

## Development of Wavelet-Kstar Algorithm Hybrid Model for the Monthly Precipitation Prediction (Case Study: Synoptic Station of Ahvaz)

FARSHAD AHMADI<sup>1</sup>, MOHAMMAD AMIN MADDAH<sup>1\*</sup>

1. Department of Hydrology and Water Resources, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

(Received: Nov. 22, 2020- Revised: Dec. 31, 2020- Accepted: Jan. 11, 2021)

### ABSTRACT

Predicting hydrological parameters, especially rainfall, has played a very important role in water resources management and planning. Therefore, the development of methods giving accurate estimates has always been of interest to researchers. In this study, precipitation data from the Ahvaz synoptic station in the period of 2018-1961 were used to develop Kstar and Gene Expression Programming wavelet hybrid models (WKstar and WGEP). The performance of the applied models was evaluated using statistical indices, including the correlation coefficient (CC), Nash-Sutcliffe (NS), Kling-Gupta (KGE) and the Willmott Index (WI). Initially, the Kstar and GEP individual models were implemented, with a delay in precipitation input up to four months ago and month numbers. Results showed that both models achieved maximum accuracy with a time delay of one month (M1 senario), but their performance was very low and unacceptable. Regarding that both models with the M1 pattern have the best performance, so the M1 inputs decomposed by five different wavelet functions to detail and approximat subsets and were represented to the models. The results showed that the performance of wavelet hybrid models was significantly improved, so that the NS index increased from 0.139 to 0.607. In addition, the best performance of WKstar and WGEP hybrid models was obtained with the inputs of the Daubechies wavelet of order four and the decomposition level two. Also, there was no significant difference between the two developed hybrid models statistically, but using the violin plot it was found that the WKstar model is more suitable for predicting precipitation at the Ahvaz synoptic station.

**Keywords:** Decomposition Level, Mother Wavelet, Willmott Index, Violin Plot.

## توسعه روش هیبریدی موجک-الگوریتم Kstar برای پیش‌بینی بارش‌های ماهانه (مطالعه موردی: ایستگاه سینوپتیک اهواز)

فرشاد احمدی<sup>۱</sup> و محمدمبین مداح<sup>۱\*</sup>

۱. گروه هیدرولوژی و منابع آب، دانشکده مهندسی آب و محیط‌زیست، دانشگاه شهید چمران اهواز، اهواز، ایران.  
(تاریخ دریافت: ۱۳۹۹/۹/۲ - تاریخ بازنگری: ۱۳۹۹/۱۰/۱۱ - تاریخ تصویب: ۱۳۹۹/۱۰/۲۲)

### چکیده

پیش‌بینی متغیرهای هیدرولوژیک و به ویژه بارش نقش بسیار مهمی در مدیریت و برنامه‌ریزی منابع آبی داشته و از این‌رو توسعه روش‌هایی که بتواند تخمین دقیقی از آن را به دست دهد همواره مورد توجه محققان بوده است. در این پژوهش از داده‌های بارش ایستگاه سینوپتیک اهواز در دوره آماری ۲۰۱۸-۱۹۶۱ برای توسعه مدل‌های هیبریدی موجک Kstar (WKstar) و برنامه‌ریزی بیان ژن (WGEP) استفاده شد. عملکرد مدل‌های به کار رفته با شاخص‌های آماری ضریب همبستگی (CC)، نش-ساتکلیف (NS)، کلینگ گوپتا (KGE) و ضریب ویلموت (WI) مورد بررسی قرار گرفت. در ابتدا مدل‌های منفرد Kstar و GEP با ورودی‌های بارش تاخیر یافته تا چهار ماه قبل و شماره ماه‌ها اجرا شدند. نتایج نشان داد که هر دو مدل با تاخیر زمانی یک ماه (الگوی M1) به بیشترین دقت رسیده اما عملکرد آنها بسیار ضعیف و غیرقابل قبول بود. با توجه به اینکه هر دو مدل با الگوی M1 بهترین عملکرد را داشته‌اند از این رو بارش‌های یک ماه قبل با استفاده از پنج تابع موجک مختلف به زیرسری‌های تقریب و جزئیات تجزیه شده و مجدداً به مدل‌ها معرفی شدند. نتایج نشان داد که عملکرد مدل‌های هیبریدی موجک نسبت به حالت منفرد بسیار بهبود یافته به طوری که شاخص NS از ۰/۱۳۹ به ۰/۶۰۷ افزایش یافت. همچنین بهترین عملکرد مدل‌های هیبریدی WKstar و WGEP با ورودی‌های تابع موجک دابجیز چهار و سطح تجزیه دو به دست آمده و از نظر آماری اختلاف معنی‌داری بین دو مدل هیبریدی توسعه یافته وجود نداشت، اما با استفاده از نمودار ویولونی مشخص گردید که مدل WKstar برای پیش‌بینی بارش‌های ایستگاه سینوپتیک اهواز مناسب‌تر می‌باشد.

واژه‌های کلیدی: سطح تجزیه، ضریب ویلموت، موجک مادر، نمودار ویولونی.

### مقدمه

بارش یکی از اجزای مهم و بسیار پیچیده چرخه هیدرولوژی است که دارای حالت‌های حدی خشکسالی و ترسالی بوده که هر کدام تاثیرات نامطلوب فراوانی بر محیط زیست و جوامع بشری دارد (Abdourahmane et al., 2019). بنابراین پیش‌بینی بارندگی برای پیش‌گیری از بلايا و همچنین برنامه‌ریزی فعالیت‌های انسانی ضروری به نظر می‌رسد. اگرچه مکانیسم‌های فیزیکی تأثیرگذار بر الگوهای بارندگی (عمدتاً گردش جهانی جو و اقیانوس‌ها) از نظر علمی کاملاً شناخته شده‌اند، اما همچنان پیش‌بینی میزان بارندگی میان مدت و بلند مدت به عنوان یک چالش در بسیاری از مناطق جهان باقی مانده است. علاوه بر این، گرم شدن کره زمین و تغییرات آب و هوایی این مهم را با شرایط پیچیده‌تری مواجه کرده و از این‌رو مدل‌های پیش‌بینی پارامترهای

هیدرولوژیک نیز از روش‌های آماری سنتی به سمت تکنیک‌های هوش مصنوعی نظیر ماشین بردار پشتیبان<sup>۱</sup> (Mirabbasi et al., 2019; Mehr et al., 2019; Ahmadi, 2020) و برنامه‌ریزی بیان ژن<sup>۲</sup> (Mehdizadeh et al., 2019; Freire et al., 2019; Sun et al., 2019) و مدل‌های هیبرید موجکی<sup>۳</sup> (Kstar) (Ekmekcioğlu et al., 2020) و مدل‌های هیبرید موجکی (Mehdizadeh et al., 2018) از مدل‌های هیبریدی ARCH-GEP و ANN-ARCH برای پیش‌بینی بارش ایستگاه‌های سینوپتیک ارومیه، رشت، گرگان، کرمان و زاهدان بهره بردند. بدین منظور

(Ahmadi and Valinia, 2020) تکامل یافته‌اند. برنامه‌ریزی بیان ژن از خانواده الگوریتم‌های فراتکاملی بوده و توانایی بالایی در مدل‌سازی فرایندهای غیرخطی و پویا داشته و در سال‌های اخیر مورد توجه پژوهش‌گران بوده است. (Mehdizadeh et al., 2018) از مدل‌های هیبریدی ARCH-GEP و ANN-ARCH برای پیش‌بینی بارش ایستگاه‌های سینوپتیک ارومیه، رشت، گرگان، کرمان و زاهدان بهره بردند. بدین منظور

\* نویسنده مسئول: Ma.maddah@scu.ac.ir

1 - Support Vector Machine (SVM)  
2 - Gene Expression Programming (GEP)  
3 - Kstar Algorithm

انرژی در سرریز استفاده نموده و گزارش کردند که روش Kstar از توانایی بسیار خوبی در مدل‌سازی پدیده‌های پیچیده برخوردار است.

