



## Investigating the effect of precipitation series decomposition on the simulation of electrical conductivity of river flow (case study: Eskandari sub-basin)

Mohammad Nazeri Tahroudi<sup>1\*</sup>, Rasoul Mirabbasi Najafabadi<sup>2</sup>

1. Postdoctoral Researcher, Department of Water Engineering, Shahrood University, Shahrood, Iran.

2. Associate Professor, Department of Water Engineering, Shahrood University, Shahrood, Iran.

\*Corresponding Author: [m\\_nazeri2007@birjand.ac.ir](mailto:m_nazeri2007@birjand.ac.ir)

### Keywords:

Approximate Signal, Joint Simulation, Random Forest, Random Tree, Wavelet Theory.

### Extended abstract

#### Introduction

By modeling and simulating the qualitative values of river flow according to meteorological parameters as well as climatic changes, changes in the meteorological phase can be considered indirectly. On the other hand, due to the random measurement of qualitative variables, it is not possible to estimate the daily time series of qualitative variables. Also, due to climate changes and the sharp decrease in river flow rate in recent years, some rivers in Iran have changed from permanent to seasonal. This issue has caused the quality variables of river flow not to be measured continuously. Based on this, simulating qualitative values based on other effective and available parameters can help fill this gap in certain conditions.

With wavelet theory being capable of improving the simulation results of meteorological and hydrological values, the present study has attempted to examine the effectiveness of this model in simulating the electrical conductivity values of surface water in response to changes in rainfall values. Therefore, the aim of this study is to investigate the effect of rainfall and rainfall decomposition series in the simulation of EC values in the Eskandari sub-basin located in the Zayandeh Rood Dam Basin using random forest and random tree models. In this study, an attempt has been made to investigate the effect of independent series decomposition on the simulation of dependent series using two learning algorithms.

### Received:

Dec/28/2022

### Revised:

Jan/09/2023

### Accepted:

Jan/10/2023

### How to cite this article:

Nazeri Tahroudi, M., & Mirabbasi Najafabadi, R. (2023). Investigating the effect of precipitation series decomposition on the simulation of electrical conductivity of river flow (case study: Eskandari sub-basin). *Journal of Drought and Climate Change Research*, 1(1), 33 - 48. [10.22077/JDCR.2023.5950.1005](https://doi.org/10.22077/JDCR.2023.5950.1005).



### Materials and methods

In this study, the effect of rainfall time series decomposition on the simulation of electrical conductivity values of the surface water of the Eskandari sub-basin in the northwest of the Zayandeh Rood Dam Basin in the statistical period of 1990-2020 was investigated by two algorithms based on random tree and random forest decomposition of observation precipitation series was done using wavelet theory, Daubechies 4 and level 2. The simulation of electrical conductivity values in the studied sub-basin was initially done by using two random tree and random forest algorithms in two phases of training and testing based on the rainfall values that correspond to the electrical conductivity values on a daily basis.

### Results and discussion

By decomposition the rainfall observation series into approximate and detail signals and increasing the simulation dimension to 4 dimensions, the simulation results of EC values in the study area revealed that in the case of the random forest algorithm, the decomposition of rainfall series was able to reduce the simulation error and improve the performance compared to the initial state of the random forest algorithm by 77% in the training phase and 10% in the testing phase. The results showed that the efficiency of the model was 0.67 and 0.73 in the training phase for random forest and random tree algorithms, respectively, and the efficiency was 0.59 and 0.55 in the testing phase for the mentioned algorithms by the Nash-Sutcliffe efficiency. In comparison to the random tree algorithm, the results of the decomposition of rainfall observation series into approximate and detail signals and their involvement in the simulation of electrical conductivity values showed that the Wavelet-RT model was able to reduce the amount of simulation error by about 54% in the training phase and about 22% in the testing phase. In general, the results showed that the decomposition of the rainfall observation series into approximate and detail signals can greatly improve the accuracy, efficiency and error rate of the simulation. By decomposing the series of observations into approximate and detail signals, the dimensions of the simulation are increased and, by its nature, the complexity of modeling and simulation increases.

### Conclusion

By decomposing the rainfall series into an approximate series and two partial series and increasing the simulation dimension to 4 dimensions, the results showed that the combination of wavelet theory with random forest and random tree algorithms was able to reduce the simulation error (RMSE) of the conductivity values and improve the algorithm's error in the training phase compared to the random forest and random tree algorithms by about 77.5 and 54%, respectively. These results in the testing phase are about 10 and 22 percent, respectively.



## بررسی اثر تجزیه سری بارش بر شبیه‌سازی هدایت الکتریکی جریان رودخانه (مطالعه

موردی: زیر حوضه اسکندری)

محمد ناظری تهرودی<sup>۱\*</sup>، رسول میرعباسی نجف آبادی<sup>۲</sup>

۱. پژوهشگر پسادکتری، گروه مهندسی آب، دانشگاه شهرکرد، شهرکرد، ایران.

۲. دانشیار گروه مهندسی آب، دانشگاه شهرکرد، شهرکرد، ایران.

\*نویسنده مسئول: [m\\_nazeri2007@birjand.ac.ir](mailto:m_nazeri2007@birjand.ac.ir)

### چکیده

### واژه‌های کلیدی:

در این مطالعه تأثیر تجزیه سری زمانی بارش در شبیه‌سازی مقادیر هدایت الکتریکی آب سطحی زیر حوضه اسکندری در شمال غرب حوضه آبریز سد زاینده‌رود در دوره آماری ۹۹-۱۳۶۹ توسط دو الگوریتم پایه درخت تصادفی و جنگل تصادفی مورد بررسی قرار گرفت. تجزیه سری مشاهداتی بارش با استفاده از تئوری موجک، دابچیز ۴ و در سطح ۲ انجام شد. در ابتدا شبیه‌سازی مقادیر هدایت الکتریکی در زیر حوضه مورد مطالعه با استفاده از دو الگوریتم درخت تصادفی و جنگل تصادفی در دو فاز آموزش و آزمایش با توجه به مقادیر بارش متناظر با مقادیر هدایت الکتریکی در مقیاس روزانه انجام گردید. نتایج بیانگر کارایی مدل ۰/۶۷/۰/۷۳ در فاز آموزش به ترتیب برای الگوریتمهای جنگل تصادفی و درخت تصادفی و کارایی ۰/۵۹/۰/۵۵ در فاز آزمایش برای الگوریتمهای یاد شده توسط آماره نش-ساتکلیف بود. با تجزیه سری بارش به یک سری تقریبی و دو سری جزئی و افزایش بعد شبیه‌سازی تا ۴ بعد، نتایج نشان داد که تلفیق تئوری موجک با الگوریتمهای جنگل تصادفی و درخت تصادفی توانسته است میزان خطای شبیه‌سازی (RMSE) مقادیر هدایت الکتریکی را در فاز آموزش نسبت به دو الگوریتم جنگل تصادفی و درخت تصادفی به ترتیب حدود ۷۷/۵ و ۵۴ درصد بهبود بخشد. این ارقام در فاز آزمایش به ترتیب حدود ۱۰ و ۲۲ درصد می‌باشد. نتایج نشان داد که با تجزیه سری مشاهداتی به سیگنال‌های جزئی و تقریبی، میزان خطا و کارایی مدل‌های مورد بررسی بهبود یافته است.

تئوری موجک، جنگل تصادفی، درخت تصادفی، سیگنال تقریبی، شبیه‌سازی توأم.

