب در ایستگاه هدف با استفاده از مشاهدات همزمان از ایستگاه های اطراف محاسبه می‌شود.. تأثیر داده‌های گمشده بر استنتاج آماری به طور بالقوه مهم است؛ به ویژه در مواردی که ایستگاه‌های هواشناسی با داده‌های گمشده، با آن ایستگاه‌هایی که داده کامل دارند متفاوت باشند. برای یک تخمین منسجم و معتبر، مدل‌سازی مقادیر از دست رفته ضروری است. حذف ساده داده‌های گمشده می‌تواند منجر به اُریبی و محدود بودن نتایج شود (Harel, 2007). برای غلبه بر این مشکل، تکنیک‌های آماری از دهه ۱۹۸۰ ظهور و شامل جایگزینی داده‌های گمشده با مقادیر قابل قبول تخمینی است که به صورت برآورد در سری زمانی با داده‌های گمشده انجام می‌گیرد. در ادبیات آماری، این تکنیک‌ها به عنوان " برآورد داده‌های گمشده" شناخته می‌شوند و استفاده از آن‌ها تعمیم‌یافته و به سایر مناطق گسترس یافته است (Rubin, 1996; Schafer, 2002; Fraser and Yan, 2007; Nunes et al,. 2010).

برای تخمین مقادیر گمشده از روش‌های مختلفی مانند میانگین حسابی ساده، روش نسبت نرمال، روش وزن‌دهی عکس فاصله، روش وزن‌دهی ضریب همبستگی و مختصات جغرافیایی و غیره، استفاده شده است. برای غلبه بر مشکل شکاف داده‌های بارندگی، از مدل خطی تعمیم‌یافته (GLM) برای ارزیابی عملکرد روش‌های برآورد در شبه جزیره مالزی استفاده شده است. بر اساس آماره‌های ارزیابی عملکرد شامل میانگین خطای مطلق و ریشه میانگین

مربعات خطا، نتایج نشان داده است که استفاده از مدل خطی تعمیم‌یافته قادر به تولید تخمین بهتر و دقیق‌تر بارندگی است (Rahman and Islam, 2014). یک مرور جامع بر تکنیک‌های رایج پُرکردن ارزش‌های گمشده داده-های بارندگی روزانه و ماهانه توسط (Egigu, 2020) انجام گرفته است. در این مرور، روش‌های درون‌یابی فضایی، تکنیک‌های رگرسیونی و الگوریتم‌های شبکه‌های عصبی مصنوعی مورد بررسی قرار گرفته‌اند. در این مجموعه مدل-ها، روش میانگین حسابی ساده (SAAM) به عنوان ساده-ترین روش برای پُرکردن داده‌های هواشناسی مطرح شده است (Chow et al,. 1988). روش نسبت نرمال(NRM) به عنوان روشی کامل‌تر برای مواقعی پیشنهاد داده شده است که اختلاف بین مقادیر داده‌های بارش بین ایستگاه هدف و ایستگاه‌های همسایه بیش از ۱۰ درصد باشد؛ در غیر این صورت پیشنهاد به استفاده از روش میانگین حسابی است (Singh, 1997). روش نسبت نرمال برای لحاظ کردن اثر فاصله در تخمین داده گمشده بارندگی توسط (Young,1992) اصلاح شده و به عنوان یک روش رایج برای تخمین داده‌های گمشده بارندگی استفاده می‌شود. روش عکس فاصله (IDW) به عنوان روشی دقیق‌تر از میانگین حسابی ساده و نسبت نرمال، مطرح شده است (Lam, 1983). روش وزن‌دهی ضریب همبستگی (CCWM)، روش وزن‌دهی نمایی عکس(IEWM)، روش نسبت بارندگی فضایی (APR)، روش کریجینگ (KM)، روش اسپلاین (TPS)، رگرسیون خطی چندگانه (MLR)، روش انتظار- حداکثر (EM)، رگرسیون حداقل مربعات ساده پارامتریک (POLSR)، رگرسیون رتبه‌بندی شده غیرپارامتریک (NPRR)، رگرسیون متعامد (OR) و رگرسیون تابعی میانگین هندسی (GMFR) از جمله روش-های معرفی شده برای تخمین داده‌های گمشده بارش در این بررسی جامع می‌باشد. با این وجود، این تکنیک‌ها، همه روش‌های برآورد و تخمین داده‌های گمشده را شامل نمی-شوند. نتیجه تجزیه و تحلیل داده‌ها به شدت به کیفیت داده‌ها و کامل بودن آن‌ها بستگی دارد. درجه مناسب بودن روش‌های تخمین مقادیر گمشده بارندگی برای مناطق اقلیمی و جغرافیایی متفاوت بایستی قبل از اجرا برای اهداف خاص، تعیین و تأیید شوند. توزیع داده‌ها نقش عمده‌ای در انتخاب تکنیک خاص برای دستیابی به نتایج بهتر ایفا می-کند. بنابراین، قبل از استفاده از هر یک از این تکنیک‌ها، یک بررسی کامل در مورد توزیع داده‌ها ضروری است.

عملکرد روش‌های مختلف در برآورد داده‌های بارش ماهانه در استان آذربایجان غربی مورد بررسی قرار گرفته است(رضازاده جودی و ستاری، ۱۳۹۵). در این پژوهش، روش‌های رگرسیون خطی ساده، رگرسیون خطی چندمتغیره، روش میانگین‌گیری ساده، روش نسبت نرمال، بهترین تخمین‌گر واحد، روش نسبت‌ها، روش برآورد چندگانه و مدل درختی M5، مورد استفاده و عملکرد آن‌ها با استفاده از ضریب همبستگی، ضریب نش- ساتکلیف، جذر میانگین مربعات خطا و میانگین خطای مطلق مورد ارزیابی قرار گرفته است. نتایج نشان داده است که از بین روش‌های کلاسیک آماری بررسی شده به ترتیب روش‌های برآورد چندگانه، نسبت نرمال و رگرسیون خطی چندمتغیره دارای نتایج نسبتاً دقیق‌تر با خطای کمتر می‌باشند. دقت برآورد داده‌های گمشده بارش ماهانه با الگوریتم ژنتیک و کلونی مورچگان در ایستگاه مشهد توسط (فرزندی و همکاران، ۱۳۹۹) بهبود یافته است. در این بررسی، ایستگاه‌هایی از کشورهای مجاور ایران با توجه به معیار فاصله، همبستگی و تکمیل بودن داده‌ها و دوره‌های دارای داده گمشده به عنوان ایستگاه‌های مبنا انتخاب و سپس ده الگوی چندگانه رگرسیونی به بارش ماهانه مشهد برازش داده شده و در نهایت پارامترهای این الگوها با روش‌های الگوریتم ژنتیک و الگوریتم کلونی مورچگان بهینه شده است نشان داده است که الگوریتم ژنتیک و کلونی مورچگان دقت برآورد داده‌های گمشده بارش را به طور چشمگیری بالا برده است. قابلیت عملکرد روش‌های برآورد نواقص آماری در رابطه با بارش در مناطق خشک ایران بررسی شده است (کوثری و همکاران، ۱۴۰۰). در این بررسی، روش‌های نسبت نرمال، رگرسیون خطی، رگرسیون چندمتغیره و عکس مجذور فاصله در برآورد نواقص آماری بارش روزانه، ماهانه و سالانه ۷۳ ایستگاه هواشناسی واقع در مناطق خشک ایران با متوسط بارش سالانه کمتر از ۲۵۰ میلی‌متر و متناسب با نسبت نقص داده از ۵ تا ۵۰ درصد مورد ارزیابی قرار گرفته