سری‌های زمانی هیدرولوژیک همواره با مشکلاتی همچون خاصیت تناوبی و وجود نویز سفید مواجه هستند. این عوامل عملکرد مدل‌های هوشمند را تضعیف نموده و برآوردها را به حالت ناریب تبدیل می‌نمایند. برای حل این مشکلات می‌توان از روش‌هایی همچون تلفیق مدل‌های هوشمند با توابع موجک بهره برد. استفاده از توابع موجک در پیش‌بینی بارش توسط محققان صورت پذیرفته که به تعدادی از جدیدترین این مطالعات اشاره می‌شود. Santos *et al.* (2019) مدل هیبریدی موجک- شبکه عصبی مصنوعی را برای پیش‌بینی بارش‌های روزانه حوضه آبریز رودخانه سان فرانسیسکو توسعه دادند. این پژوهش‌گران گزارش نمودند که استفاده از مدل هیبریدی توسعه داده شده دقت را ۳۵٪ افزایش و خطا را تا ۴۷ درصد کاهش داده است. Estévez *et al.* (2020) برای مدل‌سازی داده‌های بارش ماهانه مناطق نیمه خشک اندونزی از هیبرید موجک- شبکه عصبی مصنوعی بهره بردند. نتایج نشان داد که با تجزیه داده‌های ورودی به دو عامل تقریب و جزئیات می‌توان اطلاعات ورودی کامل‌تری را در اختیار مدل قرار داد و از این‌رو عملکرد مدل‌ها بطور قابل توجهی افزایش می‌یابد.

با توجه به پیشینه پژوهش مشاهده می‌شود که تاکنون مدل‌های هوش مصنوعی متعددی برای پیش‌بینی پارامترهای هیدرولوژیک توسعه یافته و توسط محققان مورد استفاده قرار گرفته‌اند. اما با این وجود در مطالعات اندکی از الگوریتم Kstar برای پیش‌بینی بارش بهره برده شده است. همچنین رویکرد استفاده از مدل‌های تلفیقی نیز به عنوان یک روش نظام‌مند برای بهبود عملکرد مدل‌های پایه توجهات بسیاری را به خود جلب کرده است. از این رو در مطالعه حاضر برای پیش‌بینی بارش ماهانه، تلفیق مدل‌های GEP و الگوریتم Kstar با توابع مختلف موجک مد نظر قرار گرفت.

## مواد و روش‌ها

### داده‌ها و منطقه مورد مطالعه

استان خوزستان در منتهی‌الیه جنوب غرب کشور واقع شده و از دو بخش کاملاً متفاوت جلگه‌ای و کوهستانی تشکیل یافته است. در مناطق کوهستانی و مرتفع، تابستان‌های معتدل و زمستان‌های سرد و در نواحی کوه‌پایه‌ای دارای آب و هوای نیمه بیابانی و در

ایشان برای توسعه مدل‌های مستقل ANN و GEP از داده‌های تاخیر یافته بارش استفاده نمودند. عملکرد مدل‌های مستقل و ترکیبی پیشنهادی با استفاده از آماره‌های RMSE و  $R^2$  ارزیابی شد. یافته‌های این پژوهش نشان داد که مدل‌های پیشنهادی ANN-ARCH و GEP-ARCH در تمام ایستگاه‌های مورد مطالعه با اقلیم‌های مختلف از عملکرد بسیار بهتری در مقایسه با مدل‌های مستقل برخوردار هستند. Mirabbasi *et al.* (2019) از پنج روش هوش مصنوعی شامل برنامه‌ریزی بیان ژن، ماشین بردار پشتیبان، شبکه عصبی مصنوعی، اسپلاین رگرسیونی انطباقی چند متغیره<sup>۱</sup> (MARS) و مدل درختی M5 برای پیش‌بینی بارش‌های درازمدت ایالت مادهاپا پرادش هندوستان استفاده نمودند. در پژوهش مذکور بارش تابعی از طول و عرض جغرافیایی، ارتفاع ایستگاه باران‌سنجی و شماره ماه (شماره ماه خاصیت تناوبی در داده‌های بارش منطقه را نمایندگی می‌کرد) در نظر گرفته شد. روش‌های مورد مطالعه براساس آماره‌های ضریب تعیین ( $R^2$ )، میانگین خطاهای مطلق<sup>۲</sup> (MAE) و میانگین مربعات خطاها<sup>۳</sup> (RMSE) با یکدیگر مقایسه شدند. نتایج نشان داد که عملکرد مدل‌های مورد مطالعه نزدیک به یکدیگر بوده و از توانایی لازم برای پیش‌بینی پارامتر بارش برخوردارند.

در سال‌های اخیر حجم اطلاعات ثبت شده از پدیده‌های هیدرولوژیک رو به فزونی گذاشته و برای استخراج بهتر روابط موجود در پایگاه‌های داده با حجم زیاد، می‌توان از روش‌های داده-کاوی استفاده نمود. داده‌کاوی به عنوان فرآیندی برای یافتن الگوها و ارتباطات برای ساختن مدل‌های پیش‌بینی تعریف شده است. تکنیک دسته‌بندی متداول‌ترین روش توسعه مدل‌های پیش‌بینی بوده و شامل یادگیری و رده‌بندی است. در یادگیری اطلاعات آموزشی با الگوریتم دسته‌بندی تحلیل شده و اطلاعات برای برآورد دقیق قواعد به کار می‌رود. تعدادی از روش‌های داده-کاوی پرکاربرد در زمینه‌های مختلف عبارتند از درخت تصمیم-گیری، جنگل‌های تصادفی، مدل M5 و الگوریتم Kstar. در پژوهشی Bushara (2019) از روش‌های مختلف هوش مصنوعی و الگوریتم Kstar برای پیش‌بینی ۱۳ ساله بارش‌های ۲۴ ایستگاه سینوپتیک کشور سودان استفاده نمود. ایشان عملکرد الگوریتم Kstar در مدل‌سازی بارش‌ها را بسیار مناسب گزارش نمودند.

Costache *et al.* (2020) نیز دقت بالای الگوریتم Kstar را در بررسی سیلاب‌های ناگهانی مورد تایید قرار دادند. Ekmekcioğlu *et al.* (2020) از روش‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و الگوریتم Kstar برای مدل‌سازی و پیش‌بینی پخش

عنوان تابع برازش مناسب انتخاب گردید. سپس در گام دوم، انتخاب مجموعه متغیرهای ورودی و مجموعه توابع به منظور تولید کروموزومها انجام شد. در مسئله حاضر مجموعه ورودیها متشکل از بارشهای ماهانه با تاخیرهای زمانی می‌باشد. در این مطالعه، از چهار عملگر اصلی که شامل  $\{+, -, \times, \div\}$  و نیز توابع ریاضی  $\{X^2, X^3, \sqrt[3]{X}, \sqrt{X}, \text{Log}(X)\}$  استفاده شد. گام سوم، شامل انتخاب ساختار و معماری کروموزومها می‌باشد. گام چهارم انتخاب تابع پیوندی است که در این مطالعه عمل جمع برای ایجاد پیوند بین زیرشاخه‌ها مورد استفاده قرار گرفت. در نهایت، در گام پنجم عمل‌گرهای ژنتیک و نرخ هریک از آنها انتخاب می‌شود.

### الگوریتم Kstar

الگوریتم Kstar را می‌توان به عنوان روشی برای تجزیه و تحلیل خوشه‌ای تعریف کرد که هدف اصلی آن تقسیم  $n$  مشاهده به  $K$  خوشه است به طوری که در آن، هر مشاهده به خوشه با نزدیکترین میانگین تعلق دارد (Ekmekcioğlu et al., 2020). الگوریتم Kstar را می‌توان به عنوان یک روش یادگیری مبتنی بر نمونه توصیف کرد که از تئوری آنتروپی به عنوان اندازه‌گیری فاصله استفاده می‌کند. این روش با ارائه یک رویکرد سازگار برای مدیریت خصوصیات ارزشمند واقعی، ویژگی‌های نمادین و مقادیر از دست رفته امکان دریافت اطلاعات را از داده‌های موجود به حداکثر می‌رساند. در الگوریتم Kstar فاصله از یک نمونه به نمونه دیگر می‌تواند به عنوان پیچیدگی تبدیل یک نمونه به نمونه دیگر توصیف شود (Granata et al., 2019). همانگونه که ذکر شد رگرسیون Kstar، از یک تابع فاصله آنتروپی استفاده می‌کند. فاصله آنتروپیک برای بدست آوردن نمونه‌هایی که بیشترین شباهت را از مجموعه داده‌ها به یکدیگر دارند استفاده می‌شود. در نظر بگیرید که  $a$  و  $b$  نمونه‌های مورد بررسی باشد در این صورت  $P^*$  را می‌توان به عنوان احتمال هر مسیر توصیف کرد که از  $a$  به  $b$  است. بنابراین رابطه  $P$  به شرح زیر قابل بیان می‌باشد (Ekmekcioğlu et al., 2020):

$$P^*(b|a) = \sum_{t \in p: t(a)=b}^N p(t) \quad (\text{رابطه ۱})$$

که در آن  $t$  نشان دهنده مقدار  $T$  یک مجموعه‌ای از تبدیلات داده‌هاست) بوده و  $p$  یک تابع احتمال است. با در نظر گرفتن اینکه  $P^*$  شرایط زیر را دارد:

نواحی پست جلگه‌ای هر چه به سمت جنوب و جنوب غربی پیش می‌رویم خصوصیات آب و هوا از نیمه بیابانی به بیابانی کناره‌ای تبدیل می‌شود. عمدتاً زمستان‌های این ناحیه، کوتاه و معتدل و تابستان‌ها طولانی و گرم است. میانگین بارش و دمای سالانه در استان خوزستان به ترتیب ۲۸۴ میلی‌متر و ۲۵/۳ درجه سلسیوس می‌باشد که توسط ۲۰ ایستگاه سینوپتیک تحت نظارت سازمان هواشناسی حاصل گردیده است. به منظور پیش‌بینی بارش‌های استان از داده‌های ایستگاه سینوپتیک اهواز به دلیل دارا بودن اطلاعات مناسب و طولانی مدت در دوره آماری ۲۰۱۸-۱۹۶۱ استفاده شد. شکل (۱)، موقعیت مکانی ایستگاه سینوپتیک اهواز را در کشور و استان خوزستان نشان می‌دهد.

### برنامه‌ریزی بیان ژن

روش برنامه‌ریزی بیان ژن اولین بار توسط Koza (1992) توسعه داده شد. این روش از خانواده الگوریتم‌های تکاملی محسوب می‌شود که براساس نظریه تکامل داروین توسعه یافته‌اند. الگوریتم‌های مذکور یک تابع هدف مشخص را در چهارچوب شاخص‌های کیفی تعریف نموده و در گام بعد تابع یاد شده را برای ارزیابی پاسخ‌های به دست آمده در یک فرآیند مرحله‌ای به همراه تصحیح ساختار داده‌ها به کار گرفته و در نهایت، جواب مناسب را ارائه می‌نمایند. این الگوریتم‌ها از لحاظ ساختاری می‌توانند به سه گروه اصلی تقسیم شود (Ferreira, 2002):

۱) الگوریتم‌های ژنتیک با افرادی شامل کروموزوم‌های خطی با طول ثابت و بدون بیانی پیچیده. در این سیستم‌ها کروموزوم‌ها طبق برتری خصوصیات فردیشان حفظ می‌شوند و به الگوریتم ژنتیک<sup>۱</sup> (GA) معروف هستند.

۲) الگوریتم‌های ژنتیک با افرادی شامل ساختارهای شاخه‌ای<sup>۲</sup> از اندازه‌ها و اشکال متفاوت و توانا در پذیرش تعداد بیشتری از عوامل. در این سیستم‌ها نیز ساختارهای شاخه‌ای طبق برتری خصوصیات فردیشان حفظ می‌شوند که برنامه‌ریزی ژنتیک (GP) نامیده می‌شوند.

۳) الگوریتم‌های ژنتیک با افراد کدگذاری شده به شکل کروموزوم‌های خطی با طولی ثابت و قابل بیان به شکل ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت. در این سیستم‌ها کروموزوم‌ها بواسطه برتری عوامل سببی روی فنوتیپ<sup>۳</sup> (ساختارهای شاخه‌ای) حفظ می‌شوند و به برنامه‌ریزی بیان ژن (GEP) معروف هستند.

در مطالعه حاضر برای مدل‌سازی بارش ماهانه ایستگاه سینوپتیک اهواز در گام اول، تابع جذر میانگین مربعات خطا به

صورت زیر بازنویسی می‌شوند:

(رابطه ۷)

$$P^*(b a) = P^*(i) = \frac{s}{\sqrt{2s-s^2}} \left( \frac{1-\sqrt{2s-s^2}}{1-s} \right) \text{ Where } i = |a - b|$$

(رابطه ۸)

$$K^*(b a) = K^*(i) = \frac{1}{2} \log(2s - s^2) - \log(s) + i [\log(1 - s) - \log(1 - \sqrt{2s - s^2})]$$

که در آن s پارامتر مدل بوده و بین صفر و یک متغیر است.

با در دست داشتن این روابط می‌توان با استفاده از مقادیر احتمالاتی محاسبه شده مناسب‌ترین نمونه را برای داده مورد نظر انتخاب نمود.

$$\sum_b^N P^*(b a) = 1 \quad (\text{رابطه ۲})$$

$$0 \leq P^*(b a) \leq 1 \quad (\text{رابطه ۳})$$

با توجه به روابط فوق تابع Kstar به شرح زیر بیان می‌شود:

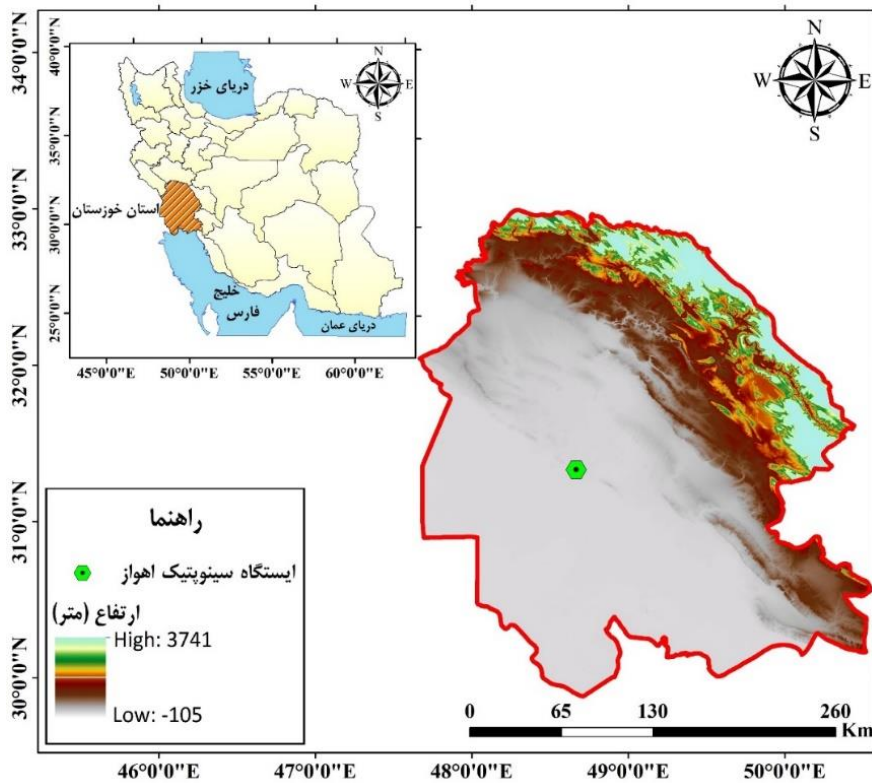
$$K^*(b a) = -\log P^*(b a) \quad (\text{رابطه ۴})$$

رابطه بالا دارای ویژگی‌های زیر است:

$$K^*(b a) \geq 0 \quad (\text{رابطه ۵})$$

$$K^*(b a) + K^*(c b) \geq K^*(c a) \quad (\text{رابطه ۶})$$

روابط ۱ تا ۶ نماینده اعداد صحیح بوده و برای اعداد پیوسته به



شکل ۱- موقعیت ایستگاه سینوپتیک اهواز

یابی وقایع با فرکانس بالا با وضوح زمانی بهتر را فراهم می‌کند (Polikar and Mastorakis, 1999). 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. 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 (Wang and Ding, 2003). A0, A1, A2 و ... ضرایب بالاترین سطوح تقریب و Dj ضریب جزئیات تابع F(x) است. در صورت ادامه مراحل تجزیه، این روند می‌تواند با تجزیه مداوم تقریب‌ها تکرار گردد (Polikar, 1996). بدین ترتیب یک موج به زیر مجموعه‌هائی مانند شکل (۲) تقسیم

