2023, Vol.9, No.4, Pages 37 to 46
Journal of Water and Sustainable Development

سال نهم، شماره ٤، ١٤٠١، صفحات ٣٧ تا ٤٦
نشریه آب و توسعه پایدار

Article Type: Applied Article

نوع مقاله: پژوهش کاربردی

Application and Evaluation of the Ability of Copulas in Estimating Daily Precipitation in the East of Lake Urmia Basin

M. Khaledi-Alamdari[1], A. Majnooni-Heris[2*], A. Fakheri Fard[3]

1, 2, 3- Ph.D. Student in Irrigation and Drainage, Associate Professor and Professor, Department of Water Science and Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran.

*(Corresponding Author Email: majnooni@tabrizu.ac.ir)

Received: 19-06-2022 Revised: 15-09-2022

Accepted: 27-09-2022 Available Online: 11-03-2023

کاربرد و ارزیابی توانایی توابع مفصل در تخمین بارش روزانه در شرق حوضه دریاچه ارومیه

محمد خالدی علمداری[1]، ابوالفضل مجنونی هریس[2*]، احمد فاخری فرد[3]

۱، ۲ و ۳- به‌ترتیب دانشجوی دکتری تخصصی آبیاری و زهکشی، دانشیار و استاد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران.

*(نویسنده مسئول، E-Mail: majnooni@tabrizu.ac.ir)

تاریخ دریافت: ۱۴۰۱/۰۳/۲۹ تاریخ بازنگری: ۱۴۰۱/۰۶/۲۴

تاریخ پذیرش: ۱۴۰۱/۰۷/۰۵ تاریخ انتشار: ۱۴۰۱/۱۲/۲۰

Abstract

The use of accurate and continuous data series is a necessary condition for most statistical and hydrological studies. Due to the importance of precipitation as one of the most important climatic and hydrological variables, in the present study, in order to predict the daily precipitation of Tabriz, Copula functions were used and the results were compared with intelligent methods and classical statistics. To predict precipitation in Tabriz station, precipitation data of Sarab, Sahand, and Maragheh stations were also used as auxiliary stations. Based on the obtained results, among all the methods studied, the M5 method with RMSE values of 3.14 mm, and the MAD method with 2.13 mm and the RF method with RMSE values of 5.18 mm and MAD 3.04 mm have the highest and lowest accuracy in estimating precipitation events, respectively. Among Archimedean copulas, the RMSE and MAD values for the Gambel function are 3.89 and 2.51 mm, respectively. Despite the range of estimation data is still very close to other methods, considering the capabilities of Copula functions, including the ability to apply multiple conditions and its probabilistic nature, which considers the behavior of the phenomenon, it can be acknowledged that in similar circumstances, the ability of Copula functions to estimate the missing data of phenomena such as rainfall is acceptable.

Keywords: Probabilistic Analysis, Missing Data, Forecasting,, Archimedean Copulas.

چکیده

استفاده از سری داده‌های صحیح و بدون داده گم‌شده، شرط لازم برای انجام بیشتر مطالعات آماری و هیدرولوژیکی است. باتوجه‌به اهمیت بارش به‌عنوان یکی از مهمترین متغیرهای اقلیمی و هیدرولوژیکی، در این پژوهش به‌منظور پیش‌بینی بارش روزانه تبریز، از توابع مفصل استفاده شده و نتایج آن با روش‌های هوشمند و آمار کلاسیک مقایسه شد. به‌منظور پیش‌بینی بارش در ایستگاه تبریز، از داده‌های بارش ایستگاه‌های سراب، سهند و مراغه نیز به‌عنوان ایستگاه‌های کمکی استفاده شد. بر اساس نتایج به‌دست آمده در بین همه روش‌های مورد بررسی، روش مدل درخت با مقادیر RMSE معادل ۳/۱۴ میلی‌متر و MAD معادل ۲/۱۳ میلی‌متر و روش جنگل تصادفی با مقادیر RMSE معادل ۵/۱۸ میلی‌متر و MAD معادل ۳/۰۴ میلی‌متر به‌ترتیب بیشترین و کمترین دقت را در برآورد رویدادهای بارش دارند. در میان مفصل‌های ارشمیدسی، تابع گامبل مقادیر RMSE و MAD به ترتیب ۳/۸۹ و ۲/۵۱ میلی‌متر می‌باشد. از آنجایی‌که محدوده خطای داده‌های تخمینی به‌دست آمده از توابع مفصل بسیار نزدیک به سایر روش‌ها می‌باشد؛ باتوجه‌به قابلیت‌های توابع مفصل از جمله توانایی اعمال شرط‌های متعدد و ماهیت احتمالاتی آن، که رفتار پدیده را در نظر می‌گیرد، می‌توان گفت در شرایط مشابه توانایی توابع مفصل در برآورد داده‌های گمشده پدیده‌های احتمالاتی مانند بارندگی مناسب است.

واژه‌های کلیدی: تحلیل احتمالاتی، داده‌های گمشده، پیش‌بینی، مفصل‌های ارشمیدسی.

HomePage: https://jwsd.um.ac.ir

How to cite this article: Khaledi-Alamdari M., Majnooni-Heris A. and Fakheri Fard A. 2023. Application and Evaluation of the Ability of Copulas in Estimating Daily Precipitation in the East of Lake Urmia Basin. Journal of Water and Sustainable Development, 9(4): 37-46. doi: http://dx.doi.org/10.22067/jwsd.v9i4.2206.1156

مقدمه

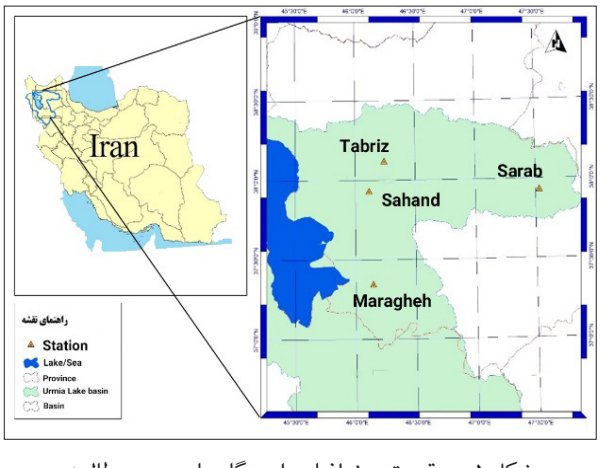
پایه و اساس انجام مطالعات هیدرولوژیکی، استفاده از داده‌های صحیح و بدون گم‌شدگی است. نتایج تجزیه و تحلیل داده‌های مترولوژیکی، هیدرولوژیکی و محیط‌زیستی بیشتر به برآورد قابل اعتماد داده‌های گم‌شده وابسته است. از سویی دیگر انجام همه مطالعات اقلیمی و هیدرولوژیکی وابسته به وجود داده‌های بارش هستند. ازاین‌رو برآورد داده‌های گم‌شده بارش اهمیت به‌سزایی در مطالعات هیدرولوژیکی دارد. در این راستا مطالعات متعددی در رابطه با برآورد داده‌های گم شده بارش در نواحی مختلف جهان انجام شده است که باتوجه‌به ماهیت روش‌های مورد استفاده و نیز خصوصیات اقلیمی هر منطقه، نتایج متفاوتی نیز به‌دست آمده است. Armanuos و همکاران (۲۰۲۰) در مطالعه‌ای با استفاده از داده‌های ۱۵ ایستگاه در بازه زمانی ۳۴ ساله، بیست و یک روش مختلف برای تخمین داده‌های گم‌شده بارش در منطقه اتیوپی را ارزیابی کردند. بر اساس نتایج به‌دست آمده، روش‌های نسبت نرمال (NR)، رگرسیون خطی چندگانه (MLR)، وزن‌دهی معکوس فاصله (IDW)، ضریب همبستگی وزنی (CCW) و میانگین حسابی (AA) مطمئن‌ترین روش‌ها در بین روش‌های مورد مطالعه هستند.