### تاریخ دریافت:

۱۴۰۱/۱۰/۰۷

### تاریخ ویرایش:

۱۴۰۱/۱۰/۱۹

### تاریخ پذیرش:

۱۴۰۱/۱۰/۲۰

## مقدمه

با مدل‌سازی و شبیه‌سازی مقادیر کیفی جریان با توجه به پارامترهای هواشناسی و همچنین تغییرات اقلیمی می‌توان به‌صورت غیرمستقیم تغییرات فاز هواشناسی را در نظر گرفت. از طرفی با توجه به اندازه‌گیری متغیرهای کیفی به‌صورت تصادفی، نمی‌توان سری زمانی روزانه‌ی متغیرهای کیفی را برآورد نمود. همچنین با توجه به تغییرات اقلیمی و کاهش شدید دبی جریان در سال‌های اخیر (Zamani et al., 2018; Pronoos Sedighi et al., 2022; Nazeri Tahroudi et al., 2022a&b)، برخی از رودخانه‌ها در کشور ایران از حالت دائمی به حالت فصلی تبدیل شده است. این موضوع سبب شده است که اندازه‌گیری متغیرهای کیفی جریان رودخانه به‌صورت مستمر انجام نشود. بر این اساس، شبیه‌سازی مقادیر کیفی با توجه به سایر پارامترهای مؤثر و در دسترس به‌خوبی می‌تواند این خلأ را در شرایط خاص برطرف نماید. در این خصوص مدل‌های تک متغیره و چند متغیره فراوانی توسعه‌یافته است که هر کدام شرایط خاص خود را دارد. در حالت تک متغیره و یا حتی دو متغیره در صورت عدم وجود متغیر وابسته دیگر که همبستگی بالا داشته باشد، می‌توان سری زمانی را با روش‌های مختلف همچون تئوری موجک تجزیه کرده و در حالت چند متغیره مدل را اجرا کرد (Bageri et al., 2022).

در سال‌های گذشته پیشرفت قابل توجهی در توسعه مدل‌های هیدرولوژیک با کارایی مناسب در پیش‌بینی مقادیر مختلف هواشناسی و هیدرولوژیک مشاهده شده است. در میان مدل‌های مختلف مفهومی و جعبه سیاه توسعه‌یافته در این دوره، مدل‌های مبتنی بر موجک ترکیبی و هوش مصنوعی از مهم‌ترین مدل‌ها در شبیه‌سازی فرآیندهای هیدرولوژیک بوده‌اند (Nourani et al., 2014). ادغام مدل‌های مبتنی بر تئوری موجک با مدل‌های مختلف از قبیل مدل‌های هوش مصنوعی می‌تواند نتایج شبیه‌سازی را بهبود دهد. رویکرد مدل‌سازی مبتنی بر شبکه عصبی مصنوعی در دو دهه گذشته توجه محققان را به خود جلب کرده است و با موفقیت در زمینه‌های مختلف

از جمله تخمین و پیش‌بینی جریان رودخانه استفاده شده است (Hu et al., 2005; Campolo et al., 2003; Nilsson et al., 2006; Wei et al., 2013; Darbandi and Pourhosseini, 2018; Meresa, 2019; Seifi and Riahi-Madvar, 2019; Feng and Niu, 2021; Hassanjabar et al., 2022; Mir-Jay- (zaee and Saraf, 2022). جایاوردنا<sup>۱</sup> و همکاران (awardena et al., 2004) از تجزیه موجک و مدل مارکوف برای شبیه‌سازی بارش در تایلند استفاده کردند. کاناس<sup>۲</sup> و همکاران (Cannas et al., 2006) از تبدیل موجک و شبکه‌های عصبی برای پیش‌بینی جریان رودخانه استفاده کرده و قطعیت و دقت آن را تأیید نمودند. باجیرائو و همکاران (Bajirao et al., 2021) کارایی مدل تلفیقی شبکه عصبی مصنوعی-موجک را در شبیه‌سازی رواناب روزانه مورد بررسی قرار دادند. نتایج نشان داد که مدل‌های ترکیبی با تئوری موجک، بهتر از مدل‌های ساده عمل می‌کنند. با توجه به دقت قابل قبول تئوری موجک در بهبود نتایج شبیه‌سازی مقادیر هواشناسی و هیدرولوژیک، در این مطالعه سعی شده است کارایی این مدل در شبیه‌سازی مقادیر هدایت الکتریکی آب سطحی با توجه به تغییرات بارش مورد بررسی قرار گیرد. لذا هدف این مطالعه بررسی تأثیر بارش و تجزیه سری بارش در شبیه‌سازی مقادیر EC در زیرحوضه اسکندری واقع در حوضه آبریز سد زاینده‌رود با استفاده از مدل‌های پایه جنگل تصادفی و درخت تصادفی می‌باشد. در این مطالعه سعی شده است تأثیر تجزیه سری مستقل در شبیه‌سازی سری وابسته با استفاده از دو الگوریتم یادگیری مورد بررسی قرار گیرد.

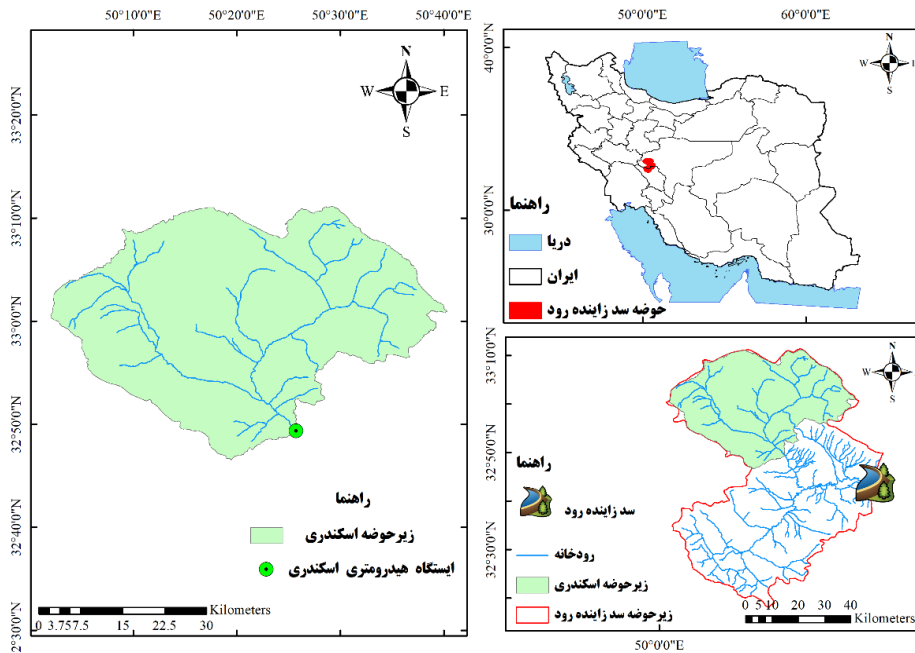
<sup>1</sup> Jayawardena

<sup>2</sup> Cannas

شکل ۱ موقعیت منطقه مورد مطالعه را نشان می‌دهد. جدول ۱ نیز خصوصیات داده‌های مورد بررسی را در دوره آماری ۹۹-۱۳۶۹ نشان می‌دهد. داده‌های مورد استفاده در این تحقیق مقادیر هدایت الکتریکی اندازه‌گیری شده در ایستگاه هیدرومتری اسکندری و بارش متناظر آن در دوره آماری ۹۹-۱۳۶۹ می‌باشد.

## مواد و روش‌ها منطقه مورد مطالعه

منطقه مورد مطالعه در این تحقیق ایستگاه هیدرومتری اسکندری، یکی از زیرحوضه‌های حوضه آبریز سد زاینده‌رود است. زیرحوضه اسکندری با مساحت ۱۵۹ کیلومترمربع در شمال غربی آن واقع شده است.