### تبدیل موجک

تئوری تبدیل موجک بر اساس تجزیه و تحلیل سیگنال با استفاده از مقیاس‌های مختلف در دامنه زمان و فرکانس پایه‌ریزی شده است. روابط مربوط به این تئوری با تعمیم مفاهیم شناخته شده در دهه ۱۹۸۰ انجام و اصطلاح موجک برای اولین بار توسط ژئوفیزیکدان فرانسوی ژان مورلت معرفی شد (Phillies et al., 1996). داده‌های لرزه‌ای تحلیل شده توسط مورلت، توانست تغییرات سریع مولفه‌های فرکانس را با گذشت زمان بررسی نموده و نشان داد که تبدیل فوریه برای اینگونه از مسائل نتایج مناسبی را ارائه نمی‌دهند. بنابراین با کمک فیزیکدان نظری کروات/کس گروسمان، مورلت تحول جدیدی را ایجاد کرد که امکان مکان

$$L = \text{Int}[\text{Log}(N)] \quad (\text{رابطه ۱۲})$$

در رابطه فوق،  $L$  تعداد سطح تجزیه،  $N$  طول سری داده‌های اولیه و  $\text{INT}$  عملگر صحیح می‌باشد. با تجزیه هر یک از سری داده‌ها، واریانس سری‌های زمانی اولیه به زیرسری‌های تقریب و جزئیات منتقل شده و هیچ‌گونه اطلاعاتی از سری زمانی اولیه از بین نخواهد رفت (Wang and Ding, 2003). بطوریکه با معکوس نمودن فرایند ارائه شده در شکل (۲) می‌توان با دقت بالایی سری زمانی اولیه را بازسازی نمود.

### تعیین الگوی ورودی برای مدل‌ها

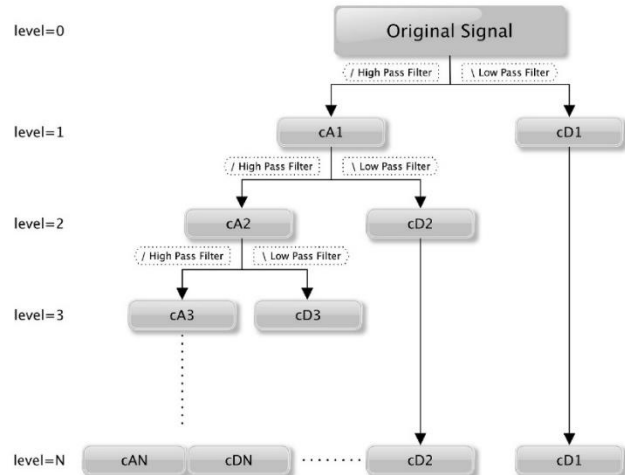
برای مدل‌سازی پدیده‌های هیدرولوژیک انتخاب مناسب‌ترین داده‌های ورودی می‌تواند تاثیر بسزایی در عملکرد روش‌های مورد استفاده داشته باشد. به طوریکه برای آموزش ماهیت ساز و کار حاکم بر پدیده، افزایش تعداد ورودی‌ها نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت نیز می‌گردد (Ahmadi, 2020). بنابراین در مطالعه حاضر سعی گردید برای پیش‌بینی بارش ماه جاری، از داده‌های بارش ماه‌های قبل به عنوان ورودی مدل‌های GEP و Kstar استفاده شود. علاوه بر داده‌های تاخیر یافته بارش، بنابه پیشنهاد Mirabbasi et al. (2019) از شماره ماه‌ها نیز در تمامی الگوهای ورودی به مدل‌ها استفاده شد. این امر کمک می‌نماید تا خاصیت فصلی و تناوبی داده‌ها به مدل معرفی گردد، بدون اینکه هزینه خاصی به فرآیند مدل‌سازی وارد شود. در نهایت ورودی‌های مدل‌های مستقل به شرح جدول (۱) آماده گردید. همانطور که در قسمت قبل نیز اشاره شد، در این مطالعه از تئوری موجک برای پیش‌بینی بارش‌ها بهره گرفته شد. بدین منظور با استفاده از توابع موجک Haar)، دابچیز ۲ (db2)، دابچیز ۴ (db4)، سیملت (Sym)، کوپفلت (Coif) و فجر-کورکین<sup>۱</sup> (F-K) الگوی ورودی که مدل‌های مستقل را به بهینه‌ترین حالت می‌رسانند، در دو سطح تجزیه شده و بدین ترتیب مدل‌های هیبریدی WGEP و WKstar ایجاد گردیدند (جدول ۱).

### ارزیابی مدل‌ها

در این تحقیق، برای ارزیابی مدل‌های موردنظر از معیارهای ضریب همبستگی (CC)، نش-ساتکلیف (NS)، کلینگ گوپتا (KGW) و ضریب ویلموت (WI) استفاده می‌شود:

$$CC = \frac{\text{cov}(O_i, P_i)}{\sqrt{\text{Var}(O_i) \times \text{Var}(P_i)}} \quad (\text{رابطه ۱۳})$$

می‌شود که آن را درخت تجزیه موجک می‌نامند (Ekmekcioğlu et al., 2020).



شکل ۲- درخت تجزیه موجک (Ekmekcioğlu et al., 2020)

نوسانی بودن و کوتاه مدت بودن از جمله ویژگی‌های مهمی است که در تعریف تابع موجک مورد توجه قرار می‌گیرد. در این صورت می‌توان گفت که  $\psi(x)$  تابع موجک است اگر و فقط اگر تبدیل فوریه آن در رابطه زیر صدق نماید:

$$\psi(0) = \int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (\text{رابطه ۹})$$

رابطه (۹)، همان شرط پذیرفتگی برای موجک  $\psi(x)$  بوده و این ویژگی (تابع با میانگین صفر)، چندان محدود کننده نبوده و می‌توان توابع بسیاری را بر اساس آن تابع موجک نامید.  $\psi(x)$  تابع موجک مادر است که توابع مورد استفاده در تحلیل، با دو عمل ریاضی انتقال و مقیاس در طول سیگنال مورد بررسی، تغییر اندازه و تغییر محل داده می‌شوند (رابطه ۱۰).

$$\varphi_{a,b}(0) = \frac{1}{\sqrt{a}} \varphi\left(\frac{x-b}{a}\right) \quad (\text{رابطه ۱۰})$$

در نهایت ضرایب موجک در هر نقطه از سیگنال ( $b$ ) و برای هر مقدار از مقیاس ( $a$ ) با استفاده از رابطه ۱۱ قابل محاسبه می‌باشد (Mallat, 1998).

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \varphi\left(\frac{t-b}{a}\right) f(t) dt \quad (\text{رابطه ۱۱})$$

به ازای مقادیر مختلف  $a$  و  $b$  مقدار  $T$  قابل محاسبه بوده و زمانی که  $T$  بیشترین مقدار مثبت را به خود اختصاص دهد در این صورت می‌توان انتظار داشت که بیشترین انطباق نیز حاصل گردد. یکی از مراحل مهم در کاربرد توابع موجک انتخاب سطح تجزیه مناسب برای تحلیل سیگنال مورد نظر می‌باشد. بدین منظور در مطالعه حاضر از رابطه پیشنهادی Wang and Ding (2003) به شرح زیر استفاده شد:

پیش‌بینی شده،  $\bar{O}$  میانگین بارش مشاهداتی،  $CC$  ضریب همبستگی بین داده‌های مشاهداتی و محاسباتی،  $\alpha$  نسبت انحراف معیار  $O_i$  و  $P_i$ ،  $\beta$  نسبت میانگین  $O_i$  و  $P_i$  و  $n$  تعداد داده‌ها می‌باشد. مدلی به عنوان مناسب‌ترین گزینه انتخاب می‌گردد که بیشترین مقدار  $CC$ ،  $NS$ ،  $KGW$  و  $WI$  را به خود اختصاص دهد (Willmott et al., 2012).

$$NS = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (\text{رابطه ۱۴})$$

$$KGW = 1 - \sqrt{(CC - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (\text{رابطه ۱۵})$$

$$WI = \left| 1 - \left[ \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right] \right|, 0 \leq WI \leq 1 \quad (\text{رابطه ۱۶})$$