است. نتایج این پژوهش نشان داده است که هر روش متناسب با شرایط میزان داده‌های گمشده از عملکرد متفاوتی برخوردار است. رگرسیون چندمتغیره زمانی‌که نقص داده‌ها زیاد نباشند از دقت بیشتری در برآورد داده-های روزانه برخوردار است ولیکن در کل به میزان نسبت داده‌های گمشده حساس است. روش نسبت نرمال، در برآورد نواقص بارش روزانه مناسب نیست ولیکن نسبت به میزان نقص داده‌ها پایدارتر از سایر روش‌ها می‌باشد. در سری‌های زمانی ماهانه عمکرد روش عکس مجذور فاصله و سپس روش نسبت نرمال مناسب تشخیص داده شده است. در سری‌های سالانه به ترتیب روش همبستگی خطی، نسبت نرمال و عکس مجذور فاصله عملکرد بهتری داشته-اند. نتایج برآورد داده بارش ماهانه در شش ایستگاه با استفاده از روش‌های میانگین حسابی، رگرسیون چندمتغیره و روش حداقل مربعات جزئی تکرارشونده غیرخطی (NLTPLS)، در استان هرمزگان نشان داده است که روش رگرسیون چندمتغیره از دقت بیشتری برخوردار است (Sattari et al,. 2017). اگرچه امروزه روش‌های مدرن و پیچیده به طور مانند یادگیری ماشین و یادگیری عمیق در تخمین بارندگی استفاده می‌شوند؛ اما این روش‌ها دارای فرمول ریاضی پیچیده، محاسبات فشرده و هزینه محاسباتی هستند که اجرای آن‌ها را با مشکل

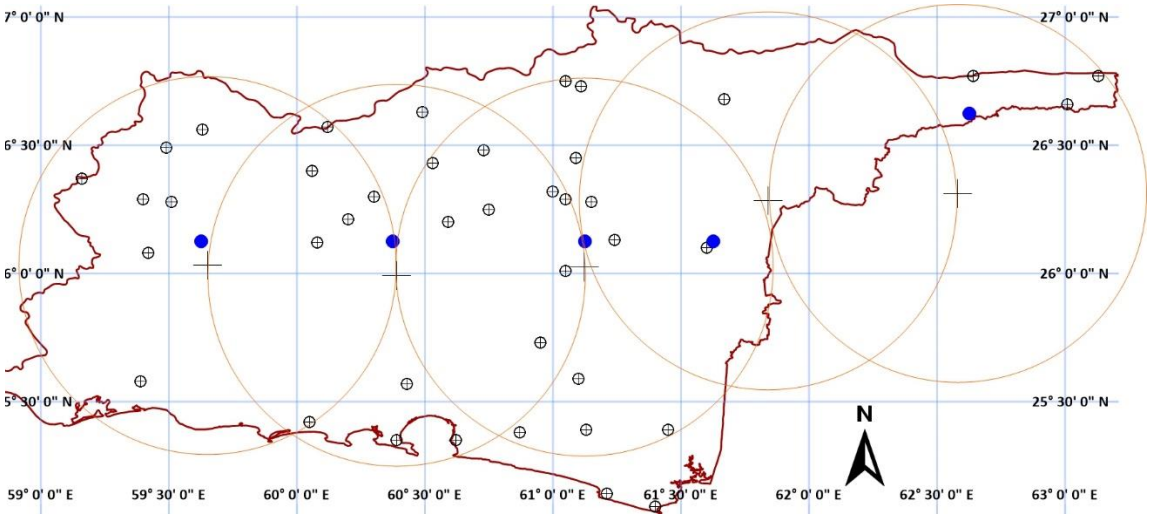
داده و روش‌شناسی

بر اساس تقسیمات هیدرولوژیکی، تعداد شش حوضه آبریز اصلی در گستره ایران وجود دارد. منطقه مورد مطالعه این پژوهش، زیر مجموعه حوضه آبریز خلیج فارس و دریای عمان می‌باشد که در گوشه جنوب‌شرقی آن واقع شده و خشک و کم‌آب می‌باشد(معروفی و همکاران، ۱۳۹۶). این حوضه آبریز، به نام رودخانه‌های بلوچستان جنوبی و رابچ – باهوکلات نیز معروف بوده و با حدود بین سدیج و مرکز پاکستان شناخته می‌شود. این حوضه آبریز دارای ویژگی-های متفاوت و منحصر به فرد بوده که مهمترین آن رژیم بارش موسمی تابستانه در بخش بزرگی از آن است. داده-های این پژوهش شامل بارش مقیاس ماهانه از ایستگاه‌های باران‌سنجی در حوضه آبریز بلوچستان جنوبی می‌باشد. این داده‌های از بایگانی شرکت سهامی آب منطقه‌ای استان سیستان و بلوچستان تهیه شده است. برای این منظور داده‌های تعداد ۳۹ ایستگاه باران‌سنجی در مرحله اول پایش شده‌اند (شکل۱). درصد گمشدگی داده‌ها، معیار مهم تصمیم‌گیری درباره امکان یا عدم امکان برآورد داده‌های گمشده و همچنین اتخاذ روش‌ها و الگوریتم‌های بهینه جهت برآورد و تخمین آن‌ها می‌باشد ادبیات پژوهشی درباره درصد گمشدگی داده‌ها و به عبارتی ارزش‌های گمشده، متنوع است. کامل‌ترین دیدگاه درباره درصد گمشدگی داده‌ها و نحوه برخورد با آن‌ها در مطالعه (Widaman, 2006) یافت می‌شود. درصد گمشدگی بر اساس یک معیار قراردادی (کمتر از ۵۰ درصد) به عنوان نسبت قابل قبول از نرخ گمشدگی در داده‌های بارش انتخاب و بر اساس آن تعداد ایستگاه‌های باران‌سنجی کاهش یافت. دوره آماری داده‌های این پژوهش از ۱۹۹۰ تا ۲۰۲۲ می‌باشد. بر اساس یک رویکرد نوآورانه، توزیع فضایی ایستگاه‌های باران‌سنجی محدود به دوایری گردید که دارای ۵۰ درصد همپوشانی

مواجه می‌کند. بنابراین، غلبه بر استفاده از روش‌های برآورد ساده‌تر به جای روش‌های پیچیده می‌باشد. مرور ادبیات پژوهش بیانگر غلبه روش‌های وزن‌دهی فضایی شامل روش‌های نسبت نرمال، روش وزن‌دهی عکس فاصله و روش وزن‌دهی ضریب همبستگی در برآورد داده‌های گمشده بارش در مقیاس‌های زمانی مختلف می‌باشد.

در جنوب‌شرق ایران، چندین پایگاه داده هواشناسی با متولیان مختلف وجود دارد. با این وجود هنوز مشکلات متعدد تداوم، کامل بودن، پُر کردن ایرادات و موارد دیگر موجود است. به طور منطقی، در مقیاس زمانی سالانه، مشکلات داده‌های بارش کمتر از مقیاس زمانی ماهانه و روزانه است. بنابراین با در نظر گرفتن داده‌های ماهانه و روزانه، تراکم ایستگاه‌های دارای داده قابل اعتماد و سری-های زمانی طولانی به شدت کاهش می‌یابد و این امر اطلاعات را محدود می‌نماید و امکان استفاده از آن‌ها در تعریف سناریوهای اقلیمی ممکن نیست. با توجه به این وضعیت، این پژوهش با اعمال اصلاحات در رایج‌ترین روش-های برآورد داده‌های گمشده بارندگی، در تلاش برای ارائه یک مجموعه داده با کیفیت خوب برای حوزه عمومی به ویژه زمینه‌های هواشناسی، اقلیم‌شناسی و هیدرولوژیکی در جنوب‌شرق ایران (حوضه آبریز بلوچستان جنوبی) است.

بودند. این دوایر فاصله ایستگاه‌ها را تعیین می‌کند که دارای قطری به طول ۱۵۰ کیلومتر بوده و دارای ۷۵ کیلومتر همپوشانی با دوایر هم‌جوار می‌باشد. این رویکرد بر اساس شباهت میانگین درازمدت بارش ایستگاه با اندک مداخله انتخاب گردیده است. میانگین درازمدت بارش سالانه ایستگاه‌های موجود در هر دایره بیشترین شباهت را به یکدیگر دارند. مبنای این تصمیم‌گیری، تجزیه و تحلیل داده‌های شبکه‌ای بارش از دو پایگاه APHRODITE و CHRIPS بوده است. از این‌رو، نزدیکترین یاخته به ایستگاه‌های باران‌سنجی در هر دایره به عنوان ایستگاه شاهد (هدف) در تجزیه و تحلیل مورد استفاده قرار گرفته است. لازم به ذکر است که در مطالعات پیشین، اصلاح داده‌های بارش شبکه‌ای با داده‌های ایستگاهی در این محدوده انجام گرفته است(رضایی و همکاران، ۱۴۰۲). بنابراین سری زمانی بارش شبکه‌ای دارای بالاترین همانندی با داده‌های مشاهداتی می‌باشد. موقعیت این یاخته‌ها با دایره توپر آبی رنگ در شکل (۱) نشان داده شده است. این یک رویکرد نوآورانه است که به ایستگاه‌های موجود در یک فاصله مشخص، وزن بیشتری در درون‌یابی فضایی داده و برآورد داده‌های بارش را مطمئن‌تر می‌نماید.