شکل ۱- موقعیت حوضه آبریز سد زاینده‌رود و زیرحوضه اسکندری  
Fig. 1. The location of Zayanderoud dam basin and the Eskandari sub-basin

جدول ۱- خصوصیات آماری داده‌های مورد بررسی  
Table 1- Statistical characteristics of the studied data

پارامتر	واحد اندازه‌گیری	انحراف معیار	میانگین	بیشینه	کمینه
Parameter	Unit	Std	Mean	Max	Min
هدایت الکتریکی	میکرو زیمنس بر سانتی‌متر	72.2	446	591	231
EC	$\mu\text{S/cm}$				
بارش	میلی‌متر	10.9	11.8	45	0.5
Rainfall	mm				

تولید می‌شود. همچنین همه بردارها از توزیع مشابهی تبعیت می‌کنند. رگرسیون درختی با استفاده از مجموعه داده‌های آموزش و  $X_n$  محاسبه شده مجموعه درخت‌هایی برابر با  $n$  را به شرح زیر تولید می‌نماید (Breiman, 2001):

$$X_n = \{h_1(x), h_2(x), \dots, h_n(x)\} \quad (1)$$

## الگوریتم جنگل تصادفی (RF)

جنگل تصادفی چندین درخت تصمیم ترکیب شده است که در ساخت آن چندین نمونه خودسازمان یافته از داده‌ها شرکت دارند (Friedman et al., 2001). در مدل جنگل بردار تصادفی  $X_n$  که مستقل از بردارهای تصادفی  $X_1, X_2, \dots, X_{n-1}$  بوده، برای درخت  $n$ ام،

است که برای زیرفضای محلی توصیف شده توسط این برگ بهینه شده است. جنگل‌های تصادفی نشان داده‌اند که عملکرد درختان تک، تصمیم را به‌طور قابل توجهی بهبود می‌بخشد: تنوع درختی با دو روش تصادفی‌سازی ایجاد می‌شود. ابتدا داده‌های آموزشی با جایگزینی برای هر درخت نمونه‌برداری می‌شود. ثانیاً، هنگام رشد یک درخت، به جای اینکه همیشه بهترین تقسیم ممکن برای هر گره را محاسبه کنیم، تنها یک زیرمجموعه تصادفی از همه ویژگی‌ها در هر گره در نظر گرفته می‌شود و بهترین تقسیم برای آن زیرمجموعه محاسبه می‌گردد (Witten and Frank, 2002; Pfahringer, 2010; Kalmegh, 2015).

### تئوری موجک

تبدیل موجک یک تبدیل ریاضی کارآمد بر پایه تبدیل فوریه بوده که در زمینه پردازش سیگنال استفاده می‌شود. موجک‌ها توابع ریاضی هستند که شکل مقیاس زمانی سری‌های زمانی و توابع آن‌ها را برای تجزیه و تحلیل سری‌های زمانی که شامل متغیرهای نا ایستا می‌شود، ارائه می‌دهند. تجزیه و تحلیل موجک فواصل طولانی مدت برای اطلاعات با فرکانس پایین و فواصل زمانی کوتاه‌تر برای اطلاعات با فرکانس بالاتر ارائه می‌دهد. تجزیه و تحلیل موجک قادر است جنبه‌های مختلفی از داده‌ها، نقاط شکست و ناپیوستگی‌های مختلف را نشان دهد که سایر روش‌های تحلیل سیگنال ممکن است قادر به نشان دادن آن نباشند. تابع موجک دارای دو ویژگی مهم نوسان و کوتاه بودن است. ضریب موجک می‌تواند در هر نقطه سیگنال (b) و برای هر مقدار مقیاس (a) با معادله زیر قابل محاسبه باشد (Mallat, 1989):

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \psi\left(\frac{t-b}{a}\right) f(t) dt \quad (4)$$

در معادله فوق،  $\psi$  تابع موجک،  $a$  برابر با مقیاس و  $b$  برابر با تبدیل است. برای مقدار  $T$  مقادیر مختلف  $a$  و  $b$  به دست می‌آید. اطلاعات بیشتر در این زمینه در تحقیقات ملات (Mallat, 1989) و سولگی و همکاران (Solgi et al., 2017) می‌باشد. متداول‌ترین توابع

$$h_n = h(x, X_n), x = \{x_1, x_2, \dots, x_p\} \quad (2)$$

بردار  $p$  بعدی فوق، یک جنگل را تشکیل داده و خروجی‌ها برای هر درخت به‌صورت زیر ارائه می‌شود:

$$y_1 = h_1(x), y_2 = h_2(x), \dots, y_n = h_n(x) \quad (3)$$

که  $\widehat{y}_n$  خروجی درخت  $n$ ام می‌باشد. برای به دست آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود (Bageri et al., 2022).

### الگوریتم درخت تصادفی<sup>۱</sup> (RT)

درخت تصادفی (RT) یک الگوریتم طبقه‌بندی کننده تحت نظارت است. RT یک الگوریتم یادگیری گروهی است که تعداد زیادی دسته‌های یادگیری را ایجاد کرده که از یک ایده بسته‌بندی برای تولید مجموعه‌ای تصادفی از داده‌ها برای ساخت درخت تصمیم استفاده می‌کند. در درخت استاندارد هر گره با استفاده از بهترین تقسیم در بین همه متغیرها تقسیم می‌شود. در یک جنگل تصادفی، هر گره با استفاده از بهترین گره در میان زیرمجموعه‌های پیش‌بینی کننده که به‌طور تصادفی در آن گره انتخاب شده‌اند، تقسیم می‌شود. درختان تصادفی توسط برایمن و کاتلر<sup>۲</sup> (Breiman and Cutler, 1993) معرفی شد. این الگوریتم می‌تواند با مسائل طبقه‌بندی و رگرسیون مقابله کند. درخت‌های تصادفی مجموعه‌ای از پیش‌بینی کننده‌های درختی هستند که به آن جنگل می‌گویند. طبقه‌بندی به این صورت عمل می‌نماید: طبقه‌بندی کننده درختان تصادفی، بردار ویژگی ورودی را می‌گیرد، آن را با هر درخت در جنگل طبقه‌بندی کرده و برچسب کلاسی را که اکثریت «رای» را دریافت کرده است، به‌عنوان خروجی ارائه می‌کند. در نمونه رگرسیون، پاسخ طبقه‌بندی کننده، میانگین پاسخ‌ها در تمام درختان جنگل است.

درختان تصادفی اساساً ترکیبی از دو الگوریتم موجود در یادگیری ماشین هستند: درختان مدل تک با ایده‌های جنگل تصادفی ترکیب می‌شوند. درختان مدل، درختان تصمیم هستند که در آن هر برگ دارای یک مدل خطی

<sup>1</sup> Random Tree

<sup>2</sup> Leo Breiman and Adele Cutler

با هر درخت در جنگل طبقه‌بندی می‌کند، و برچسب کلاسی را که اکثریت «رأی» را دریافت کرده است، خروجی می‌دهد. در صورت رگرسیون، پاسخ طبقه‌بندی کننده میانگین پاسخ‌ها در تمام درختان جنگل است (Leo, 2001; Landwehr et al., 2005; Andy, 2012).

### معیارهای ارزیابی

جذر میانگین مربعات خطا ( $RMSE$ ) و ضریب کارایی نش-ساتکلیف ( $NSE$ ) آماره‌های رایج هستند که برای انتخاب بهترین مفصل و مقایسه نتایج شبیه‌سازی استفاده می‌شوند (Nash and Sutcliffe, 1970).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - \bar{p}_i)^2} \quad (5)$$

$$NSE = 1 - \left( \frac{\sum_{i=1}^N (p_i - \bar{p}_i)^2}{\sum_{i=1}^N (p_i - p_{ave})^2} \right) \quad (6)$$

که در آن،  $p_i$  و  $\bar{p}_i$  به ترتیب برابر با مقادیر شبیه‌سازی شده و مقادیر مشاهداتی می‌باشند.  $N$  تعداد داده‌ها،  $m$  تعداد پارامترها و  $L$  حداکثر مقدار تابع احتمال برای مدل است.  $p_{ave}$  نیز برابر با متوسط مقادیر مشاهداتی می‌باشد.

### نتایج و بحث

همان‌طور که ذکر شد، در این مطالعه به شبیه‌سازی مقادیر EC در ایستگاه اسکندری در دو حالت ۱: با در نظر گرفتن تجزیه سری بارش و ۲: عدم تجزیه سری بارش پرداخته شده است. مدل پایه نیز در این مطالعه دو الگوریتم جنگل تصادفی و درخت تصادفی می‌باشد.

### نتایج شبیه‌سازی مقادیر EC در منطقه مورد مطالعه با در نظر گرفتن بارش متناظر در حالت دو بعدی

مقادیر EC ایستگاه اندازه‌گیری شده در ایستگاه هیدرومتری اسکندری در دوره آماری ۹۹-۱۳۶۹ با استفاده از دو الگوریتم درخت تصادفی و جنگل تصادفی در دو فاز آموزش و آزمایش به ترتیب به نسبت ۸۰ و ۲۰ درصد با در نظر گرفتن مقادیر بارش متناظر شبیه‌سازی شد. از بین دوره آماری مورد بررسی، مقادیر EC اندازه‌گیری شده با مقادیر متناظر بارش در مقیاس روزانه انتخاب شد. در ابتدا به بررسی همبستگی دو متغیر بارش و EC در دوره آماری مورد مطالعه با استفاده از آزمون تائو کندال پرداخته شد که نتایج به صورت شکل ۲ قابل مشاهده می‌باشد.