که در روابط فوق  $O_i$  مقادیر مشاهداتی،  $P_i$  مقادیر

جدول ۱- الگوهای مختلف ورودی به مدل‌های مستقل و هیبریدی موجکی

ردیف	الگو	آرایش الگوی ورودی
۱	M1	$R_t = f(\text{month number}, R_{t-1})$
۲	M2	$R_t = f(\text{month number}, R_{t-1}, R_{t-2})$
۳	M3	$R_t = f(\text{month number}, R_{t-1}, R_{t-2}, R_{t-3})$
۴	M4	$R_t = f(\text{month number}, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4})$
۵	M5	$R_t = f(\text{decomposition inputs via Haar wavelet, Level one})$
۶	M6	$R_t = f(\text{decomposition inputs via haar wavelet, Level two})$
۷	M7	$R_t = f(\text{decomposition inputs via db2 wavelet, Level one})$
۸	M8	$R_t = f(\text{decomposition inputs via db2 wavelet, Level two})$
۹	M9	$R_t = f(\text{decomposition inputs via db4 wavelet, Level one})$
۱۰	M10	$R_t = f(\text{decomposition inputs via db4 wavelet, Level two})$
۱۱	M11	$R_t = f(\text{decomposition inputs via Symlet wavelet, Level one})$
۱۲	M12	$R_t = f(\text{decomposition inputs via Symlet wavelet, Level two})$
۱۳	M13	$R_t = f(\text{decomposition inputs via Coiflet wavelet, Level one})$
۱۴	M14	$R_t = f(\text{decomposition inputs via Coiflet wavelet, Level two})$
۱۵	M15	$R_t = f(\text{decomposition inputs via Fejer - Korovkin wavelet, Level one})$
۱۶	M16	$R_t = f(\text{decomposition inputs via Fejer - Korovkin wavelet, Level two})$

منفرد GEP و Kstar برای مرحله تست ارائه شده است. در این نمودار شاخص‌های ارزیابی براساس مقدار متناظر آنها رنگ‌بندی شده و از بزرگ به کوچک طبقه‌بندی می‌گردند. این امر کمک شایانی در درک جایگاه شاخص مدنظر در بین کلیه شاخص‌های مورد بررسی داشته و می‌تواند اطلاعات بسیار مفیدی را در مدت زمان اندکی در اختیار تحلیل‌گران قرار دهد. با توجه به شکل (۳) مشاهده می‌شود که عملکرد مدل‌ها در حالت کلی چندان مناسب نبوده و بیشترین دقت با الگوی M1 حاصل شده است. علاوه بر این با افزایش تعداد ورودی‌ها دقت مدل‌ها روند کاهشی را نشان می‌دهد. شکل (۴) دلیل این امر را به خوبی نشان می‌دهد به طوریکه ضریب همبستگی جزئی مرتبه اول در سری داده‌ها بیشترین مقدار را به خود اختصاص داده و در نتیجه می‌تواند بیشترین اطلاعات مورد نیاز برای مدل‌ها را فراهم نماید. اما با افزایش تعداد ورودی‌ها که در واقع همان تاخیرهای ماهانه می‌باشد با حجم عظیمی از جمعیت‌های اولیه، اطلاعات کمی در اختیار مدل‌ها قرار گرفته و در نتیجه با کاهش عملکرد مواجه می‌شوند. Ahmadi (2020)، Mehdizade et al. (2020) و Mohammadi et al. (2020) در مطالعات خود نتایج مشابهی را گزارش نموده‌اند.

## نتایج و بحث

### مدل‌سازی بارش ماهانه با استفاده از مدل‌های منفرد GEP و Kstar

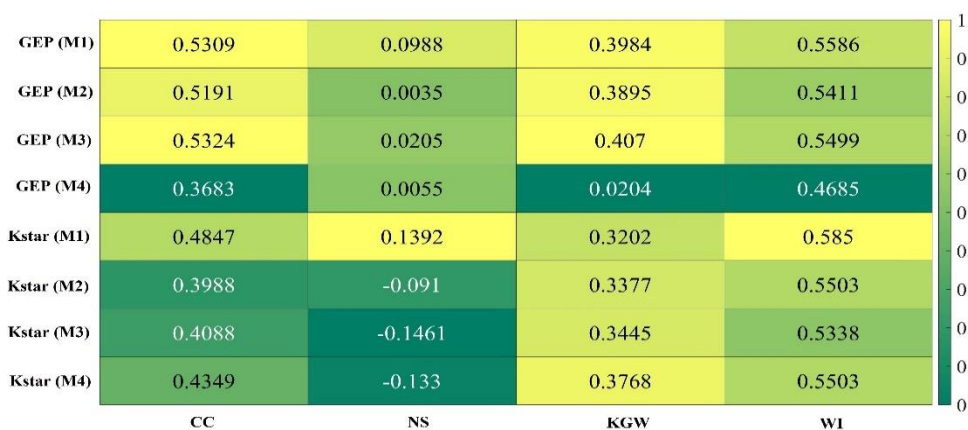
در این مطالعه به منظور پیش‌بینی بارش‌های ماهانه ایستگاه سینوپتیک اهواز، ۸۰ درصد داده‌ها (۵۵۱ ماه) برای آموزش و ۲۰ درصد مابقی برای تست (۱۴۴ ماه) در نظر گرفته شد. برای توسعه الگوریتم Kstar از نرم افزار متن باز WEKA استفاده شده و پارامترهای آن از روش آزمون و خطا تعیین گردید. پارامتر ترکیبی جهانی که بین ۱۰ تا ۱۰۰ متغیر است در این پژوهش برابر با ۱۰ به دست آمد. شایان ذکر است که با افزایش پارامتر جهانی ترکیبی از ۱۰ به ۱۰۰ عملکرد، مدل در برآورد بارش‌های ماهانه ایستگاه سینوپتیک اهواز کاهش می‌یابد. (Ekmekcioğlu et al. (2020) نیز نتایج مشابهی را در خصوص افزایش پارامتر جهانی گزارش نمودند. همچنین ایشان عدد مناسب برای این پارامتر را ۱۰ به دست آوردند. با توجه به اینکه در داده‌ها، خلا آماری وجود نداشت از این‌رو حالت گمشده در الگوریتم Kstar به صورت پیش‌فرض انتخاب و تغییری در آن صورت نگرفت.

در شکل (۳)، نمودار Heat map حاصل از اجرای مدل‌های

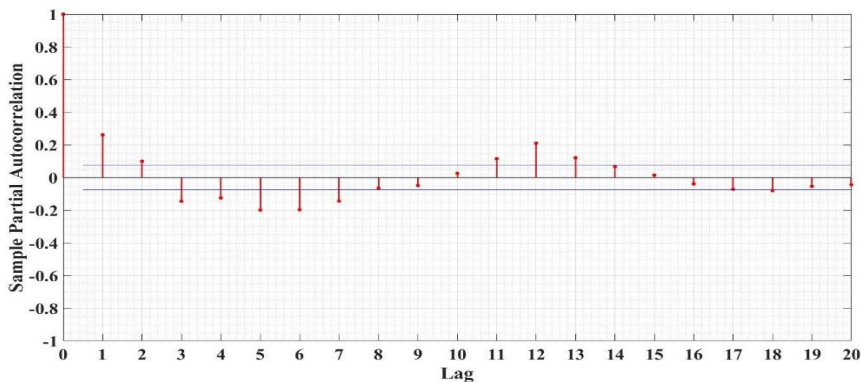
توسعه یافته و برای پدیده‌های غیرخطی و تصادفی چندان مناسب نمی‌باشد.

در شکل (۵) نمودار گرافیکی بارش‌های پیش‌بینی شده حاصل از مدل‌های منفرد GEP و Kstar ارائه شده است. با توجه این شکل مشاهده می‌شود که برآوردهای مدل GEP صرفاً در محدوده صفر تا ۶۰ میلی‌متر در حال نوسان بوده و الگوریتم Kstar نیز رفتاری مشابه داشته است.