شکل(۱):موقعیت ایستگاه‌های باران‌سنجی در حوضه آبریز بلوچستان جنوبی.

روش‌های وزن‌دهی فضایی اصلاح‌شده شامل روش عکس فاصله (IDW[1])، روش نسبت نرمال (NR[2]) و روش ضریب همبستگی (CCW[3]) جهت برآورد داده‌های گمشده در گستره حوضه آبریز بلوچستان جنوبی انتخاب گردیده‌اند. فرمول روش وزن‌دهی (Suhaila et al., 2008) به شکل زیر است:

$$X_t = \sum_{\substack{i=1 \\ i \neq t}}^{N} W_i X_i \qquad (۱)$$

در این‌جا، X_t ارزش برآورد شده از داده گمشده در ایستگاه هدف (t)، N تعداد ایستگاه‌های هم‌جوار، X_i مشاهدات در ایستگاه هم‌جوار iاُم و W_i وزن ایستگاه هم‌جوار iاُم با وزن حاصل $\sum_{i=1}^{N} W_i = 1$.

ضریب اصلاح‌شده از روش وزن‌دهی همبستگی (CCWM)

موفقیت روش وزن‌دهی عکس فاصله (IDW) به شدت به وجود همبستگی فضایی مثبت بین بین ایستگاه هدف و ایستگاه‌های همسایه بستگی دارد. بنابراین، این روش با ضرایب همبستگی به عنوان فاکتورهای وزنی جایگزین

[1] Inverse distance weighting method
[2] Normal ratio method
[3] correlation coefficient weighting method

فاصله گردید و نتایج بهتری در تخمین داده‌های گمشده نسبت به روش وزن‌دهی عکس فاصله ارائه نمود (Teegavarapu and Chandramouli, 2005). روش وزن‌دهی همبستگی(CCW) به شرح زیر است:

$$W_i = \frac{r_{it}}{\sum_{\substack{i=1 \\ i\neq t}}^{N} r_{it}} \qquad (۲)$$

در این‌جا، r_{it} ضریب همبستگی است که نسبت کوواریانس دو مجموعه داده به حاصل جذر واریانس مجموعه داده‌ها به‌دست آمده از تمام سری‌های زمانی موجود بین داده‌های ایستگاه هدف (t) و مقادیر متناظر آن که در هر ایستگاه همسایه دیگر (i) ثبت شده است. به مرور زمان، تغییرات متعددی در روش‌های موجود برای تخمین داده‌های گمشده انجام گرفته است. این اصلاحات عمدتاً بر ترکیب عوامل وزنی مختلف به منظور ارائه بینش جدید و بهبود در تخمین مقادیر گمشده بارندگی متمرکز شده‌اند. از این رو، ضریب همبستگی (r_{it}) بیانگر قدرت رابطه بین دو ایستگاه است. با این‌حال، در مطالعه حاضر تغییرات اندکی در روش CCW بوسیله آزمایش با توان‌های مختلف انجام شده است تا وزن بیشتری به روش موجود داده شود. ضریب وزنی اصلاح شده به شرح زیر است:

$$W_i = \frac{r_{it}^{p}}{\sum_{\substack{i=1 \\ i\neq t}}^{N} r_{it}^{p}} \qquad (۳)$$

در این‌جا، r_{it}^{p} ضریب همبستگی بین ایستگاه هدف (t) و ایستگاه همسایه iاُم با دامنه p از ۲ تا ۶ است.

ضریب همبستگی اصلاح‌شده با روش وزن‌دهی عکس فاصله (CIDW)

روش عکس فاصله، رایج‌ترین روشی است که برای تخمین مقادیر گمشده در داده‌های اقلیمی و هواشناسی استفاده می‌شود. روش وزن‌دهی عکس فاصله بر اساس نزدیکی ایستگاه‌های مجاور یا ایستگاه‌های اطراف به ایستگاه مورد نظر است. روش وزن‌دهی فاصله به شرح زیر ارائه شده است:

$$W_i = \frac{d_{it}^{-p}}{\sum_{\substack{i=1 \\ i\neq t}}^{N} d_{it}^{-p}} \qquad (۴)$$

در این‌جا، d_{it}^{-p} فاصله بین ایستگاه هدف (t) و ایستگاه‌های همسایه iاُم می‌باشد. با افزایش فاصله از ایستگاه هدف، وزن کاهش می‌یابد. ارزش‌های بزرگتر از p تأثیر بیشتری را به مقادیر نزدیک به ایستگاه هدف اختصاص می‌دهد. در حقیقت، این روش به شدت تحت تأثیر حداقل فاصله بین ایستگاه هدف و ایستگاه‌های همسایه قرار دارد. امّا فاکتور همبستگی نیز می‌تواند تأثیری بر نتایج برآورد بگذارد. بنابراین، این مطالعه یک تغییر جزئی از عکس فاصله را با ترکیب روش‌های IDW و CCWM برای تخمین مقدار بارندگی گمشده پیشنهاد کرده به صورت زیر بیان می‌شود:

$$W_i = \frac{r_{it}^{p} d_{it}^{-2}}{\sum_{\substack{i=1 \\ i\neq t}}^{N} r_{it}^{p} d_{it}^{-2}} \qquad (۵)$$

در این‌جا، r_{it}^{p} ضریب همبستگی بین ایستگاه هدف (t) و ایستگاه‌های همسایه iاُم با دامنه p از ۲ تا ۶ است، در حالی‌که d_{it}^{-2} فاصله بین ایستگاه هدف (t) و ایستگاه‌های همسایه iاُم می‌باشد.

نسبت نرمال اصلاح‌شده با روش عکس فاصله (NRIDW)

به طور کلی، سه روش نسبت نرمال با فاکتورهای وزنی متفاوت وجود دارد. روش نسبت نرمال قدیمی (ONRM) که برای اولین بار توسط (Paulhus and Kohler, 1952) ارائه گردید. این روش بر اساس میانگین نسبت داده‌های موجود بین ایستگاه هدف (t) و ایستگاه‌های همسایه iاُم است. با لحاظ عامل همبستگی، نسبت نرمال قدیمی توسط (Yang, 1992) اصلاح گردید. این روش اصلاح شده (NRM) ضریب همبستگی بین ایستگاه هدف و ایستگاه همسایه را به عنوان فاکتور وزنی لحاظ می‌کند. سرانجام، اثر فاصله در معادله نسبت نرمال قدیمی توسط (Tang et al. 1996) معرفی و نسبت نرمال اصلاح‌شده بر اساس ریشه

دوم فاصله (MNR-T) نام نهاده شد. روش نسبت نرمال اصلاح‌شده بر اساس همبستگی (NRM) اغلب به عنوان بهترین روش در بین روش‌های نسبت نرمال (NR) شناخته می‌شود. در حالی‌که روش IDW به دلیل سادگی آن متداول‌ترین روش مورد استفاده بوده و گاهی اوقات به عنوان بهترین روش نیز در نظر گرفته می‌شود. در این روش، ضریب همبستگی بین ایستگاه هدف و ایستگاه همسایه به عنوان فاکتورهای وزنی لحاظ می‌گردد. روش وزن‌دهی این روش به شرح زیر است:

$$W_i = \frac{(n_i-2)r_{it}^2(1-r_{it}^2)^{-1}}{\sum_{\substack{i=1\\ i\neq t}}^{N}(n_i-2)r_{it}^2(1-r_{it}^2)^{-1}} \qquad (۶)$$

همان‌گونه که قبلاً بیان شد، روش NRM به شدت تحت تأثیر همبستگی فضایی مثبت قرار دارد. در این بین، IDW تحت تأثیر حداقل فاصله بین ایستگاه هدف و ایستگاه‌های همسایه قرار دارد. ترکیبی از این دو فاکتور وزنی، احتمالاً می‌تواند در بهبود نتایج تخمین مؤثر باشد. ترکیب وزن‌ها را می‌توان به صورت زیر بیان کرد:

$$W_i = \frac{(n_i-2)r_{it}^2(1-r_{it}^2)^{-1}d_{it}^{-2}}{\sum_{\substack{i=1\\ i\neq t}}^{N}(n_i-2)r_{it}^2(1-r_{it}^2)^{-1}d_{it}^{-2}} \qquad (۷)$$

در این‌جا، r_{it}^2 مجذور ضریب همبستگی سری زمانی داده‌های ماهانه بین ایستگاه هدف و ایستگاه همسایه است؛ n_i طول داده یا تعداد نقاطی است برای محاسبه ضریب همبستگی استفاده می‌گردد؛ d_{it}^{-2} فاصله بین ایستگاه هدف (t) و ایستگاه‌های همسایه iام و W_i وزن حاصله است.