موجک به‌عنوان تابع موجک هار<sup>۱</sup> دنبال می‌شوند. تابع موجک هار ساده‌تر و یکی از اولین موجک‌ها است. تابع موجک Daubechies یکی از کارآمدترین توابع موجک در کشف ناپیوستگی‌های محلی در سیگنال‌ها است (Daubechies, 1990; Solgi et al., 2017). در این مطالعه به توصیه امیکوگلو<sup>۲</sup> و همکاران (Ekmekcioglu et al., 2020)، از تجزیه تئوری موجک سطح ۲ و دابچیز ۴ استفاده شد. یکی از مشخصات کاربردی تابع موجک، الگوریتم فیلترسازی آن است که با عبور دادن داده‌ها از فیلتر مربوطه، آن‌ها را به دو دسته تقریب و جزئیات تقسیم می‌نماید (Ahmadi and Maddah, 2021). تقریب نماینده اجزا با فرکانس پایین و جزئیات شامل اجزایی با مقیاس کوچک یا فرکانس بالا بوده و فرآیند تجزیه امواج می‌تواند تا چندین مرحله ادامه یابد.

### تفاوت الگوریتم‌های جنگل تصادفی و درخت تصادفی

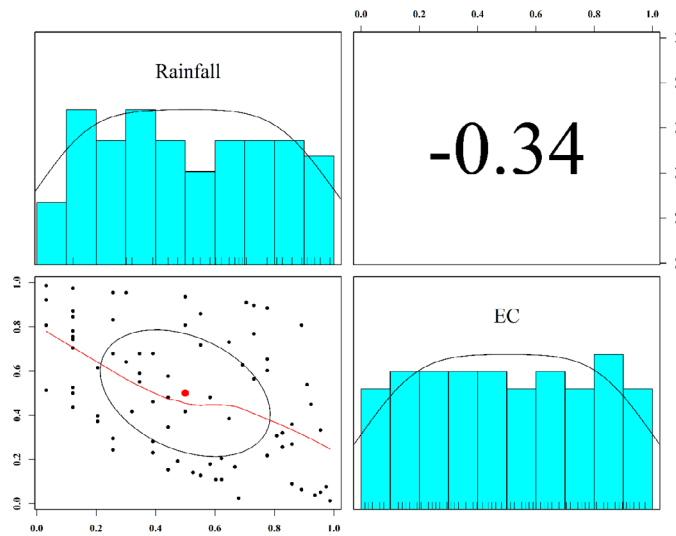
الگوریتم جنگل تصادفی ایده‌ای از تکنیک کلی جنگل‌های تصمیم‌گیری تصادفی است که یک تکنیک یادگیری مجموعه‌ای برای طبقه‌بندی و رگرسیون است که با ساختن تعداد زیادی درخت تصمیم در زمان آموزش ساخته می‌شوند. خروجی کلاس‌بندی شده از حالت طبقه‌بندی یا میانگین پیش‌بینی درختان منفرد تولید می‌شوند. جنگل‌های تصمیم تصادفی برای عادت درختان تصمیم به تطبیق بیش از حد با مجموعه آموزشی خود دقیق هستند.

الگوریتم درخت تصادفی طبقه‌بندی کننده نظارت شده است. این یک الگوریتم یادگیری گروهی است که تعداد زیادی یادگیرنده را ایجاد می‌کند. از یک ایده بسته‌بندی برای ساخت مجموعه‌ای تصادفی از داده‌ها برای ساخت درخت تصمیم استفاده می‌کند. در درخت استاندارد، هر گره با استفاده از بهترین تقسیم در بین همه متغیرها تقسیم می‌شود. در یک جنگل تصادفی، هر گره با استفاده از بهترین در میان زیرمجموعه‌های پیش‌بینی کننده که به‌طور تصادفی در آن گره انتخاب شده‌اند، تقسیم می‌شود. درختان تصادفی گروهی از پیش‌بینی کننده‌های درختی هستند که جنگل نامیده می‌شود. مکانیسم‌های طبقه‌بندی به شرح زیر است: طبقه‌بندی کننده درختان تصادفی بردار ویژگی ورودی را دریافت می‌کند، آن را

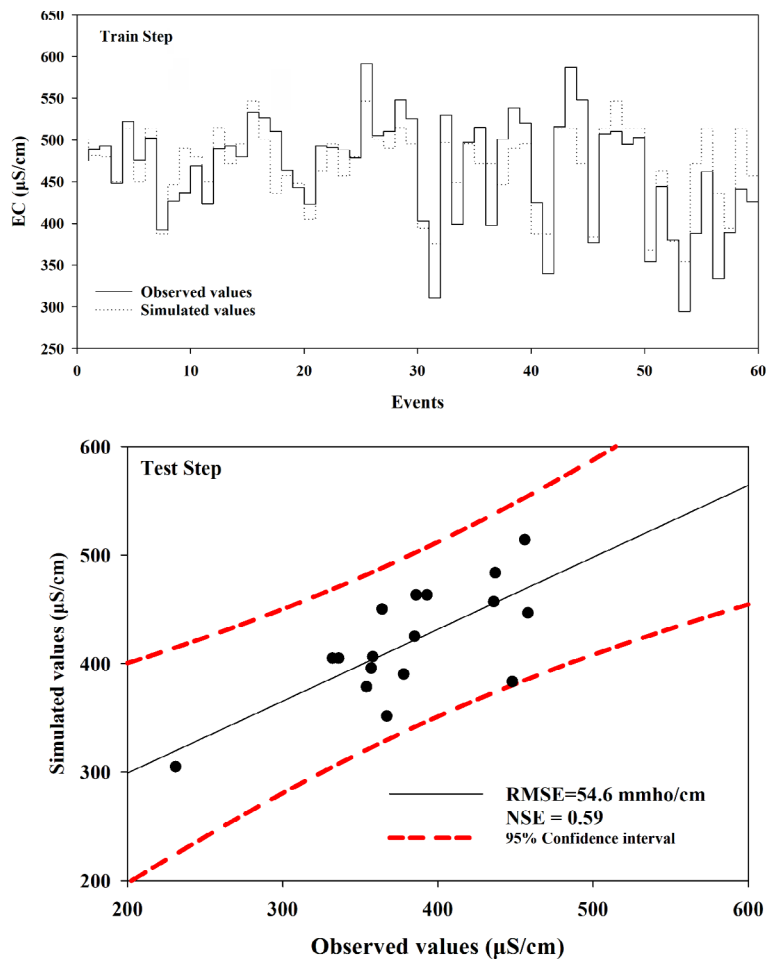
<sup>1</sup> Haar

<sup>2</sup> Ekmekcioglu





شکل ۲- نتایج بررسی همبستگی جفت متغیر بارش-هدایت الکتریکی در ایستگاه مورد مطالعه با استفاده از آماره تائو کندال  
 Fig. 2. The results of investigating the correlation of the pair variable of rainfall-electrical conductivity in the studied station using Kendall's tau statistic