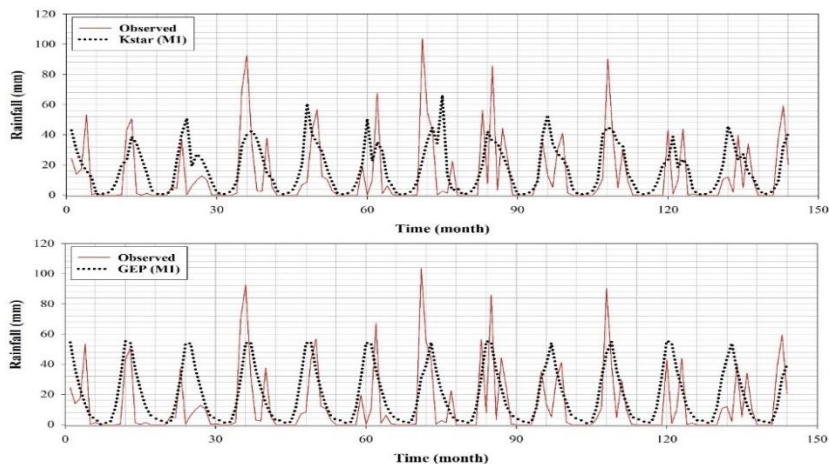
یکی از نکات مهمی که در شکل (۳) انعکاس یافته نحوه تغییرات شاخص CC می‌باشد. این شاخص بین ۰/۳۶ تا ۰/۵۴ در حال نوسان بوده اما در مقابل شاخص NS که نشان دهنده کارایی مدل‌ها می‌باشد بین ۰/۱۴- تا ۰/۱۴+ تغییر می‌کند. به عبارت دیگر ضریب همبستگی عملکرد مدل را تا حدودی مناسب ارزیابی می‌کند اما شاخص NS آن را بسیار ضعیف نشان می‌دهد. دو شاخص KGW و WI نیز تایید کننده این استنباط می‌باشد. شاخص CC براساس فرضیه نرمال بودن داده‌ها و خطی بودن آنها



شکل ۳- نمودار Heat map شاخص‌های ارزیابی عملکرد مدل‌های منفرد GEP و Kstar در مرحله آزمون



شکل ۴- نمودار خودهمبستگی جزئی سری بارش‌های ماهانه ایستگاه سینوپتیک اهواز



شکل ۵- نمودار گرافیکی مقادیر مشاهده‌ای و تخمینی توسط مدل‌های منفرد در مرحله آزمون



به عبارت دیگر عملکرد مدل GEP از ناحیه بسیار ضعیف به حالت عمل کرد قابل قبول رسیده است. در سایر شاخص‌های مورد بررسی نیز روند مشابهی وجود دارد که تایید کننده بهبود تطابق بین داده‌های مشاهداتی و محاسباتی حاصل از مدل GEP می باشد. نکته مهم دیگری نیز که باید مورد توجه قرار گیرد تاثیر سطح تجزیه در بهبود عملکرد مدل می باشد. با افزایش سطح تجزیه در کلیه مدل‌های مورد بررسی، شاخص‌ها از نظر عددی بهبود پیدا کرده‌اند. این مهم نشان دهنده تاثیر انتخاب سطح تجزیه مناسب در فرآیند مدل‌سازی می باشد.

با توجه به شکل (۶) مشاهده می شود که موجک‌ها ر کمترین دقت را در برآورد بارش‌های ماهانه داشته است. موجک‌ها ر ساختاری ساده و گسسته داشته و به همین دلیل با داده‌های بارش ماهانه که حالتی پیچیده و تصادفی دارند نتوانسته سازگاری مناسبی داشته باشد. سایر توابع موجک مورد استفاده به دلیل توانایی بالا در تطابق با پدیده‌های پیچیده نتایجی تقریباً نزدیک به هم را ارائه داده‌اند. بهترین عملکرد مدل WGEP با الگوی W(db4)(2) (موجک دابچیز چهار، با سطح تجزیه دو) به دست آمده است.

W(haar)(1)GEP	0.6035	0.2612	0.5006	0.632	1
W(haar)(2)GEP	0.6405	0.3896	0.4817	0.6255	0.9
W(db2)(1)GEP	0.7199	0.508	0.5483	0.67	0.8
W(db2)(2)GEP	0.7417	0.5371	0.5345	0.6639	0.7
W(db4)(1)GEP	0.6362	0.2821	0.5341	0.6191	0.6
W(db4)(2)GEP	0.7861	0.6068	0.6482	0.715	0.5
W(sym)(1)GEP	0.6904	0.4497	0.5949	0.6576	0.4
W(sym)(2)GEP	0.677	0.351	0.5612	0.6218	0.3
W(coiff)(1)GEP	0.7461	0.5175	0.6366	0.665	0.2
W(coiff)(2)GEP	0.716	0.4834	0.599	0.6398	0.1
W(F-K)(1)GEP	0.6983	0.442	0.5407	0.6436	0
W(F-K)(2)GEP	0.6622	0.4197	0.4168	0.6442	0
	CC	NS	KGW	WI	

شکل ۶- نمودار Heat map شاخص‌های ارزیابی عملکرد مدل WGEP در مرحله آزمون

داشته است. با مقایسه شکل‌های (۶ و ۷) مشاهده می شود که روش WKstar نسبت به مدل WGEP از شاخص NS بهتری در سطوح تجزیه دو گانه برخوردار می باشد و فقط در سطح تجزیه یک موجک F-K نتوانسته عملکردی هم‌رده با WGEP ارائه دهد. این مهم نشان می دهد که مدل Kstar با داده‌های دریافتی مشابه می تواند نتایج قابل قبول تری را ارائه دهد.

در شکل (۸) نمودارهای گرافیکی و پراکنندگی حاصل از مدل‌های WKstar و WGEP در بهترین حالت ورودی با موجک دابچیز چهار و سطح تجزیه دو ارائه شده است. با توجه به این شکل مشاهده می شود که مدل‌های هیبریدی (در مقایسه با شکل

## مدل‌سازی بارش ماهانه با استفاده از رویکرد مدل‌های هیبریدی WKstar و WGEP

به منظور کاربرد مدل‌های هیبریدی، در ابتدا می باید سری زمانی بارش ایستگاه سینوپتیک اهواز با توجه به موجک‌های مادر انتخابی و سطح تجزیه مناسب به زیرسری‌های تقریب و جزئیات شکسته شوند. برای محاسبه سطح تجزیه مناسب از رابطه (۱۲) استفاده شده و برابر با دو سطح  $(L=int(\log(695))=2)$ ، به دست آمد. در مرحله بعد، ورودی‌های بهینه بایستی تعیین گردد. با توجه به نتایج به دست آمده در قسمت قبل مشاهده می شود که الگوی M1 بهینه‌ترین عملکرد را برای مدل‌های منفرد فراهم می کند از این رو صرفاً داده‌های M1 برای تجزیه با پنج موجک مادر مورد استفاده قرار گرفت. نتایج حاصل از اجرای مدل‌های WGEP و WKStar به ترتیب در شکل‌های (۶ و ۷) ارائه شده است.

با توجه به شکل (۶) مشاهده می شود که استفاده از تجزیه موجکی عملکرد مدل GEP را به طور قابل ملاحظه‌ای افزایش داده است. به عنوان مثال در حالت منفرد و الگوی M1، معیار NS مقداری برابر با ۰/۰۹۸ به خود اختصاص داده بود اما در حالت هیبریدی، بیشترین مقدار معیار NS معادل ۰/۶۰۶ به دست آمد.