کاربست و ارزیابی عملکرد روش‌های برآورد

نحوه انتخاب ایستگاه‌های همسایه و تعداد آن‌ها در نتایج درون‌یابی فضایی جهت برآورد داده‌های گمشده بسیار مهم است. به طور معمول، مشکل محدودیت تعداد ایستگاه در طول زمان مرتفع می‌گردد. به این مفهوم که با گذشت زمان به فراوانی و تراکم ایستگاه‌ها افزوده می‌شود. اما، مشکل اساسی این است که مقادیر گمشده در بازه‌های زمانی مختلف رخ می‌دهد و با راه‌حل آن، انتخاب ایستگاه-هایی است که در آن بازه زمانی، فاقد داده گمشده می‌باشند. بسیاری از محققان در ادبیات پژوهش پیشنهاد کرده‌اند که استفاده از ۳ یا ۴ ایستگاه نزدیک برای تخمین مقادیر گمشده ایستگاه هدف کافی است. برخی دیگر پیشنهاد کرده‌اند که انتخاب ایستگاه‌های همسایه بر اساس ضریب همبستگی با ایستگاه مورد نظر باشد. یانگ (Yang, 1992) از سه ایستگاه همسایه استفاده نموده که بیشترین همبستگی را داشته‌اند؛ در حالی‌که آیشاید و همکاران (Eischeid et al. 1995; 2000) از ایستگاه‌های همسایه با حداقل همبستگی $r > 0.35$، استفاده کرده‌اند. از سوی دیگر، ترونچی و همکاران (Tronci et al. 1986) و شیا و همکاران (Xia et al. 1999) ایستگاه‌های همسایه را که در ۱۰۰ کیلومتری ایستگاه هدف قرار داشتند لحاظ نموده-اند. در همین حال، برخی از پژوهشگران، ایستگاه‌های همسایه را بر اساس شباهت در الگوهای هندسی (مانند ذوزنقه، مستطیل) سری زمانی بارندگی مشاهده شده با ایستگاه هدف خود انتخاب نموده‌اند (Teegavarapu and Chandramouli, 2005). در این مطالعه، یک ایده نوآورانه در انتخاب ایستگاه‌های همسایه برای برآورد مقادیر گمشده ایستگاه‌های هدف توسعه یافته است که ترکیبی از روش‌های یاد شده با اصلاحات جزئی می‌باشد. بر این اساس، ابتدا ایستگاه‌های باران‌سنجی را بر اساس ضرایب همبستگی بارش ماهانه گروه‌بندی گردیده و سپس بر اساس بُردِ فاصله ۷۵ تا ۱۵۰ کیلومتر مرزبندی گردیده‌اند. انتخاب بُردِ فاصله ایستگاه‌ها بر اساس آستانه ضریب همبستگی ($r > 0.55$) بوده است. بنابراین، ایستگاه‌هایی که در یک دایره با قطر ۱۵۰ کیلومتر قرار می‌گیرند دارای ضریب همبستگی بالاتر از ۰/۵۵ می‌باشند. از سوی دیگر، امکان مشارکت ایستگاه‌های یک دایره در برآورد داده‌های گمشده دایره همسایه نیز وجود دارد. برای اعمال این امکان، شرط ۵۰ درصد همپوشانی لحاظ شده است. توزیع فضایی ایستگاه-های باران‌سنجی در منطقه مورد مطالعه، مطلوب ارزیابی می‌گردد. با این وجود تراکم ایستگاه‌ها در قسمت‌های شرقی به مراتب پایین‌تر از قسمت‌های مرکزی و غربی است که

قطعاً در نتیجه ملاحظات امنیتی (مرز بین‌المللی ایران و پاکستان) می‌باشد.

یک ایستگاه برای تخمین مقادیر گمشده بارندگی انتخاب و به عنوان ایستگاه هدف تعیین می‌گردد. داده‌های بارندگی ماهانه موجود از سال ۱۹۹۰ تا ۲۰۲۲ در ایستگاه-های باران‌سنجی برای تجزیه و تحلیل بر اساس شروط بالا، استفاده می‌شوند. فرض بر این است که داده‌ها در ایستگاه هدف به منظور آزمایش روش‌های تخمین داده‌های گمشده است. این ایستگاه‌های هدف، در مرحله آزمایش، همان ایستگاه‌های مشخص شده در شکل (۱) می‌باشند. به طور کلی، مقادیر بارندگی ماهانه گمشده در کل داده‌ها، کمتر از ۵۰ درصد است. با این وجود اکثر ایستگاه‌ها کمتر از ۳۰ درصد مقادیر گمشده برخوردارند. با این‌حال، برای بررسی سازگاری نتایج برآورد، تجزیه و تحلیل به شش درصد مختلف یعنی ٪۵، ٪۱۰، ٪۱۵، ٪۲۰، ٪۲۵ و ٪۳۰ تقسیم می‌گردند تا موارد مختلف داده‌های گمشده را نشان دهد. فرض کنید ٪۵ از داده‌هایی که به طور تصادفی برای آزمایش روش‌های برآورد انتخاب شده‌اند، به عنوان داده-های گمشده در نظر گرفته شوند، سپس بخش‌های باقیمانده (٪۹۵) از داده‌ها برای محاسبه ضرایب همبستگی و میانگین نسبت بین ایستگاه هدع و ایستگاه‌های همسایه مربوط به آن استفاده می‌شوند. تجزیه و تحلیل متعاقباً برای درصدهای دیگر و همچنین فواصل مختلف بین ایستگاه هدف و ایستگاه‌های همسایه تکرار می‌شود. این تکرار، یک فرایند هدفمند برای رسیدن به بهینه‌ترین شرایط برآورد می‌باشد.

سرانجام، عملکرد روش‌های برآورد داده‌های گمشده با استفاده از شاخص مشابهت (S-index)، میانگین خطای مطلق (MAE) و ضریب همبستگی (R) مقایسه گردیده است.

$$S-index = 1 - \frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{\sum_{i=1}^{n}(|\hat{x}_i - \bar{x}| + |x_i - \bar{x}|)^2} \qquad (۸)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_i - \bar{x}| \qquad (۹)$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\hat{x}_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(\hat{x}_i - \bar{x})^2}} \qquad (۱۰)$$

در این‌جا، n تعداد کل مشاهدات، $\hat{x}_i$ ارزش برآورد شده و x_i ارزش واقعی مشاهده است.