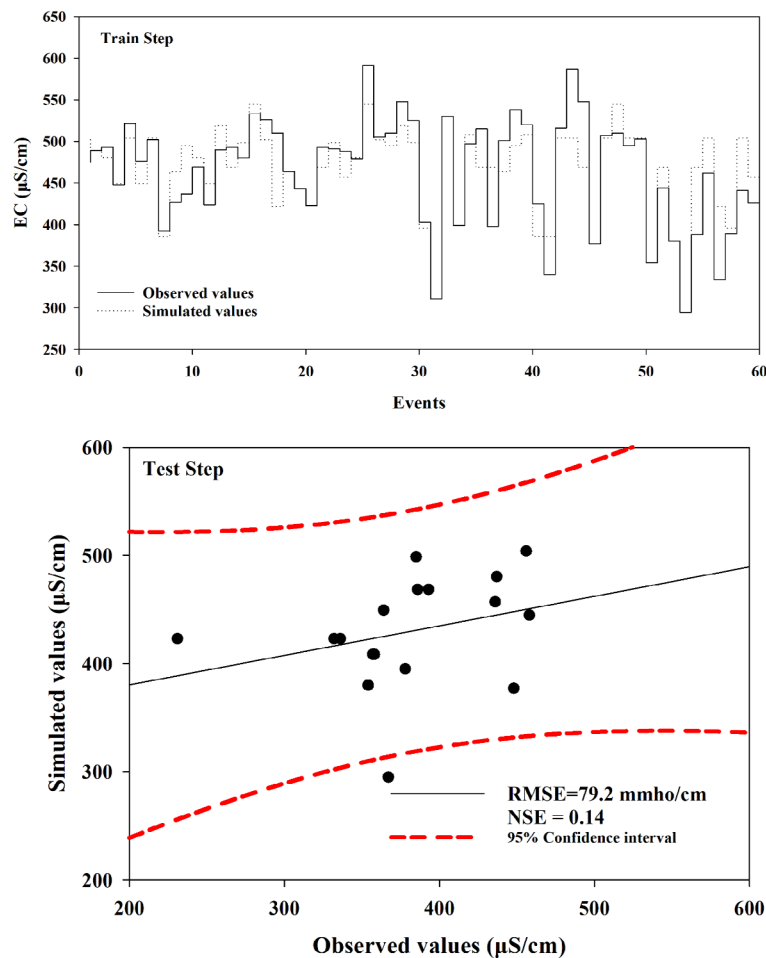
شکل ۳- نتایج شبیه‌سازی مقادیر هدایت الکتریکی با استفاده از الگوریتم جنگل تصادفی و با در نظر گرفتن مقادیر بارش متناظر در ایستگاه اسکندری

Fig. 3. Results of simulation of electrical conductivity values using random forest algorithm considering the corresponding rainfall values in Eskandari station



مقادیر شبیه‌سازی شده در محدوده اطمینان قرار داشته که نشان‌دهنده رضایت‌بخش بودن الگوریتم جنگل تصادفی در شبیه‌سازی مقادیر EC می‌باشد. در شبیه‌سازی مقادیر EC با توجه به مقادیر بارش متناظر با استفاده از الگوریتم درخت تصادفی مطابق شکل ۴، میزان خطای ۳۵/۱ میکرو زیمنس بر سانتی‌متر در فاز آزمایش برآورد شد که نسبت به الگوریتم جنگل تصادفی عملکرد بهتری در فاز آموزش داشته است. در فاز آموزش، الگوریتم درخت تصادفی توانست کارایی مدل را تا حدود ۷۳ درصد با توجه به آماره NSE برآورد کند. اما در فاز آزمایش، الگوریتم جنگل تصادفی نتوانست نسبت به الگوریتم درخت تصادفی نتایج بهتری ارائه نماید. میزان خطای شبیه‌سازی مقادیر EC در فاز آزمایش با توجه به مقادیر بارش متناظر و الگوریتم درخت تصادفی حدود ۵۷/۱ میکرو زیمنس بر سانتی‌متر و کارایی مدل نیز ۰/۵۵ برآورد شد.

همبستگی ۰/۳۴ بین دو متغیر مورد مطالعه برآورد شد. عدد منفی گزارش شده نیز حاکی از رابطه معکوس بارش و مقادیر هدایت الکتریکی در محل ایستگاه هیدرومتری اسکندری می‌باشد. نتایج شبیه‌سازی مقادیر EC با در نظر گرفتن شرایط فوق به صورت شکل‌های ۳ و ۴ ارائه گردید. با توجه به شکل ۳ می‌توان مشاهده کرد که الگوریتم جنگل تصادفی در مرحله آموزش برآزش خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده برقرار کرده است. دامنه تغییرات مقادیر مشاهداتی تا حدودی پوشش داده شده ولی نقاط کمینه و بیشینه به‌خوبی شبیه‌سازی نشده است. در فاز آموزش مقدار خطای ۳۸/۴ میکرو زیمنس بر سانتی‌متر در شبیه‌سازی مقادیر EC با در نظر گرفتن بارش متناظر توسط آماره RMSE برآورد گردید. کارایی مدل نیز با توجه به آماره NSE حدود ۰/۶۷ برآورد شد. در



شکل ۴- نتایج شبیه‌سازی مقادیر هدایت الکتریکی با استفاده از الگوریتم درخت تصادفی و با در نظر گرفتن مقادیر بارش متناظر در ایستگاه اسکندری

Fig. 4. Results of simulation of electrical conductivity values using random tree algorithm considering the corresponding rainfaell values in Eskandari station

شکل‌های ۷ و ۸ ارائه گردید. با توجه به نتایج ارائه شده در شکل ۷ می‌توان مشاهده کرد که در فاز آموزش تطابق نسبتاً خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده مستخرج از مدل Wavelet-RF وجود دارد. در فاز آموزش میزان خطای شبیه‌سازی مقادیر EC با استفاده از مدل Wavelet-RF با توجه به آماره RMSE حدود ۲۱/۶ میکرو زمینس بر سانتی‌متر می‌باشد که کارایی آن نیز با توجه به آماره NSE حدود ۰/۹۹ می‌باشد که کارایی کامل آن را نشان می‌دهد. در فاز آزمایش مطابق شکل ۷ میزان خطای شبیه‌سازی مقادیر EC با توجه به مدل Wavelet-RF حدود ۴۹/۶ میکرو زمینس بر سانتی‌متر تعیین گردید. کارایی مدل Wavelet-RF مطابق شکل ۷ حدود ۹۸ درصد می‌باشد. حدود اطمینان ۹۵ درصد ارائه شده در شکل ۷ نیز تأیید کننده‌ی دقت و کارایی مدل Wavelet-RF در شبیه‌سازی مقادیر EC می‌باشد.

مطابق شکل ۸ نیز نتایج شبیه‌سازی مقادیر EC با توجه به الگوریتم درخت تصادفی و تئوری موجک نشان دهنده‌ی تطابق خوبی بین مقادیر مشاهداتی و شبیه‌سازی شده می‌باشد. در فاز آموزش میزان خطای شبیه‌سازی مقادیر EC با توجه به مدل Wavelet-RT حدود ۲۲/۸ میکرو زمینس بر سانتی‌متر بر اساس آماره RMSE برآورد گردید که در مقایسه با مدل Wavelet-RF حدود ۵/۴ درصد بیش‌تر می‌باشد. کارایی مدل Wavelet-RT نیز حدود ۰/۸۹ بوده که در مقایسه با مدل Wavelet-RF در فاز آموزش حدود ۱۱ درصد کم‌تر می‌باشد. در فاز آزمایش نیز میزان خطای شبیه‌سازی مقادیر EC با توجه به مدل Wavelet-RT حدود ۴۶/۶ میکرو زمینس بر سانتی‌متر بوده که ۵ درصد کم‌تر از مدل Wavelet-RF می‌باشد اما کارایی مدل Wavelet-RF نسبت به مدل Wavelet-RT در فاز آزمایش در شبیه‌سازی مقادیر EC حدود ۲۸ درصد بیش‌تر می‌باشد.

نتایج شبیه‌سازی مقادیر هدایت الکتریکی در منطقه مورد مطالعه نشان داد که در فاز آموزش با توجه به آماره RMSE، مدل‌های Wavelet-RF، Wavelet-RT و let-RT به ترتیب رتبه‌های اول تا چهارم

نتایج مقایسه دو الگوریتم جنگل تصادفی و درخت تصادفی در شبیه‌سازی مقادیر EC با در نظر گرفتن بارش متناظر نشان داد که در فاز آموزش، میزان خطای الگوریتم درخت تصادفی حدود ۸/۶ درصد کم‌تر از الگوریتم جنگل تصادفی می‌باشد و در فاز آزمایش، الگوریتم جنگل تصادفی نتایج بهتری ارائه کرده و نسبت به الگوریتم درخت تصادفی، میزان خطای شبیه‌سازی را حدود ۴/۶ درصد کاهش داده است.