Abebe و همکاران (۲۰۰۰) روش‌های منطق فازی، شبکه‌های عصبی مصنوعی و نسبت نرمال را برای بازسازی داده‌های بارش روزانه مقایسه و گزارش کردند روش منطق فازی در مقایسه با دو روش دیگر، دقت بیشتری در بازسازی داده‌های بارش دارد. با بهره‌گیری از امکانات موجود، که دربرگیرنده روابط آماری و احتمالاتی توسعه داده شده می‌باشد، مدل‌های بسیاری به‌منظور شفاف‌سازی تأثیرات عوامل اقلیمی توسعه داده شده است. به‌عنوان‌مثال Yuan و Yamagata، ۲۰۱۵؛ Nguyen-Huy و همکاران، ۲۰۱۷؛ Jarvis و همکاران، ۲۰۱۸، این مدل‌ها را با فرض مشترک بودن توزیع نرمال بین متغیرها، روابط خطی را بین تعداد محدودی از شاخص‌های اقلیمی و عملکرد محصول به‌کار گرفتند. درحالی‌که یک مدل رگرسیون خطی بسیار ساده می‌باشد و تنها یک دیدگاه کلی از روند متغیرهای مورد نظر ارائه می‌دهد. این گونه مدل‌ها ممکن است به شدت تحت تأثیر داده‌های پراکنده همچون رویدادهای شدید باشند که به ایجاد رابطه ناصحیح در همبستگی بین متغیرهای در نظر گرفته شده می‌انجامد (Hassani، ۲۰۱۶). از طرف دیگر، ممکن است فرض توزیع نرمال داده‌ها درست نباشد (Nguyen-Huy و همکاران، ۲۰۱۸). برای برداشتن چنین مشکلاتی، شمار فراوانی مدل آماری مفصل[1] برای پیش‌بینی میزان بارندگی در مناطق زراعی محیط‌زیستی استرالیا تهیه شده است (Nguyen-Huy و همکاران، ۲۰۱۷). همچنین کوپلاها به‌صورت گسترده برای مدل‌سازی آماری و پیش‌بینی در رسته‌های مختلف دربرگیرنده انرژی (Bessa و همکاران، ۲۰۱۲؛ Zhang و Singh، ۲۰۱۴)، ریسک اقتصادی (Huang و همکاران، ۲۰۰۹؛ Lu و همکاران، ۲۰۱۴)، پیش‌بینی بارندگی و تغییرات اقلیم (Nguyen-Huy و همکاران، ۲۰۱۷) و هیدرولوژی (Kao و Govindaraju، ۲۰۱۰؛ Liu و همکاران، ۲۰۱۵) به‌کار گرفته شده‌اند. از کاربرد مدل‌های کوپلا می‌توان به تحلیل وقوع هم‌زمان رویدادهای شدید و متعدد آب‌وهوایی، که از آن به‌عنوان یک رویداد ترکیبی یاد می‌شود، نام برد (Manning و همکاران، ۲۰۱۸؛ Zscheischler و Seneviratne، ۲۰۱۷).

باتوجه‌به اهمیت بارش به‌عنوان یکی از مهمترین پارامترهای هواشناسی در بیشتر مطالعات اکوهیدرولوژیکی و نقش برجسته آن در توسعه پایدار، در مطالعه حاضر از توابع مفصل در پیش‌بینی بارش روزانه استفاده شده و نتایج آن با روش‌های هوشمند و آمار کلاسیک مقایسه خواهد شد.

مواد و روش‌ها

در پژوهش حاضر، بازسازی داده‌های بازرش روزانه در ایستگاه همدید تبریز مورد توجه قرار گرفته است. برای این‌منظور از فصل مشترک داده‌های رویداد بارش روزانه (روزهایی که بارش اتفاق افتاده است) چهار ایستگاه همدید شامل تبریز، سراب، سهند و مراغه در یک بازه زمانی ۳۰ ساله از سال ۱۹۹۱ تا ۲۰۲۰ استفاده شد. برای بازسازی داده‌های بارش روزانه از روش‌های آماری، روش‌های هوشمند و نیز توابع مفصل استفاده شده است. در شکل (۱) موقعیت جغرافیایی ایستگاه تبریز به‌عنوان ایستگاه هدف و ایستگاه‌های سراب، سهند و مراغه به‌عنوان ایستگاه‌های کمکی به‌منظور برآورد داده، ارائه شده است.



شکل ۱- موقعیت جغرافیایی ایستگاه‌های مورد مطالعه

• روش‌های هوشمند

- رگرسیون ماشین بردار پشتیبان[۲] (SVR)

ماشین بردار پشتیبان یکی از روش‌های یادگیری است که بر مبنای تئوری یادگیری آماری در سال ۱۹۹۲ میلادی معرفی شده است (Boser و همکاران، ۱۹۹۲). گسترش ماشین بردار پشتیبان بر اساس رگرسیون نیز در سال ۱۹۹۵ به نتیجه رسید (Vapnik، ۱۹۹۵). ماشین بردار پشتیبان مبتنی بر کمینه کردن ساختاری ریسک می‌باشد که از نظریه آموزش آماری گرفته شده است (Vapnik، ۱۹۹۸). مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده مدل طبقه‌بندی ماشین بردار پشتیبان و مدل رگرسیون بردار پشتیبان تقسیم‌بندی می‌شوند. در یک مدل رگرسیونی SVR لازم است وابستگی تابعی متغیر وابسته y به مجموعه‌ای از متغیرهای مستقل x تخمین زده شود. فرض بر این است که مانند دیگر مسائل رگرسیونی، رابطه بین متغیرهای وابسته و مستقل توسط یک تابع معین f به همراه یک مقدار اضافی خطای ناگریز که *noise* نامیده می‌شود مشخص می‌شود (رابطه ۱).

$$y=f(x)+noise \quad (1)$$

بنابراین موضوع اصلی پیدا کردن فرم تابع f است که بتواند به‌صورت صحیح موارد جدیدی را که SVR تاکنون تجربه نکرده است، پیش‌بینی کند. این تابع به‌وسیله آموزش مدل SVR بر روی یک مجموعه داده به‌عنوان مجموعه آموزش که شامل فرآیندی به‌منظور بهینه‌سازی دائمی تابع خطا است، قابل دسترسی است.