جدول(۱): مشخصات ایستگاه‌های باران‌سنجی در حوضه آبریز بلوچستان جنوبی

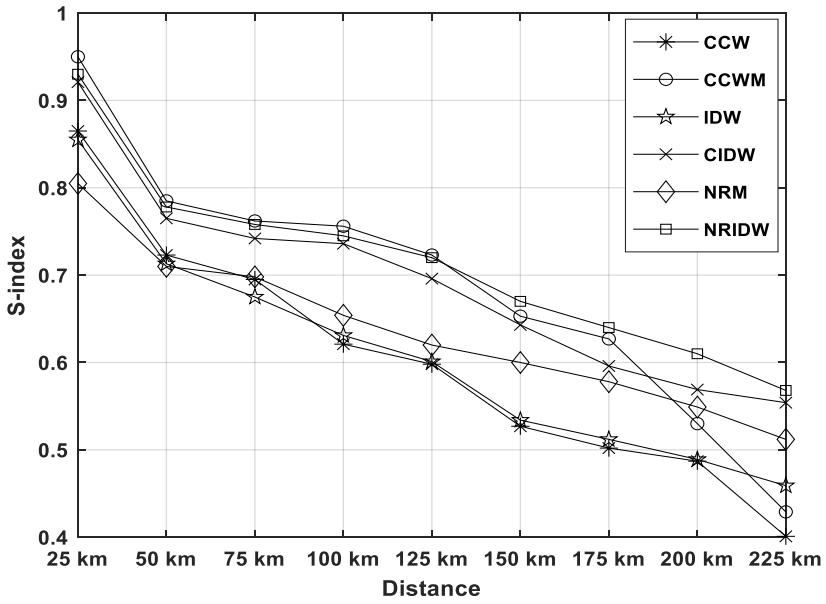
ردیف	ایستگاه	عرض‌جغرافیایی	طول جغرافیایی	ارتفاع(متر)	فاصله اقلیدوسی(کیلومتر)*	همبستگی*
۱	کَمْبِل سلیمان	۲۵/۳۸	۶۰/۸۷	۱۵	۰/۷۸۷۴ (۸۶/۶۸)	۰/۶۷
۲	سورکَمْپ	۲۵/۵۷	۶۰/۴۳	۳۱۴	۰/۸۸۹۴ (۹۲/۹۸)	۰/۶۴
۳	کَهیر	۲۵/۴۲	۶۰/۰۵	۸۸	۰/۷۷۶۳ (۸۷/۸۴)	۰/۶۳
۴	کُنارک	۲۵/۳۵	۶۰/۳۹	۱۳	۰/۷۷۵۱ (۸۶/۱۹)	۰/۵۹
۵	تیس	۲۵/۳۵	۶۰/۶۲	۱۰	۰/۸۱۱۲ (۹۹/۹۲)	۰/۶۲
۶	عورَکی	۲۵/۷۳	۶۰/۹۵	۴۵	۰/۴۳۲۰ (۴۷/۲۸)	۰/۶۷
۷	بَل ریمدان	۲۵/۳۹	۶۱/۴۵	۱۱	۰/۸۰۳۶ (۸۷/۹۷)	۰/۶۹
۸	بِرْیس	۲۵/۱۴	۶۱/۲۱	۱۰	۰/۹۸۸۶ (۱۰۹/۸۵)	۰/۵۸
۹	نِگور	۲۵/۳۹	۶۱/۱۳	۲۰	۰/۷۳۵۰ (۸۱/۷۳)	۰/۶۸
۱۰	پُلان	۲۵/۵۹	۶۱/۱	۳۰	۰/۵۴۷۲ (۵۹/۵۴)	۰/۷۴
۱۱	شارَک	۲۶/۰۱	۶۱/۰۵	۲۹۶	۰/۱۳۷۲ (۱۴/۸۲)	۰/۸۹
۱۲	بُناپ	۲۶/۷۳	۶۱/۱۱	۱۳۲۷	۰/۶۰۵۱ (۶۷/۲۸)	۰/۶۳
۱۳	اِسپتکی	۲۶/۷۵	۶۱/۰۵	۷۵۷	۰/۶۲۹۴ (۶۰/۸۹)	۰/۶۱
۱۴	لَهْاباد	۲۶/۲۸	۶۱/۱۵۱	۷۸۹	۰/۱۵۷۱ (۱۷/۴۲)	۰/۸۶
۱۵	راسک	۲۶/۱۳	۶۱/۲۴	۴۹۴	۰/۱۱۵۱ (۱۵/۴۸)	۰/۹۴

۱۶	پیشین	۲۶/۱	۶۱/۶	۷۰۰	۰/۰۳۵۳ (۳/۷۳)	۰/۹۵
۱۷	آشار	۲۶/۶۸	۶۱/۶۷	۱۰۵۰	۰/۹۵۶۵ (۶۱/۸۷)	۰/۷۱
۱۸	سیرکان	۲۶/۷۷	۶۲/۶۴	۱۰۸۲	۰/۱۴۵۷ (۱۶/۲۰)	۰/۷۸
۱۹	کُشتِگان	۲۶/۶۶۶	۶۳/۰۱	۸۸۰	۰/۳۸۶۵ (۳۸/۵۰)	۰/۶۳
۲۰	آسْکان	۲۶/۷۵	۶۳/۱۳	۱۰۲۵	۰/۵۲۵۴ (۵۲/۱۰)	۰/۷۱

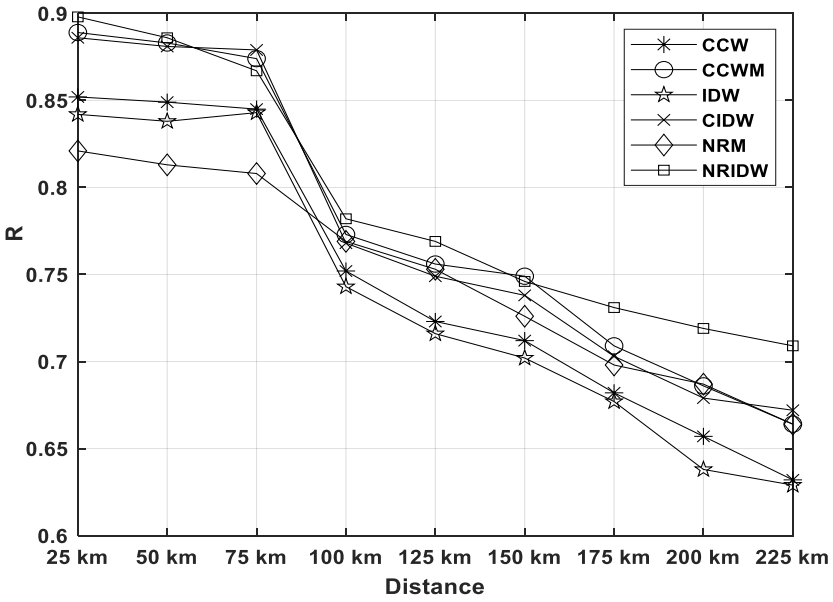
* فاصله اقلیدوسی و همبستگی بر اساس ایستگاه هدف تعیین گردیده که در شکل (۱) مشخص شده‌اند.

یافته‌ها و بحث:

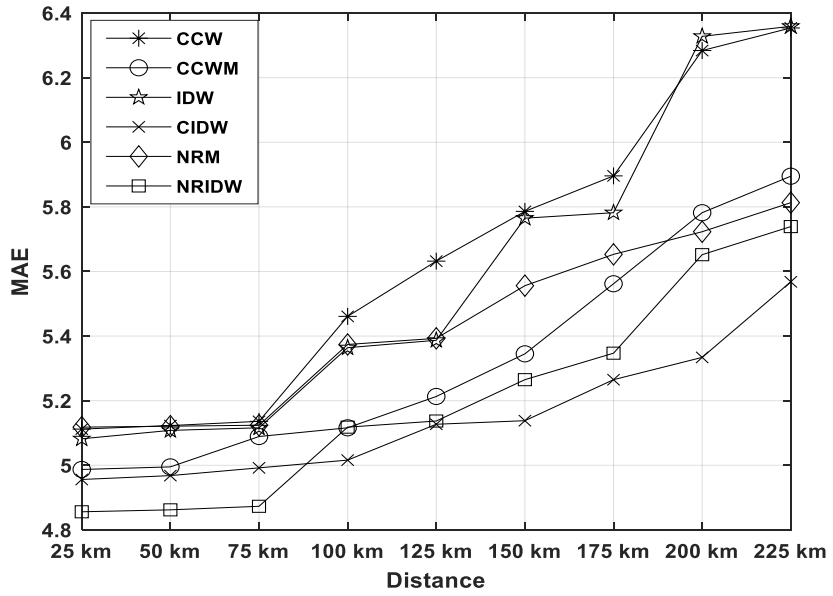
روش‌های موجود و اصلاح‌شده در بُرد فاصله بین $150 < D < 25$ کیلومتر در برآورد داده‌های گمشده ایستگاه هدف آزمایش شده‌اند. نُه فاصله (برحسب کیلومتر) برای آزمایش حساسیت روش‌ها به منظور جستجوی بُرد فاصله بهینه استفاده شده است. نتایج شاخص‌های اندازه-گیری خطا برای هر فاصله در شکل‌های (۲)، (۳) و (۴) نشان داده شده است. نتایج ٪۱۰ از داده‌های گمشده برای مقایسه فواصل مختلف برای هر شاخص اندازه‌گیری خطا انتخاب شده است.