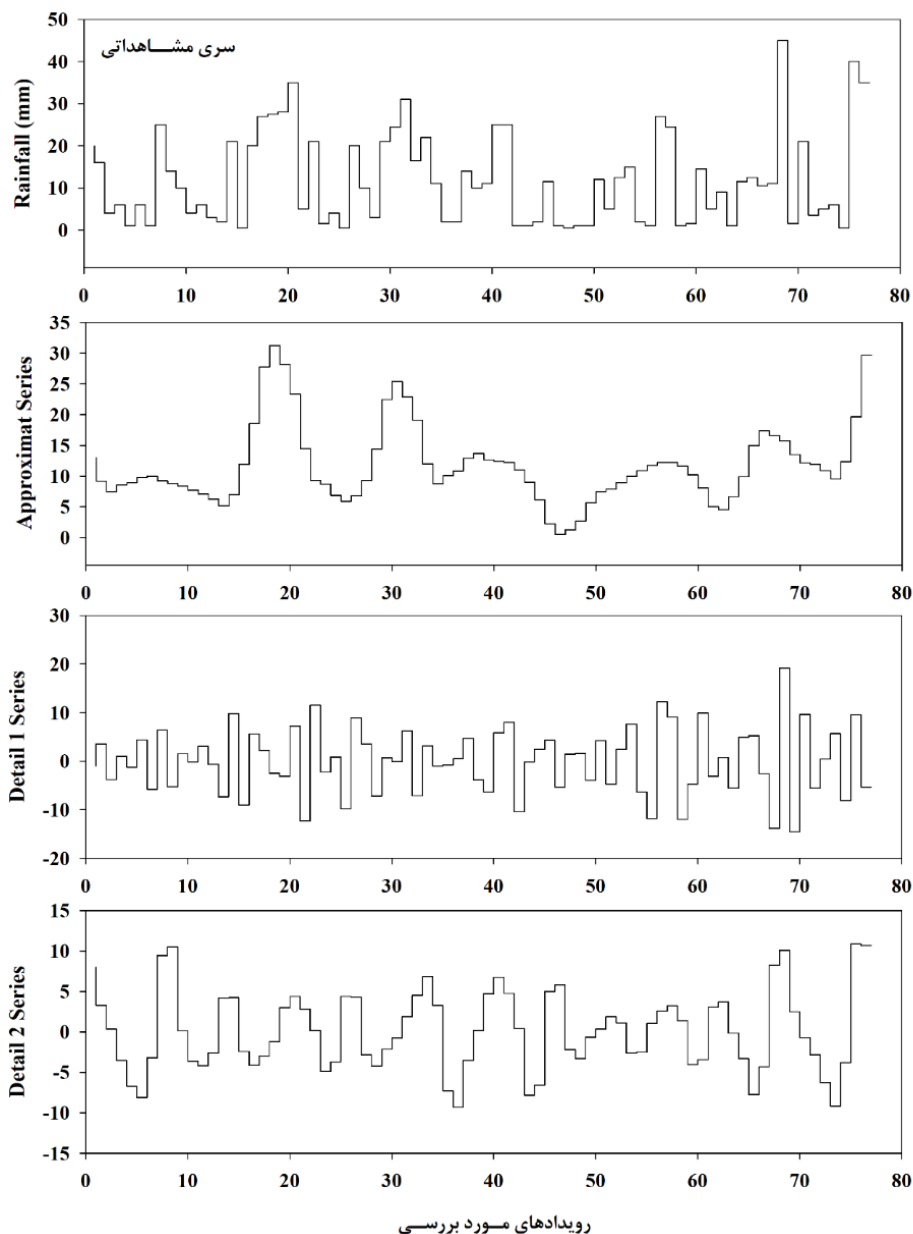
### نتایج شبیه‌سازی مقادیر EC در منطقه مورد مطالعه با در نظر گرفتن بارش تجزیه‌شده متناظر در حالت چهار بعدی

در خصوص شبیه‌سازی مقادیر EC در مطالعه حاضر در ابعاد بیش از دو بعد، از تئوری موجک استفاده شد. در این مطالعه از تئوری موجک جهت تجزیه سطح ۲ مقادیر بارش در منطقه مورد مطالعه استفاده شد. در ابتدا با استفاده از موجک دابچیز سطح ۲، سری زمانی بارش مشاهداتی تجزیه شد. شکل ۵ نتایج تجزیه سری مشاهداتی را با استفاده از تئوری موجک نشان می‌دهد. در شکل ۵ می‌توان سری مشاهداتی، سری تقریبی (Approximate) و سری‌های جزئی (De-tail) را مشاهده کرد. مجموع سری تقریبی و سری جزئی در تجزیه سیگنال مبتنی بر تئوری موجک برابر با سری مشاهداتی خواهد بود. با تجزیه مقادیر مشاهداتی بارش و تحلیل فرکانس موجود، همبستگی بین سیگنال‌های تولیدی و مقادیر مشاهداتی هدایت الکتریکی در منطقه مورد مطالعه با استفاده از آزمون تائو کندال برآورد گردید. نتایج بررسی همبستگی مقادیر مشاهداتی EC و سیگنال‌های تجزیه شده به صورت شکل ۶ ارائه شده است.

نتایج بررسی آماره تائو کندال نشان داد که بین سیگنال‌های تولیدی و مقادیر مشاهداتی EC همبستگی وجود دارد که رابطه معکوس با مقادیر مشاهداتی EC دارد. در نهایت با استفاده از سری تجزیه شده و دو الگوریتم جنگل تصادفی و درخت تصادفی، مقادیر هدایت الکتریکی در دو فاز آموزش و آزمایش شبیه‌سازی شد. نتایج شبیه‌سازی مقادیر هدایت الکتریکی در ایستگاه اسکندری با توجه به مقادیر تجزیه شده بارش با استفاده از دو الگوریتم جنگل تصادفی (Wavelet-RF) و درخت تصادفی (Wavelet-RT) به ترتیب به صورت

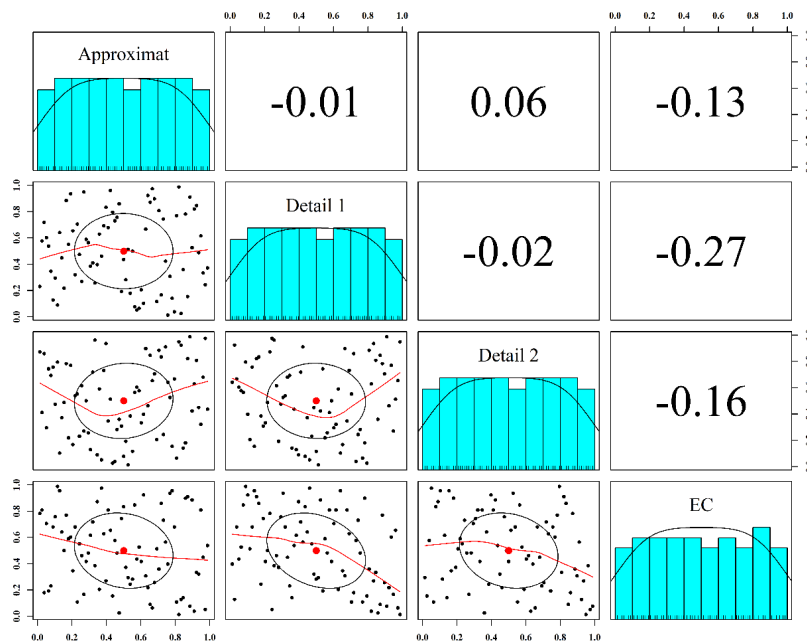
جنگل تصادفی و درخت تصادفی در شبیه‌سازی مقادیر هدایت الکتریکی زیرحوضه اسکندری در دوره آماری مورد بررسی نتایج نشان داد که با تجزیه مقادیر وابسته به سیگنال‌های تقریبی و جزئی می‌توان دقت شبیه‌سازی مقادیر هدایت الکتریکی را بهبود بخشید که با نتایج حاصل از تحقیقات (Ahmadi and Maddah, 2021) مطابقت دارد.