- شبکه‌های عصبی مصنوعی[۳] (ANN)

شبکه‌های عصبی مصنوعی از شبیه‌سازی شبکه‌های عصبی موجودات زنده الهام گرفته شده است که به‌عنوان ابزاری قدرتمند الگوی پردازش اطلاعات دارند (منهاج، ۱۳۸۴). تکنیک شبکه عصبی مصنوعی از دسته روش‌های هوشمند است که به‌طور گسترده‌ای در مدل‌سازی و پیش‌بینی فرآیندهای هیدرولوژیکی مورد استفاده قرار گرفته است. این شبکه‌ها از نورون‌ها تشکیل می‌شوند که در گروه‌هایی به نام لایه قرار گرفته و از راه اتصالات وزن‌دار به یکدیگر متصل می‌شوند. هر ساختار ساده شبکه از سه لایه تشکیل شده است: لایه ورودی، لایه پنهان و لایه خروجی. وقتی داده‌های ورودی به لایه ورودی وارد می‌شوند، از طریق شبکه عصبی عبور کرده و در لایه میانی بر روی آن‌ها پردازش انجام می‌شود تا زمانیکه خروجی در لایه خروجی به‌دست آید. هر نورون از راه اتصالات وزنی ورودی‌های زیادی را از سلول‌های عصبی دیگر دریافت می‌کند. این ورودی‌های وزنی جمع شده و یک تابع انتقالی را ایجاد می‌کنند که در نهایت خروجی نهایی نورون را تولید می‌کند (Talebizadeh و همکاران، ۲۰۱۰). باتوجه‌به اینکه شبکه عصبی مصنوعی به اطلاعات موبه‌مو در مورد روند فیزیکی به‌کار رفته در سیستم‌ها نیاز ندارد، به‌طور مؤثری برای مدل‌سازی فرآیندهای هیدرولوژیکی پیچیده استفاده می‌شود.

شبکه‌های عصبی مصنوعی انواع مختلفی دارند که متداول‌ترین آن‌ها پرسپترون چندلایه[۴] (MLP) می‌باشد و در این مطالعه از این مدل استفاده شده است. مدل MLP توسط سلول‌های عصبی ساده‌ای به نام پرسپترون تشکیل می‌شود (Kuan وWhite، ۱۹۹۴). پرسپترون با ایجاد یک ترکیب خطی باتوجه‌به وزن ورودی خود و سپس تعیین خروجی از طریق یک تابع انتقال غیرخطی، یک خروجی منفرد از چندین ورودی را محاسبه می‌کند.

- جنگل‌های تصادفی[۵] (RF)

روش جنگل‌های تصادفی را اولین بار بریمن در سال ۲۰۰۱ با توسعه درخت‌های تصمیم، به‌عنوان یک تکنیک جدید ارائه داده است که پیش‌بینی چندین الگوریتم منفرد را با هم با استفاده از قوانین مرتبط ترکیب می‌کند. این روش در بین روش‌هاي درختی، تکنیک کم‌وبیش پیچیده‌اي است که به‌منظور افزایش دقت مدل در آن چندین درخت تصمیم آموزش داده می‌شود. نتیجه به‌دست آمده پیش‌بینی گروهی از درختان تصمیم است (Breiman، ۲۰۰۱). ریشه بسیاری از تکنیک‌های آموزش گروهی بر پایه این فرض است که دقت آن‌ها از دیگر الگوریتم‌های آموزشی بالاتر است چون ترکیبی از چند مدل پیش‌بینی، دقیقتر از یک مدل می‌باشد و گروه‌ها قدرت مجموعه‌های منفرد و منحصربه‌فرد از طبقه‌ها را بیشتر می‌کنند، درحالی‌که هم‌زمان نقاط ضعف طبقه‌ها را کاهش می‌دهند (Kotsiantis و Pintelas، ۲۰۰۴).

- مدل درخت M5

زیرساخت مدل‌های درختی برپایه روش جدا سازی و چیرگی[۶] است. جایگزینی معادله رگرسیون خط به‌جای برچسب در گره‌ها، شیوه‌ای است که در مدل M5 اجرا می‌شود و می‌تواند متغیرهای عددی پیوسته را پیش‌بینی یا برآورد کند. ساخت مدل درخت در دو مرحله انجام می‌گیرد. در مرحله اول، درخت تصمیم با انشعاب‌سازی داده‌ها تشکیل می‌شود. اندازه انشعاب در مدل M5 بیشینه‌سازی کاهش انحراف معیار داده‌ها در گره فرزند است. مرحله دوم نیز دربرگیرنده کوچک کردن درخت بیش از حد بزرگ شده از راه هرس‌کردن شاخه‌ها و جایگزین‌شدن با توابع رگرسیون خطی است.

- رگرسیون فرآیند گاوسی (GPR)

مجموعه داده S با n مشاهده را در نظر بگیرید:

$S=\{(x_i,y_i)|\ i=1,\ldots,n\}$

که در آن x_i بردار ورودی با D بعد و y_i خروجی اسکالر یا هدف می‌باشد. این مجموعه متشکل از دو جز ورودی و خروجی به‌عنوان نقاط نمونه یا تجربی معرفی خواهند شد. شکل برداری داده ورودی‌های مجموعه در ماتریس $X=[x_1,x_2,\ldots,x_n]$ و

خروجی‌ها در ماتریس $Y=[y_1,y_2,\ldots,y_n]$ گردآوری می‌شوند. وظیفه رگرسیون، ایجاد یک ورودی جدید x^* به‌منظور دستیابی به توزیع پیش‌بینی شده برای مقادیر متناظر داده‌های مشاهداتی y^* و بر پایه مجموعه داده S می‌باشد. فرآیند گاوسی مجموعه‌ای از متغییرهای تصادفی است که تعداد محدودی از آن‌ها با توزیع‌های گاوسی یکی شده‌اند. توزیع گاوسی در واقع توزیع بین متغیرهای تصادفی است؛ اما فرآیند گاوسی بیانگر توزیع بین توابع می‌باشد. فرآیند گاوسی $f(x)$ توسط توابع میانگین $m(x)$ (رابطه ۲) و کواریانس به شکل رابطه ۳ تعریف می‌شود (Kuss، ۲۰۰۶):

$$m(x)=E(f(x)) \quad (۲)$$

$$k(x,x')=E(f(x)-m(x))(f(x')-m(x')) \quad (۳)$$

که در آن، $k(x,x')$ تابع کرنل بوده که در نقاط x و x' محاسبه می‌شود. فرآیند گاوسی $f(x)$ به‌صورت ذیل بیان می‌شود (رابطه ۴):