شکل(۲): ارزیابی عملکرد روش‌های برآورد در فواصل متفاوت با شاخص شباهت



شکل(۳): ارزیابی عملکرد روش های برآورد در فواصل متفاوت با شاخص همبستگی

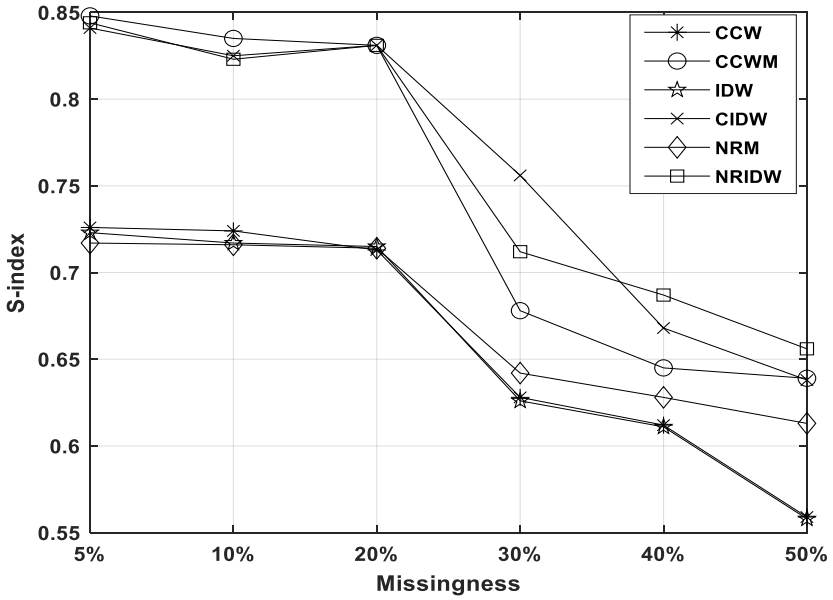


شکل(۴): ارزیابی عملکرد روش‌های برآورد در فواصل متفاوت با شاخص میانگین خطای مطلق

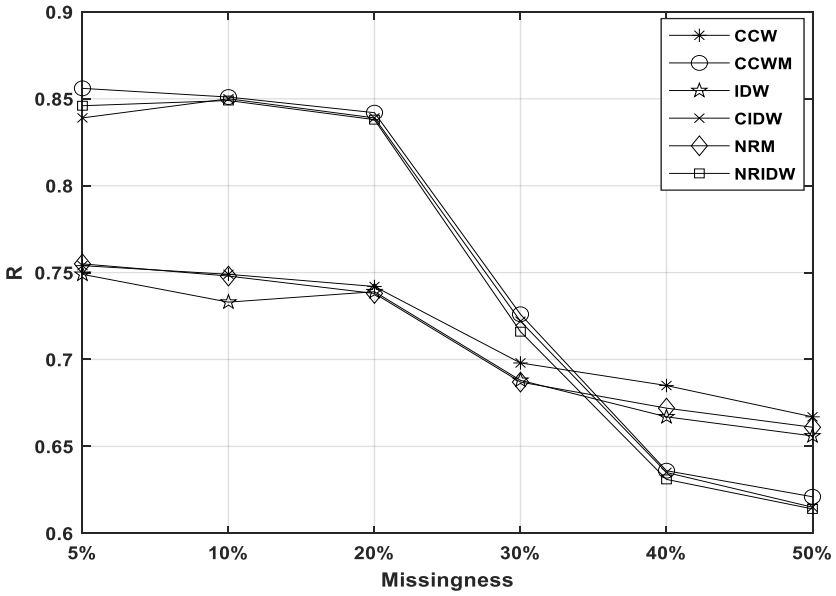
فاصله (بُرد) بهینه به عنوان فاصله خاصی تعریف می‌گردد که بیشترین مقادیر شاخص شباهت و ضریب همبستگی و همچنین کمترین میانگین خطای مطلق را به همراه دارد. تفاوت قابل ملاحظه‌ای در فواصل مختلف بر اساس شاخص‌های اندازه‌گیری خطا مشاهده می‌گردد. بنابراین، روش‌های برآورد داده‌های گمشده به فواصل حساس می‌باشند. افزایش فاصله به معنی افزایش تعداد ایستگاه‌های باران‌سنجی است. یک بُرد (فاصله) کوچک و محدود، تعداد کمتری از ایستگاه‌های باران‌سنجی را در برآورد داده‌های گمشده ایستگاه هدف مورد تجزیه و تحلیل قرار می‌دهد. با این حال، تعداد کم یا زیاد ایستگاه‌های باران‌سنجی، مشکلاتی را در تحلیل ایجاد می‌کند. به عنوان نمونه، فاصله زیاد می‌تواند تعداد زیادتری از ایستگاه‌ها را در تجزیه و تحلیل مشارکت داده، پیچیدگی و زمان محاسبات

را افزایش دهد. در سوی دیگر، فاصله محدود می‌تواند منجر به اُریبی نتایج به دلیل عدم وجود ایستگاه‌های کافی گردد. بنابراین، فاصله بهینه در این مطالعه در شعاع ۷۵ کیلومتری با تعداد ایستگاه‌های مناسب و نتایج برآورد منطقی، انتخاب گردیده است. در یک بُرد مشخص (دایره)، نزدیک‌ترین ایستگاه در فاصله تقریباً ۴ کیلومتری از ایستگاه هدف قرار دارد. فقط یک ایستگاه در فاصله دورتر از ۱۰۰ کیلومتری ایستگاه هدف قرار دارد. تمامی ایستگاه‌ها در بُردهای مشخص (دایره) دارای ضریب همبستگی ($r > 0.55$) می‌باشند (جدول۱). با این وجود، می‌توان از ایستگاه‌های موجود در دایره‌های مجاور و همسایه ایستگاه هدف نیز برای برآورد داده‌های گمشده استفاده نمود که شامل ٪۵۰ درصد همپوشانی محدوده می‌گردد. بنابراین بُرد را می‌توان تا ۲۲۵ کیلومتر افزایش داد.

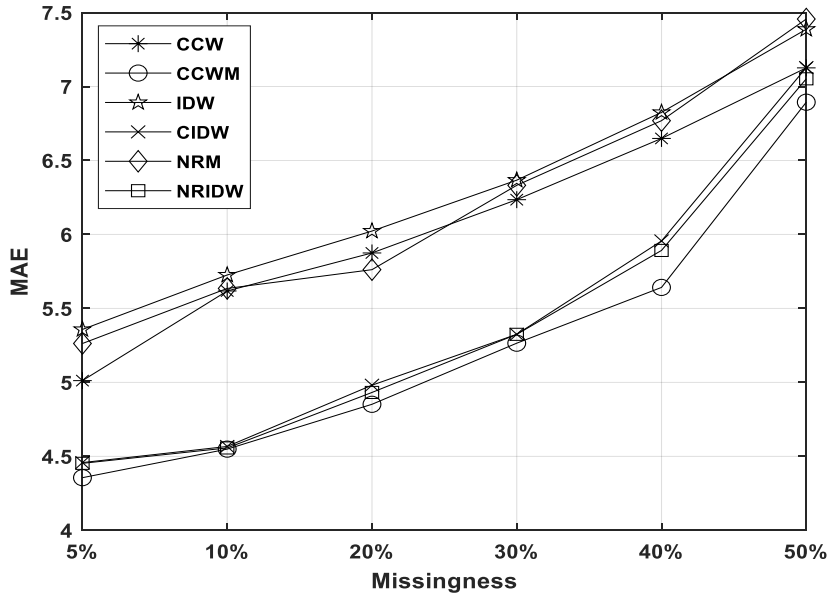
نتایج روش‌های برآورد داده‌های گمشده با ملاحظه درصدهای مختلف از گمشدگی داده‌ها در شکل‌های (۵)، (۶) و (۷) نشان داده شده است. شش درصد مختلف از گمشدگی داده‌ها که بین ٪۵ تا ٪۵۰ متغیر است برای نمایش داده‌های گمشده انتخاب شده است. روش‌های برآورد داده‌های گمشده به درصد گمشدگی داده‌ها حساس می‌باشند. عملکرد روش‌های برآورد با افزایش درصد گمشدگی داده‌های ایستگاه‌های هدف از مقدار معینی (٪۲۰)، کاهش قابل ملاحظه‌ای می‌یابد. بنابراین، مطلوب‌ترین وضعیت برای برآورد داده‌های گمشده با درصد گمشدگی کمتر از ٪۲۰ است. بر اساس شاخص ارزیابی شباهت و همبستگی، عملکرد روش‌های موجود و روش‌های اصلاح شده با درصدهای گمشدگی کمتر از ۲۰ بسیار مطلوب بوده و علاوه برآن تفاوت قابل ملاحظه‌ای ندارند. با افزایش درصد گمشدگی داده‌ها، تفاوت روش‌های برآورد بیشتر می‌گردد. روش-های وزن‌دهی فضایی اصلاح‌شده همواره عملکرد بهتری نسبت به روش‌های موجود دارند.