را کسب کردند. در فاز آزمایش که مهم‌ترین فاز در شبیه‌سازی می‌باشد، براساس آماره RMSE، مدل‌های Wavelet-RT، Wavelet-RF، RF و RT به ترتیب رتبه‌های ۱ تا ۴ را کسب کردند. در فاز آموزش کم‌ترین میزان خطا مربوط به مدل Wavelet-RF و در فاز آزمایش، کم‌ترین میزان خطا مربوط به مدل Wavelet-RT می‌باشد. با بررسی و مقایسه دو الگوریتم

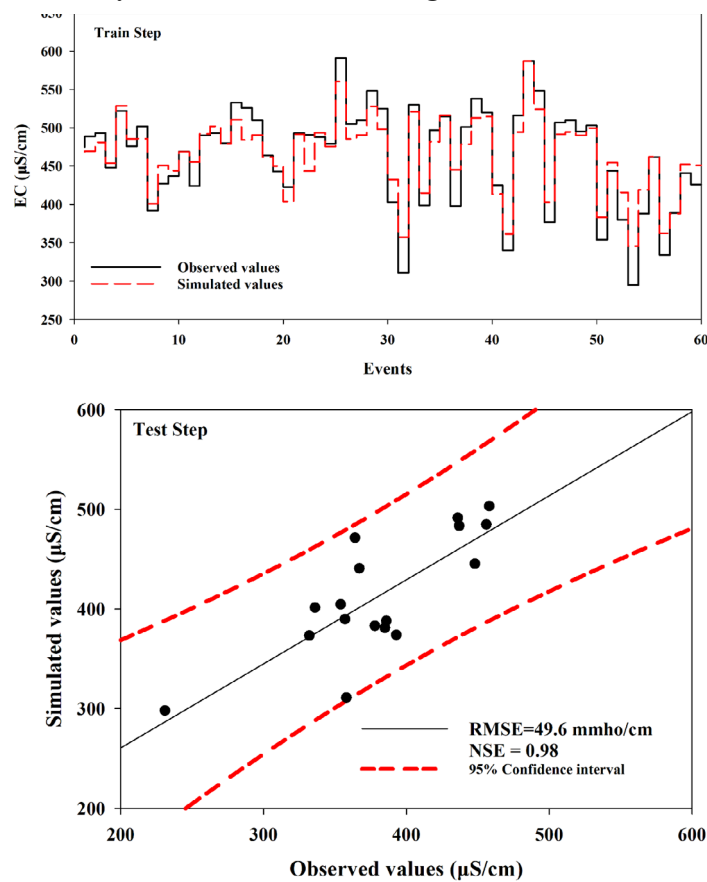


شکل ۵- نتایج تجزیه سری بارش مشاهداتی با استفاده از تئوری موجک

Fig. 5. The results of decomposition of observational rainfall series using wavelet theory

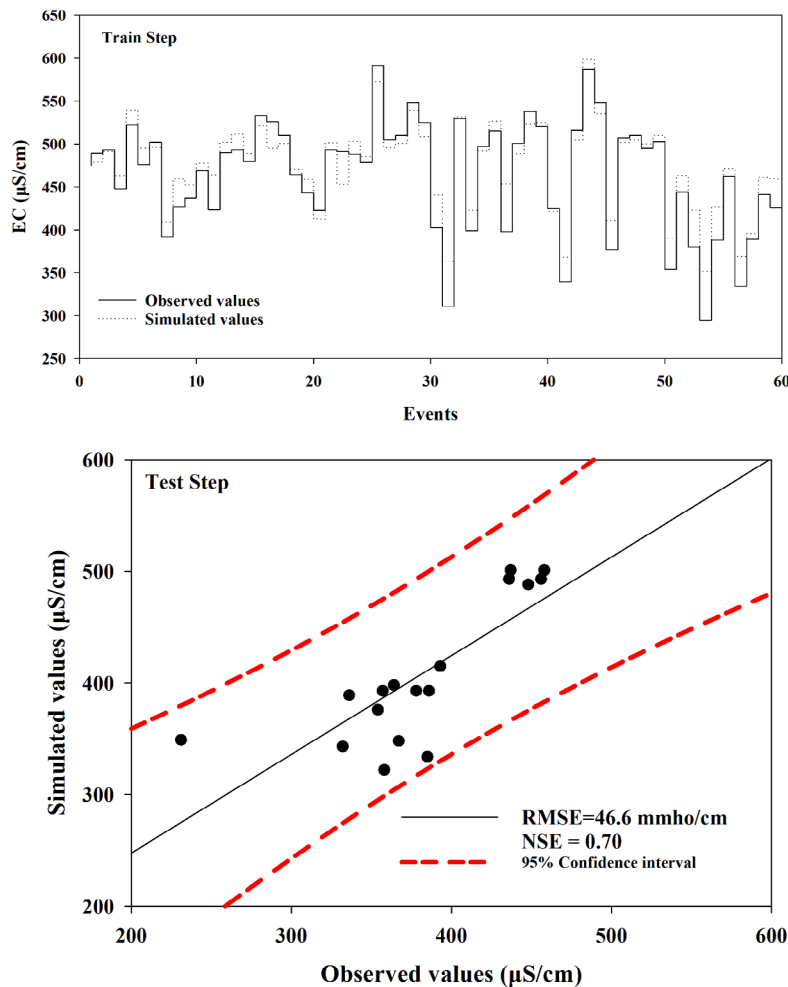


شکل ۶- نتایج بررسی همبستگی جفت متغیر بارش تجزیه شده-هدایت الکتریکی در ایستگاه مورد مطالعه با استفاده از آماره تائو کندال  
 Fig. 6. The results of investigating the correlation of the pair variable of decomposed rainfall-electrical conductivity in the studied station using Kendall's tau statistic



شکل ۷- نتایج شبیه‌سازی مقادیر هدایت الکتریکی با استفاده از الگوریتم جنگل تصادفی و با در نظر گرفتن مقادیر بارش تجزیه شده  
 متناظر در ایستگاه اسکندری (Wavelet-RF)

Fig. 7. Results of simulation of electrical conductivity values using random forest algorithm and considering the corresponding decomposed rainfall values in Eskandari station (Wavelet-RF)



شکل ۸- نتایج شبیه‌سازی مقادیر هدایت الکتریکی با استفاده از الگوریتم درخت تصادفی و با در نظر گرفتن مقادیر بارش تجزیه شده متناظر در ایستگاه اسکندری (Wavelet-RT)

**Fig. 8. Results of simulation of electrical conductivity values using random tree algorithm and considering the corresponding decomposed rainfall values in Eskandari station (Wavelet-RT)**

می‌بخشد. اما در فاز آزمایش، دقت و کارایی الگوریتم جنگل تصادفی بهتر از الگوریتم درخت تصادفی عمل کرده و میزان خطا را تا حدود ۴/۶ درصد بهبود بخشیده است. ضریب نش-ساتکلیف نیز کارایی ۰/۵۹ را در فاز آزمایش در الگوریتم جنگل تصادفی نشان می‌دهد که چندان قابل قبول نمی‌باشد.

با تجزیه سری مشاهداتی بارش به سیگنال‌های تقریبی و جزئی و افزایش بعد شبیه‌سازی تا ۴ بعد، نتایج شبیه‌سازی مقادیر EC در منطقه مورد مطالعه نشان داد که در مورد الگوریتم جنگل تصادفی، تجزیه سری بارش توانسته است میزان خطای شبیه‌سازی را نسبت به حالت اولیه الگوریتم جنگل تصادفی ۷۷ درصد در فاز آموزش و ۱۰ درصد در فاز آزمایش بهبود بخشد. در مورد الگوریتم درخت تصادفی، نتایج

### نتیجه‌گیری

در این مطالعه تأثیر تجزیه سیگنال متغیر وابسته (بارش) در شبیه‌سازی توأم مقادیر هدایت الکتریکی در ایستگاه اسکندری واقع در حوضه آبریز سد زاینده‌رود با الگوریتم‌های جنگل تصادفی و درخت تصادفی مورد بررسی قرار گرفت. مقادیر هدایت الکتریکی و بارش متناظر در این ایستگاه در دوره آماری ۹۹-۱۳۶۹ مد نظر قرار گرفت. نتایج شبیه‌سازی‌ها در دو مقیاس دو و چهار بعدی بررسی و مقایسه شد. نتایج بررسی دقت الگوریتم‌های جنگل تصادفی و درخت تصادفی در شبیه‌سازی دو بعدی مقادیر هدایت الکتریکی و بارش متناظر در حالت دو بعدی نشان داد که در فاز آموزش دقت الگوریتم درخت تصادفی بهتر از مدل جنگل تصادفی بوده و میزان خطا را حدود ۸/۶ درصد بهبود