$$f(x)\sim GP(m(x),k(x,x')) \quad (۴)$$

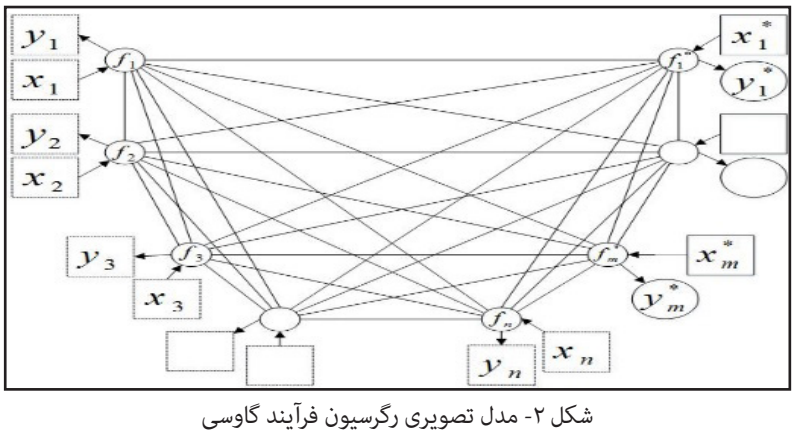
برای ساده‌سازی، مقدار تابع میانگین صفر فرض می‌شود. در فرآیند گاوسی رابطه بین بردار ورودی و هدف به‌صورت ذیل می‌باشد (رابطه ۵):

$$y_i=f(x_i)+\varepsilon \quad (۵)$$

که در آن $f(x)$ بیانگر تابع رگرسیون دلخواه و ε نیز مقدار خطای توزیع گاوسی با میانگین صفر و واریانس σ^2 می‌باشد، یعنی $\varepsilon\sim N(0,\sigma^2)$. علاوه‌بر این، فرض می‌شود $f=[f(x_1),f(x_2),\ldots,f(x_n)]^T$ رفتاری بر پایه فرآیند گاوسی داشته و به گونه‌ای که $p(f|X)=N(0,K)$ است و در آن K ماتریس کواریانس با درایه‌های $k_{i,j}=k(x_i,x_j)$ می‌باشد (رابطه ۶).

$$K(X,X)=\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \ldots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \ldots & k(x_2,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \ldots & k(x_n,x_n) \end{pmatrix} \quad (۶)$$

مقدار $k_{i,j}$ کواریانس بین مقادیر توابع نهان $f(x_i)$ و $f(x_j)$ می‌باشد. رگرسیون فرآیند گاوسی به‌منظور محاسبه توزیع پیش‌بینی شده برای مقادیر تابع f^* در نقاط تست $X^*=[x_1^*, x_2^*,\ldots, x_m^*]$ به‌کار می‌رود. مدل تصویری فرآیند گاوسی در شکل (۲) ارائه شده است. در این شکل f_i بیانگر $f(x_i)$ می‌باشد. مجموعه توابع نهان f_i که با شاخص x_i نشان داده شده‌اند، به‌طور کامل به یکدیگر پیوسته می‌باشند. هر اتصال نشان دهنده یک رابطه بین دو متغیر نهان بوده که توسط تابع کواریانس تعریف می‌شود.



شکل ۲- مدل تصویری رگرسیون فرآیند گاوسی

در شکل (۲) مربعات نشان‌دهنده متغیرهای مشاهداتی و دایره‌ها نمایانگر پارامترهای ناشناخته می‌باشند. توزیع y مشروط به اندازه‌های f بوده که با یک گاوسین همگرا و دارای خواص فیزیکی مشابه به شکل زیر ارائه می‌شود (رابطه ۷):

$$p(y|f,X)=N(f,\sigma^2_n I) \quad (۷)$$

که در آن: I ماتریس همانی می‌باشد. باتوجه‌به ویژگی‌های تابع گاوسی، توزیع حاشیه‌ای y به شکل ذیل تعیین می‌شود (رابطه ۸):

$$p(y|X)=\int p(y|f,X)p(f|X)df=N(0,K+\sigma^2_n I) \quad (۸)$$

توزیع یکپارچه شده اندازه‌های مشاهداتی که خروجی مورد نظر می‌باشند و اندازه‌های تابع در نقاط تست به‌صورت ذیل نوشته می‌شوند:

باتوجه‌به رابطه (۴) و استفاده از قواعد مدل پیش‌بینی برای به‌کار بردن محدودیت‌های روش مورد استفاده گاوسی، می‌توان توزیع شرطی را به‌دست آورد (روابط ۹ تا ۱۱):

$$p(f_*|X,y,X_*)\sim N(\bar{f}_*, cov(f_*)) \quad (۹)$$

$$\bar{f}_* = K(X_*,X)[K(X,X)+\sigma^2 I]^{-1}y \quad (۱۰)$$

$$cov(f_*)=K(X_*,X_*)-K(X_*,X)[K(X,X)+\sigma^2 I]^{-1})K(X,X_*) \quad (۱۱)$$

- مختصات جغرافیایی[۷] (روش گرافیکی)

از دسته روش‌های مورد استفاده برای بازسازی داده‌های گم‌شده، روش مختصات جغرافیایی یا روش گرافیکی می‌باشد. در این روش فاصله هر نقطه که داده گم‌شده دارد با نقاط اطرافش که

برای برآورد در نظر گرفته شده‌اند، محاسبه می‌شود. روشن است ایستگاه‌های نزدیکتر به ایستگاه مدنظر سهم بیشتری در بازسازی آن خواهند داشت؛ بنابراین لازم است ضریب وزنی بزرگتری به آن اختصاص داده شود. این ضریب وزنی با استفاده از رابطه (۱۲) محاسبه می‌شود.

$$W= 1/(x^2+ y^2) \quad (۱۲)$$

که در آن x و y به‌ترتیب طول و عرض مختصاتی ایستگاه می‌باشد. در نهایت داده‌های گم‌شده در ایستگاه هدف با استفاده از رابطه (۱۳) محاسبه می‌شود.

$$N_x = \frac{\sum_{i=1}^{n} W_i N_i}{\sum_{i=1}^{n} W_i} \quad (۱۳)$$

که در آن N_x مقدار برآورد شده داده گم‌شده در ایستگاه x، N_i اندازه داده موجود در ایستگاه i و n شناساننده شمار ایستگاه‌هایی است که برای برآورد داده گم‌شده، از داده‌های آن‌ها استفاده شده است.

- نسبت نرمال[۸]

روش نسبت نرمال ابتدا توسط (Paulhus و Kohle، ۱۹۵۲) برای تخمین داده‌های گم‌شده بارندگی به‌کار رفت و در ادامه توسط (Young، ۱۹۹۲) اصلاح شد. این روش بیشتر به میانگین نسبت داده‌های بین ایستگاه‌های شاهد و ایستگاه هدف بستگی دارد. در این روش بارش روزانه در ایستگاه هدف با نسبت میانگین بارش روزانه در ایستگاه هدف به میانگین بارش روزانه در ایستگاه‌های شاهد ضربدر بارش روزانه هم‌زمان ایستگاه شاهد هماهنگ است و که با استفاده از رابطه (۱۴) محاسبه می‌شود.

$$N_x=\frac{1}{n}\sum_{i=1}^{n}\frac{\overline{N}_x}{\overline{N}_i}N_i \quad (۱۴)$$

که در آن $\overline{N}_x$ میانگین داده‌های بارش روزانه در ایستگاه هدف، $\overline{N}_i$ میانگین داده‌های بارش روزانه در ایستگاه شاهد iام و N_i داده‌های بارش روزانه در ایستگاه iام می‌باشند.