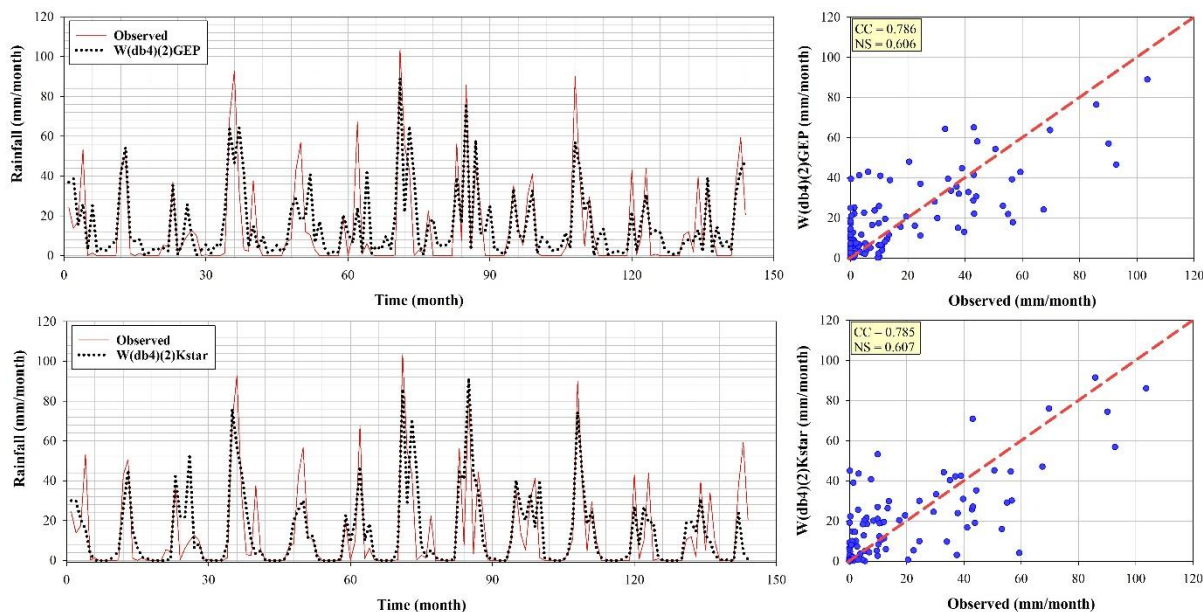
در شکل (۷) نتایج حاصل از اجرای مدل هیبریدی WKstar ارائه شده است. این شکل روندی مطابق با شکل (۶) داشته و عملکرد مدل منفرد Kstar پس از ترکیب با توابع موجک به صورت چشم‌گیری بهبود یافته است و همه موجک‌ها با سطح تجزیه دو نتوانسته‌اند بهترین اطلاعات را در اختیار مدل Kstar قرار دهند. همچنین موجک‌های هار و کویفلت با سطح تجزیه یک به بیشترین دقت دست یافته‌اند. این نتیجه با یافته‌های Solgi and Golabi (2017) مطابقت دارد. بهترین عملکرد مدل WKstar با موجک دابچیز چهار و سطح تجزیه دو بوده است که بیشترین دقت را در بین کلیه مدل‌های اجرا شده اعم از منفرد و هیبریدی

بتواند توزیع داده‌ها را نیز مورد بررسی قرار دهد. در این مطالعه برای بررسی توزیع داده‌های تولیدی توسط مدل‌های منفرد و هیبریدی از نمودار ویولونی استفاده شد (شکل ۹).

توانسته‌اند مقادیر کوچک‌تر را با دقت بهتری برآورد نمایند. همچنین برآورد مقادیر حداکثری نیز بسیار بهبود یافته و با وجود اینکه شاخص‌های ارزیابی هر دو مدل عملاً تفاوت معنی‌داری بایکدیگر ندارند اما عملکرد مدل WKstar کمی بهتر به نظر می‌رسد. برای بررسی بیشتر این موضوع نیاز به نموداری است که

W(haar)(1)Kstar	0.6765	0.4443	0.5834	0.7224
W(haar)(2)Kstar	0.6595	0.4059	0.6011	0.6992
W(db2)(1)Kstar	0.7283	0.5263	0.6234	0.7278
W(db2)(2)Kstar	0.7639	0.582	0.6577	0.7451
W(db4)(1)Kstar	0.7661	0.5794	0.7037	0.7428
W(db4)(2)Kstar	0.7849	0.6075	0.7372	0.753
W(sym)(1)Kstar	0.7283	0.5263	0.6234	0.7278
W(sym)(2)Kstar	0.7655	0.5733	0.724	0.7425
W(coiff)(1)Kstar	0.7583	0.5587	0.7175	0.7271
W(coiff)(2)Kstar	0.7319	0.5083	0.7084	0.7162
W(F-K)(1)Kstar	0.6357	0.2876	0.6183	0.7115
W(F-K)(2)Kstar	0.7327	0.4964	0.7214	0.7534
	CC	NS	KGW	WI

شکل ۷- نمودار Heat map شاخص‌های ارزیابی عملکرد مدل WGEP در مرحله آزمون



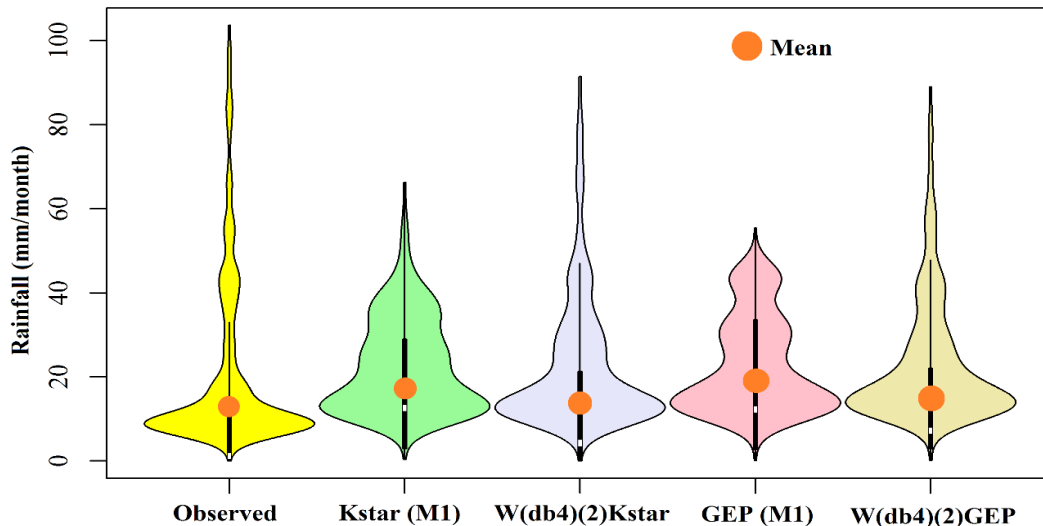
شکل ۸- نمودارهای گرافیکی و پراکنندگی مدل‌های هیبریدی در مرحله آزمون

استفاده می‌شود. در شکل (۹) نمودار ویولونی برای حالت بهینه مدل‌های منفرد و هیبریدی ارائه شده است. با توجه به این شکل مشاهده می‌شود که مدل‌های منفرد، داده‌ها را در مرحله آزمون بیش از مقدار مشاهداتی برآورد نموده‌اند و به همین دلیل میانگین بیشتری نسبت به داده‌های مشاهداتی دارند. همچنین مدل‌های منفرد نتوانسته‌اند مقادیر بیشینه را به خوبی برآورد نمایند. شایان ذکر است که مدل Kstar در برآورد بارش‌های حدی تاحدودی بهتر از مدل GEP بوده است. با بررسی ویولون‌های WKstar و WGEP مشاهده می‌شود که بیش‌برآورد این دو مدل کمتر بوده

همانگونه که ذکر گردید نتایج مدل‌های WKstar و WGEP در پیش‌بینی بارش ماهانه ایستگاه سینوپتیک اهواز، براساس معیارهای ارزیابی بسیار نزدیک به هم می‌باشد. به عبارت دیگر شاخص‌های ارزیابی آماری نمی‌توانند هیچ اطلاعاتی از نحوه توزیع داده‌ها در اختیار تحلیل‌گران قرار دهند. از این‌رو استفاده از نمودار ویولونی می‌تواند بسیار سودمند باشد. نمودار ویولونی حالت دیگری از نمودار جعبه‌ای می‌باشد. در نمودار جعبه‌ای صرفاً مینیمم، ماکزیمم، میانگین و چارک‌های داده‌ها ارائه می‌شود اما نمودار ویولونی برای بصری‌سازی توزیع داده‌ها و تراکم احتمالی آن

اختصاص داده‌اند در مدل W(db4)(2)Kstar به مقادیر مشاهداتی نزدیک‌تر می‌باشد. بنابراین با عنایت به توضیحات ارائه شده می‌توان نتیجه گرفت که مدل W(db4)(2)Kstar مناسب‌ترین روش برای پیش‌بینی بارش‌های ماهانه ایستگاه سینوپتیک اهواز می‌باشد.