- 45(1), 5–32.
- Breiman, L., & Cutler, A. (1993). A deterministic algorithm for global optimization. *Mathematical Programming*, 58(1), 179-199.
- Campolo, M., Soldati, A., & Andreussi, P. (2003). Artificial neural network approach to flood forecasting in the River Arno. *Hydrological Sciences Journal*, 48(3), 381-398.
- Cannas, B., Fanni, A., See, L., & Sias, G. (2006). Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Physics and Chemistry of the Earth, Parts A/B/C*, 31(18), 1164-1171.
- Darbandi, S., & Pourhosseini, F. A. (2018). River flow simulation using a multilayer perceptron-firefly algorithm model. *Applied Water Science*, 8(3), 1-9.
- Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*, 36(5), 961-1005.
- Ekmekcioğlu, Ö., Başakın, E. E., & Özger, M. (2022). Tree-based nonlinear ensemble technique to predict energy dissipation in stepped spillways. *European Journal of Environmental and Civil Engineering*, 26(8), 3547-3565.
- Feng, Z. K., & Niu, W. J. (2021). Hybrid artificial neural network and cooperation search algorithm for nonlinear river flow time series forecasting in humid and semi-humid regions. *Knowledge-Based Systems*, 211, 106580.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.
- Hassanjabbar, A., Nezaratian, H., & Wu, P. (2022). Climate change impacts on the flow regime and water quality indicators using an artificial neural network (ANN): a case study in Saskatchewan, Canada. *Journal of Water and Climate Change*, 13(8), 3046-3060.
- Hu, T. S., Lam, K. C., & Ng, S. T. (2005). A Modified Neural Network for Improving River Flow Prediction/ Un Réseau de Neurones Modifié pour Améliorer la Prédiction de L'Écoulement Fluvial. *Hydrological Sciences Journal*, 50(2), 299-318.
- Jain, S. K., Das, A., & Srivastava, D. K. (1999). تجزیه سری مشاهداتی بارش به سیگنال‌های تقریبی و جزئی و دخالت آن‌ها در شبیه‌سازی مقادیر هدایت الکتریکی نشان داد که مدل Wavelet-RT نسبت به الگوریتم درخت تصادفی میزان خطای شبیه‌سازی را حدود ۵۴ درصد در فاز آموزش و حدود ۲۲ درصد در فاز آزمایش بهبود داد. به‌طور کلی نتایج نشان داد که تجزیه سری مشاهداتی به سیگنال‌های تقریبی و جزئی می‌تواند دقت، کارایی و میزان خطای شبیه‌سازی را تا حد زیادی بهبود بخشد. با تجزیه سری مشاهداتی به سیگنال‌های تقریبی و جزئی، ابعاد شبیه‌سازی افزایش یافته و به طبع آن، پیچیدگی مدل‌سازی و شبیه‌سازی افزایش می‌یابد، اما کاهش ۵۴ تا ۷۷ درصدی میزان خطا در فاز آموزش و ۱۰ تا ۲۲ درصدی در فاز آزمایش این پیچیدگی را جبران می‌نماید. با توجه به گسترش مدل‌های چند متغیره، این روش جهت استفاده از مدل‌های چند متغیره توصیه می‌گردد.
- تقدیر و تشکر**
- این اثر تحت حمایت مادی صندوق حمایت از پژوهشگران و فناوران کشور (INSF) برگرفته شده از طرح شماره "۴۰۰۵۰۱۱" انجام شده است.
- منابع**
- Ahmadi, F., & Maddah, M. A. (2021). Development of Wavelet-Kstar Algorithm Hybrid Model for the Monthly Precipitation Prediction (Case Study: Synoptic Station of Ahvaz). *Iranian Journal of Soil and Water Research*, 52(2), 409-420. [In Persian].
- Andy, L. (2012). Documentation for R package random Forest. (<http://cran.r-project.org/web/packages/randomForest>)
- Bageri, F., Khalili, K., & Nazeri Tahroudi, M. (2023). Evaluation of entropy theory based on random forest in quality monitoring of ground water network. *Journal of Water and Irrigation Management*, 13(1), 123-139. [In Persian].
- Bajirao, T. S., Kumar, P., Kumar, M., Elbeltagi, A., & Kuriqi, A. (2021). Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. *Theoretical and Applied Climatology*, 145(3), 1207-1231.
- Breiman, L. (2001). *Random Forests*. *Machine Learning*

- 05022015.
- Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2022b). Application of copula-based approach as a new data-driven model for downscaling the mean daily temperature. *International Journal of Climatology*, 43(1), 240-254.
- Nilsson, P., Uvo, C. B., & Berndtsson, R. (2006). Monthly runoff simulation: Comparing and combining conceptual and neural network models. *Journal of Hydrology*, 321(1-4), 344-363.
- Nourani, V., Baghanam, A. H., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *Journal of Hydrology*, 514, 358-377.
- Pfahring, B. (2010). Random model trees: an effective and scalable regression method.
- Pronoos Sedighi, M., Ramezani, Y., Nazeri Tahroudi, M., & Taghian, M. (2022). Joint frequency analysis of river flow rate and suspended sediment load using conditional density of copula functions. *Acta Geophysica*, 1-13.
- Seifi, A., & Riahi-Madvar, H. (2019). Improving one-dimensional pollution dispersion modeling in rivers using ANFIS and ANN-based GA optimized models. *Environmental Science and Pollution Research*, 26(1), 867-885.
- Solgi, A., Pourhaghi, A., Bahmani, R., & Zarei, H. (2017). Pre-processing data using wavelet transform and PCA based on support vector regression and gene expression programming for river flow simulation. *Journal of Earth System Science*, 126(5), 1-17.
- Wei, S., Yang, H., Song, J., Abbaspour, K., & Xu, Z. (2013). A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. *Hydrological Sciences Journal*, 58(2), 374-389.
- Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, 31(1), 76-77.
- Zamani, R., Mirabbasi, R., Nazeri, M., Meshram, S. G., & Ahmadi, F. (2018). Spatio-temporal analysis of daily, Application of ANN for reservoir inflow prediction and operation. *Journal of Water Resources Planning and Management*, 125(5), 263-271.
- Jayawardena, A. W., Xu, P., & Tsang, F. L. L. (2004, July). Rainfall predication by wavelet decomposition. In *Asia Pacific Association of Hydrology and Water Resources Conference*, Singapore.
- Kalmegh, S. (2015). Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news. *International Journal of Innovative Science, Engineering & Technology*, 2(2), 438-446.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine learning*, 59(1), 161-205.
- Leo, B. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Mallat, S. G. (1989). Multiresolution approximations and wavelet orthonormal bases of  $L^2(R)$ . *Transactions of the American mathematical society*, 315(1), 69-87.
- Meresa, H. (2019). Modelling of river flow in ungauged catchment using remote sensing data: application of the empirical (SCS-CN), artificial neural network (ANN) and hydrological model (HEC-HMS). *Modeling Earth Systems and Environment*, 5(1), 257-273.
- Mirzaee, N., & Sarraf, A. (2022). Application of data fusion models in river flow simulation using signals of large-scale climate, case study: Jiroft Dam Basin. *Watershed Engineering and Management*, 13(4), 672-689.
- Nazeri Tahroudi, M., Mirabbasi, R., Ramezani, Y., & Ahmadi, F. (2022b). Probabilistic Assessment of Monthly River Discharge using Copula and OSVR Approaches. *Water Resources Management*, 36(6), 2027-2043.
- Nazeri Tahroudi, M., Mohammadi, M., & Khalili, K. (2022a). The application of the hybrid copula-GARCH approach in the simulation of extreme discharge values. *Applied Water Science*, 12(12), 1-13.
- Nazeri Tahroudi, M., Ramezani, Y., De Michele, C., & Mirabbasi, R. (2022a). Application of Copula Functions for Bivariate Analysis of Rainfall and River Flow Deficiencies in the Siminehrood River Basin, Iran. *Journal of Hydrologic Engineering*, 27(11),



seasonal and annual precipitation concentration in Jharkhand state, India. Stochastic Environmental Research and Risk Assessment, 32(4), 1085-1097.