- ضریب همبستگی وزنی[۹]

در این روش به‌منظور برآورد داده گم‌شده در ایستگاه هدف، از ضرایب همبستگی ایستگاه‌های شاهد استفاده می‌شود. کارآیی این روش به قدرت همبستگی بین ایستگاه هدف و ایستگاه‌های اطراف بستگی دارد. برای برآورد داده گم‌شده با استفاده از این روش از رابطه (۱۵) استفاده می‌شود (Teegavarapu و Chandramouli، ۲۰۰۵).

$$N_x = \sum_{i=1}^{n}\left(\frac{(r_i)}{\sum_{i=1}^{n}(r_i)}\right)N_i \quad (۱۵)$$

که در آن r_i ضریب همبستگی پیرسون بین داده‌های بارش روزانه ایستگاه هدف و ایستگاه شاهد i می‌باشد.

- نسبت نرمال با مختصات جغرافیایی

روش NRGC هر دو روش NR و GC را ترکیب می‌کند. این روش برای پیش بینی داده‌های از دست رفته بارندگی استفاده شده و بیشتر به‌عنوان یکی از بهترین روش‌ها برای تخمین داده‌های از دست رفته در نظر گرفته می‌شود. زیرا موقعیت ایستگاه‌ها را برای دستیابی به بهترین کارآیی ساماندهی می‌کند و جنبه‌های هر دو روش را با هم ترکیب می‌کند. برای روش NRGC، داده‌های از دست رفته با رابطه (۱۶) محاسبه می‌شوند.

$$N_x = \sum_{i=1}^{n}\left(\frac{\left(\frac{1}{x^2+y^2}\right)\left(\frac{\overline{N}_x}{\overline{N}_i}\right)}{\sum_{i=1}^{n}\left(\frac{1}{x^2+y^2}\right)\left(\frac{\overline{N}_x}{\overline{N}_i}\right)}\right)N_i \quad (۱۶)$$

- روش وزن‌دهی معکوس

IDW یک روش گسترده برای پر کردن داده‌های از دست رفته است. در این روش، محاسبه اندازه‌های از دست رفته بارندگی به فاصله بین ایستگاه هدف و ایستگاه‌های اطراف بستگی دارد. بیشترین وزن به نزدیکترین ایستگاه اعمال می‌شود. در این روش، داده‌های گمشده با استفاده از داده‌های مشاهده شده در ایستگاه‌های مجاور، با استفاده از رابطه (۱۷) محاسبه می‌شوند.

$$N_x = \sum_{i=1}^{n}\left(\frac{\left(\frac{1}{d_i^k}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^k}\right)}\right)N_i \quad (۱۷)$$

که در آن d فاصله هر ایستگاه از ایستگاه هدف و k حساسیت هر ایستگاه می‌باشد.

- توابع مفصل

کوپلاها توابعی هستند که یک توزیع دو یا چند متغیره را براساس دو یا چند تابع توزیع حاشیه‌ای تک متغیره تشکیل می‌دهند. با در نظر گرفتن ایستگاه مبنا و هدف تحت دو متغیر تصادفی X و Y به‌منظور تخمین اطلاعات گم شده با توابع توزیع حاشیه‌ای Fx(x) و Fy(y) باشد و با در نظر گرفتن $F_{X,Y}(x,y)$ به‌عنوان تابع توزیع تجمعی پیوسته (Joint Cumulative Distribution Function)، کوپلای C به شکل زیر تعریف می‌شود (رابطه ۱۸):

$$F_{X,Y}(x,y) = C(F_X(x),F_Y(y)) \quad (۱۸)$$

استفاده از توابع مفصل برای مدل‌سازی انعطاف بالایی دارد، چرا که برای ساخت یک مدل چند متغیره، توزیع‌های حاشیه‌ای می‌توانند به‌صورت مستقل از هم انتخاب شوند (Srinivas و همکاران، ۲۰۰۶). توابع مفصل ارشمیدسی به‌طور گسترده‌ای در تحلیل‌های دو یا چند متغیره استفاده می شود (Genest و Rivest، ۱۹۹۳).

- مفصل‌های ارشمیدسی

مفصل‌ها در طبقه‌بندی‌های مختلفی همچون T، بیضی، گاوسی و ارشمیدسی جای می‌گیرند. در این میان، خانواده مفصل‌های ارشمیدسی در علوم و مهندسی آب کاربرد بیشتری دارد. توابع مفصل ارشمیدسی در رده دو متغیره به شکل زیر تعریف می‌شود (رابطه ۱۹):

$$C_\theta(u,v)= \varphi^{-1}(\varphi(u)+\varphi(v)) \quad (۱۹)$$

φ به‌عنوان مولد مفصل معرفی شده و ویژگی‌های پیوسته، محدب و نامنفی بودن را دارد. مفصل‌های ارشمیدسی مختلف به‌ازای مقادیر مختلفی از φ ایجاد می‌شوند. نمایش ریاضی توابع پرکاربرد ارشمیدسی دربرگیرنده کلایتون، گامبل هوگارد و فرانک در جدول ذیل آورده شده است. در این توابع پارمتر تتا بیانگر درجه وابستگی بین متغیرهای وابسته را بیان می‌کند. همچنین u و v تابع پراکندگی تجمعی متغیرهای مورد بررسی می‌باشند. روابط مربوط به محاسبه تاو کندال و مولدهای هرکدارم از مفصل‌ها در جدول (۱) آورده شده است.

جدول ۱- توابع مفصل ارشمیدسی و ویژگی‌های مرتبط با آن‌ها

Copula	C(u,v)	Generator φ(t)	Generator inverse $\varphi^{-1}(t)$	Kendall's Tau
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$(t^{-\theta}-1)/\theta$	$(1+\theta t)^{-1/\theta}$	$\theta/(\theta+2)$
Gumbel	$e^{-[(-\ln u)^\theta+(-\ln v)^\theta]^{1/\theta}}$	$(-\ln(t))^\theta$	$\exp(-t)^{1/\theta}$	$(\theta-1)/\theta$
Frank	$-\frac{1}{\theta}\ln\left[1+\frac{(e^{-u\theta}-1)(e^{-v\theta}-1)}{(e^{-\theta}-1)}\right]$	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$-\frac{1}{\theta}\ln(1+\exp(-t)(\exp(-\theta)-1)^{\frac{1}{\theta}}$	$\frac{1-4\{1-D_1(\theta)\}}{\theta}$

* در رابطه مربوط به خانوده فرانک، D_1 تابع دیبای (Debye) مرتبه ۱ بوده و به‌صورت $D_1(\theta)=\int_0^\theta\left(\frac{1}{\theta((e^t-1)}\right)dt$ تعریف می‌شود. در جدول فوق، u و v توزیع های تک متغیره و θ پارامترهای وابستگی هستند.