شکل(۵): مقایسه روش‌های برآورد بر اساس شاخص شباهت با درصدهای مختلف گمشدگی



شکل(۶): مقایسه روش‌های برآورد بر اساس شاخص همبستگی با درصدهای مختلف گمشدگی



شکل(۷): مقایسه روش‌های برآورد بر اساس شاخص میانگین خطای مطلق با درصدهای مختلف گمشدگی

درصد بهبود عملکرد روش‌های اصلاح شده نسبت به روش‌های موجود در برآورد داده‌های گمشده با اعمال درصدهای مختلف از گمشدگی مورد ارزیابی قرار گرفته و نتایج آن در جدول (۲) نشان داده شده است. نتایج روش CCWM همواره با درصدهای مختلف گمشدگی بهتر از سایر روش‌ها به ویژه روش‌های موجود (اصلاح‌نشده) می‌باشد. زیرا توان‌های مختلفی با روش CCW آزمایش شده‌اند که در آن p بین ۲ تا ۶ قرار دارد. در نتیجه با افزایش توان، به ویژه با p=4، عملکرد روش CCWM بسیار بهبود یافته است. این نتایج نشان می‌دهد که با افزایش توان ضریب همبستگی، احتمال بیشتری وجود دارد که روش اصلاح‌شده نسبت به روش موجود تخمین‌های بهتری ارائه دهد. مقایسه CCMW نسبت به CCW نشان می‌دهد که بهبود از ٪۷/۰۱ تا ٪۱۷/۷۱ در شاخص شباهت، ٪۴/۲۲ تا ٪۱۲/۰۵ در شاخص همبستگی و ٪۵/۴۲ تا ٪۱۵/۳۲ بوده است. عملکرد روش CCMW با افزایش درصد گمشدگی

فراتر از ٪۲۰، کاهش می‌یابد. بنابراین روش‌های اصلاح‌شده نیز نسبت به درصد گمشدگی داده‌ها حساس می‌باشند. به طور کلی، عملکرد روش‌های اصلاح‌شده نسبت به روش‌های موجود بهبود یافته است. با این وجود، روش‌های اصلاح‌شده نیز نسبت به گمشدگی داده‌ها حساس بوده و بهترین عملکرد آن‌ها با گمشدگی کمتر از ٪۲۰ اتفاق می‌افتد.

جدول(۲): درصد بهبود عملکرد روش‌های اصلاح شده نسبت به روش‌های موجود با درصدهای مختلف گمشدگی

درصد گمشدگی	شاخت شباهت (S-index)			شاخص همبستگی(R)			شاخص میانگین خطای مطلق(MAE)		
	(CCMW/CCW)	(CIDW/IDW)	(NRIDW/NRM)	(CCMW/CCW)	(CIDW/IDW)	(NRIDW/NRM)	(CCMW/CCW)	(CIDW/IDW)	(NRIDW/NRM)
٪۵	۱۷/۷۱	۱۶/۳۲	۱۶/۸۰	۱۲/۰۵	۱۲/۰۲	۱۳/۵۳	-۱۵/۳۹	-۱۶/۸۰	-۱۳/۰۹
٪۱۰	۱۴/۹۴	۱۵/۰۶	۱۵/۳۳	۱۳/۵۰	۱۵/۹۶	۱۳/۶۲	-۱۹/۱۳	-۲۰/۲۴	-۱۹/۰۳
٪۲۰	۱۶/۳۹	۱۶/۲۲	۱۶/۵۵	۱۳/۵۵	۱۳/۵۳	۱۳/۴۸	-۱۴/۴۰	-۱۷/۳۰	-۱۷/۴۱
٪۳۰	۱۰/۹۰	۲۰/۷۷	۷/۹۶	۴/۲۲	۴/۹۴	۴/۰۱	-۱۵/۹۰	-۱۶/۳۷	-۱۵/۵۶
٪۴۰	۹/۳۹	۹/۳۳	۵/۳۹	-۶/۱۰	-۴/۸۰	-۷/۱۵	-۱۲/۹۶	-۱۲/۷۳	-۱۵/۱۵
٪۵۰	۷/۰۱	۱۴/۳۴	۱۴/۳۱	-۷/۱۱	-۶/۲۵	-۶/۹۰	-۵/۴۲	-۳/۵۱	-۳/۲۶

نتیجه‌گیری

این پژوهش یک رویکرد عملی نوآورانه و قابل اعتماد برای انتخاب بهترین روش تخمین به منظور برآورد داده‌های ناقص بارندگی ماهانه را در جنوب‌شرق ایران پیاده‌سازی نموده است. برای این منظور، روش‌های مختلفی برای یافتن بهترین روش تخمین آزمایش شده است. این روش‌ها از مجموعه روش‌های رایج وزن‌دهی فضایی می‌باشند. در این پژوهش روش‌های موجود شامل روش‌های وزن‌دهی عکس فاصله، نسبت نرمال و ضریب همبستگی مورد بررسی و بازنگری قرار گرفته اند. برخی اصلاحات یا تجدیدنظرها در روش‌های موجود انجام شده و این روش‌های اصلاح‌شده جدید برای تخمین مقادیر گمشده بارندگی آزمایش شده‌اند. اصلاحات شامل (۱) ترکیب فاکتورهای وزنی مختلف مانند ترکیب دو روش نسبت نرمال و عکس فاصله (به عنوان مثالNRIDW) که هر دو از بهترین و رایج‌ترین روش‌های موجود می‌باشند، (۲) لحاظ تأثیر همبستگی در عکس فاصله (به عنوان مثال CIDW) و در نهایت (۳) آزمایش توان‌های مختلف برای ضریب همبستگی (به عنوان مثال CCWM) بوده‌اند. این روش‌های اصلاح‌شده با درصدهای مختلف از گمشدگی داده‌ها (٪۵، ٪۱۰، ٪۲۰، ٪۳۰، ٪۴۰ و ٪۵۰) و همچنین بُرد (فاصله) متغیر آزمایش گردیده‌اند.