و توانسته‌اند میانگین مقادیر تخمینی را نزدیک به واقعیت حفظ نمایند. همچنین برآورد بارش‌های حداکثری نیز در هر دو مدل بسیار بهبود یافته اما روش W(db4)(2)Kstar در این زمینه به مراتب موفق‌تر عمل کرده است. همچنین بیشترین مقادیر بارش که حول میانگین تغییر می‌کنند (بیشترین عرض) را به خود



شکل ۹- نمودار ویولونی مدل‌های بهینه منفرد و هیبریدی برای پیش‌بینی بارش‌های ماهانه در مرحله آزمون

موجک‌ها بود. همچنین موجک F-K که کمتر در مطالعات هیدرولوژی به کار گرفته شده بود، نیز عملکردی متوسط داشته است. بهترین کارایی مدل‌های هیبریدی WKstar و WGEP با استفاده از موجک دابچیز ۴ و سطح تجزیه دو به دست آمد. مقدار آماره NS برای هر کدام از مدل‌های هیبریدی به ترتیب برابر با ۰/۶۰۷ و ۰/۶۰۸ محاسبه گردید. با توجه به معیارهای ارزیابی مورد استفاده عملاً تفاوت معنی‌داری بین مدل‌های هیبریدی مشاهده نمی‌شود اما نمودار ویولون، به خوبی اختلاف موجود را مشخص نمود و در نهایت روش W(db4)(2)Kstar به عنوان روش منتخب در پیش‌بینی بارش‌های ماهانه ایستگاه سینوپتیک اهواز پیشنهاد گردید.

### سپاس‌گزاری

بدین وسیله از حمایت مالی معاونت پژوهش و فناوری دانشگاه شهید چمران اهواز در قالب پژوهانه (GN:SCU.WH99.44131) در انجام این تحقیق تشکر و قدردانی می‌گردد.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

### نتیجه‌گیری

در این مطالعه از الگوریتم Kstar، مدل GEP و تلفیق آنها با توابع موجک برای پیش‌بینی بارش‌های ماهانه ایستگاه سینوپتیک اهواز استفاده شده و دقت آنها مورد ارزیابی قرار گرفت. نتایج حاصل نشان داد که مدل‌های منفرد Kstar و GEP در برآورد بارش‌های ماهانه از کارایی لازم برخوردار نیستند به طوری که از نظر عملکرد در وضعیت بسیار ضعیف قرار می‌گیرند. هر دو مدل‌های منفرد با الگوی M1 به بهترین کارایی خود دست یافته و مقدار آماره NS برای روش Kstar برابر با ۰/۱۳۹ و برای مدل GEP معادل با ۰/۰۰۹ به دست آمد. به منظور بهبود دقت مدل‌های مورد بررسی از یک روش ریاضی پیچیده مبتنی بر تجزیه سیگنال بهره گرفته شد. بدین منظور از موجک‌های هار، دابچیز دو، دابچیز چهار، سیملت، کویفلت و F-K استفاده شد. نتایج حاصل نشان داد که با تجزیه سری‌های زمانی بارش، اطلاعات مورد نیاز مدل‌ها با جزئیات بیشتری در اختیار آنها قرار گرفته و از این‌رو دقت در مدل‌سازی به طور قابل ملاحظه‌ای بهبود یافت. به طوریکه تخمین‌های انجام شده از حالت بسیار ضعیف به وضعیت مناسب رسیدند. در این میان ضعیف‌ترین عملکرد موجک‌ها مربوط به

### REFERENCES

Abdourahamane, Z. S., Acar, R., & Serkan, Ş. (2019). Wavelet-copula-based mutual information for rainfall forecasting applications. Hydrological

Processes, 33(7), 1127-1142.

Ahmadi, F. (2020). Evaluation of Support Vector Machine and Adaptive Neuro-Fuzzy Inference

- System Performance in Prediction of Monthly River Flow (Case Study: Nazlu chai and Sezar Rivers). *Iranian Journal of Soil and Water Research*, 51(3), 673-686. (In Farsi)
- Ahmadi, F., and Valinia, M. (2020). Prediction of Monthly River Flow Using Hybridization of Linear Time Series Models and Bayesian network (Case Study: Bakhtiari River). *Water and Irrigation Management*, 10(2), 233-245. (In Farsi)
- Bushara, N. O. (2019). Weather forecasting using soft computing models: A comparative study. *Journal of Applied Science*, 2 (2): 1-22.
- Costache, R., Pham, Q. B., Sharifi, E., Linh, N. T. T., Abba, S. I., Vojtek, M., ... & Khoi, D. N. (2020). Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote Sensing*, 12(1), 106.
- Ekmekcioğlu, Ö., Başakın, E. E., & Özger, M. (2020). Tree-based nonlinear ensemble technique to predict energy dissipation in stepped spillways. *European Journal of Environmental and Civil Engineering*, 1-19.
- Estévez, J., Bellido-Jiménez, J. A., Liu, X., & García-Marín, A. P. (2020). Monthly Precipitation Forecasts Using Wavelet Neural Networks Models in a Semiarid Environment. *Water*, 12(7), 1909.
- Ferreira, C. (2002). Genetic representation and genetic neutrality in gene expression programming. *Advances in Complex Systems*, 5(04), 389-408.
- Freire, P. K. D. M. M., Santos, C. A. G., and da Silva, G. B. L. (2019). Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting. *Applied Soft Computing*, 80: 494-505.
- Granata, F., Di Nunno, F., Gargano, R., & de Marinis, G. (2019). Equivalent discharge coefficient of side weirs in circular channel- a lazy machine learning approach. *Water*, 11(11), 2406.
- Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection* (Vol. 1). MIT press.
- Mallat, S. G. 1998. *A wavelet tour of signal processing*, San Diego. Grossmann, A., & Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15:4. 723-736.
- Mehdizadeh, S., Ahmadi, F., Mehr, A. D., & Safari, M. J. S. (2020). Drought modeling using classic time series and hybrid wavelet-gene expression programming models. *Journal of Hydrology*, 125017.
- Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2018). New approaches for estimation of monthly rainfall based on GEP-ARCH and ANN-ARCH hybrid models. *Water resources management*, 32(2), 527-545.
- Mehr, A. D. (2018). Month ahead rainfall forecasting using gene expression programming. *American Journal of Earth and Environmental*, 667, 63-70.
- Mehr, A. D., Nourani, V., Khosrowshahi, V. K., & Ghorbani, M. A. (2019). A hybrid support vector regression-firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology*, 16(1), 335-346.
- Mirabbasi, R., Kisi, O., Sanikhani, H., & Meshram, S. G. (2019). Monthly long-term rainfall estimation in Central India using M5Tree, MARS, LSSVR, ANN and GEP models. *Neural Computing and Applications*, 31(10), 6843-6862.
- Mohammadi, B., Ahmadi, F., Mehdizadeh, S., Guan, Y., Pham, Q. B., Linh, N. T. T., & Tri, D. Q. (2020). Developing novel robust models to improve the accuracy of daily streamflow modeling. *Water Resources Management*, 34(10), 3387-3409.
- Phillies, G. D. J., Gould, H., and Tobochnik, J. (1996). Wavelets: a new alternative to Fourier transforms. *Comput. Phys.* 10, 247-252.
- Polikar, R. (1996). *Fundamental concepts and an overview of the wavelet theory*. Second Edition, Rowan University, College of Engineering Web Servers, Glassboro. NJ. 08028.
- Polikar, R., & Mastorakis, N. (1999). *The story of wavelets in Physics and Modern Topics in Mechanical and Electrical Engineering*. World Scientific and Engineering Society Press, 192-197.
- Santos, C. A., Freire, P. K., Silva, R. M. D., & Akrami, S. A. (2019). Hybrid wavelet neural network approach for daily inflow forecasting using tropical rainfall measuring mission data. *Journal of Hydrologic Engineering*, 24(2), 04018062.
- Solgi, A., golabi, M. (2017). Performance Assessment of Gene Expression Programming Model Using Data Preprocessing Methods to Modeling River Flow. *Journal of Water and Soil Conservation*, 24(2), 185-201. (In Farsi)
- Sun, Y., Niu, J., & Sivakumar, B. (2019). A comparative study of models for short-term streamflow forecasting with emphasis on wavelet-based approach. *Stochastic Environmental Research and Risk Assessment*, 33(10), 1875-1891.
- Wang, W., & Ding, J. (2003). Wavelet network model and its application to the prediction of hydrology. *Nature and Science*, 1(1), 67-71.
- Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32(13), 2088-2094.