- ایجاد داده‌های تصادفی با هدف پرکردن داده‌های گمشده

به‌منظور پر کردن داده‌های گمشده یکسری داده تصادفی (v_2) بین ۱-۰ و هم سایز با داده‌های گمشده ایجاد می‌شود. سپس از روش سری v_2 ایجاد شده، مقادیر t براساس معادله زیر محاسبه خواهد شد. روابط مربوط به K(t) (توزیع کندال) در جدول (۲) آورده شده است.

$$t=K^{-1}(v_2) \quad (۲۰)$$

$$K(t)=t-(\varphi(t)/\varphi'(t)) \quad (۲۱)$$

جدول۲- اجزای مربوط به K(t) به جداسازی هر مفصل

Copula	K(t)=t-(φ(t)/φ'(t))	Generator φ(t)	φ'(t)
Gumbel	t-(ulnu/θ)	$(-\ln(t))^\theta$	$-\theta(\ln(t))^{(\theta-1)}(1/t)$
Clayton	$t-((t^{(\theta+1)}-t)/\theta)$	$(t^{-\theta}-1)/\theta$	$-\theta t t^{(-\theta-1)}$
Frank	$u-\frac{\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)}{\theta}(e^{-\theta t}-1)$	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$\theta/(1-e^{\theta t})$

• ارزیابی نتایج

به‌منظور ارزیابی دقت روش‌های برآورد بارش با اندازه‌های بارش واقعی، از شاخص‌های آماری ضریب همبستگی (r)، ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAD) و شاخص نش ساتکلیف (NSE) استفاده شد (روابط ۲۲ تا ۲۵).

$$r=\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}\right) \quad (۲۲)$$

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}} \quad (۲۳)$$

$$MAD=\frac{\sum_{i=1}^{n}|x_i-y_i|}{n} \quad (۲۴)$$

$$NSE=1-\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{\sum_{i=1}^{n}(x_i-\bar{y})^2} \quad (۲۵)$$

در این روابط x_i و y_i به‌ترتیب i امین داده واقعی و برآورد شده، $\bar{x}$ و $\bar{y}$ به‌ترتیب میانگین داده‌های واقعی و برآورد شده و n تعداد گام‌های زمانی هستند.

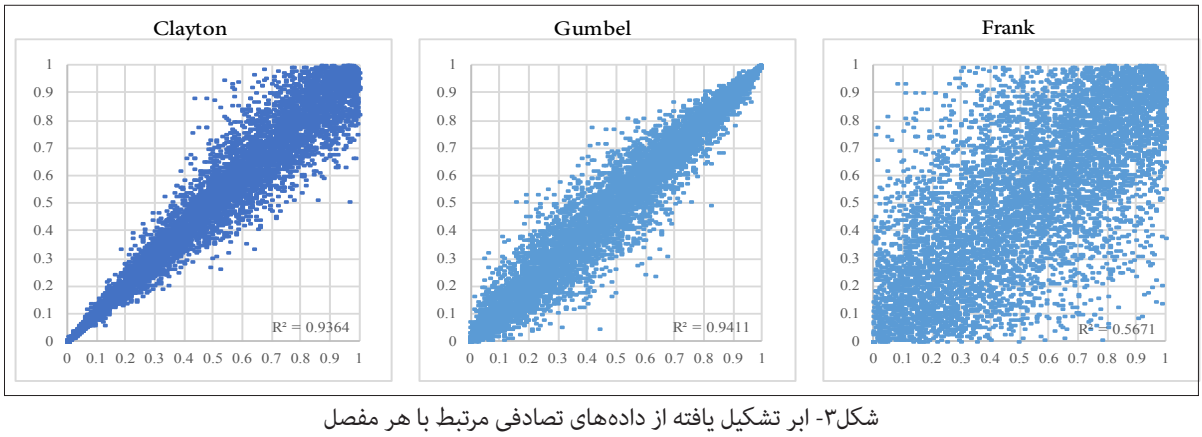
نتایج و بحث

در راستای تخمین داده‌های گم‌شده، داده‌های رویداد بارش (بزرگتر از صفر) در دو دسته آموزش و آزمون تقسیم‌بندی شدند که در میان ایستگاه‌های موجود، ایستگاه همدید تبریز بیشترین همبستگی غیر پارامتری (کندال) را با ایستگاه سهند داشت که در محاسبات و توابع مفصل شرکت داده شدند. به‌منظور تخمین داده‌های روزانه بارش ایستگاه تبریز، سه مفصل فرانک، گامبل و کلایتون بر روی توزیع متناسب با ایستگاه‌های تبریز و سهند که بیشترین همبستگی کندال را با ایستگاه تبریز داشت، برازش داده شدند که نتایج آن‌ها براساس جدول (۳) می‌باشد.

بر اساس ابر داده‌های تشکیل شده (شکل ۳)، مشخص است که مفصل گامبل همبستگی و تمرکز بالاتری پیرامون نیمساز دارد، بنابراین می‌توان بیان کرد این نوع مفصل در برآورد داده‌های گمشده روزانه بارش در ایستگاه همدید تبریز نسبت به سایر انواع مفصل برتری نسبی داشته و انتخاب می‌شود.

جدول۳- مقادیر مرتبط با توابع توزیع حاشیه هر ایستگاه

CDF	Type	Parameters	
		Tabriz	Sahand
F(x)=1-exp(-λ(x))	Exponential	λ: 0.26565	λ: 0.30376