نتایج نشان داد که روش‌های اصلاح‌شده نسبت به بُرد (فاصله) و همچنین درصد گمشدگی داده‌ها حساس می‌باشند. عملکرد روش‌های برآورد با افزایش درصد گمشدگی داده‌های ایستگاه‌های هدف از مقدار معینی (٪۲۰)، کاهش قابل ملاحظه‌ای می‌یابد. بنابراین، مطلوب‌ترین وضعیت برای برآورد داده‌های گمشده با درصد گمشدگی کمتر از ٪۲۰ است. عملکرد روش‌های موجود و روش‌های اصلاح شده با درصدهای گمشدگی کمتر از ۲۰ بسیار مطلوب بوده و علاوه برآن تفاوت قابل ملاحظه‌ای ندارند. با افزایش درصد گمشدگی داده‌ها، تفاوت روش‌های برآورد بیشتر می‌گردد. روش‌های وزن‌دهی فضایی اصلاح‌شده همواره عملکرد بهتری نسبت به روش‌های موجود دارند. همچنین، روش‌های برآورد داده‌های گمشده به فاصله ایستگاه هدف از ایستگاه‌های مجاور نیز حساس می‌باشند. فاصله بهینه در این مطالعه در شعاع ۷۵ کیلومتری با تعداد ایستگاه‌های مناسب و برآورد منطقی، تعیین گردیده است. در نهایت، عملکرد روش‌های اصلاح‌شده نسبت به روش‌های موجود، بهبود یافته است. نتایج عملکرد روش CCWM همواره با

درصدهای مختلف گمشدگی بهتر از سایر روش‌ها به ویژه روش‌های موجود (اصلاح‌نشده) می‌باشد. درصد بهبود عملکرد روش‌های اصلاح‌شده قابل ملاحظه بوده است. بنابراین، روش‌های اصلاح‌شده برای برآورد داده‌های گمشده در منطقه مورد مطالعه به خوبی عمل می‌کنند. پیشنهاد می‌گردد، اثرات ناهمواری‌ها (ارتفاعات) و همچنین طول داده‌ها در مطالعات بعدی مورد تجزیه و تحلیل قرار گیرد. از جمله مسائل دیگر که لازم است در مطالعات آتی به آن‌ها پرداخته شود، بررسی روش پیشنهادی این مطالعه در سایر حوضه‌های آبریز و مناطق بارشی ایران است. در ضمن لازم است حساسیت‌سنجی این روش‌ها به رژیم‌های متفاوت بارش انجام گیرد. ، در نهایت بررسی عملکرد روش‌های پیشنهادی نسبت به برآورد داده‌های گمشده در میان سایر متغیرهای اقلیمی مانند دما نیز توصیه می‌شود

.

سپاسگزاری

از بازبینان ناشناس بابت ارائه نظرات سازنده در راستای بهبود کیفیت مقاله صیمیمانه قدردانی می‌گردد. این پژوهش از طرف شرکت سهامی آب منطقه‌ای استان سیستان و بلوچستان در قالب پروژه تقاضا محور پشتیبانی شده که نویسندگان مراتب قدردانی خود را اعلام می‌دارند.

منابع

رضازاده جودی، *ع.ر.*، م.ت. ستاری.۱۳۹۵. ارزیابی عملکرد روش‌های مختلف در برآورد داده‌های بارش ماهانه، تحقیقات کاربردی علوم جغرافیایی، دوره ۱۶، شماره ۴۲، صص ۱۷۶-۱۵۵.

رضایی، محسن.، م، اژدری‌مقدم.، غ، ر، عزیزیان.، ع، ا.، شمسی‌پور.۱۴۰۲. ارزیابی پایگاه‌های داده‌های شبکه‌ای بارش در حوزه آبخیز بلوچستان جنوبی، محیط زیست و مهندسی آب، سال۹، شماره ۱، صص ۷۶-۶۶.

فرزندی، محبوبه.، ح، ثنایی‌نژاد.، ب، قهرمان.، م، سرمدی. ۱۳۹۹. افزایش دقت برآورد داده‌های گمشده بارش ماهانه با الگوریتم ژنتیک و کلونی مورچگان، پژوهش‌های اقلیم‌شناسی، دوره ۱۱، شماره ۴۲، صص ۶۰-۴۷.

کوثری، م.ر.، اسمعیل‌زاده حسینی، میترااالسادات، ا، ح.، م، میری.۱۴۰۰. بررسی کارایی روش‌های برآورد نواقص آماری در رابطه با پارامتر بارش در مناطق خشک ایران، فیزیک زمین و فضا، دوره ۴۷، شماره ۲، صص ۳۳۲-۳۱۵.

معروفی، صفر.، ر، نوروز ولاشدی.،ف، گلکار. ۱۳۹۶. مدل‌سازی و پیش‌بینی بارش ماهانه در حوضه آبریز بلوچستان جنوبی، تحقیقات جغرافیایی، دوره ۳۲، شماره ۱، صص ۱۶۲-۱۴۹.

Abebe, A. J., Solomatine, D. P., & Venneker, R. G. W .2000. Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. *Hydrological Sciences Journal*, *45*(3), 425-436.

Chow, V. T., Maidment, D. R., & Mays, L. W. 1988. Design Storms-Chapter 14. *Applied Hydrology, McGraw-Hill Book Company*.

Coulibaly, P., & Evora, N. D. 2007. Comparison of neural network methods for infilling missing daily weather records. *Journal of hydrology*, *341*(1-2), 27-41.

de Carvalho, J. R. P., Almeida Monteiro, J. E. B., Nakai, A. M., & Assad, E. D. 2017. Model for multiple imputation to estimate daily rainfall data and filling of faults. *Revista Brasileira de Meteorologia*, *32*, 575-583.

Egigu, M. L. 2020. Techniques of Filling Missing Values of Daily and Monthly Rain Fall Data: A Review. *SF J Environ Earth Sci. 2020; 3 (1), 1036*.

Eischeid, J. K., Baker, C. B., Karl, T. R., & Diaz, H. F. 1995. The quality control of long-term climatological data using objective data analysis. *Journal of applied meteorology and climatology*, *34*(12), 2787-2795.

Fraser, G., & Yan, R. 2007. Guided multiple imputation of missing data: using a subsample to strengthen the missing-at-random assumption. *Epidemiology*, 246-252.

Harel, O. 2007. Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*, *4*(1), 75-89.

Lam, N. S. N. 1983. Spatial interpolation methods: a review. *The American Cartographer*, *10*(2), 129-150.

Nascimento, M. L., Gonçalves, K. C., & Mendonça, M. J. 2023. Spatio-Temporal Instrumental Variables Regression with Missing Data: A Bayesian Approach. *Computational Economics*, *62*(1), 29-47.

Nunes, L. N., Klück, M. M., & Fachel, J. M. G. 2010. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. *Revista Brasileira de Epidemiologia*, *13*(4), 596-606.

Paulhus, J. L., & Kohler, M. A. 1952. Interpolation of missing precipitation records. *Monthly Weather Review*, *80*(8), 129-133.

Rahman, M. G., & Islam, M. Z. 2014. Fimus: A framework for imputing missing values using co-appearance, correlation and similarity analysis. *Knowledge-Based Systems*, *56*, 311-327.

Ramos-Calzado, P., Gómez-Camacho, J., Pérez-Bernal, F., & Pita-López, M. F. 2008. A novel approach to precipitation series completion in climatological datasets: application to Andalusia. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *28*(11), 1525-1534.

Rubin, D. B. 1996. Multiple imputation after 18+ years. *Journal of the American statistical Association*, *91*(434), 473-489.

Sattari, M. T., Rezazadeh-Joudi, A., & Kusiak, A. 2017. Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*, *48*(4), 1032-1044.

Schafer, J. L., & Graham, J. W. 2002. Missing data: our view of the state of the art. *psychological methods*, *7*(2), 147.

Schneider, T. 2001. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of climate*, *14*(5), 853-871.

Singh, V. P. 1997. The use of entropy in hydrology and water resources. *Hydrological processes*, *11*(6), 587-626.

Suhaila, J., Sayang, M. D., & Jemain, A. A. 2008. Revised spatial weighting methods for estimation of missing rainfall data. *Asia-Pacific Journal of Atmospheric Sciences*, *44*(2), 93-104.

Tang, W. Y., Kassim, A. H. M., & Abubakar, S. H. 1996. Comparative studies of various missing data treatment methods-Malaysian experience. *Atmospheric Research*, *42*(1-4), 247-262.

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), 191-206.

Tronci, N., Molteni, F., & Bozzini, M. 1986. A comparison of local approximation methods for the analysis of meteorological data. *Archives for Meteorology Geophysics and Bioclimatology Series B Theoretical and Applied Climatology*, *36*(2), 189-211.

Widaman, K. F. 2006. Best practices in quantitative methods for developmentalists: III. Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development*. 7(1), 210-211.

Xia, Y., Fabian, P., Stohl, A., & Winterhalter, M. 1999. Forest climatology: estimation of missing values for Bavaria, Germany. *Agricultural and forest meteorology*, *96*(1-3), 131-144.

Young, K. C. 1992. A three-way model for interpolating for monthly precipitation values. *Monthly Weather Review*, *120*(11), 2561-2569.