شکل۳- ابر تشکیل یافته از داده‌های تصادفی مرتبط با هر مفصل

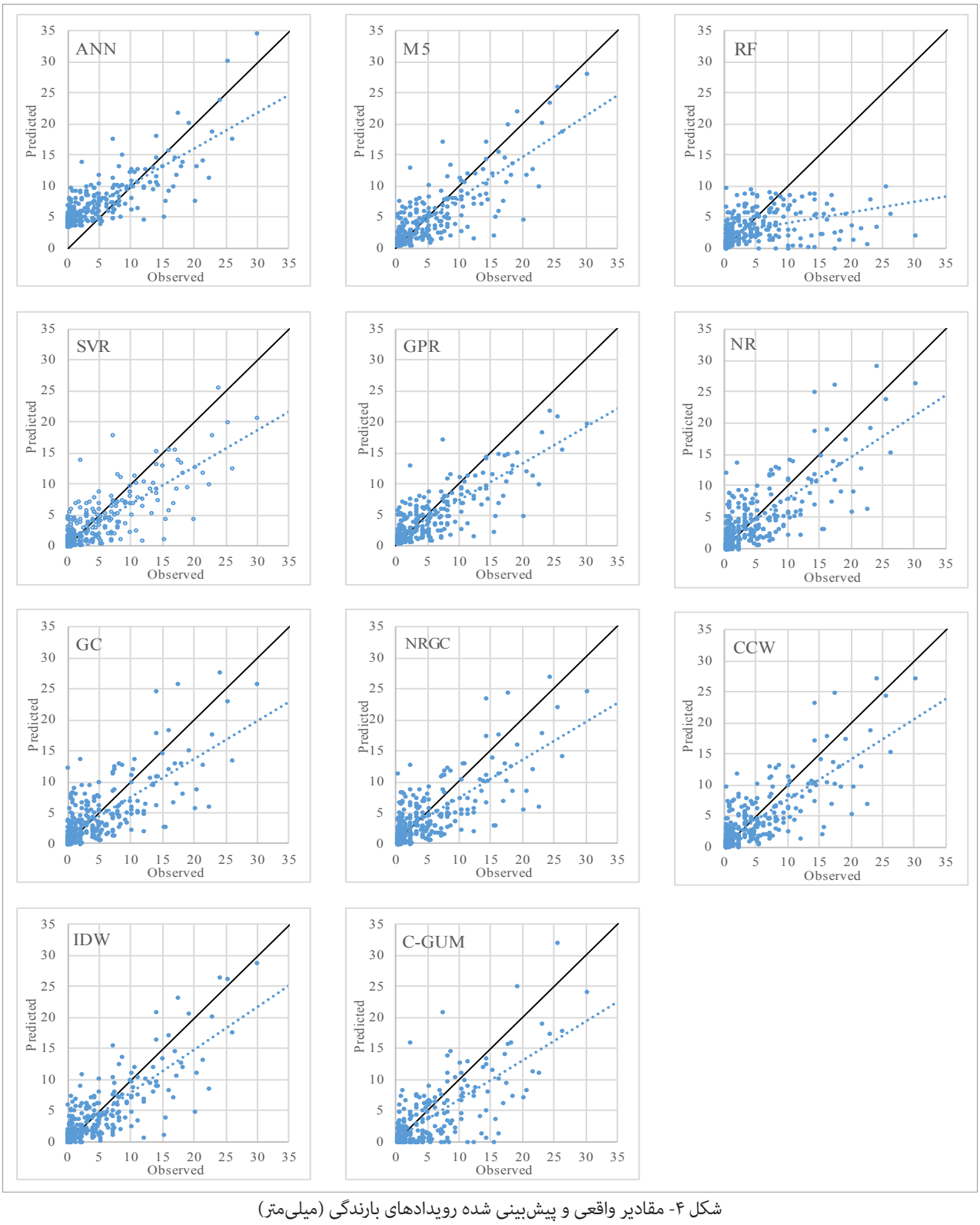
در جدول (۴) مقادیر آماره‌های ارزیابی روش‌های برآورد داده‌های بارش روزانه نشان داده شده است. براین‌اساس، در بین روش‌های هوشمند، به‌طور مشترک روش‌های M5 با مقادیر RMSE برابر ۳/۱۴ میلی‌متر و MAD برابر ۲/۱۳ میلی‌متر و GPR با اندازه‌های RMSE معادل ۳/۱۲ میلی‌متر و MAD برابر ۲/۱۷ میلی‌متر بالاترین دقت و روش RF با اندازه‌های RMSE برابر ۵/۱۸ میلی‌متر و MAD برابر ۳/۰۴ میلی‌متر و ضریب همبستگی ۰/۳۹ کمترین دقت را در برآورد رویداد بارش روزانه دارند. در بین روش‌های آماری نیز روش IDW با اندازه‌های RMSE برابر ۳/۲۱ میلی‌متر و MAD برابر ۲/۱۲ میلی‌متر به‌عنوان بهترین روش برآورد رویداد بارش روزانه معرفی شد. در نقطه مقابل، روش GC با مقادیر RMSE برابر ۳/۷۳ میلی‌متر و MAD برابر ۲/۵۵ میلی‌متر، کمترین دقت را در برآورد رویداد بارش روزانه دارد. قابل توجه است که NSE تابع مفصل گامبل در آستانه مقبولیت واقع شده و نتایج ارزیابی این روش بسیار نزدیک به سایر روش‌های برتر از نظر برآورد می‌باشد.

جدول ۴- مقادیر شاخص‌های ارزیابی روش‌های مورد استفاده

Method	r	RMSE	MAD	NS
ANN	۰,۸۱	۴,۲۸	۳,۷۶	۰,۳۹
M۵	۰,۸۲	۳,۱۴	۲,۱۳	۰,۶۷
RF	۰,۳۹	۵,۱۸	۳,۰۴	۰,۱۰
SVR	۰,۸۱	۳,۳۶	۲,۰۹	۰,۶۲
GPR	۰,۸۲	۳,۱۲	۲,۱۷	۰,۶۶
NR	۰,۷۵	۳,۶۶	۲,۵۳	۰,۵۵
GC	۰,۷۴	۳,۷۳	۲,۵۵	۰,۵۴
NRGC	۰,۷۵	۳,۶۳	۲,۴۸	۰,۵۶
CCW	۰,۷۸	۳,۴۵	۲,۳۴	۰,۶۰
IDW	۰,۸۱	۳,۲۱	۲,۱۲	۰,۶۶
C-GUM	۰,۷۴	۳,۸۹	۲,۵۱	۰,۵۰

- شکل مقادیر پیش‌بینی و مقادیر واقعی

در شکل (۴) اندازه رویدادهای واقعی بارش در برابر اندازه برآورد شده با استفاده از هر یک از روش‌های مورد بررسی نشان داده شده است. تراکم بیشتر نقاط پیرامون خط نیمساز نشان دهنده کارآیی بیشتر و دقت بالاتر هر روش در برآورد داده بارش است و هرچه خط روند داده‌ها به خط نیمساز نزدیکتر باشد به‌معنای دقت بالاتر روش خواهد بود. براین‌اساس روش RF و پس از آن ANN با اختلاف مشهودی کمترین دقت را در پیش‌بینی رویداد بارش دارند.



شکل ۴- مقادیر واقعی و پیش‌بینی شده رویدادهای بارندگی (میلی‌متر)

باتوجه‌به نتایج به دست آمده، توابع مفصل هرچند نسبت به برخی روش‌های آماری و هوشمند خطای بیشتری دارند، بااین‌حال محدوده خطای داده‌های برآوردی بسیار نزدیک به سایر روش‌ها است. ازاین‌رو باتوجه‌به قابلیت‌های توابع مفصل از جمله توانایی به‌کار بردن شرط‌های بیشمار و ماهیت روش که براساس احتمالات پایه‌گذاری شده، می‌توان بیان کرد در شرایط مشابه توانایی توابع مفصل در برآورد داده‌های گمشده پدیده‌های تصادفی و احتمالاتی همچون بارندگی و سایر پارامترهای اقلیمی و هیدرولوژیکی قابل قبول و مورد اطمینان می‌باشد، چرا که از توزیع آماری ویژه هر ایستگاه و ایستگاه مفصل شده بهره می‌گیرد. با استفاده از این برتری، مفصلی که تعیین شده می‌تواند شرط‌های احتمالاتی مد نظر، همچون آستانه بارش و یا بیشینه و کمینه بارندگی را در برآورد داده‌های گمشده لحاظ نماید. این قابلیت در پیش‌بینی و تحلیل ریسک نیز بسیار اهمیت دارد و با اتکا بر آن می‌توان مدیریت و برنامه‌ریزی آینده را باتوجه به شرط‌های احتمالاتی مد نظر مشخص و فراهم کرد. همچنین باتوجه‌به اینکه کاربرد توابع مفصل در برآورد داده گم‌شده و یا بازسازی داده‌ها کمتر مورد بررسی پژوهشگران قرار گرفته، ازاین‌رو توصیه می‌شود از این توابع برای بازسازی دیگر داده‌های هیدرولوژیکی استفاده و نتایج آن بررسی شود. همچنین باتوجه‌به اثر اقلیم هر منطقه در انتخاب و معرفی روش برگزیده، بحث اقلیم نیز باید مورد توجه قرار گیرد.

نتیجه‌گیری

باتوجه‌به اهمیت بارش در بیشتر مطالعات هیدرولوژیکی، در پژوهش حاضر، کاربرد توابع مفصل و مقایسه نتایج به‌دست آمده از آن در پیش‌بینی رویدادهای بارش روزانه با روش‌های هوشمند و آماری ارزیابی شد. به‌طور کلی نتایج به‌دست آمده برای هر سه دسته روش‌های آماری و هوشمند و مفصل در پیش‌بینی رویدادهای بارشدقت قابل قبول را نشان می‌دهد. به‌طور کلی در نتایج روش M5 بالاترین دقت را در برآورد رویدادهای بارش روزانه نشان داد. در بین روش‌های آماری نیز روش IDW به‌عنوان روش برگزیده معرفی شد. در نقطه مقابل روش RF کمترین دقت برآورد داده‌های بارش روزانه را دارد. همچنین باتوجه‌به قابلیت‌های توابع مفصل دربرگیرنده توانایی به‌کار بردن شرط‌های بیشمار و ماهیت احتمالاتی آن، استفاده از آن برای پیش‌بینی متغیرهای هیدرولوژیکی توصیه می‌شود. درحالی‌که توابع مفصل، با در نظر گرفتن چندجانبه رفتار و خود پارامتر و همچنین دوره‌های بازگشت آن، نسبت به سایر روش‌های کلاسیک و هوشمند برتری دارد. همچنین باتوجه‌به اینکه کاربرد توابع مفصل در پیش‌بینی و برآورد داده‌های هیدرولوژیکی کمتر مورد توجه قرار گرفته است، ازاین‌رو ارزیابی کارآیی توابع مفصل در برآورد متغیرهای مختلف هیدرولوژیکی و با در نظر گرفتن شرایط اقلیمی و تأثیر آن در معرفی مدل مناسب و مقایسه نتایج حاصل از توابع مفصل با روش‌های آماری و هوشمند توصیه می‌شود.

پی‌نوشت

1-Copula
2-Support Vector Regression
3-Artificial Neural Networks
4-Multi-Layer Perceptron
5-Random Forests
6-Divide and Conquer
7-Geographical Coordinates (GC)
8-Normal Ratio (NR)
9-Correlation Coefficient Weighted (CCW)

منابع

منهاج، م.ب. ۱۳۸۴. مبانی شبکه‌های عصبی (هوش محاسباتی) جلد اول. مرکز نشر دانشگاه صنعتی امیرکبیر. چاپ سوم. تهران، ایران.

Abebe A.J., Solomatine D.P. and Venneker R.G. 2000. Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. Hydrological Sciences Journal, 45(3): 425-36.

Armanuos A.M., Al-Ansari N. and Yaseen Z.M. 2020. Cross assessment of twenty-one different methods for missing precipitation data estimation. Atmosphere, 11(4): 389.

Bessa R.J., Miranda V., Botterud A., Zhou Z. and Wang J. 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renewable Energy, 1;40(1): 29-39.

Boser B.E., Guyon I.M. and Vapnik V.N. 1992. A training algorithm for optimal margin classifiers. InProceedings of the fifth annual workshop on Computational learning theory. Pennsylvania, Pittsburgh, USA.

Breiman L. 2001. Random forests. Machine learning, 45(1): 5-32.

Nguyen-Huy T., Deo R.C., An-Vo D.A., Mushtaq S. and Khan S. 2017. Copula-statistical precipitation forecasting model in Australia's agro-ecological zones. Agricultural water management, 191: 153-72.

Nguyen-Huy T., Deo R.C., Mushtaq S., An-Vo D.A. and Khan S. 2018. Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach. European journal of agronomy, 98: 65-81.

Paulhus J.L. and Kohler M.A. 1952 Interpolation of missing precipitation records. Monthly Weather Review, 80(8): 129-33.

Srinivas S., Menon D. and Meher Prasad A. 2006. Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. Journal of transportation engineering, 132(12): 945-55.

Talebizadeh M., Morid S., Ayyoubzadeh S.A. and Ghasemzadeh M. 2010. Uncertainty analysis in sediment load modeling using ANN and SWAT model. Water Resources Management, 24(9): 1747-61.

Teegavarapu R.S. and 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.

Vapnik V.N. 1998. Statistical learning theory. Wiley, New York.

Vapnik V. 1995. The nature of statistical learning theory. Springer, New York.

Young K.C. 1992. A three-way model for interpolating for monthly precipitation values. Monthly Weather Review, 120(11): 2561-9.

Yuan C. and Yamagata T. 2015. Impacts of IOD, ENSO and ENSO Modoki on the Australian winter wheat yields in recent decades. Scientific reports, 5(1): 1-8.

Zhang L. and Singh V.P. 2014. Trivariate flood frequency analysis using discharge time series with possible different lengths: Cuyahoga river case study. Journal of Hydrologic Engineering, 19(10): 05014012.

Zscheischler J. and Seneviratne S.I. 2017. Dependence of drivers affects risks associated with compound events. Science advances, 3(6): e1700263.

Genest C. and Rivest L.P. 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American statistical Association, 88(423): 1034-43.

Hassani B.K. 2016. Dependencies and relationships between variables. Scenario Analysis in Risk Management. Springer International Publishing Switzerland.

Huang J.J., Lee K.J., Liang H. and Lin W.F. 2009. Estimating value at risk of portfolio by conditional copula-GARCH method. Insurance: Mathematics and economics, 45(3): 315-24.

Jarvis C., Darbyshire R., Eckard R., Goodwin I. and Barlow E. 2018. Influence of El Niño-Southern oscillation and the Indian Ocean Dipole on winegrape maturity in Australia. Agricultural and Forest Meteorology, 248: 502-10.

Kao S.C. and Govindaraju R.S. 2010. A copula-based joint deficit index for droughts. Journal of Hydrology, 380(1-2): 121-34.

Kotsiantis S. and Pintelas P. 2004. Combining bagging and boosting. International Journal of Computational Intelligence, 1(4): 324-33.

Kuan C.M. and White H. 1994. Artificial neural networks: An econometric perspective. Econometric reviews, 13(1): 1-91.

Kuss M. 2006. Gaussian process models for robust regression, classification, and reinforcement learning. Ph. D. dissertation, Technische Universität Darmstadt, Darmstadt, Germany.

Liu Z., Zhou P., Chen X. and Guan Y. 2015. A multivariate conditional model for streamflow prediction and spatial precipitation refinement. Journal of Geophysical Research: Atmospheres, 120(19): 10-16.

Lu X.F., Lai K.K. and Liang L. 2014. Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model. Annals of operations research, 219(1): 333-57.

Manning C., Widmann M., Bevacqua E., Van Loon A.F., Maraun D. and Vrac M. 2018. Soil moisture drought in Europe: a compound event of precipitation and potential evapotranspiration on multiple time scales. Journal of Hydrometeorology, 19(8): 1255-71.